Appendix

EC.1. Proofs for Section 4

We need several results from convex analysis to prove Lemma 1. For any convex function g on $\mathbb{R}$, let dom $g = \{x \in \mathbb{R} : g(x) < \infty\}$ be its effective domain. The following theorems are from Rockafellar (1997), specialized to convex functions g with dom $g = \mathbb{R}$. The definitions of a proper convex function and a closed convex function can be found on p.24 and p.52 therein respectively.

THEOREM EC.1 **(a.k.a. Rockafellar (1997), Corollary 10.1.1)**. *A convex function finite on all of $\mathbb{R}$ is necessarily continuous.*

THEOREM EC.2 **(a.k.a. Rockafellar (1997), Theorem 24.1)**. *Let g be a closed proper convex function on $\mathbb{R}$, such that dom $g = \mathbb{R}$. Then g'_+ exists and is a finite non-decreasing function on $\mathbb{R}$. Moreover, g'_+ is right-continuous, i.e., $\lim_{z \searrow x} g'_+(z) = g'_+(x)$ for any $x \in \mathbb{R}$.*

THEOREM EC.3 **(a.k.a. Rockafellar (1997), Corollary 24.2.1)**. *Let g be a finite convex function on a non-empty open real interval I. Then*

$$g(y) - g(x) = \int_x^y g'_+(t)dt$$

for any x and y in I.

THEOREM EC.4 **(a.k.a. Rockafellar (1997), Theorem 24.2)**. *Let φ be a non-decreasing function from $\mathbb{R}$ to $[-\infty, \infty]$ such that $\varphi(b)$ is finite for some $b \in \mathbb{R}$. Then the function given by*

$$g(x) = \int_b^x \varphi(t)dt$$

is a well-defined closed proper convex function on $\mathbb{R}$.

Proof of Lemma 1. Throughout this proof, without loss of generality let $a = 0$ (by replacing $f(x)$ with $f(x+a)$, and $h(x)$ with $h(x+a)$ respectively). Note that optimizations (1) and (2) do not depend on $f(x)$ for $x < 0$. For the purpose of applying Theorems EC.1–EC.4 more directly, let us extend f to $\mathbb{R}^-$, by defining $f(x) = \eta - \nu x$ for $x < 0$ (this extension of f is a mathematical artifact and does not necessarily match the given true density).

Let $\mathcal{F}_1$ be the feasible region in (1), and $\mathcal{F}_2$ be the feasible region in (2). We show that $\mathcal{F}_1 = \mathcal{F}_2$.

Proof of $\mathcal{F}_1 \subset \mathcal{F}_2$: Since $f(x) < \infty$ for at least one $x \in \mathbb{R}$ (e.g., take $x = 0$) and $f(x) \geq 0 > -\infty$ for all $x \in \mathbb{R}$, we get that f is proper (Rockafellar (1997), p.24).

Next, we argue that $f(x) < \infty$ for all $x \in \mathbb{R}$. Suppose on the contrary that $f(x_0) = \infty$ for some $x_0 > 0$. If $f(y) < \infty$ for some $y > x_0$, then $((y - x_0)/y)f(0) + (x_0/y)f(y) = ((y - x_0)/y)\eta + (x_0/y)f(y) < \infty = f(x_0)$, contradicting (1d). But if $f(y) = \infty$ for all $y > x_0$, then $\int_0^\infty f(t)dt = \infty$, contradicting (1a). Therefore, $f(x) < \infty$ for all $x \in \mathbb{R}$, and with (1e), we conclude that f is finite.

Since f is finite on $\mathbb{R}$, Theorem EC.1 implies that f is continuous and hence lower semi-continuous. Since f is proper, lower semi-continuity is the same as closedness (Rockafellar (1997), p.52). Hence f is closed.

Therefore, together with the convexity condition in (1d), Theorem EC.2 implies the existence of f'_+ that satisfies (2c). Moreover, Theorem EC.3 implies (2f).

Next, with the monotonicity of f'_+ by (2c), we have $f'_+(x) \geq f'_+(0) = -\nu$ for all $x \geq 0$, thus implying the first inequality of (2d). To prove the second inequality in (2d), suppose in the contrary that $f'_+(x_0) > 0$ for some $x_0 > 0$. Since $f'_+(x) \geq f'_+(x_0) > 0$ for all $x > x_0$ by (2c), we have, from (2f), $f(x) = \int_0^x f'_+(t)dt + \eta \to \infty$, implying that $\int_0^\infty f(x)dx = \infty$ and contradicting (1a). Hence the second inequality in (2d) holds. We have therefore shown (2d).

Lastly, suppose that $f'_+(x) \not\to 0$. Then, since (2d) holds, there exists a sequence $x_k \to \infty$ such that $f'_+(x_k) \to c < 0$. But since f'_+ is monotone by (2c), $\lim_{x\to\infty} f'_+(x)$ exists and must equal c. But then, from (2f), $f(x) = \int_0^x f'_+(t)dt + \eta \to -\infty$, violating (1e). Thus (2e) holds.

The constraints (2a) and (2b) follow immediately from (1a) and (1b). We therefore conclude that $\mathcal{F}_1 \subset \mathcal{F}_2$.

Proof of $\mathcal{F}_2 \subset \mathcal{F}_1$: Since f'_+ is bounded on $\mathbb{R}$ by (2d), Theorem EC.4 and (2c) (with $f'_+(x)$ defined as $-\nu$ for $x < 0$) implies that the f defined by (2f) is convex on $\mathbb{R}$, giving (1d).

Suppose $f(x_0) < 0$ for some $x_0 > 0$. Then, since $f'_+ \leq 0$ by (2d), (2f) implies $f(x) < 0$ for all $x \geq x_0$. Thus $\int_0^\infty f(x)dx = -\infty$, contradicting (2a). Therefore, (1e) holds.

The constraint (2d) implies (1c) immediately. The condition (2f) implies $f(0) = f(0+)$. Thus, combining with (2b), we get that (1b) holds. Finally, note that (2a) is the same as (1a). We conclude that $\mathcal{F}_2 \subset \mathcal{F}_1$. □

To prove Theorem 2, we first introduce the following lemma:

Lemma EC.1. *If f is a feasible solution of (1), equivalently (2), then $xf(x) \to 0$ and $x^2 f'_+(x) \to 0$ as $x \to \infty$.*

Proof of Lemma EC.1. We need the observations that $f(x)$ is non-increasing by (2d), $f(x) \geq 0$ for all $x \geq a$ by (1e), and that f is integrable on $[a,\infty)$ with $\int_a^\infty f(x)dx = \beta$ by (1a). Denote $F(x) = \int_a^x f(t)dt$ and $g(x) = xf(x) - F(x)$. Consider, for $a \vee 0 \leq x_1 \leq x_2$,

$$\begin{aligned} g(x_2) - g(x_1) &= x_2 f(x_2) - x_1 f(x_1) - (F(x_2) - F(x_1)) \\ &\leq x_2 f(x_2) - x_1 f(x_1) - f(x_2)(x_2 - x_1) \quad \text{since } f(x) \text{ is non-increasing} \\ &= x_1[f(x_2) - f(x_1)] \\ &\leq 0 \quad \text{again since } f \text{ is non-increasing} \end{aligned}$$

Therefore g is non-increasing for $x \geq a \vee 0$, and since $xf(x) \geq 0$ and $0 \leq F(x) \leq \beta$ for $x \geq a \vee 0$, we have g bounded from below on the same range. This implies that g must converge to a limit, say c, as $x \to \infty$. In other words, $xf(x) - F(x) \to c$, and since $F(x) \to \beta$, we have $xf(x) \to c + \beta$. Since $xf(x) \geq 0$ for $x \geq a \vee 0$, there are two cases: $c + \beta > 0$ or $c + \beta = 0$. The first case implies that $xf(x) \geq \epsilon > 0$ for some ϵ for all large enough x. This means $f(x) \geq \epsilon/x$ for all large enough x, and hence $\int_a^\infty f(x)dx = \infty$, which contradicts (1a). Therefore $xf(x)$ must converge to 0. This proves the first part of the lemma.

