Supplement to "Combinatorial bootstrap inference in partially identified incomplete structural models"

(*Quantitative Economics*, Vol. 6, No. 2, July 2015, 499–529)

Marc Henry
Penn State

Romuald Méango
IFO Institute

Maurice Queyranne
Sauder School, UBC and CORE

S1. Min-cut–Max-flow theorem

The network flow problem refers to the optimal way to route quantities through a given exogenous network to meet supply, demand, and capacity constraints.

Definition S1 (Directed graph). A directed graph is a pair $G = (V, E)$, where $V = \{1, \dots, n\}$ is a finite set of points, called nodes, and E is a set of ordered pairs $e = (v^1, v^2)$ of elements of V, called arcs.

Definition S2 (Path). Fix a graph $G = (V, E)$ and a sequence $v^1, \dots, v^r$ of nodes in V. A path is a sequence of arcs $e^1, \dots, e^{r-1}$, such that $e^i = (v^i, v^{i+1})$ for each $i = 1, \dots, r-1$.

Definition S3 (Network). A network is a directed graph $G = (V, E)$ with two distinct nodes s and t designated as source and sink, respectively, such that there is at least one path between s and t, and such that each arc $e \in E$ is endowed with a positive number $c(e)$ called its capacity. A network will be denoted $N = (V, E, c)$.

A flow problem is defined on a network.

Definition S4 (Flow). A flow is an assignment of weights $f(e)$ to each arc $e \in E$ of a network $N = (V, E, c)$, which satifies the following contraints:

(1) Capacity constraint: For any $e \in E$, $0 \le f(e) \le c(e)$.

Marc Henry: marc.henry@psu.edu
Romuald Méango: meango@ifo.de
Maurice Queyranne: maurice.queyranne@sauder.ubc.ca
The authors thank David Straley for providing the data and Daniel Stubbs for computing assistance.

DOI: 10.3982/QE377

(2) Flow conservation: For any $v \in V \setminus \{s, t\}$, the sum of flows $f(e)$ on arcs $e = (v, \cdot)$ starting at v is equal to the sum of flows $f(e')$ on arcs $e' = (\cdot, v)$ ending at v.

The sum of $f(\cdot, t)$ flowing into the sink t will be denoted F and will be called the value of the flow.

The maximum flow problem is the problem of finding a flow that maximizes the sum of $f(\cdot, t)$ flowing into the sink t. It is the dual of a problem known as minimum cut.

DEFINITION S5 (Cut). A cut in a network is a node partition (S, T), i.e., $S \cup T = V$ and $S \cap T = \emptyset$, such that $s \in S$ and $t \in T$. The capacity $c(S, T)$ of a cut (S, T) is the sum of capacities $c(v, w)$ for all arcs (v, w) such that $v \in S$ and $w \in T$ (i.e., arcs leaving S).

The minimum cut problem is the problem of finding a cut in a network with minimum capacity.

Observe the following facts:

(1) By definition of a cut (S, T), the net flow across a cut, i.e., the sum of $f(v, w)$ for all arcs (v, w) such that $v \in S$ and $w \in T$ (i.e., arcs leaving S), is equal to the value of the flow, i.e., the amount F reaching the sink t.

(2) By the capacity constraints and the previous observation, the value F of the flow is smaller than the capacity of any cut.

(3) From the previous observation, it immediately follows that if there is a flow and a cut such that value F is equal to the capacity of the cut, then the flow is maximum and the cut is minimum.

We can now state the main theorem, which is from Ford and Fulkerson (1957).

THEOREM S1 (Min-cut–max-flow). *In any network, the value of the maximum flow equals the capacity of the minimum cut.*

Before we prove the theorem and present the Ford–Fuljerson algorithm to find a maximum flow, two more definitions are needed.

DEFINITION S6 (Residual network). A residual network is obtained from a network endowed with a flow by removing all arcs (v, w) such that $f(v, w) > 0$ and replacing them by an arc (w, v) with capacity equal to $f(v, w)$ and, when $f(v, w) < c(v, w)$, arc (v, w) with capacity equal to $c(v, w) - f(v, w)$.

DEFINITION S7 (Augmenting path). An augmenting path in a network endowed with a flow is a path in the residual network.

The Ford–Fulkerson algorithm to find a maximum flow in a network consists in finding augmenting paths and increasing flow along the latter. Its convergence is based on the proof below.

PROOF OF THEOREM S1. The method of proof is to show the equivalence of the following three statements:

(1) Flow F is maximum.

(2) There is no augmenting path.

(3) There exists a cut with capacity F.

By definition, if the flow is maximum, there is no augmenting path. By previous observation, if there exists a cut with capacity F, then the cut is minimal and the flow is maximum. There remains to show that if there is no augmenting path, then there is a cut with capacity F. Let S be the set of vertices v such that there is a path from s to v in the residual network. Then we have the following scenarios:

- The set S contains s.
- Since there is no augmenting path, $t \notin S$.
- All arcs e leaving S in the original network have $f(e) = c(e)$.

Hence, $(S, V \setminus S)$ is a cut with capacity F as required. □

S1.1 *A note on the validity of the bootstrap*

There seems to be a common misconception that the use of the bootstrap in partially identified settings is invalid because of the "parameter-on-the-boundary problem." We thought it useful to add a little heuristic note to dispel this misconception, recasting the issue in the terms of Problem 2, page 10 of the Supplementary Material to Bugni (2010). Consider the latter with the maximum as a choice for the criterion G. Hence we are looking to approximate the limiting distribution of

$$\varGamma_n = \sqrt{n}\max\big[\max(E_n Y_1, 0), \max(E_n Y_2, 0)\big].$$

If the approximating statistic is constructed in the manner

$$\varGamma_n^*(\text{naive}) = \sqrt{n}\max\big[\max\big(E_n^* Y_1, 0\big), \max\big(E_n^* Y_2, 0\big)\big],$$

Andrews (2000) shows that we have an invalid procedure in the sense that $\varGamma_n^*(\text{naive})$ may have a different limiting distribution from $\varGamma_n$ if the parameter is on the boundary ($EY_1 = 0$, say).

Instead, rewrite $\varGamma_n$ as

$$\begin{aligned}\varGamma_n = \max\big[&\max\big(\sqrt{n}[E_n Y_1 - EY_1] + \sqrt{n}EY_1, 0\big),\\ &\max\big(\sqrt{n}[E_n Y_2 - EY_2] + \sqrt{n}EY_2, 0\big)\big].\end{aligned}$$

Under the null hypothesis that $EY_i \leq 0$, $i = 1, 2$, EY_i may be strictly negative, in which case, for sufficiently large n, the term drops from the maximum, or is zero, in which case the remaining term is $\max(\sqrt{n}[E_n Y_i - EY_i], 0)$. Hence, heuristically, the limiting distribution is

$$\varGamma = \max\big[\max(Z_1, 0)1\{EY_1 = 0\}, \max(Z_2, 0)1\{EY_2 = 0\}\big], \tag{S1.1}$$

with Z_1, Z_2 standard normal. Using the law of iterated logarithms, (S1.1) is also the limiting distribution of

$$\begin{aligned} &\max\bigl[\max\bigl(\sqrt{n}[E_n Y_1 - EY_1], 0\bigr)1\{E_n Y_1 > \tau_n/\sqrt{n}\}, \\ &\quad \max\bigl(\sqrt{n}[E_n Y_2 - EY_2], 0\bigr)1\{E_n Y_2 > \tau_n/\sqrt{n}\}\bigr], \end{aligned} \tag{S1.2}$$

where τ_n converges to 0 at the right rate.

