2020

A problem of distributive justice, solved by the lasso

Prof. Philip T. Reiss

A problem of distributive justice, solved by the lasso

Philip T. Reiss
Department of Statistics, University of Haifa

October 7, 2019

Abstract

The problem of dividing an estate among creditors, when their claims total more than the value of the estate, was posed in the Talmud and has been analyzed in the game theory literature. Here we reveal a close connection between schemes for estate division and linear regression solution paths obtained by least angle regression or by the lasso. We focus primarily on the division scheme known as constrained equal awards, but also consider a more complex approach described by Aumann and Maschler (1985).

Keywords: constrained equal awards, estate division, least angle regression, Talmud

1 Introduction

When a person dies with debts totaling more than the value of the estate, the question arises of how to divide the estate among the creditors. This problem of distributive justice, raised in the Talmud, was revisited from a game-theoretic perspective by Aumann and Maschler (1985), in work that has spawned much subsequent research in the mathematical social sciences literature.

At first glance there is nothing statistical about the estate division problem, but in this note we reveal a surprising connection to one of the most influential developments in statistical research of the last quarter-century: the lasso ("least absolute shrinkage and selection operator") method for linear regression (Tibshirani, 1996). Specifically, we show that the standard Talmudic solution to the allocation problem can be formulated as the lasso solution to a particular linear regression problem.

After fixing notation in the following section, in Section 3 we present the standard Talmudic formula for estate division, known today as *constrained equal awards* (CEA). In Section 4 we demonstrate how the CEA solution corresponds to the least angle regression (LARS) solution (Efron et al., 2004) to a regression problem. LARS is closely related to the lasso, and equivalent to it in some cases—including our case, as we explain in Section 5. In Section 6 we relate LARS and lasso to to a more complex division scheme described by Aumann and Maschler (1985). Section 7 describes how estate divisions can be computed using software for LARS and the lasso. This is admittedly of limited practical use—at least for CEA solutions, which are very straightforward to compute—but may have some pedagogical value. Some concluding remarks are offered in Section 8.

2 Notation

We assume there are n creditors $\mathcal{C}_1, \ldots, \mathcal{C}_n$, with claims $y_1, \ldots, y_n$ respectively, and for simplicity we assume throughout that

$$0 < y_1 \leq \ldots \leq y_n. \tag{1}$$

Let E be the value of the estate, assumed to be less than $\sum_{i=1}^{n} y_i$, i.e., there is a shortfall

$$S \equiv \sum_{i=1}^{n} y_i - E > 0 \tag{2}$$

in paying the debts. A division of the estate can be defined as an n-tuple $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ such that $0 \leq \varepsilon_i \leq y_i$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} \varepsilon_i = E$. If $\beta_i \in [0, y_i]$ is the shortfall (amount unpaid) for creditor $\mathcal{C}_i$, then we clearly have

$$y_i = \beta_i + \varepsilon_i, \quad i = 1, \ldots, n. \tag{3}$$

This recalls traditional regression notation, but with the conventional interpretations reversed: here the ε_i's are the "solutions" (the payouts to the creditors), whereas the β_i's are more like "residuals." The precise connection with a formal linear regression model will be explained in Section 4.

3 A Talmudic solution: Constrained equal awards

The Talmud is a fundamental text of Judaism that was set in its final form in late antiquity. Our estate division problem is posed in the Talmudic tractate *Ketubot* (p. 93a ff.). It should be noted that the Talmud establishes rules, related to when the debts were incurred, that would in many practical cases imply that some creditors take precedence over others; but here we assume that no such precedence applies.

We focus on the simple solution accepted as normative by codifiers of Jewish law, from Rabbi Isaac Alfasi (11th century) onward, and referred to by Aumann and Maschler (1985) as constrained equal awards (CEA). The CEA solution is

$$\varepsilon_i = \min\{\alpha, y_i\}, \tag{4}$$

for $i = 1, \ldots, n$, for some value $\alpha \geq 0$. Aumann and Maschler (1985) prove straightforwardly that allocation (4), along with the constraint $\sum_{i-1}^{n} \varepsilon_i = E$, implies a unique value of α.

To see how to construct the specific CEA solution to a given estate division problem, it is helpful to consider an example, such as the one presented by Maimonides (*Mishneh*

Total estate E	$\mathcal{C}_1$ ($y_1 = 100$)	$\mathcal{C}_2$ ($y_2 = 200$)	$\mathcal{C}_3$ ($y_3 = 300$)
200	66⅔	66⅔	66⅔
300	100	100	100
400	100	150	150
500	100	200	200
600	100	200	300

Table 1: CEA payouts $\varepsilon_1, \varepsilon_2, \varepsilon_3$ to three creditors claiming amounts (5), for selected estate values E.

Torah, Laws of Lender and Borrower 20:4), of debts in the amounts of

$$(y_1, y_2, y_3) = (100, 200, 300). \tag{5}$$

The allocation is presented by Maimonides in a "dynamic" form, as proceeding in stages as the estate value E increases—or as we might call it, a *solution path*:

1. The first 300 or portion thereof is divided equally among the three creditors.
2. The next 200 or portion thereof is divided equally between $\mathcal{C}_2$ and $\mathcal{C}_3$.
3. The last 100 or portion thereof is given to $\mathcal{C}_3$.

