Imposing Structure by Prior Knowledge in Semiparametric Analysis

Stefan Sperlich

Département des sciences économiques
Université de Genève

April 14, 2011

Overview

- Pros and Cons for non- and semiparametric methods
  - Powerful data-analytic tools
  - Problems: Curse of dimensionality, bandwidth, boundary, bias
  - Justified doubts, e. g. concerning forecasting performance
- *Hypothesis*: With **suitable incorporation of prior knowledge** in the statistical modeling process these methods can improve in many (economic) fields
- Will consider
  - Prediction of American stock returns by parametric priors
  - Prediction of Danish stock returns with generated regressors
  - Marshallian demand analysis with (parametric) restrictions
  - Hicksian demand anal. with generated regressors and par. restrictions
- Will concentrate on kernel based local-polynomial regression

Motivation

- *Propose* different ways to include prior knowledge in semiparametrics
- *Idea*: economic theory should directly guide the modeling process
- *Statistical advantages*: dimension, variance or bias reduction by importing more structure
- *Typical examples*: PLM, SIM, additivity (GAM), monotonicity (for regression), symmetry (for densities), ...
- But still, in econometric-theory literature the general tendency in the literature is to *relax* functional forms, not vice verse.

First Thoughts

- However, **on the one hand** we know already for parametric forecasting that it improves if weak restrictions on the signs of coefficients and return forecasts are imposed, see However, Campbell, Thompson (2008)
- or that incorporating information about the order of integration can result in large efficiency gains, see Lewellen (2004); Torous, Valkanov, Yan (2004); Campbell, Yogo (2006).
- and **on the other hand** for many economic model like consumer demand systems plenty of model restrictions have to be imposed to guarantee reasonable and interpretable outcomes
- take symmetry and non-negativity of the Slutsky matrix, adding-up for the equations and homogeneity of the functions which automatically causes dimension reduction

Four Case studies as for illustration

Nonparametric Prediction of Stock Returns

- **Preliminaries**:
  - A *validated* R^2, a measure for the quality of prediction
  - A bootstrap test for significant forecast power
- Improved prediction through parametric prior smoothing
- Prediction with predicted bonds

Semiparametric Analysis of Consumer Demand

- **Preliminaries**:
  - A system of preferences and demand
  - Integrability conditions
- Estimating the indirect utility under constraints
- Simplifications by use of generated regressors

Predicting Stock Index and Returns

using prior knowledge

for implicit modeling

A performance measure: the validated R^2

What is an appropriate performance measure for prediction purposes?

- The classical and adjusted R^2s are good for **in-sample**, bad for **out-sample** prediction
- as still very popular in finance ... looked for modification
- but would like to know how well the estimate works **outside** the considered moderate sample
- Replace **total variation** and **not explained variation** by its **cross validated** analogs
- Certainly, the CV can be adapted to sample size and autocorrelation AR function

The definition of our performance measure in detail

We consider $Y_t = g(X_t) + \xi_t$ and define

$$R_V^2 = 1 - \frac{\sum_t \{Y_t - \hat{g}_{-t}\}^2}{\sum_t \{Y_t - \bar{Y}_{-t}\}^2},$$

Properties:

- $R_V^2 \in (-\infty, 1]$ where $R_V^2 < 0$ if we cannot predict better than the mean
- Measures how well a given model and estimation principle predicts compared to another (here: to the CV mean)
- CV punishes overfitting, i.e. pretending a functional relationship that is not really there (leads to $R_V^2 < 0$)

Can we beat the historical mean?

- Parametric null hypothesis vs. non-/semiparametric alternative
- H_0: $Y_t = \bar{Y} + \xi_t$ vs. H_1: $Y_t = g(\boldsymbol{X}_{t-1}) + \varepsilon_t$
- Construct B bootstrap samples $\{Y_1^b, \ldots, Y_{\mathcal{T}}^b\}$ with **residuals under the null**

$$Y_t^b = Y_t + \hat{\varepsilon}_t^0 \cdot u_t^b, \qquad \hat{\varepsilon}_t^0 = Y_t - \hat{g}_{-t}$$

  with iid zero-mean variance-one rv u_t^b.
- In each bootstrap iteration b calculate $R_V^{2,b}$
- Determine **quantiles** of empirical distribution of R_V^2 under the Null

$$F^*(u) = \frac{1}{B} \sum_b \mathbb{1}_{\{R_V^{2,b} \leq u\}}$$

Working with Parametric Priors

Incorporating parametric prior knowledge

Include **prior information** in analysis coming from

- (Simple) empirical data analysis or statistical modeling
- Good economic model

Basic idea: Nonparametric estimator **multiplicatively** guided by, for example, parametric model

$$g(x) = g_\theta(x) \cdot \frac{g(x)}{g_\theta(x)}$$

Essential fact:

- Prior captures characteristics of **shape** of $g(x)$
- Second factor **less** variable than original function
- Nonparametric estimator of **correction factor** $\frac{g(x)}{g_\theta(x)}$ with **better results** and **less bias**

Improved smoothing through prior knowledge

One idea to solve problems of **fully** nonparametric models:

- Curse of dimensionality
- Boundary problems
- Bandwidth (incl. local vs global) problems
- etc.

