NONPARAMETRIC ANALYSIS OF LABOUR SUPPLY USING RANDOM FIELDS

IAN CRAWFORD

Nuffield College, University of Oxford

Abstract. This paper studies labour supply in panel data by means of random fields. In doing so it describes a way of uniting classical revealed preference techniques and econometric prediction by means of a best linear, unbiased prediction procedure based on Goldberger (1962) and known as Weiner-Kolmogorov prediction or Universal Kriging in the spatial statistics literature. This, it is argued, retains the best features of both revealed preference and statistical approaches. In an application to the consumption and labour supply decisions of NYC taxi drivers this paper makes a number of empirical points: first that behaviour which is, on the basis of conventional revealed preference-based measures, rational, can be shown to be highly economically implausible; secondly that modelling labour supply using parsimonious relevant conditioning can solve the puzzle by providing predictions which match the data, are theoretically-consistent yet are behaviourally and economically sensible; thirdly it shows that modelling behaviour at the level at the which the theory is designed to apply (which is to say at the level of the individual) can given greater insights into behaviour and heterogeneity than modelling population moments or quantiles; lastly than the practice of assuming monotonic scalar heterogeneity when modelling cross sectional data may give a strongly misleading impression of both behaviour and preference heterogeneity.

1. Introduction

Empirical microeconomics offers two broad approaches to modelling the demand behaviour of price-taking individuals. The first is the statistical approach. This focusses on the identification and estimation of demand functions. These can then be used to check for consistency with economic theory, to investigate counterfactuals and to carry out welfare analysis.[1] The second is classical revealed preference analysis.[2] This approach is essentially phenomenological – it aims fully to characterise the implications of economic theory in terms of the observable properties of finite collections of observed data points rather than in terms of shape restrictions on unobservable and, therefore, to-be-estimated demand curves. Like statistical demand analysis, revealed preference analysis can be used to test economic theory, to forecast behaviour and to carry out welfare analysis.

The statistical approach has significant strengths but also some weaknesses. Its strengths include the ability to make point predictions, to deal with errors of various kinds, to allow behaviour to depend on (non-economic) factors beyond income and prices, to accommodate non-theory-consistent observed behaviour, and to make statistical inferences. Its weakness include often providing a poor fit, and that the connection between the statistical model and the behaviour of individuals is not immediate. The lack of fit is to some degree a necessity built into the approach – parametric models sacrifice fit to deliver interpretability and nonparametric methods sacrifice fit to trade off bias/variance. Nonetheless, very poor fit can be unsettling when doing forecasting or welfare exercises. The disconnect between the statistical model and individual behaviour comes from the fact that when theory meets data, the focus shifts away from the behaviour of the individual to the behaviour of features of statistical distributions; we normally model sample analogues of population conditional expectations, for example, or conditional quantiles. The link between the behaviour of these statistics and the behaviour of an individual person is not immediate, but it can be established by making assumptions. Typically, these concern the properties of the joint distribution of observables and some unobservable auxiliary parameters (variously thought of as shocks, error terms and unobserved heterogeneity) which are appended to the model. But economic theory is silent on these auxiliary parameters and data can say nothing about unobservables. As a consequence, these assumptions are largely untestable, not manifestly plausible and represent an 'untidy veil' between economic theory and data.[3]

This work was substantially completed whilst the author was visiting Oxford Erskine Fellow at the University of Canterbury in Christchurch, New Zealand and I am very grateful to the School of Economics and Finance and to the Erskine Program for its hospitality.

[1]For example, Deaton and Muellbauer (1980), Blundell, Banks and Lewbel (1997), Hausman and Newey (1995, 2016), Lewbel and Pendakur (2009), Blundell, Horowitz and Parey (2017).

[2]Afriat (1967), Diewert (1973) Varian (1983).

[3]Dan McFadden in his presidential address to the Econometric Society in 1985 used this phrase to distinguish nonparametric from parametric methods. This paper argues that as long as the error term is present, the veil remains in place even with nonparametric statistical demand systems.

Conditional mean regressions, for example, will recover interpretable individual demands only when unobserved heterogeneity is additively separable in the regression model. But as Brown and Walker (1989) and Lewbel (2001) have shown, demand functions generated from random utility functions do not typically yield demand functions where the unobserved taste shifters are additive. Alternative approaches based on conditional quantile demands, whilst they represent a significant advance in terms of identification of individual demands under non-separable heterogeneity (see Matzkin (2003, 2008)), also require assumptions regarding the dimensionality of heterogeneity (that it is scalar) and monotonicity in order to recover individual demands.[4] The scalar heterogeneity assumption is restrictive and not testable empirically, and in the absence of scalar heterogeneity, quantile demands can no longer be interpreted as individual demands (see Blundell, Horowitz and Parey (2017) for a discusison of this).

Classical revealed preference analysis, by contrast, is designed to apply directly to individual-level data (for example, either panel data or experimentally-generated data). The results of this approach are therefore immediately economically interpretable and meaningful because the individual is the level at which the theory applies and at which it should therefore be tested, interpreted and used.[5] But the difficulties with revealed preference methods are threefold and significant. Firstly, since they are based on systems of inequalities they generally provide set-valued counterfactual demands and welfare measures. These are "sharp", in the sense that they exhaust the empirical content of the economic theory, but they may also be blunt, in the sense of not ruling much out. This means that they are often thought, with some justification, to be close-to-useless for serious practical application. Secondly, there is little-to-no scope within revealed preference theory for introducing variables beyond prices and incomes. Revealed preference analysis only allows us to explain behaviour in terms of income and substitution effects, yet other variables might, quite plausibly, be expected to have an impact on behaviour. Since consumer theory has little to say about variables which are neither choices nor constraints, there is no clear way to exploit any other variables which might nevertheless be important influences on behaviour. Thirdly classical revealed preference analysis is, at the level of the individual, deterministic. This means that it is hard to allow for non-theory-consistent behaviours, or to construct measures of forecast uncertainty or to make statistical inferences.

These two approaches have developed somewhat in parallel but a number of papers have combined them to some degree by applying the revealed preference machinery to statistical objects estimated using cross section data. Blundell, Browning and Crawford (2003, 2008), for example, test and impose revealed preference requirements on conditional expectations (nonparametrically estimated Engel curves). They assume that, on each budget constraint, the distribution of demand was generated from the demand of a single consumer with an additive error. A more general approach is taken in Blundell, Kristensen and Matzkin (2014, 2017) who assume that the distribution of demand is generated from a distribution of unobserved tastes and then estimate conditional quantiles of this distribution. Both are applied to cross section data in which there is limited price variation. Consequently they inherit from revealed preference methods the drawback that they only provide bounds on counterfactual behaviour and welfare. Also in a cross-sectional data context, Hoderlein and Stoye (2014) and Kitamura and Stoye (2018) employ a stochastic revealed preference approach[6] and therefore focus not on the behaviour of individuals but on behaviour of distributions of demands. Arguably, applying the revealed preference machinery to conditional mean or quantile functions conflates the problems inherent in both approaches whilst applying them to distributions loses the focus on individual behaviour.

What largely separates the statistical and the classical revealed preference approaches is the error term. The statistical approach has one, the classical revealed preference approach does not. Standard practice in the statistical demand literature is to assume that the consumer's demands are a convolution of a rational systematic component, and a random shock. The rational component is motivated by theory. The stochastic component is motivated as follows:

> "The ... interpretation is that the true utility used by consumers to make choices is deterministic, but due to the researcher's inability to formulate individual behavior precisely, an additional stochastic term is added, thus making utility stochastic from the researcher's point of view (see Manski 1977; McFadden 1981, 1984). This is the interpretation followed in the economics literature" - Nevo (*Annual Reviews of Economics*, 2011, p. 59)

This echoes almost identical views expressed nearly 70 years earlier by Haavelmo:

> "Observable economic variables do not satisfy exact relationships (except, perhaps, some trivial identities). Therefore, if we start out with such a theoretical scheme, we have - for the purpose

[4]The economic content of this monotonicity assumption that individuals' preferences satisfy single-crossing so that they cannot re-rank along the budget constraint as prices and incomes move; a quantile then identifies a unique individual (assuming a fixed population).

[5]Methodological individualism is, evidently, the first "meta-axiom" of neoclassical economics (Varoufakis (2012)).

[6]McFadden and Richter (1971) and McFadden (2005).

of application - to add some stochastical elements, to bridge the gap between the theory and the facts." - Haavelmo, (*Econometrica*, 1944.)

As both authors make clear, the random shock is not part of the model but an indication by the researcher that complete knowledge of the correct model is out of reach. In contrast, classical revealed preference theory assumes that demand data provides direct observations of points on an individual's demand curve and therefore does not generally append error terms.[7] It is this which creates a disconnect between the statistical literature on demand behaviour and the theory-driven revealed preference approach.

In this paper I describe and apply a statistical model of demand which interpolates observed choices. There is no error term, as such, and so the approach sits well with revealed preference analysis. Demand is predicted at some new, previously unobserved price-budget, or a sequence of such budgets, by minimising the variance of the prediction error subject to an unbiasedness constraint. The minimand is not the sum of squared residuals. Rather it is the variance of the prediction error. The object of interest is therefore a Best Linear Unbiased Predictor (BLUP) in the sense of Goldberger (1962). It will be shown below that the best linear unbiased predictor is a locally-weighted sum of the observed demands at the sample points. As well as being an exact interpolator when the prediction point is also a sample point, the variance of the predictor shrinks to zero as we consider a sequence of prices which converge to an observed sample point.

Because the model exactly matches the data at the observed prices there is no error term and the approach is a natural econometric counterpart to classical revealed preference methods. However it also enjoys a number of advantages compared to the revealed preference approach. Firstly it provides point, rather than set-valued, predictions of demand curves. Secondly it provides a measure of forecasting uncertainty regarding predictions at un-sampled budgets. Thirdly it can make use of variables beyond prices and incomes to describe behaviour and thus it enables a richer understanding of behaviour. Fourthly, because the method can be applied to a sequence of counterfactual price-budgets it fits perfectly with the approach to measuring welfare through finite approximations to differential equations developed by Vartia (1983). Thus it retains the elegance of revealed preference but increases their power and applicability by overcoming the principal drawbacks encountered in applications.

The plan of the paper is as follows. In section 2 I describe the statistical model and the predictor. In section 3 I provide two illustrative examples. In section 4 I apply the method to tthe consumption and labour supply decisions of a sample of NYC cab drivers. This sample of individuals provides an interesting empirical study for a number of reasons. Firstly it is a panel and therefore we observe individuals repeatedly. Secondly the canonical consumption-leisure model, being low-dimensional, makes it relatively straightforward to illustrate the results visually and show how the methods outlined are working (e.g. to show the entire labour supply curve, not just marginal/own-price responses or price-constant Engel curves). Thirdly these individuals are not necessarily rational; V. Crawford and Meng (2011) provide evidence of significant reference-dependence in their behaviour. As a result these data provide a challenging environment in which to try to apply revealed preference theory. Section 4 follows up a number of substantive topics which, I argue, only the combination of panel data and a statistical approach which allows individual-level economic analysis can satisfactorily explore. After describing the data in section 4.1, section 4.2 provides initial estimates of the labour supply functions for each individual driver. These strongly indicate that relatively small changes in real wages are apparently associated with locally very abrupt changes in hours of work and questions whether this is really plausible. Section 4.3 conducts a reveal preference analysis and shows that the behaviour of these drivers (both the observed and modelled behaviour) is not rational and considered whether imposing full rationality can provide a more economically and behaviourally plausible set of estimates (as argued by Blundell, Horowitz and Parey (2017)). It shows that this is not the case – the observed behaviour is, by conventional measures, close to rational but even behaviour which is constrained to be fully rational from a revealed preference point of view does not necessarily result in economic plausibility. Indeed it may look anything but rational in the more everyday sense of the word. Section 4.4 introduces conditional demand functions. This allows that the decision-making environment is richer than can be described by just the changing budget constraint and it considers how best to choose appropriate conditioning variables. It suggests that combining a conditional revealed preference-based approach with Selten's (1991) measure of predictive success provides a principled way of conducting a specification search which gives a role to economic theory. A number of possible conditioning variables are investigated empirically, including a number which have been suggested in the literature on reference-dependence (Camerer et al (1997), V. Crawford and Meng (2011)). It shows that Selten-optimal conditioning of demands results in individual-level labour supply functions which are *both* rational in the technical economic sense and also rational in the everyday sense of being reasonably sensible. Section 4.5 turns to heterogeneity and

[7]Varian (1985) considers the impact of classical measurement error. The ideas described in this paper also accommodate measurement errors of this kind. However, while the data used here, and in many modern consumption panel datasets, are undoubtedly subject to measurement problems, but classical measurement error is not the first order problem. See section 4 for a discussion.

revisits the taxi drivers' individual-level labour supply functions and plots their wage-offer curves. This reveals a rich picture individual behaviour. Section 4.6 uses these individual wage offer curves to examine the assumption of quantile invariance/single crossing. It shows that heterogeneity in individual wage offer curves cannot, in these data, be captured by the monotonicity assumption and demonstrates that if researchers take quantile demands to be indicative of the demands of an individual, then they gain a strongly misleading impression of individual behaviour as being simpler/smoother than it really is. Section 4.7 recovers the individual level welfare effects of wages changes - these too show empirically important re-ranking of individuals as wages change. Section 5 concludes and discusses further extensions.

2. The predictor

2.1. **Background.** The method used here is known as Wiener–Kolmogorov prediction or, more commonly, *universal kriging*.[8] The essence of the approach in the context of the generalised linear model was laid out in Goldberger (1962) who derived the best linear unbiased predictor (BLUP) in a general linear model in which the errors were non-spherical. The solution was to recognise that in the presence of interdependence in the disturbances, the pattern of sample residuals contains additional information which can be used to modify the regression function so as to reduce the prediction variance. The method used in this paper is standard in the geo-statistics and engineering literature, but to fix ideas it is useful first to review Goldberger's basic result and consider the generalised linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{V}$. The problem of interest is one of predicting a new drawing of the regressand given the vector of regressors. The actual draw will be

$$y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0$$

Given interdependence of disturbances in the sample, that is, in view of the fact that the variance-covariance matrix of the sample disturbances $\boldsymbol{V}$ is not simply proportional to the identity matrix, it is not reasonable to assume that the prediction disturbance ε_0 is independent of the sample disturbances $\boldsymbol{\varepsilon}$. Instead Goldberger assumes

$$\begin{aligned}\mathbb{E}[\varepsilon_0] &= 0\\ \mathbb{E}\left[(\varepsilon_0)^2\right] &= \sigma_{\mathbf{0}}^{\mathbf{2}}\\ \mathbb{E}[\varepsilon_0\boldsymbol{\varepsilon}] &= \mathbf{v}_0\end{aligned}$$

where $\mathbf{v}_0$ is the covariance between the prediction disturbance and the sample disturbances.

