

# Gorman revisited:

## Nonparametric conditions for exact linear aggregation

Laurens Cherchye\*, Ian Crawford†, Bram De Rock‡and Frederic Vermeulen§

March 26, 2013

### Abstract

In the tradition of Afriat (1967), Diewert (1973) and Varian (1982), we provide a revealed preference characterisation of exact linear aggregation. This guarantees that aggregate demand can be written as a function of prices and aggregate income alone, while abstracting from income-distributional aspects. We also establish nonparametric conditions for individual consumption to be representable in terms of Gorman Polar Form preferences. Our results are simple and complement those of Gorman (1953, 1961). We illustrate the practical usefulness of our results by means of an empirical application to a Spanish balanced microdata panel.

**Acknowledgments:** Some of our results in the current paper already appeared in an earlier manuscript entitled “Aggregation without the aggravation? Nonparametric analysis of the representative consumer”, which should now be considered superseded. We are very grateful to Richard Blundell, Don Brown, Martin Browning, Erwin Diewert, Federico Echenique, Arthur Lewbel, Peter Neary, Krishna Pendakur and seminar participants at Bergen, CIREQ Montreal, DIW Berlin, Keele, Malaga, Oxford, Paris Dauphine and Tilburg for helpful comments and suggestions.

## 1 Introduction

Many models in the theoretical and empirical literature on macroeconomics, international trade and industrial organisation assume, at least implicitly, that aggregate demand is invariant to changes in the income distribution over individual consumers and, hence, that aggregate demand depends only on prices and aggregate income. The best known theoretical results on this topic are probably those of Gorman (1953, 1961), who made explicit the conditions on microeconomic consumer behaviour under which aggregate demand can be written as a function of prices and aggregate income alone.<sup>1</sup>

---

\*Center for Economic Studies, University of Leuven. E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven.be. Laurens Cherchye gratefully acknowledges financial support from the Research Fund KU Leuven through the grant STRT1/08/004.

†Institute for Fiscal Studies and University of Oxford, Department of Economics, Manor Road Building, Manor Road, Oxford OX1 3UQ, United Kingdom. E-mail: ian.crawford@economics.ox.ac.uk.

‡ECARES and ECORE, Université Libre de Bruxelles. Avenue F.D. Roosevelt 50, 1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges the European Research Council (ERC) for his Starting Grant.

§Department of Economics, University of Leuven, Naamsestraat 69, B-3000 Leuven, Belgium. E-mail: frederic.vermeulen@kuleuven.be. Frederic Vermeulen gratefully acknowledges financial support from the Research Fund KU Leuven through the grant STRT/12/001.

<sup>1</sup>Exact linear aggregation is to be distinguished from exact nonlinear aggregation, where aggregate demand is a function of some representative level of aggregate income, which itself can be a function of the distribution of income over the individuals. See, for example, Muellbauer (1975, 1976) and Jerison (1994) for more discussion.

Specifically, Gorman showed that exact linear aggregation is possible *if and only if* consumers have preferences of the Gorman Polar Form such that the corresponding linear Engel curves have common slopes. The empirical literature on consumer behaviour, however, has consistently tended to show that these conditions do not hold in practice. Lewbel and Pendakur (2009), for example, provide strong parametric evidence of nonlinear Engel curve behaviour whilst Blundell, Chen and Kristensen (2007) consider semi- and nonparametric evidence for this non-linearity.

In this paper, we revisit the problem that Gorman addressed. We too seek necessary and sufficient conditions for exact linear aggregation. However, we do this from a rather different perspective, that of the nonparametric revealed preference tradition of Samuelson (1938, 1948), Afriat (1967), Diewert (1973) and Varian (1982). Instead of describing the restrictions on behaviour in terms of the derivatives of certain functions (the slopes of Engel curves, for example), this approach works by characterising them in terms of a finite system of inequalities involving the consumer's observed choices only.

Finding a nonparametric equivalent to Gorman's aggregation theorems is, of course, of a certain amount of theoretical interest, but this is not our only motivation: we are also interested in empirical implementation. In particular we are interested in whether it may be possible empirically to analyse microdata for its aggregation properties without resorting to regression analysis. Regression analysis, in the words of Daniel McFadden in his presidential address to the Econometric Society, "interposes an untidy veil between econometric analysis and the propositions of economic theory". In fact McFadden was discussing parametric regression but nonparametric regression is not immune from the same observation. If the implications of economic theory are described in terms of the shapes of functions implied by the theory (e.g. Engel curves) then any empirical investigation of the theory requires those functions to be estimated from data. As a result, the conclusions from such an exercise necessarily rest jointly on the validity of the hypothesis at stake plus a number of crucial auxiliary statistical assumptions necessary to deliver consistent estimates of the functions of interest. This is the case whether the estimates are parametric, semi-parametric or nonparametric. For example both Lewbel and Pendakur (2009) and Blundell, Chen and Kristensen (2007) provide evidence based on pooled cross-section data - they therefore need to make a number of carefully chosen auxiliary assumptions about the form of unobserved heterogeneity and how it enters the model in order to deliver their estimates. Both studies also need to follow an instrumental variables strategy which also brings with it a set of important identifying assumptions. Revealed preference methods do not require the identification or estimation of structural functions. Instead they involve only inequality restrictions on the observables alone. As a result they are, to a great extent, free of the need for auxiliary hypotheses.<sup>2</sup> They therefore allow researchers to focus with much greater clarity on the hypothesis at the core. Furthermore, they are applicable when there are only very few observations and, hence, when statistical methods would be infeasible or uninformative.

The main contribution of this paper is twofold. Firstly, we establish the nonparametric counterparts of Gorman's aggregation conditions. We start by providing a revealed preference characterisation of Gorman Polar Form preferences for an individual consumer. We then propose an easy-to-implement necessary and sufficient test for Gorman's conditions for exact linear aggregation. Secondly, we demonstrate the practical usefulness of our results through an empirical investigation using a balanced microdata panel of Spanish households. Our first main conclusion here will be that we strongly reject exact linear aggregation when focusing on the set of all rational households in our

---

<sup>2</sup>Measurement error is an important exception to this but it can be handled in a fairly straightforward manner based on Varian (1985).

sample. Our second main result is that it is primarily due to heterogeneity in the marginal utility of income. Interestingly, this heterogeneity appears to be unrelated to the standard set of controls typically used as conditioning variables in microeconomic studies of household behaviour.