To prove the second part, we need the observation that $f'_+(x)$ is non-decreasing for $x \geq a$ by (2c), and is non-positive for $x \geq a$ by (2d). Also, by (2f) we have $f(x) = \int_a^x f'_+(t)dt + \eta$ for $x \geq a$. Let $\bar{F}(x) = \int_x^\infty f(t)dt$ for $x \geq a$, which is finite and converges to 0 by (1a). We now define $\tilde{g}(x) =$

$-x^2 f'_+(x) + 2\tilde{F}(x)$, where $\tilde{F}(x) = -\int_x^\infty t f'_+(t)dt$, for $x \geq a$. Note that $x f'_+(x)$ is integrable on $[a, \infty)$ because the absolute continuity of f, and $\lim_{x\to\infty} x f(x) \to 0$ as we have just proved, which allows integration by parts yielding

$$\tilde{F}(x) = -\int_x^\infty t f'_+(t)dt = -tf(t)|_x^\infty + \int_x^\infty f(t)dt = xf(x) + \bar{F}(x) < \infty \tag{EC.1}$$

For any $(a \vee 0) \leq x_1 \leq x_2$,

$$\begin{aligned}
\tilde{g}(x_2) - \tilde{g}(x_1) &= x_1^2 f'_+(x_1) - x_2^2 f'_+(x_2) - 2\tilde{F}(x_1) + 2\tilde{F}(x_2) \\
&\leq x_1^2 f'_+(x_1) - x_2^2 f'_+(x_2) + f'_+(x_2)(x_2^2 - x_1^2) \quad \text{since } f'_+(x) \text{ is non-decreasing} \\
&= x_1^2 (f'_+(x_1) - f'_+(x_2)) \\
&\leq 0 \quad \text{again since } f'_+(x) \text{ is non-decreasing}
\end{aligned}$$

Therefore, $\tilde{g}(x)$ is non-increasing for $x \geq a$. Note that $-x^2 f'_+(x) \geq 0$ for $x \geq a$. Also, from (EC.1), since $\lim_{x\to\infty} x f(x) \to 0$ and $\bar{F}(x) \to 0$, we have $\tilde{F}(x) \to 0$ as $x \to \infty$ and hence also bounded for large enough x. Therefore $\tilde{g}$ is bounded from below. This implies that $\tilde{g}$ must converge to a limit, say $\tilde{c}$, as $x \to \infty$. Since $\tilde{F}(x) \to 0$, we have $-x^2 f'_+(x) \to \tilde{c}$. Since $-x^2 f'_+(x) \geq 0$ for $x \geq a$, there are two cases: either $\tilde{c} > 0$ or $\tilde{c} = 0$. The former case implies that $-x f'_+(x) \geq \epsilon / x$ for some $\epsilon > 0$ and large enough x, and so $\tilde{F}(x) = -\int_x^\infty x f'_+(x)dx = \infty$ for $x \geq a$, which contradicts (EC.1). Therefore $-x^2 f'_+(x) \to 0$. This proves the second part of the lemma. □

Proof of Theorem 2. Throughout this proof, without loss of generality let $a = 0$. By Lemma EC.1, we can introduce the extra conditions $xf(x) \to 0$ and $x^2 f'_+(x) \to 0$ as $x \to \infty$ into formulation (2). In other words, formulation (2) is equivalent to (letting $a = 0$)

$$\begin{aligned}
\max_f \quad & \int_0^\infty h(x) f(x) dx \\
\text{subject to} \quad & \int_0^\infty f(x)dx = \beta && \text{(EC.2a)} \\
& f(0) = \eta && \text{(EC.2b)} \\
& f'_+(x) \text{ exists and is non-decreasing and right-continuous for } x \geq 0 && \text{(EC.2c)}
\end{aligned}$$

$$-\nu \le f'_+(x) \le 0 \text{ for } x \ge 0 \tag{EC.2d}$$

$$f'_+(x) \to 0 \text{ as } x \to \infty \tag{EC.2e}$$

$$f(x) = \int_0^x f'_+(t)dt + \eta \text{ for } x \ge 0 \tag{EC.2f}$$

$$xf(x) \to 0 \text{ and } x^2 f'_+(x) \to 0 \text{ as } x \to \infty \tag{EC.2g}$$

For convenience, we let $\tilde{H}(x) = \int_0^x h(u)du$ and $H(x) = \int_0^x \tilde{H}(u)du$. Consider the objective function of (EC.2). Since $\tilde{H}$ is continuous and f is absolutely continuous with $f(x) = \int_0^x f'_+(t)dt + \eta$ by (EC.2f), we have, using integration by parts,

$$\int_0^\infty h(x)f(x)dx = \tilde{H}(x)f(x)\Big|_0^\infty - \int_0^\infty \tilde{H}(x)f'_+(x)dx = -\int_0^\infty \tilde{H}(x)f'_+(x)dx \tag{EC.3}$$

where the second equality follows from (EC.2g) and that $\tilde{H}(x) = O(x)$ as $x \to \infty$ since h is bounded. As H is continuous and f'_+ has bounded variation by (EC.2d) and (EC.2c), we have, using integration by parts again, that (EC.3) is equal to

$$-H(x)f'_+(x)\Big|_0^\infty + \int_0^\infty H(x)df'_+(x) = \int_0^\infty H(x)df'_+(x) \tag{EC.4}$$

where the equality follows from (EC.2g) and that $H(x) = O(x^2)$ as $x \to \infty$ since h is bounded.