The resulting division is displayed in Table 1 for selected values of E, and in Figure 1 for all $E \in [0, 600]$.

In the notation of (3), the stagewise equal distribution illustrated by this example can be formulated as an iterative algorithm, in which distribution to $\mathcal{C}_i$ is represented by increasing the corresponding ε_i at a given step. At the ℓth iteration, the amount yet to be distributed is denoted by $R_{\ell-1}$; $\mathcal{C}_1, \ldots, \mathcal{C}_{\ell-1}$ have already been paid in full, and part or all of the remaining $R_{\ell-1}$ is divided equally among $\mathcal{C}_\ell, \ldots, \mathcal{C}_n$. Let $\boldsymbol{\varepsilon}^\ell = (\varepsilon_1^\ell, \ldots, \varepsilon_n^\ell)^T$ denote the allocation to $\mathcal{C}_1, \ldots, \mathcal{C}_n$ up to and including the ℓth iteration. We define $y_0 = 0$ and $\boldsymbol{\varepsilon}^0 = \mathbf{0} \in \mathbb{R}^n$.

Algorithm 1 (CEA division of an estate).

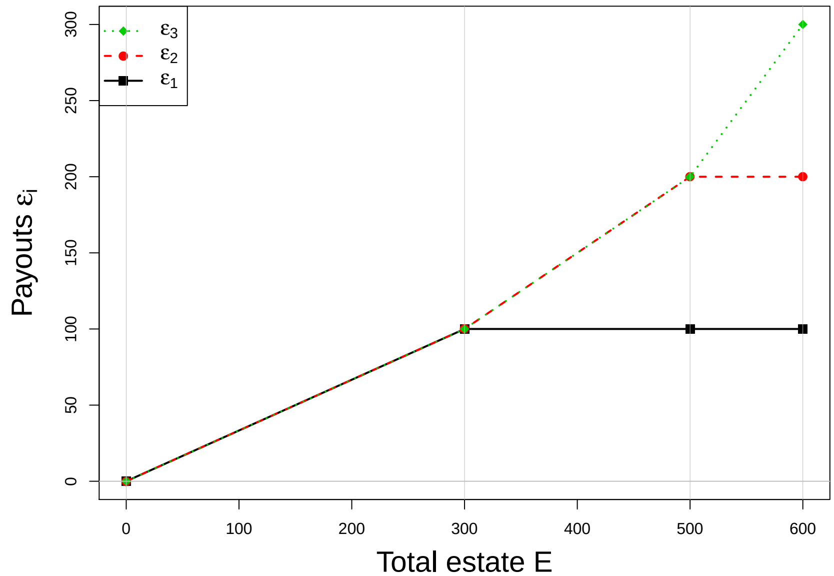

Figure 1: CEA payouts $\varepsilon_1, \varepsilon_2, \varepsilon_3$, for our example with debts (5), as functions of total estate E.

1. Set $\ell = 0$ and $R_0 = E$.

2. Increment ℓ by 1.

3. Let $R_\ell = R_{\ell-1} - (n - \ell + 1)(y_\ell - y_{\ell-1})$.

4. If $R_\ell \geq 0$ then the amount remaining is sufficient to distribute $y_\ell - y_{\ell-1}$ to each of $\mathcal{C}_\ell, \ldots, \mathcal{C}_n$, i.e., to increase $\varepsilon_\ell, \ldots, \varepsilon_n$ by that amount. Thus:

   (a) For $i = \ell, \ldots, n$, we update

   $$\varepsilon_i^\ell = \sum_{m=1}^{\ell} (y_m - y_{m-1}) = y_\ell; \tag{6}$$

   in particular, $\varepsilon_\ell^\ell = y_\ell$, i.e., $\mathcal{C}_\ell$ has been paid in full.

   (b) For $i < \ell$, the allocation to $\mathcal{C}_i$ is not updated since $\mathcal{C}_i$ was previously paid in full, i.e.,

   $$\varepsilon_i^\ell = \varepsilon_i^{\ell-1} = y_i. \tag{7}$$

(c) If $\ell = n$, all debts have been paid in full and the algorithm terminates with payout vector $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^n$; otherwise return to step 2.

5. If $R_\ell < 0$, apportion $R_{\ell-1}$ equally among $\mathcal{C}_\ell, \ldots, \mathcal{C}_n$: the algorithm then terminates with a shortfall, with payouts

$$\begin{aligned} \varepsilon_i &= \varepsilon_i^{\ell-1} + R_{\ell-1}/(n-\ell+1) \\ &= y_{\ell-1} + R_{\ell-1}/(n-\ell+1) \quad \text{for } i = \ell, \ldots, n \end{aligned}$$

and $\varepsilon_i = y_i$ as in (7) for $i < \ell$. □

The interim payout at iteration $\ell \in \{0, \ldots, n\}$ follows from (6), (7):

$$\varepsilon_i^\ell = y_{\min\{i,\ell\}} \text{ for } i = 1, \ldots, n. \tag{8}$$

Summing over i shows that the payout vectors $\boldsymbol{\varepsilon}^0, \ldots, \boldsymbol{\varepsilon}^n$ correspond to estate values

$$E^0 = 0,\ E^1 = ny_1, \text{ and } E^\ell = \left(\sum_{i=1}^{\ell-1} y_i \right) + (n-\ell+1)y_\ell \text{ for } \ell = 2, \ldots, n. \tag{9}$$

These are the break points defining the linear segments of the payout functions—in our example, 0, 300, 500 and 600 (see Figure 1).