Dimension and bias (or variance) reduction:

$$g(x_1) + c = (g_\theta(x_2) + c) \cdot \frac{g(x_1) + c}{g_\theta(x_2) + c} = \tilde{g}(x_1, x_2)$$

Consider also **higher** dimensions for x_1 and x_2 with possibly **overlapping** covariates

Miscellaneous ...

Local Problem: Prior crosses x-axis

- More robust estimates with suitable trimming (censoring or truncation)
- **Shift** by a distance c so that new prior strictly greater than zero and does not intersect the x-axis

$$g(x) + c = (g_\theta(x) + c) \cdot \frac{g(x) + c}{g_\theta(x) + c}$$

- For **increasing** c more and more equal to **usual** local-polynomial: **Diminishes** effect of guide
- Idea goes back to Glad (1998)

Illustration with yearly Stock Price Index

Description of data

- Annual **American** stock market data
- January values of the Standard and Poor Composite Stock Price Index (period: 1871–2009)
- More details: Shiller (1989, 2005)
- Variables: stock price index, dividend and earnings accruing to index, short- and long-term interest rates, consumer price index (inflation), ten-year government bond, etc.

First Step: One-dimensional

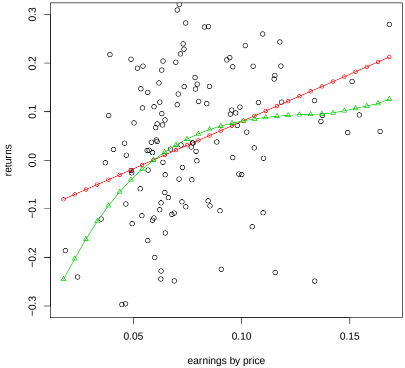
Table: Predictive power: R_V^2 and p-values

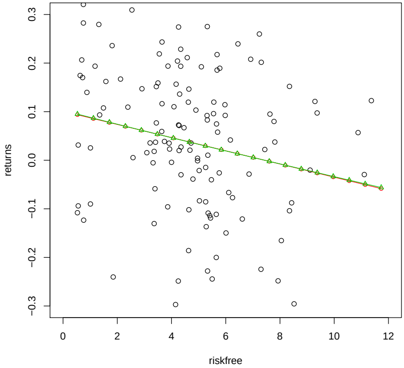
	S	d	e	r	L	inf	b
par	-1.0	1.0	8.0	2.7	-1.1	-1.4	-0.4
nonpar	-1.2	0.9	11.8	2.5	-0.8	-1.6	-0.7
	(0.596)	(0.193)	(0.005)	(0.079)	(0.571)	(0.759)	(0.573)

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t \quad \text{vs.} \quad Y_t = g(X_{t-1}) + \xi_t$$

- OLS and local-linear kernel-regression
- Only **earnings** and **risk-free** with predictive power
- Factor **1.5** increase for earnings

One-dimensional case - graphs





Second Step: Two-dimensional

Table: Predictive power: R^2_V and p-values

	e, S	e, d	e, r	e, L	e, inf	e, b
par	6.8	6.9	12.2	7.3	9.2	8.8
nonpar	8.5 (0.003)	12.6 (0.003)	13.7 (0.000)	11.0 (0.004)	11.0 (0.000)	11.3 (0.000)

$$Y_t = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X}_{t-1} + \varepsilon_t \quad \text{vs.} \quad Y_t = g(\boldsymbol{X}_{t-1}) + \xi_t$$

- Par: Improved prediction with **extra** variable *inf*, *b*, *r*
  Model $\{e, r\}$ even better than one-dim. nonpar.
- Nonpar: **All** shown models beat significantly historical mean and seem to **improve** prediction compared to 2-dim par.
- Increase in predictive power of ”only” **12**% compared to first step

Second Step: Two-dim. guided by prior

Table: Predictive power: R^2_V

	e, S	e, d	e, r	e, L	e, inf	e, b
nonpar	8.5	12.6	13.7	11.0	11.0	11.3
prior	6.6	13.5	12.1	12.8	9.5	8.0

- Here we include always **same variables** for prior (linear regression) and correction (nonpar.)
- Increase of 7% for $\{e, d\}$ and **16**% for $\{e, L\}$
- Slightly decrease for the rest:
  Poor prior or already adequate 2-dim. fit ?

