

# Identification of Time-Varying Transformation Models with Fixed Effects, with an Application to Unobserved Heterogeneity in Resource Shares

Irene Botosaru, Chris Muris, and Krishna Pendakur\*

October 17, 2021

## Abstract

We provide new results showing identification of a large class of fixed- $T$  panel models, where the response variable is an unknown, weakly monotone, time-varying transformation of a latent linear index of fixed effects, regressors, and an error term drawn from an unknown stationary distribution. Our results identify the transformation, the coefficient on regressors, and features of the distribution of the fixed effects. We then develop a full-commitment intertemporal collective household model, where the implied quantity demand equations are time-varying functions of a linear index. The fixed effects in this index equal logged *resource shares*, defined as the fractions of household expenditure enjoyed by each household member. Using Bangladeshi data, we show that women’s resource shares decline with household budgets and that half of the variation in women’s resource shares is due to unobserved household-level heterogeneity.

---

\*This paper partly supersedes and replaces an earlier working paper by Botosaru and Muris entitled “Binarization for panel models with fixed effects”. Email: botosari@mcmaster.ca, muerisc@mcmaster.ca, pendakur@sfu.ca. We would like to thank Stéphane Bonhomme, P.A. Chiappori, Ian Crawford, Iván Fernández-Val, Dalia Ghanem, Bryan Graham, Hiro Kasahara, Shakeeb Khan, Valerie Lechene, Arthur Lewbel, Corrine Low, Maurizio Mazzocco, Jack Porter and Elie Tamer for useful discussions and suggestions; conference participants at the ASSA 2018, Berkeley-Stanford Jamboree 2017, IWEEE 2018, Seattle-Vancouver Econometrics Conference 2016 and 2017, and seminar participants at Cornell, CREST, Harvard, Oxford, Penn State, Queen’s University, Rotterdam, Tilburg, University of Bristol, TSE, UBC, UCL, UC Davis, UCLA, UPenn and Vanderbilt for comments and suggestions. We thank Shirleen Manzur for her research assistance on this project, and gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada under grants IDG 430-2015-00073, IG 435-2020-0026, and IG 435-2021-0778, and from the Institute for Advanced Study at the University of Bristol. This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

**JEL codes:** C14; C23; C41. **Keywords:** panel data, fixed effects, incidental parameter, time-varying transformation, collective household, full commitment, resource shares, gender inequality

# 1 Introduction

We provide sufficient conditions for point-identification in a general class of fixed- $T$  time-varying nonlinear panel models. This class has a response variable equal to a time-varying weakly monotonic transformation of a linear index of regressors, fixed effects, and error terms. In contrast, almost all existing results for this class of models require time-invariance of the transformation. Our theorems imply novel identification results for time-varying versions of some commonly used models.

Specifically, we consider models of the type:

$$Y_{it} = h_t(\alpha_i + X_{it}\beta - U_{it}), i = 1, \dots, n, t = 1, \dots, T \geq 2, \quad (1.1)$$

where  $h_t$  is an unknown weakly monotonic transformation,  $\alpha_i$  are unobserved individual-specific effects,  $X_{it}$  is a vector of strictly exogenous observed explanatory variables with coefficients  $\beta$ , and  $U_{it}$  is an error term drawn from a stationary distribution.

Our setting has the following four features:

1. fixed- $T$  —in fact  $T = 2$  is sufficient for our results;
2.  $h_t$  may be **time-varying** and **weakly** monotonic;
3.  $h_t$  and the distribution of  $U_{it}$  may be nonparametric;
4. and, the fixed effects  $\alpha_i$  are **unrestricted**.

The panel model literature is very large, with many papers showing identification of  $\beta$ , and sometimes of  $h_t$ , in models with two or three of the four features above (see, e.g., Arellano and Bonhomme (2011) for an overview). Ours is the first paper to study identification of models with these four features, all of which are demanded by our microeconomic model and empirical application. We refer to (1.1) together with the four features above as the *fixed-effects linear transformation* (FELT) model, following the terminology in Abrevaya (1999).

Our contribution is at least three-fold. First, we provide sufficient conditions for the identification of  $h_t$  and  $\beta$  for models in the FELT class. Second, for the case where  $h_t$  is strictly monotonic, we provide results on the identification of some features of the distribution of fixed effects. Third, we provide a full-commitment intertemporal

collective household model whose implied quantity demand equations lie in the FELT class, and estimate the model using Bangladeshi panel data.

We provide sufficient conditions for the point-identification of  $h_t$  and  $\beta$  in (1.1) for two non-nested cases: one where  $U_{it}$  is drawn from an arbitrary but stationary distribution, and one where it is drawn from the logistic distribution. Our approach provides a systematic way of analyzing models nested in the FELT class. An immediate implication of our work is that extensions to time-varying and/or nonparametric counterparts of well-known models can now be shown to be identified. For example, the following models are all nested in the FELT class and are now identified: ordered choice with time-varying thresholds; censored regression with time-varying censoring points; the multiple-spell generalized accelerated failure-time (GAFT) duration model; and the Box-Cox panel model with time-varying parameters.

For the case where  $h_t$  is strictly monotonic we provide additional identification results for features of the distribution of fixed effects  $\alpha_i$ . For the case where  $U_{it}$  is drawn from an arbitrary but stationary distribution, we identify the conditional mean (up to location) and the conditional variance of the distribution of  $\alpha_i$ . For the case where it is drawn from the logistic distribution, we identify all finite moments of the distribution of  $\alpha_i$ . To the best of our knowledge, there are no results in the fixed- $T$ , fixed-effects, nonlinear panel literature that cover this aspect of model identification. We also sketch the identification of marginal effects for the case when  $h_t$  is invertible.

Our theoretical work builds on established results from Doksum and Gasko (1990) and Chen (2002) who show that cross-sectional transformation models can in general be *binarized* into a set of related binary choice models. In this paper, we show that we can binarize in a panel setting when the transformation  $h_t$  varies in an arbitrary way over time.

The key innovation underlying our theoretical work is *time-varying* binarization. For an arbitrary threshold  $y_t$ , we define the following binary random variable:

$$\begin{aligned} D_t(y_t) &\equiv 1 \{Y_t \geq y_t\} \\ &= 1 \{U_t \leq \alpha + X_t\beta - h_t^-(y_t)\}, \end{aligned} \tag{1.2}$$

where  $h_t^-$  is the generalized inverse of  $h_t$  and where the equality follows from specification (1.1) and weak monotonicity of  $h_t$ . Varying the threshold  $y_t$  in (1.2) across time periods converts a FELT model into a collection of binary choice models. This conver-

sion is what we call time-varying binarization.<sup>1</sup> However, to the best of our knowledge, previous papers that used binarization in a panel setting, e.g., Chen (2010a), Chernozhukov et al. (2020), restrict the thresholds to be equal across time periods, i.e.  $y_t = y$  for all  $t$ . Essentially, it is the relaxation of this restriction that enables us to show identification of time-varying transformations  $h_t$ .

Once a FELT model has been converted into a collection of binary choice models via time-varying binarization, we invoke Chamberlain (1980) and Manski (1987) to show identification of the resulting binary choice models. We then re-assemble the identified models to obtain identification of  $h_t$  and  $\beta$  in the FELT model. Omitting the fact that any FELT model can be transformed into *many* binary choice models obtains identification of  $\beta$  only.

We provide a full-commitment intertemporal collective household model that implies time-varying quantity demand functions of the form (1.1). The nonlinear quantity demand functions in our model are time-varying because quantity demands depend on prices, and prices are unobserved but vary across the waves of our panel. The fixed effects in our economic model have a clear interpretation: they are logged *resource shares*, defined as the fractions of total household expenditure consumed by each of its members. Resource shares are not directly observable, but are important because unequal resource shares across household members signal within-household inequality. Our econometric results then imply that the quantity demand equations, the mean (up to location), and the variance of the resource shares are all identified. This is useful because it allows us to characterize the variation of, or inequality in, resource shares.

Our intertemporal collective household model —along with the identification results above— permits the use of short panel data to study resource shares within households. Previous cross-sectional methods to measure resource shares have imposed the identifying restriction that resource shares do not vary with household budgets, e.g., Dunbar et al. (2013), and have relied on a random-effects model for unobserved heterogeneity in resource shares, e.g., Dunbar et al. (2021). Our framework relaxes both these restrictions, and shows how panel data can enrich the study of resource shares. Ours are the first empirical estimates of a full-commitment collective household model in a short panel, and we demonstrate the importance of accounting

---

<sup>1</sup>Muris (2017) uses time-varying binarization in a panel ordered logit model where the transformation is time-invariant and parametric.

for both observed and unobserved heterogeneity in women’s resource shares.

Using a two-period Bangladeshi panel dataset on household expenditures, we show that less than half of the variation in women’s resource shares can be explained by observed covariates. This means that there is much more inequality within households than would be suggested by variation in observed factors. We also find that women’s resource shares are negatively correlated with household budgets. This means that women in poorer households have larger resource shares, and are therefore less poor than their household budgets would suggest. Further, that resource shares are found to covary with household budgets suggests caution in using the cross-sectional identifying restriction of independence suggested by Dunbar et al. (2013).

Section 2 provides a review of the related literature. In Sections 3 and 4, we provide our main identification results. Section 5 introduces our collective household model, which uses data described in Section 6. We present estimates of women’s resource shares in Section 7. All proofs, descriptive statistics for the data, additional robustness results and estimation details are in the Appendix.

## 2 Existing Literature

We describe in detail how we connect to, first, the econometrics literature, and, second, to the literature on collective household models.

### 2.1 Fixed- $T$ Nonlinear Panel Models with Fixed Effects

The literature on panel models is vast. Despite the vastness of this literature, we are not aware of any paper that delivers all four features discussed above. Below, we highlight the key differences between our approach and approaches in the literature that *lack* one or more of our key features.

**Feature 1: We show identification in fixed- $T$  panel models.** The incidental parameter problem occurs in fixed-effect panel models with a finite number of time periods, see Neyman and Scott (1948). Ours is a fixed- $T$  approach, with  $n \rightarrow \infty$  and works even if  $T = 2$ . A large literature analyzes the behavior of fixed effects procedures under the alternative assumption that the number of time periods goes to infinity, e.g., Hahn and Newey (2004), Arellano and Hahn (2007), Arellano and Bonhomme (2009), Fernández-Val (2009), Fernández-Val and Weidner (2016), and

Chernozhukov et al. (2020). In this setting, it is generally possible to identify each fixed effect, and consequently, the distribution of fixed effects. In our model, we show identification of specific moments of this distribution even though the number of time periods is fixed.

**Feature 2: We allow weakly monotonic time-varying transformation  $h_t$ .** Abrevaya (1999) provides a consistent estimator of  $\beta$  (the “leapfrog” estimator) in the FELT model under the restriction that the transformations are *strictly* monotonic.<sup>2</sup> Abrevaya (2000) considers a model that allows for *weak* monotonicity but restricts the transformations to be time-invariant (and allows for nonseparable errors). He provides a consistent estimator for  $\beta$  only.<sup>3</sup> A literature on duration models also considers time-invariant transformations that are weakly monotonic due to censoring, e.g., Lee (2008), Khan and Tamer (2007), Chen (2010a,b); Chen and Zhou (2012); Chen (2012), and Chen and Zhou (2012); we review this below.

A more recent literature has focused on identification issues in a class of panel models with potentially *non-monotonic* but time-invariant structural functions (or strong assumptions on how those functions vary over time), e.g., Hoderlein and White (2012), Chernozhukov et al. (2013), Chernozhukov et al. (2015). These papers focus on (partial) identification of partial effects, and the approaches employed there preclude identification of the structural function(s) or of the distribution of fixed effects.

**Feature 3: We allow nonparametric transformations and nonparametric errors.** Bonhomme (2012) proposes a general-purpose likelihood-based approach to obtain identification for models with parametric  $h_t$  and parametric  $U_{it}$ , even allowing for dynamics. Our model requires strictly exogenous regressors, precluding many dynamic structures. But, our Theorems 1 and 2 apply even when  $h_t$  is nonparametric,  $U_{it}$  is nonparametric, or both are nonparametric.

The setting with parametric transformations and parametric errors covers many models previously shown to be identified, including the *time-invariant* fixed-effect panel versions of: binary choice (e.g., Rasch (1960), Chamberlain (1980), Magnac (2004), and Chamberlain (2010)); the linear regression model with normal errors;

---

<sup>2</sup>Chen (2010c) in Remark 6 discusses a version of Abrevaya (1999) that allows for some weak monotonicity due to censoring. He focuses on  $\beta$  and does not discuss identification of  $h_t$ , although his Remark 1 sketches an approach for estimation of  $h_t = h$  for all  $t$ .

<sup>3</sup>Chernozhukov et al. (2020) uses a distribution regression technique that is closely related to our binarization approach, and consequently accommodates weakly monotonic transformations. However, theirs is a large- $T$  setting.

and the ordered logit model (e.g., Das and van Soest (1999); Baetschmann et al. (2015); Muris (2017)). Application of our results immediately shows identification of the time-varying versions of these models. This result is novel for the ordered logit model, where our results imply identification of time-varying thresholds.

Parametric transformation models with nonparametric errors are widely studied, starting with Manski (1987) for the binary choice fixed effects model. (Aristodemou (2021) provides partial identification results for ordered choice with nonparametric errors.) Parametric panel data censored regression models also fit into our framework, and were studied intensively starting with Honoré (1992), see also, e.g., Charlier et al. (2000), Honoré and Kyriazidou (2000), Chen (2012). These papers show identification of  $\beta$  for the linear model with time-invariant censoring and nonparametric errors. In this context, our results show identification of models that were not previously known to be identified. In particular, the model is identified even if the transformation is nonparametric (as opposed to linear or Box-Cox) and time-varying and/or where the censoring cutoff is time-varying.<sup>4</sup>

Duration models can be recast as transformation models with nonparametric transformations (see Ridder (1990)). Consequently, the large literature on identification of duration models is related to our work.

Consider the multiple-spell mixed proportional hazards (MPH) model with spell-specific baseline hazard, analyzed in Honoré (1993).<sup>5</sup> This model can be obtained from FELT by letting (i)  $h_t^{-1}(v) = \log \left\{ \int_0^v \lambda_{0t}(u) du \right\}$ , where  $\lambda_{0t}$  is the baseline hazard for spell  $t$ , (ii)  $\alpha_i$  and  $U_{it}$  are independent across  $t$ , and (iii)  $U_{it}$  is independent of  $X_i$  and distributed as EV1. Honoré (1993) derives sufficient conditions for the identification of this model (Lee (2008) provides consistent estimators under other parametric error distributions). Our theorems immediately provide the novel result that this model is identified when the error terms are drawn from a nonparametric distribution.

Consider the single-spell generalized accelerated failure time (GAFT) model in-

---

<sup>4</sup>Many papers in the literature on censored regression have focused on endogeneity. For example, Honoré and Hu (2004) allow for endogenous covariates, and Khan et al. (2016) study the case of endogenous censoring cutoffs. Our results do not cover the case of endogenous regressors or cutoffs. Horowitz and Lee (2004) and Lee (2008) consider dependent censoring, where the censoring cutoff depends on observed covariates and the error term follows a parametric distribution. We do not consider dependent censoring.

<sup>5</sup>Horowitz and Lee (2004) show identification of this model under the restriction that the baseline hazard is the same for all spells, analogous to time-invariant  $h_t$ . Chen (2010b) considers the same model, but relaxes the restriction that errors are type 1 EV, but shows identification of only the common parameter vector  $\beta$ .

roduced by Ridder (1990) (see also van den Berg (2001)) that has non EV1 errors, and is consistent with a duration model. Just like the MPH model, it can be extended to a multiple-spell setting, see, e.g., Evdokimov (2011). Abrevaya (1999) shows that  $\beta$  in the multiple-spell GAFT model is consistently estimated. However, he does not show identification of the transformation  $h_t$ , which can be seen as dual to identification of the spell-specific baseline hazard function.<sup>6</sup> Evdokimov (2011) considers identification of a related version of the multiple-spell GAFT with spell-specific baseline hazard, but requires continuity of  $\alpha_i$  and at least 3 spells ( $T \geq 3$ ). Our results show identification of both  $\beta$  and  $h_t$  in the multiple-spell GAFT model, imposing no restrictions on  $\alpha_i$  and requiring just 2 spells ( $T = 2$ ).

**Feature 4: We allow for unrestricted fixed effects.** A related literature considers restrictions on the joint distribution of  $(\alpha_i, X_{i1}, \dots, X_{iT})$ . For example, Altonji and Matzkin (2005) impose exchangeability on this joint distribution, and Bester and Hansen (2009) restricts the dependence of  $\alpha_i$  on  $(X_{i1}, \dots, X_{iT})$  to be finite-dimensional. In our model this joint distribution is unrestricted.

A further group of papers establishes identification of panel models, including the distribution of  $\alpha_i$ , by using techniques from the measurement error literature that: (i) impose various assumptions on  $\alpha_i$ , such as full support and/or continuous distribution; (ii) assume serial independence of  $U_{it}$ ; and (iii) restrict the conditional distribution of  $(\alpha_i, U_{i1}, \dots, U_{iT})$  conditional on  $(X_{i1}, \dots, X_{iT})$ , see, Evdokimov (2010), Evdokimov (2011), Wilhelm (2015), and Freyberger (2018). In contrast, our results on the identification of the conditional variance of  $\alpha_i$  do not require (i). All our other results, including identification of the dependence of  $\alpha_i$  on observed covariates, are free of assumptions like (i), (ii) and (iii).