Consider now a linear predictor of y_0 denoted by $\hat{y}_0 = \boldsymbol{\omega}'\mathbf{Y}$ where $\boldsymbol{\omega}$ are weights. The objective is to choose the linear predictor which minimises the variance of the prediction $\sigma^2_{\hat{y}_0} = \mathbb{E}\left[(\hat{y}_0 - y_0)^2\right]$ subject to the unbiasedness constraint $\mathbb{E}[\hat{y}_0 - y_0] = 0$.[9] Goldberger (1962) describes the constrained optimisation problem and shows that the BLUP can be recovered by solving (see Goldberger (1962) p. 371, eq'n (3.9))

$$\begin{bmatrix}\mathbf{V} & \mathbf{X}\\ \mathbf{X}' & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\omega}\\ -\boldsymbol{\lambda}\end{bmatrix} = \begin{bmatrix}\mathbf{v}_0\\ \mathbf{x}_0\end{bmatrix}$$

for the weights $\boldsymbol{\omega}$ and the Lagrange multipliers $\boldsymbol{\lambda}$. This gives the following expression for the optimal weights

$$\hat{\boldsymbol{\omega}} = \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0 + \mathbf{V}^{-1}\left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{v}_0$$

and the BLUP can be written as

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\lambda}}_0' \cdot \left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]$$

where

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}Y\\ \hat{\boldsymbol{\lambda}}_0 &= \mathbf{V}^{-1}\cdot\mathbf{v}_0\end{aligned}$$

Thus the BLUP in this case is composed of two terms; the first term is the GLS predictor; the second is an estimate of the prediction disturbance itself. This latter component requires knowledge of the interdependence of disturbances along with the GLS residuals (which are estimates of the sample disturbances). Goldberger (1962) assumed that the covariance structure of the disturbances was either known or behaved according to some pre-defined model.

[8]See, for example, Cressie (1993), Stein (1999).

[9]Minimising the mean squared prediction error of an unbiased predictor is identical to minimising the prediction variance.

The approach used here is based on Goldberger's insight but is adapted to a spatial model in which the dependence structure is data-driven rather than assumed known *a priori*.

2.2. **Preliminaries.** Suppose we observe a set of N observations of prices and demands for an individual denoted: $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1,\ldots,N}$. Let $\boldsymbol{\rho}_i$ denote the prices normalised by the total budget so that $\boldsymbol{\rho}_i \cdot \mathbf{q}_i = 1$. Most of the rest of this paper will work with these normalised prices. Suppose further that this individual's demand behaviour is described by a system of demand curves $\mathbf{Q}(\boldsymbol{\rho})$. Following the classical revealed preference approach, the data are taken to represent direct observations on the individual's demand curves $\mathbf{Q}(\boldsymbol{\rho})$:[10]

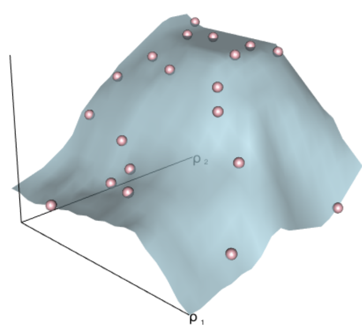
Assumption $\mathbf{q}_i = \mathbf{Q}(\boldsymbol{\rho}_i)$.

Consider demand over some region of interest in price-space

$$\{\mathbf{Q}(\boldsymbol{\rho}) \;:\; \boldsymbol{\rho} \in D\}$$

where D is a compact subset of $\mathbb{R}^K_{++}$. Define this to be the population of interest – that is, the population is the set of demands at all possible normalised prices in D. The data are taken to be a random sample from this population. The aim is to predict demand at some normalised price vector denoted $\boldsymbol{\rho}_0$, or at a sequence of such price vectors.

FIGURE 2.1. A demand curve: population and sample.



To fix ideas and to emphasise the spatial framework which follows, consider Figure 2.1. This shows a demand curve for a good which is part of a two-good demand system. In this case $\boldsymbol{\rho} \in \mathbb{R}^2_{++}$ and a point on the demand curve corresponds demand at the location given by the own- and cross-price coordinates $\boldsymbol{\rho}' = [\rho^1, \rho^2]$. The vertical axis shows quantity demanded against the normalised own- and cross-prices on the horizontals. The surface represents the demand curve and is treated as the population from which samples are generated. The sample is a collection of demands and prices and is represented by the points shown on the surface. The aim is to predict the surface at some new location $\boldsymbol{\rho}_0$ or a set of new locations.

It is assumed that the demand curve can be thought of a deterministic general shape/trend plus local fluctuations. The local fluctuations around the global shape/trend are directly analogous to the disturbances in Goldberger (1962). These local variations in the demand function will be described by a random field.

Assumption $Q(\boldsymbol{\rho})$ can be decomposed into a global shape function $\mu(\boldsymbol{\rho})$, and a real-valued random field $Z(\boldsymbol{\rho})$ such that $Q(\boldsymbol{\rho}) = \mu(\boldsymbol{\rho}) + Z(\boldsymbol{\rho})$ where the random field $Z(\boldsymbol{\rho})$ possesses finite first and second moments.

A random field is a multivariate generalisation of a stochastic process. Unlike a univariate stochastic process where the locations of the random variables are thought of as ordered points in time, the locations here are vectors identifying points in a multidimensional space. As with univariate stochastic processes, estimation and prediction of random fields is based upon auto-correlations. Qualitatively, the assumption used in this paper will be that the field values are more closely correlated when they are closer together. i.e. when price-budget differences are small. In a time series context, auto-correlations are necessarily chronological and often strongly directional (e.g. backward-looking only) in nature. In the case of random fields they may operate any direction.

[10]Measurement error and its impact are discussed below.

The use of a random field in this context is *not* intended to mean that the phenomenon of interest − choice behaviour − has resulted from a random process; it is there to allow the construction of a basis for inference and to quantify the uncertainty associated with the prediction at previously unobserved locations. This random field assumption, in the context of this model, merely provides a statistically-founded approach to a dataset consisting of samples collected in normalised price-space. Such an approach is absent from classical revealed preference methodology.

In order to be able to predict quantities from a single realisation of a random field or any stochastic process, some kind of stationarity assumption regarding the dependence between sample points is required.

Definition. A *stationary* random field is such that, for any vector $\mathbf{h} \in \mathbb{R}^K$ and for any set of points $\boldsymbol{\rho}_i,\ i = 1, ..., N$,

$$\begin{aligned} &Prob\left[Z(\boldsymbol{\rho}_1 + \mathbf{h}) \in [a_1, b_1], ..., Z(\boldsymbol{\rho}_N + \mathbf{h}) \in [a_N, b_N]\right] \\ &= Prob\left[Z(\boldsymbol{\rho}_1) \in [a_1, b_1], ..., Z(\boldsymbol{\rho}_N) \in [a_N, b_N]\right] \end{aligned}$$

Full stationarity means that none of the statistical properties of the random field change with location; that is they are the same everywhere in the region of interest in price space. Stationarity can also be summarised by saying that the finite dimensional distributions of a stationary field are translation invariant. A weaker form of stationarity (second order or weak stationarity) requires only that the mean of the random field is constant and the covariance function is independent of the absolute location and instead only depends on the separating vector. The form of stationarity required here is yet weaker still. It is characterised in terms of the variogram of the random field.

Definition. The *variogram* of a random field is $2\gamma\left(\boldsymbol{\rho}, \boldsymbol{\rho} + \mathbf{h}\right) := Var\left[Z(\boldsymbol{\rho}) - Z(\boldsymbol{\rho} + \mathbf{h})\right]$ where $\mathbf{h} \in \mathbb{R}^K$.

The variogram of a random field is the variance of the difference in field values at two locations:

$$Var\left[Z(\mathbf{x}) - Z(\mathbf{y})\right]$$

The quantity

$$\gamma(\mathbf{x}, \mathbf{y}) := \frac{1}{2} Var\left[Z(\mathbf{x}) - Z(\mathbf{y})\right]$$

is the semi-variogram of Z. The stationarity assumption used here is characterised in terms of the semi-variogram.

Definition. A random field is *intrinsically* stationary if

$$\left.\begin{matrix} \mathbb{E}\left[Z(\boldsymbol{\rho})\right] = \mathbb{E}\left[Z(\boldsymbol{\rho} + \mathbf{h})\right] \\ \gamma(\boldsymbol{\rho}, \boldsymbol{\rho} + \mathbf{h}) = \gamma(\mathbf{h}) \end{matrix}\right\} \forall \boldsymbol{\rho}, \boldsymbol{\rho} + \mathbf{h} \in D$$

An intrinsically stationary random field has a constant mean and a semi-variogram function which only depends on the separation between locations, and not on their absolute positions. Any stationary or weakly stationary field is also intrinsically stationary but the reverse is not true. Under the stronger assumption of second order/weak stationarity the use of the semi-variogram to characterise dependence would be equivalent to using covariances or auto-correlations (see, for example, Cressie (1993) for a covariance/second-order stationarity-based version of what follows).

Assumption The random field $Z(\boldsymbol{\rho})$ is intrinsically stationary and mean-zero.

This intrinsic stationarity assumption means that $Z(\boldsymbol{\rho})$ the semi-variogram reduces to

$$\frac{1}{2}\mathbb{E}\left[\left(Z(\boldsymbol{\rho}) - Z(\boldsymbol{\rho} + \mathbf{h})\right)^2\right] = \gamma(\mathbf{h})$$

Give our assumptions on the random field, the variance of the increment between $\boldsymbol{\rho}$ and $\boldsymbol{\rho} + \mathbf{h}$ is finite and its value only depends on the separating vector $\mathbf{h}$. It does not depend on its absolute position in D, i.e. $Var[Z(\boldsymbol{\rho}+\mathbf{h}) - Z(\boldsymbol{\rho})] < \infty,\ \forall \boldsymbol{\rho}, \boldsymbol{\rho} + \mathbf{h} \in D$ and is only a function of $\mathbf{h}$.

2.3. **Prediction.** Given a new normalised price vector $\boldsymbol{\rho}_0$ we want to evaluate the predictor $\hat{Q}(\boldsymbol{\rho}_0)$ at that point subject to the unbiasedness requirement:

$$\mathbb{E}\left[\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right] = 0$$

Minimising the mean squared prediction error of an unbiased predictor is identical to minimising the prediction variance:

$$\begin{aligned} mspe(\hat{Q}(\boldsymbol{\rho}_0)) &:= \mathbb{E}\left[\left(\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right)^2\right] \\ &= \underbrace{\mathbb{E}\left[\left(\hat{Q}(\boldsymbol{\rho}_0) - \mathbb{E}\left[\hat{Q}(\boldsymbol{\rho}_0)\right]\right)^2\right]}_{variance} + \underbrace{\left(\mathbb{E}\left[\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right]\right)^2}_{bias} \end{aligned}$$

Therefore the prediction problem is to solve

$$\min_{\hat{Q}(\boldsymbol{\rho}_0)} \mathbb{E}\left[\left(\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right)^2\right] \quad \text{subject to} \quad \left(\mathbb{E}\left[\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right]\right)^2 = 0$$

Confining ourselves to linear predictors

$$\hat{Q}(\boldsymbol{\rho}_0) := \sum_{i=1}^{N} \omega_i Q(\boldsymbol{\rho}_i)$$

or in matrix notation $\hat{Q}(\boldsymbol{\rho}_0) = \boldsymbol{\omega}'\mathbf{Q}$ with weights $\boldsymbol{\omega}$ on each sample point such that $\sum_{i=1}^{N} \omega_i = 1$.

Given the assumption that $Q(\boldsymbol{\rho})$ can be decomposed into a global shape function $\mu(\boldsymbol{\rho})$, and a real-valued random field $Z(\boldsymbol{\rho})$ such that

$$Q(\boldsymbol{\rho}) = \mu(\boldsymbol{\rho}) + Z(\boldsymbol{\rho}),$$

the linear predictor and the true value are

$$\hat{Q}(\boldsymbol{\rho}_0) = \sum_{i=1}^{N} \omega_i \left[\mu(\boldsymbol{\rho}_i) + Z(\boldsymbol{\rho}_i)\right] \text{ and } \mathrm{Q}(\boldsymbol{\rho}_0) = \mu(\boldsymbol{\rho}_0) + Z(\boldsymbol{\rho}_0)$$

The bias is

$$\begin{aligned} \mathbb{E}\left[\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right] &= \mathbb{E}\left[\sum_{i=1}^{N} \omega_i \left[\mu(\boldsymbol{\rho}_i) + Z(\boldsymbol{\rho}_i)\right] - \left[\mu(\boldsymbol{\rho}_0) + Z(\boldsymbol{\rho}_0)\right]\right] \\ &= \sum_{i=1}^{N} \omega_i \left(\mathbb{E}\left[\mu(\boldsymbol{\rho}_i)\right] + \underbrace{\mathbb{E}\left[Z(\boldsymbol{\rho}_i)\right]}_{=0}\right) - \mathbb{E}\left[\mu(\boldsymbol{\rho}_0)\right] + \underbrace{\mathbb{E}\left[Z(\boldsymbol{\rho}_0)\right]}_{=0} \\ &= \sum_{i=1}^{N} \omega_i \mu(\boldsymbol{\rho}_i) - \mu(\boldsymbol{\rho}_0) \end{aligned}$$

where the final step uses the assumption that $\mu(\boldsymbol{\rho})$ is deterministic and so $\mathbb{E}\left[\mu(\boldsymbol{\rho}_i)\right] = \mu(\boldsymbol{\rho}_i)$. This global shape/trend component is modelled using a linear combination of known basis functions $\{f_0(\boldsymbol{\rho}), ..., f_J(\boldsymbol{\rho})\}$ with $f_0(\boldsymbol{\rho}) = 1$ (e.g. polynomials or trigonometric functions)

$$\mu(\boldsymbol{\rho}) = \sum_{j=0}^{J} \beta_j f_j(\boldsymbol{\rho})$$

with unknown coefficients vector $\underset{(J\times 1)}{\boldsymbol{\beta}} = [\beta_0, ..., \beta_J]'$.