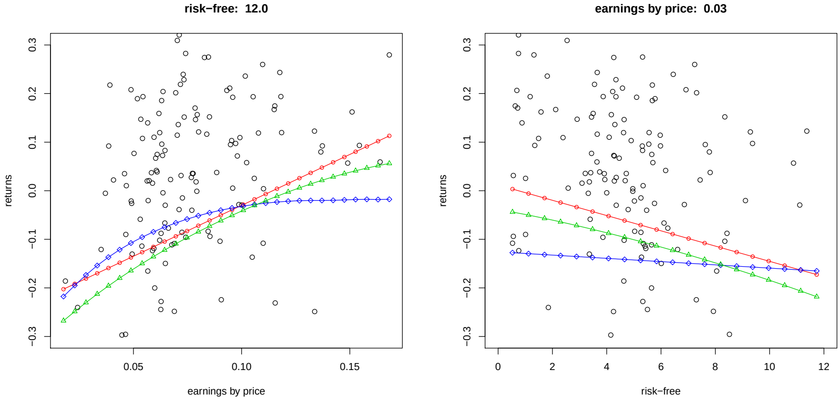
Third Step: Different variables for prior

Table: Predictive power: R^2_V

	S	d	e	r	L	inf	b
e	8.8	7.6	9.3	15.8	10.7	11.4	11.8
e, L	9.9	13.1	14.2	18.5	13.3	13.3	11.4

- Prior: 1-dim. linear regression
- Correction, i.e. nonparametric factor:
  - 1-dim: $\{e, S\}$, $\{e, r\}$, $\{e, inf\}$, $\{e, b\}$ **improve** compared to fully nonpar.
  - 2-dim: e. g. $\{e, L\}$ improvement of **29**%
  - Best: $\{e, L, r\}$ (**35**% to fully nonpar, **131**% to par)

Third Step: Graphs of best model



Working with Predicted Factors

The prediction framework

- The excess stock return:

$$S_t = \log\{(P_t + D_t)/P_{t-1}\} - r_{t-1}$$

with **dividends** D_t paid during year t, **stock price** P_t at the end of year t, and **short-term interest rate** r_t

$$r_t = \log(1 + R_t/100)$$

with **discount rate** R_t

- Covariates with predictive power in simple regression: **dividend-price ratio**, **earnings-price ratio**, or **interest rates**
- In nonpar. regression

$$Y_t = g(d_{t-1}, S_{t-1}) + \varepsilon_t$$

with **dividend-price ratio** d_{t-1} and **excess stock returns** S_{t-1}

Prediction with Bonds ?

There exists some economic motivation:

- Usually: **separate analysis** of stocks and bonds (**positively correlated**)
- Same year's bond yield is basically the **prediction error**
- **FED-Model**: Direct comparison of stocks and bonds
- Are stocks and bonds driven by the **same factors/informations**?
- To what extent they move together (**co-movement**)?
- Economic theory: prices are driven by **fundamentals**, investors should focus on **forward** earnings and profitability

Use **unknown** bond of the **current** year as **further** covariate

Prediction with constructed regressors

Consider now the two-step procedure

- **1. Step:** Construct **bond yields** with nonparametric model

$$b_t = m(v_{t-1}) + \zeta_t,$$

  where v_{t-1} vector of regressors (e. g. last years **bond yield**, **interest rate**, **dividend-price ratio**, or **excess stock returns**)

- **2. Step:** Include **pilot estimate** $\hat{b}_t$ in local-linear kernel-regression

$$Y_t = g(\hat{b}_t, w_{t-1}) + \varepsilon_t.$$

- Note:
  - Bandwidth choice (CV) in each or only in the final step
  - Simple linear model automatically embedded (estimated **without bias**)

Statistical framework

Exists some statistical motivation?

- Let $\tilde{g}$ be the function of (unknown) actual bond

$$Y_t - g(\hat{b}_t) \;=\; Y_t - \tilde{g}(b_t) + \tilde{g}(b_t) - g(\hat{b}_t) \;\simeq\; \tilde{\varepsilon}_t + g'(\hat{b}_t)(b_t - \hat{b}_t)$$

- Second term quite **predictable** (empirical study)
- Maybe a closer look to the **prediction error** clarifies the relation of bond and stock prediction
- Asymptotically for **dependent data** (algebraic α-mixing): as had we observed the real bond

Theorem

$$\mid g_{LL}(\hat{b}_t) - \tilde{g}(b_t) \mid \to 0$$

Why using constructed regressors in Statistics ?

- Interpret first stage as **optimal nonparametric transformation**
- Mapping the long-term interest rate to current bond yield

$$L_{t-1} \longrightarrow \hat{b}_t$$

- Subsequent nonparametric smoother of transformed variable is characterized by **less bias**
- Practical example of method of Park et al. (1997) which improves nonparametric regression with simple transformation techniques
- Small difference: We use additional variable, in their work they estimate on the original scale

The old stories ...

Why not **directly** v_{t-1} in stock prediction?