For (EC.2a), we can write

$$\int_0^\infty f(x)dx = \int_0^\infty \frac{x^2}{2} df'_+(x) \tag{EC.5}$$

by merely viewing $h \equiv 1$ in (EC.3) and (EC.4). Also, since $f(x) \to 0$ as $x \to \infty$ by (EC.2g), we can use integration by parts again to write

$$f(0) = -\int_0^\infty f'_+(x)dx = -xf'_+(x)\Big|_0^\infty + \int_0^\infty x df'_+(x) = \int_0^\infty x df'_+(x) \tag{EC.6}$$

where the third equality follows from (EC.2g) again. Therefore, (EC.2) can be written as

$$\max_f \quad \int_0^\infty H(x)df'_+(x)$$

$$\text{subject to} \quad \int_0^\infty \frac{x^2}{2} df'_+(x) = \beta \tag{EC.7a}$$

$$\int_0^\infty x df'_+(x) = \eta \tag{EC.7b}$$

$$f'_+(x) \text{ exists and is non-decreasing and right-continuous for } x \geq 0 \quad \text{(EC.7c)}$$

$$-\nu \leq f'_+(x) \leq 0 \text{ for } x \geq 0 \quad \text{(EC.7d)}$$

$$f'_+(x) \to 0 \text{ as } x \to \infty \quad \text{(EC.7e)}$$

$$f(x) = \int_0^x f'_+(t)dt + \eta \text{ for } x \geq 0 \quad \text{(EC.7f)}$$

$$xf(x) \to 0 \text{ and } x^2 f'_+(x) \to 0 \text{ as } x \to \infty \quad \text{(EC.7g)}$$

We show that (EC.7g) is redundant. By (EC.7a), we have $\int_0^\infty (x^2/2)df'_+(x) < \infty$ and hence $\int_x^\infty (t^2/2)df'_+(t) \to 0$ as $x \to \infty$. Now, for $x \geq 0$, we have

$$\int_x^\infty \frac{t^2}{2} df'_+(t) \geq \frac{x^2}{2} \int_x^\infty df'_+(t) = -\frac{x^2}{2} f'_+(x) \geq 0$$

where the first inequality follows since $f'_+(x)$ is non-decreasing by (EC.7c), the equality follows from (EC.7e), and the last inequality from (EC.7d). Therefore, $-(x^2/2)f'_+(x) \to 0$ as $x \to \infty$. This shows that the second part of (EC.7g) is redundant.

By (EC.7b), and since $f'_+(x)$ is monotone, we can use integration by parts to get

$$\eta = \int_0^\infty x df'_+(x) = xf'_+(x)\Big|_0^\infty - \int_0^\infty f'_+(x)dx = -\int_0^\infty f'_+(x)dx \quad \text{(EC.8)}$$

where the last equality follows since we have proved $-(x^2/2)f'_+(x) \to 0$ and so $xf'_+(x) \to 0$ as $x \to \infty$. Now, using (EC.7f) and (EC.8), we write

$$f(x) = \int_0^x f'_+(t)dt + \eta = \int_0^x f'_+(t)dt - \int_0^\infty f'_+(t)dt = -\int_x^\infty f'_+(t)dt \quad \text{(EC.9)}$$

Since $-(x^2/2)f'_+(x) \to 0$ as $x \to \infty$, we have $f'_+(x) = o(1/x^2)$. So $-\int_x^\infty f'_+(t)dt = o(1/x)$. Then (EC.9) implies the first part of (EC.7g) is redundant.

Therefore, (EC.7) can be written as

$$\begin{aligned} \max_f \quad & \int_0^\infty H(x)df'_+(x) \\ \text{subject to} \quad & \int_0^\infty \tfrac{x^2}{2} df'_+(x) = \beta \\ & \int_0^\infty x df'_+(x) = \eta \\ & f'_+(x) \text{ exists and is non-decreasing and right-continuous for } x \geq 0 \\ & -\nu \leq f'_+(x) \leq 0 \text{ for } x \geq 0 \\ & f'_+(x) \to 0 \text{ as } x \to \infty \end{aligned} \quad \text{(EC.10)}$$

and the constraint (EC.7f) in (EC.2) states that f can be recovered from $f(x)=\int_0^x f'_+(t)dt+\eta$. Note that this definition of f must necessarily have a right derivative coinciding with the obtained $f'_+(x)$.

Finally, let $p(x)=f'_+(x)/\nu+1$. Then (EC.10) can be rewritten as

$$\begin{aligned}\max_p \quad & \nu\int_0^\infty H(x)dp(x)\\ \text{subject to} \quad & \int_0^\infty x^2dp(x)=\tfrac{2\beta}{\nu}\\ & \int_0^\infty xdp(x)=\tfrac{\eta}{\nu}\\ & p(x) \text{ non-decreasing and right-continuous for } x\geq 0\\ & 0\leq p(x)\leq 1 \text{ for } x\geq 0\\ & p(x)\to 1 \text{ as } x\to\infty\end{aligned} \tag{EC.11}$$

or equivalently

$$\begin{aligned}\max_p \quad & \nu\int_{-\infty}^\infty H(x)dp(x)\\ \text{subject to} \quad & \int_{-\infty}^\infty x^2dp(x)=\tfrac{2\beta}{\nu}\\ & \int_{-\infty}^\infty xdp(x)=\tfrac{\eta}{\nu}\\ & p(x) \text{ non-decreasing and right-continuous for } x\in\mathbb{R}\\ & 0\leq p(x)\leq 1 \text{ for } x\in\mathbb{R}\\ & p(x)\to 1 \text{ as } x\to\infty\\ & p(x)=0 \text{ for } x<0\end{aligned} \tag{EC.12}$$

since $H(x)=x=x^2=0$ at $x=0$. One can uniquely identify, up to measure zero, a non-decreasing, right-continuous p such that $\lim_{x\to\infty}p(x)=1$ and $p(x)=0$ for $x<0$ with a probability measure supported on $\mathbb{R}^+$. Hence (EC.12) is equivalent to (5). This concludes the result. □

To prove Theorem 3, we need several results from Winkler (1988) stated below.

THEOREM EC.5 **(Winkler (1988) Theorem 2.1(b))**. *Let $\mathcal{X}$ be a measurable space with σ-field $\mathcal{F}$ and suppose that $\mathcal{P}$ is a simplex of probability measures whose extreme points are Dirac measures. Fix measurable functions $f_1,\dots,f_n$ and real numbers $c_1,\dots,c_n$. Consider the set*

$$\mathcal{H}=\left\{q\in\mathcal{P}: f_i \text{ is } q\text{-integrable and } \int f_i dq=c_i,\ 1\leq i\leq n\right\}$$

Then $\mathcal{H}$ is convex and

$$ex\ \mathcal{H} = \Bigg\{ q \in \mathcal{H} : q = \sum_{i=1}^{m} t_i \cdot \delta(x_i),\ t_i > 0,\ \sum_{i=1}^{m} t_i = 1,\ x_i \in \mathcal{X},\ 1 \le m \le n+1, \text{ the vectors } (f_1(x_i), \ldots, f_n(x_i), 1),\ 1 \le i \le m, \text{ are linearly independent} \Bigg\}$$

where ex $\mathcal{H}$ denotes the set of all extreme points of $\mathcal{H}$.

Theorem EC.6 **(Adapted from Winkler (1988) Theorem 3.2)**. *Let $\mathcal{X}$ be a Hausdorff space, $\mathcal{F}$ be the Borel σ-field and $\mathcal{P}_r(\mathcal{X})$ be the set of regular probability measures on $\mathcal{X}$. Let*

$$\mathcal{H} = \left\{ q \in \mathcal{P}_r(\mathcal{X}) : f_i \text{ is } q\text{-integrable and } \int f_i dq = c_i,\ 1 \le i \le n \right\}$$

Every measure affine functional J on $\mathcal{H}$ fulfills

$$\sup\{J(q) : q \in \mathcal{H}\} = \sup\{J(q) : q \in ex\ \mathcal{H}\}$$

Theorem EC.6 is precisely Theorem 3.2 in Winkler (1988), except replacing the inequalities with equalities for the moments that define $\mathcal{H}$, which is immediate (and is pointed out by the comment right after the theorem in Winkler (1988)).