To put this into more compact matrix notation, let

$$\mathbf{Q} := \begin{bmatrix} Q(\boldsymbol{\rho}_1) \\ \vdots \\ Q(\boldsymbol{\rho}_N) \end{bmatrix} \quad \mathbf{X} := \begin{bmatrix} 1 & f_1(\boldsymbol{\rho}_1) & \cdots & f_J(\boldsymbol{\rho}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_1(\boldsymbol{\rho}_N) & \cdots & f_J(\boldsymbol{\rho}_N) \end{bmatrix} \quad \mathbf{Z} := \begin{bmatrix} Z(\boldsymbol{\rho}_1) \\ \vdots \\ Z(\boldsymbol{\rho}_N) \end{bmatrix}$$

then

$$\mathbf{Q} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}.$$

and the bias is then expressed as

$$\mathbb{E}\left[\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right] = \sum_{i=1}^{N} \omega_i \mu(\boldsymbol{\rho}_i) - \mu(\boldsymbol{\rho}_0) = \boldsymbol{\omega}'\mathbf{X}\boldsymbol{\beta} - \mathbf{x}_0'\boldsymbol{\beta}$$

where $\mathbf{x}_0 = [1, f_1(\boldsymbol{\rho}_0), ..., f_J(\boldsymbol{\rho}_0)]'$. The unbiasedness condition therefore reduces to

$$\boldsymbol{\omega}'\mathbf{X}\boldsymbol{\beta} - \mathbf{x}_0'\boldsymbol{\beta} = 0$$

and so it suffices that

$$\boldsymbol{\omega}'\mathbf{X} - \mathbf{x}_0 = 0. \tag{2.1}$$

This does not depend on the coefficients $\boldsymbol{\beta}$.

Turning now to the objective function, the prediction variance is

$$\begin{aligned}
\sigma_e^2 &:= Var\left[\hat{Q}(\boldsymbol{\rho}_0) - Q(\boldsymbol{\rho}_0)\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{N}\omega_i Q(\boldsymbol{\rho}_i) - Q(\boldsymbol{\rho}_0)\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{N}\omega_i \left[\mu(\boldsymbol{\rho}_i) + Z(\boldsymbol{\rho}_i)\right] - \left[\mu(\boldsymbol{\rho}_0) + Z(\boldsymbol{\rho}_0)\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{N}\omega_i Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_0) + \underbrace{\sum_{i=1}^{N}\omega_i \mu(\boldsymbol{\rho}_i) - \mu(\boldsymbol{\rho}_0)}_{=0}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{N}\omega_i Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_0)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_j) - 2\sum_{i=1}^{N}\omega_i Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_0) + \left[Z(\boldsymbol{\rho}_0)\right]^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_j) - 2\sum_{i=1}^{N}\omega_i Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_0) + \left[Z(\boldsymbol{\rho}_0)\right]^2\right]
\end{aligned}$$

Inserting

$$\sum_{i=1}^{N}\omega_i \frac{\left[Z(\boldsymbol{\rho}_i)\right]^2}{2}\underbrace{\sum_{j=1}^{N}\omega_j}_{=1} + \sum_{j=1}^{N}\omega_j \frac{\left[Z(\boldsymbol{\rho}_j)\right]^2}{2}\underbrace{\sum_{i=1}^{N}\omega_i}_{=1} = \sum_{i=1}^{N}\omega_i \left[Z(\boldsymbol{\rho}_i)\right]^2$$

into the first term gives

$$\sigma_e^2 = \mathbb{E}\left[\begin{array}{c} -\sum_{i=1}^{N}\omega_i \frac{\left[Z(\boldsymbol{\rho}_i)\right]^2}{2}\underbrace{\sum_{j=1}^{N}\omega_j}_{=1} + \sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_j) - \sum_{j=1}^{N}\omega_j \frac{\left[Z(\boldsymbol{\rho}_j)\right]^2}{2}\underbrace{\sum_{i=1}^{N}\omega_i}_{=1} \\ + \sum_{i=1}^{N}\omega_i \left[Z(\boldsymbol{\rho}_i)\right]^2 - 2\sum_{i=1}^{N}\omega_i Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_0) + \left[Z(\boldsymbol{\rho}_0)\right]^2 \underbrace{\sum_{i=1}^{N}\omega_i}_{=1} \end{array}\right]$$

Now simplifying

$$\sigma_e^2 = \mathbb{E}\left[\begin{array}{c} -\sum_{i=1}^{N}\omega_i \frac{\left[Z(\boldsymbol{\rho}_i)\right]^2}{2}\sum_{j=1}^{N}\omega_j + \sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_j) - \sum_{j=1}^{N}\omega_j \frac{\left[Z(\boldsymbol{\rho}_j)\right]^2}{2}\sum_{i=1}^{N}\omega_i \\ + \sum_{i=1}^{N}\omega_i \left[Z(\boldsymbol{\rho}_i)\right]^2 - 2\sum_{i=1}^{N}\omega_i Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_0) + \left[Z(\boldsymbol{\rho}_0)\right]^2 \sum_{i=1}^{N}\omega_i \end{array}\right]$$

$$\sigma_e^2 = \mathbb{E}\left[\begin{array}{c} -\sum_{i=1}^{N}\omega_i \frac{\left[Z(\boldsymbol{\rho}_i)\right]^2}{2}\sum_{j=1}^{N}\omega_j + \sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_j) - \sum_{j=1}^{N}\omega_j \frac{\left[Z(\boldsymbol{\rho}_j)\right]^2}{2}\sum_{i=1}^{N}\omega_i \\ + \sum_{i=1}^{N}\omega_i \left\{\left[Z(\boldsymbol{\rho}_i)\right]^2 - 2Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_0) + \left[Z(\boldsymbol{\rho}_0)\right]^2\right\} \end{array}\right]$$

$$\sigma_e^2 = \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{-\frac{[Z(\boldsymbol{\rho}_i)]^2}{2} + Z(\boldsymbol{\rho}_i)Z(\boldsymbol{\rho}_j) - \frac{[Z(\boldsymbol{\rho}_j)]^2}{2}\right\} + \sum_{i=1}^{N}\omega_i\left[Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_0)\right]^2\right]$$

$$\sigma_e^2 = \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{-\frac{[Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_j)]^2}{2}\right\} + 2\sum_{i=1}^{N}\omega_i\frac{[Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_0)]^2}{2}\right]$$

$$\sigma_e^2 = \mathbb{E}\left[-\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{[Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_j)]^2}{2}\right\} + 2\sum_{i=1}^{N}\omega_i\frac{[Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_0)]^2}{2}\right]$$

We now use the assumption of intrinsic stationarity to introduce the semi-variogram

$$\sigma_e^2 = -\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j\underbrace{\frac{\mathbb{E}\left[\left(Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_j)\right)^2\right]}{2}}_{\gamma(\boldsymbol{\rho}_i,\boldsymbol{\rho}_j)} + 2\sum_{i=1}^{N}\omega_i\underbrace{\frac{\mathbb{E}\left[\left(Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_0)\right)^2\right]}{2}}_{\gamma(\boldsymbol{\rho}_i,\boldsymbol{\rho}_0)}$$

$$\sigma_e^2 = -\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j\gamma(\boldsymbol{\rho}_i,\boldsymbol{\rho}_j) + 2\sum_{i=1}^{N}\gamma(\boldsymbol{\rho}_i,\boldsymbol{\rho}_0)$$

In matrix notation

$$\sigma_e^2 := -\boldsymbol{\omega}'\boldsymbol{\Gamma}\boldsymbol{\omega} + 2\boldsymbol{\omega}'\boldsymbol{\gamma}_0 \tag{2.2}$$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma(\boldsymbol{\rho}_1,\boldsymbol{\rho}_1) & \cdots & \gamma(\boldsymbol{\rho}_1,\boldsymbol{\rho}_N) \\ \vdots & \ddots & \vdots \\ \gamma(\boldsymbol{\rho}_N,\boldsymbol{\rho}_1) & \cdots & \gamma(\boldsymbol{\rho}_N,\boldsymbol{\rho}_N) \end{bmatrix} \qquad \boldsymbol{\gamma}_0 = \begin{bmatrix} \gamma(\boldsymbol{\rho}_1,\boldsymbol{\rho}_0) \\ \vdots \\ \gamma(\boldsymbol{\rho}_N,\boldsymbol{\rho}_0) \end{bmatrix}$$

Using equations (2.1) and (2.2) the problem of minimising the variance of the prediction error subject to the unbiasedness constraint becomes the constrained optimisation problem

$$\min_{\boldsymbol{\omega}} -\boldsymbol{\omega}'\boldsymbol{\Gamma}\boldsymbol{\omega} + 2\boldsymbol{\omega}'\boldsymbol{\gamma}_0 \text{ subject to } \boldsymbol{\omega}'\mathbf{X}\text{-}\mathbf{x}_0 = \mathbf{0}$$

The Lagrangian is

$$L(\boldsymbol{\lambda},\boldsymbol{\omega}) = -\boldsymbol{\omega}'\boldsymbol{\Gamma}\boldsymbol{\omega} + 2\boldsymbol{\omega}'\boldsymbol{\gamma}_0 - (\boldsymbol{\omega}'\mathbf{X} - \mathbf{x}_0)\,\boldsymbol{\lambda}$$

with the Lagrange multiplier parameter vector $\boldsymbol{\lambda} := [\lambda_0, ..., \lambda_N]$.

It is then straightforward (for example, Cressie, (1993)) to show that this has a closed-form solution as long as the number of basis functions is less than the number of sample points:

$$\hat{\boldsymbol{\omega}} = \boldsymbol{\Gamma}^{-1}\left[\boldsymbol{\gamma}_0 - \mathbf{X}\left(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - \mathbf{x}_0\right)\right]$$

$$\hat{\boldsymbol{\lambda}} = \left(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - \mathbf{x}_0\right)$$

It follows that the best linear unbiased predictor is

$$\hat{Q}(\boldsymbol{\rho}_0) = \hat{\boldsymbol{\omega}}'\mathbf{Q}$$

and the variance of the prediction is

$$\hat{\sigma}_e^2(\boldsymbol{\rho}_0) = \begin{bmatrix} \hat{\boldsymbol{\omega}} & \hat{\boldsymbol{\lambda}} \end{bmatrix}\begin{bmatrix} \boldsymbol{\gamma}_0 \\ \mathbf{x}_0 \end{bmatrix}$$

The *key* feature of this procedure, given the ultimate aim of applying the classical revealed preference machinery, is that it is an exact interpolator. This is because when the prediction point is also a sample point the vector $\begin{bmatrix} \boldsymbol{\gamma}_0 & \mathbf{f}_0 \end{bmatrix}$ is identical to one of the columns of $\boldsymbol{\Gamma}$ and so the unique solution of the simultaneous equation system sets one of the weights to one and the rest to zero. Furthermore, the variance of the predictor at such a point is zero since all of the Lagrange multipliers are zero. This reflects the basic assumption in the approach that, at the sample points, there is no uncertainty as we directly observe the demand.

In order to take this procedure to the data we need to understand the interdependence between field values at different locations. (Recall Goldberger (1962) assumed that the covariance function was known in order to apply the BLUP procedure to the generalised linear model). This means modelling the semi-variogram function to recover $\gamma(\boldsymbol{\rho}_i,\boldsymbol{\rho}_j)$ at the sample locations to populate the Γ matrix, and, by using the stationarity assumption, using it to generate the values of $\gamma(\boldsymbol{\rho}_i,\boldsymbol{\rho}_0)$ to create the $\boldsymbol{\gamma}_0$ vector. The intrinsic stationarity assumption means that

$$\gamma\left(\boldsymbol{\rho}_i,\boldsymbol{\rho}_j\right) = \frac{1}{2}\mathbb{E}\left[\left(Z(\boldsymbol{\rho}_i) - Z(\boldsymbol{\rho}_j)\right)^2\right] = \gamma\left(\mathbf{h}\right)$$

i.e., that the semi-variogram is only function of the separation between the sampling points. The semi-variogram an therefore be modelled as the conditional expectation

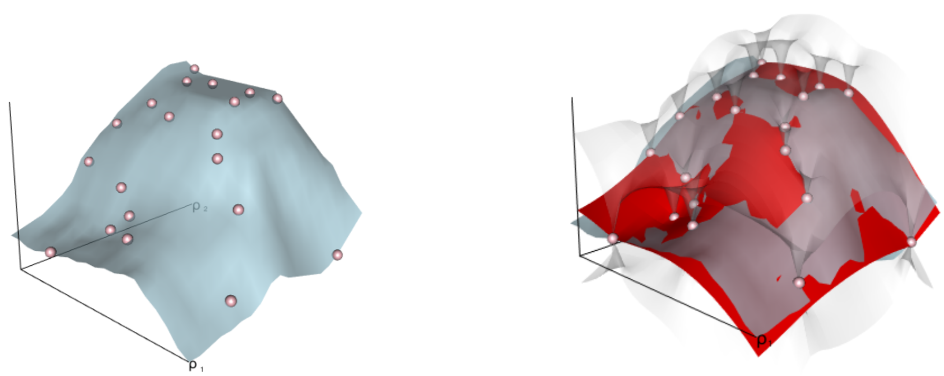
$$\mathbb{E}\left[\frac{1}{2}\left(Z(\boldsymbol{\rho}_i)-Z(\boldsymbol{\rho}_j)\right)^2\middle|\,\mathbf{h}\right]$$

The function $\gamma(\mathbf{h})$ needs to satisfy a number of properties to ensure that the prediction variance is non-negative. These are: (*i*) $\gamma(\mathbf{h}) \geq 0$, (*ii*) $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$, (*iii*) $\lim_{|\mathbf{h}|\to\infty} \frac{\gamma(\mathbf{h})}{|\mathbf{h}|^2} \leq 0$ and (*iv*) $\sum_{i=1}\sum_{j=1}^{N} a_i a_j \gamma(\boldsymbol{\rho}_i, \boldsymbol{\rho}_i) \leq 0$ for any finite sequence of points $\{\boldsymbol{\rho}_i\}_{i=1,\ldots,N}$ and for any finite sequence of real numbers $\{a_i\}_{i=1,\ldots,N}$ such that $\sum_i a_i = 0$, i.e. γ is a conditionally negative semi-definite function (Cressie (1993)). Not every arbitrary function is therefore a valid semi-variogram function and it is convenient to use a parametric functional form for $\gamma(\mathbf{h})$ which satisfies these conditions (see for example Webster and Oliver (2007)).