- Multi-dim. estimation suffers from the **curse of dimensionality** in several aspects:
  - Dimension of the covariates
  - Interpretability
- To circumvent curse of dimensionality **more structure** proposed:
  - Additivity
  - Semiparamteric modeling
- Use obtained structural information (not necessarily additive) as a kind of **dimension** and **complexity reduction**
- Reduce variation and **improve** predictive power in the R_V^2–sense

Illustration with Danish data

Description of the data

- Annual **Danish** stock and bond market data (period: 1922–1996)
- Value weighted portfolio of individual stocks (chosen to obtain maximum coverage of the market index of CSE)
- CSE open during the second world war
- Corrections for stock splits and new equity issues below market prices
- More details: *Lund and Engsted (1996)*.
- Variables: stock price index, dividend accruing to index, bond, short- and long-term interest rates

Danish data: **Bond** prediction

Table: R_V^2 of different **Bond** models 1923–1996

v_{t-1}	S	L	r	S,r	S,r,b
par	11.6%	24.0%	22.3%	33.1%	37.4%
nonpar	16.3%	23.9%	26.8%	33.0%	37.4%

- Bonds seem to be **predictable** in an adequate way
- Actually with **both**, parametric and nonparametric models
- ... there exist also some (here largely neglected) literature on parametric bond prediction ...

Danish data: **Stock** prediction

Table: R_V^2 of different **Stock** models 1923–1996

$w_{t-1} \backslash v_{t_1}$	d	r	b	d,L	d,r	r,b
par	-6.3%	-5.7%	-4.0%	-5.8%	-7.2%	0.5%
nonpar	-1.4%	-3.6%	5.9%	-6.0%	-7.4%	-8.6&
$\hat{b}_t$	8.3%	1.4%	10.6%	-3.8%	2.9%	-3.6%
$\hat{b}_t, v_{t-1}$	13.9%	16.3%	8.9%	28.3%	21.6%	20.3%

- All parametric models with **negative** R_V^2
- Very good results in general for **diagonal** $w_{t-1} = v_{t-1}$
- **Improvement** of prediction from $R_V^2 = 5.9\%$ to
  - $R_V^2 = 28.9\%$ for $\hat{g}(\hat{b}_t, d_{t-1}, S_{t-1}, L_{t-1})$ and $\hat{b}_t = \hat{p}(d_{t-1}, L_{t-1})$
  - $R_V^2 = 30.3\%$ for $\hat{g}(\hat{b}_t, d_{t-1}, S_{t-1}, L_{t-1})$ and $\hat{b}_t = \hat{p}(d_{t-1})$

Consumer Demand Analysis

estimating expenditure equations

starting from dual problems

Preliminary Remarks to Preferences and Demand

- for utility U, nominal total expenditures X, prices $\boldsymbol{P}$, quantities $\boldsymbol{Q}$, shares $\boldsymbol{W}$
- Consider the consumer problem:
  1. $Max\ U = v(\boldsymbol{Q})$ subject to $\boldsymbol{P}'\boldsymbol{Q} = X$
  2. $Min\ X = \boldsymbol{P}'\boldsymbol{Q}$ subject to $v(\boldsymbol{Q}) = U$
- with solution: $Q_i = g_i(X, \boldsymbol{P}) = h_i(U, \boldsymbol{P}),\ i = 1, \ldots, M$
  1. the *Marshallian* (or *uncompensated*) demands and
  2. the *Hicksian* (or *compensated*) demands respectively
- substituting into original problems gives
  1. the *indirect utility function* $U = V(X, \boldsymbol{P})$, and
  2. the *cost function* $X = C(U, \boldsymbol{P})$ respectively

Some Properties of the cost function

1. homogeneous in prices, i.e. $C(U, \theta \boldsymbol{P}) = \theta C(U, \boldsymbol{P}) \; \forall \theta > 0$
2. concave in prices
3. increasing in U and at least one P_i, nondecreasing in all

Assuming differentiability we get by Shephard's Lemma

$$\frac{\partial C(U, \boldsymbol{P})}{\partial P_i} = h_i(U, \boldsymbol{P}) = Q_i = g_i(X, \boldsymbol{P}) = -\frac{\partial V / \partial P_i}{\partial V / \partial X}$$

called Roy's identity, and for budget shares

$$w_i = \frac{\partial \ln C(U, \boldsymbol{P})}{\partial \ln P_i} = \frac{\partial \ln V / \partial \ln P_i}{\sum \partial \ln V / \partial \ln P_k}$$

Properties of Integrability

Often one concentrates on

1. homogeneity $g_i(\theta X, \theta \boldsymbol{P}) = g_i(X, \boldsymbol{P}) = h_i(U, \boldsymbol{P}) = h_i(U, \theta \boldsymbol{P})$
2. symmetry $\partial h_i(U, \boldsymbol{P})/\partial P_j = \partial h_j(U, \boldsymbol{P})/\partial P_i$
3. Negativity $\{\partial h_i/\partial P_j\}_{i,j}$ is neg. semidef.