4 Least angle regression

4.1 Estate division as a formal linear model

Algebraically, the n equations (3) are equivalent to the familiar linear model equation

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{10}$$

with $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, in the specific case

$$\boldsymbol{X} = \boldsymbol{I}_n, \tag{11}$$

i.e., the $n \times n$ identity matrix. In Section 4.2 we describe least angle regression (abbreviated as LARS in the original paper by Efron et al. (2004), and as LAR by some other authors such as Hastie et al. (2015)), which derives a "path" of solutions $\hat{\boldsymbol{\beta}}$ for a linear model of the general form (10). Then, in Section 4.3, we demonstrate how the CEA division can be derived from LARS in the special case (11). The connection with the lasso is explained in Section 5.

4.2 The general algorithm

For the following outline of the LARS algorithm, let $\boldsymbol{X}$ in (10) be any $n \times m$ matrix whose columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are linearly independent and satisfy

$$\|\boldsymbol{x}_j\| = 1 \text{ for } j = 1, \ldots, m, \tag{12}$$

where (here and below) $\|\cdot\|$ denotes the ℓ_2 norm. It is sometimes assumed that $\boldsymbol{y}$ and the $\boldsymbol{x}_j$'s are centered (to avoid the need for an intercept), but these assumptions are not made here.

LARS is a variant of forward selection that produces a path of coefficient vectors $\boldsymbol{\beta}$, from the zero vector to the ordinary least squares estimate $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$, in a manner that is optimal in a particular geometric sense. Efron et al. (2004) present the algorithm as a series of updates to the active set $\mathcal{A} \subset \{1, \ldots, m\}$ of included predictors and to the associated fitted value vector $\widehat{\boldsymbol{\mu}}_{\mathcal{A}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathcal{A}}$. At each stage (and with possible sign changes to the $\boldsymbol{x}_j$'s), (i) the included predictor vectors ($\boldsymbol{x}_j$ for $j \in \mathcal{A}$) are those that form the smallest angle with the residual vector (hence the algorithm's name); and (ii) the fitted value vector is moved in the direction that is equiangular to each of these vectors. The algorithm proceeds as follows.

Algorithm 2 (Least angle regression).

1. Begin with $\mathcal{A} = \varnothing$ and $\widehat{\boldsymbol{\mu}}_{\mathcal{A}} = \boldsymbol{0} \in \mathbb{R}^n$.

2. Let $\widehat{c}_j = \boldsymbol{x}_j^T(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}_{\mathcal{A}})$ for $j = 1, \ldots, m$.

3. Update $\mathcal{A}$ to $\mathcal{A} \cup \{j \in \mathcal{A}^c : |\widehat{c}_j| = \max_{k \in \mathcal{A}^c} |\widehat{c}_k|\}$.

4. Let $\boldsymbol{X}_{\mathcal{A}}$ be the matrix whose columns are $\text{sign}(\widehat{c}_j)\boldsymbol{x}_j$ for $j \in \mathcal{A}$, and let $\boldsymbol{u}_{\mathcal{A}} = \boldsymbol{X}_{\mathcal{A}}(\boldsymbol{X}_{\mathcal{A}}^T\boldsymbol{X}_{\mathcal{A}})^{-1}\boldsymbol{1}$ divided by its norm, where $\boldsymbol{1}$ is a vector of $|\mathcal{A}|$ 1's. Update $\widehat{\boldsymbol{\mu}}_{\mathcal{A}}$ to $\widehat{\boldsymbol{\mu}}_{\mathcal{A}} + \widehat{\gamma}\boldsymbol{u}_{\mathcal{A}}$, where $\widehat{\gamma} > 0$ is defined below.

5. If $\widehat{\boldsymbol{\mu}}_{\mathcal{A}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{OLS}$, stop; otherwise return to step 2. □

By (12), the $\widehat{c}_j$'s are the cosines of the angles formed by the $\boldsymbol{x}_j$'s with the residual vector $\boldsymbol{y} - \widehat{\boldsymbol{\mu}}_{\mathcal{A}}$, times the length of the latter. Thus, in step 3, we add to $\mathcal{A}$ the one or more thus-far-excluded predictors that, with possible changes of sign, form the smallest angle with the residual vector. Efron et al. (2004) show that, when $\widehat{\boldsymbol{\mu}}_{\mathcal{A}}$ is updated to $\widehat{\boldsymbol{\mu}}_{\mathcal{A}} + \gamma \boldsymbol{u}_{\mathcal{A}}$ for small positive γ, the cosines of the angle between the residual vector and $\text{sign}(\widehat{c}_j)\boldsymbol{x}_j$ for $j \in \mathcal{A}$ are equal, and are larger than the cosines for $\pm\boldsymbol{x}_j$ with $j \in \mathcal{A}^c$. The value $\widehat{\gamma}$ in step 4, given explicitly by (2.13) of Efron et al. (2004), is the smallest positive number such that, after $\widehat{\boldsymbol{\mu}}_{\mathcal{A}}$ is updated to $\widehat{\boldsymbol{\mu}}_{\mathcal{A}} + \widehat{\gamma}\boldsymbol{u}_{\mathcal{A}}$, $\pm\boldsymbol{x}_j$ for one or more $j \in \mathcal{A}^c$ will form the same smallest angle with the residual vector as do $\text{sign}(\widehat{c}_j)\boldsymbol{x}_j$ for $j \in \mathcal{A}$.