2.4. **Two Examples.**

2.4.1. *Example 1: An irregular, theory-consistent demand surface.* To illustrate the ideas outlined consider Figure 2.2. Panel (A) replicates Figure 2.1 above and shows the demand for a good on the vertical axis against normalised own- and cross-prices on the horizontals.[11] This represents the demand for one good in a two-good demand system. The demand for the other good (not illustrated) is implied by adding-up. The behaviour represented by this demand curve is theory-consistent in the sense that the corresponding two-good demand system satisfies the weak axiom of revealed preference (WARP) everywhere. WARP is both necessary and sufficient for utility maximisation when there are only two goods.[12] This examples uses the same 20 randomly drawn sample observations on the demand curve shown previously in Figure 2.1. As is evident, they are not particularly evenly spread across the demand curve.

FIGURE 2.2. Prediction of an irregular but theory-consistent demand curve.



(A) The true demand curve and sample points

(B) True and estimated demand curves and confidence intervals

Panel (B) shows the estimated demand curve (in red) overlaid on the true (population) demand. The confidence intervals are also shown as partially transparent surfaces. Note that in most cases it would be sensible to confine the area of prediction to being within the range of the variation seen in the data as nonparametric prediction methods

[11] This demand curve model is, in fact, part of Maungawhau (Mt Eden), a volcano in the Auckland volcanic field. This data set gives high resolution topographic information for a section of the volcano. The "own" and "cross" price values are just the geographical eastings-westing locations. At the resolution at which it was measured this part of Maungawhau satisfies the Weak Axiom of Revealed Preference and is therefore consistent with being as-if generated by the maximisation of a well-behaved utility function (as well as plate tectonics). Maungawhau represents a more interesting testbed for the methods described in this paper than data simulated from a standard economic model of preferences. This is because even models with moderately elaborate functional forms like the trans-log, the cost function of which is able to approximate any twice-differentiable cost function (Diewert (1981)), tend to be very smooth and regular in comparison; the random field predictor fits extremely well as a result. Maungawhau is more of a challenge whilst still representing a rational demand system. It also serves, once again, to emphasise the spatial character of demand curves as surfaces rather than lines.

[12] Varian (1982).

such as this would not be expected to perform well outside of this range. Nonetheless in the interests of seeing how the method does I have estimated the demand curve over the full range of values for which the true model is known.

The model matches the demand curve closely in areas where the observations are dense – and of course it perfectly fits the demand at the sample points themselves. But it does less well in regions further away from the sample points. The confidence interval on the forecasts similarly shrinks to zero as it approaches the known (sample) points but gets wider in areas in which the data are sparse. This gives the confidence interval a characteristic dimpled shape.

2.4.2. *Example 2: Revealed preference bounds on a marginal demand curve.* The second example relates the predictor to classical revealed preference bounds. Consider the following data,

$$\text{normalised prices} := \begin{bmatrix} 0.5 & 1.14 & 0.48 & 1.00 & 0.74 \\ 0.70 & 0.70 & 1.39 & 0.54 & 0.70 \end{bmatrix} \quad \text{quantities} := \begin{bmatrix} 1.11 & 0.34 & 0.81 & 0.65 & 0.76 \\ 0.49 & 0.56 & 0.44 & 0.65 & 0.63 \end{bmatrix}$$

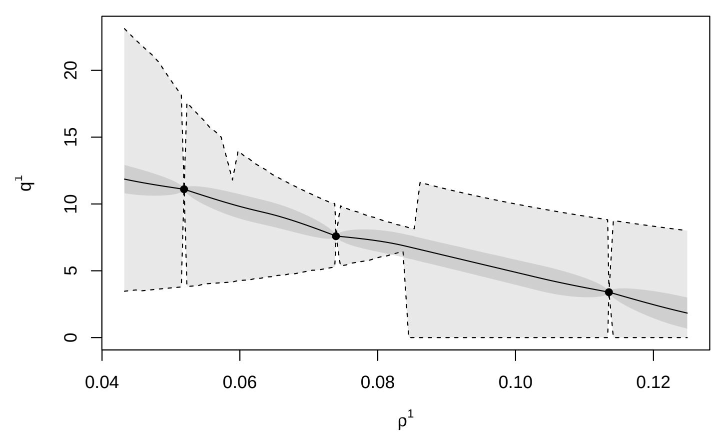
and consider the demand for good 1 as the price of good 1 is varied while the price of good 2 is held constant at 0.70. The support set[13] is

$$S(\boldsymbol{\rho}_0) = \left\{ \mathbf{q}_0 | \left\{ \boldsymbol{\rho}_0, \mathbf{q}_0, \{\boldsymbol{\rho}_i, \mathbf{q}_i\}_{i=1,\ldots,N} \right\} \text{ satisfies WARP} \right\}$$

and is calculated for each $\boldsymbol{\rho}_0$ is the sequence. The bounds on the demand for good 1 are shown in Figure 2.3 as the shaded area bounded by the dashed-lines.

Note that this example is constructed specifically so that the sequence of prices along the forecast demand curve passes through three observed prices. This is designed as a best-case example for revealed preference analysis because at those prices the revealed preference demands are able to collapse to the observed demand. Elsewhere the RP bounds are wide. This is true even at prices arbitrarily close to the observed points. Any demand function which lies within these bounds is WARP-consistent, however implausible behaviourally, even a step-function. But while no admissible function will, by definition, violate a model of maximising behaviour many will violate common sense and our intuition of what demand curves should look like.

FIGURE 2.3. The bounds on the marginal demand curve from Revealed Preference and the modelled demand curve



The solid line interpolating the observed demands is the BLUP of the demand curve. The darker shaded area around it is the 95% confidence interval. As with the first example, the confidence interval on the forecasts shrinks as it approach the known (sample) points creating the dimpled effect. The improvement over the best-case revealed preference bounds is evident.

In this case the forecast demands are within the revealed preference bounds. Indeed they satisfy WARP everywhere and are thus consistent with having been generated by the same single set of rational preferences which generated the data. This is not necessarily always automatically the case; for example, when the data themselves are not consistent with rationality. The next section shows how to impose rationality on the forecast demands.

[13] See Varian (1982).

2.5. **Adding restrictions.** Thus far there has been no attempt to impose any of the requirements of economic theory on the predicted demand curves. Adding up and homogeneity are easily dealt with, and indeed the predictor described above will necessarily retain these properties, but the global consistency of the demand system with a single preference ordering is not guaranteed without additional restrictions. Revealed preference axioms, being global in nature, are particularly advantageous because they have a different flavour from the usual local smoothness assumptions. While knowing how the derivatives of a function behave in one area of its domain tells you little about how they behave elsewhere this is *not* true of a function restricted by revealed preference axioms: any ensemble of data points, however remote from each other, must still jointly satisfy the relevant revealed-preference restrictions in combination with both each other and the rest of the surface.

Let Æ(.) denote a function which computes the Afriat (cost) efficiency of a sequence of demands and budgets. Afriat Efficiency is a conventional and economically meaningful index of how close the data are to being consistent with utility maximisation. It measures how much of the budget the data-generating individual appears to waste by failing to optimise. If this individual's observed demands satisfy the revealed preference restrictions then $Æ(\{\boldsymbol{\rho}_i, \mathbf{q}(\boldsymbol{\rho}_i)\}_{i=1,\ldots,N}) = 1$; otherwise $Æ(\{\boldsymbol{\rho}_i,\mathbf{q}(\boldsymbol{\rho}_i)\}_{i=1,\ldots,N}) \in (0, 1]$.

Given the individual's *observed* level of cost-efficiency a reasonable and conservative approach which does not conflict with what is observed would be to have the efficiency of the estimated demand system match it. For example, if they are observed to behave consistently with utility maximisation up to a level of cost efficiency of 0.97, then it seems sensible to wish to forecast their demand behaviour to also be 0.97-efficient. Suppose, therefore, that using the methods from the previous section their demands are predicted over a sequence of $(M - N) > 0$ previously unobserved budgets, plus the Nobserved budgets themselves. Denote the set of locations, predictions and prediction error variance matrices by $\{\boldsymbol{\rho}_i, \hat{\mathbf{q}}(\boldsymbol{\rho}_i), \hat{\boldsymbol{\Omega}}_i\}_{i=1,\ldots,M}$. Note that these are a mixture of estimates at new budgets (and their corresponding prediction error variances), plus, since the predictor exactly matches the demands at observed budgets, the original data themselves. These data-points will have forecast variances of zero. Following Blundell *et al* (2008), a restricted predictor can be generated using the Gaussian quasi-likelihood ratio or minimum distance criterion function

$$\min_{\{\mathbf{q}^r_m\}_{m=1,\ldots,M}} \sum_{m=1}^{M} (\mathbf{q}^r_m - \hat{\mathbf{q}}_m(\boldsymbol{\rho}_m))'\hat{\boldsymbol{\Omega}}_m^{-1}(\mathbf{q}^r_m - \hat{\mathbf{q}}_m(\boldsymbol{\rho}_m))' \text{ subject to } Æ(\{\boldsymbol{\rho}_m, \mathbf{q}^r_m\}_{m=1,\ldots,M}) = Æ(\{\boldsymbol{\rho}_i, \mathbf{q}_i\}_{i=1,\ldots,N})$$

Members of the set $\{\boldsymbol{\rho}_i, \hat{\mathbf{q}}(\boldsymbol{\rho}_i), \hat{\boldsymbol{\Omega}}_i\}_{i=1,\ldots,M}$ which are in fact observations will not move as a result of this procedure because their prediction error variance is zero and hence the loss or penalty associated with moving them is infinite. However, predicted demands at previously-unobserved budgets will move, and the less precisely they are forecast the more they will move.

The solution to this weighted minimum-distance method defines an ensemble of demands and forecasts which will which are unique almost everywhere and which mutually satisfy the revealed preference restrictions at the same observed cost-efficiency as the individual whose behaviour is being modelled. Note that it is also possible, with an obvious small adaptation of the procedure above, to impose full rationality on the M predicted points on the demand curve even if the N data points are not rational. The overall cost efficiency of the combined data and predictions will still be that of the data themselves (this is due to the fact that the cost-efficiency of a dataset cannot be less than the cost efficiency of any sub-set of that dataset) but it will often be useful for the welfare analysis of counterfactuals, for example, to have the predicted demands be fully and mutually consistent with well-behaved preferences.

3. Application and Further Extensions

This section studies the consumption and labour supply decisions of a sample of New York cab drivers. These data were collected by Farber (2005, 2008) and they provide an interesting and informative empirical study for two main reasons. Firstly the canonical consumption-leisure model of labour supply, being low-dimensional, makes it straightforward to illustrate the results visually. This means that we are able to inspect the entire demand surface *in toto*; we are not obliged to focus on selected slices through this surface, for example marginal demand curves (plots of own- and cross-price effects holding other prices and the budget constant), or Engel curves (which hold both prices constant). This gives a much more complete picture of how these methods are working. Secondly these individuals are not necessarily rational; V. Crawford and Meng (2011) provide evidence of significant reference-dependence in their behaviour. As a result these data provide a more challenging and interesting environment in which to try to apply revealed preference theory. This section uses these individuals to follow up a number of substantive topics which, I argue, only the combination of panel data and this statistical approach which allows individual-level economic analysis can satisfactorily explore.

3.1. **Data.** Farber (2005, 2008) collected 538 trip sheets for 15 drivers between June 1999 and May 2001. Each trip sheet records the driver's name, hack number, and date and the details of each fare. For each fare the data record the start time, start location, end time, end location and fare.

The object of interest here is the drivers' demands in consumption-leisure space (denoted (c, l)). The time a driver spends working is the sum of time spent driving (d) with a fare-paying passenger aboard, the time spend waiting (v) for the next passenger when a driver earns nothing and the time on rest breaks (b). The analogue of the wage is earnings per hour driving. A consumer price index for NYC/NJ published by the BLS[14] denoted p is used to measure the price of consumption. The basic budget constraint for each driver is $pc = wd$ where w is earnings per hour earning (driving). Given $d + v + b + l = T$ this gives the budget constraint in (c, l) space as

$$pc = w(T - v - b) - wl$$

The principal variable affecting the amount a driver can earn is the amount of time he has to spend waiting for a fare (v); when the streets are busy or when the weather is inclement, the flow of customers is higher and the less time a driver has to spend waiting for a fare. Shift-to-shift changes in this waiting time act as a shift-specific fixed cost. Variations in this waiting time induce shift-to-shift variations in the intercept of the budget constraint. Fare themselves are regulated but they too vary from job-to-job according to the length of the journey and the time it takes to complete it, as well as the destination (e.g. airport runs are fixed fares and journeys outside of the five boroughs are charged according to different – higher – tariffs). Together these effects create shift-to-shift variation in realised wages which are idiosyncratic to each driver and which induce crossing of the consumption-leisure budget constraint.

The data is the unbalanced panel $\left\{\mathbf{p}_i^j, \mathbf{q}_i^j\right\}_{i=1,\ldots,N_i}^{j=1,\ldots,15}$ where j indexes individuals and

$$\mathbf{p}_i^j = \begin{bmatrix} w_i^j \\ p_i \end{bmatrix} = \begin{bmatrix} \text{hourly wage} \\ \text{numeraire} \end{bmatrix}; \quad \mathbf{q}_i^j = \begin{bmatrix} l_i^j \\ c_i^j \end{bmatrix} = \begin{bmatrix} \text{24 - work hours - breaks - waiting} \\ \text{consumption} \end{bmatrix}$$

The data are ultimately from cab meters which are calibrated, sealed and subject to checking by the Taxi and Limousine Commission. Thus it would seem plausible that they are relatively immune from significant measurement errors of the typical classical kind where the observed value is thought of as the true value plus a mean-zero perturbation.[15] Nonetheless they are subject to measurement problems of a different kind. This relates to trip sheets which may be missing entirely. There is no independent record of whether a driver worked on a particular day other than can be inferred by the presence/absence of a trip sheet. If a trip sheet is not available for a specific driver on a given day, it is not possible to know whether that driver did not work on that day or if they did work and the trip sheet was not provided. This prevented Farber (2008) from looking at an intertemporal model and also explains, if not justifies, the use of a static model in this paper as well. In the present context missing observations will mean that the scatter of observations across the demand curves are more sparse than the otherwise might be. Trip sheets which are missing at random will still result in a random spatial point pattern which will provide coverage of the normalised price-space, but if missing observations are more likely on certain shifts (for example which earnings are high or low) this will not be the case. In such regions of sparse data the confidence intervals will reflect this, however, I have not looked further into this.