Since $\sqrt{n}[E_n Y_i - EY_i]$ and $\sqrt{n}[E_n^* Y_i - E_n Y_i]$ have the same limiting distribution (as also pointed out in Andrews (2000, p. 402, line 6)), (S1.2) has the same limiting distribution as its version where $\sqrt{n}[E_n Y_i - EY_i]$ is replaced by $\sqrt{n}[E_n^* Y_i - E_n Y_i]$. Then we have the bootstrap procedure proposed by Bugni (2010), which is also proposed, up to minor modification, by Chernozhukov, Hong, and Tamer (2007, Remark 4.2) and by Galichon and Henry (2009) in the context of incomplete models.

The crucial point is that $\sqrt{n}[E_n^* Y_i - E_n Y_i]$ was used to (correctly) approximate $\sqrt{n}[E_n Y_i - EY_i]$, instead of trying to approximate (incorrectly) $\max(E_n Y, 0)$ by $\max(E_n^* Y, 0)$.

In Section 3.2 of the current paper, the bootstrap approximation only relies, as does Bugni (2010), on the fact that the bootstrapped empirical process $\sqrt{n}(P_n^* - P_n)$ has the same limiting distribution as the empirical process $\sqrt{n}(P_n - P)$, in the sense of Theorem 3.6.3 of van der Vaart and Wellner (1996), for instance. This is what allows the classical bootstrapping of the Kolmogorov–Smirnov (KS) specification test, for instance, where

$$\sup_x \sqrt{n}\bigl|P_n^*(-\infty, x) - P_n(-\infty, x)\bigr|$$

has the same limiting distribution as

$$\sup_x \sqrt{n}\bigl|P_n(-\infty, x) - P(-\infty, x)\bigr|$$

as shown in Bickel and Freedman (1981, Corollary 4.2, p. 1205), for instance. Incidentally, note that the maximum that we have in Section 3.2 is like the KS supremum or like the *outer* maximum in the expression of Γ_n above, *not the inner max operators* in Γ_n (this may be the source of the common misconception). More precisely, we are only using the fact that the finite family

$$\bigl(\sqrt{n}\bigl(P_n^*(A) - P_n(A)\bigr)\bigr)_{A \subseteq \mathcal{Y}}$$

has the same joint limiting distribution as the family

$$\bigl(\sqrt{n}\bigl(P_n(A) - P(A)\bigr)\bigr)_{A \subseteq \mathcal{Y}}.$$

We also index by x, but with a finite support, so the reasoning is the same.

S2. Elderly care provision

We estimate the determinants of long term care option choices for elderly parents in American families. The model we use closely follows that proposed by Engers and Stern (2002), who present these choices as the result of a family bargaining game. The family members decide simultaneously whether to participate in a family reunion where the care option maximizing the participants' utility is chosen. Profits are then split among these participants according to some benefit-sharing rule.

The data consist of a sample of 1212 elderly Americans with two children drawn from the National Long Term Care Survey, sponsored by the National Institute of Aging and conducted by the Duke University Center for Demographic Studies under Grant U01-AG007198 (Duke (1999)). Elderly people were interviewed in 1984 about their living and care arrangements. The survey questions include gender and age of their children, the distance between homes of the elderly parent and each of the children, the disability status of the elderly parent (where disability is referred to as problems with "activities of daily living or instrumental activities of daily living" (ADL)), and the number of days per week each of the children devotes to the care of the elderly parent. The dependent variable is the care provision for the parent. The parent is asked to list children (either at home or away from home) and how much each provides help. If only one child is listed as providing significant help, that child is designated the primary care giver. If more than one child is listed, the one providing the most time is designated the primary care giver. If the elderly parent lives in a nursing home, then the nursing home is the primary care giver. If no child is listed and the parent does not live in a nursing home, then the parent is designated as "living alone." Table S1 presents the list of variables used in the analysis: they include parent characteristics, characteristics of the children, and the care option chosen.

Table S1. List of covariates.

Variables	Equal to 1 if	Percentage of Sample
Care Option	Living with child 1	26.81
	Living with child 2	6.75
	Living in nursing home	19.92
	Living home alone	46.54
Parent Variables		
DA	Highly disabled	33.81
DM	Living with the spouse	40.36
Children Variables		
$DD0$	Living with parent	11.55
$DD1$	Distance from parent: 31 min and more	49.45
DS	Female	49.26

S3. Data summary statistics

The dependent variable is the care option chosen by the family. We classify it in four categories: child 1 (resp. child 2) means that the firstborn child (resp. the second born child) provides the most care in terms of days spent helping the parents. Our data exhibit a strong dominance of the choice of child 1 over the choice child 2. Child 1 is listed as primary care giver in 26.81% of the families, while child 2 is only listed in 6.75% of our observations. On average, the older child provides 3.50 (3.12) days of care, while the second child provides 0.48 (1.34) days. When child 1 is the chosen option, the average numbers of days spent by the primary care giver climbs to 4.32 (2.92), while the other child provides on average only 0.17 (0.77) days of care. When child 2 is the preferred option, the firstborn will spend on average only 0.29 (1.29) days of care, while the second child provides 1.74 (2.15) days of care. The third care option—for the parent to enter a nursing home—represents 19.92% of the families in our sample. The remaining option—for parents to live alone—includes all the cases where no child is listed and the parent does not live in a nursing home (it could well be the case that another individual, other than the children, provides care for the parent). The parent is living alone in the remaining 46.74% of the sample, which makes it the most frequent option in our sample. We use two types of explanatory variables: those relative to the elderly parent and those relative to children. The first variable relative to the elderly parent is their disability status based on six types of daily living activities: bed, bathing, dressing, eating, toileting, walking inside. As much as 65% of the parents in the sample population suffer at least one disability, of which 51.5% have four or more. In the following discussion, we introduce a categorical variable DA, which has value 1 when the parent shows at least four disabilities and value 0 otherwise. The second variable relative to the parent characteristics is a categorical variable for the presence of the parent's spouse in the household. We denote it DM. Previous studies show a positive effect on the incentive to remain home when the parent lives with his or her spouse. In our sample, 40.36% live with a spouse. We consider three characteristics of each child: their distance to the parent, their birth order, and their gender. The distance can be viewed as a cost for a child to provide days of care to the parent. The survey measures the time required to travel from each child's residence to the parent's residence. About 50% of the children in the sample live at a travel distance of 30 minutes or more, and 12% of the children live in the same household as their parent. Among care givers, the distribution of distance is quite different. Among the firstborn children in charge of the parent, 25% live in the same household, and the percentage of those living more than 30 minutes away drops to 31%. Similarly, among primary care giver who are second born children, the percentage of those living more than 30 minutes away is only 25%. However, the number of those living in the same household as the parents is 7%. As noted by Engers and Stern (2002), the observations where the parent lives in the same household as the child induces some statistical problems. It is unclear if the child is the actual care giver for the parent or if the parent is actually the care giver for the child. To capture the effect of distance on children's incentives, we will therefore conduct estimation conditional on the children living in a different household than the parent's household. In our analysis, we include a dummy variable for each one of the children: $DD1_i = 1$ if child i lives 30 minutes away or more, and $= 0$ otherwise. The child's gender is an important variable in most of the analysis. Previous work (see, for example, Horowitz (1982), Treas, Gronvold, and Bergston (1980)) suggests that being female makes a child most likely to provide care. However, Stern (1995) and Engers and Stern (2002) indicate that the data set suffers from significant misclassification of gender (see Stern (1995) for a detailed discussion of the problems with the data set). In the specification, we introduce a random error term to account for misclassification of gender.

S4. The game

The observable choice of care option is modeled as in Engers and Stern (2002) as the outcome of a family bargaining game. Before going further, we introduce some notation. We will index family members as follows: parent, 0; firstborn child, 1; second born child, 2. The payoff to family member i, $i = 0, 1, 2$, is the sum of three terms.