Notes:

- Slutsky or substitution matrix: $\{\frac{\partial h_i}{\partial P_j}\}_{i,j} = \left\{\frac{\partial g_i}{\partial X} Q_j + \frac{\partial g_i}{\partial P_j}\right\}_{i,j}$
- Engel curves describing quantities (or shares) as functions of total expenditure / income

Here, we will also focus on **separability** and **possible linearity**

An Almost Ideal Case

Set $u = \ln U$, $\boldsymbol{p} = \ln \boldsymbol{P}$
Assume log-cost function may be written as

$$\ln C^{AI}(\boldsymbol{p}, u) = f_1(\boldsymbol{p}) + f_2(\boldsymbol{p})u$$

e.g. $f_1(\boldsymbol{p} = \boldsymbol{p}'\boldsymbol{a}) + \frac{1}{2}\boldsymbol{p}'\boldsymbol{A}\boldsymbol{p}$ and $f_2(\boldsymbol{p}) = 1 + \boldsymbol{p}'\boldsymbol{b}$
Invert to get indirect utility as

$$V^{AI}(\boldsymbol{p}, x) = \frac{x - f_1(\boldsymbol{p})}{f_2(\boldsymbol{p})}$$

Then, the AI compensated expenditure-share system is

$$\omega^{AI}(\boldsymbol{p}, u) = \boldsymbol{a} + \boldsymbol{p}'\boldsymbol{A} + \boldsymbol{b}u$$

and the uncompensated expenditure-share system

$$\boldsymbol{w}^{AI}(\boldsymbol{p}, x) = \boldsymbol{a} + \boldsymbol{p}'\boldsymbol{A} + \boldsymbol{b}\frac{x - f_1(\boldsymbol{p})}{f_2(\boldsymbol{p})}$$

Starting from the Indirect Utility Model

Our indirect utility Model

Define **indirect utility** $V(\boldsymbol{p}, x)$ to give maximum utility attained by a consumer when faced with

- **log–prices** $\boldsymbol{p} = (p^1, \ldots, p^M)$
- **log–total expenditure** x

A **partially linear** indirect utility function

$$V(\boldsymbol{p}, x) = x - \boldsymbol{f}(x)^\top \boldsymbol{p} - \frac{1}{2}\boldsymbol{p}^\top \boldsymbol{A}\boldsymbol{p}$$

- $\boldsymbol{f} = (f^1, \ldots, f^M)^\top$ unknown differentiable functions of log–total expenditure
- $\boldsymbol{A} = \{a^{kl}\}_{k,l=1}^{M}$ parameters

Our indirect utility Model

Define **indirect utility** $V(\boldsymbol{p}, x)$ to give maximum utility attained by a consumer when faced with

- **log–prices** $\boldsymbol{p} = (p^1, \ldots, p^M)$
- **log–total expenditure** x

Extension to **varying coefficients**

$$V(\boldsymbol{p}, x) = x - \boldsymbol{f}(x)^\top \boldsymbol{p} - \frac{1}{2}\boldsymbol{p}^\top \boldsymbol{A}(x)\boldsymbol{p}$$

- $\boldsymbol{f} = (f^1, \ldots, f^M)^\top$, $\boldsymbol{A}(x) = \{a^{kl}(x)\}_{k,l=1}^M$ unknown differentiable functions of log–total expenditure

The Regression Model

With **Roy's identity**

$$w^k(\boldsymbol{p}, x) = -\frac{\partial V(\boldsymbol{p}, x)/\partial p^k}{\partial V(\boldsymbol{p}, x)/\partial x}$$

we get expenditure shares as functions of total expenditure and all prices

$$\boldsymbol{w}(\boldsymbol{p}, x) = \frac{\boldsymbol{f}(x) + \boldsymbol{A}\boldsymbol{p}}{1 - \nabla_x \boldsymbol{f}(x)^\top \boldsymbol{p}}$$

Rationality restrictions:

- Slutsky-symmetry if $\boldsymbol{A} = \boldsymbol{A}^\top$
- For homogeneity use $\tilde{x} = x - p^M$ and $\tilde{p}^k = p^k - p^M$ for all k
- Adding-up by construction $w^M(\tilde{\boldsymbol{p}}, \tilde{x}) = 1 - \sum_{k=1}^{M-1} w^k(\tilde{\boldsymbol{p}}, \tilde{x})$

Estimation

Consider $M-1$ expenditure share equations

$$\boldsymbol{w}(\tilde{\boldsymbol{p}}, \tilde{x}) = \frac{\boldsymbol{f}(\tilde{x}) + \boldsymbol{A}\tilde{\boldsymbol{p}}}{1 - \nabla_{\tilde{x}}\boldsymbol{f}(\tilde{x})^\top \tilde{\boldsymbol{p}}}$$

Basic idea:

- **Iteratively** solving minimization problems for nonparametric part (adapted kernel smoothing)
- **Symmetry-restricted** least squares for parametric coefficients
- Local-polynomial approximation

$$\boldsymbol{f}(t) \approx \boldsymbol{f}(\tilde{x}) + \nabla_{\tilde{x}}\boldsymbol{f}(\tilde{x})(t-\tilde{x}) \approx \boldsymbol{\alpha}(\tilde{x}) + \boldsymbol{\beta}(\tilde{x})(t-\tilde{x})$$