Finally, special regressor approaches (see the review in Lewbel (2014)) have identifying power in transformation models with fixed effects. They require the availability of a continuous variable that is independent of the fixed effects. With such a variable, one can show identification of transformation models in the cross-sectional case (Chiappori et al. (2015)) and in the panel data case, e.g., Honoré and Lewbel (2002), Ai and Gan (2010), Lewbel and Yang (2016), Chen et al. (2019). Our results do not invoke a special regressor. Further, we are not aware of any special regressor-based

---

<sup>6</sup>Khan and Tamer (2007) establish consistency of an estimator of the regression coefficient in GAFT under the restriction that the baseline hazard is the same for all spells, analogous to time-invariant  $h_t$ .



papers that identify time-varying transformations or the distribution of fixed effects.<sup>7</sup>

**We also show identification of some aspects of the distribution of fixed effects.** Correlated random effects models identify the distribution of individual effects, but at the cost of restricting their distribution. To our knowledge, we are the first to show identification of moments of this distribution in a nonlinear panel model, when that distribution is unrestricted.

We show the practical importance of these innovations in our empirical work below. Identification of the conditional mean and variance of the distribution of fixed effects in a context with time-varying transformations is essential to our investigation of women’s access to household resources in rural Bangladeshi.

## 2.2 Microeconomic Models of Collective Households

Dating back at least to Becker (1962), collective household models are those in which the household is characterized as a collection of individuals, each of whom has a well-defined objective function, and who interact to generate household level decisions such as consumption expenditures. *Efficient collective household* models are those in which the individuals in the household are assumed to reach the (household) Pareto frontier. Chiappori (1988, 1992) showed that, like in earlier results in general equilibrium theory, the assumption of Pareto efficiency is very strong. Essentially, it implies that the household can be seen as maximizing a weighted sum of individual utilities, where the weights are called *Pareto weights*. This in turn implies that the household-level allocation problem is observationally equivalent to a decentralized, person-level, allocation problem.

In this decentralized allocation, each household member is assigned a *shadow budget*. They then demand a vector of consumption quantities given their preferences and their personal shadow budget, and the household purchases the sum of these demanded quantities (adjusted for shareability/economies of scale and for public goods within the household). For the special case of an assignable good, which is demanded by a single known household member, the household purchases exactly what that person demands given their shadow budget.

---

<sup>7</sup>We conjecture that the existence of a special regressor would be sufficient to identify time-varying nonparametric transformations, and, with strict monotonicity, the distribution of fixed effects. However, we think that a setting with completely unrestricted fixed effects is useful in a variety of empirical applications, including our own.

*Resource shares*, defined as the ratio of each person’s shadow budget to the overall household budget, are useful measures of individual consumption expenditures. If there is intra-household inequality, these resource shares would be unequal. Consequently, standard per-capita calculations (assigning equal resource shares to all household members) would yield invalid measures of individual consumption and poverty (see, e.g., Dunbar et al. (2013)). In this paper, we show identification of the conditional mean (up to location) and conditional variance of the distribution of resource shares in a panel data context.

There are many ways to identify resource shares with cross-sectional data. A common identifying assumption (used by, e.g., Dunbar et al. (2013)) is that resource shares are independent of household budgets in a cross-sectional sense. This identifying restriction has been used to estimate resource shares, within-household inequality and individual-level poverty in many countries (Dunbar et al. (2013) and Dunbar et al. (2021) in Malawi; Bargain et al. (2014) in Cote d’Ivoire; Calvi (2019) in India; Vreyer and Lambert (2016) in Senegal; Bargain et al. (2018) in Bangladesh). In our model, we show identification of the response of the conditional mean of resource shares to observed covariates, even if resource shares are correlated with (lifetime) household budgets. Consequently, we can test this identifying restriction.

Dunbar et al. (2013) does not accommodate unobserved heterogeneity in resource shares. Two newer papers, Chiappori and Kim (2017) and Dunbar et al. (2021) consider identification in cross-sectional data with unobserved household-level heterogeneity in resource shares. Like Chiappori and Kim (2017) and Theorem 1 in Dunbar et al. (2021), our work investigates identification of the distribution of resource shares up to an unknown normalization. However, the results in those papers are of the random effects type. That is, the authors impose the restriction that the conditional distribution of resource shares is independent of the household budget. In this paper, we consider a panel data setting with household-level unobserved heterogeneity in resource shares, without any restriction on the distribution of resource shares. Further, we show sufficient conditions for identification of the conditional variance of (logged) resource shares.

The literature cited above considered one-period micro-economic models. But, many interesting questions about households, and the distribution of resources within households, are dynamic in nature. For example: how do household members share risk?; how do household investments relate to individual consumption?; how can we

use information from multiple time periods to estimate resource shares when there is unobserved household-level heterogeneity?

Chiappori and Mazzocco (2017) review the literature on collective household models in an intertemporal setting. These models generally come in two flavors—limited commitment or full commitment—depending on whether or not the household can commit to a *permanent* Pareto weight at the moment of household formation. Full-commitment models answer “yes”, and limited-commitment model answer “no”. Limited commitment models have commanded the most theoretical attention. Much effort has gone into testing the full-commitment model against a limited-commitment alternative, e.g., Ligon (1998); Mazzocco (2007); Mazzocco et al. (2014); Voena (2015).

Fewer papers study the identification of Pareto weights or resource shares in an intertemporal context. Lise and Yamada (2019) use a long panel of Japanese household consumption data to estimate how Pareto weights (which are dual to resource shares) depend on observed covariates and on unanticipated shocks. Their model does not allow for correlated unobserved heterogeneity, and they require many observations for each household (so as to see when Pareto weights change). They find evidence that Pareto weights do change, so that the full-commitment model does not hold in Japan.

Our model is one of full-commitment, allows for correlated unobserved household-level heterogeneity, and is identified in a short (e.g., 2 period) panel. So we provide a complement to the approach of Lise and Yamada (2019) for cases where the data are not rich enough to estimate a limited-commitment model. The key cases here are where the panel is too short to see changes in Pareto weights, or where the observed covariates leave too much room for unobserved heterogeneity. The cost of covering these cases is the assumption of full commitment.

Full commitment models are more restrictive, but may be useful nonetheless. Chiappori and Mazzocco (2017) write “In more traditional environments (such as rural societies in many developing countries), renegotiation may be less frequent since the cost of divorce is relatively high, threats of ending a marriage are therefore less credible, and noncooperation is less appealing since households members are bound to spend a lifetime together.” We use a full commitment setting to estimate resource shares for rural Bangladeshi households.

In this paper, we adapt the general full-commitment framework of Chiappori and Mazzocco (2017) to the scale economy and sharing model of Browning et al. (2013).

Then, like Dunbar et al. (2013) do in their cross-sectional analysis, we identify resource shares on the basis of household-level demand functions for assignable goods. In our general model, observed household-level quantity demand functions for assignable goods depend on resource shares, and resource shares depend a time-invariant factor (a fixed effect) representing the initial (and permanent) Pareto weights of household members.

We then provide a parametric form for utility functions that results in demand equations for assignable goods that are nonlinear in shadow budgets, and have logged shadow budgets that are linear in logged household budgets and a fixed effect. Further, demand equations are time-varying because prices vary over time. Such demand equations fall into the FELT class, and are therefore identified in our short-panel setting. The parametric model also gives meaning to the fixed effect: it equals a logged resource share, so its distribution is an economically interesting object. So, our microeconomic theory demands an econometric model that allows for time-varying transformations and that can identify moments of the conditional distribution of fixed effects.

### 3 Identification

Dropping the  $i$  subscript, let  $Y = (Y_1, \dots, Y_T)'$  and  $X = (X'_1, \dots, X'_T)'$ . We then rewrite FELT as a latent variable model:

$$\begin{aligned} Y_t &= h_t(Y_t^*) = h_t(\alpha + X_t\beta - U_t), \\ U_t|\alpha, X &\sim F_t(u|\alpha, X), \end{aligned} \tag{3.1}$$

and denote the supports of  $Y_t$ ,  $Y_t^*$ ,  $X_t$  by  $\mathcal{Y} \subseteq \mathbb{R}$ ,  $\mathcal{Y}^* = \mathbb{R}$ , and  $\mathcal{X} \subseteq \mathbb{R}^K$ , respectively.<sup>8</sup>

We provide sufficient conditions for identification of  $(\beta, h_t)$ .<sup>9</sup> We consider two non-nested cases. The first case does not impose parametric restrictions on the distribution of  $U_t$ , requiring only that it is conditionally stationary. In this case, the idiosyncratic errors may be serially dependent and heteroskedastic. The second case

---

<sup>8</sup>The supports may be indexed by  $t$ .

<sup>9</sup>The results in this section were previously circulated in the working paper Botosaru and Muris (2017). That paper also introduces four estimators, depending on whether the outcome variable is discrete or continuous, and on whether the stationary distribution of the error term is nonparametric or logistic. In this paper, we use a GMM estimator instead.

assumes that  $U_t$ ,  $t = 1, \dots, T$ , are serially independent, standard logistic, and strictly exogenous. It may appear that the second case is a special case of the first. However, the second case requires weaker assumptions on the distribution of the regressors (c.f. Assumption 3 below) while imposing stronger assumptions on the error distribution. For both cases, we maintain the assumption below:

**Assumption 1.** [*Weak monotonicity*] For each  $t$ , the transformation  $h_t : \mathcal{Y}^* \rightarrow \mathcal{Y}$  is unknown, non-decreasing, right continuous, and non-constant.

This assumption allows us to work with the generalized inverse  $h_t^- : \mathcal{Y} \rightarrow \mathcal{Y}^*$ , defined as:

$$h_t^-(y) \equiv \inf \{y^* \in \mathcal{Y}^* : y \leq h_t(y^*)\},$$

with the convention that  $\inf(\emptyset) = \inf(\mathcal{Y})$ .

### 3.1 Identification strategy: time-varying binarization

It is well-known (see e.g. Doksum and Gasko (1990) and Chen (2002)) that cross-sectional transformation models can be *binarized* into a set of binary choice models. Binarization has been used previously in panel settings (see e.g. Chen (2010a), Chernozhukov et al. (2020)), but those approaches have restricted the threshold to be equal across time periods. An exception is Muris (2017), who uses time-varying thresholds in a panel ordered logit model with a time-invariant and parametric transformation.

We now describe *time-varying binarization*. For an arbitrary threshold  $y_t \in \underline{\mathcal{Y}} \equiv \mathcal{Y} \setminus \inf \mathcal{Y}$ ,<sup>10</sup> we define the following binary random variable:

$$\begin{aligned} D_t(y_t) &\equiv 1 \{Y_t \geq y_t\} \\ &= 1 \{U_t \leq \alpha + X_t\beta - h_t^-(y_t)\}, \end{aligned} \tag{3.2}$$

where the equality follows from specification (1.1) and weak monotonicity of  $h_t$ . Varying the threshold  $y_t$  in (3.2) across time periods converts any FELT model into a collection of binary choice models.

Two time periods are sufficient for our identification results, so we let  $T = 2$  in what follows. Consider the vector of binary variables

$$D(y_1, y_2) \equiv (D_1(y_1), D_2(y_2)),$$

---

<sup>10</sup>We use  $\underline{\mathcal{Y}}$  instead of  $\mathcal{Y}$  because  $D_t(\inf \mathcal{Y}) = 1$  almost surely for all  $t$ .

for any two points  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ . Our identification strategy for  $(\beta, h_1, h_2)$  is based on the observation that  $D(y_1, y_2)$  follows a fixed effects binary choice model for *any*  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ . This result is summarized in Lemma 1 below.

The proof of identification proceeds in three steps. First, we show identification of  $\beta$  and of  $h_2^-(y_2) - h_1^-(y_1)$  for arbitrary  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ . In the resulting binary choice model relating  $D_t(y)$  to  $X$ , the difference  $h_2^-(y_2) - h_1^-(y_1)$  is the coefficient on the differenced time dummy, and  $\beta$  is the regression coefficient on  $X_2 - X_1$ . For a given binary choice model, identification of  $\beta$  and of  $h_2^-(y_2) - h_1^-(y_1)$  follows Manski (1987) for the nonparametric version of FELT, and Chamberlain (2010) for the logistic version. This result is summarized in Theorem 1 below.

Second, we show that varying the pair  $(y_1, y_2)$  over  $\underline{\mathcal{Y}}^2$  obtains identification of

$$\{h_2^-(y_2) - h_1^-(y_1), (y_1, y_2) \in \underline{\mathcal{Y}}^2\}.$$

Third, we show that identification of this set of differences obtains identification of  $h_1$  and  $h_2$  under a normalization assumption on  $h_1^-$ . This result is presented in Theorem 2.

Figures 3.1 and 3.2 illustrate the intuition behind our identification strategy for two arbitrary functions,  $h_1$  and  $h_2$ , both accommodated by FELT. The line with kinks and a flat part represents an arbitrary function  $h_1$ , while the solid curve represents an arbitrary function  $h_2$ . Consider Figure 3.1. Pick a  $y_1 \in \underline{\mathcal{Y}}$  on the vertical axis. For all  $y \leq y_1$ ,  $h_1(y)$  gets mapped to zero, while for all  $y > y_1$ , it gets mapped to one. Now pick a  $y_2 \in \underline{\mathcal{Y}}$ . For all  $y \leq y_2$ ,  $h_2(y)$  gets mapped to zero, while for all  $y > y_2$ , it gets mapped to one. This gives rise to a fixed effects binary choice model for  $(D_1(y_1), D_2(y_2))$ , also plotted in the figure as the grey solid lines. Our first result in Theorem 1 identifies the difference  $h_1^-(y_1) - h_2^-(y_2)$  at arbitrary points  $(y_1, y_2)$ , as well as the coefficient  $\beta$ . It is clear that normalizing  $h_1^-(\cdot)$  at an arbitrary point identifies the function  $h_2^-(y_2)$  at an arbitrary point  $y_2$ . This is captured in Figure 3.2. There, for an arbitrary  $y_0$ ,  $h_1^-(y_0) = 0$ . Then, as  $y_2$  is arbitrary, Figure 3.2 shows that moving  $y_2$  on its support traces out the generalized inverse  $h_2^-$  on its domain. Theorem 2 wraps up this argument by showing that  $h_1$  and  $h_2$  are identified from their generalized inverses.

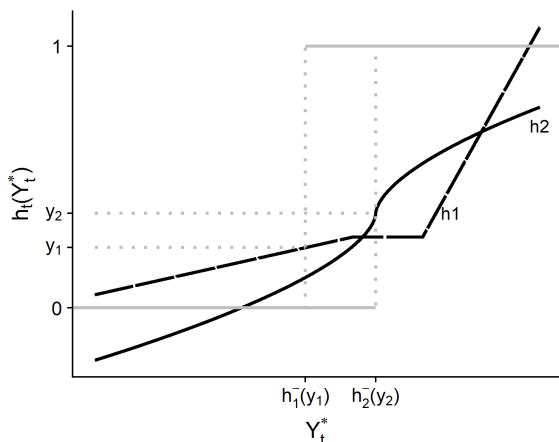


Figure 3.1: FELT functions  $h_1$  and  $h_2$ .

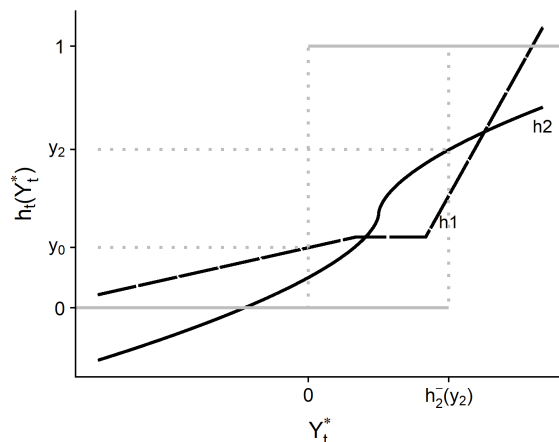


Figure 3.2: Normalization and tracing.

### 3.1.1 Gains from *time-varying* binarization

Using different thresholds in different time periods has advantages over what can be achieved using a time-invariant threshold  $y_1 = y_2 = y$ . First, time-varying binarization increases the information available, helping to mitigate possible thin-set identification issues.<sup>11</sup> Second, marginal effects can be identified without identifying the conditional distribution of the fixed effects. These marginal effects can be time-varying in an unrestricted way. Third, the time-varying transformation function can be identified, which, beyond being of possible independent interest, leads to the identification of features of the conditional distribution of the fixed effects. We informally discuss the first two points below. The third point is presented in detail in the remainder of Section 3 and in Section 4. None of the results in this section are possible with time-invariant thresholds.

**Non-constant thresholds mitigate thin set identification.** We show in the remainder of Section 3 that identification of  $\beta$  and  $h_t$  depends on the properties of the subpopulation of switchers, i.e. those with  $D_2(y_2) \neq D_1(y_1)$ . This is suggestive of thin set identification, see e.g. Khan and Tamer (2010); Lewbel (2019). Identification requires that the set  $\{D_2(y_2) \neq D_1(y_1), y_1, y_2 \in \mathcal{Y}\}$  have strictly positive probability, which depends on the properties of the joint distribution of  $(U_1, U_2)$  and of the transformation functions  $(h_1, h_2)$ . Here, we provide an informal argument for why time-varying binarization may mitigate thin-set identification by considering a

<sup>11</sup>We would like to thank an anonymous referee for pointing this out.

simple example.