The remainder of this paper is structured as follows. Section 2 contains our main theoretical results, which contain a revealed preference characterisation of individual Gorman Polar Form preferences and an easy-to-implement necessary and sufficient nonparametric test for exact linear aggregation. Section 3 presents our empirical application. Section 4 concludes.

## 2 Exact linear aggregation: a nonparametric characterisation

In this section, we start by briefly reviewing the revealed preference conditions for rational consumption behaviour of individual consumers in terms of the well-known Afriat inequalities. Then, we investigate the conditions needed to guarantee exact linear aggregation, i.e. aggregate demand only depends on aggregate income and is not affected by how the income is actually distributed across consumers. From the functional derivative-based literature, we know that this independence result applies *if and only if* consumers have preferences of the Gorman Polar Form and linear Engel curves with common slopes.

We will proceed in three steps to define our revealed preference counterparts of these Gorman-type aggregation conditions. Firstly, we derive a revealed preference characterisation of individual preferences of the Gorman Polar Form. Secondly, we show the remarkable and important result that if observed prices are nonproportional, then having data that satisfy the Afriat inequalities is equivalent to the data satisfying the Gorman Polar Form conditions. In practice, this data requirement is very weak: it is met for the application that we present in the next section and, indeed, we are not aware of observational (non-experimentally generated) data on consumer behaviour which exhibits price-proportionality. Finally, we present the revealed preference versions of Gorman's aggregation conditions stated above. This will define an easy-to-apply linear test for exact linear aggregation.

### 2.1 Gorman Polar Form preferences

Suppose that we have a balanced microdata panel of consumers indexed by  $h = 1, \dots, H$  observed over a number of periods indexed  $t = 1, \dots, T$ . Following Gorman (1953), we make the classical assumption that the law of one price holds and that prices are strictly positive  $K$ -vectors ( $\mathbf{p}_t \in \mathbb{R}_{++}^K$ ). For each consumer  $h$  we observe non-negative quantities  $\mathbf{q}_t^h \in \mathbb{R}_+^K$ . We will denote these microdata by  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}^{h \in \eta}$ , with  $\eta = \{1, \dots, H\}$  and  $\tau = \{1, \dots, T\}$  being the index sets for consumers and periods, respectively. We will use  $\mathbf{Q}_t = \sum_{h \in \eta} \mathbf{q}_t^h$  to denote the aggregate demand vector in period  $t$ , so that the macrodata are  $\{\mathbf{p}_t, \mathbf{Q}_t\}_{t \in \tau}$ . Aggregate income is denoted by  $Y_t$  and is equal to  $\mathbf{p}_t' \sum_{h \in \eta} \mathbf{q}_t^h = \mathbf{p}_t' \mathbf{Q}_t$ .

**Individual rationality.** We will assume that all the consumers are rational in the sense that observed demand results from the maximisation of a well-behaved utility function subject to an individual budget constraint. Throughout, we will assume that utility functions are well-behaved (i.e. nonsatiated, monotonically increasing, concave and continuous). We can formally define

individual rationality as follows:

**Definition 1 (Individual rationalisation)** *A well-behaved utility function  $u^h$  provides an individual rationalisation of the data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  if for each observation  $t \in \tau$  we have  $u^h(\mathbf{q}_t^h) \geq u^h(\mathbf{q})$  for all  $\mathbf{q}$  with  $\mathbf{p}'_t \mathbf{q} \leq \mathbf{p}'_t \mathbf{q}_t^h$ .*

For our following discussion it is useful to be more specific about the empirical content of individual rationalisation. A core result in the revealed preference approach to demand is that there exists a utility function that provides an individual rationalisation of the data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  if and only if the data satisfy the well-known Afriat inequalities. This is formally captured by Afriat's Theorem (Varian, 1982; based on Afriat, 1967):

**Theorem 1 (Afriat's Theorem)** *The following statements are equivalent:*

(1.A). *There exists an individual rationalisation of the data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$ .*

(1.B). *For all  $s, t \in \tau$ , there exist numbers  $u_s^h, u_t^h \in \mathbb{R}_+$  and  $\beta_t^h \in \mathbb{R}_{++}$  that meet the Afriat inequalities*

$$u_s^h \leq u_t^h + \beta_t^h \mathbf{p}'_t (\mathbf{q}_s^h - \mathbf{q}_t^h).$$

The theorem thus implies that any data set  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  that can be rationalised by a well-behaved utility function needs to satisfy the Afriat inequalities. These Afriat inequalities are linear inequalities that are expressed in the unknowns  $u_t^h$  and  $\beta_t^h$  and that can easily be verified.<sup>3</sup> They also allow us to obtain an explicit construction of the utility levels and the marginal utility of income associated with each observation  $t$ : they define a utility level  $u_t^h$  and a marginal utility of income  $\beta_t^h$  (associated with the observed income  $\mathbf{p}'_t \mathbf{q}_t^h$ ) for each observed  $\mathbf{q}_t^h$ . Importantly, as has been demonstrated by Varian (1982), and later by Blundell, Browning and Crawford (2003, 2008), the above insights can be used to formally evaluate policy reforms in terms of individual welfare.

**Gorman Polar Form preferences.** We next define what it means for the data of an individual consumer to be rationalisable with the Gorman Polar Form. The Gorman Polar Form is usually defined in terms of an indirect utility function  $w^h$ . Let  $y^h$  represent the income of consumer  $h$ . The indirect utility function  $w^h$  is connected with the direct utility function  $u^h$  in the following way:

$$w^h(\mathbf{p}, y^h) = \max_{\mathbf{q}^h} \{u^h(\mathbf{q}^h) \mid \mathbf{p}' \mathbf{q}^h \leq y^h\}.$$

We can now state the next definition.