The first term V_{ij} is the value to parent 0 and to child i of care option j, where $j \in 1, 2$ means child j becomes the primary care giver, $j = 0$ means the parent remains alone, and $j = 3$ means the parent is moved to a nursing home. The matrix $V = (V_{ij})_{ij}$ is known to both children and the parent. We suppose it takes the form

$$V_{ij} = \gamma_{ij} + W\beta_{ij} + Z_j\psi_{ij},$$

where W indicates the characteristics of the parents (DA and DM), Z_j indicates the characteristics of care option j (DS, $DD1$, and $DD2$), and $X = (W, Z)$. The object of inference $\theta = (\gamma_{ij}, \beta_{ij}, \psi_{ij})'$ is unknown to the analyst.

Example S1. Consider the following family, in which the matrix where the given value of X and θ result in V takes the form

$$V = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 4 & -1 & 1 \\ 0 & -1 & 4 & 1 \end{bmatrix}.$$

Rows indicate family member $i = 0, 1, 2$ and columns represents care giving options $j = 0, 1, 2, 3$, in that order. In this example, the parent is indifferent between all the care options, except the one where she has to move to the nursing home. Each child prefers to be the primary care giver to any other care option, followed by living in a nursing home, the parent living at home, and being taken care of by the other child, in that order.

The second term in the payoff results from the family bargaining process as follows. We assume that it is always in the interest of the parent to attend the family reunion. However, child i ($i = 1, 2$) can refrain from participating in the meeting. By choosing not to participate, a member of the family agrees on whatever is decided but can neither ensure the role of primary care giver nor be involved in any side payment. Both children simultaneously decide whether to participate in the long term care decision. Suppose M is the set of children who participate. The option chosen is option $j \in M \cup \{0, 3\}$, which maximizes the participants' total utility $\sum_{i \in M} V_{ij}$. It is assumed that participants abide by the decision and that benefits are then shared equally[1] among parent and children participating in the decision through a monetary transfer s_i, which is the second term in the children's payoff.

The third term ϵ_i in the payoff is a random benefit from participation, which is 0 for children who decide not to participate and is distributed according to absolutely continuous distribution $\nu(\cdot|\theta)$ for each child who does participates. All children observe the realizations of ϵ, whereas the analyst only knows its distribution.

Therefore, the payoff matrix of the participation game can be determined in the following way.

- If both children decide to participate, denoted PP, child i's payoff (for $i = 1, 2$) is

$$\begin{aligned} \Pi_i &= \varepsilon_i + w^{PP} \\ &= \varepsilon_i + \frac{1}{3} \times \left(\max_{j \in \{0,1,2,3\}} \sum_{k \in \{0,1,2\}} V_{kj} \right), \end{aligned}$$

where w^{PP} is the share of the overall benefit that each child gets when both children participate. Note that we have three participants and an equal sharing rule, which explains the term $\frac{1}{3}$.

- If neither child participates, denoted NN, child i's payoff is

$$\Pi_i = V_{ij} \quad \text{with } j \text{ such that } j = \arg\max\{V_{00}; V_{03}\}.$$

Indeed, since none of the children participates, the parent (0) picks the best option among the two available (0 and 3). We will denote this quantity w_i^{NN}.

- If only child 1 participates, (option denoted PN), child 1's payoff is

$$\begin{aligned} \Pi_1 &= \varepsilon_1 + w_1^{PN} \\ &= \varepsilon_1 + \frac{1}{2} \times \left(\max_{j \in \{0,1,3\}} \sum_{i \in \{0,1\}} V_{ij} \right), \end{aligned}$$

where w_1^{PN} is the share of the overall benefit that player 1 gets when she is the only one to participate, and child 2's payoff is

$$\Pi_2 = V_{2j} : j = \arg \max_{j \in \{0,1,3\}} \sum_{i \in \{0,1\}} V_{ij}.$$

Following previous notation, we will call this quantity w_2^{PN}.

- For option NP, w_i^{NP}, $i = 1, 2$, are defined similarly.

The payoff matrix can then be written as in Table S2.

[1] Other benefit-sharing rules can be explored. Engers and Stern (2002) study different possible rules: two Pareto optimal rules and one based on the Shapley value.

TABLE S2.

Child 1	Child 2: N	Child 2: P
N	w_1^{NN}, w_2^{NN}	$w_1^{NP}, \varepsilon_2 + w_2^{NP}$
P	$\varepsilon_1 + w_1^{PN}, w_2^{PN}$	$\varepsilon_1 + w^{PP}, \varepsilon_2 + w^{PP}$

We derive best responses (in pure strategies) $\mathrm{br}_i(s_{3-i})$ for child i to a strategy s_{3-i} played by her sibling. Then we have

$$\mathrm{br}_1(P) = \begin{cases} P & \text{if } \varepsilon_1 \geq w_1^{NP} - w^{PP}, \\ N & \text{if } \varepsilon_1 \leq w_1^{NP} - w^{PP}, \end{cases}$$

$$\mathrm{br}_1(N) = \begin{cases} P & \text{if } \varepsilon_1 \geq w_1^{NN} - w_1^{PN}, \\ N & \text{if } \varepsilon_1 \leq w_1^{NN} - w_1^{PN}, \end{cases}$$

$$\mathrm{br}_2(P) = \begin{cases} P & \text{if } \varepsilon_2 \geq w_2^{PN} - w^{PP}, \\ N & \text{if } \varepsilon_2 \leq w_2^{PN} - w^{PP}, \end{cases}$$

$$\mathrm{br}_2(N) = \begin{cases} P & \text{if } \varepsilon_2 \geq w_2^{NN} - w_2^{NP}, \\ N & \text{if } \varepsilon_2 \leq w_2^{NN} - w_2^{NP}. \end{cases}$$

These lead to the following Nash equilibria in pure strategies:

$$(N, N) \quad \Leftrightarrow \quad \varepsilon_1 \leq w_1^{NN} - w_1^{PN} \quad \text{and} \quad \varepsilon_2 \leq w_2^{NN} - w_2^{NP},$$
$$(P, P) \quad \Leftrightarrow \quad \varepsilon_1 \geq w_1^{NP} - w^{PP} \quad \text{and} \quad \varepsilon_2 \geq w_2^{PN} - w^{PP},$$
$$(N, P) \quad \Leftrightarrow \quad \varepsilon_1 \leq w_1^{NP} - w^{PP} \quad \text{and} \quad \varepsilon_2 \geq w_2^{NN} - w_2^{NP},$$
$$(P, N) \quad \Leftrightarrow \quad \varepsilon_1 \geq w_1^{NN} - w_1^{PN} \quad \text{and} \quad \varepsilon_2 \leq w_2^{PN} - w^{PP}.$$

Therefore, the equilibrium correspondence $\varepsilon \rightrightarrows G(\varepsilon|x;\theta)$ depends on the rankings of the terms $w_i^r - w_i^s$, $i \in 1, 2$, and $r, s \in \{NN, NP, PN, PP\}$.

It can also be shown that there exists a unique Nash equilibrium in mixed strategies as follows. A mixed profile $(\alpha P + (1-\alpha)N; \beta P + (1 - \beta N))$ is a Nash equilibrium if and only if

$$\alpha = \frac{\varepsilon_2 - (w_2^{NN} - w_2^{NP})}{(w_2^{PN} - w^{PP}) - (w_2^{NN} - w_2^{NP})} \quad \text{and} \quad \beta = \frac{\varepsilon_1 - (w_1^{NN} - w_1^{PN})}{(w_1^{NP} - w^{PP}) - (w_1^{NN} - w_1^{PN})},$$

the denominators are nonzero, and

$$\min\{w_1^{NN} - w_1^{PN}; w_1^{NP} - w^{PP}\} < \varepsilon_1 < \max\{w_1^{NN} - w_1^{PN}; w_1^{NP} - w^{PP}\},$$
$$\min\{w_2^{NN} - w_2^{NP}; w_2^{PN} - w^{PP}\} < \varepsilon_2 < \max\{w_2^{NN} - w_2^{NP}; w_2^{PN} - w^{PP}\}.$$

Each action profile results in a (almost surely) unique care option choice; hence for each participation shock ϵ, we can derive $G(\epsilon|X;\theta)$ as the set of probability measures on the set of care options $\{0, 1, 2, 3\}$ induced by mixed strategy profiles, which are probabilities on the set of participation profiles $\{NN, NP, PN, PP\}$.

