Linear Estimation of Structural and Causal Effects for Nonseparable Panel Data

Whitney K. Newey

MIT and NBER

Fisher-Schultz Lecture of the Econometric Society
European Meeting of the Econometric Society

August 31, 2023

This paper builds on and supersedes Chernozhukov, Hausman, and Newey (2019), "Demand Analysis with Many Prices," NBER working paper 26424.

The current paper is coauthored with V. Chernozhukov, B. Deaner, Y. Gao, and J.A. Hausman.

INTRODUCTION

Models that are nonseparable (i.e. not additively separable) in observables and unobservables are important for economics; see Matzkin (2023) and references therein.

The observed variables are quantities of economic interest such as prices.

Unobserved variables represent preferences and/or technology.

An example is demand, where the share of some good is

$$S = s(p_1, ..., p_K, y, \eta_1, \eta_2, ...),$$

where p_k are prices of goods, y is total expenditure, and $\eta_1, \eta_2, ...$ are possibly infinite number of unobserved taste variables, e.g. Lewbel (2001).

$$S = s(p_1, ..., p_K, y, \eta_1, \eta_2, ...),$$

Equivalent to stochastic revealed preference models of McFadden and Richter (1971), McFadden (2005), Kitamura and Stoye (2018), when choice is single valued.

Nonseparable models are also equivalent to treatment effect models where $\eta' s$ can represent potential outcomes.

In economic models it seems important to allow price effects, income effects, and other effects to vary over individuals in general ways, motivating interest in models that are not additively separable and have general, multi-dimensional heterogeneity.

$$S = s(p_1, ..., p_K, y, \eta_1, \eta_2, ...).$$

Endogeneity, where there is correlation between observables and unobservables, is often a problem.

Observables may be choice variables that are depend on preferences and/or technology.

Observables may be determined by an equilibrium condition that depends on preferences and/or technology.

Panel data can be used to identify and estimate counterfactual effects when observables and unobservables are correlated.

A fundamental identifying assumption for panel data is that the conditional distribution of unobserved heterogeneity in each time period, conditional on regressors in all time periods, does depend on the time period; see Chamberlain (1982, 1992), Manski (1987), Honore (1992), Pesaran and Smith (1995), Abrevaya (2000), Chernozhukov, Fernández-Val, Hahn, and Newey (2009, 2013), Graham and Powell (2012), Pakes and Porter (2016), Shi, Shum, and Song (2018).

This assumption enables identification of counterfactual effects from variation in observables over time.

In a linear model the linear projection version of this assumption is equivalent to existence of an additive fixed effect and regressors being uncorrelated with the idiosyncratic at all leads and lags, see Chenozhkov et al. (2013).

In nonseparable models this condition is different than the correlated random effects assumptions of Chamberlain (1980), Wooldridge (2002,2005), Altonji and Matzkin (2005), Arellano and Bonhomme (2012).

This condition is motivated by time stationarity of preferences and/or technology.

Some time effects can be allowed through observables.

Some time stationarity is important for using variation in observables over time to help identify counterfactual effects.

In this paper we use linear regression methods to estimate counterfactual effects for nonseparable models under the time stationarity condition.

We approximate a smooth nonseparable model by a linear model with coefficients that depend only on the heterogeneity and so inherit the time stationarity property, giving regression coefficients that do not vary over time.

We estimate each individual's coefficients by individual ridge regression to regularize for possible singularity (i.e. nonidentification) or near singularity of the second moment matrix of individual specific regressors.

We estimate average effects by debiasing the average linear combinations of individual ridge regression coefficients.

The debiased average ridge estimator is an empirical Bayes estimator of a common prior mean.

It varies between the linear fixed effects estimator and the average of individual specific least squares effects as the ridge regularization λ varies between ∞ and 0.

Its conditional expectation is the true conditional effect if coefficients do not vary over individuals.

In general its expectation is a weighted average of individual effects with matrix weights varying by the strength of identification of the individual effects.

The use of this estimator is motivated by an approximation where number T of time periods grows with the number of individuals n.

As in Chernozhukov et al. (2009, 2013) the identified set may shrink quickly as T grows with n.

As the number of regressors also grows with T the approximation error from a linear regression for each individual should also shrink quickly.

These two features enable accurate large sample inference about effects of interest for nonseparable models using the debiased average ridge estimator.

We bypass the need for identification and inference methods under partial identification.

Importantly, we also provide ways of checking the extent of partial identification in practice to help determine whether the large T theory applies.

We provide quantile plots of the ratio of size of regularized and true linear combinations.

Three examples of this estimator are

1) Estimating average exact surplus and deadweight loss for consumer demand, including an empirical example;

2) Estimating the effect of changing taxes on taxable income for nonlinear budget sets, including an empirical example;

3) Estimating treatment and policy effects from panel data.

An extensive application to consumer demand using scanner data is given.

In demand estimation panel data controls for endogeneity from imperfect market competition and from individual price indices for groups of goods.

Also the methods allow for zero demand by simply including the observations a zero expenditure share outcome.

For expenditure share outcomes, with log of prices and total expenditure as regressors, the individual ridge estimates shrink towards own price elasticity -1, cross price elasticities 0, and expenditure elasticity 1.

Average equivalent variation and deadweight loss for taxes on milk and soda are estimated.

Also find that weak axiom of revealed preference is rejected for 95% of individuals while most individuals are estimated to have negative own price elasticities when time stationary preferences are allowed.

Asymptotic theory given for continuous regressors in linear model with number of regressors much less than T.

Asymptotic theory for a binary regressor is also given, where the identified set shrinks fast enough with T so that standard inference is correct in large samples.

General result that combines continuous and discrete regressors is under construction.

This paper differs from Chernozhukov Hausman, and Newey (2019) in

1) The approximation of general nonseparable models.

2) Estimation of structural and causal effects, rather than just average coefficients.

3) Asymptotic theory that allows for approximation of a general nonseparable model and nonidentification with fixed T with the identified set shrinking fast enough with T so that inference is asymptotically correct.

4) Providing measures of the extent of identification for applications.

THE MODEL AND EFFECTS OF INTEREST

Panel data, observations indexed by individual i and time period t, $(i = 1, ..., n; t = 1, ..., T)$.

Outcome variable S_{it} and right hand side variable X_{it}.