In most realistic settings, only one predictor is added to $\mathcal{A}$ in each iteration of step 3 (there are no "ties"). Thus the first step is in the predictor direction forming the smallest angle with $\boldsymbol{y}$; then there are two predictor directions forming the minimal angle with the residual vector, and the second step is equiangular to these two directions; then there are three predictor directions forming the smallest angle with the residual vector, and the third step is equiangular to these three; and so on.

4.3 A special case that is equivalent to Algorithm 1

Assume (1) as above, and consider the case (11) of model (10). Although not ordinarily encountered in practice, this case is discussed in some detail in Section 4.1 of Efron et al. (2004). When $\boldsymbol{X} = \boldsymbol{I}_n$, the fitted value vectors $\widehat{\boldsymbol{\mu}}_{\mathcal{A}}$ at the iterations of the LARS algorithm are equal to the interim parameter estimates $\hat{\boldsymbol{\beta}}$, which we denote by

$$\hat{\boldsymbol{\beta}}^0 = \boldsymbol{0}, \hat{\boldsymbol{\beta}}^1, \ldots, \hat{\boldsymbol{\beta}}^n = \hat{\boldsymbol{\beta}}_{OLS}.$$

(Although there are fewer than n iterations in the event of ties, this notation remains valid if we view $r > 1$ predictors entering the active set together as r "iterations" with the same $\hat{\boldsymbol{\beta}}$ repeated r times.) By Lemma 1 of Efron et al. (2004), for $k = 0, \ldots, n$ and $i = 1, \ldots, n$, the ith element of $\hat{\boldsymbol{\beta}}^k$ is given by $\hat{\beta}_i^k = (y_i - y_{n-k})_+$. Hence $\hat{\boldsymbol{\varepsilon}}^k \equiv \boldsymbol{y} - \hat{\boldsymbol{\beta}}^k$ has components $\hat{\varepsilon}_i^k = y_i - \hat{\beta}_i^k = \min\{y_i, y_{n-k}\} = y_{\min\{i,n-k\}}$.

For $\ell = 0, \dots, n$, let $\tilde{\boldsymbol{\varepsilon}}^\ell = \hat{\boldsymbol{\varepsilon}}^{n-\ell}$, i.e., the $\tilde{\boldsymbol{\varepsilon}}^\ell$'s are simply the $\hat{\boldsymbol{\varepsilon}}^k$'s numbered in reverse. Thus for each i, ℓ, $\tilde{\boldsymbol{\varepsilon}}^\ell$ has ith element

$$\tilde{\varepsilon}_i^\ell = y_{\min\{i,\ell\}}. \tag{13}$$

Exactly the same expression was obtained in (8) for the interim payouts ε_i^ℓ in Algorithm 1. Moreover, since both algorithms proceed in piecewise linear stages, they give the same values for intermediate points. We have thus proved:

Proposition 1 *The payout vectors $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ produced by Algorithm 1, as E increases from 0 to $\sum_{i=1}^n y_i$, are identical to the vectors of residuals* (13) *produced along the reverse LARS path for model* (10)*,* (11)*.*

Note that the ascending order in (1) serves to simplify the common expression (8), (13) for the CEA payouts and the LARS residuals, but is not required in order for Proposition 1 to hold.

4.4 Duality

In the terminology of Aumann and Maschler (1985, p. 202), Proposition 1 implies that CEA divides the estate E into $\varepsilon_1, \dots, \varepsilon_n$ in a manner that is *dual* to the division of the total shortfall S into $\beta_1, \dots, \beta_n$ by LARS with $\boldsymbol{X} = \boldsymbol{I}_n$. This duality is illustrated by Figure 2. The left panel displays the CEA payouts ε_i for the debts (5) in the example of Section 3. This is identical to Figure 1, except the horizontal axis is reversed so that the estate E decreases from 600 at left to 0 at right, or equivalently, the total shortfall $S = \sum_{i=1}^n y_i - E = 600 - E$ increases from 0 to 600. The right panel shows the corresponding individual shortfalls $\beta_i = y_i - \varepsilon_i$ $(i = 1, 2, 3)$. The crux of the duality is that the CEA shortfall functions $\beta_1, \beta_2, \beta_3$ shown in the right panel are precisely the LARS coefficient paths for response vector (5) and design matrix $\boldsymbol{I}_3$. Indeed, the right panel of Figure 2 was produced by a slightly modified version of the default plotting function in the `lars` package (Hastie and Efron, 2013) for R (R Core Team, 2019).