**Definition 2 (Gorman Polar Form rationalisation)** *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  are rationalisable by the Gorman Polar Form if there exists a utility function  $u^h$  that provides an individual rationalisation of the data and if there exists an associated indirect utility function  $w^h(\mathbf{p}, y^h) = \frac{y^h - a^h(\mathbf{p})}{b^h(\mathbf{p})}$ , with  $a^h(\mathbf{p}) \in \mathbb{R}$  and  $b^h(\mathbf{p}) \in \mathbb{R}_{++}$  for all  $\mathbf{p}$  and the functions  $a^h$  and  $b^h$  linearly homogeneous of degree 1.*

In this definition, the price index  $a^h(\mathbf{p})$  is often interpreted as subsistence expenditure - although this interpretation is not always valid (see Pollak, 1971, p. 403, fn. 4) - while the price index  $b^h(\mathbf{p})$  is interpreted as the inverse of the marginal utility of income.

<sup>3</sup>Varian (1982) showed that any data set  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  that can be rationalised by a well-behaved utility function will also satisfy the *Generalized Axiom of Revealed Preference* (GARP). We opted to keep our statement of Afriat's Theorem in line with our following discussion, which will focus on Afriat inequalities.

Before moving on it is worth pointing out the well-established fact that the Gorman Polar Form does not necessarily give rise to well-behaved preferences in all parts of the quantity-space: in general, well-behaved preferences only apply to a limited range of possible income values; and therefore Gorman Polar Form preferences are defined in terms of boundary conditions for the possible income levels.<sup>4</sup> To keep the exposition simple, our following analysis only considers income values that lie within this range, and, thus, we will not explicitly consider income boundary values (but, importantly, boundary conditions do appear in the proof of Theorem 2; see Appendix 1). We can then state the characterisation.<sup>5</sup>

**Theorem 2** *The following statements are equivalent:*

(2.A). *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  are rationalisable by the Gorman Polar Form.*

(2.B). *For all  $s, t \in \tau$ , there exist numbers  $w_s^h, w_t^h \in \mathbb{R}_+$ ,  $a_t^h \in \mathbb{R}$  and  $b_t^h \in \mathbb{R}_{++}$  such that*

$$w_s^h \leq w_t^h + \frac{1}{b_t^h} \mathbf{p}'_t (\mathbf{q}_s^h - \mathbf{q}_t^h), \quad (2.B.1)$$

$$w_t^h = \frac{(\mathbf{p}'_t \mathbf{q}_t^h) - a_t^h}{b_t^h}, \quad (2.B.2)$$

$$a_t^h = \delta a_s^h \text{ and } b_t^h = \delta b_s^h \text{ if } \mathbf{p}_t = \delta \mathbf{p}_s \text{ for } \delta \in \mathbb{R}_{++}. \quad (2.B.3)$$

Similar to Theorem 1, the numbers in this result have certain structural interpretations. Condition (2.B.1), for example, is an Afriat inequality which allows us to construct a utility level ( $w_t^h$ ) and marginal utility of income ( $1/b_t^h$ ) for each observation  $t$ . We can interpret every  $w_t^h$  as an indirect utility value (the function value  $w^h(\mathbf{p}, y^h)$  in Definition 2, which equals the utility value  $u^h(\mathbf{q}^h)$  under rational consumer behaviour). Condition (2.B.2) then states the Gorman Polar Form restriction, with the numbers  $a_t^h$  and  $b_t^h$  corresponding to the price indices  $a^h(\mathbf{p})$  and  $b^h(\mathbf{p})$  in Definition 2 evaluated at  $\mathbf{p}_t$ . Condition (2.B.3), finally, imposes linear homogeneity of these price indices.

Two further notes are in order. Firstly, the Gorman Polar Form characterisation in Theorem 2 is nonlinear in  $a_t^h$  and  $b_t^h$ . However, in our proof of Theorem 2 we show that it can be equivalently expressed in linear form. This makes it computationally very convenient. Secondly, in the absence of proportional price movements, Gorman Polar Form preferences provide *no additional restrictions* over and above the standard Afriat inequalities stated in Theorem 1.<sup>6</sup> In other words, Gorman Polar Form preferences and rational preferences are nonparametrically (in the revealed preference sense) *equivalent*: for data in which proportional prices movements are not observed their empirical implications are identical. This result is formally stated as follows:

<sup>4</sup>See, for example, Pollak (1971) and Blackorby, Boyce and Russell (1978) for a more detailed discussion on the local nature of Gorman Polar Form preferences. As we explain in Appendix 1, similar boundary conditions are needed to prove Theorem 2.

<sup>5</sup>An alternative revealed preference characterisation of the Gorman Polar Form can be found in work in progress by Brown and Shannon. In a certain sense, the work of these authors is complementary to ours as Brown and Shannon characterise Gorman Polar Form preferences in terms of so-called ‘dual’ Afriat numbers (which have an interpretation in terms of indirect utility functions; see Brown and Shannon, 2000), whereas our analysis starts from the original ‘primal’ Afriat numbers (to be interpreted in terms of direct utility functions). We thank Don Brown for revealing this to us in a private conversation.

<sup>6</sup>Specifically, under nonproportional prices condition (2.B.3) becomes redundant. Then, one can easily verify that, for any given solution for the Afriat inequalities (2.B.1), there also exists a solution for condition (2.B.2).

**Corollary 1** *The following statements are equivalent when prices  $\mathbf{p}_t \neq \delta \mathbf{p}_s$  ( $\delta \in \mathbb{R}_{++}$ ) for all  $s, t \in \tau$ :*

- (A). *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  are rationalisable by the Gorman Polar Form.*
- (B). *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$  satisfy the standard Afriat inequalities in Theorem 1.*

This is an important result. It implies that if the data can be rationalised by a well-behaved utility function and observed prices are nonproportional, then we can *always* construct an indirect utility function which exactly rationalises the data with the Gorman Polar Form. This is perhaps surprising as the Gorman Polar Form is usually thought of as a very demanding restriction. However, empirically it seems that this is *only* the case when proportional prices are observed in the data. In such a case, the Gorman Polar Form is *extremely* demanding as we can directly observe points on an Engel curve and this Engel curve must be perfectly straight. However, we are not aware of *any* observational (non-experimentally generated) consumer panel data in which proportional prices changes are ever observed. Thus, it turns out that, empirically, the Gorman Polar Form is without additional content from a revealed preference point of view.<sup>7</sup> This implies that, not only is it the case that if the data is rationalisable by a non-trivial utility function it can also be rationalised by a well-behaved utility function (as demonstrated in Afriat's Theorem), it is also the case that it can be rationalised by a utility function in the Gorman Polar Form class.