Proposition EC.1 **(Winkler (1988) Proposition 3.1)**. *Let $\mathcal{X}$, $\mathcal{F}$ and $\mathcal{H}$ be given as in Theorem EC.6 and the function g on $\mathcal{X}$ be integrable for every $q \in \mathcal{H}$ (possibly with integral values ∞ or $-\infty$). Then the functional G on $\mathcal{H}$ defined by $G(q) = \int_{\mathcal{X}} g dq$ is measure affine.*

Proof of Theorem 3. By Examples 2.1(a) in Winkler (1988), the set $\mathcal{P}$ in Theorem EC.5 can be chosen to be the set of all regular probability measures. On Polish space every probability measure is regular. Therefore, on the space $\mathbb{R}^+$, which is Polish, we can take $\mathcal{P}$ in Theorem EC.5 as the set of all probability measures. The $\mathcal{H}$ in Theorems EC.5 and EC.6 then coincide. By Proposition EC.1, the objective $\nu\mathbb{E}[H(X)]$ in $OPT(\mathcal{P}^+)$ is measure affine. Therefore, using Theorems EC.5 and EC.6, and noting that

$$\mathcal{H} \supseteq \left\{ q \in \mathcal{H} : q = \sum_{i=1}^{m} t_i \cdot \delta(x_i),\ t_i > 0,\ \sum_{i=1}^{m} t_i = 1,\ x_i \in \mathcal{X},\ 1 \le m \le n+1 \right\} \supseteq \text{ex } \mathcal{H}$$

for the coincided $\mathcal{H}$ in Theorems EC.5 and EC.6, we conclude the theorem. □

Proof of Proposition 1. If program (5) is consistent, then by Theorem 3, either an optimal solution in $\mathcal{P}_3^+$ exists, which corresponds to the first case of the lemma, or there exists a feasible sequence $\mathbb{P}^{(k)} \in \mathcal{P}_3^+$ such that $Z(\mathbb{P}^{(k)}) \to Z^*$. Let $\mathbb{P}^{(k)} \sim (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, p_1^{(k)}, p_2^{(k)}, p_3^{(k)})$. Suppose that x_i's are all bounded above by a number, say M. Then, since $[0,M]^3 \times \mathcal{S}_3$ is a compact set, by Bolzano-Weierstrass Theorem we must have a subsequence of $(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, p_1^{(k)}, p_2^{(k)}, p_3^{(k)})$, say $(x_1^{(k_j)}, x_2^{(k_j)}, x_3^{(k_j)}, p_1^{(k_j)}, p_2^{(k_j)}, p_3^{(k_j)})$ converge to $(x_1^*, x_2^*, x_3^*, p_1^*, p_2^*, p_3^*)$ in $[0,M]^3 \times \mathcal{S}_3$. Since $H(x)$ is continuous by construction, we have $Z(\mathbb{P}^{(k_j)}) = \nu \sum_{i=1}^3 H(x_i^{(k_j)}) p_i^{(k_j)} \to \nu \sum_{i=1}^3 H(x_i^*) p_i^* = Z(\mathbb{P}^*)$, where $\mathbb{P}^* \sim (x_1^*, x_2^*, x_3^*, p_1^*, p_2^*, p_3^*)$. As $Z(\mathbb{P}^{(k_j)})$ is a subsequence of $Z(\mathbb{P}^{(k)})$, $Z(\mathbb{P}^*)$ must be equal to Z^*, and so $\mathbb{P}^*$ is an optimal solution, which reduces to the first case in the lemma. Therefore, for the second case, we should focus on the scenario that at least one $x_i^{(k)}$ satisfies $\limsup_{k\to\infty} x_i^{(k)} = \infty$.

Without loss of generality, we fix the convention that $x_1^{(k)} \le x_2^{(k)} \le x_3^{(k)}$. If at least one of $x_i^{(k)}$ satisfies $\limsup_{k\to\infty} x_i^{(k)} = \infty$, we must have $\limsup_{k\to\infty} x_3^{(k)} = \infty$. In order that $\mathbb{P}^{(k)}$ is feasible, $\mathbb{E}^{(k)}[X] = \mu$ holds and so $x_1^{(k)} \le \mu$ for all k. We now distinguish two cases: either $x_2^{(k)}$ is uniformly bounded, say by a large number $M \ge \mu$, or $\limsup_{k\to\infty} x_2^{(k)} = \infty$ also. Consider the first case. First, we find a subsequence $x_3^{(k_j)} \nearrow \infty$. Since $(x_1^{(k_j)}, x_2^{(k_j)}) \in [0,M]^2$ which is compact, we can choose a further subsequence $k_{j'}$ such that $(x_1^{(k_{j'})}, x_2^{(k_{j'})}, x_3^{(k_{j'})}) \to (x_1^*, x_2^*, \infty)$ where $(x_1^*, x_2^*) \in [0,M]^2$. Now, since $(p_1^{(k_{j'})}, p_2^{(k_{j'})}, p_3^{(k_{j'})}) \in \mathcal{S}_3$ which is also compact, we can choose another further subsequence $k_{j''}$ such that $(p_1^{(k_{j''})}, p_2^{(k_{j''})}, p_3^{(k_{j''})}) \to (p_1^*, p_2^*, p_3^*) \in \mathcal{S}_3$. Note that by the constraint $\mathbb{E}^{(k)}[X^2] = p_1^{(k_{j''})} {x_1^{(k_{j''})}}^2 + p_2^{(k_{j''})} {x_2^{(k_{j''})}}^2 + p_3^{(k_{j''})} {x_3^{(k_{j''})}}^2 = \sigma$, we must have $p_3^{(k_{j''})} = (\sigma - p_1^{(k_{j''})} {x_1^{(k_{j''})}}^2 - p_2^{(k_{j''})} {x_2^{(k_{j''})}}^2) / {x_3^{(k_{j''})}}^2 \le \sigma / {x_3^{(k_{j''})}}^2 \to 0$. In conclusion, in this case, we end up being able to find a sequence of measures $\mathbb{P}^{(k)'} \sim (x_1^{(k)'}, x_2^{(k)'}, x_3^{(k)'}, p_1^{(k)'}, p_2^{(k)'}, p_3^{(k)'})$ with $(x_1^{(k)'}, x_2^{(k)'}, x_3^{(k)'}, p_1^{(k)'}, p_2^{(k)'}, p_3^{(k)'}) \to (x_1^*, x_2^*, \infty, p_1^*, p_2^*, 0)$ where $x_1^*, x_2^* \in \mathbb{R}^+$ and $(p_1^*, p_2^*) \in \mathcal{S}_2$.