Some authors (e.g., Chun et al., 2001; Herrero and Villar, 2001) have considered a *constrained equal losses* (CEL) rule, which is the dual of the CEA rule. The previous paragraph can be reformulated as saying that the LARS coefficient paths give all CEL

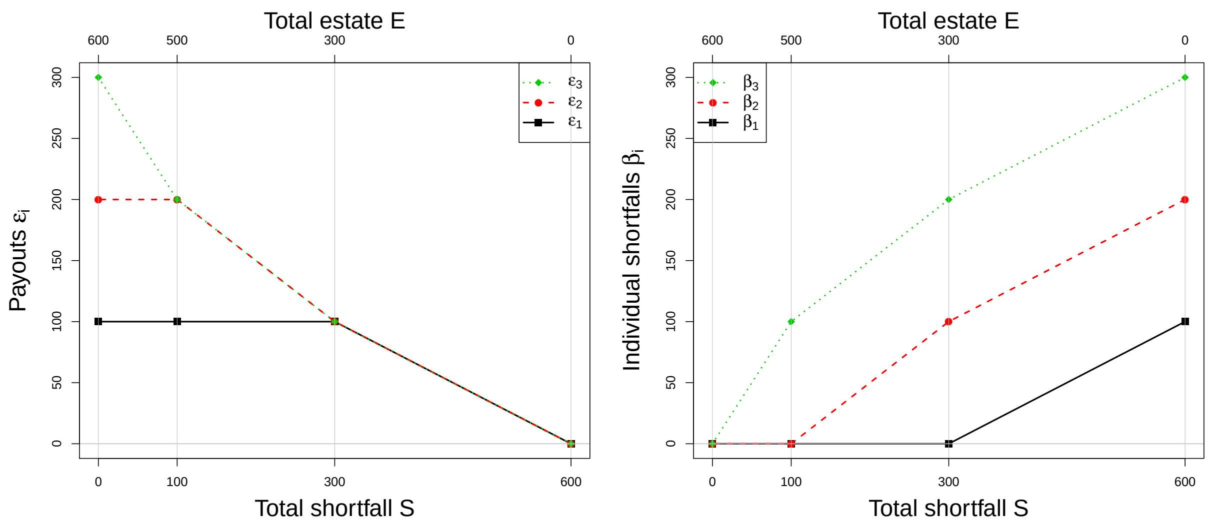

Figure 2: *Left:* CEA payouts $\varepsilon_1, \varepsilon_2, \varepsilon_3$ for the example of Section 3, as in Figure 1 but with the horizontal axis reversed. *Right:* Individual shortfalls $\beta_i = y_i - \varepsilon_i$ $(i = 1, 2, 3)$ for this example.

divisions of the total shortfall S. But what is the meaning of S in the context of regression model (10), (11)? To answer this, we consider next the LARS-lasso connection.

5 Connection with the lasso

The LARS algorithm is of interest primarily because of its connection with the lasso estimator for model (10), given by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta}(\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2) \quad \text{subject to } \|\boldsymbol{\beta}\|_1 \equiv \sum_{i=1}^{n} |\beta_i| \leq S \tag{14}$$

for some $S > 0$ (Tibshirani, 1996). While LARS and lasso do not coincide in general, Efron et al. (2004, p. 417) show that with one modification to the LARS algorithm, its solution path is the same as the set of lasso solutions as the maximum ℓ_1 norm S increases. LARS and lasso *do* coincide (without this modification) under a "positive cone condition" (Efron et al., 2004, p. 425), which holds in particular if the columns of $\boldsymbol{X}$ are orthogonal, as they are under our assumption (11). Since LARS and lasso are equivalent in our setting, we can restate Proposition 1 in terms of the lasso:

Proposition 2 *The payout vectors* $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ *produced by Algorithm 1, as* E *increases from 0 to* $\sum_{i=1}^n y_i$*, are identical to the residual vectors* $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \hat{\boldsymbol{\beta}}$ *resulting from the lasso problem* (14) *with design matrix* (11) *and* ℓ_1 *constraint parameter* $S = \sum_{i=1}^n y_i - E$ *(i.e.,* S *decreasing from* $\sum_{i=1}^n y_i$ *to 0).*

Note that the total shortfall S in a given estate division problem equals the ℓ_1 constraint parameter S in the corresponding lasso problem.

Proposition 2 leads to a geometric view of the CEA solution. For general $\boldsymbol{X}$, the contours of the sum of squared errors $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2$ are hyperellipsoids centered at $\hat{\boldsymbol{\beta}}_{OLS}$, and when such a hyperellipsoid is just large enough to touch the constraint region $\|\boldsymbol{\beta}\|_1 \leq S$, the point of contact is the lasso estimate $\hat{\boldsymbol{\beta}}$ (Tibshirani, 1996, Section 2.3). When this contact occurs at a corner of the constraint region, the result is a sparse estimate, i.e., one with some components equal to zero.