S5. Specification

We provide estimates for two specifications of the utility matrix presented in this section.

S5.1 *Specification 1*

The first matrix of interest is

$$V(X;\theta)=\begin{bmatrix} \begin{pmatrix} \beta_{00} \\ +\beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & \begin{pmatrix} \alpha \\ +\psi_s DS_1 \end{pmatrix} & \psi_s DS_2 & 0 \\ \begin{pmatrix} \beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_1 \\ +\beta_{ac} DA \end{pmatrix} & 0 & 0 \\ \begin{pmatrix} \beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & 0 & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_2 \\ +\beta_{ac} DA \end{pmatrix} & 0 \end{bmatrix}.$$

Recall that the columns indicate the options in the order $\{0, 1, 2, 3\}$, and the rows represent each member of the family in the order parent, child 1, child 2. For example, the value the firstborn child (family member 1) living less than 30 minutes away from the parent's home attaches to the fact that she takes care of a nondisabled, nonmarried parent is measured by β_{11}, whereas for a disabled parent, it is $\beta_{11} + \beta_{ac}$. Note some implications of the model:

(1) When the parent is unmarried and has no serious disability ($DA = 0$), the children are indifferent between the option where they are not in charge of the parent.

(2) The presence of problems with ADL affects the utility of the family in two ways. First, the utility of each member is affected by the term β_{ah} if the option "live alone" is preferred. Second, if a child is chosen as primary care giver, she will bear the "cost" (under the hypothesis that $\beta_{ac} < 0$) of taking care of the parent with a disability. This cost is transferred to the nursing home if this option is the one preferred by the family.

(3) The parameter β_m associated to the variable DM measures the additional utility for all family member to choose the option "live alone" when a spouse is present in the same household.

(4) The effect of distance on children's utility is introduced in the valuation of the primary care giver. The variable ψ_d measures a cost for child i to travel (30 minutes or more) to provide care for the parent on a regular basis.

(5) As mentioned earlier, we introduce the parameter α that allows for a preference of the parent for the oldest child. The parameter α measures the incremental utility for the parent of being taken care of by the firstborn child as compared with the second born

child. This specification allows for the presence of favoritism, as defined by Li, Rosenzweig, and Zhang (2010). In their words: "favoritism exists if the parent derives more utility by spending the same time with [...] one child versus another. Such favoritism could be based on the child's endowment," the endowment here being simply birth order. The gender effect is introduced in the same manner.

S5.2 *Specification 2*

We summarize the second specification of the utilities that we study in the matrix

$$V(X;\theta)=\begin{bmatrix} \begin{pmatrix} \beta_{00} \\ +\beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & \begin{pmatrix} \alpha \\ +\psi_s DS_1 \end{pmatrix} & \psi_s DS_2 & 0 \\ \delta V_{00} & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_1 \\ +\beta_a DA \end{pmatrix} + \delta V_{03} & \delta V_{04} & 0 \\ \delta V_{00} & \delta V_{03} & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_2 \\ +\beta_a DA \end{pmatrix} + \delta V_{04} & 0 \end{bmatrix}.$$

The main difference with the first model is the introduction of the term δV_{0i} in the evaluation of option j by child i. This term measures the degree of altruism of the children. An individual shows altruism when their utility depends directly on someone else's. The children are altruistic if $\delta > 0$.

We introduce the idiosyncratic part of the utility through two terms: first, random participation benefits for each child ε_i, $i \in \{1,2\}$; second, error terms u_j, $j \in \{0,1,2,3\}$, which measure unobserved components of the family evaluation of the alternatives. We assume

$$\varepsilon_i \sim \text{i.i.d. } \mathcal{N}(\mu,\sigma_\varepsilon); \qquad u_j \sim \text{i.i.d. } \mathcal{N}(0,\sigma_u).$$

As mentioned above, we introduce in the specification a random error term to account for misclassification of the gender of children. Recall that we include a parameter δ, measuring the preference of the parent for a female caretaker, so that we have

$$V_{0,i} = V'_{0,i} + \psi_s DS_i \quad \text{for } i \in \{1,2\},$$

where $V'_{0,i}$ sums up the remaining terms entering the parent's utility. Suppose that $(DS_i)^*$ is the reported value of child i's gender, when DS_i is the true value of gender. It follows that

$$V_{0,i} = V'_{0,i} + \psi_s (DS_i)^* + \xi_i \quad \text{for } i \in \{1,2\} \text{ with } \xi_i = \psi_s\big(DS_i - (DS_i)^*\big).$$

Supposing that true gender is misreported with probability p_ξ and that measurement error is independent of true gender, we have $\xi_i = 0$ with probability $1 - p_\xi$ and $\psi_s(1-2DS_i)$ with probability p_ξ. Failing to take into account this measurement error may lead to biased estimates of the parameter ψ_s and may worsen the identification issue.

S6. Estimation methodology

The methodology proposed in the paper allows the construction of the identified set based on the hypothetical knowledge of the true distribution of the data. As described in Section 3 of the paper, we account for sampling uncertainty and control the level of confidence by constructing set functions $A \mapsto \overline{P}(A|X)$, which dominate $P(A|X)$ (uniformly over $A \subseteq \{0, 1, 2, 3\}$ and X) with probability $1-\alpha$ (the chosen level of confidence; here 0.95). We implement the method detailed in Section 3 of the paper with a number of bootstrap replications $B = 2500$. Second, we obtain the model likelihood by simulating the valuation matrix and computing the equilibrium correspondence from the payoff matrix, for given values of X and θ. The procedure for a given X and θ is as follows:

- We generate and store R draws of ε from the distribution ν_θ. Here, $R = 5000$ and ν_θ is normally distributed with mean μ and variance σ_ε^2, where $(\mu, \sigma_\varepsilon^2)$ belong to the parameter θ.
- For each value ε^r, we compute the valuation matrix $V(X, \varepsilon^r, \theta)$ and the corresponding payoff matrix.
- Then we determine the equilibrium correspondence $G(\varepsilon|X, \theta)$ from the analytical results derived in the preceding section. The Gambit software provides an alternative for computing numerically the set NE for more complex games.
- The last step of the simulation is to compute an estimator of the model likelihood $\mathcal{L}$ defined in (3.3) of the paper as $\hat{\mathcal{L}}(A|X; \theta) = \frac{1}{R}\sum_{r=1}^{R} \min_{\sigma \in G(\varepsilon^r|X;\theta)} \sigma(A)$.

Having constructed those two elements, the identified set comprises all values of θ such that for all observed values of the explanatory variables, the minimum over $A \subseteq \{0, 1, 2, 3\}$ of the function $\overline{P}(A|X; \theta) - \hat{\mathcal{L}}(A|X; \theta)$ is nonnegative, as explained in Section 3 of the paper. We construct an n-dimensional grid to conduct the search over the parameter space. Each value of the parameter can be tested in a fraction of a second on a standard laptop, and a region of small dimensionality (1–4) can be constructed in a few hours, again on a standard laptop without parallel processing. However, estimation time grows exponentially with the number of parameters induced by the model. In our case, each specification involves a 12-dimensional parameter space. Parallel processing becomes, therefore, necessary. We use an open-MP procedure for parallel processing, which is perfectly suited to the method we propose. The computation resources have been provided by the Réseau Québécois de Calcul de Haute Performance (RQCHP). All computations were made under the system Cottos, which provides up to 128 computation nodes (1024 CPU cores) equipped with two Intel Xeon E5462 quad-core processors at 3 GHz. Under one node, approximately 10^7 parameters points can be tested in 24 hours.