## 2.2 Exact linear aggregation

We can now use these insights to provide the revealed preference counterpart of Gorman's conditions for exact linear aggregation. As stressed above, exact linear aggregation implies that aggregate demand only depends on prices and aggregate income and is thus independent of the income distribution. Formally, this implies that aggregate demand can be written in the simple form  $\mathbf{Q} = \mathbf{g}(\mathbf{p}, Y)$ , where  $\mathbf{g}(\mathbf{p}, Y)$  is a vector-valued demand equation that only depends on aggregate income  $Y = \sum_{h \in \eta} y^h$  and prices  $\mathbf{p}$ . In other words, *any* income distribution of a given  $Y$  gives rise to the same aggregate demands  $\mathbf{Q}$ .

Gorman proved that such exact linear aggregation holds if and only if consumers' preferences are of the Gorman Polar Form with common slopes for the linear Engel curves. This is translated formally in the following definition:

**Definition 3 (Exact linear aggregation)** *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}^{h \in \eta}$  satisfy the conditions for exact linear aggregation if, for each  $h \in \eta$ , the data are rationalisable by the Gorman Polar Form and, moreover, the associated indirect utility functions are given by  $w^h(\mathbf{p}, y^h) = \frac{y^h - a^h(\mathbf{p})}{b(\mathbf{p})}$ , i.e. we have  $b^h(\mathbf{p}) = b(\mathbf{p})$ .*

In terms of Definition 2, exact linear aggregation requires a common  $b(\mathbf{p})$  index for all consumers (i.e.  $b^h(\mathbf{p}) = b(\mathbf{p})$  for all  $h$ ). The idea is that the marginal utility of income must be independent of income variations across consumers but can vary with prices. Using Theorem 2, we get the following characterisation of exact linear aggregation in revealed preference terms.

---

<sup>7</sup>At this point it is worth recalling that we focus on preferences taking the Gorman Polar Form for income values within bounded ranges, which here means that the equivalence in Corollary 1 has a local nature by construction.

**Theorem 3** *The following statements are equivalent for the data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}^{h \in \eta}$ :*

(3.A). *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}^{h \in \eta}$  satisfy the conditions for exact linear aggregation.*

(3.B). *For all  $s, t \in \tau$  and  $h \in \eta$ , there exist numbers  $w_s^h, w_t^h \in \mathbb{R}_+$ ,  $a_t^h \in \mathbb{R}$  and  $b_t \in \mathbb{R}_{++}$  such that*

$$w_s^h \leq w_t^h + \frac{1}{b_t} \mathbf{p}'_t (\mathbf{q}_s^h - \mathbf{q}_t^h), \quad (3.B.1)$$

$$w_t^h = \frac{(\mathbf{p}'_t \mathbf{q}_t^h) - a_t^h}{b_t}, \quad (3.B.2)$$

$$a_t^h = \delta a_s^h \text{ and } b_t = \delta b_s \text{ if } \mathbf{p}_t = \delta \mathbf{p}_s \text{ for } \delta \in \mathbb{R}_{++}. \quad (3.B.3)$$

As compared to Theorem 2, the key requirement is that the Afriat number  $b_t$  is common across consumers who face the same prices (i.e.  $b_t^h = b_t$  for all  $h$ ). Referring to Definition 2, this effectively imposes Gorman Polar Form preferences with a common  $b(\mathbf{p})$  index for all consumers. We note, finally, that our characterisation in Theorem 3 can be linearised in a directly similar way as our earlier characterisation in Theorem 2. As such, it implies an easy-to-apply linear test for exact linear aggregation.

Interestingly, the characterisation in Theorem 3 also generalises several special cases that generate the same independence of the income distribution. Two important examples are Varian's (1983) revealed preference characterisation of identical homothetic preferences (where  $a^h(\mathbf{p}) = 0$  in Definition 2) and Brown and Calsamiglia's (2007) revealed preference characterisation of quasi-linear preferences (where  $a^h(\mathbf{p}) = -p^i \phi(\mathbf{p})$  and  $b^h(\mathbf{p}) = p^i$ , with  $p^i$  the price of the numeraire and  $\phi$  a function that is homogeneous of degree one).

As a final result, we connect the characterisation in Theorem 3 with Corollary 1 we can see that if observed prices are nonproportional, then a necessary and sufficient condition for exact linear aggregation is that each consumer satisfies the standard Afriat inequalities with a common marginal utility of income. This is formally stated in the following result:

**Corollary 2** *The following statements are equivalent when prices  $\mathbf{p}_t \neq \delta \mathbf{p}_s$  ( $\delta \in \mathbb{R}_{++}$ ) for all  $s, t \in \tau$ :*

(A). *The data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}^{h \in \eta}$  satisfy the conditions for exact linear aggregation.*

(B). *For all  $s, t \in \tau$  and  $h \in \eta$ , there exist numbers  $w_s^h, w_t^h \in \mathbb{R}_+$  and  $b_t \in \mathbb{R}_{++}$  such that*

$$w_s^h \leq w_t^h + \frac{1}{b_t} \mathbf{p}'_t (\mathbf{q}_s^h - \mathbf{q}_t^h).$$

### 3 An application

In the previous section we established the revealed preference conditions for exact linear aggregation. Importantly, our characterisation can be linearised in unknowns, which makes it easily testable. We will next illustrate our revealed preference based aggregation results by means of an empirical application. Here, it is worth to recall from our discussion in the Introduction that revealed preference methods are intrinsically "nonparametric": in contrast to the more standard functional-derivative based methods, they do not need auxiliary parametric or statistical assumptions. As such, this empirical revealed preference analysis should thus lead to robust conclusions.