In our $\boldsymbol{X} = \boldsymbol{I}_n$ case corresponding to CEA, we have $\hat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{y}$ and the hyperellipsoidal contours are hyperspheres centered at $\boldsymbol{y}$. The lasso solution $\hat{\boldsymbol{\beta}}$ is then the point closest to $\boldsymbol{y}$ in the constraint region $\|\boldsymbol{\beta}\|_1 \leq S$. It is clear from (9) that $\hat{\boldsymbol{\beta}}$ is *not* sparse in the above sense—meaning that $\hat{\beta}_i > 0$ for each i, i.e. none of the creditors is paid in full—if and only if

$$E < ny_1. \tag{15}$$

An equivalent condition to (15) is that the point

$$(y_1 - E/n, \ldots, y_n - E/n)^T \tag{16}$$

has all positive coordinates; and this point turns out to be the orthogonal projection of $\boldsymbol{y}$ onto the nearest facet of the polytope $\|\boldsymbol{\beta}\|_1 \leq S$, and hence equals the lasso estimate $\hat{\boldsymbol{\beta}}$. By (2), condition (15) is also equivalent to

$$\sum_{i=2}^n (y_i - y_1) < S. \tag{17}$$

Figure 3 illustrates these equivalent conditions for the $n = 2$ case. The dark grey diamond is the constraint region. Condition (17) reduces here to $y_2 < y_1 + S$, so that for (y_1, y_2) in that region (shown in medium grey), the lasso estimates $\hat{\beta}_1, \hat{\beta}_2$ are the positive numbers $y_1 - E/2, y_2 - E/2$ given by (16). On the other hand, $y_2 \geq y_1 + S$ (the light grey region) yields the sparse lasso estimate $(\hat{\beta}_1, \hat{\beta}_2) = (0, S)$.

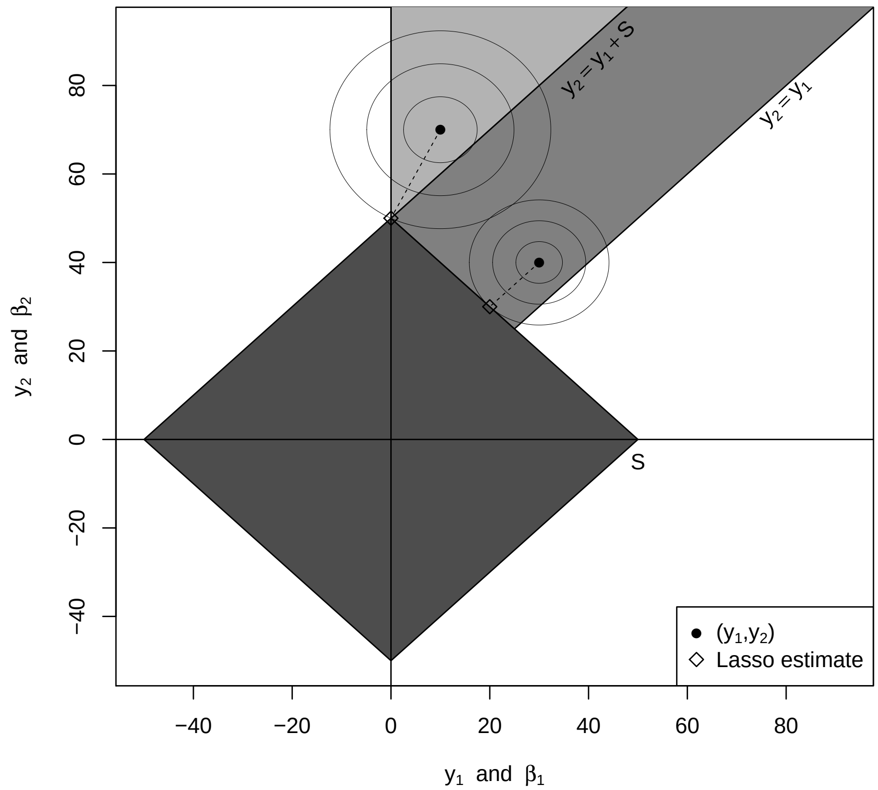

Figure 3: A geometric view of lasso estimates with $\boldsymbol{X} = \boldsymbol{I}_2$ and $0 < y_1 \leq y_2$, with axes representing both the responses y_1, y_2 and the coefficients β_1, β_2. The lasso coefficients $(\hat{\beta}_1, \hat{\beta}_2)$ are the points in the constraint region $|\beta_1| + |\beta_2| \leq S = 50$ (dark grey) nearest to a given response pair (y_1, y_2); the associated residuals give the CEA solution for claims y_1, y_2 and total shortfall $S = y_1 + y_2 - E$. For (y_1, y_2) in the $y_2 < y_1 + S$ (medium grey) region, $\hat{\beta}_1, \hat{\beta}_2$ are positive, as in the example shown with $y_1 = 30, y_2 = 40, E = 20$. For $y_2 \geq y_1 + S$ (light grey), $(\hat{\beta}_1, \hat{\beta}_2) = (0, S)$, as in the example shown with $y_1 = 10, y_2 = 70, E = 30$.

6 The "CG-consistent" division

The paper of Aumann and Maschler (1985) focuses primarily not on CEA, but on an alternative division proposed in the above-cited Talmudic passage and attributed to Rabbi Nathan. Under this scheme, with claims (5) as above, estates $E = 100, 200, 300$ are divided as shown in Table 2. Over the centuries, commentators have struggled to find a

Total estate E	$\mathcal{C}_1$ $(y_1 = 100)$	$\mathcal{C}_2$ $(y_2 = 200)$	$\mathcal{C}_3$ $(y_3 = 300)$
100	33⅓	33⅓	33⅓
200	50	75	75
300	50	100	150

Table 2: Payouts $\varepsilon_1, \varepsilon_2, \varepsilon_3$ according to the CG-consistent scheme.

common underlying rule that would explain these allocations: equal division for $E = 100$, proportional division (half payment of each claim) for $E = 300$, and neither of the above for $E = 200$.

Aumann and Maschler (1985) solve the mystery by appealing to a ruling in the Mishnah, the earlier and more concise of the two components of the Talmud. Suppose there are two competing claims on a garment: one litigant claims to own it outright while the other claims half-ownership. The Mishnah (*Bava Metzia* 1:1) rules that $\frac{3}{4}$ of the garment's value is awarded to the first and $\frac{1}{4}$ to the second. The rationale is that the second claimant tacitly concedes half the garment to the first, leaving the remaining half in dispute; splitting this half equally produces the $\frac{3}{4}$–$\frac{1}{4}$ division. Aumann and Maschler (1985) offer a general mathematical formulation of this idea, which they call the *contested garment* (CG) principle. They then show (Theorem A) that any allocation problem has a unique solution such that for each pair (i, j), the awards $\varepsilon_i, \varepsilon_j$ agree with the division of $\varepsilon_i + \varepsilon_j$ according to the CG principle; and that the allocations in Table 2 are all instances of this solution.

The key point for our purposes is that by Theorem B of Aumann and Maschler (1985), their algorithm for finding the path of "CG-consistent" solutions for given $y_1, \dots, y_n$ is closely related to Algorithm 1 for CEA. Specifically:

- When $E \leq \sum_{i=1}^{n} y_i/2$, the CG-consistent division is equal to the CEA solution for a modified problem, in which the claims $y_1, \dots, y_n$ are replaced by $y_1/2, \dots, y_n/2$.

- The solutions for $E > \sum_{i=1}^{n} y_i/2$ are such that the CG-consistent rule is *self-dual*. Intuitively, self-duality means that a rule "treats losses and awards in the same way"; formally, this means that if claims $y_1, \dots, y_n$ yield the division $\varepsilon_1, \dots, \varepsilon_n$ of estate E, then the same claims on estate $\sum_{i=1}^{n} y_i - E$ result in the division $y_1 - \varepsilon_1, \dots, y_n - \varepsilon_n$.

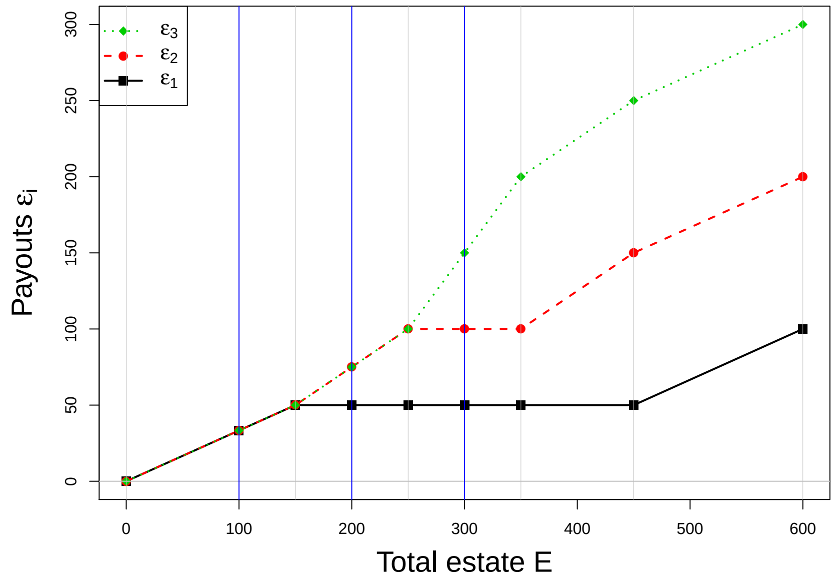

Figure 4: CG-consistent payouts $\varepsilon_1, \varepsilon_2, \varepsilon_3$. The thin vertical lines indicate steps of the algorithm, while the thicker ones mark the estate values given in Table 2.

The connection with CEA implies that the CG-consistent solutions can also be derived by LARS, or by the lasso. The result for our example appears in Figure 4. For discussion of the rationale behind self-duality and the qualitative change that occurs at $E = \sum_{i=1}^{n} y_i/2$, the reader is referred to Aumann and Maschler (1985, pp. 202–206).