Estimation - Notes

The **local problem** is then

$$\min_{\alpha(\tilde{x}),\, \beta(\tilde{x}),\, \boldsymbol{A}} \sum_{i=1}^{N} \boldsymbol{e}_i^\top \boldsymbol{\Omega}\, \boldsymbol{e}_i$$

$$\boldsymbol{e}_i \quad \equiv \quad \boldsymbol{w}_i - \frac{\alpha(\tilde{x}) + (\tilde{x}_i - \tilde{x})\,\beta(\tilde{x}) + \boldsymbol{A}\tilde{\boldsymbol{p}}_i}{1 - \beta(\tilde{x})^\top \tilde{\boldsymbol{p}}_i}$$

with $(M-1) \times (M-1)$ weighting matrix $\boldsymbol{\Omega}$

Key idea:

- Local-polynomial model for numerator
- **Lower-order** local-polynomial in denominator
- Get starting values from reference group where $\tilde{\boldsymbol{p}} = \boldsymbol{0}$

Starting from the Log-Cost Model

Our log-cost Model

Redef. $\{W_i^1, ..., W_i^M, P_i^1, ..., P_i^M, X_i\}_{i=1}^n$ random vector giving the expenditure shares, log-prices, and log-expenditures

Extend the (homothetic) translog model to

$$\ln C(\boldsymbol{p}, u) = u + \boldsymbol{p}'\overline{\beta}(u) + \frac{1}{2}\boldsymbol{p}'\boldsymbol{A}\boldsymbol{p}$$

Dual indirect utility function is

$$u = V(\boldsymbol{p}, x) = x - \boldsymbol{p}'\overline{\beta}(u) - \frac{1}{2}\boldsymbol{p}'\boldsymbol{A}\boldsymbol{p}$$

cannot be solved for u analytically

Shephard's Lemma gives compensated expenditure share

$$\omega(\boldsymbol{p}, u) = \overline{\beta}(u) + \boldsymbol{A}\boldsymbol{p}$$

The (Almost) Observable Demand System

- Properties yield Restrictions: $\boldsymbol{\iota}'\overline{\boldsymbol{\beta}}(u) = 1$ and $\boldsymbol{A}'\boldsymbol{\iota} = \boldsymbol{0}_M$ are sufficient for homogeneity, $\boldsymbol{A} = \boldsymbol{A}'$ for symmetry
- re-scale prices s.th. $\overline{\boldsymbol{p}} = \boldsymbol{0}_M$, then $V(\overline{\boldsymbol{p}}, x) = x$
- log real expenditure, $x^R = R(\boldsymbol{p}, x)$, with reference $\overline{\boldsymbol{p}}$, then

$$V(\boldsymbol{p}, x) = V(\overline{\boldsymbol{p}}, x^R), \qquad x^R = R(\boldsymbol{p}, x) = \ln C(\overline{\boldsymbol{p}}, V(\boldsymbol{p}, x))$$

  what yields $R(\boldsymbol{p}, x) = V(\boldsymbol{p}, x)$.
- Thus, uncompensated shares can be defined by substituting x^R for u
- in compensated demand system:

$$\begin{aligned} \boldsymbol{w}(\boldsymbol{p}, x) &= \omega(\boldsymbol{p}, V(\boldsymbol{p}, x)) = \omega(\boldsymbol{p}, V(\overline{\boldsymbol{p}}, R(\boldsymbol{p}, x))) \\ &= \overline{\boldsymbol{\beta}}(V(\overline{\boldsymbol{p}}, R(\boldsymbol{p}, x))) + \boldsymbol{A}\boldsymbol{p} = \boldsymbol{\beta}\left(x^R\right) + \boldsymbol{A}\boldsymbol{p} \end{aligned}$$

Estimation of parametric part

Consider for each product j the sample

$$w_i^j - w_k^j = \beta^j(x_i^R) - \beta^j(x_k^R) + \boldsymbol{a}^j(\boldsymbol{p}_i - \boldsymbol{p}_k) + \epsilon_i^j - \epsilon_k^j \ , \forall\ i \neq k\ .$$

Weighting inversely to $|x_i^R - x_k^R|$ cancels β^j, and estimator is

$$\begin{aligned} \hat{\boldsymbol{A}}_{RSF} &= \hat{H}_{PP}^{-1}\hat{H}_{PW} \\ \hat{H}_{PW} &= \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{k=i+1}^{n-1} (\boldsymbol{p}_i - \boldsymbol{p}_k)(\boldsymbol{w}_i - \boldsymbol{w}_k)^T \hat{v}_{ik} \\ \text{and } \hat{H}_{PP} & \quad \text{analogously, where } \hat{v}_{ik} = K_h(\hat{x}_i^R - \hat{x}_k^R) \end{aligned}$$

$$\sqrt{n}\left(\hat{\boldsymbol{a}}_{RSF}^j - \boldsymbol{a}^j\right) \to N\left(0, E\left[\Sigma_{P|X^R}^{-1}\right] E\left[\boldsymbol{P}_X \sigma_{jj}(X, \boldsymbol{P}) \boldsymbol{P}_X'\right] E\left[\Sigma_{P|X^R}^{-1}\right]\right)$$

where $\boldsymbol{P}_X := \boldsymbol{P} - E[\boldsymbol{P}|X^R]$.