S7. Results

We perform the estimation of the two previously introduced specifications under different values of the mean and variance of the error term. To alleviate the computational burden, we first test the significance of some of the individual parameters by checking whether the hyperplanes defined by $\theta_i = 0$—where θ_i is a component of θ—intersect the 95% confidence region. In practice, this amounts to building a constrained confidence region under the null hypothesis. We fail to reject the null hypothesis if the estimation procedure returns a nonempty set. We then obtain a constrained confidence region for the remaining parameters.

S7.1 *Specification 1*

For each value of mean and variance of the error term, we find a nonempty intersection between the confidence region and the hyperplane defined by $\beta_{11} = 0$. This means we fail to reject (at the 5% level) the null hypothesis that there is no additional constant disutility for a child to take care of an elderly parent. Since this hypothesis is not rejected, we obtain a constrained confidence region for the remaining parameters. We then obtain confidence regions for different values of β_{11} and discuss the latter's effect on the regions. We note that the null hypothesis $H_0 : \beta_{00} = 0$ is always rejected. Hence, when we control for all other effects, parents are not indifferent between the first two options. They show a clear preference in favor of living in their own home (option called "living alone") instead of living in a nursing home (β_{00} is always positive). The results we present are then for given values of β_{00}. We provide an insight of how different values of this parameter change the results.

We report the range for each parameter in Table S3. Note that the identified set is not a compact set. In particular, β_{ac}, β_{ah}, β_m, and ψ are allowed to diverge to $-\infty$.

The following results are generally consistent with expectations and previous results on the subject.

(1) The existence of several problems with the parent's functional ability is a key determinant of the decision to enter a nursing home: β_{ah} and β_{ac} are both negative and can both be (very) large. The negative sign of β_{ah} captures the fact that a parent's disability increases the value of care provided by the family or a specialized institution.

TABLE S3. Parameters range for estimation of Specification 1 at $\beta_{11} = 0$, $\beta_{ac} = -\beta_m$ and for different values of the error terms and of β_{00}.

Parameters	Min	Max	Min	Max	Min	Max	Min	Max
β_{00}	2	2	3	3	1	1	1	1
β_{11}	0	0	0	0	0	0	0	0
β_{ah}	$-\infty$	−3.57	$-\infty$	−3.57	$-\infty$	−2.86	$-\infty$	−2.14
$\beta_{ac} = -\beta_m$	$-\infty$	−2.86	$-\infty$	−2.86	$-\infty$	−3.57	$-\infty$	−3.57
α	0.00	8.00	0.00	8.00	1.00	5.00	0.00	4.00
ψ_s	0.00	5.00	0.00	4.00	1.00	4.00	0.00	2.00
ψ_d	$-\infty$	−2.86	$-\infty$	−2.14	$-\infty$	−1.43	$-\infty$	−3.57
μ	−1	−1	0	0	−1	−1	0	0
σ_ε	1	1	1	1	1	1	1	1
σ_u^2	1	1	1	1	0.25	0.25	0.25	0.25
p_ξ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

In addition, $\beta_{ac} < 0$ means that the disability entails a utility cost for the child if he is chosen as primary care giver. In other words, holding all else equal, parental disability decreases the value for family members of being primary care giver by at least 2.86 utils. Placing the elderly parent in a nursing home transfers this cost to the institution. If we think that the expansion of life expectancy observed in past years is correlated with the appearance of more functional ability problems, this may partly explain the trend away from care provision by children toward other alternatives.

(2) Parameter β_m associated with the parent living with a spouse is positive and large. This implies that married parents are more likely to stay at home. In families where the parent is disabled, the effect of living with the spouse somehow compensates the disutility created by the disability factor and, all other variables controlled, preserves the incentive for parents to live at home.

(3) While we cannot rule out parents being indifferent to the gender or birth order of their primary care giver, estimation shows a tilt of the confidence interval toward positive values for both parameters, with a possible positive and large magnitude of the parameter α. In case $\mu = -1$ and $\sigma_u^2 = 0.25$, the data reveal that parents exhibit a preference for an older and a female care giver.

(4) Children living more than 30 minutes from the parents are less likely to provide care than those living closer to the parents. Distance has a (possibly strong) disutility effect on children's incentives to participate in the care decision. All else being equal, moving 30 minutes away from the parents reduces by at least 1.43 utils, the utility of a child when providing care.

(5) Note that $H_0 : \sigma_u^2 = 0$ is rejected. The magnitude of the unobserved idiosyncratic term u is linked to the value of the parameter β_{00}. Values of β_{00} higher than 1 are rejected for $\sigma_u^2 = 0.25$ and are the only values admissible if $\sigma_u = 1$. Recall that β_{00} measures the preference of the parent for living in their own home. Higher values of this parameter then induce higher probabilities for the choice of the option "living alone," i.e., $\mathbb{P}$(option "living alone") $\to 1$. Higher magnitude of the unobserved heterogeneity is therefore needed to make sense of the difference in choices that we observe in the data. Alternatively, the unobserved heterogeneity need not have a "large" variance when β_{00} is close to 0, which suggests that the model explains the variance in the data well.

The shape of the confidence region also conveys a considerable amount of information. In Figures S1 and S2, we plot two-dimensional and three-dimensional projections and cuts of the confidence region for column 2 of Table S3, i.e., $\mu_\varepsilon = 0$, $\sigma_\varepsilon^2 = 1$, and $\sigma_u^2 = 1$.

Of great interest is the projection of the identified set in the plane β_{ah}, β_m. Figure S2(a) reveals a linear relation between the two parameters of the type $\beta_{ah} = -\beta_m$. The estimation rejects models for which the absolute values of the two parameters are significantly different. The data suggest therefore, that the disutility induced by the disability of the parent can be entirely compensated by the presence of a spouse in the same household.

Notice the triangular shape of the region plotted in Figure S2(b) that entails the simultaneous rejection of large values of ψ_s and α. This finding means that only one of the