### 3.1 The data

The data we use are drawn from the Spanish Continuous Family Expenditure Survey (ECPF). This is one of the few surveys with detailed expenditure information for a panel of households. The ECPF is a quarterly budget survey of Spanish households which interviews about 3200 households every quarter. We focus on a subsample of couples (with or without children), in which the husband is in full-time employment in a non-agricultural activity while the wife is out of the labour force. This choice is driven by the fact that we want to minimise the impact of possible nonseparabilities between consumption and leisure.<sup>8</sup> Given the construction of the ECPF, households can be interviewed for up to eight consecutive quarters. However, our sample would be rather small if we would focus on those households observed for a full eight periods. Therefore, we have drawn a balanced panel of 342 households which are observed 5 times in order to balance the desire for a reasonable number of observations both across households and time. In what follows, we focus on a set of 15 nondurable commodity groups.<sup>9</sup> We note that the movements in the observed price vectors are nonproportional.

Our following analysis proceeds in two steps. First we check, individual household by individual household, whether observed behaviour is rationalisable by the Gorman Polar Form, albeit with heterogeneous  $b^h(\mathbf{p})$  indices for the different households. In the second step we then pool the data across households to investigate the conditions for exact linear aggregation.

### 3.2 Rationalisability by the Gorman Polar Form

For every household in our sample we test whether their behaviour is rationalisable by preferences of the Gorman Polar Form. There are two important points to note about this procedure. Firstly, since there are no proportional price movements observed in the data this is (by Corollary 1) equivalent to testing whether the data satisfy the standard Afriat inequalities. Secondly, we deal with each household individually and so we allow for complete preference heterogeneity within the Gorman class: households may differ with respect to whether they are rationalisable (by this class) and the precise form of their preference within this class. The results are presented in Table 1.

TABLE 1: Tests of Gorman Polar Form rationalisation

	Pass	Fail
N	326	16
Proportion	0.953 (0.011)	0.047 (0.053)

It turns out that the behaviour of 95% of the households in our data is exactly rationalisable by preferences of the Gorman Polar Form. A little under 5% of the data (16 households) are not rationalisable by well-behaved preferences at all. The standard errors on these proportions are

<sup>8</sup>Thus, our empirical analysis uses households/families as consumers. As Samuelson (1947, p.224) pointed out, “Attention should also be called to the fact that even the classical economist does not literally have the individual in mind, so much as the family; of course, some hardy souls will pursue the will-o'-the-wisp of sovereignty within the family so as to reduce even these collective indifference curves to an individualistic basis”. See Cherchye, De Rock and Vermeulen (2007, 2011) for a revealed preference characterisation of Chiappori’s (1988) model of collective household consumption that starts from an individualistic basis.

<sup>9</sup>Household preferences over these commodity groups are assumed to be stable over time and separable from durable goods. The commodity groups are (i) food and non-alcoholic drinks at home; (ii) alcohol; (iii) tobacco; (iv) energy at home (heating by electricity); (v) services at home (heating: not electricity, water, furniture repair); (vi) nondurables at home (cleaning products); (vii) non-durable medicines; (viii) medical services; (ix) transportation; (x) petrol; (xi) leisure (cinema, theatre, clubs for sports); (xii) personal services; (xiii) personal nondurables (toothpaste, soap); (xiv) restaurants and bars and (xv) travelling (holiday).

reported in parentheses - we see that the small number of failures we observe in this sample is subject to sampling variation: it might easily be the case that another similarly sized random sample of households from this population would contain no failures at all. Since it is a necessary condition for exact linear aggregation that individual households act as if they are utility maximisers, we therefore do not include these 16 households in the further analysis.<sup>10</sup>

### 3.3 Testing for exact linear aggregation

We now turn to the main question - given microdata with rational agents does the observed behaviour support exact linear aggregation? In other words, is their aggregate demand independent of the income distribution? To do this we check the condition that is given in Theorem 3. At this point the test requires pooling across households so that we can investigate the commonality of the marginal utility of incomes within the sample. Specifically, we need to check for the data pooled across households whether there exist numbers  $w_t^h \in \mathbb{R}_+$  for  $h = 1, \dots, 326$  and  $t = 1, \dots, 5$  and  $b_t \in \mathbb{R}_{++}$  for  $t = 1, \dots, 5$  such that,

$$w_s^h \leq w_t^h + \frac{1}{b_t} \mathbf{p}'_t (\mathbf{q}_s^h - \mathbf{q}_t^h),$$

for all observations  $s, t \in \tau$  and all households. To do this we use phase one of the simplex method linear programming algorithm, which efficiently determines whether or not a system of linear inequalities has a basic feasible solution. We find that this condition is rejected. Despite the fact that these households all satisfy the necessary condition (Gorman Polar Form preferences) perfectly, and despite the very flexible nature of revealed preference tests, it seems that the additional restriction required for aggregation (that within a period all households have the same  $b_t$  parameter) is too much: data cannot bear the weight of the theory required for exact linear aggregation.

### 3.4 Preference heterogeneity

Given that the behaviour of our remaining households is precisely consistent with the idea that they have preferences of the Gorman Polar Form albeit with different preferences within that class, we now turn to the issue of heterogeneity. The focus is on the heterogeneity which is relevant for aggregation, namely heterogeneity with respect to the marginal utility of income.