Table 1 reports descriptive statistics for each driver. Figure 4.1 in the Appendix depicts the dataset in its entirety as budget constraints and choices for each of the drivers. Two features are noteworthy. Firstly the budget constraints cross frequently. Secondly, there is a great deal of price and behavioural variation on display.

[14]https://www.bls.gov/regions/new-york-new-jersey/news-release/consumerpriceindex_newyorkarea.htm

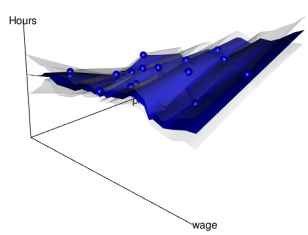
[15]If they are subject to perturbation it is not of a classical mean-zero kind. See 'New York Cabs Gouged Riders Out of Millions', *New York Times*, 12 March 2010.

Table 1. Descriptive Statistics, by driver

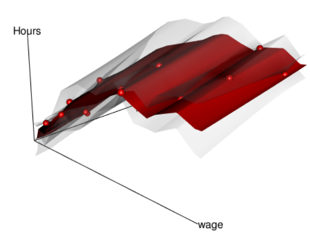
Driver	Shifts	Income ($)	Driving (hrs)	Breaks (hr's)	Waiting (hr's)	Wage ($)
1	39	157.58	4.32	0.90	2.53	36.41
2	14	97.10	2.78	2.41	1.11	34.68
3	40	147.51	4.52	0.39	1.76	33.02
4	23	144.96	3.98	2.11	2.48	38.12
5	24	160.71	4.42	0.74	2.05	36.69
6	37	172.44	5.13	0.86	2.64	34.23
7	19	162.02	5.47	0.54	1.70	30.61
8	45	133.19	3.90	1.65	2.45	33.83
9	13	157.95	4.03	0.55	2.13	39.44
10	17	165.84	4.49	0.64	2.57	37.37
11	70	172.01	4.56	0.93	2.28	37.72
12	72	203.05	5.84	0.60	2.69	35.07
13	33	163.51	4.63	0.97	2.29	36.01
14	46	156.23	4.80	0.67	2.30	32.73
15	46	128.97	3.66	0.24	1.66	35.62

3.2. **Initial Labour Supply Estimates.** The labour supply curve is estimated for each driver individually. The results for all of the drivers are shown, in full, in Figure 4.2 in the Appendix. The consumption demand functions corresponding to these labour supply functions are implied by adding up. They are not illustrated but are readily available from the author.

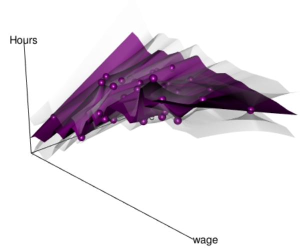
Figure 3.1. Labour supply curves



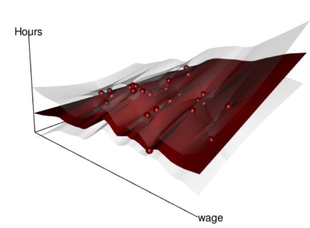
(a) Driver 1



(b) Driver 2



(c) Driver 14



(d) Driver 15

Figure 3.1 illustrates the results for four drivers. The vertical axis measures labour supplied in hours. The horizontal axes plot normalised wages and the price of consumption. Generally labour supply curves are upward-sloping with respect to wages – though there are obvious local departure from this (for example, Driver 14's wage-responses are severely non-monotonic at low values of the price of consumption). Another feature which is evident is the cross-price effect and the heterogeneity within it; for Drivers 1 and 15 there is evidence that the cross price effect is negative at low wages but positive at higher wages; for Driver 14 the opposite appears to be the case; for Driver 2 there seems to be little cross-price responsiveness at any wage level. However the principal property which stands out with these results is that the data, and hence the estimated function which fits them, are such that small changes in the incentives provides by prices are apparently associated with locally very large changes in behaviour. This is a fact, to a greater or lesser extent, with every driver in the sample (see Figure 4.2 in the Appendix). Such locally large changes in hours worked as a consequence of small variations in wages cannot, in these data and for the reasons discussed above, be simply ascribed to measurement error. Nonetheless they seem barely credible economically or behaviourally.

3.3. **Revealed Preference Analysis.** The behaviour revealed by these estimates, and indeed the raw data themselves, has prompted the question (Farber 2005, 2008 and V. Crawford and Meng 2011) of whether these drivers behave as rational economic agents, in which case their patterns of behaviour seen can be explained by income and substitution effects alone, or whether a more elaborate model such as reference-dependence is required.

To shed light on this, Table 2 reports the results of driver-by-driver revealed preference tests. Only one of the drivers (Driver 2) displays behaviour which is fully consistent with utility maximisation; the rest violate WARP to some degree. To answer the question of quite how badly they violate WARP, the table also reports the Afriat Efficiency index for each driver. The efficiency measures all indicate that, whilst they are not perfectly theory-consistent, the drivers only waste around 1% to $1\frac{1}{2}$% of their budgets (around \$1 to \$2 per shift) when they fail to optimise perfectly. In other words the departures from full rationality are, on this evidence, economically immaterial. One natural conclusion is that these drivers are in fact very close to being as-if neo-classical. Nonetheless, the labour supply functions which fit these data look anything but rational in the more everyday sense of the word.

TABLE 2. Revealed Preference Results, by driver

Driver	WARP	Afriat Efficiency (Æ)
1	Fail	0.9953
2	Pass	1
3	Fail	0.9978
4	Fail	0.9893
5	Fail	0.9977
6	Fail	0.9834
7	Fail	0.9956
8	Fail	0.9909
9	Fail	0.9899
10	Fail	0.9930
11	Fail	0.9830
12	Fail	0.9859
13	Fail	0.9926
14	Fail	0.9826
15	Fail	0.9844

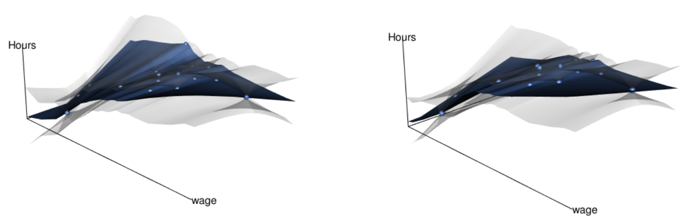
Table 2 shows that Driver 2 is unique insofar as he is observed to behave fully rationally. His labour supply function, shown in Panel (B) in Figure 3.1, is one of the best-behaved in the sense that it is relatively smooth. It has been argued that rationality yields well-behaved estimates of the demand function and price responsiveness[16], so it may be the case that the roughness in labour supply responses displayed by other drivers is a consequence of their non-rational, albeit only slightly inefficient, behaviour.

To see whether failure of rationality is the problem and whether imposing rationality can improve the labour supply functions recovered from the data, consider, as an illustration, Driver 10. He fails WARP but his Afriat Efficiency score is 0.993. Conventionally this is taken to indicate that he is close-to-rational (see for example Choi,

[16]Blundell, Horowitz and Parey (2012)

Kariv, Mueller and Silverman (2014)). An alternative measure of this closeness-to-rationality is the proportion of his observed behaviour which is WARP-consistent (his Houtman-Maks index) which is 0.8823. In fact his WARP violation is caused by just two observations. Let us assume that he had a bad couple of shifts, remove these as aberrations and re-estimate, his labour supply function using only his WARP-consistent behaviour. The results are shown in Figure 3.2. The left-hand panel shows his initial labour supply curve estimated using all of his data. The right-hand panel shows the labour supply function using only his WARP-consistent behaviour. Evidently removing violations improves the aesthetics and plausibility of the function somewhat. It also, it should be noted, improves the economics: it produces a consumption-leisure demand system which is globally rational: all of the points evaluated on the function are consistent which having been generated by the same utility function as the rational observations from Driver 10 himself.

FIGURE 3.2. Labour supply curves for Driver 10

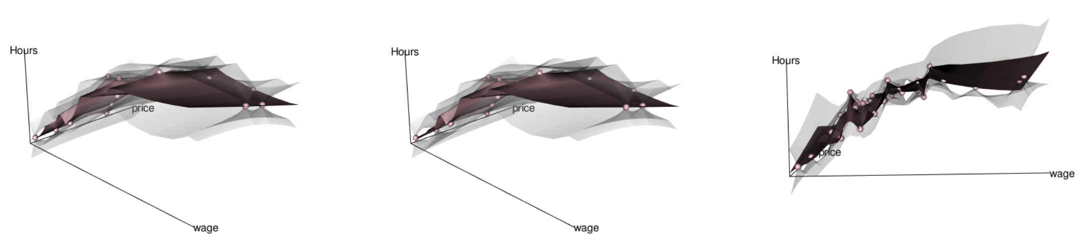


(A) Driver 10

(B) Driver 10 - WARP consistent

But now consider Driver 3. Driver 3 is observed 40 times. He too fails WARP and is therefore not perfectly neo-classical either. His Afriat Efficiency score is even better than Driver 10 at 0.9978, as is his Houtman-Maks index of 0.975. In fact his WARP violation is caused by a single observation. Removing this one observation and re-estimating subject to the constraint that surface satisfies WARP results in the labour supply function shown in 3.3:

FIGURE 3.3. Labour supply curves for Driver 3



(A) Driver 3

(B) Driver 3 - WARP-consistent, view #1

(C) Driver 3 - WARP-consistent, view #2

Panel (A) shows his labour supply using all of his observed behaviour. The middle (B) and right-hand (C) panels show two views of his labour supply estimated using only his rational behaviour. As might be expected, the removal of a single observation does not radically affect the estimates and it is hard to see much of a difference. Nonetheless, there is some improvement in the function - it is a little smoother - and this function (unlike the panel

(A)) is also globally rational and WARP-consistent with the data used. However, panel (C) which gives a side-on view shows that rationality alone does not resolve the issue; there are still regions in which small changes in wages are associated with implausibly abrupt changes in hours worked. Revealed preference analysis has no problem 'rationalising' demand elasticities which vary locally from close-to-zero to close to infinite (see Figure 2.3 above). Thus it seems that rationality is not enough to give us estimates which anyone can be confident in. This due to a hard-to-shake intuition that (i) if the incentives provided by relative prices only change a little, then behaviour should only respond a little, and (ii) that this should essentially be the case everywhere in price-space.

This may point to a general problem with the way in which revealed-preference methods are applied: often if a dataset satisfies the revealed preference characterisation of the model of interest it is tempting to (albeit heuristically) take the model to be true and then to engage in applied work on the basis of that model. But these estimates should give us pause: a failure to reject the model does not mean that the model is correct and the labour supply functions we estimate, which exactly correspond to the revealed preference data, and shows evidence of behaviour which, whilst model-consistent, is not economically sensible. Secondly, imposing rationality in the form of revealed preference conditions on the estimated demand curves, whilst it will result in forecasts which are globally rational, may not (and the evidence in this case is, *do not*) result in labour supply behaviour in which one can tenably believe. Essentially the problem may be that consumer theory, as described by the revealed preference machinery, requires behaviour to be explainable through income and substitution effects. Whilst this explanation may be possible, it can appear rather forced.

3.4. **Conditional Preferences and Conditional Demand.** If these drivers are close-to-rational, and imposing rationality itself does not result in a credible model fo consumption-leisure chocies, then something else must be going on. In other words the decision-making environment is richer than simply a changing budget; something else (denoted z in what follows) may be changing too and affecting preferences and choices. This factor (or factors) could be operating either through the constraints (e.g. some additional limitation on choices beyond the budget constraint) or through preferences. In what follow it will be assumed that the channel is via preferences and that it augments the individual's preferences as a taste shifter: $u(\mathbf{q}, z)$.

Economic theory provides little guidance on how conditioning variables should enter preferences other than they should be non-separable in order to have any observable behavioural implications at all. In terms of revealed preference conditions the only requirement is that the data should be "conditionally rational".

> **Definition**. A dataset $\{\mathbf{p}_i, \mathbf{q}_i, \mathbf{z}_i\}_{i=1,\ldots,N}$ is *conditionally rational* if $\{\mathbf{p}_j, \mathbf{q}_j | \mathbf{z}_j = \mathbf{z}\}_{j\in\{1,\ldots,N\}}$ satisfies GARP for all $\mathbf{z} \in \{\mathbf{z}_i\}_{i=1,\ldots,N}$.

To make progress this paper takes the idea that "a good theory will successfully predict events that are difficult to predict" (Binmore and Shaked (2010)) and adopts a loss function which is a measure of predictive success proposed by Selten (1991).[17] The essential idea focusses on a comparison between the observed choices made by an individual and the set of choices admissible under the theory model of interest. In the present case our model of interest is utility maximisation under conditional rationality. Selten (1991) observes that if the set of choices consistent with a model is very large relative to the set of choices which the agent could possibly display then one should be little surprised if much of observed behaviour lies in the theory-consistent set − it could hardly have done otherwise. For example, if we are testing the utility maximising model and the collection of budget constraints on which the consumer is observed to chose never cross then all feasible choices, necessarily, satisfy GARP: it would be impossible to make an observation which wasn't theory consistent. Such data should not be taken as a strong validation of the theory. This suggests that empirical fit alone is not a sufficient basis for ranking the performance of alternative theories: if it were then no theory could out-perform a meaningless, anything-goes, theory which can rationalise any behaviour. A better approach would be to consider the trade-off between some measure of fit and some sort of measure of how demanding the theory is. Following Selten (1991) let a denote the size of the theory-consistent subset, *relative* to the outcome space for the model of interest; this is called the *area* in Selten's terminology. In the case of utility-maximisation the outcome space is the set of all possible choice-ensembles which satisfy the budget constraints and the theory-consistent subsets are those ensembles which satisfy the RP condition. The area is measured by calculating the proportion of feasible choices which are also theory-consistent. The relative area of the empty set is zero and the relative area of all outcomes is one so $a \in [0, 1]$. Selten (1991) argues that demanding theories are characterised by small values for a. Areas close to zero indicate that choice is highly constrained by the theory (few of the feasible choices would be admissible under the model). Areas close to one indicate the reverse (anything goes).