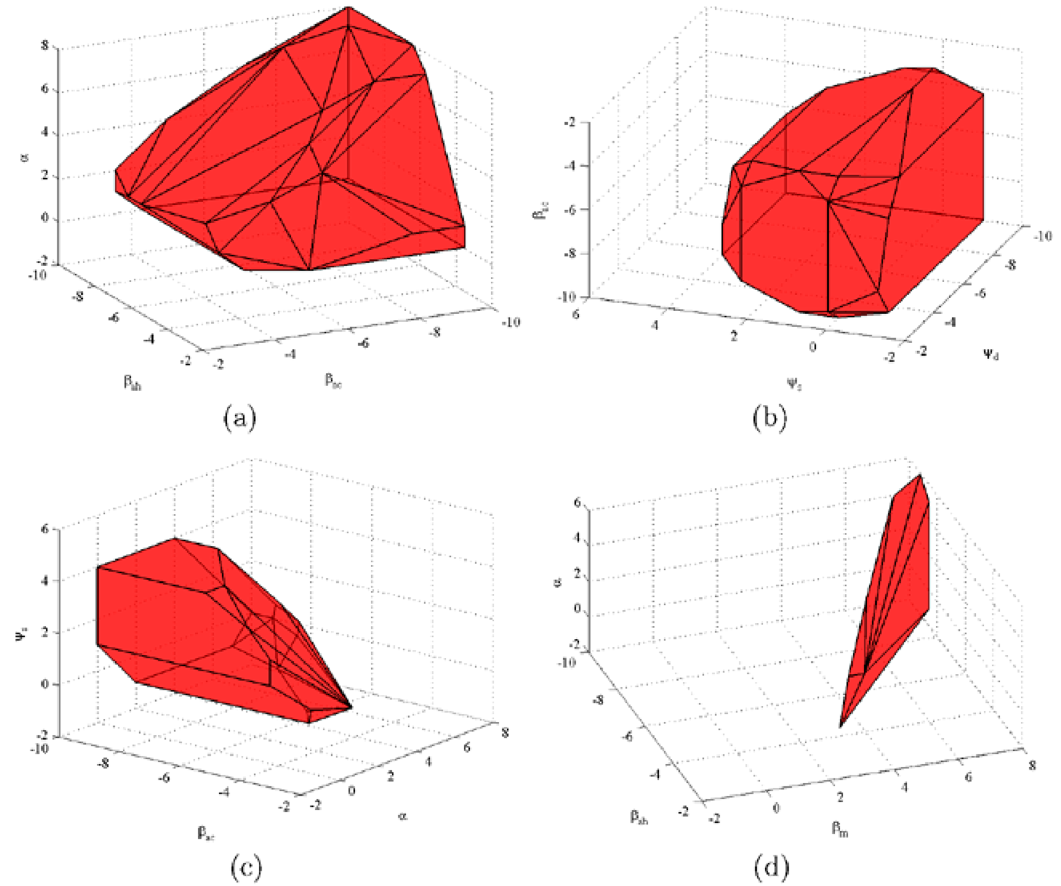

FIGURE S1. Three-dimensional representations of the confidence region for Specification 1 at $\beta_{00} = 3$, $\beta_{11} = 0$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$, $p_\xi = 0.1$.

effects (gender or birth order) can be large, not both. In other words, firstborn daughters are not the only possible care givers. Note also that both effects can be very small, though not jointly insignificant.

We observe similar types of constraints for the pairs (α, β_{ah}), (α, β_{ac}), (α, ψ_d), (ψ_s, β_{ac}), and (ψ_s, ψ_d), as large values of parameters α or ψ_s are only permitted when the other parameters are jointly large (see Figure S2(c)–(f)). For example, we obtain a constrained confidence region at $\beta_{ac} = -3.5$. The ranges for the two parameters, α and ψ_s, are tighter, as $\alpha \in [1, 2]$ and $\psi \in [0, 1]$.

Figure S3 shows the effect of the variation of parameter β_{11} on ψ_s and α. Recall that β_{11} represents a fixed cost or benefit for the child chosen as care giver. We observe negative relations between β_{11} and ψ_s, and between β_{11} and α. Negative values of ψ_s and α are only admissible for positive values of β_{11}. Hence a model where parents exhibit no favoritism for a daughter and/or a firstborn, or favoritism for a son and/or a second born, will be consistent with our data if and only if there exists a strictly positive con-

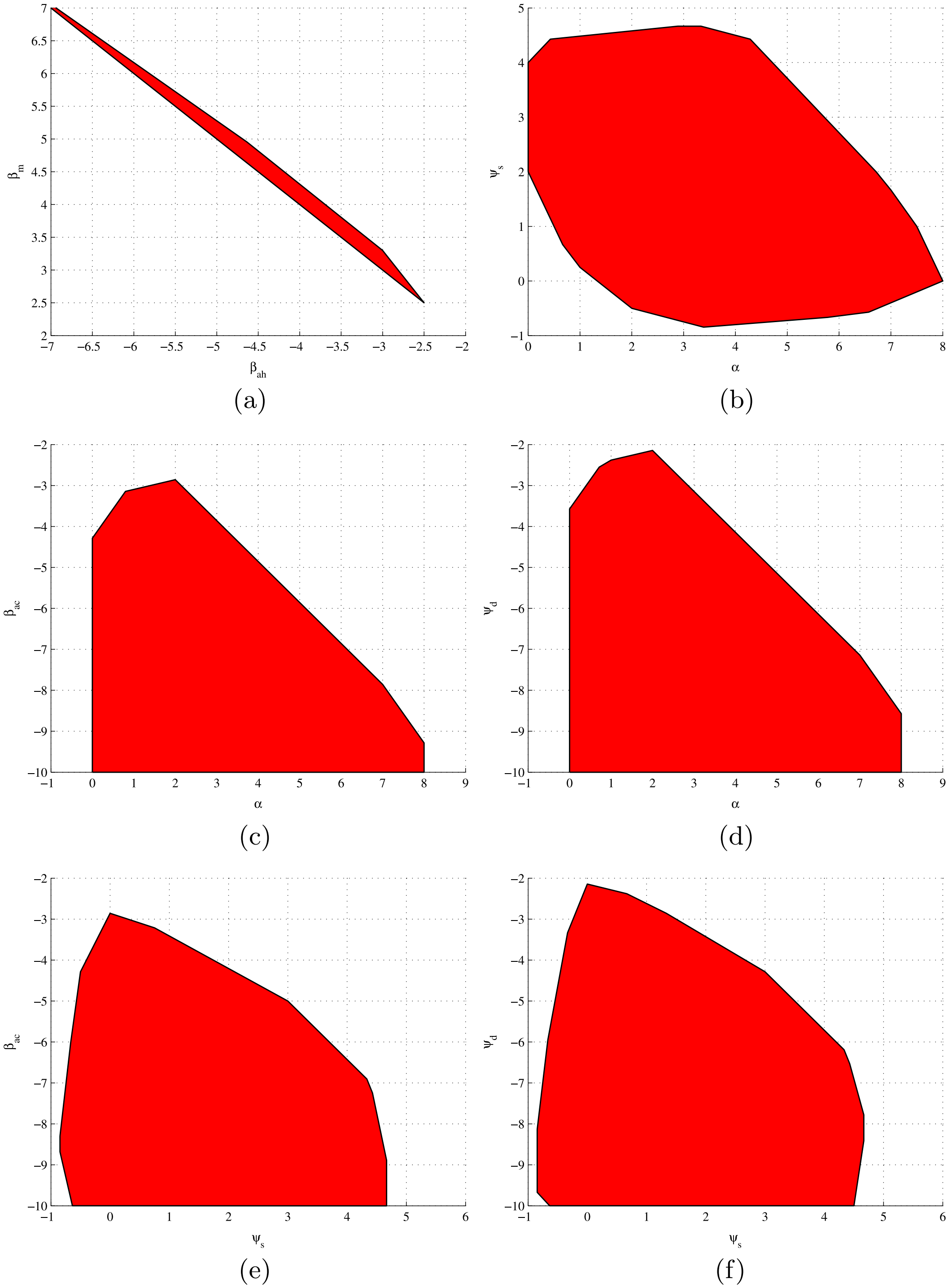

FIGURE S2. Two-dimensional representations of the confidence region for Specification 1 at $\beta_{00} = 3$, $\beta_{11} = 0$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$, $p_\xi = 0.1$.