Estimation of the nonparametric part

Have in mind x^R is predicted, so make use of constructed regressors

As $\boldsymbol{A}$ is estimated with parametric rate, use ordinary loc.lin.

$$\hat{\theta}(x^R) = \operatorname{argmin} \sum_{i=1}^{n} \left\{ (w_i^j - \hat{\boldsymbol{a}}^j \boldsymbol{p}_i) - \theta_1 - \theta_2(\hat{x}_i^R - x^R) \right\}^2 K_h(\hat{x}_i^R - x^R)$$

Then we get

$$\sqrt{(nh \wedge ng_n)} \left\{ \hat{\boldsymbol{\beta}}(x^R) - \boldsymbol{\beta}(x^R) - B_\beta(x^R) \right\} \longrightarrow N\left(0, \Sigma_\beta(x^R)\right)$$

$$B_\beta(x^R) = \frac{h^2}{2} \mu_2(K) \boldsymbol{\beta}''(x^R) - B_X(x^0, \boldsymbol{p}^0) \boldsymbol{\beta}'(x^R)$$

where $\mu_l(K) = \int v^l K(v) dv$ and

$$\frac{1}{nh \wedge ng_n} \Sigma_\beta(x^R) = \frac{1}{nh} p^{-1}(x^R) \|K\|_2^2 \Sigma_\epsilon(x^R) \oplus \sigma_X^2(x^0, \boldsymbol{p}^0) \boldsymbol{\beta}'^2(x^R)$$

Varying Price Effects

If second-order price effects are not independent of utility:

$$\ln C(\boldsymbol{p}, u) = u + \boldsymbol{p}'\overline{\beta}(u) + \frac{1}{2}\boldsymbol{p}'\overline{\boldsymbol{A}}(u)\boldsymbol{p}$$

Indirect utility and compensated expenditure-shares are

$$\begin{aligned} u &= V(\boldsymbol{p}, x) = x - \boldsymbol{p}'\overline{\beta}(u) - \frac{1}{2}\boldsymbol{p}'\overline{\boldsymbol{A}}(u)\boldsymbol{p} \\ \omega(\boldsymbol{p}, u) &= \overline{\beta}(u) + \overline{\boldsymbol{A}}(u)\boldsymbol{p} \end{aligned}$$

Again, at base prices one has

$$\overline{\beta}(u) = \boldsymbol{\beta}\left(x^R\right) \;, \; \boldsymbol{A}\left(x^R\right) = \overline{\boldsymbol{A}}(u) = \overline{\boldsymbol{A}}(V(\boldsymbol{p}, x)) = \overline{\boldsymbol{A}}(V(\overline{\boldsymbol{p}}, x^R))$$

Therefore, we get $\boldsymbol{w}(\boldsymbol{p}, x^R) = \boldsymbol{\beta}\left(x^R\right) + \boldsymbol{A}(x^R)\boldsymbol{p}$

Combine consistency results on varying coeffs with those on generated regressors

Estimation of Model with varying Price Effects

Following Cleveland, Grosse, Shyu (1991), and Sperlich (2009)

$$\sum_{i=1}^{n} \left[W_i^j - \beta_0^j - \beta_1^j (\hat{x}_i^R - x_0^R) - \left\{ \boldsymbol{a}_0^j + \boldsymbol{a}_1^j (\hat{x}_i^R - x_0^R) \right\}' \boldsymbol{P}_i \right]^2 K_h(\hat{x}_i^R - x_0^R)$$

$$\widehat{\beta^j}(x_0^R) := \beta_0^j \ , \ \widehat{\boldsymbol{a}^j}(x_0^R) = (\hat{a}_1^j, \ldots, \hat{a}_M^j)'(x_0^R) := \boldsymbol{a}_0^j$$

V1 $E[(p^j)^{2s}] < \infty$ for some $s > 2$, $\forall j$. Second derivative of $r_{jk}(x^R) := E[p^j p^k | x^R]$ is cont. and bounded from zero