In what follows, a number of potential conditioning variables are considered. For each one, the theoretical restriction is conditional rationality. Thus the procedure is simply to partition the data on the basis of the conditioning

[17]See also Selten and Krischker (1983) and, perhaps, Beatty and Crawford (2011).

variable and test that WARP holds for all of the resulting subsets of the data. We calculate a, the size of the theoretically admissible set, for all values of z by numerical integration (see Beatty and Crawford (2011) for details). The obvious point here is that if the conditioning variable takes on many values then there will be many subsets within which the RP condition must hold, but these subsets will be small. Consequently the theoretically-consistent area within each will be tend to be large. At the extreme, where every observation is associated with a unique value of z, then it will be impossible to violate the conditional rationality requirement and hence $a = 1$; the conditioning variable will make the theory completely undemanding. Selten's criteria will penalise such conditioning variables as they will fail the 'a good theory predicts things which are hard to predict' maxim. This therefore more-or-less rules ou,t at the outset, using an unobserved, continuously varying, taste shifter to rationalise behaviour. Instead we focus on observables.

Table 3. Conditioning variables

	Conditioning Variable	Coded as	Used for Driver
1.	Hours worked on the last shift	Individual-specific quintiles	1,7
2.	Earnings on the last shift	Individual-specific quintiles	-
3.	Hourly wage on the last shift	Individual-specific quintiles	13
4.	Day/night shift	Dummy	-
5.	Weekend/Weekday	Dummy	-
6.	School night/Friday or Saturday night	Dummy	-
7.	Raining or Snowing/Dry	Dummy	5,9
8.	Dry/Rain/Snow	Banded	3,4
9.	Winter/Spring-Autumn/Winter	Banded	-
10.	Temperature	Banded: $<$32F, 32F-50F,50F-70F,70F-90F, $>$90F	8,10
11.	Temperature	Degrees Fahrenheit	6,11,14
12.	Leave-me-out expected earnings	$	12,15
13.	None	-	2

A range of possible conditioning variables are investigated. These are listed in columns 1 and 2 of Table 3. A number of them are inspired by the literature on reference-dependence. Reference-dependent models have played an important role in empirical analysis of workers', consumers', and investors' choice behavior since Camerer *et al* (1997). Camerer *et al* found a negative elasticity of hours-worked with respect to earnings per hour and proposed an informal model in the spirit of prospect theory, in which drivers have daily income targets that play the role of Kahneman and Tversky's (1979) reference points. More recent work by Farber (2005, 2008) using these data allowed reference-dependence in the form of income-targeting. Farber found that a reference-dependent model, with a daily income target treated as a latent variable, fits better than a reference-independent model; but he also concluded that estimates of latent targets were too unstable to yield a useful reference-dependent model. V. Crawford and Meng (2011) reconsidered Farber's data in light of Kőszegi and Rabin's (2006) model of reference-dependence in which an agent's reference point for each good is equal to its theoretical rational expectation.[18] V. Crawford and Meng (2011) took a driver to have a daily target for hours as well as income, and used natural sample proxies to approximate rational-expectations targets instead of treating the targets as latent variables.

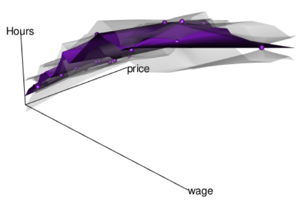
Given the central importance of reference-dependence in explaining the behaviour of these individuals, a number of conditioning variables have been selected in the spirit of that literature, albeit without fully embracing the behavioural model. These include a number of models in which an individual driver's behaviour depends on the outcome realised on their last shift (variously their earnings, hours worked and the realised hourly wage). As well as being motivated by reference-dependence (where the reference point is the last outcome) these conditioning variables can also be thought of in terms of short-memory habits/intertemporal non-separability. In some cases these have been coded by individual-specific quintile: the last shift may have resulted in, for example, earnings which were: very poor, below par, normal, quite good, or very good. This coding is for two reasons. The first is that it seems highly plausible that drivers really think this way – that the don't resolve their assessment of the last shift to the precision of the dollar, but rather bin it in these somewhat rougher terms. The other reason relates to the Selten

[18]The feedback loop between preferences and rationally expected decisions creates a game between a person and his own future self, which game can have multiple equilibria. These equilibria are ranked ex ante, and Kőszegi and Rabin (2006, p. 1135) use a refinement, "preferred personal equilibrium," to focus on the one that is best for the person.

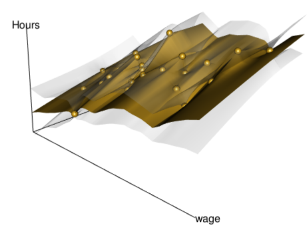
index for conditional rationality which is the criteria for model selection. What governs the predictive strength of conditional revealed preference restrictions is really no different to standard revealed preference: the number of observations at each level of the conditioning variable is important. Thus a relative coarse binning of the data is more likely to produce a meaningful (in the Selten sense) model. As a check on this, and also in order to allow for reference-dependent-like behaviour which is even closer to the literature and in particular the Kőszegi and Rabin's rational expectations model of reference-dependence, drivers' behaviour is also conditioned on the natural sample proxy used by V. Crawford and Meng (2011): leave-me-out expected earnings. The reasons for the selection of the other conditioning variables relate to environmental considerations which will affect the pleasantness, or otherwise, of working (e.g. weather-related variables) or the opportunity costs (whether it is the weekend, for example, when time spent working might mean time not spent with family or friends), or both.

The individual-level modelling approach adopted in this paper lends itself to bespoke models rather than one size (approximately) fits all and so the Selten index is calculated for each of these potential conditioning variables and the best variable for each driver is selected. The drivers for whom each conditioning variable provides the best model of the data are listed in the third column of Table 3. Conditioning variables inspired by reference-dependence are successful with one third of the drivers whilst environmental variables account for the rest. Interestingly day/night shift or weekend effects etc. do not seem to be the best explanation for any of the drivers. The general exception is Driver 2, the driver who who satisfied WARP. This driver does not require any further conditioning and the Selten-index selects the unconditional demand for him.

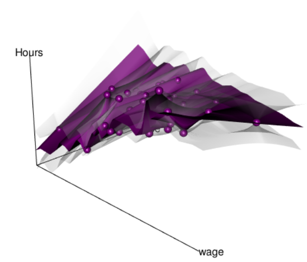
Figure 3.4. Conditional and unconditional labour supply functions



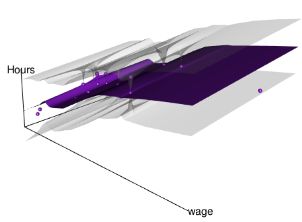
(a) Driver 8. Unconditional labour supply



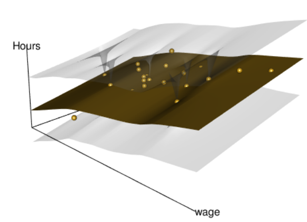
(b) Driver 13. Unconditional labour supply.



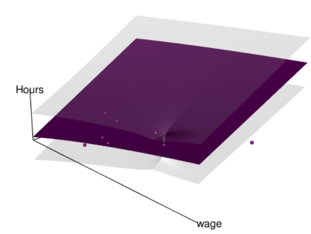
(c) Driver 14. Unconditional labour supply



(d) Driver 8. Conditional labour supply



(e) Driver 13. Conditional labour supply.

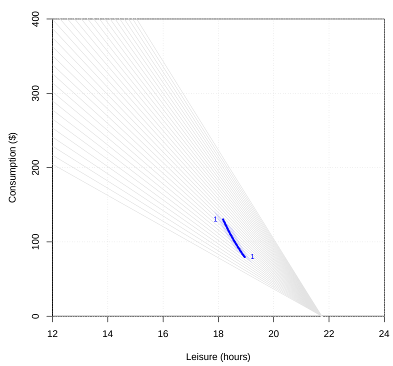


(f) Driver 14. Conditional labour supply

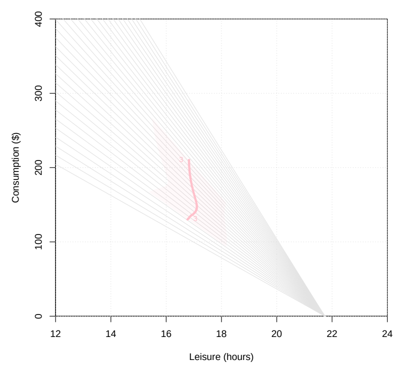
Using these conditioning variables the conditional demand systems are re-estimated for each driver. The full set of labour supply results are shown in Figure 4.3 in the Appendix. Figure 3.4 shows the results for three of the drivers. Panels (A), (B) and (C) re-present their unconditional labour suply funcitons (for comparison) and panels (D) , (E) and (F) show the conditional curves. Note that the conditional labour supply functions now vary with respect to three variables: normalised wages, the normalised price of consumption and the conditioning variable. They are illustrated in normalised price space for a fixed level of the conditioning variable (the middle value for the banded variables, e.g. conditional on a typical day for earnings in the last shift, and the modal value for the rest).[19] The effects of conditioning are apparent. In each case the resulting labour supply curves are much better behaved in two senses: firstly they all fully satisfy conditional revealed preference restrictions, and secondly they are much more plausible than the unconditional surfaces because they generally do not exhibit locally unstable behaviour. The confidence intervals are however wider - this is due the relative sparseness of data in four dimensions relative to three dimensions - the intuition for this is similar to the curse of dimensionality in kernel-based multivariate non-parametric regression analysis.

3.5. **Positive Analysis: Individual Wage-Offer Curves.** Using these estimates it is straightforward to construct traditional wage-offer curves for each individual and track them as the relative prices change. This places all of the individual drivers on the same, counterfactual budget constraints and allows for easy comparison.

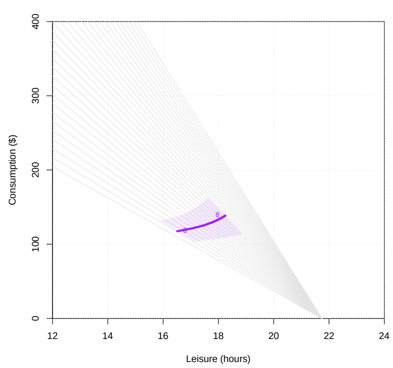
Figure 3.5. Wage Offer Curves



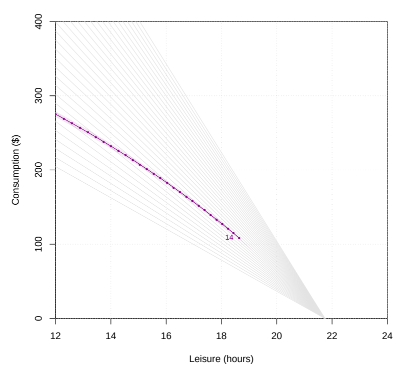
(A) Driver 1



(B) Driver 3



(C) Driver 8



(D) Driver 14

Figure 3.5 illustrates the wage offer curves for four drivers. The full set are available for inspection in Figure 4.4 in the Appendix, but these four individuals show the main kinds of behaviour on display quite clearly. Note that the counterfactual budget constraints on which they are predicted are identical but that the wage offer curves for different individuals are of different lengths: the demands for each individual have only been estimated within the convex hull of the range of price variation observed in the data for that individual to avoid extrapolation.

[19]These can, of course, also be illustrate for any value of the conditional variable and these are available from the author.

Panel (A) shows an individual (Driver 1) whose demand for leisure fall as the wage rises relative to the price of consumption and whose consumption rises. In other words this driver substitutes away from leisure to consumption and is induced to work longer hours as the return increases. Panel (B) shows an individual (Driver 3) whose hours of work are essentially fixed with respect to the wage. Figure (C) shows an individual (Driver 8) whose demand for leisure increases as wages rise and for whom consumption is complimentary to leisure. The final figure, panel (D), also exhibits an increases demand for leisure (reduced labour supply) in response to a wage increase, but for this driver consumption and leisure are gross substitutes.

3.6. **Quantile invariance and individual identity.** Figure 3.6 places all of the wage offer curves onto the same set of budget constraints. The result is an unclear muddle designed to make a clear point: there is *significant* re-ranking of the individuals along the budget constraint as the wage changes. The wage-offer curves cross frequently and individuals who, at low wages, might be characterised as having a relatively strong taste for consumption and low taste for leisure on the basis of their ranking along the budget constraint, alter their relative position as wages rise. This would seem to constitute evidence that an individual's position along the budget constraint is not necessarily a good way of characterising preference heterogeneity; it is, *prima facie*, better thought of as principally being driven by choice because it changes with the choice set.

FIGURE 3.6. Wage offer curves: all drivers

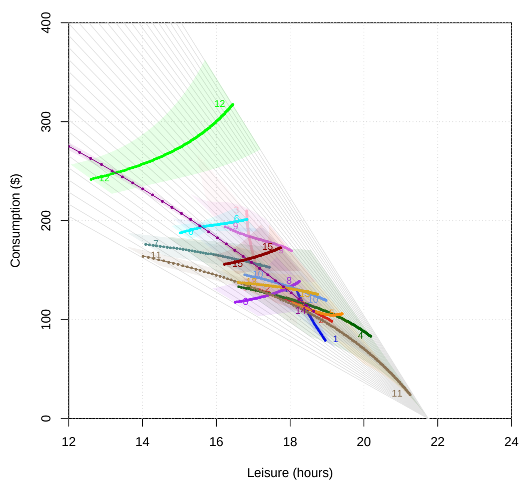
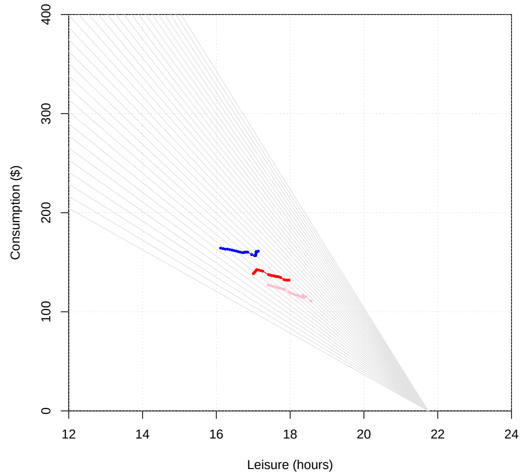


Figure 3.7 shows the wage-offer curves corresponding to the 33rd, 50th and 66th percentiles of the distribution on each budget constraint - effectively, since in this case we do have a fixed population of drivers and we know their identities, these are the 5th and 8th and the 11th ranked driver (in terms of either consumption or leisure) on each budget constraint. These quantile demands are recovered over a somewhat restricted range of budget constraints compared to Figure 3.6 (roughly speaking the middle of the variation show in that figure). This is because this is the range of variation for which we can recover quantiles for a fixed set of individuals (recall we do not extrapolate the wage-offer curve for any driver outside of the range where they have been previously observed therefore the set of drivers on each budget constraint can change as the budget varies taking individuals out of their observed range).