**Partitioning on observables.** To the extent that heterogeneity in the  $b_t$  parameter might be driven by observables, stratification is a flexible and fully nonparametric way in which to allow for this - the idea being that exact linear aggregation might be valid when applied to sub-groups of demographically similar households, even though when applied to the data *in toto* it is rejected. To investigate this further we allocated the data to relatively homogeneous groups on the basis of observables such as their age profiles, schooling level, household size and number of children. This resulted in 52 groups of which 34 groups contain more than one household (see Appendix 2 for the frequency distribution of the group sizes). We then test the conditions for exact linear aggregation once more but this time *within* each of these groups. Although the number of different households within a group can be as small as two, as long as there is more than one household the conditions for exact linear aggregation always fail. This despite the fact that the strength of revealed preference

<sup>10</sup>An alternative would be to impose rationalisability on them. Blundell, Browning and Crawford (2008) describe a way to do this. In this case, and in view of the very small number of such households, we opted for simplicity.

tests in general (weakly) increases with the number of observations, so that reducing the number of households involved in a test, by considering only those with similar observables, should make it easier to rationalise behaviour.

**Partitioning on unobservables.** We also consider a second partitioning exercise. As before, the idea is to partition our set of 326 rational households into a number of subsets, such that each subset contains households which, together, are exactly linearly aggregable. Instead of partitioning on observables we use an algorithm which searches for the most parsimonious grouping, i.e. one that minimises the number of groups required to exclusively and exhaustively partition the data into groups within which behaviour is aggregable. Of course, the only way to determine this precisely is to form the set of all subsets of the data (and there are  $2^{326}$  of these) and to check all of them. Since this is computationally infeasible we adapt an efficient algorithm developed by Crawford and Pendakur (2013), and calculate bounds on the minimal number of aggregable groups.

We find that the fewest number of groups we need to rationalise the data is somewhere between 90 and 103. Clearly  $52 \notin [90, 103]$  and this is why our attempt to rationalise the data on the previous demographics-based partition failed. On average, there are fewer than 4 households per aggregable subset, and the largest group we were able to construct consisted of 19 households. Essentially, this outcome says that we need a substantial number of groups (including many singletons) to rationalise our data in terms of exact linear aggregation.

**Interpretation.** Generally, our results appear to indicate that heterogeneity in the marginal utility of income is both essentially idiosyncratic and economically meaningful in the sense that it is sufficient, firstly, to be easily detectable with a nonparametric/revealed preference test and, secondly, to prevent exact linear aggregation from going through.

It is interesting to interpret this finding in the light of recent developments in the macroeconomics literature. The advent of more powerful computers and improved numerical methods did not only give birth to the field of microeconometrics, but it also allowed macroeconomists to shift attention towards rich heterogeneous agents models. Nowadays, one of the main workhorse models for studying heterogeneity in macroeconomics is what Heathcote, Storesletten and Violante (2009) call the “standard incomplete markets” (SIM) model (see also Ríos-Rull, 1995, Ljungqvist and Sargent, 2004, and Krusell and Smith, 2006). The SIM model is characterised by a continuum of individuals (households), who have different preferences and who differ with respect to characteristics like productivity or health status. These individuals then are faced with independent uninsurable shocks in their endowments, which lead to behavioural changes at the micro level and, ultimately, also at the macro level. We may take our results as providing empirical support for using such macroeconomic models with a continuum of households (like the SIM model).

Related to this, our results can also be interpreted as revealed preference evidence against the existence of a representative agent.<sup>11</sup> If the conditions for linear exact aggregation are satisfied, then there exists a representative agent for which the aggregate demand can be modelled as the outcome of rational, maximising behaviour given prices and aggregate income. Importantly, this representative agent also allows for normative conclusions: the agent’s preferences can properly be

---

<sup>11</sup>Notwithstanding the shift towards heterogeneous agents models, it is fair to say the representative consumer still plays an important role in a number of modern macroeconomic models and in macroeconomic textbooks (see, for example, Clarida, Gali and Gertler, 1999, Woodford, 2003, Uhlig, 2010, and Gourio, 2012).

represented by an aggregate social welfare function.<sup>12</sup> As such, our empirical results also complement the overwhelming evidence against the existence of such a representative agent (see, for example, Kirman, 1992, and Carroll, 2000). More specifically, they add that even the existence of a limited set of representative agents seems to be a very unrealistic hypothesis. As indicated above, because our test is intrinsically nonparametric, it provides robust evidence in support of this conclusion.

## 4 Conclusion

Many economic models use the assumption that aggregate demand depends solely on prices and aggregate income and, thus, abstract from income-distributional effects. Although the conditions for the existence of such a distribution-independent aggregation have been argued to be demanding, it is fair to say that existing evidence is solely based on Gorman's well-known exact linear aggregation results within a functional-derivative based framework. To test Gorman's conditions for exact linear aggregation (which boil down to consumers having preferences of the Gorman Polar Form with an equal marginal utility of income), one needs to make many additional assumptions to bring these conditions to the data.

In this paper, we revisited the problem of exact linear aggregation by bringing in tools from the revealed preference literature. These tools are based solely on the data at hand and do not need any additional parametric or statistical assumptions. As such, they allow for robustly analysing the empirical validity of exact linear aggregation. In addition to a few interesting and rather important side results (like a revealed preference characterisation of Gorman Polar Form preferences for an individual consumer), we proposed a revealed preference test for exact linear aggregation. Interestingly, the test is linear and thus easy-to-apply in practice.

We demonstrated the practical usefulness of our revealed preference characterisation by means of an empirical application to a Spanish balanced microdata panel. Our main conclusion is that we could not find any evidence suggesting the existence of a limited set of household types for which aggregate demand can be modelled as independent of the income distribution. That is, our conditions for exact linear aggregation are not satisfied for our sample, and the same result holds even when considering small groups of households defined in terms of observable characteristics. Moreover, even an algorithmic approach designed to group the data as efficiently as possible into aggregable groups failed to find a parsimonious grouping. We may interpret these results as providing empirical support for (e.g. SIM) macroeconomic models that are based on a continuum of consumers rather than a limited set of representative agents.

---

<sup>12</sup>See, for example, Dow and Werlang (1988), Kirman (1992) and Jerison (1994) for more discussion on representative agents and how this relates to exact linear aggregation.