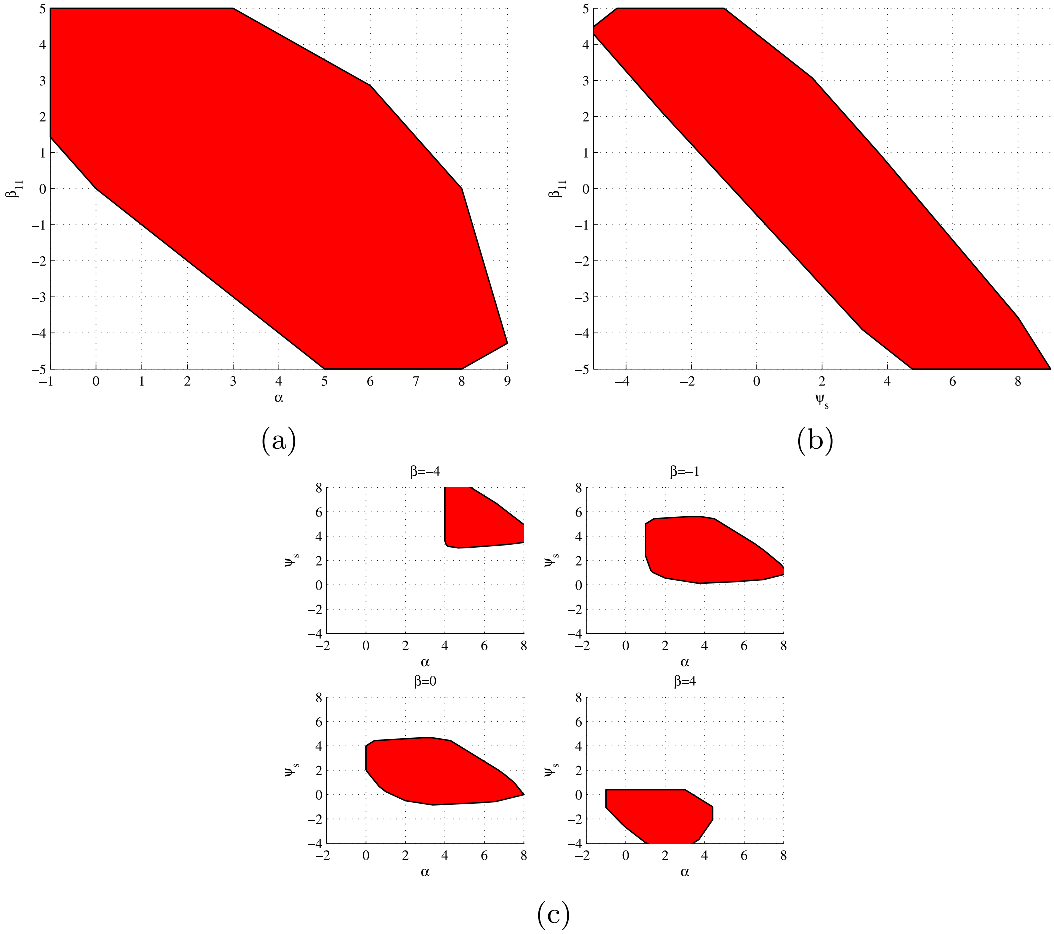

Figure S3. Parameter β_{11} in relation to other parameters: Specification 1 at $\beta_{00} = 3$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$, $p_\xi = 0.1$.

stant benefit for a child to be caregiver. In the case where $\beta_{11} \ll 0$, the null hypothesis of indifference to gender and birth order of the primary care giver is rejected, and the minimum value of both effects (gender and birth order) increases with more negative values of the fixed cost (see Figure S3(c)).

S7.2 *Specification 2*

As discussed earlier, the second specification analyzes primarily altruistic behavior on the part of the children. The parent's utility enters in the children's evaluation through the parameter δ. The greater is δ, the more the parent's preferences influence the family's decision. Notice that in the case where $\delta = 0$, both children value all options identically when they are not in charge of the parent, irrespective of the parent characteristics, in particular as pertains to disability.

TABLE S4. Parameter ranges for Specification 2 at $\beta_{11} = 0$ and $\beta_a = -\beta_m$, and for different values of the error terms.

Parameters	Min	Max	Min	Max	Min	Max	Min	Max
$\beta_{0,0}$	1.13	2.07	1.13	2.53	0.67	0.67	0.67	1.13
$\beta_{1,1}$	0	0	0	0	0	0	0	0
δ	0.11	1.00	0.11	1.00	0.00	0.67	0.00	0.78
$\beta_a = -\beta_m$	$-\infty$	-4.29	$-\infty$	-4.29	$-\infty$	-3.57	$-\infty$	-2.86
α	0.25	3.25	0.25	4.00	0.25	4.00	0.25	4.00
ψ_s	0.25	4.75	0.25	4.00	1.00	4.75	0.25	3.25
ψ_d	$-\infty$	-2.14	$-\infty$	-2.14	$-\infty$	-2.14	$-\infty$	-1.43
μ	-1	-1	0	0	-1	-1	0	0
σ_ε	1	1	1	1	1	1	1	1
σ_u^2	1	1	1	1	0.25	0.25	0.25	0.25
p_ξ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Not surprisingly, as in the previous specification, we reject $H_0 : \beta_{00} = 0$ (at the 5% level) as β_{00} is always positive, and we fail to reject (at the 5% level) the null hypothesis that there is no additional constant disutility for a child to take care of the elderly parent ($H_0 : \beta_{11} = 0$). The effects of β_{11} on the remaining parameters is the same as discussed above. We obtain constrained confidence region for the remaining parameters. We report the range for each parameters in Table S4. Note again that the identified set is not a compact set, as β_{ac}, β_{ah}, β_m, and ψ are allowed to diverge to $-\infty$.

The data seem to comply with both types of models: altruistic and nonaltruistic. However, the hypothesis of nonaltruism is rejected for some values of the error term, namely $\sigma_u^2 = 1$. That is, in the case where the variance of the unobserved heterogeneity is higher ($\sigma_u^2 = 1$ compared to $\sigma_u^2 = 0.25$), the behavior of children can only be rationalized in a framework where they act according to altruistic motives. Alternatively, high degrees of altruism ($\delta > 0.78$) are not permitted for a low variance of u. This feature can be understood by relating to the parameter β_{00}. We discussed in the previous section the fact that σ_u^2 is related to β_{00}, as large values of σ_u^2 imply large values of β_{00} (see the first line of Table S4). In Figure S4, we plot three-dimensional cuts of the confidence region. Figure S5(a) shows a projection of the confidence region in the space (β_{00}, δ). We can observe a negative dependence between the two parameters. Both parameters cannot be both very large or both very small. This suggests that a model compatible with the data is either one where children are predominantly selfish and parents have a strong preference for living alone, or one where children are fairly altruistic and parents have only a weak preference for living alone, or a middle point between the two.

In this specification, we reject the joint null hypothesis of indifference of the parent to gender and birth order of the primary care giver when there exists no constant cost or benefit associated with providing care ($\beta_{11} = 0$). The estimation suggests indeed a preference of parents for an older child and for a daughter ($\alpha > 0$ and $\psi_s > 0$). Negative values of β_{11} (i.e., existence of constant disutility for care givers) are associated with