Even though the wage-offer curves of the individual drivers are highly heterogeneous, somewhat non-linear and cross frequently, the quantile wage-offer curves are *necessarily* much more smooth/regular. This is because quantile curves *cannot* cross even though individuals' curves can. In order to avoid this re-ranking the identity of the driver simply switches. One implication of this is that, if researchers take quantile demands to be indicative of (or indeed the same as) the demands of an individual, then they gain a misleading impression of individual behaviour as being simpler/smoother than perhaps it really is.

FIGURE 3.7. Quantile wage-offer curves



3.7. **Welfare Analysis: Cost-of-living indices.** Welfare analysis can be carried out using Marshallian demands by following the methods outlined in Vartia (1983).[20] Holding utility constant implies the differential equation:

$$\frac{dx(t)}{dt} = \sum Q_k(\boldsymbol{p}(t), x(t))\frac{dp_k(t)}{dt}$$

where $t \in [0,1]$ and $\mathbf{p}(t)$ are linear paths for normalised prices and $x(t)$ denotes the value of the expenditure function/compensated income along this path. Integrating over the path gives

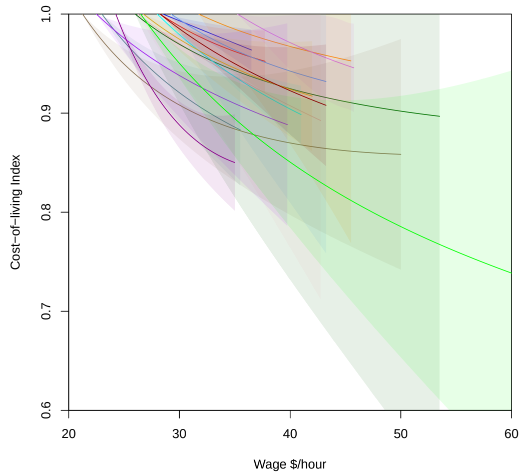
$$x(t) - x(0) = \sum_k \int Q_k(\boldsymbol{p}(t), x(t))\frac{dp_k(t)}{dt}dt$$

Along the price path the compensating income follows a differential equation in which the arguments are demands and prices. Vartia's algorithm linearly approximates this over a sequence of discrete intervals between the initial and final prices and allows the level of compensating income to change at each intermediate point. Vartia (1983) shows that the overall welfare effect is well approximated by summing over these small intermediate changes. The methods described in this paper, being essentially point-wise, fit naturally into Vartia's approach without modification.

Figure 3.8 shows the the cost-of-living effects, calculated using Vartia's method, for each driver along each wage-offer curve. As wages rise the cost-of-living falls. Each series is based at the starting point of that individual's wage-offer curve. As a result they start at different point because any given wage rate may lie outside of the range observed for an individual driver. Since these paths are generated by wage offer curves which exhibit significant re-ranking these welfare effects also reflect re-ranking: the driver who is made least well off by a $2 increase in the wage may do (relatively) better when the wage is raised by $4 for example.

[20] Balk (1995) notes that the idea was also contained in a paper by Malmqvist (1993) which was originally drafted in the 1950's. See also Hausman and Newey (1995).

FIGURE 3.8. Cost-of-living indices, by driver as a function of wages



4. Conclusion & Discussion

This paper describes a fruitful way to bridge the gap which exists at present between statistical demand models and deterministic, classical revealed preference theory. The key idea is that local variations around the global shape which describes an individual's general behaviour can be treated *as if* they were the realisation of a random field which is intrinsically stationary. The interdependence of the field values at different locations can be used to improve the prediction variance as per Goldberger (1962). The result is a statistical interpolant which is also a best linear unbiased predictor. In simulated examples the method is shown to work well in parts of the region of interest in which observations are closely spaced and less well when the data are sparse. The relationship to revealed-preference bounds is also illustrated.

In an application to NYC cab-drivers, individual-level labour supply curves, wage-offer curves and welfare measures are recovered. These show that neither rational, nor close-to-rational behaviour guarantee economically plausible price responses. However, conditioning on observed taste shifters can yield a parsimonious and still theory-consistent account of the data. This paper suggests a methods for doing this which gives an important role to economic theory based on Selten's measure of predictive success. The labour supply functions which emerge paint a rich picture of heterogeneous behaviours which exhibit a substantial amount of reranking of individuals along budget constraints. As rankings change with the choice set it is argued that characterising individual-level preference heterogeneity by choice quantiles may not successfully link quantile behaviour to underlying individual behaviour.

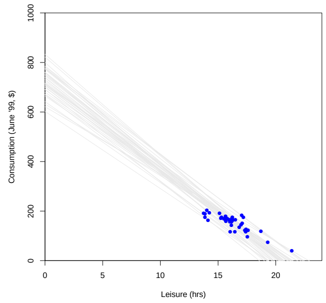
Because it does not rely on error terms in the normal way, the approach developed here is a natural econometric counterpart to classical revealed preference methods based on individual-level data, but it enjoys three main advantages compared to the revealed preference approach. Firstly it provides point, rather than set-valued, predictions of demand curves and of welfare measures. Secondly it allows the researcher to introduce conditioning variables rather than forcing everything to work via the mechanisms of price and income effects. Thirdly, it provides a measure of uncertainty regarding predictions at un-sampled prices. Thus it retains the elegance of revealed preference theory but overcomes the principal drawbacks encountered in applications.

A number of avenues remain to be explored. Classical measurement error can be incorporated straightforwardly[21] but non-classical forms are a more important, and currently neglected, issue in modern consumption panels. The semi-variogram model plays an important role and this paper uses a simple linear model based on Euclidean distance which treats the own- and the cross-price directions symmetrically. It would be useful to consider a model general model which allows the dependence to vary by direction. It may be interesting to compare the results to non-BLUP alternatives. Finally, alternatives to Selten's criteria might also be usefully considered e.g. the Kullback-Leibler divergence provides different penalisation and also has a natural interpretation in terms of information-losses associated with conditioning.

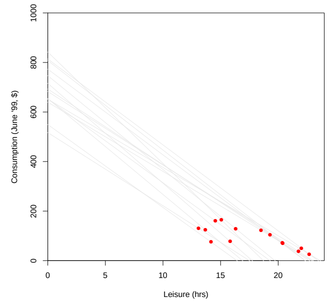
[21]See Cressie (1993), Stein (1999).

Appendix

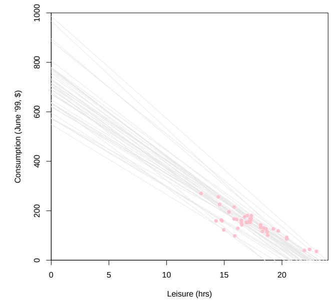
Figure 4.1. NYC Taxi driver data: by driver



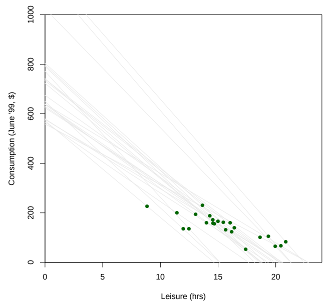
(a) Driver 1



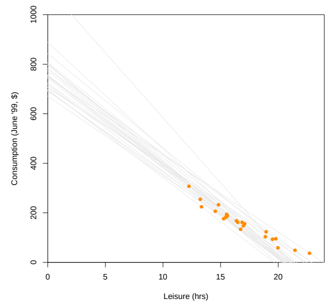
(b) Driver 2



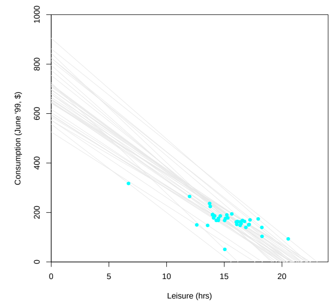
(c) Driver 3



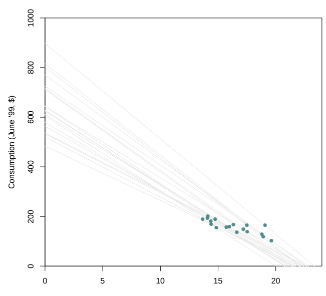
(d) Driver 4



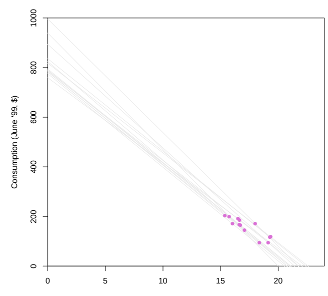
(e) Driver 5



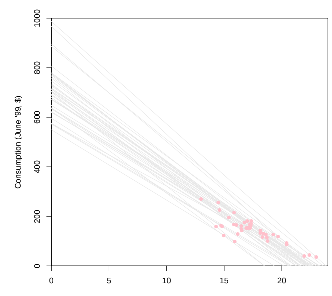
(f) Driver 6



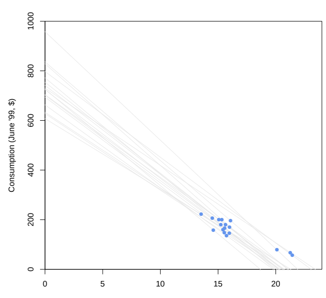
(g) Driver 7



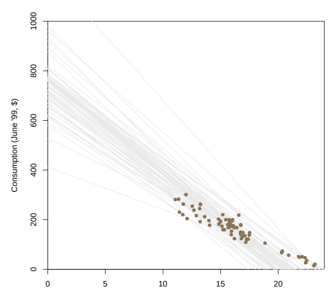
(h) Driver 8



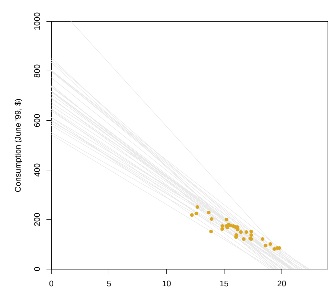
(i) Driver 9



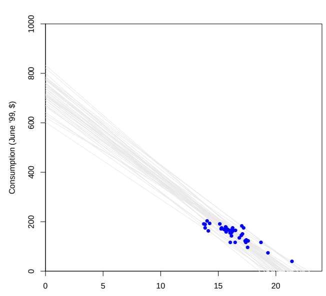
(j) Driver 10



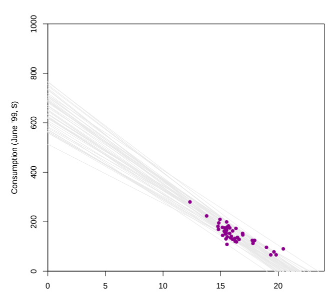
(k) Driver 11



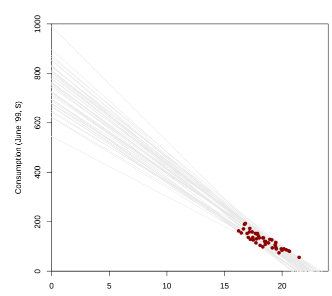
(l) Driver 12



(m) Driver 13

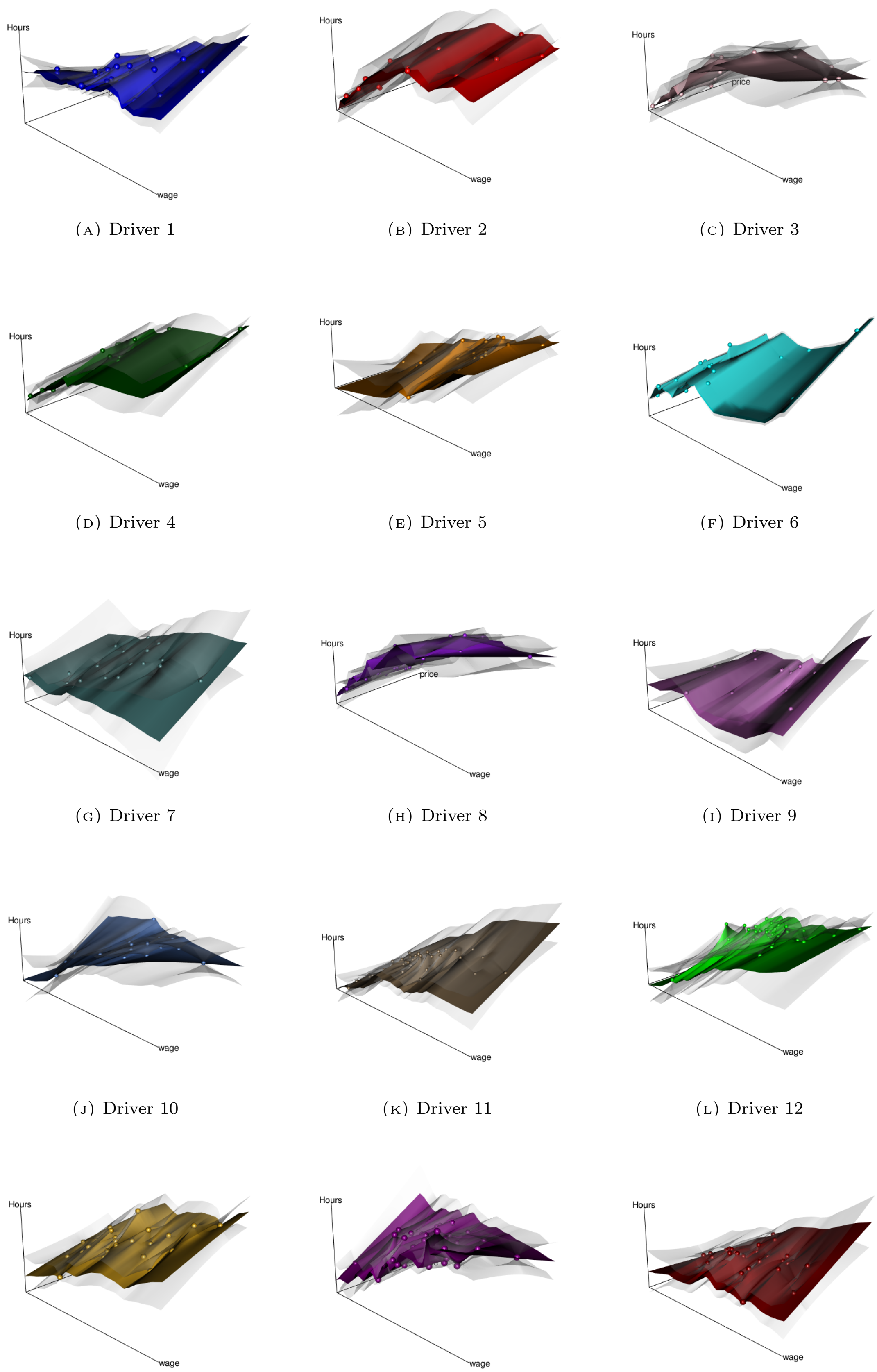


(n) Driver 14



(o) Driver 15

Figure 4.2. Labour supply curves by driver



(A) Driver 1

(B) Driver 2

(C) Driver 3

(D) Driver 4

(E) Driver 5

(F) Driver 6

(G) Driver 7

(H) Driver 8

(I) Driver 9

(J) Driver 10

(K) Driver 11

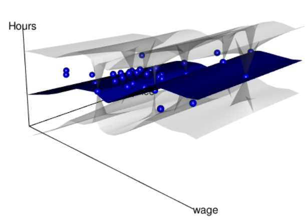
(L) Driver 12

(M) Driver 13

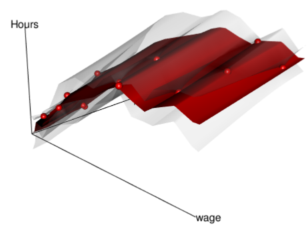
(N) Driver 14

(O) Driver 15

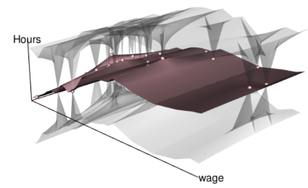
FIGURE 4.3. Conditional Labour supply curves by driver