For the second case, namely when $\limsup_{k\to\infty} x_i^{(k)} = \infty$ for both $i = 2$ and 3. We can argue similarly that there is a sequence of measures $\mathbb{P}^{(k)'} \sim (x_1^{(k)'}, x_2^{(k)'}, x_3^{(k)'}, p_1^{(k)'}, p_2^{(k)'}, p_3^{(k)'})$, such that $x_2^{(k)'}, x_3^{(k)'} \to \infty$ and $p_2^{(k)'}, p_3^{(k)'} \to 0$. In other words, $(x_1^{(k)'}, x_2^{(k)'}, x_3^{(k)'}, p_1^{(k)'}, p_2^{(k)'}, p_3^{(k)'}) \to (x_1^*, \infty, \infty, 1, 0, 0)$ where $x_1^* \in \mathbb{R}^+$. □

Proof of Lemma 2. It follows from Jensen's inequality that for any $\mathbb{P}\in\mathcal{P}^+$, $\mathbb{E}[X^2]\geq\mathbb{E}[X]^2$, which gives $\sigma\geq\mu^2$ in (5). On the other hand, if $\sigma\geq\mu^2$, it is also rudimentary to find $\mathbb{P}\in\mathcal{P}_2^+$ with $\mathbb{E}[X]=\mu$ and $\mathbb{E}[X^2]=\sigma$. Substituting $\mu=\eta/\nu$ and $\sigma=2\beta/\nu$, we get $\eta^2\leq 2\beta\nu$. Lastly, $\mathbb{E}[X^2]=\mathbb{E}[X]^2$ if and only if $\mathbb{P}$ is a point mass. The equivalent statements regarding program (1) follows from Theorem 2. □

Proof of Proposition 2. Consider a sequence $f^{(k)}(x), x\geq a$ given by

$$f^{(k)}(x)=\begin{cases}\eta-\nu(x-a) & \text{for } a\leq x\leq x_1^{(k)}+a\\ \eta-\nu x_1^{(k)}-\nu p_2^{(k)}(x-a-x_1^{(k)}) & \text{for } x_1^{(k)}+a\leq x\leq x_2^{(k)}+a\\ 0 & \text{for } x_2^{(k)}+a\leq x\end{cases}$$

where

$$\begin{aligned}&x_1^{(k)}=\mu-\gamma^{(k)} \text{ and } \gamma^{(k)}=\tfrac{\sigma-\mu^2}{x_2^{(k)}-\mu}\\ &x_2^{(k)}\to\infty\\ &p_1^{(k)}=1-p_2^{(k)}\\ &p_2^{(k)}=\tfrac{\sigma-\mu^2}{{x_2^{(k)}}^2-2\mu x_2^{(k)}+\sigma}\end{aligned}\tag{EC.13}$$

and μ,σ are defined in (4).

We claim that $f^{(k)}$ is feasible for (1) for large enough k. This can be argued by verifying that $(x_1^{(k)},x_2^{(k)},p_1^{(k)},p_2^{(k)})\in\mathcal{P}_2^+$ is feasible for (5) and invoking the one-to-one correspondence between the feasible solutions in (1) and (5) depicted in Theorem 2. Here we provide an alternate direct verification. It is obvious that for large enough $x_2^{(k)}$, $f^{(k)}$ is non-negative and convex. Moreover, $f^{(k)}(a)=f^{(k)}(a+)=\eta$ and ${f_+^{(k)}}'(a)\geq-\nu$. To show $\int_a^\infty f(x)dx=\beta$, we first verify that

$$p_1^{(k)}x_1^{(k)}+p_2^{(k)}x_2^{(k)}=\mu\tag{EC.14}$$

and

$$p_1^{(k)}{x_1^{(k)}}^2+p_2^{(k)}{x_2^{(k)}}^2=\sigma\tag{EC.15}$$

for large k. In fact, we will do so by showing that $\gamma^{(k)}$ and $p_2^{(k)}$ displayed in (EC.13) are the unique choices that satisfy (EC.14) and (EC.15) and also $x_1^{(k)}=\mu-\gamma^{(k)}$ and $p_1^{(k)}=1-p_2^{(k)}$. With the latter conditions, (EC.14) and (EC.15) can be written as

$$(1-p_2^{(k)})(\mu-\gamma^{(k)})+p_2^{(k)}x_2^{(k)}=\mu$$

and

$$(1-p_2^{(k)})(\mu-\gamma^{(k)})^2+p_2^{(k)}{x_2^{(k)}}^2=\sigma$$

respectively, which further gives

$$p_2^{(k)}\left(\gamma^{(k)}+x_2^{(k)}-\mu\right)-\gamma^{(k)}=0 \tag{EC.16}$$

and

$$p_2^{(k)}\left({x_2^{(k)}}^2-\left(\mu-\gamma^{(k)}\right)^2\right)+\left(\mu-\gamma^{(k)}\right)^2=\sigma \tag{EC.17}$$

From (EC.16) we have

$$p_2^{(k)}=\frac{\gamma^{(k)}}{\gamma^{(k)}+x_2^{(k)}-\mu} \tag{EC.18}$$

Putting (EC.18) into (EC.17), we get

$$\frac{\gamma^{(k)}}{\gamma^{(k)}+x_2^{(k)}-\mu}\left({x_2^{(k)}}^2-\left(\mu-\gamma^{(k)}\right)^2\right)+\left(\mu-\gamma^{(k)}\right)^2=\sigma$$

which can be simplified to

$$\gamma^{(k)}\left(x_2^{(k)}+\mu-\gamma^{(k)}\right)+\left(\mu-\gamma^{(k)}\right)^2=\sigma$$

giving

$$\gamma^{(k)}=\frac{\sigma-\mu^2}{x_2^{(k)}-\mu} \tag{EC.19}$$

Plugging (EC.19) into (EC.18), we have

$$p_2^{(k)}=\frac{\sigma-\mu^2}{(\sigma-\mu^2)+\left(x_2^{(k)}-\mu\right)^2} \tag{EC.20}$$

thus recovering $\gamma^{(k)}$ and $p_2^{(k)}$ in (EC.13).

Therefore,

$$\begin{aligned}
\int_a^\infty f^{(k)}(x)dx &= \int_a^{x_1^{(k)}+a}[\eta-\nu(x-a)]dx+\int_{x_1^{(k)}+a}^{x_2^{(k)}+a}[\eta-\nu x_1^{(k)}-\nu p_2^{(k)}(x-a-x_1^{(k)})]dx \\
&= \eta x_2^{(k)}-\frac{\nu}{2}{x_1^{(k)}}^2-\nu x_1^{(k)}(x_2^{(k)}-x_1^{(k)})-\frac{\nu p_2^{(k)}}{2}(x_2^{(k)}-x_1^{(k)})^2 \\
&= \eta x_2^{(k)}+\frac{\nu p_1^{(k)}}{2}{x_1^{(k)}}^2+\frac{\nu p_2^{(k)}}{2}{x_2^{(k)}}^2-\nu p_1^{(k)}x_1^{(k)}x_2^{(k)}-\nu p_2^{(k)}{x_2^{(k)}}^2 \text{ using } p_1^{(k)}=1-p_2^{(k)} \\
&= \eta x_2^{(k)}+\frac{\nu\sigma}{2}-\nu x_2^{(k)}\mu \text{ using (EC.14) and (EC.15)} \\
&= \beta \text{ using } \eta-\nu\mu=0 \text{ and } \beta=\nu\sigma/2
\end{aligned}$$

Hence $f^{(k)}$ is feasible for (1) for large enough k.

Now, the objective value evaluated at $f^{(k)}$ is

$$\int_a^{x_1^{(k)}+a} h(x)(\eta-\nu(x-a))dx+\int_{x_1^{(k)}+a}^{x_2^{(k)}+a} h(x)(\eta-\nu x_1^{(k)}-\nu p_2^{(k)}(x-a-x_1^{(k)}))dx \tag{EC.21}$$

The first term in (EC.21) is bounded since $x_1^{(k)}\to\mu$. We focus on the second term. By the assumption, we can find $C>0$ such that $h(x)\geq Cx^\epsilon$ for all $x\geq a$. Then, for large enough k,

$$\begin{aligned}
&\int_{x_1^{(k)}+a}^{x_2^{(k)}+a} h(x)(\eta-\nu x_1^{(k)}-\nu p_2^{(k)}(x-a-x_1^{(k)}))dx\\
&\geq C\int_{x_1^{(k)}+a}^{x_2^{(k)}+a} x^\epsilon(\eta-\nu x_1^{(k)}-\nu p_2^{(k)}(x-a-x_1^{(k)}))dx\\
&\geq C\int_{x_1^{(k)}+a}^{x_2^{(k)}+a} [(\eta-\nu p_1^{(k)}x_1^{(k)}+\nu p_2^{(k)}a)x^\epsilon-\nu p_2^{(k)}x^{\epsilon+1}]dx\\
&=(\eta-\nu p_1^{(k)}x_1^{(k)}+\nu p_2^{(k)}a)\frac{x^{\epsilon+1}}{\epsilon+1}\Big|_{x_1^{(k)}+a}^{x_2^{(k)}+a}-\nu p_2^{(k)}\frac{x^{\epsilon+2}}{\epsilon+2}\Big|_{x_1^{(k)}+a}^{x_2^{(k)}+a}\\
&=(\eta-\nu p_1^{(k)}x_1^{(k)}+\nu p_2^{(k)}a)\frac{(x_2^{(k)}+a)^{\epsilon+1}}{\epsilon+1}-(\eta-\nu p_1^{(k)}x_1^{(k)}+\nu p_2^{(k)}a)\frac{(x_1^{(k)}+a)^{\epsilon+1}}{\epsilon+1}\\
&\quad-\nu p_2^{(k)}\frac{(x_2^{(k)}+a)^{\epsilon+2}}{\epsilon+2}+\nu p_2^{(k)}\frac{(x_1^{(k)}+a)^{\epsilon+2}}{\epsilon+2}
\end{aligned} \tag{EC.22}$$

Note that since $p_1^{(k)}\to1$, $x_1^{(k)}\to\mu$, $p_2^{(k)}\to0$ and $\eta-\nu\mu=0$, the second term in (EC.22) converges to 0. Moreover, since $p_2^{(k)}\to0$, the fourth term also converges to 0. Consider the first term in (EC.22). In particular,

$$\begin{aligned}
\eta-\nu p_1^{(k)}x_1^{(k)}+\nu p_2^{(k)}a&=\eta-\nu(1-p_2^{(k)})(\mu-\gamma^{(k)})+\nu p_2^{(k)}a\\
&=p_1^{(k)}\nu\gamma^{(k)}+\nu p_2^{(k)}(\mu+a)
\end{aligned}$$

by using $\eta-\nu\mu=0$ and $p_1^{(k)}=1-p_2^{(k)}$. Substituting $\gamma^{(k)}=(\sigma-\mu^2)/(x_2^{(k)}-\mu)$ and $p_2^{(k)}=\Theta(1/{x_2^{(k)}}^2)$, and using $p_1^{(k)}\to1$, we have

$$(\eta-\nu p_1^{(k)}x_1^{(k)}+\nu p_2^{(k)}a)\frac{(x_2^{(k)}+a)^{\epsilon+1}}{\epsilon+1}=(p_1^{(k)}\nu\gamma^{(k)}+\nu p_2^{(k)}(\mu+a))\frac{(x_2^{(k)}+a)^{\epsilon+1}}{\epsilon+1}=\frac{\nu(\sigma-\mu^2){x_2^{(k)}}^\epsilon}{\epsilon+1}(1+o(1))$$

On the other hand, for the third term in (EC.22), substituting $p_2^{(k)}=(\sigma-\mu^2)/({x_2^{(k)}}^2-2\mu x_2^{(k)}+\sigma)$, we have

$$-\nu p_2^{(k)}\frac{(x_2^{(k)}+a)^{\epsilon+2}}{\epsilon+2}=-\frac{\nu(\sigma-\mu^2){x_2^{(k)}}^\epsilon}{\epsilon+2}(1+o(1))$$

Thus, (EC.22) is equal to

$$\left(\frac{1}{\epsilon+1}-\frac{1}{\epsilon+2}\right)\nu(\sigma-\mu^2)x_2^{(k)^\epsilon}(1+o(1))\to\infty$$

and hence the optimal value of (1) is ∞. □

EC.2. Proofs for Section 5

To prove Proposition 3, we borrow the following result:

LEMMA EC.2 **(Adapted from Theorem 5.1 in Birge and Dulá (1991))**. *Consider $OPT(\mathcal{P}[0,\tilde{c}])$ for any $0<\tilde{c}<\infty$. Suppose H is convex with derivative H' convex on $(0,c)$ and concave on $(c,\tilde{c})$ for some $0\le c\le\tilde{c}$. If $OPT(\mathcal{P}[0,\tilde{c}])$ is consistent, then an optimal solution exists and lies in $\mathcal{P}_2[0,\tilde{c}]$.*

This lemma follows from Theorem 5.1 in Birge and Dulá (1991) that applies to the associated dual problem.

Proof of Proposition 3. By Theorem 3, $OPT(\mathcal{P}^+)$ has the same optimal value as $OPT(\mathcal{P}_3^+)$. By Lemma EC.2, for every $\mathbb{P}$ feasible for $OPT(\mathcal{P}_3^+)$, which necessarily has bounded support say on $[0,M]$ for some $M>0$, there exists $\mathbb{P}'\in\mathcal{P}_2[0,M]$ with the same first and second moments such that $Z(\mathbb{P}')\ge Z(\mathbb{P})$. Since $\mathcal{P}_2^+\subset\mathcal{P}_3^+$, this implies that $OPT(\mathcal{P}_3^+)$ has the same optimal value as $OPT(\mathcal{P}_2^+)$, which concludes the proposition. □

Proof of Proposition 4. Proof of 1: Let the optimal probability measure in $\mathcal{P}_2^+$ be represented by (x_1,x_2,p_1,p_2). Note that $x_1\neq x_2$ since otherwise $\sigma=\mu^2$. Adopting a similar line of analysis as in Birge and Dulá (1991), we let $x_1<x_2$ without loss of generality. For a two-support-point distribution to be feasible, we must have $x_1<\mu$. Feasibility also enforces that $p_1x_1+p_2x_2=\mu$, $p_1x_1^2+p_2x_2^2=\sigma$ and $p_1+p_2=1$. Hence $p_2=1-p_1$, which gives $p_1x_1+(1-p_1)x_2=\mu$ and $p_1x_1^2+(1-p_1)x_2^2=\sigma$. From the first equation we get $p_1=(x_2-\mu)/(x_2-x_1)$. Putting this into $p_1x_1^2+(1-p_1)x_2^2=\sigma$, we further get $x_2=(\sigma-\mu x_1)/(\mu-x_1)$. Now, putting this in turn into $p_1=(x_2-\mu)/(x_2-x_1)$, we obtain $p_1=(\sigma-\mu^2)/(\sigma-2\mu x_1+x_1^2)$ and hence $p_2=1-p_1=(\mu-x_1)^2/(\sigma-2\mu x_1+x_1^2)$. Therefore, Z^* is given by

$$\max_{x_1\in[0,\mu)}\nu(p_1H(x_1)+p_2H(x_2))=\max_{x_1\in[0,\mu)}\nu\left(\frac{\sigma-\mu^2}{\sigma-2\mu x_1+x_1^2}H(x_1)+\frac{(\mu-x_1)^2}{\sigma-2\mu x_1+x_1^2}H\left(\frac{\sigma-\mu x_1}{\mu-x_1}\right)\right)$$

which is exactly $\max_{x_1\in[0,\mu)} W(x_1)$.