Consider the case  $X_1 = X_2 = x$  and  $U_1 = U_2 = U$ , and take two points  $y_1, y_2 \in \underline{\mathcal{Y}}$ . Then the conditional probability below

$$\begin{aligned}
& P(D_2(y_2) \neq D_1(y_1) | \alpha, X_1 = x, X_2 = x) \\
&= P(D_1(y_1) = 1, D_2(y_2) = 0 | \alpha, X_1 = x, X_2 = x) \\
&\quad + P(D_1(y_1) = 0, D_2(y_2) = 1 | \alpha, X_1 = x, X_2 = x) \\
&= P(U \leq x\beta - h_1^-(y_1), U > \alpha + x\beta - h_2^-(y_2) | \alpha, X_1 = x, X_2 = x) \\
&\quad + P(U > x\beta - h_1^-(y_1), U \leq \alpha + x\beta - h_2^-(y_2) | \alpha, X_1 = x, X_2 = x)
\end{aligned}$$

is different from zero only if  $h_1^-(y_1) \neq h_2^-(y_2)$ . To appreciate the usefulness of time-varying binarization in this example, further let  $h_1(\cdot) = h_2(\cdot) = h(\cdot)$ . By Assumption 1,  $h(\cdot)$  is not constant so that the probability above is non-zero as long as  $y_1 \neq y_2$ . This would not be the case by considering a constant threshold, e.g.  $y_1 = y_2 = y$ .

**Time-varying thresholds identify time-varying marginal effects.** In Sections 3.2 and 3.3, we discuss in detail the identification of  $\beta$  and  $h_t$ . Here, we sketch an argument for identification of the marginal effects given identification of  $(\beta, h_1, h_2)$ , assuming that  $(h_1, h_2)$  are invertible. For details, including the case when  $h_t$  is non-invertible, we refer the reader to Botosaru and Muris (2017, Section 3.2).<sup>12</sup>

Let

$$Y_t(x) \equiv h_t(\alpha + x\beta - U_t), \quad t = 1, 2 \quad (3.3)$$

be the counterfactual outcome at time  $t$  under  $X_t = x$ . Under the assumption that the transformation functions are invertible, we have for any  $s, t \in \{1, 2\}$  that

$$\begin{aligned}
P(Y_t(x) \leq y | \alpha, X) &= P(h_t(\alpha + x\beta - U_t) \leq y | \alpha, X) \\
&= P(h_t(\alpha + x\beta - U_s) \leq y | \alpha, X) \quad (3.4)
\end{aligned}$$

$$= P(\alpha + x\beta - U_s \leq h_t^{-1}(y) | \alpha, X) \quad (3.5)$$

$$= P(\alpha + X_s\beta - U_s \leq h_t^-(y) + (X_s - x)\beta | \alpha, X) \quad (3.6)$$

$$= P(Y_s \leq h_s(h_t^-(y) + (X_s - x)\beta) | \alpha, X), \quad (3.7)$$

---

<sup>12</sup>For other recent work on marginal effects, see Chen et al. (2020) for cross-sectional models, and Davezies et al. (2021) for panel models.



where (3.4) follows by the stationarity assumption on the stochastic errors; (3.5) follows by weak monotonicity of  $h_t$ ; (3.6) follows by adding and subtracting  $X_s\beta$  on both sides of the inequality; and (3.7) follows by the model specification and strict monotonicity of the transformation functions. Integrating out with respect to the distribution of  $\alpha$  conditional on  $X$  obtains

$$P(Y_t(x) \leq y | X) = P(Y_s \leq h_s(h_t^-(y) + (X_s - x)\beta) | X). \quad (3.8)$$

Once the distribution of the counterfactual outcomes has been identified, the average structural function (ASF) is identified. For example, when  $h_t$  is invertible, the ASF is given by

$$ASF_t(x) \equiv E(Y_t(x)),$$

and the marginal effect for a continuous regressor is  $\frac{\partial ASF_t(x)}{\partial x_k}$ .

Importantly, this identification result demands time-varying binarization. Without time-varying binarization, one would not identify the functions  $(h_1, h_2)$  and one would not be able to compute (3.8).

**Time-varying thresholds identify the conditional distribution of  $\alpha$ .** It is essential for our empirical application to identify the conditional mean and variance of the fixed effects. This requires the identification of the transformation functions. It is not sufficient to use time-invariant thresholds, which only obtains  $\beta$  and  $h_2^-(y) - h_1^-(y)$ . See Section 4 for a detailed discussion on the identification of features of the conditional distribution of  $\alpha$  given  $X$ .

## 3.2 Nonparametric errors

In this section, we provide nonparametric identification results for  $(\beta, h_1, h_2)$ . Parts of our identification proof build on Manski (1987), who in turn builds on Manski (1975, 1985).

**Assumption 2.** *[Error terms]*

- (i)  $F_1(u|\alpha, X) = F_2(u|\alpha, X) \equiv F(u|\alpha, X)$  for all  $(\alpha, X)$ ;
- (ii) The support of  $F(u|\alpha, X)$  is  $\mathbb{R}$  for all  $(\alpha, X)$ .

Assumption 2 places no parametric distributional restrictions on the distribution

of  $U_{it}$  and allows the stochastic errors  $U_{it}$  to be correlated across time. The first part of the assumption, 2(i), is a stationarity assumption, requiring time-invariance of the distribution of the error terms conditional on the trajectory of the observed regressors and on the unobserved heterogeneity. This assumption excludes lagged dependent variables as covariates. Additionally, as noted by, e.g., Chamberlain (2010), although it allows for heteroskedasticity, it restricts the relationship between the observed regressors and  $U_{it}$  by requiring that even when  $x_1 \neq x_2$ ,  $U_1$  and  $U_2$  have equal skedasticities. This type of stationarity assumption is common in linear and nonlinear panel models, e.g., Chernozhukov et al. (2013) and references therein.

Assumption 2(ii) requires full support of the error terms. It guarantees that, for any pair  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ , the probability of being a *switcher* is positive. In our context, being a switcher refers to the event  $D_1(y_1) + D_2(y_2) = 1$ , so that Assumption 2 guarantees that  $P(D_1(y_1) + D_2(y_2) = 1) > 0$ . This assumption is similar to Assumption 1 in Manski (1987).

Let  $\Delta X \equiv X_2 - X_1$  and for an arbitrary pair  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ , define

$$\gamma(y_1, y_2) \equiv h_2^-(y_2) - h_1^-(y_1). \quad (3.9)$$

**Lemma 1.** *Suppose that  $(Y, X)$  follows the model in (3.1). Let Assumptions 1 and 2 hold. Then for all  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ ,*

$$\text{med}(D_2(y_2) - D_1(y_1) | X, D_1(y_1) + D_2(y_2) = 1) = \text{sgn}(\Delta X \beta - \gamma(y_1, y_2)). \quad (3.10)$$

*Proof.* The proof builds on Manski (1985, 1987), and is presented in Appendix A.1. □

Let  $W \equiv (\Delta X, -1)'$  and  $\theta(y_1, y_2) \equiv (\beta, \gamma(y_1, y_2))$ , so that (3.10) can be written as

$$\text{med}(D_2(y_2) - D_1(y_1) | X, D_1(y_1) + D_2(y_2) = 1) = \text{sgn}(W\theta(y_1, y_2)).$$

For identification of  $\theta(y_1, y_2)$  we impose the following additional assumptions.

**Assumption 3.** *[Covariates]*

(i) *The distribution of  $\Delta X$  is such that at least one component of  $\Delta X$  has positive Lebesgue density on  $\mathbb{R}$  conditional on all the other components of  $\Delta X$  with probability one. The corresponding component of  $\beta$  is non-zero.*

(ii) The support of  $W$  is not contained in any proper linear subspace of  $\mathbb{R}^{K+1}$ .

Assumption 3(i) requires that the change in one of the regressors be continuously distributed conditional on the other components. Assumption 3(ii) is a full rank assumption. These assumptions are standard in the binary choice literature concerned with point identification of the parameters.

Assumption 3 resembles Assumption 2 in Manski (1987), the difference being that our assumption concerns  $W$ , which includes a constant that captures a time trend. The presence of this constant requires sufficient variation in  $X_t$  over time. No linear combination of the components of  $X_t$  can equal the time trend.

**Assumption 4.** [Normalization- $\beta$ ] For any  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ ,  $\theta(y_1, y_2) \in \Theta = \mathcal{B} \times \mathbb{R}$ , where  $\mathcal{B} = \{\beta : \beta \in \mathbb{R}^K, \|\beta\| = 1\}$ .

Assumption 4 imposes a normalization on  $\beta$ , namely that the norm of the regression coefficient equals 1. Scale normalizations are standard in the binary choice literature, and are necessary for point identification when the distribution of the error terms is not parameterized. Normalizing  $\beta$  (instead of  $\theta$ ) avoids a normalization that would otherwise depend on the choice of  $(y_1, y_2)$ . In this way, the scale of  $\beta$  remains constant across different choices of  $(y_1, y_2)$ . Alternatively, one can normalize the coefficient on the continuous covariate (cf. Assumption 3(i)) to be equal to one. In our economic model in Section 5 the latter assumption holds automatically.<sup>13</sup>

**Theorem 1.** Suppose that  $(Y, X)$  follows the model in (3.1), and let the distribution of  $(Y, X)$  be observed. Let Assumptions 1, 2, 3, and 4 hold. Then, for an arbitrary pair  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ ,  $\theta(y_1, y_2)$  is identified.

*Proof.* The proof proceeds by showing that FELT can be converted into a binary choice model for an arbitrary pair  $(y_1, y_2)$ , and then builds on Theorem 1 in Manski (1987), which in turn uses results in Manski (1985). See Appendix A.2.  $\square$

So far, we have identified the regression coefficient  $\beta$  and the difference in the generalized inverses at arbitrary pairs  $(y_1, y_2)$ . We consider now identification of the functions  $h_1$  and  $h_2$  on  $\underline{\mathcal{Y}}$ .

---

<sup>13</sup>There are models with sufficient structure on the transformation  $h_t$  where identification is possible without a normalization on the regression coefficient. Examples include the linear regression model, the censored linear regression model in Honoré (1992), and the interval-censored regression model in Abrevaya and Muris (2020).

**Assumption 5.** [Normalization- $h_1$ ] For some  $y_0 \in \underline{\mathcal{Y}}$ ,  $h_1^-(y_0) = 0$ .

Such a normalization is standard in transformation models, see, e.g., Horowitz (1996). Without this normalization, all identification results hold up to  $h_1^-(y_0)$ . We normalize the function in the first time period only, imposing no restrictions on the function in the second period beyond that of weak monotonicity (cf. Assumption 1). In Section 4, we show that this normalization assumption is not necessary for our results on the identification of the conditional mean or of the conditional variance of the fixed effects.

**Theorem 2.** *Suppose that  $(Y, X)$  follows the model in (3.1), and let the distribution of  $(Y, X)$  be observed. Under Assumptions 1, 2, 3, 4, and 5, the transformations  $h_1$  and  $h_2$  are identified.*

*Proof.* The proof proceeds by identifying the generalized inverses of monotone functions, which obtains identification of the pre-images of  $h_1$  and  $h_2$ . This obtains identification of the functions themselves. See Appendix A.3.  $\square$

### 3.3 Logit errors

In this section, we show identification of  $(\beta, h_1, h_2)$  when the error terms are assumed to follow the standard logistic distribution. The logistic case is not nested in the nonparametric case. In particular, when the errors are logistic, we do not require a continuous regressor. However, we require conditional serial independence of the error terms.<sup>14</sup>

**Assumption 6.** [Logit] (i)  $F_1(u|\alpha, X) = F_2(u|\alpha, X) = \Lambda(u) = \frac{\exp(u)}{1+\exp(u)}$ , and  $U_1$  and  $U_2$  are independent; (ii)  $E(W'W)$  is invertible.

Assumption 6(i) strengthens Assumption 2 by requiring the errors to follow the standard logistic distribution and to be serially independent. Note that one consequence of this assumption, which specifies the variance of the error terms to be equal to 1, is to eliminate the need to normalize  $\beta$ . On the other hand, Assumption 6(ii) imposes weaker restrictions on the observed covariates relative to Assumption 3, since it does not require the existence of a continuous covariate. Sufficient variation in  $\Delta X$

---

<sup>14</sup>See Chamberlain (2010) and Magnac (2004) for more details about identification under nonparametric versus logistic errors in the panel data binary choice context.

is sufficient to obtain identification of the vector  $\beta$  when the error terms follow the standard logistic distribution.

**Theorem 3.** *Suppose that  $(Y, X)$  follow the model in (3.1), and let the distribution of  $(Y, X)$  be observed. Let Assumptions 1 and 6 hold. Then, for an arbitrary pair  $(y_1, y_2) \in \underline{\mathcal{Y}}^2$ ,  $\theta(y_1, y_2)$  is identified. Additionally, letting Assumption 5 hold, then the transformations  $h_1(\cdot)$  and  $h_2(\cdot)$  are identified.*

*Proof.* See Appendix A.4. □

## 4 Conditional distribution of fixed effects

If  $(h_1, h_2)$  are invertible, we can use the previous identification theorem to identify features of the distribution of the fixed effects conditional on observed regressors. These features are the change in the conditional mean function of  $\alpha$  and the conditional variance of  $\alpha$  conditional on  $X_1, X_2$ . These results are relevant since in our collective household model, the fixed effects represent the log of resource shares, and both the standard deviation of these resource shares and the response of their conditional mean to covariates are key parameters of interest in the empirical literature. As this is relevant to our application, we note here that a normalization assumption, such as 4, on the demand function in the first period is not necessary for these results on the resource shares because, e.g., we only need their deviation with respect to the mean of the fixed effects.

In this section, we provide sufficient conditions for the identification of the change in the conditional mean function of the fixed effects, defined as:

$$\mu(x) \equiv E[\alpha | X = x], \text{ for all } x \in \mathcal{X}, \quad (4.1)$$

as well as for the conditional variance of the fixed effects. For these results, the normalization assumption 5 is not necessary. To provide intuition for this, let

$$c_1 \equiv h_1^{-1}(y_0),$$

at an arbitrary  $y_0 \in \underline{\mathcal{Y}}$  and  $g_t(y) \equiv h_t^{-1}(y) - c_1$  for all  $y \in \underline{\mathcal{Y}}$ . Note that Theorem 1 recovers

$$\tilde{U}_t \equiv \alpha - U_t = h_t^{-1}(Y_t) - X_t\beta,$$

up to  $c_1$ , so that the joint distribution of  $(\tilde{U}_1, \tilde{U}_2, X_1, X_2)$  is identified up to  $c_1$ . By placing restrictions on the distribution of  $(\alpha, U_1, U_2, X_1, X_2)$ , we can then recover our features of interest.

**Theorem 4.** (i) *Let the assumptions of Theorem 1 hold, and additionally assume that (4a)  $(h_1, h_2)$  are strictly increasing, and (4b) let  $m \in \mathbb{R}$  be an unknown constant such that  $E(U_t | X = x) = m$ , for all  $x \in \mathcal{X}$ . Then, for any  $x, x' \in \mathcal{X}$ , the change in the conditional mean function  $\mu(x) - \mu(x')$  is identified and given by*

$$\mu(x) - \mu(x') = E[g_t(Y_t) - X_t\beta | X = x] - E[g_t(Y_t) - X_t\beta | X = x'].$$

*Proof.* See Appendix A.5. □

*Remark 1.* As opposed to our main identification result in Theorem 2, Theorem 4 does not use a normalization on the functions  $(h_1, h_2)$ . If we were to impose the normalization in Assumption 5, the conditional mean function  $\mu(x)$  would be identified for all  $x \in \mathcal{X}$ . This result provides justification for nonparametric regression of  $\alpha$  on observables (up to location).

*Remark 2.* Under slightly weaker conditions, we can obtain the projection coefficients of  $\alpha$  on  $X_t$ . This is of interest for our empirical application. Recall that the joint distribution of  $(\tilde{U}_1, \tilde{U}_2, X_1, X_2)$  is identified up to  $c_1$ . Then, assuming  $\text{Cov}(U_s, X_t) = 0$ , we can identify the projection coefficient of  $\alpha$  on  $X_t$  from

$$[\text{Var}(X_t)]^{-1} \text{Cov}(\alpha, X_t) = [\text{Var}(X_t)]^{-1} \text{Cov}(\alpha - U_s, X_t) = [\text{Var}(X_t)]^{-1} \text{Cov}(\tilde{U}_s, X_t).$$

Second, define the conditional variance of the fixed effects as

$$\sigma_\alpha^2(x) \equiv \text{Var}[\alpha | X = x], \text{ for all } x \in \mathcal{X}. \quad (4.2)$$

For this second result, we strengthen our assumptions to include, among others, serial independence of the error term. This allows us to pin the persistence in unit  $i$ 's time series on  $\alpha_i$  instead of on serial dependence in the errors.

**Theorem 5.** *Let the assumptions of Theorem 1 hold and assume that (5a)  $(h_1, h_2)$  are strictly increasing, and (5b)  $\text{Cov}[\alpha, U_t | X = x] = 0$  for all  $x \in \mathcal{X}$  and  $t$ , and (5c)  $\text{Cov}[U_1, U_2 | X = x] = 0$  for all  $x \in \mathcal{X}$ . Then for all  $x \in \mathcal{X}$ , the conditional variance*

function  $\sigma_\alpha^2(x)$  is identified and given by:

$$\sigma_\alpha^2(x) = \text{Cov}(g_2(Y_2) - X_2\beta, g_1(Y_1) - X_1\beta | X = x).$$

*Proof.* See Appendix A.5. □

It may be possible to obtain the entire conditional distribution of the fixed effects under the assumption that  $(\alpha, U_1, U_2)$  are mutually independent by using arguments similar to those in Arellano and Bonhomme (2012).