7 Implementation

The supplementary material provides R code that uses standard LARS/lasso implementations to divide estates according to either CEA or the CG-consistent approach. As acknowledged in the Introduction, there is little practical need for such software, but some readers may find it helpful for understanding the connection between estate division and the lasso. The code for the CEA solution can be run in two ways, depending on whether one wishes to find the solutions for the full path of estate totals E, or for a specific E:

1. When E is not supplied by the user, the `lars` package (Hastie and Efron, 2013) is employed to obtain the full solution path, i.e., the payout vectors $\boldsymbol{\varepsilon}^0, \ldots, \boldsymbol{\varepsilon}^n$, given

by (8), for estate values (9); allocations for all other $E \in (0, \sum_{i=1}^{n} y_i)$ can be derived from these by linear interpolation.

2. When E is supplied, the `glmnet` package (Friedman et al., 2010) is used to obtain the specific solution $\varepsilon_1, \ldots, \varepsilon_n$ for that estate total. This is complicated somewhat by the fact that `glmnet` performs the unconstrained minimization

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \left(\frac{1}{2n} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|_1 \right)$$

for a penalty parameter $\lambda \geq 0$, instead of the constrained minimization (14). The code calls a generic root-finding function to derive the λ corresponding to the constraint parameter $S = \sum_{i=1}^{n} y_i - E$. (One can, again, get the solution more simply from the LARS steps by linear interpolation; the only advantage of `glmnet` here is to illustrate the connection with the lasso.)

For simplicity, the CG-consistent solution is implemented only with the `lars` package.

8 Discussion

Efron et al. (2004) describe LARS and lasso as "moderately greedy" forward stepwise procedures. We have shown that, for $\boldsymbol{X} = \boldsymbol{I}_n$, the residuals from these procedures coincide with CEA division (Algorithm 1) in reverse. If Algorithm 1 is maximally egalitarian in that it distributes equally among all n creditors, then among the $n - 1$ who have not yet been paid in full, and so on, then running this procedure in reverse, as LARS does, yields a coefficient path that is the opposite of egalitarian and thus "greedy" in a novel sense. This observation may have some pedagogical utility for instructors explaining LARS and lasso (cf. the comments of Hurley and Rickard, 2009, on the inverse relationship between equitable distribution and sparsity). Alternatively, instead of presenting LARS as assigning coefficients in an anti-egalitarian manner (namely, by the CEL rule mentioned in Section 4.4), at least for the special case $\boldsymbol{X} = \boldsymbol{I}_n$, an instructor might portray LARS as here assigning residuals in an egalitarian manner (namely, by CEA).

We have been concerned with two algorithms for "solving" the formal linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, but $\boldsymbol{\beta}$ has not denoted a parameter to be estimated and $\boldsymbol{\varepsilon}$ has not been a random vector. This paper, then, might be seen as an extreme expression of the shift from generative "data modeling" to "algorithmic modeling" that has occurred since the manifesto of Breiman (2001)—a change that is much in evidence in the text of Hastie et al. (2015) on the lasso and its extensions.

This paper has considered *how* to attain CEA and CG-consistent allocations, but has left aside the question of *why* such divisions might be viewed as just solutions. The latter question is treated, from both legal and mathematical standpoints, by O'Neill (1982), Aumann and Maschler (1985), Dagan (1996), Herrero and Villar (2001) and Chun et al. (2001). These authors also consider alternative distribution schemes.

Acknowledgments. This work was supported in part by Israel Science Foundation grant 1777/16. My sincere thanks to David Azriel and to the Editor, Associate Editor and referees for their very insightful comments on the manuscript. Thanks as well to Robert Aumann for bringing the estate division problem to my attention in 1995, while he was visiting the University of Toronto for a conference in his honour. Coincidentally, Robert Tibshirani was then also at the University of Toronto, and it was around that time that his celebrated lasso paper (Tibshirani, 1996) was accepted for publication.

SUPPLEMENTARY MATERIAL

R code for constrained equal awards: R functions `cea()` and `cg()` to find CEA and CG-consistent solutions via LARS and lasso implementations, as described in Section 7. (.R files)

References

Aumann, R. J. and M. Maschler (1985). Game theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory 36*(2), 195–213.

Breiman, L. (2001). Statistical modeling: The two cultures (with discussion). *Statistical Science 16*(3), 199–231.

Chun, Y., J. Schummer, and W. Thomson (2001). Constrained egalitarianism: A new solution for claims problems. *Seoul Journal of Economics 14*(3), 269–297.

Dagan, N. (1996). New characterizations of old bankruptcy rules. *Social Choice and Welfare 13*(1), 51–59.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics 32*(2), 407–499.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software 33*(1), 1–22.

Hastie, T. and B. Efron (2013). *lars: Least angle regression, lasso and forward stagewise*. R package version 1.2.

Hastie, T., R. Tibshirani, and M. Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

Herrero, C. and A. Villar (2001). The three musketeers: Four classical solutions to bankruptcy problems. *Mathematical Social Sciences 42*(3), 307–328.

Hurley, N. and S. Rickard (2009). Comparing measures of sparsity. *IEEE Transactions on Information Theory 55*(10), 4723–4741.

O'Neill, B. (1982). A problem of rights arbitration from the Talmud. *Mathematical Social Sciences 2*(4), 345–371.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B 58*(1), 267–288.