Proof of 2: Let $\mathbb{P}^{(k)} \sim (x_1^{(k)}, x_2^{(k)}, p_1^{(k)}, p_2^{(k)})$ be a feasible sequence with $Z(\mathbb{P}^{(k)}) \to Z^*$. Without loss of generality let $x_1^{(k)} \le x_2^{(k)}$. Since $p_1^{(k)}x_1^{(k)} + p_2^{(k)}x_2^{(k)} = \mu$, we must have $x_1^{(k)} \le \mu$. Then we must have a subsequence $x_2^{(k_i)} \to \infty$, since otherwise $(x_1^{(k)}, x_2^{(k)}, p_1^{(k)}, p_2^{(k)})$ would lie in a compact set and there would exist a subsequence $(x_1^{(k_i')}, x_2^{(k_i')}, p_1^{(k_i')}, p_2^{(k_i')}) \to (x_1^*, x_2^*, p_1^*, p_2^*)$, where $Z(\mathbb{P}^{(k_i')}) = \nu\sum_{j=1}^2 p_j^{(k_i')} H(x_j^{(k_i')}) \to \nu\sum_{j=1}^2 p_j^* H(x_j^*)$ by the continuity of H, violating the non-existence of optimal solution. By $p_1^{(k_i)}{x_1^{(k_i)}}^2 + p_2^{(k_i)}{x_2^{(k_i)}}^2 = \sigma$, we have $p_2^{(k_i)} = (\sigma - p_1^{(k_i)}{x_1^{(k_i)}}^2)/{x_2^{(k_i)}}^2 \to 0$, and $p_2^{(k_i)}x_2^{(k_i)} = (\sigma - p_1^{(k_i)}{x_1^{(k_i)}}^2)/x_2^{(k_i)} \to 0$. Thus $p_1^{(k_i)} = 1 - p_2^{(k_i)} \to 1$ and $x_1^{(k_i)} = (\mu - p_2^{(k_i)}x_2^{(k_i)})/p_1^{(k_i)} \to \mu$. Therefore,

$$Z(\mathbb{P}^{(k_i)}) = \nu\left(p_1^{(k_i)}H(x_1^{(k_i)}) + p_2^{(k_i)}H(x_2^{(k_i)})\right) = \nu\left(p_1^{(k_i)}H(x_1^{(k_i)}) + \frac{\sigma - p_1^{(k_i)}{x_1^{(k_i)}}^2}{{x_2^{(k_i)}}^2} H(x_2^{(k_i)})\right)$$
$$\to \nu(H(\mu) + \lambda(\sigma - \mu^2))$$

Proof of 3: First, we show that $W(x_1) \to \nu(H(\mu) + \lambda(\sigma-\mu^2))$ as $x_1 \nearrow \mu$. Consider the second term of $W(x_1)$ given by

$$\lim_{x_1\nearrow\mu} \frac{\nu(\mu - x_1)^2}{\sigma - 2\mu x_1 + x_2^2} H\left(\frac{\sigma-\mu x_1}{\mu - x_1}\right) = \lim_{x_1\nearrow\mu} \frac{\nu(\sigma-\mu x_1)^2}{\sigma - 2\mu x_1 + x_2^2}\left(\frac{\mu - x_1}{\sigma - \mu x_1}\right)^2 H\left(\frac{\sigma-\mu x_1}{\mu - x_1}\right) = \nu\lambda(\sigma - \mu^2)$$

and the claim follows. Combining Parts 1 and 2 of this proposition, we must have $Z^* = \max_{x_1\in[0,\mu]} W(x_1)$. □

EC.3. Proofs for Section 6

We first show a result in parallel to Theorem 2 for the case of (11):

THEOREM EC.7. *Suppose h is bounded. Then the optimal value of* (11) *is the same as*

$$\begin{aligned} \max_{\mathbb{P}} \quad & \overline{\nu}\mathbb{E}[H(X)] \\ \textit{subject to} \quad & \underline{\mu} \le \mathbb{E}[X] \le \overline{\mu} \\ & \underline{\sigma} \le \mathbb{E}[X^2] \le \overline{\sigma} \\ & \mathbb{P} \in \mathcal{P}^+ \end{aligned} \tag{EC.23}$$

Here the decision variable is a probability distribution $\mathbb{P}\in\mathcal{P}^+$, and $\mathbb{E}[\cdot]$ is the corresponding expectation. Moreover, there is a one-to-one correspondence between the feasible solutions to (11) *and* (EC.23)*, given by $f'_+(x+a)=\overline{\nu}(p(x)-1)$ for $x\in\mathbb{R}^+$, where f'_+ is the right derivative of a feasible solution f of* (11) *such that $f(x)=\int_a^x f'_+(t)dt+\eta$ for $x\geq a$, and p is a probability distribution function that is associated with a feasible probability measure over $\mathbb{R}^+$ in* (EC.23)*.*

Proof of Theorem EC.7. Note that formulation (11) can be written as

$$\begin{aligned}\max_{\underline{\beta}\leq\beta\leq\overline{\beta},\underline{\eta}\leq\eta\leq\overline{\eta}}\max_f\quad & \int_a^\infty h(x)f(x)dx\\ \text{subject to}\quad & \int_a^\infty f(x)dx=\beta\\ & f(a)=f(a+)=\eta\\ & f'_+(a)\geq-\overline{\nu}\\ & f(x)\text{ convex for }x\geq a\\ & f(x)\geq 0\text{ for }x\geq a\end{aligned}\tag{EC.24}$$

The inner maximization is exactly (1), and thus by Theorem 2 we can reformulate (EC.24) as

$$\begin{aligned}\max_{\underline{\beta}\leq\beta\leq\overline{\beta},\underline{\eta}\leq\eta\leq\overline{\eta}}\max_{\mathbb{P}}\quad & \overline{\nu}\mathbb{E}[H(X)]\\ \text{subject to}\quad & \mathbb{E}[X]=\frac{\eta}{\overline{\nu}}\\ & \mathbb{E}[X^2]=\frac{2\beta}{\overline{\nu}}\\ & \mathbb{P}\in\mathcal{P}^+\end{aligned}$$

which is equivalent to (EC.23). □

For convenience, we denote $\widetilde{OPT}(\mathcal{D})$ as the program

$$\begin{aligned}\max_{\mathbb{P}}\quad & \overline{\nu}\mathbb{E}[H(X)]\\ \text{subject to}\quad & \underline{\mu}\leq\mathbb{E}[X]\leq\overline{\mu}\\ & \underline{\sigma}\leq\mathbb{E}[X^2]\leq\overline{\sigma}\\ & \mathbb{P}\in\mathcal{D}\end{aligned}$$

where $\mathcal{D}$ is a collection of probability measures on $\mathbb{R}$. For example, (EC.23) can be written as $\widetilde{OPT}(\mathcal{P}^+)$. Let $\tilde{Z}(\mathbb{P})=\overline{\nu}\mathbb{E}[H(X)]$ be the objective function in $\mathbb{P}$.