S_{it} is an unknown function of X_{it} and unobserved individual heterogeneity η_{it},

$$S_{it} = s(X_{it}, \eta_{it}).$$

For demand X_{it} is individual prices and expenditure and S_{it} is expenditure share.

Unobserved heterogeneity η_{it} represent preferences, technology, or counterfactuals in a causal model; η_{it} may be infinite dimensional.

The function $s(x_t, \eta_t)$ is unknown.

$$S_{it} = s(X_{it}, \eta_{it}),$$

Nonseparable model in allowing for general functional interactions between the observed variables X_{it} and unobserved heterogeneity η_{it}.

Use conditional time stationarity of the η_{it} to identify and estimate effects of changing X_{it} on S_{it}.

Let $X_i = (X_{i1}', ..., X_{iT}')'$ be the vector of regressors for all time periods and $x_i = (x_1', ..., x_T')'$ a possible realization of X_i.

Time stationarity condition is

Assumption 1: The distribution of η_{it} conditional on X_i does not depend on t.

$$S_{it} = s(X_{it}, \eta_{it}),$$

Assumption 1: The distribution of η_{it} conditional on X_i does not depend on t.

Time effects allowed when X_{it} is a function of t, but need some restrictions for identification.

Taxable income example has time trend but reserve fuller discussion of time effects for another time.

Parameters of interest are averages of $s(x_{it}, \eta_{it})$ over x_{it}, η_{it} for $x_{it} \neq X_{it}$.

Example 1: Average policy and causal effects.

In the language of counterfactuals let $s(X_{it}^{+}, \eta_{it})$ be a potential outcome at $x_{it} = X_{it}^{+}$ and $s(X^{-}, \eta_{it})$ at $x_{it} = X_{it}^{-}$.

$$\theta_0 = \frac{1}{T} \sum_{t=1}^{T} E[s(X_{it}^{+}, \eta_{it}) - s(X_{it}^{-}, \eta_{it})].$$

This is an average effect of changing x_{it} from X_{it}^{-} to X_{it}^{+}.

Example 2: Average equivalent variation and deadweight loss bounds.

Important objects of interest for estimating welfare effects of taxes.

Here S_{it} is expenditure share of a commodity, $X_{it} = (P_{it}, Z_{it})$, P_{it} is price, Z_{it} a vector of covariates including total expenditure Y_{it} and possibly prices of other goods.

Let P_{it}^{-} be initial price for individual i at time t, Δ_{it} be change in price, π be a bound on the income effect over all individuals and $\omega_t(X_i)$ a weight function.

Let $U_i = (U_{i1}, ..., U_{iT})'$ be a T vector of $U(0,1)$ i.i.d. simulation draws.

A bound on equivalent variation averaged over time periods and individuals is

$$\theta_0 = E[\frac{1}{T}\sum_{t=1}^{T}\omega_t(X_i)V_t(X_i,U_i,s)],$$
$$V_t(X_i,U_i,s) = \exp(-\pi(P_{it}^{-}+\Delta_{it}U_{it}))\times$$
$$\Delta_{it}\frac{Y_{it}}{P_{it}^{-}+\Delta_{it}U_{it}}s(P_{it}^{-}+\Delta_{it}U_{it},Z_{it},\eta_{it}),$$

By Hausman and Newey (2016) θ_0 is an upper (lower) bound on weighted average of equivalent variation for a change from P_{it}^{-} to $P_{it}^{-}+\Delta_{it}$ when π is a lower (upper) bound on the income effect for every individual.

Corresponding average deadweight loss bound in paper.

Example 3: Average of heterogeneous taxable income elasticities

Structural economic model for taxable income S_{it} as function of budget set and heterogeneity from Blomquist, Kumar, and Newey (2022).

Isoelastic utility and a certain distribution for heterogeneity give

$$S_{it} = s(X_{it}, \eta_{it}) = \beta_1(\eta_{it}) + \sum_{j=2}^{4} X_{jit}\beta_j(\eta_{it}).$$

$\beta_2(\eta_{it})$ is taxable income elasticity for individual i, X_{2it} is log of slope of last budget segment, X_{3it} is the difference of logs of the slope of the first and last segment, and $X_{4it} = t$.

Average taxable income elasticity is

$$\theta_0 = E[\beta_2(\eta_{it})] = E[\frac{1}{T}\sum_{t=1}^{T}\{s(X_{it} + e_2, \eta_{it}) - s(X_{it}, \eta_{it})\}],\ e_2 = (0, 1, 0, 0)'.$$

A general form for parameter of interest including Examples 1 - 3 is for known functions $H_t^+(X_i)$ and $H_t^-(X_i)$,

$$\theta_0 = E[\frac{1}{T}\sum_{t=1}^{T}\{H_t^+(X_i)s(X_{it}^+,\eta_{it}) + H_t^-(X_i)s(X_{it}^-,\eta_{it})\}].$$

Example 1 is special case where $H_t^+(X_i) = 1$ and $H_t^-(X_i) = -1$.

Example 2 has $H_t^-(X_i) = 0$ and

$$\begin{aligned} X_{it}^+ &= (\bar{P}_{it} + \Delta_{it}U_{it}, Z_{it}), \\ H_t^+(X_i) &= \omega_t(X_i)\exp(-\pi(\bar{P}_{it} + \Delta_{it}U_{it}))\Delta_{it}\frac{Y_{it}}{\bar{P}_{it} + \Delta_{it}U_{it}}. \end{aligned}$$

Example 3 has $X_{it}^+ = X_{it} + e_2$, $X_{it}^- = X_{it}$, $H_t^+(X_i) = 1$, and $H_t^-(X_i) = -1$.

DEBIASED AVERAGE RIDGE ESTIMATTION

For now we bypass identification of θ_0 to focus on estimation.

Time stationarity of Assumption 1 is an "exclusion restriction" that is sufficient for identification when certain rank conditions are satisfied.

As always rank conditions can be checked in the data and we will return to this important problem in what follows.

Also we focus on large T inference where identified sets are small enough that partial identification issues can be ignored.

Importantly, we also provide ways of checking the extent of identification in practice to help determine whether the large T theory is good approximation.

Estimation is based on approximating the unknown, nonseparable outcome function $s(x_t, \eta_t)$ by a linear combination of a $J \times 1$ vector of known functions $b(x_t) = (1, b_2(x_t), ..., b_J(x_t))'$ (where we assume throughout that $b_1(x_t) = 1$).