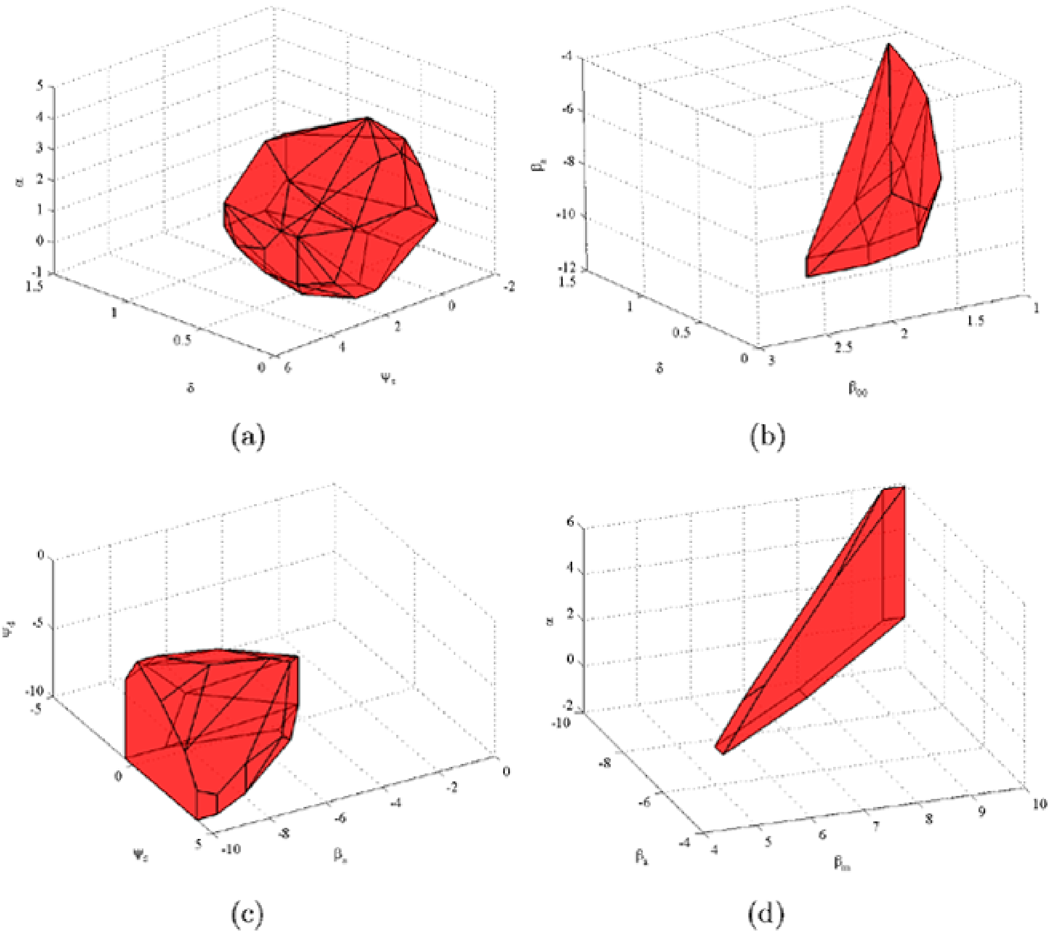

FIGURE S4. Three-dimensional representations of the confidence region for Specification 2 at $\beta_{11} = 0$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$, $p_\xi = 0.1$.

larger minimum values of α and β, hence stronger preference for firstborn and female care givers.

Confidence intervals for the other parameters do not differ significantly from the previous specification. We again reject the hypotheses that distance (between parent and caretaker) and parent's disability are insignificant, as the range for ψ_d is $\psi_d \in (-\infty, -1.43]$ and the range for β_a is $\beta_a \in (-\infty, -2.86]$. We also reject the hypothesis that living with a spouse does not increase the value for a parent to remain autonomous.

We observe very similar shapes of the confidence region as for the previous specification. Most of the analysis above still applies to the second specification. Notice that in the nonaltruistic case, i.e., $\delta = 0$, only very large effects of the parent's disability are consistent with the data (see Figure S5(b)). In other words, parental disability reduces the utility of providing care for the nonaltruistic type of children more than for the altruistic type.

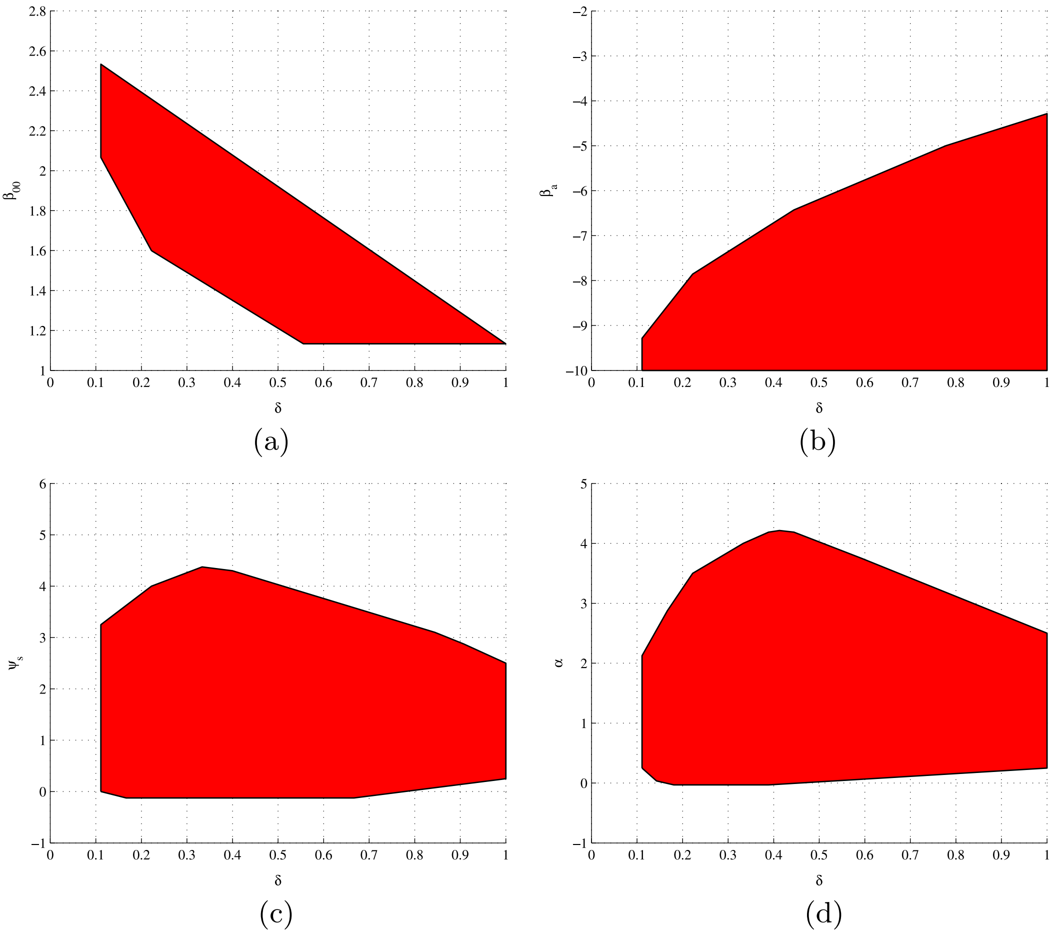

FIGURE S5. Parameter δ in relation to other parameters: Specification 2 at $\beta_{11}=0$, $\mu=0$, $\sigma_\varepsilon=1$, $\sigma_u=1$, $p_\xi=0.1$.

References

Andrews, D. (2000), "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space." *Econometrica*, 68, 399–405. [3, 4]

Bickel, P. and D. Freedman (1981), "Some asymptotic theory for the bootstrap." *The Annals of Statistics*, 9, 1196–1217. [4]

Bugni, F. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78, 735–753. [3, 4]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75, 1243–1284. [4]

Duke (1999), "National long term care survey." Public use data set produced and distributed by the Duke University Center for Demographic Studies with funding from the National Institute on Aging under Grant U01-AG007198. [5]

Engers, M. and S. Stern (2002), "Long-term care and family bargaining." *International Economic Review*, 43, 73–114. [5, 6, 7, 8]

Ford, L. and D. Fulkerson (1957), "A simple algorithm for finding maximal network flows and an application to the Hitchcock problem." *Canadian Journal of Mathematics*, 9, 210–218. [2]

Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters." *Journal of Econometrics*, 152, 186–196. [4]

Horowitz, A. (1982), "The role of families in providing care to the frail and chronically ill elderly in the community." Final report submitted to the Health Care Financing Administration, New York. [7]

Li, H., M. Rosenzweig, and J. Zhang (2010), "Altruism, favoritism, and guilt in the allocation of family resources: Sophie's choice in Mao's mass send-down movement." *Journal of Political Economy*, 118, 1–38. [11]

Stern, S. (1995), "Estimating family long term care decisions in the presence of endogenous child characteristics." *Journal of Human Resources*, 30, 551–580. [7]

Treas, J., R. Gronvold, and V. Bergston (1980), "Filial destiny? The effect of birth order on relations with aging parents." In *Annual Scientific Meeting of the Gerontological Society of America*. [7]

van der Vaart, A. and J. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer, New York. [4]

Submitted July, 2013. Final version accepted June, 2014.