Proposition EC.2. *The optimal value of* $\widetilde{OPT}(\mathcal{P}^+)$ *is identical to that of* $\widetilde{OPT}(\mathcal{P}_3^+)$*.*

Proof of Proposition EC.2. For $\mathbb{P}$ feasible in $\widetilde{OPT}(\mathcal{P}^+)$, let $\mu = \mathbb{E}[X]$ and $\sigma = \mathbb{E}[X^2]$ be its first and second moments. By Theorem 3 there must exist $\mathbb{P}' \in \mathcal{P}_3^+$ with the corresponding expectations $\mathbb{E}'[X] = \mu$ and $\mathbb{E}'[X^2] = \sigma$ such that $\tilde{Z}(\mathbb{P}) \leq \tilde{Z}(\mathbb{P}')$. Since $\mathcal{P}_3^+ \subset \mathcal{P}^+$, we conclude that the optimal value of $\widetilde{OPT}(\mathcal{P}^+)$ is identical to that of $\widetilde{OPT}(\mathcal{P}_3^+)$. □

Proof of Theorem 5. Theorem 5 follows from Theorem EC.7 and Proposition EC.2, in the same way as the proof of Theorem 1. □

Proposition EC.3. *Under Assumption 3,* $\widetilde{OPT}(\mathcal{P}^+)$ *has the same optimal value as* $\widetilde{OPT}(\mathcal{P}_2^+)$*.*

Proof of Proposition EC.3. We know from Proposition EC.2 that $\widetilde{OPT}(\mathcal{P}^+)$ has the same optimal value as $\widetilde{OPT}(\mathcal{P}_3^+)$. Any $\mathbb{P}$ feasible for $\widetilde{OPT}(\mathcal{P}_3^+)$ must necessarily have bounded support, say on $[0, M]$. By Lemma EC.2 there must exist $\mathbb{P}' \in \mathcal{P}_2^+$, with the same first and second moments as $\mathbb{P}$, such that $\tilde{Z}(\mathbb{P}) \leq \tilde{Z}(\mathbb{P}')$. Since $\mathcal{P}_2^+ \subset \mathcal{P}_3^+$, this implies that $\widetilde{OPT}(\mathcal{P}_3^+)$ has the same optimal value as $\widetilde{OPT}(\mathcal{P}_2^+)$, which concludes the proposition. □

The following explains the origin of the two subproblems in (13):

Lemma EC.3. *Under Assumption 1, and let* $\overline{\sigma} \geq \underline{\mu}^2$*. The optimal value of* $\widetilde{OPT}(\mathcal{P}_2^+)$ *is given by* $\tilde{Z}^* = \max\{\tilde{Z}_1^*, \tilde{Z}_2^*\}$*, where* $\tilde{Z}_1^*$ *is the optimal value of*

$$\begin{aligned} \max_{\mathbb{P}} \quad & \overline{\nu}\mathbb{E}[H(X)] \\ \textit{subject to} \quad & \mathbb{E}[X] = \overline{\mu} \\ & \underline{\sigma} \leq \mathbb{E}[X^2] \leq \overline{\sigma} \\ & \mathbb{P} \in \mathcal{P}_2^+ \end{aligned} \tag{EC.25}$$

and $\tilde{Z}_2^*$ *is the optimal value of*

$$\begin{aligned} \max_{\mathbb{P}} \quad & \overline{\nu}\mathbb{E}[H(X)] \\ \textit{subject to} \quad & \underline{\mu} \leq \mathbb{E}[X] \leq \overline{\mu} \\ & \mathbb{E}[X^2] = \overline{\sigma} \\ & \mathbb{P} \in \mathcal{P}_2^+ \end{aligned} \tag{EC.26}$$

respectively.

Proof of Lemma EC.3. We argue that to solve $\widetilde{OPT}(\mathcal{P}_2^+)$, it suffices to restrict attention to the feasible region $\{\mathbb{P}\in\mathcal{P}_2^+:\mathbb{E}[X]=\overline{\mu},\underline{\sigma}\leq\mathbb{E}[X^2]\leq\overline{\sigma}\}\cup\{\mathbb{P}\in\mathcal{P}_2^+:\underline{\mu}\leq\mathbb{E}[X]\leq\overline{\mu},\mathbb{E}[X^2]=\overline{\sigma}\}$. Since $h\geq 0$, $\tilde{Z}^*\geq 0$. There is nothing to prove if $\tilde{Z}^*=0$. So suppose $\tilde{Z}^*>0$. There exists $\mathbb{P}\sim(x_1,x_2,p_1,p_2)\in\mathcal{P}_2^+$ with one of the x_i's having $H(x_i)>0$ and $p_i>0$. Now suppose $\mathbb{P}$ satisfies $\mathbb{E}[X]<\overline{\mu}$ and $\mathbb{E}[X^2]<\overline{\sigma}$. We can increase x_i so that $\mathbb{E}[X]\leq\overline{\mu}$ and $\mathbb{E}[X^2]\leq\overline{\sigma}$ remain satisfied, and $\tilde{Z}^*(\mathbb{P})$ is at least as large as before since $H(x)$ is non-decreasing. Hence any $\mathbb{P}$ such that $\mathbb{E}[X]<\overline{\mu}$ and $\mathbb{E}[X^2]<\overline{\sigma}$ must have $\tilde{Z}(\mathbb{P})\leq\tilde{Z}(\mathbb{P}')$ for some $\mathbb{P}'\in\{\mathbb{P}\in\mathcal{P}_2^+:\mathbb{E}[X]=\overline{\mu},\underline{\sigma}\leq\mathbb{E}[X^2]\leq\overline{\sigma}\}\cup\{\mathbb{P}\in\mathcal{P}_2^+:\underline{\mu}\leq\mathbb{E}[X]\leq\overline{\mu},\mathbb{E}[X^2]=\overline{\sigma}\}$. This proves the lemma. □

Proof of Theorem 6. Lemma EC.3 allows one to consider only the programs (EC.25) and (EC.26) when solving $\widetilde{OPT}(\mathcal{P}_2^+)$. Theorem 6 then follows from Lemma 3, Theorem EC.7 and Proposition EC.3, using the same line of arguments in the proof of Theorem 4. □

References

Birge JR, Dulá JH (1991) Bounding separable recourse functions with limited distribution information. *Ann. Oper. Res.* 30(1):277–298.

Rockafellar RT (1997) *Convex Analysis.* (Princeton University Press, Princeton, NJ).

Winkler G (1988) Extreme points of moment sets. *Math. Oper. Res.* 13(4):581–587.