V2 Second derivatives of $\boldsymbol{A}(x^R)$ are cont. and bounded

Set $a_0^j(x^R) := \beta^j(x^R)$, $P_i^0 \equiv 1$ for all i

Set $\alpha_k := \left(a_0^k, a_1^k, \cdots, a_M^k\right)'$ for $k = 1, \ldots, M$. Then it holds

$$\sqrt{(nh \wedge ng_n)}\left\{\hat{\alpha}_k - \alpha_k - B_k(x^R)\right\} \longrightarrow N\left(0, \Sigma_{\alpha_k}(x^R)\right)$$

with

$$B_k(x^R) = \frac{h^2}{2}\mu_2(K)\alpha_k'' - B_X(x^0, p^0)\alpha_k'$$

The covariance structure is given by

$$\frac{1}{nh}p^{-1}(x^R)\|K\|_2^2 \Omega \Sigma_{\epsilon k,k}(x^R) \oplus \sigma_X^2(x^0, p^0)(\alpha_k')^2$$

respectively by

$$\frac{1}{nh}p^{-1}(x^R)\|K\|_2^2 \Omega_{j,j} \Sigma_{\epsilon}(x^R) \oplus \sigma_X^2(x^0, p^0){\gamma'}_j^2$$

where $\Omega^{-1} := E\left[(P^0, P^1, \ldots, P^M)'(P^0, P^1, \ldots, P^M)|x^R\right]$
and $\gamma_j = (a_j^1, a_j^2, \ldots, a_j^M)$, $j = 0, \ldots, M$

Consistent initial estimator for $\hat{x}_i^R$

- Def. *log nominal expenditure*, $x^N = N(\boldsymbol{p}, x)$ as level of expend. at $\boldsymbol{p}$ which yields same level of utility as x at $\overline{\boldsymbol{p}}$.
- Again, is implicitly defined by

$$x = V(\overline{\boldsymbol{p}}, x) = V(\boldsymbol{p}, x^N) = x^N - \boldsymbol{p}'\overline{\beta}\{V(\overline{\boldsymbol{p}, x})\} - \frac{1}{2}\boldsymbol{p}'\boldsymbol{A}\boldsymbol{p}$$

$$\Longleftrightarrow \qquad x^N = N(\boldsymbol{p}, x) = x + \boldsymbol{p}'\beta(x) + \frac{1}{2}\boldsymbol{p}'\boldsymbol{A}\boldsymbol{p}$$

  Further, note $x^R = R(\boldsymbol{p}, x) = N^{-1}(\boldsymbol{p}, x)$.
- Monotonic increasing costs in utility give monotonic increase of $R(\boldsymbol{p}, x)$ and $N(\boldsymbol{p}, x)$ in x for each $\boldsymbol{p}$, i.e. we can invert N. Further, for each $\boldsymbol{p}$ fixed, and $t = \hat{N}(\boldsymbol{p}, x)$, $\hat{R}(\boldsymbol{p}, t) = \hat{N}^{-1}(\boldsymbol{p}, t)$,

$$\sup_t |\hat{R}(\boldsymbol{p}, t) - R(\boldsymbol{p}, t)| \quad \leq \quad \sup_t |\frac{d}{dt}R(\boldsymbol{p}, t)| \sup_v |N(\boldsymbol{p}, v) - \hat{N}(\boldsymbol{p}, v)|$$

  so initial estimate for function N would do

Initial Estimators for β and $\boldsymbol{A}$

- Recall that for $\overline{\boldsymbol{p}} = \mathbf{0}_M$ we have $x^R = R(\overline{\boldsymbol{p}}, x) = x$, s.th.

$$E[\boldsymbol{W}_i | X_i = x, \boldsymbol{P}_i = \overline{\boldsymbol{p}}] = \beta(x) = \beta(x^R)$$

  Use smoother for people facing $\overline{\boldsymbol{p}}$ (or including *neighbors*)

- $\boldsymbol{A}$ is matrix of log-price derivatives of compensated expenditure share eqns, i.e. of compensated semi-elasticities. In general, can be expressed in terms of observables:

$$\Upsilon(\boldsymbol{p}, x) = \boldsymbol{\nabla}_{pp} \boldsymbol{w}(\boldsymbol{p}, x) + \boldsymbol{\nabla}_x \boldsymbol{w}(\boldsymbol{p}, x) \boldsymbol{w}(\boldsymbol{p}, x)'$$

  Therefore, a consistent estimator for $\boldsymbol{A}$ is given by

$$\hat{\boldsymbol{A}}_0 = \frac{1}{n} \sum_{i=1}^{n} \widehat{\Upsilon}(\boldsymbol{P}_i, X_i)$$

  with estimating $\boldsymbol{w}(\boldsymbol{p}, x)$ and its derivatives nonparametrically

Inference and Restrictions

- Easy to impose homogeneity,
- straight forward to impose symmetry,
- but hard to guarantee concavity / negativity without overdoing.
- Restricted estimators provide directly specification tests, usually based on bootstrap or subsampling.
- In application rejected for example symmetry but - different to parametric models - could analyze why !
- Typical criticism

  Hard to implement and calculate

  Dependence on bandwidth choice
- Extensions

  IV methods for problems of endogeneity