## 4.1 Logit errors

An alternative approach to the identification of features of the conditional distribution of  $\alpha$  is available for the logistic version of the model.<sup>15</sup> In this case, for any  $m \in \mathbb{N}$ , the  $m^{\text{th}}$  moment  $E[\alpha^m | X = x]$  can be identified given the previously identified  $(h_t, \beta)$  and one period of data. The result is stated for the case with a normalization on  $h_1$ .

**Theorem 6.** *Let all assumptions of Theorem 3 hold, and assume that  $h_t$  is invertible for at least one  $t$ . Then  $E[\alpha^m | X = x]$  is identified for any  $m \in \mathbb{N}$ .*

*Proof.* See Appendix A.5. □

## 5 Microeconomic model

In this section, we construct a model of an efficient full-commitment intertemporal collective (FIC) household. We provide sufficient conditions implying that observed household-level quantity demand equations are time-varying functions of fixed effects that lie in the FELT class.

Essentially, we adapt the intertemporal collective household model of Chiappori and Mazzocco (2017) to the model of household scale economies given in the static household model of Browning et al. (2013). We assume efficiency, i.e. that the household members together reach the Pareto frontier. Consequently, our model does not account for inefficiency due to e.g. consumption externalities or information asymmetries.

---

<sup>15</sup>We are grateful to an anonymous referee for pointing this out.

We use subscripts  $i, j, t$ . Let  $i = 1, \dots, n$  index households and assume the household has a time-invariant composition, with  $N_{ij}$  members of type  $j$ . Let  $j = m, f, c$  for men, women and children. Let  $t = 1, 2$ . Let  $z$  be a vector of time-varying household-level demographic characteristics, and let the numbers of household members of each type,  $N_{im}, N_{if}$  and  $N_{ic}$ , be (time-invariant) elements of  $z_{it}$ . Like Chiappori and Mazzocco (2017), this is a model with uncertainty, so we use the superscript  $s = 1, 2$  to index states in the second period only. As in Chiappori and Mazzocco (2017), the use of 2 time periods and 2 states is for illustration only; the model goes through with any finite number of states or periods. Similarly, inclusion of a risky asset would not change the features of the model that we use.

Indirect utility,  $V_j(p, x, z)$ , is the maximized value of utility given a budget constraint defined by prices  $p$  and budget  $x$ , given characteristics  $z$ . Let  $V_j$  be strictly concave in the budget  $x$ . Indirect utility depends on time only through its dependence on the budget constraint and time-varying demographics  $z$ . Let  $v_{ijt} \equiv V_j(p_t, x_{it}, z_{it})$  denote the utility level of a person of type  $j$  in household  $i$  in period  $t$ .

Household decisions are made on the basis of individual shadow budget constraints, reflecting the economic environment within the household. These shadow constraints are characterized by a shadow price vector faced by all members and shadow budgets which may differ across household members. Let  $A_{it} \equiv A(z_{it})$  be a diagonal matrix that gives the shareability of each good, and let it depend on demographics  $z_{it}$  (including the numbers of household members). More shareable goods have lower shadow prices of within-household consumption.

For nonshareable goods, the corresponding element of  $A_{it}$  equals 1; for shareable goods, it is less than 1, possibly as small as  $1/N_i$  where  $N_i$  is the number of household members. Goods may be partly shareable, with an element of  $A_{it}$  between  $1/N_i$  and 1. With market prices  $p_t$ , within-household shadow prices are given by the linear transformation  $A_{it}p_t$ .

Browning et al. (2013) also allow for inequality in shadow budgets. Let  $\eta_{ijt}$  be the *resource share* of type  $j$  in household  $i$  in time period  $t$ . It gives the fraction of the household budget consumed by that type. Each person of the  $N_{ij}$  people of type  $j$  consumes  $\eta_{ijt}/N_{ij}$  of the household budget  $x_{it}$ , so they each have a shadow budget of  $\eta_{ijt}x_{it}/N_{ij}$ .

Each household member faces a shadow budget of  $\eta_{ijt}x_{it}/N_{ij}$  and shadow prices



of  $A_{it}p_t$ , so that, within the household, indirect utility is given by

$$v_{ijt} = V_j(A_{it}p_t, \eta_{ijt}x_{it}/N_{ij}, z_{it}). \quad (5.1)$$

Let  $V_{x_j}(p, w, z) \equiv \partial V_j(p, w, z)/\partial w$  be the monotonically decreasing marginal utility of person  $j$  with respect to their shadow budget. Then,  $V_{x_j}(A_{it}p_t, \eta_{ijt}x_{it}/N_{ij}, z_{it})$  is the value of their marginal utility evaluated at their shadow budget constraint.

Let  $p_2^s, z_{i2}^s, \bar{x}_i^s$  for  $s = 1, 2$  be the possible realizations of state-dependent variables that occur with household-specific probabilities  $\pi_i^1$  and  $\pi_i^2$  (which sum to 1). Here,  $\bar{x}_i^s$  is the state-specific lifetime wealth of household  $i$ , revealed in period 2.

Pareto weights  $\phi_{ij}$  vary arbitrarily across households, and depend, for example, on the household-specific expectation of lifetime wealth at the moment of household formation. Because ours is a full commitment model, Pareto weights do not have a time subscript because they are fixed at the moment of household formation, when the full-commitment contract is set. This household-level time-invariant variable will form the basis of our fixed-effects variation.

As in Chiappori and Mazzoco (2017), the assumption of efficiency implies that we can represent the household's decisions via the Bergson-Samuelson Welfare Function,  $W_i$ , for the household is

$$W_i \equiv \sum_{j=1}^J N_{ij} \phi_{ij} \left[ v_{ij1} + \sum_{s=1}^2 \rho_i \pi_i^s v_{ij2}^s \right]. \quad (5.2)$$

The term in square brackets is the expected lifetime utility, discounted by the discount factor  $\rho_i$ , of each member of type  $j$  in household  $i$ . Each member of type  $j$  gets the Pareto weight  $\phi_{ij}$ . The assumption of efficiency implies that the household reaches the Pareto Frontier; the Pareto weights pinpoint which point on the Frontier is chosen by the household.

Next, substitute indirect utility (5.1) for utility  $v_{ij1}$  and  $v_{ij2}^s$  into (5.2),<sup>16</sup> and form the Lagrangian using the intertemporal budget constraint with interest rate  $\tau$ ,  $x_{i1} + x_{i2}^s/(1 + \tau) = \bar{x}_i^s$ , and the adding-up constraints on resource shares,  $\sum_j \eta_{ij1} = \sum_j \eta_{ij2}^s = 1$ . Each household  $i$  chooses  $x_{i1}$ ,  $\eta_{ij1}$  and  $\eta_{ij2}^s$  to maximize  $W_i$ . Solving this

---

<sup>16</sup>In contrast, Chiappori and Mazzoco (2017) substitute *direct* utility for utility  $v_{ij1}$  and  $v_{ij2}^s$  using a model of pure private and pure public goods. In that model, each individual's utility is given by their direct utility function, which is a function of their (unobserved) consumption of a vector of private goods and their (observed) consumption of a vector of public goods.

optimization problem obtains for any two types,  $j$  and  $k$ , that:

$$\frac{\phi_{ij}}{\phi_{ik}} = \frac{V_{xk}(A_{i1}p_1, \eta_{ik1}x_{i1}/N_{ik}, z_{i1})}{V_{xj}(A_{i1}p_1, \eta_{ij1}x_{i1}/N_{ij}, z_{i1})} = \frac{V_{xk}(A_{i2}p_2^s, \eta_{ik2}^s x_{i2}^s / N_{ik}, z_{i2}^s)}{V_{xj}(A_{i2}p_2^s, \eta_{ij2}^s x_{i2}^s / N_{ij}, z_{i2}^s)}, \quad (5.3)$$

for  $s = 1, 2$  for all  $i$ . That is, the household chooses resource shares so as to equate ratios of marginal utilities with ratios of Pareto weights.

An *assignable good* is one where we observe the consumption of that good by a specific person (or type of person). Assuming the existence of a scalar-value demand function  $q_j(p, x, z)$  for an assignable and non-shareable good (e.g., food or clothing) for a person of type  $j$ , Browning et al. (2013) show that the household's quantity demand,  $Q_{ijt}$ , for the assignable good for each of the  $N_{ij}$  people of type  $j$  is given by

$$Q_{ijt} = q_j(A_{it}p_t, \eta_{ijt}x_{it}/N_{ij}, z_{it}).$$

Assuming that the assignable good is a normal good implies that  $q_j$  is strictly increasing in its second argument, and is therefore strictly monotonic.

Since  $p_t$  is unobserved and varies over time,<sup>17</sup> we can express  $Q_{ijt}$  as a time-varying function of observed data. Defining  $\tilde{q}_{jt}(x, z) = q_j(p_t, x, z)$ , we have

$$Q_{ijt} = \tilde{q}_{jt}(\eta_{ijt}x_{it}/N_{ij}, z_{it}). \quad (5.4)$$

This is the structural demand equation that we ultimately bring to the data.

The model above determines resource share via the first-order conditions (5.3), and these resource shares depend on time-invariant Pareto weights  $\phi_{ij}$ . But these resource shares are a vector of implicit functions, which may be hard to work with. To make the model tractable, we impose sufficient structure on utility functions to find closed forms for resource shares.

In our empirical example below, we work with data that have time-invariant demographic characteristics, so let  $z_{it} = z_i$  be fixed over time. This implies that the shareability of goods embodied in  $A$  is time-invariant:  $A_{it} = A_i = A(z_i)$ . We will

---

<sup>17</sup>The assumption that prices are unobserved but vary *only* with time is important for our empirical application below, where we observe Bangladeshi households in 2 time periods. It implies moment conditions for differenced demand functions that do not depend on prices. If prices vary with any observed variables, then these could be worked into the moment conditions, by conditioning on those variables. However, if prices vary with unobserved variables, our estimation strategy would not work.

estimate the demand equation for women's food in nuclear households comprised of 1 man, 1 woman and 1 – 4 children, so  $N_{if} = N_{im} = 1$ .

Let indirect utilities be in the price-independent generalized logarithmic (PIGL) class (Muellbauer (1975, 1976)) given by

$$V_j(p, x, z) = C_j(p, z) + (B(p, z)x)^{r(z)} / r(z). \quad (5.5)$$

Here,  $V_j$  is homogeneous of degree 1 in  $p, x$  if  $C_j$  is homogeneous of degree 0 in  $p$  and  $B$  is homogeneous of degree  $-1$  in  $p$ .  $V$  is increasing in  $x$  if  $B(p, z)$  is positive and  $V$  is concave in  $x$  if  $r(z) < 1$ . In terms of preferences, this class is reasonably wide. It gives quasihomothetic preferences if  $r(z) = 1$ , and PIGLOG preferences as  $r(z) \rightarrow 0$  (this includes the Almost Ideal Demand System of Deaton and Muellbauer (1980)).

The functions  $C_j$  vary across types  $j$ , and so the model allows for preference heterogeneity between types, e.g., between men and women. The restrictions that  $B(p, z)$  and  $r(z)$  don't vary across  $j$  and that  $r(z)$  does not depend on prices  $p$  are important: as we see below, they imply that resource shares are constant over time.

Substituting into the BCL model, observed demographics and period  $t$  budgets, we have that marginal utilities are given by

$$V_{xj}(A_{it}p_t, \eta_{ijt}x_{it}/N_{ij}, z_{it}) = B(A_{it}p_t, z_{it})^{r(z_i)} (\eta_{ijt}x_{it}/N_{ij})^{r(z_i)-1}.$$

For  $r(z_i) \neq 1$ , and for any pair of types  $j, k$ , we substitute into (5.3) and rearrange to get

$$\left( \frac{\eta_{ikt}}{\eta_{ijt}} \right) = \frac{N_{ik}}{N_{ij}} \left( \frac{\phi_{ij}}{\phi_{ik}} \right)^{1/(r(z_i)-1)}. \quad (5.6)$$

The household chooses resource shares in each period and each state to satisfy (5.6). Since the right-hand side has no variation over time or state, this implies that, given PIGL utilities (5.5), the resource shares in a given household  $i$  are independent of period  $t$  and state  $s$ . However, resource shares do vary with both observed and unobserved variables across households  $i$ .

Let the fixed resource shares that solve the first-order conditions with PIGL demands be denoted  $\bar{\eta}_{ij}$ , and define

$$\alpha_i \equiv \ln(\bar{\eta}_{if}/N_{if}) = \ln \bar{\eta}_{if}, \quad (5.7)$$

to equal to the logged resource share of the woman in the household.

Let there be a multiplicative Berkson (1950) measurement error denoted  $\exp(-U)$  which multiplies the budget, so that if we observe  $x$ , the actual budget is  $x/\exp(U)$ . The measurement error is i.i.d. across time and households, which implies stationarity of  $U$ . Here, the measurement error does not affect resource shares, but does affect the distribution of observed quantity demands. Then household demand for women's food (the assignable good),  $Q_{ift}$ , is given by

$$Q_{ift} = \tilde{q}_{ft}(\exp(\alpha_i) x_{it} / \exp(U_{it}), z_i).$$

This is a FELT model, conditional on covariates:

$$Y_{it} = h_t(Y_{it}^*, z_i), \tag{5.8}$$

where  $h_t(Y_{it}^*, z_i) = \tilde{q}_{ft}(\exp(Y_{it}^*), z_i)$ ,  $Y_{it}^* = \alpha_i + X_{it} - U_{it}$ , and  $X_{it} = \ln x_{it}$  is the logged household budget.

The assumption that the assignable good is normal means that the time-varying functions  $h_t$  are strictly monotonic in  $Y_{it}^*$ . One could additionally impose that the demand functions  $\tilde{q}_{ft}$  come from the application of Roy's Identity to the indirect utility function (5.5). These demand functions equal a coefficient times the shadow budget plus a coefficient times the shadow budget raised to a power, where the coefficients are time-varying and depend on  $z_i$ .<sup>18</sup> We do not impose that additional structure here; instead, we show in Section 7 that the estimated demand curves given by FELT are close to the PIGL shape restrictions.

Here, the time-dependence of  $h_t$  is economically important; it is driven by the price-dependence of preferences and by the fact that prices are common to all households  $i$  but vary over time  $t$ . Further, the fixed effects  $\alpha_i$  are economically meaningful

---

<sup>18</sup>Individual demands are derived by the application of Roy's Identity to (5.5), and are:

$$q_{jt}(x, z) = c_{jt}(z) x^{1-r(z)} + b_t(z) x$$

where  $c_{jt}(z) = -\frac{\nabla_p C_j(p_t, z)}{B(p_t, z)^r}$ ,  $b_t(z) = -\nabla_p \ln B(p_t, z)$ . This notation makes clear that we have time-varying demand functions, due to the fact that prices vary over time. In our application, prices in each period are not observed, so we allow the ( $z$ -dependent) functions  $c_{jt}$  and  $b_t$  to vary over time. We require that the assignable good be normal, meaning that its demand function is globally increasing in  $x$ . This form for demand functions is globally increasing if  $c_{jt}(z)$ ,  $b_t(z)$  and  $1 - r(z)$  are all positive.

parameters: they are equal to the logged women’s resource shares in each household. The standard deviation of the logs is a common inequality measure, and the standard deviation of  $\alpha_i$  is identified by FELT given strict monotonicity, as we show in Section 4. Further, the covariation of  $\alpha_i$  with observed regressors is identified.

## 6 Data

We use data from the 2012 and 2015 Bangladesh Integrated Household Surveys. This data set is a household survey panel conducted jointly by the International Food Policy Research Institute and the World Bank. In this survey, a detailed questionnaire was administered to a sample of rural Bangladeshi households. This data set has two useful features for our purposes: 1) it includes person-level data on food intakes and household-level data on total household expenditures; and 2) it is a panel, following roughly 6000 households over two (nonconsecutive) years. The former allows us to use food as the assignable good to identify our collective household model parameters. The latter allows us to model household-level unobserved heterogeneity in women’s resource shares.

The questionnaire was initially administered to 6503 households in 2012, drawn from a representative sample frame of all rural Bangladeshi households. Of these, 6436 households remained in the sample in 2015. In these data, expenditures on food include imputed expenditure from home production. We drop households with a discrepancy between people reported present in the household and the personal food consumption record, and households with no daily food diary data. Of the remaining data, 6205 households have total expenditures reported for both 2012 and 2015.

In this paper, we focus on households that do not change members between periods.<sup>19</sup> There are 1920 households whose composition is unchanged between 2012 and 2015. Roughly half of these households have more than one adult man or more than one adult women. To simplify the interpretation of estimated resource shares we focus on nuclear households. This leaves 871 nuclear households comprised of one man, one woman and 1 to 4 children, where children are defined to be 14 years old or younger.

---

<sup>19</sup>That is, we exclude households with births, deaths, new members by marriage or adoption, etc. Although a full-commitment model can accommodate such changes in household composition, it is easier to think through the meaning of a person’s resource share if the composition is held constant.

Our household-level annual expenditure,  $x_{it}$ , is the sum of total expenditure on, and imputed home-produced consumption of, the following categories of consumption: rent, food, clothing, footwear, bedding, non-rent housing expense, medical expenses, education, remittances, religious food and other offerings (jakat/ fitra/ daan/ sodka/ kurbani/ milad/ other), entertainment, fines and legal expenses, utensils, furniture, personal items, lights, fuel and lighting energy, personal care, cleaning, transport and telecommunication, use-value from assets, and other miscellaneous items. These spending levels derive from one-month and three-month duration recall data in the questionnaire, and are grossed up to the annual level. Estimation uses  $X_{it} = \ln x_{it}$ , the natural logarithm of annual consumption.

The assignable good,  $Y_{it}$ , is annual expenditure on food for the woman. The surveys contains a one-day (24-hour) food diary with data on *person-level* quantities (measured in kilograms) of food consumption in 7 categories: Cereals, Pulses, Oils; Vegetables; Fruits; Proteins; Drinks and Others. These consumption quantities include home-produced food and purchased food and gifts. They include both food consumed in the home (both cooked at home and prepared ready-to-eat food), as well as food consumed outside the home (at food carts or restaurants). These one-day food quantities are transformed into one-day food expenditures by multiplying by estimated village-level unit-values (following Deaton (1997)), and from this, the woman’s share of one-day household food expenditure is calculated. Finally, the woman’s annual expenditure on food is calculated as her share of one-day food expenditure multiplied by the household’s total annual food expenditure (from recall data).

Our model is also conditioned on a set of time-invariant demographic variables  $z_i$ . We include several types of observed covariates in  $z_i$  that may affect both preferences and resource shares: 1) the age in 2012 of the adult male; 2) the age in 2012 of the adult female; 3) the average age in 2012 of the children; 4) the average education in years of the adult male; 5) the average education in years of the adult female; 6) an indicator that the household has 2 children; 7) an indicator that the household has 3 or 4 children; and 8) the fraction of children that are girls.<sup>20</sup>

For the first five of these demographic variables, in order to reduce the support

---

<sup>20</sup>Since household membership is fixed for all households in our sample, age, number and gender composition are time-invariant by construction. However, education level of men and women are time-varying in roughly 20% of households. For our time-invariant education variables, we use the average education across the two observed years.

of the regressors, we top- and bottom-code each variable so that values above (below) the 95<sup>th</sup> (5<sup>th</sup>) percentiles equal the 95<sup>th</sup> (5<sup>th</sup>) percentile values. For all seven of these variables, we standardize the location and scale so that their support is  $[0, 1]$ . This support restriction simplifies our monotonicity restrictions when it comes to estimation, as explained in Section 7.

We do not trim the data for outliers in the budget or food quantity demands. Instead, we trim the support of the estimated nonparametric regression functions to account for fact that these estimators are high-variance near their boundaries.

Table 2 in the Appendix, Section C gives summary statistics on these data.

## 7 Estimation of Resource Shares

Given our identification arguments, estimation could be based on composite versions of the maximum score estimator or the conditional logit estimator (see Botosaru and Muris (2017), Theorem 5, and Theorem 4 for the case of non-invertible  $h_t$ ). However, for the empirical application in this paper, it is desirable to condition on household characteristics. This extension is not covered by Botosaru and Muris (2017). For this reason, we follow a sieve GMM approach that facilitates the inclusion of a large vector of demographic conditioning variables  $z$  and the imposition of strict monotonicity on the demand functions (aka normality of the assignable good).

The women’s food demand equation (5.8) is a FELT model, conditional on observed covariates  $z$ . Denote the inverse demand functions  $g_t(Y_{it}, z_{it}) = h_t^{-1}(\cdot, z_{it})$ . Given (5.8) a two-period setting with  $t = 1, 2$ , and time-invariant demographics  $z_{it} = z_i$ , we have

$$\alpha_i + X_{it} - U_{it} = g_t(Y_{it}, z_i), \quad (7.1)$$

implying the conditional moment condition

$$E [g_2(Y_{i2}, z_i) - g_1(Y_{i1}, z_i) - \Delta X_{it} | X_{i1}, X_{i2}, z_i] = 0, \quad (7.2)$$

using stationarity of the conditional distribution of the Berkson errors  $\exp(-U_{it})$ .<sup>21</sup>

---

<sup>21</sup>Our proposed estimator does not follow directly from our identification argument. Identification of  $(g_1, g_2)$  based on the conditional moment restriction in 7.2 is related to identification in linear nonparametric conditional moment models, see, e.g., Carrasco et al. (2007), Darolles et al. (2011). This involves assumptions such as completeness and measurable separability. Such an analysis is beyond the scope of this paper.

We approximate the inverse demand functions,  $g_t$ ,  $t = 1, 2$ , using Bernstein polynomials. This allows us to impose monotonicity in a straightforward way, see, e.g., Wang and Ghosh (2012). Let  $k = 0, \dots, K$  index the univariate Bernstein functions

$$\mathcal{B}_k(u, K) = \binom{K}{k} u^k (1-u)^{K-k}, \quad u \in [0, 1],$$

where  $K$  is the degree of the Bernstein polynomial. Let  $l = 0, \dots, L$  index the elements of  $z_i$ , and let the first (index 0) element of  $z_i$  be a constant equal to 1, that is  $z_i = [1, z_{i1}, \dots, z_{iL}]$ . Our approximation of  $g_t(Y_{it}, z_i)$  is then given by:

$$g_t(Y_{it}, z_i) \approx \sum_{k=0}^K \beta_{kt}(z_i) \mathcal{B}_k(Y_{it}, K) \approx \sum_{l=0}^L \sum_{k=0}^K z_i^{(l)} \beta_{kt}^{(l)} \mathcal{B}_k(Y_{it}, K). \quad (7.3)$$

Here the coefficients on the Bernstein polynomials are linear functions of  $z_i$ . When there are no covariates (that is,  $L = 0$ ), the expression above reduces to the standard Bernstein polynomial approximation  $g_t(Y_{it}) \approx \sum_{k=0}^K \beta_{kt}^{(0)} \mathcal{B}_k(Y_{it}, K)$ . Note that the dependence of the coefficients  $\beta_{kt}^{(l)}$  on the demographics may vary with time. In this way, the relationship between the (nonlinear) demand  $Y_{it}$  and the latent budget  $Y_{it}^*$  depends on demographic characteristics and on prices, through  $t$ . Since Bernstein polynomials are defined on the unit interval, we normalize  $Y_{it}$  to be uniform on  $[0, 1]$  by applying its empirical distribution function.<sup>22</sup>

To form GMM estimators, we construct the following unconditional moments:

$$E \left[ \left( \sum_{l=0}^L \sum_{k=0}^K z_i^{(l)} \left( \beta_{k2}^{(l)} \mathcal{B}_k(Y_{i2}, K) - \beta_{k1}^{(l)} \mathcal{B}_k(Y_{i1}, K) \right) - \Delta X_{it} \right) \mathcal{B}_{k'}(x_{it}, K) z_i^{(l')} \right] = 0$$

for  $k' = 0, \dots, K$ ,  $l' = 0, \dots, L$ , and  $t = 1, 2$ , and

$$E \left[ \left( \sum_{l=0}^L \sum_{k=0}^K z_i^{(l)} \left( \beta_{k2}^{(l)} \mathcal{B}_k(Y_{i2}, K) - \beta_{k1}^{(l)} \mathcal{B}_k(Y_{i1}, K) \right) - \Delta X_{it} \right) X_{it} z_i^{(l')} \right] = 0,$$

for  $l' = 0, \dots, L$  and  $t = 1, 2$ . We include the second condition (where the logged

---

<sup>22</sup>We use this normalization for the estimation of the Bernstein coefficients, but we present our results in terms of untransformed  $Y_{it}$ . These results are obtained by applying the inverse transformation to the function estimated with transformed data.



household budget  $X_{it}$  is exogenous) because we ultimately wish to consider the correlation of  $\alpha_i$  and  $X_{it}$ . For a given order of the sieve approximation,  $K$ , the equations above amount to a parametric linear GMM problem.

We impose increasingness on the estimates of the functions  $g_1$  and  $g_2$  by imposing  $\beta_{kt}^{(l)} \geq \beta_{k-1,t}^{(l)}$ , for all  $l$ , for all  $t$ , and for  $k \geq 2$ . This results in a quadratic programming problem with linear inequality restrictions, which we implement in R using the `quadprog` package.

We implement our estimation method over a range of smoothing parameter values, that is,  $K \in \{1, 2, \dots, 12\}$ , in search of a region where the estimates are not very sensitive to small variations in the smoothing parameter. We select the mid-point of that region, so our main results use  $K = 8$ , see Appendix B. We present results for  $K \in \{1, 4, 8, 10\}$  in Appendix D. We use the nonparametric bootstrap for inference. Because the nonparametric bootstrap may not be valid for this case, we take our estimated confidence intervals with caution, see Appendix B.

We present estimates for the demand functions,  $\hat{h}_t(x, \bar{z}) = \hat{g}_t^{-1}(\cdot, \bar{z})$  for  $t = 1, 2$ , where  $\bar{z}$  represents a household with 1 child and has other observed characteristics that are the average of those of 1-child households. The functions  $h_t$  (demand functions) are not of direct interest, but identification of the  $h_t$  supports identification of moments of the distribution of fixed effects  $\alpha_i$ . In our context,  $\alpha_i$  equals the log of the resource share of the woman in household  $i$ .

We characterize several interesting features of the distribution of resource shares. Recall from Theorems 4 and 5 that identification of features of this distribution does not impose a normalization on assignable good demand functions, and only identifies the distribution of logged resource shares (fixed effects) up to location. Consequently, we only identify features of the resource share distribution up to a scale normalization. This is related to identification results in Chiappori and Ekeland (2009) which show identification up to location using assignable goods.

Let  $\hat{g}_{it} = \hat{g}_t(Y_{it}, z_i)$  equal the predicted values of the inverse demand functions at the observed data. Recall that  $g_t(Y_{it}, z_i) = \alpha_i + X_{it} - U_{it}$ , so we can think of  $\hat{g}_{it} - X_{it}$  as a prediction of  $\alpha_i - U_{it}$ . We then compute the following summary statistics of interest, leaving the dependence of  $\hat{g}_{it}$ ,  $t = 1, 2$ , on  $z_i$  implicit:

1. an estimate of the standard deviation of  $\alpha_i$  given by

$$\hat{std}(\alpha) = \sqrt{\hat{cov}((\hat{g}_{i2} - X_{i2}), (\hat{g}_{i1} - X_{i1}))},$$

where  $c\hat{ov}$  denotes the sample covariance. The standard deviation of logs is a standard (scale-free) inequality measure. So this gives a direct measure of inter-household variation in women's resource shares.

2. an estimate of the standard deviation of the projection error,  $e_i$ , of  $\alpha_i$  on  $\bar{X}_i = \frac{1}{2}(X_{i1} + X_{i2})$  and  $Z_i$ . Consider the projection

$$\alpha_i = \gamma_1 \bar{X}_i + \gamma_2 Z_i + e_i,$$

where  $Z_i$  contains a constant. We are interested in the standard deviation of  $e_i$ . To obtain this parameter, we compute estimators for  $\gamma_1, \gamma_2$  from the pooled linear regression of  $\hat{g}_{it} - X_{it}$  on  $\bar{X}_i$  and  $Z_i$ . Call these estimators  $\hat{\gamma}_1, \hat{\gamma}_2$ . Then, as in  $\hat{std}(\alpha)$  in (1), an estimate of the standard deviation of  $e_i$  is given by:

$$\hat{std}(e_i) = \sqrt{c\hat{ov}((\hat{g}_{i2} - X_{i2} - \hat{\gamma}_1 \bar{X}_i - \hat{\gamma}_2 Z_i), (\hat{g}_{i1} - X_{i1} - \hat{\gamma}_1 \bar{X}_i - \hat{\gamma}_2 Z_i))}.$$

This object measures the amount of variation in  $\alpha_i$  that cannot be explained with observed regressors. If it is zero, then we don't really need to account for household-level unobserved heterogeneity in resource shares. If it is much larger than zero, then accounting for household-level unobserved heterogeneity is important.

3. an estimate of the standard deviation of  $\alpha_i + X_{it}$  for  $t = 1, 2$ , computed as

$$\hat{std}(\alpha_i + X_{it}) = \sqrt{\hat{var}(\alpha_i) + \hat{var}(X_{it}) + 2c\hat{ov}(\alpha_i, X_{it})},$$

where

$$\begin{aligned} c\hat{ov}(\alpha_i, X_{i1}) &= cov(\hat{g}_{i1} - X_{i1}, X_{i1}), \\ c\hat{ov}(\alpha_i, X_{i2}) &= cov(\hat{g}_{i2} - X_{i2}, X_{i2}), \end{aligned}$$

and  $\hat{var}(\alpha_i) = (\hat{std}(\alpha))^2$  and  $\hat{var}(X_{it})$ ,  $t = 1, 2$ , is observed in the data. Since  $\alpha_i + X_{it}$  is a measure of the woman's shadow budget,  $\hat{std}(\alpha_i + X_{it})$  is a measure of inter-household inequality in women's shadow budgets. This inequality measure is directly comparable to the standard deviation of  $X_i$  (shown in Table 1), which measures inequality in household budgets.

4. an estimate of the covariance of  $\alpha_i, X_{it}$  for  $t = 1, 2$ , denoted  $c\hat{ov}(\alpha_i, X_{it})$ . This object is of direct interest to applied researchers using cross-sectional data to identify resource shares. If this covariance is non-zero, then the independence of resource shares and household budgets is cast into doubt, and identification strategies based on this restriction are threatened.

Of these, the first 2 summary statistics are about the variance of fixed effects, and are computed using data from both years. Their validity requires serial independence of the measurement errors  $U_{it}$ . In contrast, the second 2 summary statistics are about the correlation of fixed effects with the household budget, and are computed at the year level. They are valid with stationary  $U_{it}$ , even in the presence of serial correlation.

We also consider the multivariate relationship between resource shares, household budgets and demographics. Recall that the fixed effect  $\alpha_i$  subject to a location normalization; this means that resource shares are subject to a scale normalization. So, we construct an estimate of the woman's resource share in each household as  $\hat{\eta}_i = \exp\left(\frac{1}{2}(\hat{g}_{i1} - X_{i1}) + (\hat{g}_{i2} - X_{i2})\right)$ , normalized to have an average value of 0.33. Then, we regress estimated resource shares  $\hat{\eta}_i$  on  $\bar{X}_i$  and  $Z_i$ , and present the estimated regression coefficients, which may be directly compared with similar estimates in the cross-sectional literature.

The estimated coefficient on  $X_i$  gives the conditional dependence of resource shares on household budgets, and therefore speaks to the reasonableness of the restriction that resource shares are independent of those budgets (an identifying restriction used in the cross-sectional literature). Finally, using the estimate of the variance of fixed effects, we construct an estimate of  $R^2$  in the regression of resource shares on observed covariates. This provides an estimate of how much unobserved heterogeneity matters in the overall variation of resource shares.

## 8 Results

Figure 8 shows our estimates of  $h_1$  and  $h_2$  (or, equivalently, of  $g_1$  and  $g_2$ ) for  $K = 8$ , for a family with two children with mean values,  $\bar{z}$ , of the other demographics. The figures have food quantities  $q_t$  on the vertical axis and  $\hat{g}_t(q_t)$  on the horizontal axis, so the horizontal axis is like a predicted logged household budget. Solid lines give

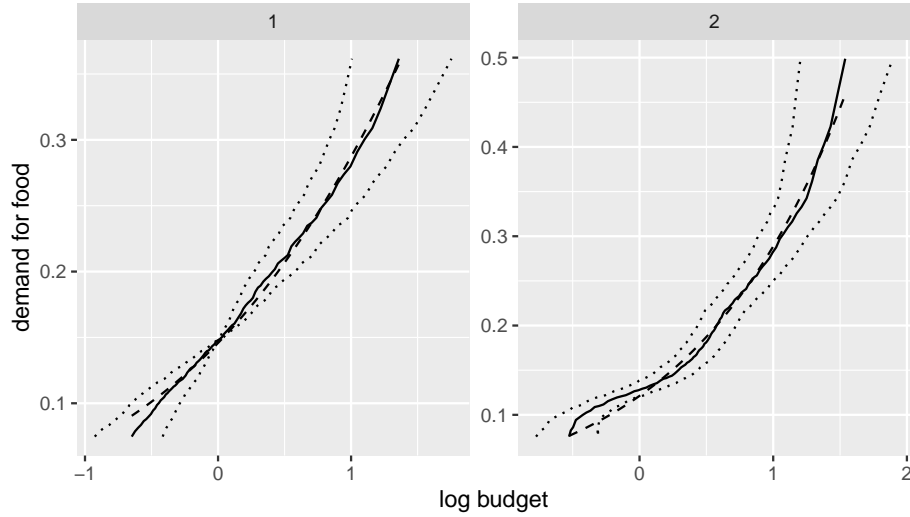


Figure 8.1: Estimated demand functions. Solid line is the nonparametric estimate, evaluated at the mean value of the demographics. The dotted lines indicate the 95% confidence interval. The dashed line is the PIGL closest to the nonparametric estimate. Left panel is for period 1, right panel is for period 2.

the nonparametric estimates, and 95% pointwise confidence bands for the nonparametric estimates are denoted by dotted lines. Additionally, to provide reassurance that the PIGL utility model—which implies the FELT demand curves—fits the data adequately, we display the PIGL demand curve closest to the FELT estimates in each time period with dashed lines.<sup>23</sup>

Note that since  $g_t$  are identified only up to location (of  $g_1$ ), we normalize the average of  $\hat{g}_t$  to half the geometric mean of household budgets at  $t=1$ ,  $\bar{x}_1$ . Because estimated nonparametric regression functions can be ill-behaved near their boundaries, we truncate the estimated functions at the 5th and 95th percentiles of the distribution of  $q_t$  in each  $t$ . The key message from Figure 8 is that these estimated demand curves are somewhat nonlinear, estimated reasonably precisely, and not too far from PIGL. The estimated PIGL curvature parameter is  $r(\bar{z}) = 0.06$ , which means that food demands are close to PIGLOG (as in Banks et al. (1997)).

Table 1 gives our summary statistics (items 1-4 above), with bootstrapped 95%

<sup>23</sup>We compute these PIGL demand curves by nonlinear least squares estimation of a pooled  $q_{it}$  on  $\hat{g}_{it}$ , where the demand curves have the form  $q_{it} = c_t x^{1-r} + b_t x$ . We estimate the model on a grid of 198 points, one for each interior percentile of  $q_{it}$  in each period  $t = 1, 2$ .

	Estimate	q(2.5)	q(97.5)
<hr/> $n = 871$ <hr/>			
<b>Variability of fixed effects <math>\alpha_i</math></b>			
$\hat{std}(\alpha_i)$	0.2647	0.1518	0.3707
$\hat{std}(e_i)$	0.1637	0.1262	0.1931
$\hat{cov}(\alpha_i, X_{i1})$	-0.0901	-0.1292	-0.0429
$\hat{cov}(\alpha_i, X_{i2})$	-0.1034	-0.1418	-0.0561
$\hat{std}(\alpha_i + X_{i1})$	0.3537	0.2720	0.4605
$\hat{std}(\alpha_i + X_{i2})$	0.3763	0.3010	0.4760
<hr/> <b>Regression estimates</b> <hr/>			
$R^2: \bar{X}_i, z_i$ on $\eta_i$	0.5205	0.3361	0.6508
$\bar{X}_i$	-0.0452	-0.0700	-0.0196
age–woman	-0.2513	-0.4437	-0.0545
age–man	0.2845	0.1389	0.4191
2 children	-0.0502	-0.1694	0.0413
3 or 4 children	-0.1248	-0.2407	0.0140
avg age of children	-0.0095	-0.1979	0.2151
fraction girl children	0.0136	-0.0881	0.1246
education–woman	-0.0325	-0.1519	0.1198
education–men	-0.1330	-0.2589	-0.0015

Table 1: Estimates.

confidence intervals, for our estimates with 8 Bernstein polynomials (see the Appendix for other lengths of the Bernstein sieve). In the lower panel, we provide estimated regression coefficients, also with bootstrapped 95% confidence intervals, where we regress estimated resource shares  $\hat{\eta}_i$  on log-budgets  $\bar{X}_i$  and demographics  $z_i$ .

Starting with the top panel of Table 1, the standard deviation of  $\alpha_i$  is a measure of inter-household dispersion in women’s resource shares. If this dispersion is very small, then variation in resource shares does not induce much inequality, and we can reasonably use the household-level income distribution as a proxy for person-level inequality. However, if the dispersion is large, then household-level measures of inequality could be very misleading. Note here that we focus on inequality among women, not on gender inequality, precisely because the location of resource shares is not identified.

The estimated value is roughly 0.26, with a 95% confidence interval covering roughly 0.15 to 0.37. To get a sense of the magnitude for the standard deviation of logged resource shares, suppose that women’s resource shares were lognormally

distributed. Then our estimated standard deviation of 0.26 is consistent with 95% of the distribution of the resource shares lying in the range  $[0.25, 0.75]$ , which represents quite a bit of heterogeneity across households.

The next row of Table 1 considers how much of the variation in  $\alpha_i$  we can explain with observed covariates. The standard deviation of  $e_i$  gives a measure of the unexplained variation, and gives us an idea of whether household-level unobserved heterogeneity is an important feature of the data. If the standard deviation of  $e_i$  is very small, then fixed effects are not needed—conditioning on observed covariates would be sufficient. Our estimate of the standard deviation of the unexplained variation in  $\alpha_i$  is about 0.16. This is large relative to the overall estimated standard deviation of 0.26, and suggests that accounting for household-level unobserved heterogeneity is quite important.

Dunbar et al. (2021) use a random effects cross-sectional data approach to get at these same issues in a Malawian dataset. They find that the standard deviation of (unlogged) men’s resource shares in Malawi is 0.28, and that the standard deviation of the component of resource shares driven by observed covariates is 0.16. Our findings thus echo the findings in cross-sectional work (large variance in resource shares across households, only about half of which is explained by observed covariates) in a fixed-effects panel-data setting where unobserved heterogeneity in resource shares may be correlated with the household budget.

The next two rows give the covariance of  $\alpha_i$  and  $X_{it}$ . Here, we see that log resource shares  $\alpha_i$  strongly and statistically significantly negatively covary with observed household budgets (the implied correlation coefficients are close to  $-0.8$ ). This means that women in poor households are somewhat less poor than they appear (on the basis of their household budget), and women in richer households are somewhat less rich than they appear.

To fix ideas, we note that roughly 50 and 40 per cent of the households in our sample were under the World Bank societal poverty line (see World Bank (2018)) for a lower-middle income country of US\$3.50 per person per day in 2012 and 2015, respectively. Thus, the bottom half of this population is near subsistence. In contrast, the 95<sup>th</sup> percentile of the daily per-capita household expenditure distribution in our sample is roughly \$10 (or triple the poverty line) in both years. Since the correlation of resource shares with household expenditures is negative, those households under the poverty line (half the population) have relatively high women’s resource shares.

Under the assumption that women’s resource shares are smaller than men’s (recalling that the level of resource shares is not identified), this implies that households that are closer to subsistence have a more equal gender distribution of resources.<sup>24</sup>

The next two rows give the estimated standard deviation of women’s log shadow budgets. This is a scale-free parameter: it does not depend on the location normalization of  $\alpha_i$  (which corresponds to a scale normalization of shadow budgets). The estimated standard deviations are 0.35 and 0.38 in the two periods, respectively. We can compare these with the standard deviation of log-budgets, reported in Table 1, of 0.49 and 0.53. The point estimates suggest that there is *less* inequality in women’s shadow budgets than in household budgets. Although the confidence intervals are large, the test of the hypothesis that the standard deviation of log-budgets equals the standard deviation of log-shadow budgets rejects in both years.<sup>25</sup>

Thus, if we take these results at face value, there is *less* consumption inequality among women than household-level analysis would suggest. However, another implication of this is that there is *more* gender inequality than household level data would suggest. The reason is that household-level analysis of gender inequality pins gender inequality on over-representation of one gender in poorer households. In our data, all households have 1 man and 1 woman, so household-level analysis of gender inequality would show zero gender inequality. But, because women in richer households have smaller resource shares, this induces gender inequality even in these data.

Finding correlation between  $\alpha_i$  and household budgets is not sufficient to invalidate previous identification strategies for cross-sectional settings that rely on independence between resource shares and household budgets. The reason is that the independence required is conditional on other observed covariates. To get a handle on this, the bottom panel of Table 1 presents estimates of coefficients in a linear regression of normalized (to average 0.33) estimated resource shares  $\hat{\eta}_i$  on log-budgets  $\bar{X}_i$  and other covariates  $z_i$ .

The figure below shows the scatterplot of predicted resource shares versus the log household budget. Here, we see a lot of variation in resource shares, and it is clearly correlated with household budgets. The overall variation here provides an estimate

---

<sup>24</sup>This is consistent with findings like Bertrand et al. (2015) showing the importance of gender identity norms in the within-household distribution, but with an added constraint that all household members must meet a subsistence requirement.

<sup>25</sup>For  $H_0 : Var(X_{it}) = Var(\alpha_i + X_{it})$ , we have the following estimated test statistics and (confidence intervals). Period 1: 0.110(0.0236, 0.165); Period 2: 0.137(0.0527, 0.191).

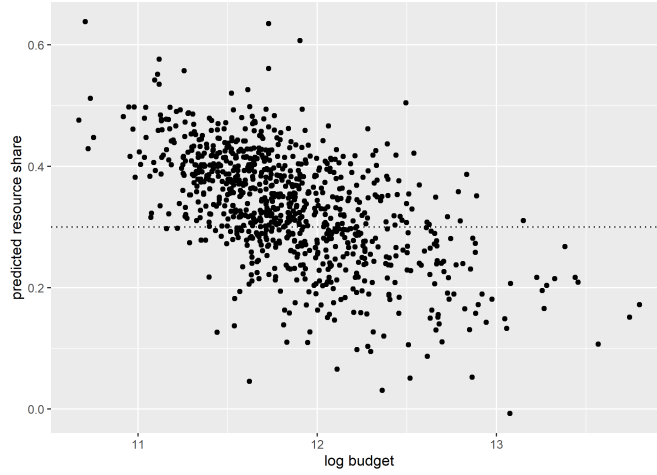


Figure 8.2: Scatterplot of predicted resource shares and log budget.

of the explained sum of squares in an infeasible regression of true resource shares  $\eta_i$  on  $\bar{X}_i$  and  $z_i$ .<sup>26</sup> We may construct an estimate of the total sum of squares of resource shares from our estimate of the standard deviation of  $\alpha_i$ . This yields an estimate of  $R^2$  in the infeasible regression, which we interpret as the fraction of variation in resource shares explained by observables.

In the first row of the bottom panel, we see that observed variables explain roughly half the variation in resource shares (the estimate of  $R^2$  is 0.52). This magnitude of explained variation is very close to that reported in Dunbar et al. (2021) in their cross-sectional estimate based on Malawian data. However, whereas the estimate in Dunbar et al. (2021) is conditional on the assumption that unobserved heterogeneity in resources shares is independent of the household budget, our estimate allows for correlation of resource shares with the household budget. This large magnitude of unexplained variation (roughly half) suggests that accounting for unobserved heterogeneity in resource shares is quite important.

Consider first the coefficient on  $\bar{X}_i$ . The estimated coefficient is  $-0.045$  and is statistically significantly different from 0. This means that, even after conditioning on other covariates (many of which are highly correlated with the budget), we still

---

<sup>26</sup>This artificial regression is infeasible because we observe (through  $g_{it}$ ) a prediction of  $\alpha_i + U_{it}$ , not of  $\alpha_i$  itself. However, because we have an estimate of the variance of  $\alpha_i$ , we can construct an estimate of the variance of  $\eta_i$  (subject to the scale normalization that it has a mean of 0.33). In our regression, the LHS variable is  $\exp \hat{g}_{it}$ , which is a prediction of  $\eta_i u_{it}$ . Since  $u_{it}$  are uncorrelated with  $\bar{X}_i$  and  $z$  by assumption, the explained sum of squares from this regression applies to  $\eta_i$ , and we can use it to form an estimate of  $R^2$ , which we report, along with a bootstrapped confidence interval.



see a significant relationship between resource shares and household budgets.

However, the magnitude of this effect is small. Conditional on  $z_i$ , the standard deviation of  $X_{it}$  is 0.43 in year 1 and 0.47 in year 2. Thus, comparing two households with identical  $z$  but which are one standard deviation apart in terms the household budget, we would expect the woman in the poorer household to have a resource share 2 percentage points higher than the woman in the richer household. Thus, the bulk of the variation that makes the standard deviation of women’s shadow budgets smaller than that of household budgets is not running through the dependence of resource shares on household budgets, but rather through the dependence of resource shares on other covariates that are correlated with household budgets.

We get a very precise estimate of the conditional dependence of resource shares on household budgets. Overall, then, we see that women’s resource shares are statistically significantly correlated with household budgets, even conditional on other observed characteristics. But, the estimated difference in resource shares at different household budgets is quite small. So, we take this as evidence that the identifying restrictions used by Dunbar et al. (2013) (and Dunbar et al. (2021)) may be false, though perhaps not *very* false. It does suggest that alternative identifying restrictions—such as those developed here with a panel model—may be useful.

The rows of Table 1 give several other coefficients that are comparable to other estimates in the literature. Calvi (2019) finds that women’s resource shares in India decline with the age of the woman. In these Bangladeshi data, we find evidence that women’s resource shares are strongly negatively correlated with the age of women and positively correlated with the age of men.

Dunbar et al. (2013) find that women’s resource shares in Malawi decline with the number of children. Here, we also see that pattern: households with 2 children have women’s resource shares 5 percentage points less than households with 1 child; households with 3 or 4 children have resource shares 12 percentage points less. Dunbar et al. (2013) also find that Malawian women’s resource shares are higher in households with girls than households with boys. We do not see evidence of this in rural Bangladesh: the estimated coefficient on the fraction of children that are girls statistically insignificantly different from 0.

In the Appendix, we also provide estimates analogous to Table 1 using a different assignable good: clothing. Under the model, using different assignable goods should

yield the same estimates of resource shares.<sup>27</sup> This is roughly what we find in our estimates using women’s clothing.

Our estimates use 8th order Bernstein polynomials to approximate the inverse demand functions. In Appendix D, we present estimates using Bernstein polynomials of order  $K = 1, 4, 8, 10$  and show that our finite-dimensional parameter estimates have roughly the same value for  $K \geq 8$ .

In summary, in these rural Bangladeshi households, we find evidence that women’s resource shares have substantial dependence on household-level unobserved heterogeneity and are slightly negatively correlated with household budgets. The former suggests that random-effects type approaches to the estimation of resource shares may be inadequate. The latter suggests that consumption inequality faced by women is actually smaller than household-level consumption inequality. It also suggests that cross-sectional identification strategies invoking independence of resource shares from household budgets, such as Dunbar et al. (2013), could be complemented by panel-based identification strategies such as ours.

## References

- Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics*, 93(2):203–228.
- Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics*, 95(1):1–23.
- Abrevaya, J. and Muris, C. (2020). Interval censored regression with fixed effects. *Journal of Applied Econometrics*, 35(2):198–216.
- Ai, C. and Gan, L. (2010). An alternative root-n consistent estimator for panel data binary choice models. *Journal of Econometrics*, 157(1):93–100.
- Altonji, J. G. and Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4):1053–1102.
- Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67:1341–1383.
- Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68:399–405.
- Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77(2):489–536.

---

<sup>27</sup>Food is a plausible *assignable* good (because if one person eats it, nobody else can), but it may not be non-shareable (because there may be scale economies in cooking). In contrast, clothing may be plausibly non-shareable, but it may not be assignable (because, e.g., mothers and daughters might wear each others’ clothes). See the Appendix for details on clothing estimates.

- Arellano, M. and Bonhomme, S. (2011). Nonlinear panel data analysis. *Annual Review of Economics*, 3(1):395–424.
- Arellano, M. and Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies*, 79(3):987–1020.
- Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. In Blundell, R., Newey, W., and Persson, T., editors, *Advances in Economics and Econometrics*, pages 381–409.
- Aristodemou, E. (2021). Semiparametric identification in panel data discrete response models. *Journal of Econometrics*, 220(2):253–271.
- Baetschmann, G., Staub, K. E., and Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society A*, 178(3):685–703.
- Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, 79(4):527–539.
- Bargain, O., Donni, O., and Kwenda, P. (2014). Intrahousehold distribution and poverty: Evidence from cote d’ivoire. *Journal of Development Economics*, 107:262–276.
- Bargain, O., Lacroix, G., and Tiberti, L. (2018). Validating the collective model of household consumption using direct evidence on sharing. Partnership for Economic Policy Working Paper 2018-06.
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5):9–49.
- Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association*, 45(250):164–180.
- Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.
- Bester, C. A. and Hansen, C. (2009). Identification of marginal effects in a non-parametric correlated random effects model. *Journal of Business and Economic Statistics*, 27(2):235–250.
- Bonhomme, S. (2012). Functional differencing. *Econometrica*, 80(4):1337–1385.
- Botosaru, I. and Muris, C. (2017). Binarization for panel models with fixed effects. cemmap working paper CWP31/17.
- Browning, M., Chiappori, P. A., and Lewbel, A. (2013). Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies*, 80(4):1267–1303.
- Calvi, R. (2019). Why are older women missing in india? the age profile of bargaining power and poverty. *Journal of Political Economy*. Forthcoming.
- Carrasco, M., Florens, J.-P., and Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6B, chapter 77. Elsevier, 1 edition.

- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225–238.
- Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica*, 78(1):159–168.
- Charlier, E., Melenberg, B., and van Soest, A. (2000). Estimation of a censored regression panel data model using conditional moment restrictions efficiently. *Journal of Econometrics*, 95:25–56.
- Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, 70(4):1683–1697.
- Chen, S. (2010a). An integrated maximum score estimator for a generalized censored quantile regression model. *Journal of Econometrics*, 155(1):90–98.
- Chen, S. (2010b). Non-parametric identification and estimation of truncated regression models. *Review of Economic Studies*, 77(1):127–153.
- Chen, S. (2010c). Root-N-consistent estimation of fixed-effect panel data transformation models with censoring. *Journal of Econometrics*, 159(1):222–234.
- Chen, S. (2012). Distribution-free estimation of the Box-Cox regression model with censoring. *Econometric Theory*, 28(3):680–695.
- Chen, S., Khan, S., and Tang, X. (2019). Exclusion Restrictions in Dynamic Binary Choice Panel Data Models: Comment on “Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors”. *Econometrica*, 87(5):1781–1785.
- Chen, S., Khan, S., and Tang, X. (2020). Identification and estimation of weakly separable models without monotonicity.
- Chen, S. and Zhou, X. (2012). Semiparametric estimation of a truncated regression model. *Journal of Econometrics*, 167(2):297–304.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, volume 6B, chapter 76, pages 5549–5632. Elsevier.
- Chernozhukov, V., Fernández-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2):535–580.
- Chernozhukov, V., Fernández-Val, I., Hoderlein, S., Holzmann, H., and Newey, W. (2015). Nonparametric identification in panels using quantiles. *Journal of Econometrics*, 188(2):378–392.
- Chernozhukov, V., Fernández-Val, I., and Weidner, M. (2020). Network and panel quantile effects via distribution regression. *Journal of Econometrics*.
- Chiappori, P., Komunjer, I., and Kristensen, D. (2015). Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, 188:22–39.
- Chiappori, P. A. (1988). Rational household labor supply. *Econometrica*, 56(1):63–90.
- Chiappori, P. A. (1992). Collective labor supply and welfare. *Journal of Political Economy*, 100(3):437–467.
- Chiappori, P. A. and Ekeland, I. (2009). The microeconomics of efficient group behavior: Identification. *Econometrica*, 77(3):763–799.
- Chiappori, P. A. and Kim, J. H. (2017). A note on identifying heterogeneous sharing

- rules. *Quantitative Economics*, 8(1):201–218.
- Chiappori, P.-A. and Mazzoco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, 55(3):985–1045.
- Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79:1541–1565.
- Das, M. and van Soest, A. (1999). A panel data model for subjective information on household income growth. *Journal of Economic Behavior & Organization*, 40(4):409–426.
- Davezies, L., D’Haultfoeuille, X., and Laage, L. (2021). Identification and estimation of average marginal effects in fixed effects logit models.
- Deaton, A. and Muellbauer, J. (1980). *Economics and consumer behavior*. University Press, Cambridge.
- Deaton, A. S. (1997). *The Analysis of Household Surveys : A Microeconomic Approach to Development Policy*. Washington, D.C. : The World Bank.
- Doksum, K. A. and Gasko, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review*, 58:243–252.
- Dunbar, G., Lewbel, A., and Pendakur, K. (2021). Identification of random resource shares in collective households without preference similarity restrictions. *Journal of Business and Economic Statistics*, 39(2):402–421.
- Dunbar, G. R., Lewbel, A., and Pendakur, K. (2013). Children’s resources in collective households: identification, estimation, and an application to child poverty in Malawi. *American Economic Review*, 103(1):438–471.
- Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working paper.
- Evdokimov, K. (2011). Nonparametric identification of a nonlinear panel model with application to duration analysis with multiple spells. Working paper.
- Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85.
- Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel data models with large N, T. *Journal of Econometrics*, 192(1):291–312.
- Freyberger, J. (2018). Nonparametric panel data models with interactive fixed effects. *The Review of Economic Studies*, 85(3):1824–1851.
- Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.
- Hoderlein, S. and White, H. (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics*, 168(2):300–314.
- Honoré, B. (1992). Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60(3):533–565.
- Honoré, B. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies*, 60(1):241–246.

- Honoré, B. and Hu, L. (2004). Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122(2):293–316.
- Honoré, B. and Kyriazidou, E. (2000). Panel discrete choice models with lagged dependent variables. *Econometrica*, 78(4):839–874.
- Honoré, B. and Lewbel, A. (2002). Semiparametric binary choice panel data models without strictly exogenous regressors. *Econometrica*, 70(5):2053 – 2063.
- Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64(1):103–137.
- Horowitz, J. L. and Lee, S. (2004). Semiparametric estimation of a panel data proportional hazards model with fixed effects. *Journal of Econometrics*, 119(1):155–198.
- Khan, S., Ponomareva, M., and Tamer, E. (2016). Identification of panel data models with endogenous censoring. *Journal of Econometrics*, 194(1):57–75.
- Khan, S. and Tamer, E. (2007). Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics*, 136(1):251–280.
- Khan, S. and Tamer, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, 78(6):2021–2042.
- Lee, S. (2008). Estimating panel data duration models with censored data. *Econometric Theory*, 24(5):1254–1276.
- Lewbel, A. (2014). An overview of the special regressor method. In *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pages 38–62. Oxford University Press.
- Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *JOURNAL OF ECONOMIC LITERATURE*, 57(4):835–903.
- Lewbel, A. and Yang, T. (2016). Identifying the average treatment effect in ordered treatment models without unconfoundedness. *Journal of Econometrics*, 195:1–22.
- Ligon, E. (1998). Risk sharing and information in village economies. *The Review of Economic Studies*, 65(4):847–864.
- Lise, J. and Yamada, K. (2019). Household sharing and commitment: Evidence from panel data on individual expenditures and time use. *The Review of Economic Studies*, 86(5):2184–2219.
- Magnac, T. (2004). Panel binary variables and sufficiency: Generalizing conditional logit. *Econometrica*, 72(6):1859–1876.
- Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205–228.
- Manski, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3):313–333.
- Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2):357–362.
- Mazzocco, M. (2007). Household intertemporal behaviour: A collective characterization and a test of commitment. *The Review of Economic Studies*, 74(3):857–895.
- Mazzocco, M., Ruiz, C., and Yamaguchi, S. (2014). Labor supply and household dynamics. *American Economic Review*, 104(5):354–59.

- Muellbauer, J. (1975). Aggregation, income distribution and consumer demand. *The Review of Economic Studies*, 42(4):525–543.
- Muellbauer, J. (1976). Community preferences and the representative consumer. *Econometrica*, (5):979–999.
- Muris, C. (2017). Estimation in the fixed effects ordered logit model. *The Review of Economics and Statistics*, 99(3):465–477.
- Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen:Denmark Paedagogiske Institut.
- Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies*, 57(2):167–181.
- van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5, chapter 55, pages 3381–3460. Elsevier.
- Voena, A. (2015). Yours, mine and ours: Do divorce laws affect the intertemporal behavior of married couples? *American Economic Review*, 105(8):2295–2332.
- Vreyer, P. D. and Lambert, S. (2016). Intra-household inequalities and poverty in senegal. mimeo, Paris School of Economics.
- Wang, J. and Ghosh, S. (2012). Shape restricted nonparametric regression with bernstein polynomials. *Computational Statistics and Data Analysis*, 56:2729–2741.
- Wilhelm, D. (2015). Identification and estimation of nonparametric panel data regressions with measurement error. Cemmap working paper CWP34/15.
- World Bank (2018). Poverty and shared prosperity 2018: piecing together the poverty puzzle. <https://www.worldbank.org/en/publication/poverty-and-shared-prosperity-2018>.

# ONLINE APPENDICES

## A Proofs

### A.1 Proof of Lemma (1)

*Proof.* Define  $\bar{D} = 1 \{D_1(y_1) + D_2(y_2) = 1\}$ . The proof consists in showing the following:

$$\begin{aligned}
 & \text{med} (D_2(y_2) - D_1(y_1) | X, \bar{D} = 1) & (A.1) \\
 = & \text{sgn} (P(D(y_1, y_2) = (0, 1) | X, \bar{D} = 1) - P(D(y_1, y_2) = (1, 0) | X, \bar{D} = 1)) & (A.2) \\
 = & \text{sgn} \left( \frac{P(D(y_1, y_2) = (0, 1), \bar{D} = 1 | X)}{P(\bar{D} = 1 | X)} - \frac{P(D(y_1, y_2) = (1, 0), \bar{D} = 1 | X)}{P(\bar{D} = 1 | X)} \right) & (A.3) \\
 = & \text{sgn} (P(D(y_1, y_2) = (0, 1), \bar{D} = 1 | X) - P(D(y_1, y_2) = (1, 0), \bar{D} = 1 | X)) & (A.4) \\
 = & \text{sgn} (P(D(y_1, y_2) = (0, 1) | X) - P(D(y_1, y_2) = (1, 0) | X)) & (A.5) \\
 = & \text{sgn} (P(D_2(y_2) = 1 | X) - P(D_1(y_1) = 1 | X)) & (A.6) \\
 = & \text{sgn} (\Delta X \beta - \gamma(y_1, y_2)) & (A.7)
 \end{aligned}$$

where (A.2) follows since the random variable  $D_2(y_2) - D_1(y_1) \in \{-1, 1\}$ , which implies that

$$\begin{aligned}
 & \text{med} (D_2(y_2) - D_1(y_1) | X, \bar{D} = 1) \\
 = & \begin{cases} 1 & \text{if } P(D(y_1, y_2) = (0, 1) | X, \bar{D} = 1) > P(D(y_1, y_2) = (1, 0) | X, \bar{D} = 1) \\ -1 & \text{if } P(D(y_1, y_2) = (0, 1) | X, \bar{D} = 1) < P(D(y_1, y_2) = (1, 0) | X, \bar{D} = 1) \end{cases} ,
 \end{aligned}$$

(A.3) follows from the definition of conditional probability, (A.4) follows since the sign function is not affected by scaling both quantities by the same positive factor (the denominator), (A.5) follows by the definition of  $\bar{D}$ , and (A.6) follows since:

$$\begin{aligned}
 P(D_2(y_2) = 1 | X) &= P(D(y_1, y_2) = (0, 1) | X) + P(D(y_1, y_2) = (1, 1) | X) \\
 P(D_1(y_1) = 1 | X) &= P(D(y_1, y_2) = (1, 0) | X) + P(D(y_1, y_2) = (1, 1) | X)
 \end{aligned}$$

Finally, (A.7) follows from Assumption 2(ii), which implies that, e.g.,

$$P(D_2(y_2) = 1 | \alpha, X) > P(D_1(y_1) = 1 | \alpha, X) \Leftrightarrow \alpha + X_2 \beta - h_2^-(y_2) > \alpha + X_1 \beta - h_1^-(y_1) .$$

Integrating both sides over the conditional distribution of  $\alpha$  given  $X$  obtains:

$$\begin{aligned}
 P(D_2(y_2) = 1 | X) > P(D_1(y_1) = 1 | X) &\Leftrightarrow X_2 \beta - h_2^-(y_2) > X_1 \beta - h_1^-(y_1) \\
 &\Leftrightarrow \Delta X \beta - \gamma(y_1, y_2) > 0 .
 \end{aligned}$$



Result (3.10) now follows.  $\square$

## A.2 Proof of Theorem 1

*Proof.* Following Manski (1985), it suffices to show that for an arbitrary  $\theta \in \Theta$ ,  $\theta \neq \theta_0 \equiv \theta_0(y_1, y_2)$ ,

$$P(W\theta < 0 \leq W\theta_0) + P(W\theta_0 < 0 \leq W\theta) > 0. \quad (\text{A.8})$$

Our proof follows very closely that in Manski (1985), with  $W\theta$  taking the role of  $xb$  and  $W\theta_0$  taking the role of  $x\beta$ . However, our scale normalization is different.

Without loss of generality, let  $X_K$  be the continuous regressor in Assumption 3(i). Separate  $\Delta X = (\Delta X_{-K}, \Delta X_K)$  where the first component  $\Delta X_{-K}$  represents all covariates except the  $K$ -th one. Similarly, for any  $\theta = (\beta, \gamma) \in \Theta$ , separate  $\beta = (\beta_{-K}, \beta_K)$ . Furthermore denote  $W_{-K} = (\Delta X_{-K}, -1)$  and  $\theta_{-K} = (\beta_{-K}, \gamma)$ .

Assume that the associated regression coefficient  $\beta_{0,K} > 0$ . The case  $\beta_{0,K} < 0$  follows similarly. Let  $\theta = (\beta, \gamma) \in \Theta$ ,  $\theta \neq \theta_0$ . As in Manski (1985, p. 318), consider three cases: (i)  $\beta_K < 0$ ; (ii)  $\beta_K = 0$ ; (iii)  $\beta_K > 0$ .

**Cases (i) and (ii).**  $\beta_K \leq 0$ . The proof is identical to that in Manski (1985), with  $X\beta$  replaced by  $W\theta$ . The fact that we use a different scale normalization does not come into play.

**Case (iii).**  $\beta_K > 0$ . note that

$$P(W\theta < 0 \leq W\theta_0) = P\left(-\frac{W_{-K}\theta_{0,-K}}{\beta_{0,K}} < \Delta X_K < -\frac{W_{-K}\theta_{-K}}{\beta_K}\right).$$

$$P(W\theta_0 < 0 \leq W\theta) = P\left(-\frac{W_{-K}\theta_{-K}}{\beta_K} < \Delta X_K < -\frac{W_{-K}\theta_{0,-K}}{\beta_{0,K}}\right).$$

By assumption 4,  $\frac{\beta_{-K}}{\beta_K} \neq \frac{\beta_{0,-K}}{\beta_{0,K}}$ , which shows that the first  $K$  components of the vector  $\theta$  are not a scalar multiple of the first  $K$  components of the vector  $\theta_0$ . Therefore,  $\theta$  is not a scalar multiple of  $\theta_0$ . In particular,  $\frac{\theta_{0,-K}}{\beta_{0,K}} \neq \frac{\theta_{-K}}{\beta_K}$ . Additionally, assumption 3(ii) implies that  $P\left(\frac{W_{-K}\theta_{0,-K}}{\beta_{0,K}} \neq \frac{W_{-K}\theta_{-K}}{\beta_K}\right) > 0$ . Hence at least one of the two probabilities above is positive so that (A.8) holds.  $\square$

## A.3 Proof of Theorem 2

*Proof.* Under Assumption 5,  $h_1^-(y_0) = 0$ . Using the pair  $(y_0, y_2)$  for binarization thus obtains identification of

$$\begin{aligned} \gamma(y_0, y_2) &= h_2^-(y_2) - h_1^-(y_0) \\ &= h_2^-(y_2). \end{aligned}$$

By varying  $y_2 \in \underline{\mathcal{Y}}$ , we identify the function  $h_2^-$  from the binary choice models associated with  $\{D(y_0, y_2) = (D_1(y_0), D_2(y_2)), y_2 \in \underline{\mathcal{Y}}\}$ .

The pairs  $(y_0, y_2)$  and  $(y_1, y_2)$  identify the difference

$$\begin{aligned}\gamma(y_0, y_2) - \gamma(y_1, y_2) &= (h_2^-(y_2) - h_1^-(y_0)) - (h_2^-(y_2) - h_1^-(y_1)) \\ &= h_1^-(y_1).\end{aligned}$$

By varying  $y_1 \in \underline{\mathcal{Y}}$  we therefore identify  $h_1^-$ .

Thus, the functions  $h_1^-$  and  $h_2^-$  are identified. Because of monotonicity of  $h_t$  (Assumption 1), and because  $\mathcal{Y}$  is known,  $h_t^-$  contains all the information about the pre-image of  $h_t$ . Knowledge of the pre-image of a function is equivalent to knowledge of the function itself. Therefore,  $h_t$  can be identified from  $h_t^-$ .  $\square$

## A.4 Proof of Theorem (3)

*Proof.* For the panel data binary choice model with logit errors, we obtain

$$P(D_2(y_2) = 1 | \bar{D}(y_1, y_2) = 1, X, \alpha) \tag{A.9}$$

$$= \frac{P(D_2(y_2) = 1, \bar{D}(y_1, y_2) = 1 | X, \alpha)}{P(\bar{D}(y_1, y_2) = 1 | X, \alpha)} \tag{A.10}$$

$$= \frac{P(D_1(y_1) = 0, D_2(y_2) = 1 | X, \alpha)}{P(\bar{D}(y_1, y_2) = 1 | X, \alpha)} \tag{A.11}$$

$$= \frac{P(D_1(y_1) = 0, D_2(y_2) = 1 | X, \alpha)}{P(D_1(y_1) = 0, D_2(y_2) = 1 | X, \alpha) + P(D_1(y_1) = 1, D_2(y_2) = 0 | X, \alpha)} \tag{A.12}$$

$$= \frac{1}{1 + \frac{P(D_1(y_1)=1, D_2(y_2)=0 | X, \alpha)}{P(D_1(y_1)=0, D_2(y_2)=1 | X, \alpha)}} \tag{A.13}$$

$$= \Lambda(\Delta X \beta - \gamma(y_1, y_2)) \tag{A.14}$$

where A.10 follows from the definition of a conditional probability; A.11 follows because  $D_2 = 1$  and  $\bar{D} = 1$  are equivalent to  $D_1 = 0$  and  $D_2 = 1$ ; A.12 follows because  $D_1 + D_2 = 1$  happens precisely when either  $(D_1, D_2) = (1, 0)$  or  $(D_1, D_2) = (0, 1)$ ; A.13 follows by dividing by the numerator; and the final expression follows by the argument below.