The linear combination coefficients $\beta(\eta_t) = (\beta_1(\eta_t), ..., \beta_J(\eta_t))'$ are unknown functions that can depend nonparametrically on η

The approximation is

$$s(x_t, \eta_t) \approx b(x_t)'\beta(\eta_t).$$

Idea is that for x_t bounded and $s(x_t, \eta_t)$ a smooth enough function of x_t, various approximation theorems (e.g. Jackson theorems) imply that there are functions $b(x_t)$ not depending on η_t (e.g. power series) and associated coefficients $\beta(\eta_t)$ where the approximation error is small uniformly in x_t and η_t.

Analogous approximation for demand functions in Hausman and Newey (2016).

$$s(x_t, \eta_t) \approx b(x_t)'\beta(\eta_t).$$

Under time stationarity as in Assumption 1, for any $\tilde{X}_i$ that is conformable with and depends only on X_i,

$$\begin{aligned} E[s(\tilde{X}_{it}, \eta_{it})|X_i] &\approx E[b(\tilde{X}_{it})'\beta(\eta_{it})|X_i] = b(\tilde{X}_{it})'E[\beta(\eta_{it})|X_i] \\ &= b(\tilde{X}_{it})'\bar{\beta}(X_i),\ \bar{\beta}(X_i) := E[\beta(\eta_{it})|X_i]. \end{aligned}$$

Time stationarity implies $\bar{\beta}(X_i)$ does not depend on t, i.e same coefficients $\bar{\beta}(X_i)$ appear in the approximation for each time period t.

There is corresponding approximation for the parameters of interest, e.g. in Example 1:

$$\begin{aligned} E[s(X_{it}^+, \eta_{it}) - s(X_{it}^-, \eta_{it})|X_i] &\approx E[\{b(X_{it}^+) - b(X_{it}^-)\}'\beta(\eta_{it})|X_i] \\ &= \{b(X_{it}^+) - b(X_{it}^-)\}'E[\beta(\eta_{it})|X_i] \\ &= \{b(X_{it}^+) - b(X_{it}^-\}'\bar{\beta}(X_i). \end{aligned}$$

$$E[s(X_{it}^{+}, \eta_{it}) - s(X_{it}^{-}, \eta_{it})|X_i] \approx \{b(X_{it}^{+}) - b(X_{it}^{-})\}'\bar{\beta}(X_i).$$

By iterated expectations

$$\theta_0 = E[\frac{1}{T}\sum_{t=1}^{T} E[s(X_{it}^{+}, \eta_{it}) - s(X_{it}^{-}, \eta_{it})|X_i]] \approx E[\frac{1}{T}\sum_{t=1}^{T}\{b(X_{it}^{+}) - b(X_{it}^{-})\}'\bar{\beta}(X_i)]$$
$$= E[a_i'\bar{\beta}(X_i)],\ a_i = \frac{1}{T}\sum_{t=1}^{T}[b(X_{it}^{+}) - b(X_{it}^{-})].$$

Thus, in Example 1 θ_0 is approximately $E[a_i'\bar{\beta}(X_i)]$ for a_i as given.

For the general parameter θ_0 the approximation is

$$\theta_0 \approx E[a_i'\bar{\beta}(X_i)],\ a_i = \frac{1}{T}\sum_{t=1}^{T}\{H_t^{+}(X_i)b(X_{it}^{+}) + H_t^{-}(X_i)b(X_{it}^{-})\}.$$

Smoothness of the outcome $S_{it} = s(X_{it}, \eta_{it})$ in X_{it} is not necessary for a good approximation..

It is sufficient that

$$S_{it} = \bar{s}(X_{it}, \eta_{it}) + v_{it},\ \ E[v_{it}|X_i] = 0,$$

where $\bar{s}(x_{it}, \eta_{it})$ is smooth in x_{it}.

In particular, a panel binary choice model will not generally be smooth in regressors but the approximation still works if there are idiosyncratic components of η_{it} with smooth densities.

For example suppose S_{it} is binary with

$$S_{it} = 1(\eta_{2it} X_{it} > \eta_{1it}),\ \ \eta_{1it} \text{ i.i.d. and independent of } X_i, \eta_{2it}.$$

Then for the CDF G of η_{1it}, $E[S_{it}|X_i] = G(\eta_{2it} X_{it}) = \bar{s}(X_{it}, \eta_{it})$ will be smooth in X_{it} when the pdf of η_{1it} is smooth.

$$\theta_0 \approx E[a_i'\bar{\beta}(X_i)],\ a_i = \frac{1}{T}\sum_{t=1}^{T}\{H_t^+(X_i)b(X_{it}^+) + H_t^-(X_i)b(X_{it}^-)\}.$$

Can estimate θ_0 by replacing expectation by a sample average and plugging in estimator of $\bar{\beta}(X_i)$.

The approximation implies

$$E[S_{it}|X_i] \approx E[b(X_{it})'\beta(\eta_{it})|X_i] = b(X_{it})'\bar{\beta}(X_i).$$

Could try estimating $\bar{\beta}(X_i)$ from a linear regression of $S_i = (S_{i1}, ..., S_{iT})'$ on $B_i = [b(X_{i1}), ..., b(X_{iT})]'$.

Could be high multicollinearity, so regularize using ridge regression.

Let $Q_i = B_i'B_i/T$, $D_i = diag(0, I)$, $\lambda > 0$; does not regularize constant.

A ridge regression estimator of $\bar{\beta}(X_i)$ is

$$\hat{\beta}_i = (Q_i + \lambda D_i)^{-1}B_i'S_i/T.$$

We also debias to mitigate shrinkage bias from the ridge regularization.

Let A_i denote a square T-dimensional matrix with a_i' as its first row and its other rows being corresponding rows of the identity matrix.

Also, let

$$W_i = (Q_i + \lambda D_i)^{-1} Q_i \tag{1}$$

Estimator of θ_0 is

$$\hat{\theta} = \bar{a}'(\overline{AW})^{-1}\overline{A\beta},\ \bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i,\ \overline{AW} = \frac{1}{n}\sum_{i=1}^{n} A_i W_i,\ \overline{A\beta} = \frac{1}{n}\sum_{i=1}^{n} A_i \hat{\beta}_i. \tag{2}$$

An estimator $\hat{V}$ for the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta_0)$ can be obtained via delta method as

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_i^2,\ \hat{\psi}_i = (a_i - \bar{a})'(\overline{AW})^{-1}\overline{A\beta} + \bar{a}'(\overline{AW})^{-1} A_i[\hat{\beta}_i - W_i(\overline{AW})^{-1}\overline{A\beta}].$$

Example 3 illustrates that parameters of interest may include elements of $\theta_0 = E[\beta_{it}]$.

Like choosing a_i to be unit vectors and stacking these unit vectors into an identity matrix, leading to A_i being an identity matrix.

Debiased average ridge estimators $\hat{\theta}$ of $E[\bar{\beta}(X_i)]$ and $\hat{V}$ of the asymptotic variance V of $\sqrt{n}(\hat{\theta} - \theta_0)$ are then given by

$$\hat{\theta} = (\overline{W})^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i,\ \hat{V} = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i\hat{\psi}_i',\ \hat{\psi}_i = (\overline{W})^{-1}(\hat{\beta}_i - W_i\hat{\theta}).$$

PROPERTIES OF THE ESTIMATORS

The average slope has some properties that help in interpreting it.

Property A: The debiased average slope is an empirical Bayes estimator of a common prior mean for Gaussian likelihood and prior with variance $1/\lambda$; the debiased average slope can also be interpreted as the limit of a iterative procedure where the prior mean is updated to be the previous debiased average.

Property B: As λ goes to infinity the coefficients in $\hat{\beta}$ of nonconstant variables converges to the fixed effects estimator.

This result is consistent with $1/\lambda$ being the variance of the prior distribution of the slope coefficients.

As the prior variance shrinks the common prior for the slopes dominates the Bayes estimator and makes the estimated coefficients not vary over individuals.

The structural/causal effect estimator also has interesting basic bias mitigation properties.

First, for any λ the estimator is unbiased when $\bar{\beta}(X_i)$ does not vary with X_i.

Property C: If $\bar{\beta}(X_i) = \beta$ does not vary with i then $E[\hat{\theta}|X_1, ..., X_n] = \beta$.

Matrix weighting by A_i is important for this property.

The second bias mitigation property is

Property D: If Q_i is nonsingular for every observation i then as $\lambda \longrightarrow 0$,

$$E[\hat{\theta}|X_1, ..., X_n] \longrightarrow \frac{1}{n}\sum_{i=1}^{n} a_i \bar{\beta}(X_i).$$

This is minimal property for a regularization method.

It is important to understand the nature of $\hat{\theta}$ when $\bar{\beta}(X_i)$ does vary with X_i and when Q_i is singular for some observations i.

Nonsingularity of Q_i is the rank condition for identification of $\bar{\beta}(X_i)$, as usual for regression coefficients, so singular Q_i means that $\bar{\beta}(X_i)$ is not identified.

Consequently the linear combination $a_i'\bar{\beta}(X_i)$ is not identified either, except in the exceptional case that a_i is orthogonal to the null space of Q_i.

Nevertheless, $\hat{\beta}$ still estimates an identified object that may be of interest.

Consider the simple example of one time varying regressor and the object of interest is the average across individuals of the coefficient of that variable.

We suppose that the outcome variable is given by

$$S_{it} = \beta_{1it} + \beta_{2it} * X_{2it}.$$

We consider the average ridge estimator where $\lambda D_i = diag(0, \lambda)$.

Because the first diagonal element of this matrix is zero the slope estimator $\hat{\beta}_{2i}$ for observation i will be the ridge estimator from a regression of $Y_{it} - \bar{Y}_i$ on $X_{2it} - \bar{X}_2$.

Then for $\tilde{Q}_i = \sum_{t=1}^{T}(X_{2it} - \bar{X}_2)^2/T$, standard ridge calculations give

$$E[\hat{\beta}_{2i}|X_i] = w_i * \bar{\beta}_{2i}(X_i), \; w_i = \frac{\tilde{Q}_i}{\tilde{Q}_i + \lambda}.$$

Averaging across i then gives

$$E[\hat{\theta}_2|X_1, ..., X_n] = \frac{\sum_{i=1}^n w_i \bar{\beta}_2(X_i)}{\sum_{i=1}^n w_i}.$$

This is a weighted average of $\bar{\beta}_2(X_i)$ where the weight is zero if $\bar{\beta}_2(X_i)$ is not identified (i.e. X_{2it} does not vary over time t) and where more weight is given to observations where $\tilde{Q}_i$ is larger.

The limit as $\lambda \longrightarrow 0$ is the average of $\bar{\beta}_2(X_i)$ over observations where $\tilde{Q}_i > 0$.

In general,

$$\begin{aligned} E[\hat{\theta}|X_1, ..., X_n] &= \frac{1}{n}\sum_{i=1}^{n} a_i^{\lambda}\bar{\beta}(X_i), \\ a_i^{\lambda\prime} &= \bar{a}'(\overline{AW})^{-1}A_iW_i, \\ W_i &= (Q_i + \lambda D_i)^{-1}Q_i. \end{aligned}$$

The regularized linear combination coefficients a_i^{λ} have a known form that can be shown to be a projection of a transformation of the true ones on the orthogonal complement of the null space of Q_i.

Not necessarily 0 when Q_i is singular and not between zero and one.

Comparing a_i^λ and a_i across observations does give a way of evaluating extent of identification.

Here we consider quantile plots of

$$\sqrt{\frac{a_i^{\lambda\prime} a_i^\lambda}{a_i' a_i}}, \ (i = 1, ..., n).$$

Could and we will consider other measures of discrepancy between a_i^λ and a_i.

Important to be able to look at the data and evaluate the extent of nonidentification condition (i.e. rank condition not being satisfied).

Graham and Powell (2012) considered average effects only keeping observations where $\det(Q_i)$ was large enough.

Their average coefficients are unbiased when $\bar{\beta}(X_i)$ is constant but average effects are not.

When estimating average effects it seems good to regularize in a way specific to $a_i's$ of interest, as ridge does.

Also, their regularization has a different purpose, to provide consistent asymptotically normal estimates with continuous regressors and number of regressors equal to number of time periods.

Our interest in regularization is to allow discreteness in regressors while bypassing partial identification and assessing the extent of identification in applications.

APPLICATION TO CONSUMER DEMAND

We estimate demand for types of goods with completely flexible specifications of heterogeneity and functional form.

We model demand at an intermediate level of multi-stage budgeting, that is demand for a type of good in a bundle of multiple types.

Modeling type demand provides valid estimates of welfare and policy effects under preference separability conditions, e.g. see Deaton and Muellbauer (1980).

Very common in practice to impose some separability, such as excluding labor supply and intertemporal choice when modeling demand.

Nonparametric separability seems parsimonious, directed approach when evaluating policy involving a type of consumer good, like a tax on soda, rather than imposing logit demand with random coefficients on many specific goods.

Apply to estimating demand for groceries using the Nielsen retail scanner data to construct price indices, and using the Nielsen Homescan Panel to track purchases and household characteristics.

The data include 3396 households from Houston-area zip in the years 2007-2014.

We restrict our analysis to the 2864 households included for at least 12 months.

We model monthly expenditure .

We construct price indices based on monthly total expenditures per good type, and on the quantity purchased per month.

We model 15 good types: soda, milk, soup, water, butter, cookies, eggs, orange juice, ice cream, bread, chips, salad, yogurt, coffee, and cereal.

As in Burda, Harding, and Hausman (2008, 2012) we chose these groups because they made up a relatively large proportion of total grocery expenditure.

The price index for each good type is a weighted geometric average of the actual purchase prices (expenditure divided by quantity) over all purchases made by the household in the month, with weights equal to the proportion of expenditure on a specific item associated with a unique item code.

The price index is

$$\ln(p_\ell) = \sum_{a=1}^{A_\ell} w_{a\ell} \ln(p_{a\ell})$$

a denotes a particular item code, A_j is the number of codes for the j^{th} type of good, $w_{a\ell}$ is the proportion of expenditure on commodity ℓ that is spent on code a, and $p_{a\ell}$ is expenditure on code a divided by quantity of code a for a month.

This is a Tornqvist price index which was shown by Diewert (1976) to be exact for a quadratic utility specification and a second order approximation to the exact price index for any utility.

Deaton and Muellbauer (1980, pp 132-133) showed that with weak separability this price index appears in share equations for a Rotterdam demand specification (i.e. log quantity as a linear function of log prices and log expenditure) and indicate that it leads to a good approximation when prices within a group tend to move together.

Including zero expenditures makes it necessary to impute price indices for time periods where an individual purchased none of a particular good.

If a household had purchased the good before, then price indices are imputed as the most recent price faced by the household in a past purchase.

Rarely, a good is never purchased prior to a given month, in which case its imputed price is the average price of from nearby stores where purchases are made.

We checked for differences in results between using all households and the 2864 that were present for at least a year and found no statistically significant differences, suggesting attrition and selection bias do not play a large role in this data.

We checked time stationarity and by comparing fixed effects elasticity estimates with and without quarterly time dummies and found very small differences in elasticities.

Including time dummies made less difference for 2010-2014 (after great recession) so used just those years.

The frequency of household-month observations with zero total expenditures varies by good: for some goods, particularly milk and soda, most households record purchases each month, while other goods, such as orange juice and ice cream, are purchased more infrequently.

Zero purchases could be related to good storage, which we do not model.

Our empirical results are for goods with low frequency of zero purchases, namely soda and milk.

The model allows preferences to vary over time as familiar from discrete choice panel data, where there is idiosyncratic data

WARP fails to hold for 95% of individuals for 15 goods considered over 60 months, which could happen because of time variation in preference

Minimal regularization of individual specific demands have negative own price effect for 65% of observations and only 1.5 percent for soda and 2.3 percent for milk would reject a one sided test of downward sloping demand.

These results are consistent with utility maximization with time varying preferences.

Average equivalent variation and deadweight loss are valid in this case.

Table 1: Soda Consumer Surplus Estimates (Linear)

	Upper Bounds Income Quantiles			Lower Bounds Income Quantiles		
λ	Upper	Lower	All	Upper	Lower	All
1000	11.78	10.56	10.5	11.78	10.56	10.5
	(0.464)	(0.551)	(0.22)	(0.464)	(0.551)	(0.22)
100	11.78	10.56	10.5	11.78	10.56	10.5
	(0.464)	(0.551)	(0.22)	(0.464)	(0.551)	(0.22)
10	11.8	10.56	10.51	11.8	10.56	10.51
	(0.465)	(0.551)	(0.219)	(0.465)	(0.551)	(0.219)
1	11.85	10.58	10.56	11.85	10.58	10.56
	(0.466)	(0.55)	(0.219)	(0.466)	(0.55)	(0.219)
0.05	11.92	10.61	10.61	11.92	10.61	10.61
	(0.469)	(0.551)	(0.22)	(0.469)	(0.551)	(0.22)
0.005	11.93	10.63	10.63	11.93	10.63	10.63
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
0.0005	11.93	10.64	10.63	11.93	10.64	10.63
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
5×10^{-5}	11.93	10.64	10.63	11.93	10.64	10.63
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)

Table 2: Soda Consumer Surplus Estimates (Quadratic)

	Upper Bounds Income Quantiles			Lower Bounds Income Quantiles		
λ	Upper	Lower	All	Upper	Lower	All
1000	12.08	10.66	10.69	12.08	10.66	10.69
	(0.471)	(0.556)	(0.22)	(0.471)	(0.556)	(0.22)
100	12.08	10.66	10.69	12.08	10.66	10.69
	(0.471)	(0.556)	(0.22)	(0.471)	(0.556)	(0.22)
10	12.07	10.66	10.69	12.07	10.66	10.69
	(0.471)	(0.556)	(0.22)	(0.471)	(0.556)	(0.22)
1	12.02	10.65	10.68	12.02	10.65	10.68
	(0.47)	(0.554)	(0.22)	(0.47)	(0.553)	(0.22)
0.05	11.98	10.66	10.67	11.98	10.66	10.67
	(0.471)	(0.552)	(0.22)	(0.471)	(0.552)	(0.22)
0.005	11.96	10.67	10.66	11.96	10.67	10.66
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
0.0005	11.94	10.67	10.66	11.94	10.67	10.66
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
5×10^{-5}	11.93	10.68	10.66	11.93	10.68	10.66
	(0.47)	(0.553)	(0.22)	(0.47)	(0.553)	(0.22)

Table 3: Soda Consumer Surplus Estimates (Cubic)

	Upper Bounds Income Quantiles			Lower Bounds Income Quantiles		
λ	Upper	Lower	All	Upper	Lower	All
1000	11.98	10.64	10.64	11.98	10.64	10.64
	(0.467)	(0.553)	(0.219)	(0.467)	(0.553)	(0.219)
100	11.97	10.64	10.64	11.97	10.64	10.64
	(0.467)	(0.553)	(0.219)	(0.467)	(0.553)	(0.219)
10	11.97	10.64	10.64	11.97	10.64	10.64
	(0.467)	(0.552)	(0.219)	(0.467)	(0.552)	(0.219)
1	11.96	10.64	10.65	11.96	10.64	10.65
	(0.468)	(0.551)	(0.219)	(0.468)	(0.551)	(0.219)
0.05	11.95	10.66	10.65	11.95	10.66	10.65
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
0.005	11.94	10.67	10.66	11.94	10.67	10.66
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
0.0005	11.93	10.68	10.66	11.93	10.68	10.66
	(0.47)	(0.552)	(0.22)	(0.47)	(0.552)	(0.22)
5×10^{-5}	11.92	10.69	10.66	11.92	10.68	10.66
	(0.471)	(0.552)	(0.221)	(0.471)	(0.552)	(0.221)

Table 3: Soda Deadweight Loss Estimates (Cubic)

	Upper Bounds			Lower Bounds		
	Income Quantiles			Income Quantiles		
λ	Upper	Lower	All	Upper	Lower	All
1000	0.4385	0.4047	0.3944	0.4383	0.4046	0.3942
	(0.0202)	(0.0242)	(0.0105)	(0.0202)	(0.0242)	(0.0105)
100	0.4382	0.4044	0.3941	0.438	0.4042	0.3939
	(0.0202)	(0.0242)	(0.0105)	(0.0202)	(0.0242)	(0.0105)
10	0.4357	0.4017	0.3918	0.4354	0.4015	0.3916
	(0.0199)	(0.024)	(0.0104)	(0.0199)	(0.024)	(0.0104)
1	0.4302	0.3947	0.3867	0.4299	0.3946	0.3865
	(0.0193)	(0.0238)	(0.0101)	(0.0193)	(0.0238)	(0.0101)
0.05	0.4264	0.3851	0.3819	0.4261	0.385	0.3817
	(0.0193)	(0.024)	(0.0102)	(0.0193)	(0.024)	(0.0102)
0.005	0.4303	0.3803	0.3795	0.43	0.3801	0.3793
	(0.0207)	(0.0254)	(0.011)	(0.0207)	(0.0254)	(0.011)
0.0005	0.4394	0.3818	0.381	0.4391	0.3816	0.3808
	(0.0247)	(0.0278)	(0.0126)	(0.0247)	(0.0278)	(0.0126)
5×10^{-5}	0.4427	0.3804	0.3822	0.4425	0.3802	0.382
	(0.0279)	(0.0291)	(0.0142)	(0.0279)	(0.0291)	(0.0142)

Table 3: Milk Consumer Surplus Estimates (Cubic)

	Upper Bounds Income Quantiles			Lower Bounds Income Quantiles		
λ	Upper	Lower	All	Upper	Lower	All
1000	9.135	7.018	7.58	9.135	7.018	7.58
	(0.341)	(0.329)	(0.137)	(0.341)	(0.329)	(0.137)
100	9.133	7.018	7.58	9.133	7.018	7.58
	(0.341)	(0.328)	(0.137)	(0.341)	(0.328)	(0.137)
10	9.123	7.026	7.582	9.123	7.026	7.582
	(0.341)	(0.327)	(0.137)	(0.341)	(0.327)	(0.137)
1	9.113	7.056	7.59	9.113	7.056	7.59
	(0.34)	(0.326)	(0.137)	(0.34)	(0.326)	(0.137)
0.05	9.135	7.076	7.606	9.135	7.076	7.606
	(0.341)	(0.326)	(0.138)	(0.341)	(0.326)	(0.138)
0.005	9.149	7.075	7.613	9.149	7.075	7.613
	(0.343)	(0.326)	(0.138)	(0.343)	(0.326)	(0.138)
0.0005	9.145	7.071	7.603	9.145	7.071	7.603
	(0.345)	(0.326)	(0.139)	(0.345)	(0.326)	(0.139)
5×10^{-5}	9.15	7.062	7.595	9.15	7.062	7.595
	(0.347)	(0.327)	(0.139)	(0.347)	(0.327)	(0.139)

Table 3: Milk Deadweight Loss Estimates (Cubic)

	Upper Bounds			Lower Bounds		
	Income Quantiles			Income Quantiles		
λ	Upper	Lower	All	Upper	Lower	All
1000	0.2003	0.1648	0.1685	0.2002	0.1647	0.1684
	(0.0136)	(0.0119)	(0.00758)	(0.0136)	(0.0119)	(0.00758)
100	0.2002	0.1643	0.1684	0.2002	0.1643	0.1683
	(0.0135)	(0.0118)	(0.00752)	(0.0135)	(0.0118)	(0.00752)
10	0.1984	0.1605	0.1663	0.1984	0.1605	0.1663
	(0.0129)	(0.0113)	(0.00713)	(0.0129)	(0.0113)	(0.00713)
1	0.1868	0.1458	0.155	0.1867	0.1458	0.155
	(0.012)	(0.0105)	(0.00663)	(0.012)	(0.0105)	(0.00663)
0.05	0.1588	0.1224	0.1324	0.1587	0.1224	0.1324
	(0.0159)	(0.0157)	(0.00891)	(0.0159)	(0.0157)	(0.00891)
0.005	0.148	0.1174	0.1241	0.148	0.1174	0.124
	(0.0267)	(0.024)	(0.0135)	(0.0267)	(0.024)	(0.0135)
0.0005	0.1585	0.1254	0.1344	0.1585	0.1254	0.1344
	(0.0419)	(0.034)	(0.0198)	(0.0419)	(0.034)	(0.0198)
5×10^{-5}	0.1574	0.1412	0.1435	0.1574	0.1412	0.1435
	(0.0519)	(0.0439)	(0.0251)	(0.0519)	(0.0439)	(0.0251)

Figure 2: Quantile Plot, Soda

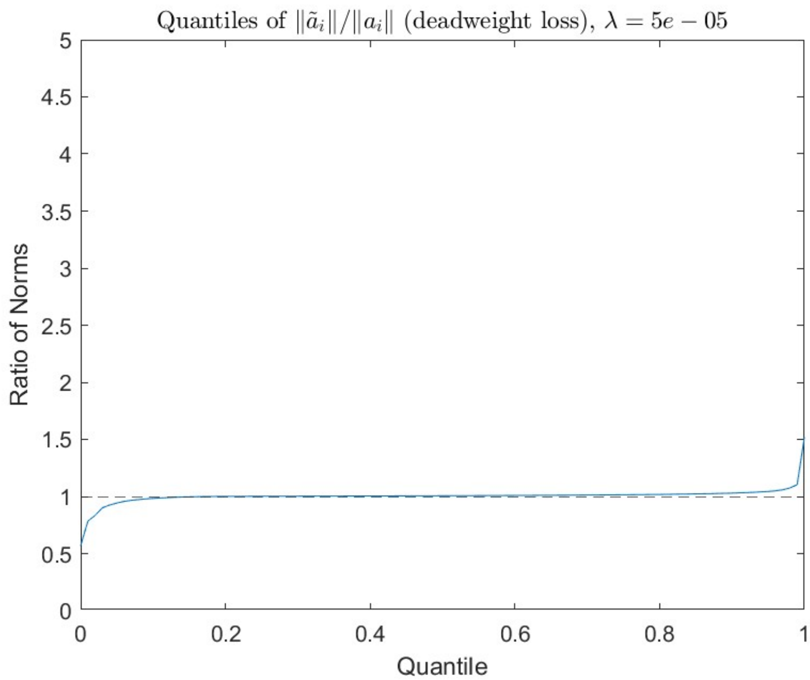


Figure 3: Quantile Plot, Milk

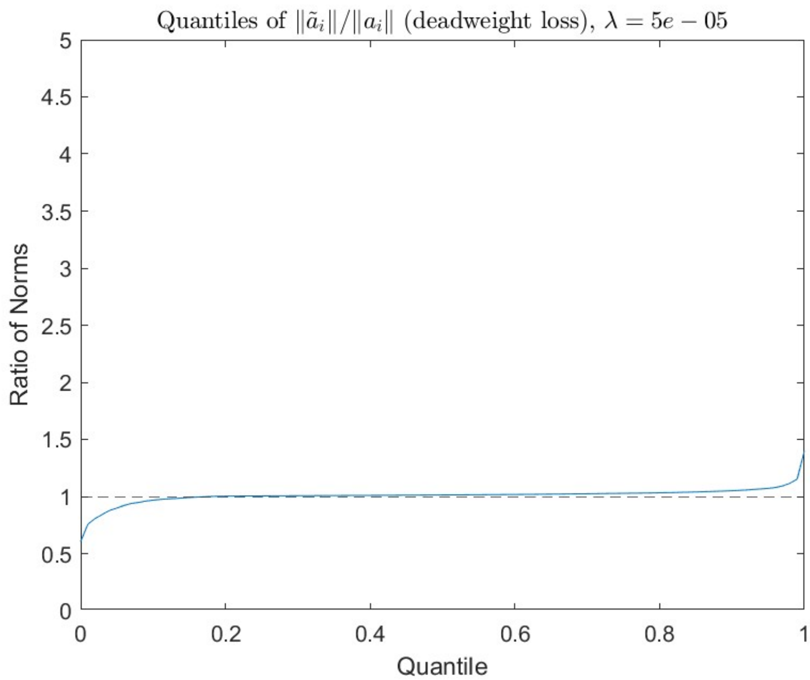


Figure 4: Quantile Plot, Soda

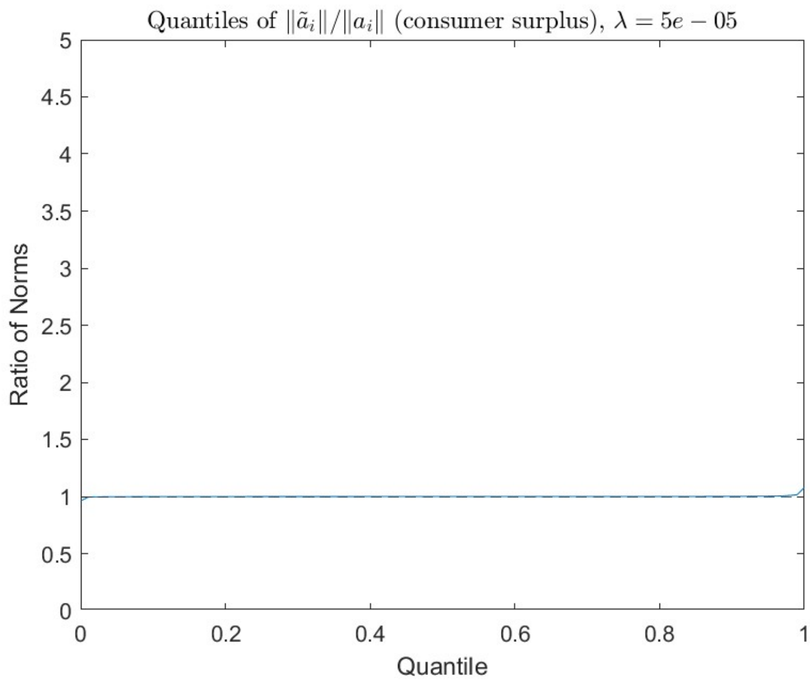
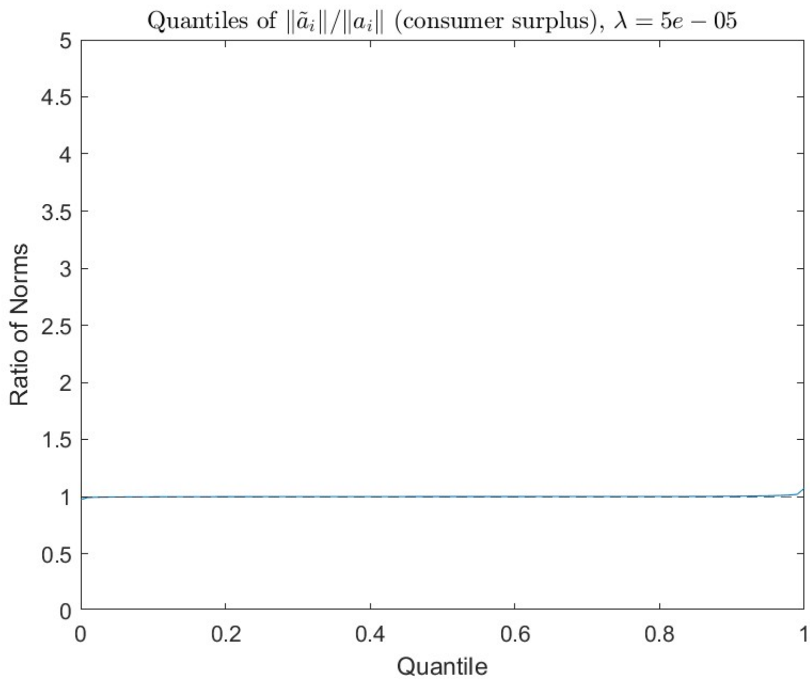


Figure 5: Quantile Plot, Milk



APPLICATION TO HETEROGENOUS TAXABLE INCOME ELASTICITY

Model is

$$S_{it} = s(X_{it}, \eta_{it}) = \beta_1(\eta_{it}) + \sum_{j=2}^{6} X_{jit}\beta_j(\eta_{it}).$$

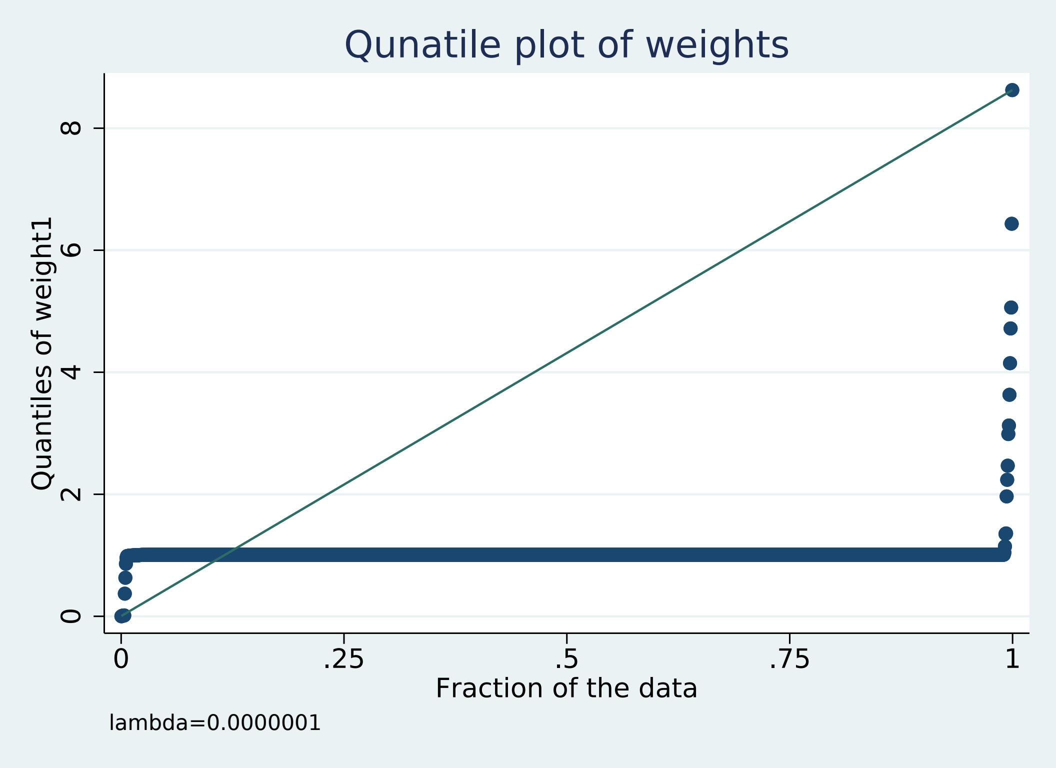
where S_{it} is log taxable income, $\beta_2(\eta_{it})$ is taxable income elasticity for individual i, X_{2it} is log of slope of last budget segment, X_{3it} is the difference of logs of the slope of the first and last segment, X_{4it} is log of intercept (nonlabor income) of first segment, X_{5it} is difference of log of interecept of last and first segments, and $X_{6it} = t$.

The time trend allows for individual specific productivity growth.

Data is PSID.

Report results in following table and graph.

Lambda	Non-debised slope elasticity	Debised slope elasticity	Standard Error	Non-debised income elasticity	Debised income elasticity	Standard Error
0	0.780617	0.787642	0.415832	0.01345	0.007462	0.062251
1.00E-07	0.677571	0.685251	0.37199	0.001554	0.001543	0.053651
1.00E-06	0.597343	0.605638	0.288264	0.007321	0.007399	0.035524
1.00E-05	0.474162	0.486876	0.205531	0.026225	0.027226	0.02018
0.0001	0.34046	0.368166	0.148268	0.016726	0.020582	0.014406
0.001	0.295437	0.3787	0.107041	-0.0011	0.001402	0.010159
0.01	0.220099	0.531998	0.093561	-0.00267	-0.00328	0.009117
0.1	0.09549	0.759263	0.091503	-0.00083	0.003373	0.009558
0.2	0.067323	0.847766	0.094317	-0.00053	0.006233	0.009713
0.3	0.053496	0.898555	0.097036	-0.00041	0.007672	0.009782
0.4	0.044842	0.93253	0.099342	-0.00034	0.008516	0.009827
0.5	0.038785	0.957175	0.10128	-0.00029	0.009057	0.009861
1	0.023564	1.021592	0.107563	-0.00018	0.010123	0.009993
2	0.013386	1.066017	0.113199	-0.0001	0.010501	0.010166
3	0.009376	1.083669	0.115836	-7.2E-05	0.010553	0.010268



Qunatile plot of weights
Quantiles of weight1
0
2
4
6
8
0
.25
.5
.75
1
Fraction of the data
lambda=0.0000001

SUMMARY

We give linear estimators of structural and causal effects in nonseparable panel data.

These estimators are based on linear random coefficients approximations to nonseparable models.

Via regularization these estimators bypass partial identification when there are many time periods.

Importantly, we give ways of quantifying the extent of partial identification in applications.

Large sample theory is given for large n and T.

We find little sensitivity of the results to the degree of approximation in applications to estimation of average welfare effects and to average taxable income elasticities with heterogeneity.