# Appendix 1

## Proof of Theorem 2.

As a preliminary step, we provide an equivalent linear formulation of the conditions in (2.B). Let  $\alpha_t^h = -a_t^h/b_t^h$  and  $\beta_t^h = 1/b_t^h$ . Then we get the following linear reformulations of the conditions (2.B.1) – (2.B.3):

$$w_s^h \leq w_t^h + \beta_t^h \mathbf{p}'_t (\mathbf{q}_s^h - \mathbf{q}_t^h), \quad (2.B.1')$$

$$w_t^h = \alpha_t^h + \beta_t^h (\mathbf{p}'_t \mathbf{q}_t^h), \quad (2.B.2')$$

$$\alpha_t^h = \alpha_s^h \text{ and } \beta_t^h = \beta_s^h / \delta \text{ if } \mathbf{p}_t = \delta \mathbf{p}_s \text{ for } \delta \in \mathbb{R}_{++}. \quad (2.B.3')$$

(2.A)  $\Rightarrow$  (2.B) : Condition (2.B.1') readily follows Theorem 1 for a utility function  $u^h$  that rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$ . Then, we can use  $w_t^h = \max_{\mathbf{q}} \{u^h(\mathbf{q}) \mid \mathbf{p}'_t \mathbf{q} \leq \mathbf{p}'_t \mathbf{q}_t^h\}$  (using  $\mathbf{p}'_t \mathbf{q}_t^h = y_t^h$ ). Given this, Definition 2 directly implies (2.B.2') and (2.B.3') when using  $\alpha_t = -a^h(\mathbf{p}_t)/b^h(\mathbf{p}_t)$  and  $\beta_t^h = 1/b^h(\mathbf{p}_t)$ .

(2.B)  $\Rightarrow$  (2.A) : Consider

$$u^h(\mathbf{q}) = \min_t [w_t^h + \beta_t^h \mathbf{p}'_t (\mathbf{q} - \mathbf{q}_t^h)].$$

Varian (1982) has shown that this utility function rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t^h\}_{t \in \tau}$ . Using (2.B.2'), we have

$$u^h(\mathbf{q}) = \min_t [\alpha_t^h + \beta_t^h \mathbf{p}'_t \mathbf{q}]. \quad (1)$$

Let us then verify whether the function  $u^h$  meets Definition 2. Consider some arbitrary prices  $\mathbf{p}_0$  and income  $y_0^h$ . As a preliminary step, we recall that

$$w^h(\mathbf{p}_0, y_0^h) = \max_{\mathbf{q}} \{u^h(\mathbf{q}) \mid \mathbf{p}'_0 \mathbf{q} \leq y_0^h\}.$$

Thus, using (1), we get

$$w^h(\mathbf{p}_0, y_0^h) = \max_{\mathbf{q}} \left\{ \min_t [\alpha_t^h + \beta_t^h \mathbf{p}'_t \mathbf{q}] \mid \mathbf{p}'_0 \mathbf{q} \leq y_0^h \right\}.$$

Dropping the min operator, we can equivalently state

$$w^h(\mathbf{p}_0, y_0^h) = \max_{w, \mathbf{q}} \{w \mid w \leq \alpha_t^h + \beta_t^h \mathbf{p}'_t \mathbf{q} \ (t \in \tau), \ \mathbf{p}'_0 \mathbf{q} \leq y_0^h\},$$

which obtains the linear program

$$\begin{aligned} w^h(\mathbf{p}_0, y_0^h) &= \max_{w \in \mathbb{R}, \mathbf{q} \in \mathbb{R}_+^N} w \\ &\text{s.t.} \\ w - \beta_t^h \mathbf{p}'_t \mathbf{q} &\leq \alpha_t^h \ (t \in \tau), \\ \mathbf{p}'_0 \mathbf{q} &\leq y_0^h. \end{aligned} \quad (2)$$

The dual linear program is given as

$$\begin{aligned}
w^h(\mathbf{p}_0, y_0^h) &= \min_{\theta_t \in \mathbb{R}_+, (t \in \tau), \lambda \in \mathbb{R}_+} \sum_{t=1}^T \alpha_t^h \theta_t + \lambda y_0^h & (3) \\
&\text{s.t.} \\
&\sum_{t=1}^T \theta_t = 1, \\
&-\sum_{t=1}^T \theta_t \beta_t^h \mathbf{p}_t + \lambda \mathbf{p}_0 \geq 0.
\end{aligned}$$

Let  $\theta_t^*$  ( $t \in \tau$ ) and  $\lambda^*$  define the optimum of program (3). In general, these optimal values are independent of  $y_0^h$  when  $y_0^h$  respects boundary conditions that limit the domain of  $y_0^h$ . In practice, the boundary values for  $y_0^h$  can be determined by standard methodology for sensitivity analysis of linear programming. (Technically, these bounds will correspond to the range of  $y_0^h$  (as the objective coefficient of  $\lambda$ ) for which the optimal basic feasible solution of the linear program (3) remains constant.) These boundary conditions parallel the usual conditions that apply to indirect utility functions representing Gorman Polar Form preferences; see our discussion following Definition 2 in the main text.

Thus, because the solution of the problem (3) is independent of  $y_0^h$  (under the stated boundary conditions), we conclude that the function  $w^h$  in (3) meets the requirement in Definition 2 for

$$\lambda^* = 1/b^h(\mathbf{p}_0) \text{ and } -a^h(\mathbf{p}_0)/b^h(\mathbf{p}_0) = \sum_t \theta_t^* \alpha_t^h.$$

Specifically, for  $w^*$  the optimal value of linear program (3) (or, equivalently, (2)), we get

$$\begin{aligned}
w^h(\mathbf{p}_0, y_0^h) &= w^* \\
&= \lambda^* y_0^h + \sum_{t=1}^T \theta_t^* \alpha_t^h \\
&= \frac{y_0^h - a^h(\mathbf{p}_0)}{b^h(\mathbf{p}_0)}.
\end{aligned}$$

Inspection of problems (2) and (3) reveals that the price indices  $a^h$  and  $b^h$  are linearly homogenous of degree 1 (if again the same income boundary conditions hold).■

### Proof of Corollary 1.

Conditions (2.B.2) and (2.B.3) in Theorem 2 are void if  $\mathbf{p}_t \neq \delta \mathbf{p}_s$  ( $\delta \in \mathbb{R}_{++}$ ) for all  $s, t \in \tau$ . As such, rationalisability by Gorman Polar Form only requires consistency with the condition (2.B.1).■