Note that  $\frac{P(D_1(y_1)=1, D_2(y_2)=0 | X, \alpha)}{P(D_1(y_1)=0, D_2(y_2)=1 | X, \alpha)}$  equals

$$\frac{P(D_1(y_1) = 1|X, \alpha)P(D_2(y_2) = 0|X, \alpha)}{P(D_1(y_1) = 0|X, \alpha)P(D_2(y_2) = 1|X, \alpha)} \quad (\text{A.15})$$

$$= \frac{\Lambda(\alpha + X_1\beta - h_1^-(y_1)) [1 - \Lambda(\alpha + X_2\beta - h_2^-(y_2))]}{[1 - \Lambda(\alpha + X_1\beta - h_1^-(y_1))] \Lambda(\alpha + X_2\beta - h_2^-(y_2))} \quad (\text{A.16})$$

$$= \frac{\exp(\alpha + X_1\beta - h_1^-(y_1))}{\exp(\alpha + X_2\beta - h_2^-(y_2))} \quad (\text{A.17})$$

$$= \exp((X_1 - X_2)\beta - (h_1^-(y_1) - h_2^-(y_2))), \quad (\text{A.18})$$

where A.15 follows from serial independence of  $(U_1, U_2)$  conditional on  $(X, \alpha)$ ; A.16 from the logit model specification; and A.17 follows from

$$\Lambda(u) / (1 - \Lambda(u)) = \exp(u).$$

The discussion above implies that A.9 does not depend on  $\alpha$ . Hence,

$$\begin{aligned} p(X, y_1, y_2) &\equiv P(D_2(y_2) = 1 | \bar{D}(y_1, y_2) = 1, X) \\ &= \Lambda(\Delta X\beta - \gamma(y_1, y_2)) \\ &= \Lambda(W\theta(y_1, y_2)). \end{aligned}$$

and note that  $p(X, y_1, y_2)$  is identified from the distribution of  $(Y, X)$ , which is assumed to be observed. Then

$$\theta(y_1, y_2) = [E(W'W)]^{-1} E(W'\Lambda^{-1}(p(X, y_1, y_2)))$$

by invertibility of  $\Lambda$  and the full rank assumption on  $E[W'W]$ . This establishes identification of  $\beta$  and  $\gamma(y_1, y_2)$ . The proof in Section A.3 applies, which shows the identification of  $h_1$  and  $h_2$ .  $\square$

## A.5 Proof of Theorems 4, 5, and 6

*Proof.* Proof of Theorems 4 and 5. (a) Note that, without Assumption 5, it follows immediately from the proof of Theorem 2 that we can only identify

$$\{h_t^{-1}(y) - c_1, y \in \mathcal{Y}, t = 1, 2\},$$

i.e. we identify the functions  $g_t(y)$ ,  $t = 1, 2$ .

Because the functions  $g_1, g_2$  are identified, and because the distribution of  $(Y, X)$  is observable, we can identify the distribution of the left hand side of the relation

below:

$$\begin{aligned} g_t(Y_t) - X_t\beta &= h_t^{-1}(Y_t) - X_t\beta - c_1 \\ &= \alpha - U_t - c_1, \end{aligned}$$

for  $t = 1, 2$ .

It follows that for all  $x \in \mathcal{X}$ ,

$$\begin{aligned} \mu(x) &= E[\alpha | X = x] \\ &= E[\alpha - U_t | X = x] + E[U_t | X = x] \\ &= E[h_t^{-1}(Y_t) - X_t\beta | X = x] + m \\ &= E[g_t(Y_t) - X_t\beta | X = x] + c_1 + m. \end{aligned}$$

is identified up to the constants  $c_1$  and  $m$ .

The difference in conditional means at any two values  $x, x' \in \mathcal{X}$  is therefore identified and given by:

$$\begin{aligned} \mu(x) - \mu(x') &= \\ &= (E[g_t(Y_t) - X_t\beta | X = x] + c_1 + m) - (E[g_t(Y_t) - X_t\beta | X = x'] + c_1 + m) \\ &= E[g_t(Y_t) - X_t\beta | X = x] - E[g_t(Y_t) - X_t\beta | X = x']. \end{aligned}$$

(b) To see that the conditional variance is identified, note that for all  $x \in \mathcal{X}$ ,

$$\begin{aligned} Cov(g_2(Y_2) - X_2\beta, g_1(Y_1) - X_1\beta | X = x) &= \\ &= Cov(h_2^{-1}(Y_2) - X_2\beta - c_1, h_1^{-1}(Y_1) - X_1\beta - c_1 | X = x) \\ &= Cov(\alpha - U_1 - c_1, \alpha - U_2 - c_1 | X = x) \\ &= Var(\alpha | X = x) - Cov(\alpha, U_1 | X = x) - Cov(\alpha, U_2 | X = x) + Cov(U_1, U_2 | X = x) \\ &= Var(\alpha | X = x) = \sigma_\alpha^2(x), \end{aligned}$$

where the first equality follows from the definition of  $g_t$ ; the second from the model; the third equality follows from the linearity of the covariance; and the fourth equality uses assumption (4c) and (4d).  $\square$

*Proof.* Proof of Theorem 6. Let  $t$  be the time period for which  $h_t$  is invertible. By

the model equations and the binomial theorem, we have that for any  $m \in \mathbb{N}$ ,

$$\begin{aligned} E \left[ (h_t^{-1}(Y_t) - X_t\beta)^m \mid X = x \right] &= E \left[ (\alpha - U_t)^m \mid X = x \right] \\ &= \sum_{k=0}^m \binom{m}{k} E \left[ \alpha^k (-U_t)^{m-k} \mid X = x \right] \\ &\equiv P_m(x). \end{aligned}$$

Since  $h_t$  and  $\beta$  are identified by Theorem 3, so is  $P_m(x)$ .

Setting  $V_t = -U_t$ , using the law of iterated expectations, and because  $U_t \mid \alpha, X$  does not depend on  $(\alpha, X)$ ,

$$\begin{aligned} \sum_{k=0}^m \binom{m}{k} E \left[ \alpha^k V_t^{m-k} \mid X = x \right] &= \sum_{k=0}^m \binom{m}{k} E \left[ \alpha^k \mid X = x \right] E \left[ V_t^{m-k} \right] \\ &\equiv \sum_{k=0}^m \binom{m}{k} a_k(x) l_{m-k}, \end{aligned} \quad (\text{A.19})$$

where

$$a_k(x) = E \left[ \alpha^k \mid X = x \right]$$

are the quantities of interest,  $k \in \mathbb{N}$ , and

$$l_{m-k} = E \left[ V_t^{m-k} \right]$$

are the known moments of the standard logistic distribution.

For an arbitrary  $0 \leq j \leq m$ , compute the difference for A.19 evaluated at  $j+1$  and  $j$ :

$$\begin{aligned} P_{j+1}(x) - P_j(x) &= \sum_{k=0}^{j+1} \binom{j+1}{k} a_k(x) l_{j+1-k} - \sum_{k=0}^j \binom{j}{k} a_k(x) l_{j-k} \\ &= a_{j+1}(x) l_0 + \sum_{k=0}^j a_k(x) \left[ \binom{j+1}{k} l_{j+1-k} - \binom{j}{k} l_{j-k} \right] \\ &= a_{j+1}(x) + \sum_{k=0}^j a_k(x) b_{kj}, \end{aligned}$$

where  $l_0 = E(V_t^0) = 1$ , and

$$b_{kj} \equiv \binom{j+1}{k} l_{j+1-k} - \binom{j}{k} l_{j-k}$$

are the coefficients on  $a_k(x)$  for a given  $j$ . For each  $k, j$ ,  $b_{kj} \neq 0$ . This follows from the properties of the logistic distribution. It is obvious that the 0th, 1st, and 2nd moments are  $l_0 = 1, l_1 = 0, l_2 = \pi^2/3$ . For moments beyond the second, recall that  $l_n = \pi^2(2^n - 2)|B_n|$ , where  $B_n$  is the  $n$ -th Bernoulli number. This implies that  $l_n > 0$  for all even  $n$ , and  $l_n = 0$  for all odd  $n \geq 3$ . The  $b_{kj}$  are differences between adjacent moments, i.e.  $(l_j, l_{j-1})$ , multiplied by binomial coefficients. By construction, one of them is zero and one of them is nonzero, so that  $b_{kj} \neq 0$  for all  $k, j$ .

Thus, the  $a_k(x), k = 1, \dots, m$  are solutions to the  $m$ -dimensional linear system that gathers

$$P_{j+1}(x) - P_j(x) - b_{0j} = a_{j+1}(x) + \sum_{k=1}^j a_k(x) b_{kj},$$

where we used that  $a_0(x) b_{0j} = b_{0j}$  and then moved this known quantity to the lefthandside.

In matrix form, the resulting system of  $m$  equations in the  $m$  unknowns  $a_k(x)$  is:

$$\begin{bmatrix} P_1(x) - P_0(x) - b_{01} \\ P_2(x) - P_1(x) - b_{02} \\ \vdots \\ P_{m-1}(x) - P_{m-2}(x) - b_{0,m-1} \\ P_m(x) - P_{m-1}(x) - b_{0,m} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ b_{11} & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & 1 & \ddots & \vdots \\ b_{1,m-2} & b_{2,m-2} & b_{3,m-2} & 1 & 0 \\ b_{1,m-1} & b_{2,m-1} & \cdots & b_{m-1,m-1} & 1 \end{bmatrix} \begin{bmatrix} a_1(x) \\ a_2(x) \\ \vdots \\ a_{m-1}(x) \\ a_m(x) \end{bmatrix}$$

The matrix of coefficients is a unit lower triangular matrix, which means that the  $a_k(x)$  can be uniquely solved for. This concludes the proof.  $\square$

## B Details on GMM estimation

*Degree of Bernstein polynomial.* Implementing this method requires the selection of the degree of the Bernstein polynomial,  $K$ . While developing a formal selection rule for this parameter would be desirable, it is beyond the scope of the present paper. Nonetheless, we adopt an informal selection rule for the number of Bernstein basis functions – the smoothing parameter – based on the following observation. In our semiparametric setting, the estimators are known to have the same asymptotic distribution for a range of smoothing parameters (see, e.g., Chen (2007)). When the number of Bernstein basis functions is small, the bias dominates and the estimates exhibit a decreasing bias as the number of terms increases. On the other hand, when the number of basis functions is large, the statistical noise dominates. We implement our estimation method over a range of smoothing parameter values, that is,  $K \in \{1, 2, \dots, 12\}$ , in search of a region where the estimates are not very sensitive to small variations in the smoothing parameter. We select the mid-point of that region, so our main results use  $K = 8$ . We present results for  $K \in \{1, 4, 8, 10\}$  in the

Appendix. In the Appendix table, “X” denotes a case where an estimated variance is negative.

*Confidence bands.* The confidence bands presented here are computed via the nonparametric bootstrap. We acknowledge that these may not be valid as the shape constraints may be binding, see for example Andrews (1999, 2000). Furthermore, they would have to reflect the choice of the order of the Bernstein polynomial. The theory to construct valid uniform confidence bands for our estimator is not yet available. We leave this as a topic for future research. Two possible solutions are to not impose monotonicity on the estimator for  $h_t$  or to use a subset of households with fixed  $-z$  and implement the estimator proposed by Botosaru and Muris (2017). Both monotonicity and accounting for household characteristics are essential components of our empirical application, so we do not pursue these approaches.

We use 1,000 bootstrap replications for all our results. We report pointwise 95% confidence bands. All reported estimates in the main text are bias-corrected using the bootstrap.

## C Descriptive Statistics

$n = 871$ Variable	raw data				top- and bottom-coded and normalized	
	Mean	Std Dev	Min	Max	Mean	Std Dev
$X_{i1}$	11.68	0.49	10.24	13.68		
$X_{i2}$	12.02	0.53	10.70	14.08		
$\Delta X_i$	0.34	0.43	-0.94	2.31		
$Y_{i1}$	1561	1277	0	13500		
$Y_{i2}$	1912	1727	0	15000		
age of woman	32.76	7.64	19.00	90.00	0.35	0.25
age of man	39.77	10.76	15.00	105.00	0.37	0.25
2 children	0.47	0.50	0.00	1.00	0.47	0.50
3 or 4 children	0.19	0.39	0.00	1.00	0.19	0.39
age of children	8.37	2.75	1.00	14.00	0.50	0.25
fraction girl children	0.46	0.39	0.00	1.00	0.46	0.39
education of woman	4.10	3.43	0.00	10.00	0.44	0.37
education of man	3.41	3.74	0.00	10.00	0.34	0.37

Table 2: Summary statistics.





# D Additional estimates

	1			4			8			10		
	Estimate	q(2.5)	q(97.5)	Estimate	q(2.5)	q(97.5)	Estimate	q(2.5)	q(97.5)	Estimate	q(2.5)	q(97.5)
<b>Variability in unobservables: Food</b>												
sd_a	0.3574	0.2258	0.5025	0.3197	0.2210	0.4376	0.2647	0.1518	0.3707	0.2286	0.0462	0.3592
sd_e	0.1947	0.1401	0.2377	0.1845	0.1413	0.2202	0.1637	0.1262	0.1931	0.1631	0.1308	0.1908
sd_ax1	-0.0506	-0.0934	-0.0085	-0.0805	-0.1300	-0.0246	-0.0901	-0.1292	-0.0429	-0.0781	-0.1166	-0.0339
sd_ax2	-0.0549	-0.0979	-0.0129	-0.0832	-0.1313	-0.0277	-0.1034	-0.1418	-0.0561	-0.0930	-0.1310	-0.0497
cov_ax1	0.5159	0.4067	0.6465	0.4272	0.3279	0.5763	0.3537	0.2720	0.4605	0.3635	0.2846	0.4583
cov_ax2	0.5444	0.4489	0.6637	0.4636	0.3757	0.6107	0.3763	0.3010	0.4760	0.3822	0.3104	0.4781
<b>Regression coefficients: Food</b>												
$R^2 : \bar{X}_i, z_i$ on $\eta_i$	0.6749	0.5292	0.8088	0.6341	0.4179	0.8026	0.5205	0.3361	0.6508	0.5369	0.3797	0.6447
$\bar{X}_i$	-0.0353	-0.0643	-0.0083	-0.0237	-0.0534	0.0039	-0.0452	-0.0700	-0.0196	-0.0481	-0.0731	-0.0234
age_women	-0.1095	-0.4028	0.2182	-0.2807	-0.5228	-0.0498	-0.2513	-0.4437	-0.0545	-0.3438	-0.5891	-0.1424
age_men	-0.0284	-0.3205	0.2123	0.1146	-0.0846	0.2811	0.2845	0.1389	0.4191	0.2904	0.1306	0.4392
children2	-0.1948	-0.3296	-0.0860	-0.1955	-0.3180	-0.0884	-0.0502	-0.1694	0.0413	-0.0821	-0.2229	0.0167
children3p	-0.2134	-0.3372	-0.1132	-0.2231	-0.3440	-0.0991	-0.1248	-0.2407	0.0140	-0.1159	-0.2466	0.0197
age_children	-0.3004	-0.5448	-0.0620	-0.0819	-0.2875	0.1477	-0.0095	-0.1979	0.2151	0.0949	-0.0910	0.3209
frac_girl	-0.0338	-0.1724	0.1209	0.0674	-0.0487	0.2049	0.0136	-0.0881	0.1246	-0.0322	-0.1466	0.0854
edu_women	0.0250	-0.1319	0.1803	0.0012	-0.1546	0.2036	-0.0325	-0.1519	0.1198	-0.1096	-0.2278	0.0132
edu_men	0.0079	-0.1213	0.1533	-0.1385	-0.2764	0.0091	-0.1330	-0.2589	-0.0015	-0.0202	-0.1501	0.1298
<b>Variability in unobservables: Cloth</b>												
sd_a	0.4729	X	0.7989	0.4224	0.2622	0.6944	0.2690	0.1472	0.4265	0.2443	0.1142	0.3749
sd_e	0.2350	X	0.3936	0.2131	0.1139	0.2875	X	X	0.1210	X	X	X
sd_ax1	-0.1254	-0.2157	-0.0341	-0.0503	-0.1049	0.0304	-0.1394	-0.1745	-0.0994	-0.1659	-0.1964	-0.1340
sd_ax2	-0.1185	-0.2186	-0.0207	-0.0903	-0.1583	-0.0043	-0.1832	-0.2211	-0.1406	-0.1903	-0.2234	-0.1537
cov_ax1	0.4633	X	0.7828	0.5667	0.4144	0.8386	0.1927	X	0.4568	X	X	0.3213
cov_ax2	0.5203	X	0.8248	0.5319	0.3654	0.8218	X	X	0.4190	X	X	0.3217
<b>Regression coefficients: Cloth</b>												
$R^2 : \bar{X}_i, z_i$ on $\eta_i$	11.2428	10.9458	14.5224	0.3753	-0.0344	1.0855	0.8791	0.6307	1.2056	1.1260	0.9142	1.3417
$\bar{X}_i$	0.1716	0.0975	0.2610	0.0015	-0.0509	0.0574	-0.0925	-0.1245	-0.0591	-0.1488	-0.1738	-0.1200
age_women	0.2658	-0.4693	1.1146	-0.3608	-0.6580	-0.0759	-0.2666	-0.4136	-0.0894	-0.1123	-0.2224	0.0056
age_men	0.1706	-0.3577	0.8011	0.5223	0.3033	0.8700	0.2860	0.1753	0.4254	0.1840	0.0961	0.2934
children2	-0.5101	-0.8971	-0.1832	0.0950	-0.0863	0.3863	0.0495	-0.0380	0.1572	0.0623	0.0008	0.1369
children3p	-0.3017	-0.7165	0.2456	-0.1110	-0.2746	0.1671	-0.0377	-0.1448	0.2903	-0.0549	-0.1511	0.2868
age_children	-1.1070	-1.7016	-0.5342	-0.0087	-0.2421	0.2433	0.0449	-0.0820	0.1722	-0.0925	-0.2044	-0.0050
frac_girl	-0.0222	-0.3256	0.3410	-0.3711	-0.5782	-0.2411	0.0926	0.0196	0.1633	0.0620	0.0028	0.1103
edu_women	-0.0821	-0.5592	0.3867	-0.0727	-0.2605	0.1185	-0.2170	-0.3154	-0.1348	-0.1029	-0.1815	-0.0414
edu_men	-0.5244	-0.9658	-0.1072	0.0746	-0.1260	0.4180	0.0420	-0.0497	0.1742	0.0167	-0.0537	0.1036

Table 3: Summary statistics for log resource shares and their projection on demographics and budget.