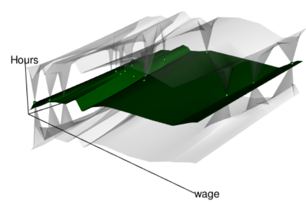
(A) Driver 1



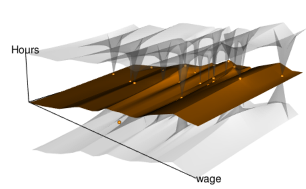
(B) Driver 2



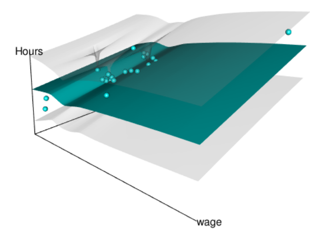
(C) Driver 3



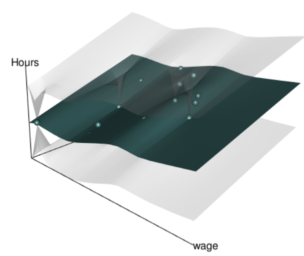
(D) Driver 4



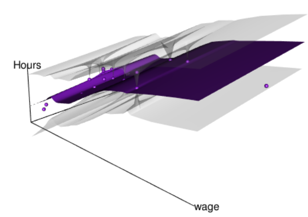
(E) Driver 5



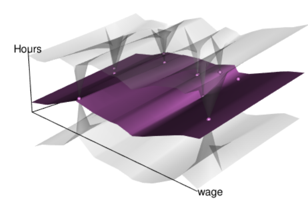
(F) Driver 6



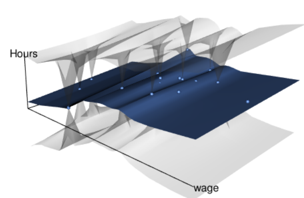
(G) Driver 7



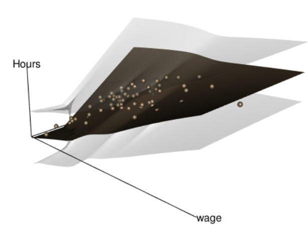
(H) Driver 8



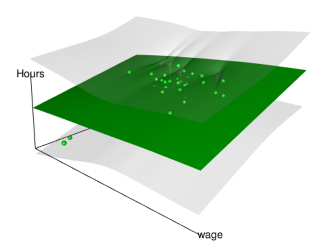
(I) Driver 9



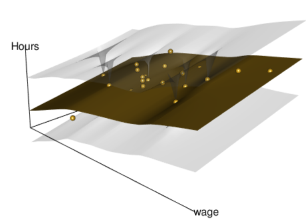
(J) Driver 10



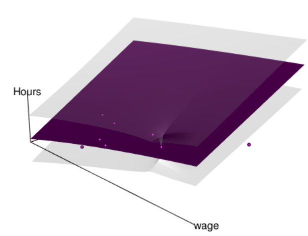
(K) Driver 11



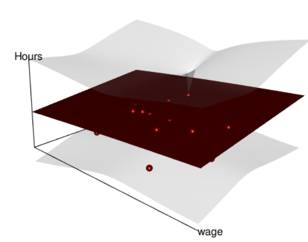
(L) Driver 12



(M) Driver 13

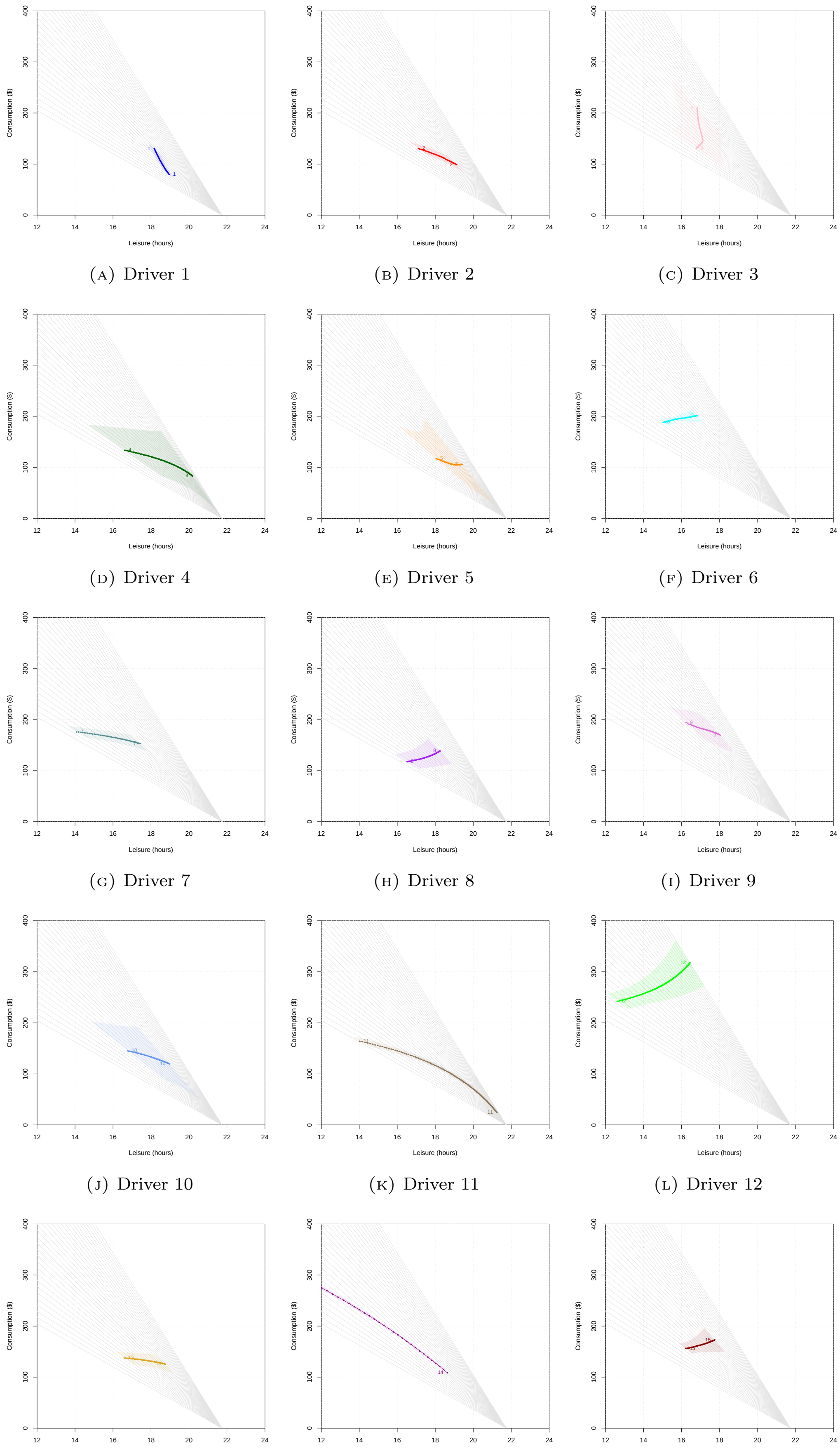


(N) Driver 14



(O) Driver 15

Figure 4.4. Wage Offer Curves, by driver



(a) Driver 1 (b) Driver 2 (c) Driver 3

(d) Driver 4 (e) Driver 5 (f) Driver 6

(g) Driver 7 (h) Driver 8 (i) Driver 9

(j) Driver 10 (k) Driver 11 (l) Driver 12

(m) Driver 13 (n) Driver 14 (o) Driver 15

References

[1] Afriat, S.N. (1967),"The construction of a utility function from expenditure data", *International Economic Review,* 8, 76-77.

[2] Balk,. B. M. (1995), "Approximating a Cost-of-Living Index from Demand Functions: A Retrospect", *Economics Letters*, 49, 147-155.

[3] Banks, J., R. Blundell and A. Lewbel, (1997), "Quadratic Engel Curves and Consumer Demand", *Review of Economics and Statistics*, 79 (4), 527-539

[4] Beatty, T. and Ian Crawford. 2011. "How Demanding Is the Revealed Preference Approach to Demand?" *American Economic Review*, 101 (6), 2782-95.

[5] Blundell, R., M. Browning and I. Crawford (2003), "Nonparametric Engel curves and revealed preference", *Econometrica*, 71, 205-240.

[6] Blundell, R., M. Browning and I. Crawford (2008), "Best nonparametric bounds on demand responses", *Econometrica*, 76, 1227-1262.

[7] Blundell, R., J. Horowitz and M. Parey (2017): "Nonparametric Estimation of a Heterogeneous Demand Function under the Slutsky Inequality Restriction," *Review of Economics and Statistics*, 99(2), 291-304.

[8] Blundell, R., D. Kristensen, and R.L. Matzkin (2014) "Bounding Quantile Demand Functions using Revealed Preference Inequalities", *Journal of Econometrics* 179, 112-127.

[9] Blundell, R., D. Kristensen, and R.L. Matzkin (2014) "Individual counterfactuals with multidimensional unobserved heterogeneity", cemmap working paper, CWP60/17.

[10] Brown, B. W., and M. B. Walker (1989), "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica* 57, 815– 829.

[11] Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 112, 407-441.

[12] Choi, S., S. Kariv, W. Müller and D. Silverman (2014), "Who Is (More) Rational?" *American Economic Review*, 104(6), 1518-1550

[13] Crawford, V.P., and JJ. Meng (2011): "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–1932.

[14] Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons, Inc.

[15] Deaton, A. S., and J. Muellbauer, "An Almost Ideal Demand System," *American Economic Review* 70 (1980), 312-336.

[16] Diewert, W.E. (1973), "Afriat and revealed preference theory", *Review of Economic Studies*, 40, 419-426.

[17] Farber, Henry S. (2005): "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113(1), 46–82.

[18] Farber, Henry S. (2008): "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 98(2), 1069-1082.

[19] Hausman, J. A. and W. K. Newey (1995), "Nonparametric estimation of exact consumers surplus and deadweight loss." *Econometrica*, 63 (6), 1445–1476.

[20] Hausman, J.A. and W.K. Newey (2016) "Individual Heterogeneity and Average Welfare", *Econometrica* 84, 1225–1248

[21] Hoderlein, S., and J. Stoye (2014) "Revealed Preferences in a Heterogeneous Population", *Review of Economics and Statistics* 96, 197-213.

[22] Kőszegi, B. and M. Rabin (2006) "A Model of Reference-Dependent Preferences", *Quarterly Journal of Economics*, 121, 4, 1133–1165.

[23] Kitamura, Y. and J. Stoye (2018) "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86: 1883-1909.

[24] Lewbel, A. (2001), "Demand systems with and without errors", *American Economic Review*, 91, 611-618.

[25] Lewbel, Arthur, and Krishna Pendakur. 2009. 'Tricks with Hicks: The EASI Demand System.' *American Economic Review*, 99, 827-63.

[26] Malmqvist, S. (1993), "Index Numbers and Demand Functions", *The Jounral of Productivity Analysis*, 4, 251-260.

[27] Matzkin, R.L. (2003), "Nonparametric estimation of nonadditive random functions" *Econometrica*, 71, 1339-1375

[28] Matzkin, R.L. (2008) "Identification in nonparametric simultaneous equations" *Econometrica*, 76, 945-978.

[29] McFadden, D.L. and M.K. Richter (1971) "On the Extension of a Set Function on a Setof Events to a Probability on the Generated Booleanσ-Algebra," Tech. rep., Universityof California, Berkeley.

[30] McFadden, D.L. (2005) "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26: 245-264.

[31] Scheuerer, M., R. Schaback and M. Schlather (2013) 'Interpolation of spatial data – A stochastic or a deterministic problem?', *European Journal of Applied Mathematics*, Volume 24, Issue 4 August 2013 , pp. 601-629.

[32] Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York

[33] Varian, H. (1982), "The nonparametric approach to demand analysis", *Econometrica*, 50, 945-973

[34] Varian, H. (1985), "Non-parametric analysis of optimizing behavior with measurement error", *Journal of Econometrics* 30, 445-458

[35] Vartia, Y. O. (1983), "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions", *Econometrica*, 51, 79-98

[36] Varoufakis, Y. (2012) "A Most Peculiar Failure. On the dynamic mechanism by which the inescapable theoretical failures of neoclassical economics reinforce its dominance" UADPhilEcon, Working paper, Department of Economics, University of Athens.