### Proof of Theorem 3.

This follows from Theorem 2 (i.e. each household is rationalisable by the Gorman Polar Form) and the result of Gorman (i.e. the marginal utility of income is household independent, which is captured by the common  $b_t$  (i.e.  $b_t^h = b_t$  for all  $h$ )). ■

### Proof of Corollary 2.

The result follows from combining Corollary 1 with Theorem 3.■

## Appendix 2

TABLE 2: Household sub-groups

Group size	Frequency
1	18
2	6
3	4
4	6
5	1
6	2
7	3
8	2
9	1
10	2
11	1
14	1
20	1
22	1
24	1
35	1
51	1

## References

- [1] Afriat, S. (1967), “The construction of utility functions from expenditure data”, *International Economic Review*, 8, 67-77.
- [2] Blackorby, C., R. Boyce and R. Russell (1978), “Estimation of demand systems generated by the Gorman polar form: a generalization of the S-branch utility tree”, *Econometrica*, 46, 345-363.
- [3] Blundell, R., M. Browning and I. Crawford (2003), “Nonparametric Engel curves and revealed preference”, *Econometrica*, 71, 205-240.
- [4] Blundell, R., M. Browning and I. Crawford (2008), “Best nonparametric bounds on demand responses”, *Econometrica*, 76, 1227-1262.
- [5] Blundell, R., X. Chen and D. Kristensen (2007), “Nonparametric IV estimation of shape-invariant Engel curves”, *Econometrica*, 75, 1613-1669.
- [6] Brown, D. and C. Calsamiglia (2007), “The nonparametric approach to applied welfare analysis”, *Economic Theory*, 31, 183-188.
- [7] Brown, D. and C. Shannon (2000), “Uniqueness, stability, and comparative statics in rationalizable Walrasian markets”, *Econometrica*, 68, 1529-1540.
- [8] Carroll, C., (2000), “Requiem for the representative consumer? Aggregate implications of microeconomic consumption behavior”, *American Economic Review*, 90(2), 10-115

- [9] Cherchye, L., B. De Rock and F. Vermeulen (2007), “The collective model of household consumption: a nonparametric characterization”, *Econometrica*, 75, 553-574.
- [10] Cherchye, L., B. De Rock, and F. Vermeulen (2011), “The revealed preference approach to collective consumption behavior: testing and sharing rule recovery”, *Review of Economic Studies*, 78, 176-198.
- [11] Chiappori, P.-A. (1988), “Rational household labor supply”, *Econometrica*, 56, 63-89.
- [12] Clarida, R., J. Gali and M. Gertler (1999), “The science of monetary policy: A New Keynesian perspective”, *Journal of Economic Literature*, 37, 1661-1707.
- [13] Crawford, I. and K. Pendakur (2013), “How many types are there?”, *Economic Journal*, 123, 77-95.
- [14] Diewert, E. (1973), “Afriat and revealed preference theory”, *Review of Economic Studies*, 40, 419-425.
- [15] Dow, J. and S. Werlang (1988), “The consistency of welfare judgments with a representative consumer”, *Journal of Economic Theory*, 44, 265-280.
- [16] Gorman, W. (1953), “Community preference fields”, *Econometrica*, 21, 63-80.
- [17] Gorman, W. (1961), “On a class of preference fields”, *Metroeconomica*, 13, 53-56.
- [18] Gourio, F. (2012), “Disaster risk and business cycles”, *American Economic Review*, 102, 2734-66.
- [19] Heathcote, J., K. Storesletten and G. Violante (2009), “Quantitative macroeconomics with heterogeneous households”, *Annual Review of Economics*, 1, 319-354.
- [20] Jerison, M. (1994), “Optimal income distribution rules and representative consumers”, *Review of Economic Studies*, 61, 739-771.
- [21] Kirman, A. (1992), “Whom or what does the representative individual represent?”, *Journal of Economic Perspectives*, 6, 117-36.
- [22] Krusell, P. and A. Smith (2006), “Quantitative macroeconomic models with heterogeneous agents” in R. Blundell, W. Newey, T. Persson, *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, Cambridge.
- [23] Lewbel, A. and K. Pendakur (2009), “Tricks with Hicks: The EASI demand system”, *American Economic Review*, 99, 827-863.
- [24] Ljungqvist, L. and T. Sargent (2004), *Recursive Macroeconomic Theory*, MIT Press, Cambridge.
- [25] Muellbauer, J. (1975), “Aggregation, income distribution and consumer demand”, *Review of Economic Studies*, 62, 525-543.
- [26] Muellbauer, J. (1976), “Community preferences and the representative consumer”, *Econometrica*, 44, 979-999.

- [27] Pollak, R. (1971), “Additive utility functions and linear Engel curves”, *Review of Economic Studies*, 38, 401-414.
- [28] Ríos-Rull, J.-V. (1995), “Models with heterogeneous agents” in T. Cooley (ed.), *Frontiers of Business Cycle Research*, Princeton University Press, Princeton.
- [29] Samuelson, P. (1938), “A note on the pure theory of consumer behaviour”, *Economica*, 5, 61-71.
- [30] Samuelson, P. (1947), *Foundations of Economic Analysis*, Harvard University Press, Cambridge MA.
- [31] Samuelson, P. (1948), “Consumption theory in terms of revealed preference”, *Economica*, 15, 243-253.
- [32] Uhlig, H. (2010), “Understanding the impact of fiscal policy. Some fiscal calculus”, *American Economic Review*, 100, 30-34.
- [33] Varian, H. (1982), “The nonparametric approach to demand analysis”, *Econometrica*, 50, 945-972.
- [34] Varian, H. (1983), “Nonparametric tests of consumer behaviour”, *Review of Economic Studies*, 50, 99-110.
- [35] Varian, H. (1985), “Nonparametric analysis of optimizing behavior with measurement error”, *Journal of Econometrics*, 30, 445-458.
- [36] Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton.