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 versa.
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 $R^2$s 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

\[
R^2_V = 1 - \frac{\sum_t \{ Y_t - \hat{g}_t \}^2}{\sum_t \{ Y_t - \bar{Y}_t \}^2},
\]

Properties:

- \( R^2_V \in (-\infty, 1] \) where \( R^2_V < 0 \) 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 \( R^2_V < 0 \))
Can we beat the historical mean?

- Parametric null hypothesis vs. non-/semiparametric alternative

\[ H_0: Y_t = \bar{Y} + \xi_t \quad \text{vs.} \quad H_1: Y_t = g(X_{t-1}) + \varepsilon_t \]

- Construct \( B \) bootstrap samples \( \{Y^b_1, \ldots, Y^b_T\} \) with residuals under the null:

\[
Y^b_t = Y_t + \hat{\varepsilon}^0_t \cdot u^b_t, \quad \hat{\varepsilon}^0_t = Y_t - \hat{g}_{-t}
\]

with iid zero-mean variance-one rv \( u^b_t \).

- In each bootstrap iteration \( b \) calculate \( R^2_{V} \)

- Determine quantiles of empirical distribution of \( R^2_{V} \) under the Null:

\[
F^*(u) = \frac{1}{B} \sum_b 1_{\{R^2_{V} \leq u\}}
\]
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
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)
- 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: $R^2_V$ and p-values

<table>
<thead>
<tr>
<th></th>
<th>$S$</th>
<th>$d$</th>
<th>$e$</th>
<th>$r$</th>
<th>$L$</th>
<th>$\text{inf}$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>par</td>
<td>-1.0</td>
<td>1.0</td>
<td>8.0</td>
<td>2.7</td>
<td>-1.1</td>
<td>-1.4</td>
<td>-0.4</td>
</tr>
<tr>
<td>nonpar</td>
<td>-1.2</td>
<td>0.9</td>
<td>11.8</td>
<td>2.5</td>
<td>-0.8</td>
<td>-1.6</td>
<td>-0.7</td>
</tr>
<tr>
<td></td>
<td>(0.596)</td>
<td>(0.193)</td>
<td>(0.005)</td>
<td>(0.079)</td>
<td>(0.571)</td>
<td>(0.759)</td>
<td>(0.573)</td>
</tr>
</tbody>
</table>

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

- OLS and local-linear kernel-regression
- Only earnings and risk-free with predictive power
- Factor 1.5 increase for earnings
One-dimensional case - graphs

S.Sperlich (Université de Genève)  Structure guided by prior knowledge  April 14, 2011  16 / 51
Second Step: Two-dimensional

<table>
<thead>
<tr>
<th></th>
<th>e, S</th>
<th>e, d</th>
<th>e, r</th>
<th>e, L</th>
<th>e, inf</th>
<th>e, b</th>
</tr>
</thead>
<tbody>
<tr>
<td>par</td>
<td>6.8</td>
<td>6.9</td>
<td>12.2</td>
<td>7.3</td>
<td>9.2</td>
<td>8.8</td>
</tr>
<tr>
<td>nonpar</td>
<td>8.5</td>
<td>12.6</td>
<td>13.7</td>
<td>11.0</td>
<td>11.0</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.000)</td>
<td>(0.004)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

\[ Y_t = \beta_0 + \beta^\top X_{t-1} + \varepsilon_t \quad \text{vs.} \quad Y_t = g(X_{t-1}) + \xi_t \]

- **Par:** Improved prediction with \textbf{extra} variable \textit{inf}, \textit{b}, \textit{r}
  
  Model \{e, r\} even better than one-dim. nonpar.

- **Nonpar:** \textbf{All} shown models beat significantly historical mean and seem to \textbf{improve} prediction compared to 2-dim par.

- Increase in predictive power of "only" \textbf{12\%} compared to first step
Second Step: Two-dim. guided by prior

**Table:** Predictive power: $R_V^2$

<table>
<thead>
<tr>
<th></th>
<th>$e, S$</th>
<th>$e, d$</th>
<th>$e, r$</th>
<th>$e, L$</th>
<th>$e, \text{inf}$</th>
<th>$e, b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonpar</td>
<td>8.5</td>
<td>12.6</td>
<td>13.7</td>
<td>11.0</td>
<td>11.0</td>
<td>11.3</td>
</tr>
<tr>
<td>prior</td>
<td>6.6</td>
<td>13.5</td>
<td>12.1</td>
<td>12.8</td>
<td>9.5</td>
<td>8.0</td>
</tr>
</tbody>
</table>

- 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>
<thead>
<tr>
<th></th>
<th>S</th>
<th>d</th>
<th>e</th>
<th>r</th>
<th>L</th>
<th>inf</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>8.8</td>
<td>7.6</td>
<td>9.3</td>
<td>15.8</td>
<td>10.7</td>
<td>11.4</td>
<td>11.8</td>
</tr>
<tr>
<td>e, L</td>
<td>9.9</td>
<td>13.1</td>
<td>14.2</td>
<td>18.5</td>
<td>13.3</td>
<td>13.3</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Table: Predictive power: $R^2_V$

- 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

- **Risk-free: 12.0**
- **Earnings by price: 0.03**
Working with Predicted Factors
The prediction framework

- The excess stock return:

\[ S_t = \log\left(\frac{(P_t + D_t) / P_{t-1}}{1 - r_{t-1}}\right) \]

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

\[ r_t = \log\left(1 + \frac{R_t}{100}\right) \]

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

   \[ Y_t = g(\hat{b}_t, w_{t-1}) + \epsilon_t. \]

   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

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

- 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 $\alpha$-mixing): as had we observed the real bond

**Theorem**

$$| g_{LL}(\hat{b}_t) - \tilde{g}(b_t) | \to 0$$
Why using constructed regressors in Statistics?

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

\[ L_{t-1} \rightarrow \hat{b}_t \]

- 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 $R^2_V$–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: $R^2_V$ of different **Bond** models 1923–1996

<table>
<thead>
<tr>
<th>$v_{t-1}$</th>
<th>S</th>
<th>L</th>
<th>r</th>
<th>S,r</th>
<th>S,r,b</th>
</tr>
</thead>
<tbody>
<tr>
<td>par</td>
<td>11.6%</td>
<td>24.0%</td>
<td>22.3%</td>
<td>33.1%</td>
<td>37.4%</td>
</tr>
<tr>
<td>nonpar</td>
<td>16.3%</td>
<td>23.9%</td>
<td>26.8%</td>
<td>33.0%</td>
<td>37.4%</td>
</tr>
</tbody>
</table>

- 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: $R^2_V$ of different Stock models 1923–1996

<table>
<thead>
<tr>
<th>$w_{t-1}$ \ $v_t$</th>
<th>d</th>
<th>r</th>
<th>b</th>
<th>d,L</th>
<th>d,r</th>
<th>r,b</th>
</tr>
</thead>
<tbody>
<tr>
<td>par</td>
<td>-6.3%</td>
<td>-5.7%</td>
<td>-4.0%</td>
<td>-5.8%</td>
<td>-7.2%</td>
<td>0.5%</td>
</tr>
<tr>
<td>nonpar</td>
<td>-1.4%</td>
<td>-3.6%</td>
<td>5.9%</td>
<td>-6.0%</td>
<td>-7.4%</td>
<td>-8.6%</td>
</tr>
<tr>
<td>$\hat{b}_t$</td>
<td>8.3%</td>
<td>1.4%</td>
<td>10.6%</td>
<td>-3.8%</td>
<td>2.9%</td>
<td>-3.6%</td>
</tr>
<tr>
<td>$\hat{b}<em>t$, $v</em>{t-1}$</td>
<td>13.9%</td>
<td>16.3%</td>
<td>8.9%</td>
<td>28.3%</td>
<td>21.6%</td>
<td>20.3%</td>
</tr>
</tbody>
</table>

- All parametric models with **negative** $R^2_V$
- Very good results in general for **diagonal** $w_{t-1} = v_{t-1}$
- **Improvement** of prediction from $R^2_V = 5.9\%$ to
  - $R^2_V = 28.9\%$ for $\hat{g}(\hat{b}_t, d_{t-1}, S_{t-1}, L_{t-1})$ and $\hat{b}_t = \hat{p}(d_{t-1}, L_{t-1})$
  - $R^2_V = 30.3\%$ 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
for utility $U$, nominal total expenditures $X$, prices $P$, quantities $Q$, shares $W$

Consider the consumer problem:

1. $\text{Max } U = v(Q)$ subject to $P'Q = X$
2. $\text{Min } X = P'Q$ subject to $v(Q) = U$

with solution: $Q_i = g_i(X, P) = h_i(U, P)$, $i = 1, \ldots, M$

1. the Marshallian (or uncompensated) demands and
2. the Hicksian (or compensated) demands respectively

substituting into original problems gives

1. the indirect utility function $U = V(X, P)$, and
2. the cost function $X = C(U, P)$ respectively
Some Properties of the cost function

1. homogeneous in prices, i.e. \( C(U, \theta P) = \theta C(U, P) \forall \theta > 0 \)

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

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

called Roy’s identity, and for budget shares

\[
w_i = \frac{\partial \ln C(U, P)}{\partial \ln P_i} = \frac{\partial \ln V/\partial \ln P_i}{\sum \partial \ln V/\partial \ln P_k}
\]
Properties of Integrability

Often one concentrates on

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

Notes:

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

Here, we will also focus on \textbf{separability} and \textbf{possible linearity}
An Almost Ideal Case

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

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

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

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

Then, the AI compensated expenditure-share system is

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

and the uncompensated expenditure-share system

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

Starting from the Indirect Utility Model
Our indirect utility Model

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

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

A partially linear indirect utility function

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

- \( f = (f^1, \ldots, f^M)^\top \) unknown differentiable functions of log–total expenditure
- \( A = \{a_{kl}\}_{k,l=1}^M \) parameters
Our indirect utility Model

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

- log–prices \( p = (p^1, \ldots, p^M) \)
- log–total expenditure \( x \)

Extension to **varying coefficients**

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

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

With Roy’s identity

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

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

\[ w(p, x) = \frac{f(x) + Ap}{1 - \nabla_x f(x)^\top p} \]

Rationality restrictions:

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

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

Basic idea:

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

$$f(t) \approx f(\tilde{x}) + \nabla_{\tilde{x}}f(\tilde{x})(t - \tilde{x}) \approx \alpha(\tilde{x}) + \beta(\tilde{x})(t - \tilde{x})$$
The local problem is then

\[
\min_{\alpha(\tilde{x}), \beta(\tilde{x}), A} \sum_{i=1}^{N} e_i^T \Omega e_i
\]

\[
e_i \equiv w_i - \frac{\alpha(\tilde{x}) + (\tilde{x}_i - \tilde{x}) \beta(\tilde{x}) + A \tilde{p}_i}{1 - \beta(\tilde{x})^T \tilde{p}_i}
\]

with \((M - 1) \times (M - 1)\) weighting matrix \(\Omega\)

Key idea:

- Local-polynomial model for numerator
- Lower-order local-polynomial in denominator
- Get starting values from reference group where \(\tilde{p} = 0\)
Starting from the Log-Cost Model
Redef. \{W^1_i, \ldots, W^M_i, P^1_i, \ldots, P^M_i, X_i\}_{i=1}^n \text{ random vector giving the expenditure shares, log-prices, and log-expenditures}

Extend the (homothetic) translog model to

\[
\ln C(p, u) = u + p'\bar{\beta}(u) + \frac{1}{2}p'A p
\]

Dual indirect utility function is

\[
u = V(p, x) = x - p'\bar{\beta}(u) - \frac{1}{2}p'A p
\]

cannot be solved for \(u\) analytically

Shephard’s Lemma gives compensated expenditure share

\[
\omega(p, u) = \bar{\beta}(u) + Ap
\]
Properties yield Restrictions: $\iota'\beta'(u) = 1$ and $A'\iota = 0_M$ are sufficient for homogeneity, $A = A'$ for symmetry.

re-scale prices s.th. $\bar{p} = 0_M$, then $V(\bar{p}, x) = x$

log real expenditure, $x^R = R(\mathbf{p}, x)$, with reference $\bar{p}$, then

$$V(\mathbf{p}, x) = V(\bar{p}, x^R), \quad x^R = R(\mathbf{p}, x) = \ln C(\bar{p}, V(\mathbf{p}, x))$$

what yields $R(\mathbf{p}, x) = V(\mathbf{p}, x)$.

Thus, uncompensated shares can be defined by substituting $x^R$ for $u$.

in compensated demand system:

$$w(\mathbf{p}, x) = \omega(\mathbf{p}, V(\mathbf{p}, x)) = \omega(\mathbf{p}, V(\bar{p}, R(\mathbf{p}, x)))$$

$$= \bar{\beta}(V(\bar{p}, R(\mathbf{p}, x))) + Ap = \beta(x^R) + Ap$$

what yields $R(\mathbf{p}, x) = V(\mathbf{p}, x)$. Thus, uncompensated shares can be defined by substituting $x^R$ for $u$.
Consider for each product $j$ the sample

$$w_i^j - w_k^j = \beta^j(x_i^R) - \beta^j(x_k^R) + a^j(p_i - p_k) + \epsilon_i^j - \epsilon_k^j, \forall \ i \neq k.$$  

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

$$\hat{A}_{RSF} = \hat{H}_{PP}^{-1} \hat{H}_{PW}$$

$$\hat{H}_{PW} = \left( \begin{array}{c} n \end{array} \right)^{-1} \sum_{i=1}^{n} \sum_{k=i+1}^{n-1} (p_i - p_k)(w_i - w_k)^T \hat{v}_{ik}$$

and $\hat{H}_{PP}$ analogously, where $\hat{v}_{ik} = K_h(\hat{x}_i^R - \hat{x}_k^R)$

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


S.Sperlich (Université de Genève) Structure guided by prior knowledge April 14, 2011 44 / 51
Estimation of the nonparametric part

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

As $A$ is estimated with parametric rate, use ordinary loc.lin.

$$
\hat{\theta}(x^R) = \arg\min \sum_{i=1}^{n} \left\{ (w^j_i - \hat{a}^j_i p^i) - \theta_1 - \theta_2(\hat{x}_i^R - x^R) \right\}^2 K_h(\hat{x}_i^R - x^R)
$$

Then we get

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

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

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

$$
\frac{1}{nh \wedge ng_n} \Sigma_\beta(x^R) = \frac{1}{nh} \rho^{-1}(x^R) \|K\|_2^2 \Sigma_\epsilon(x^R) \oplus \sigma^2_X(x^0, p^0) \beta'^2(x^R)
$$
Varying Price Effects

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

\[
\ln C(p, u) = u + p'\bar{\beta}(u) + \frac{1}{2}p'\bar{A}(u)p
\]

Indirect utility and compensated expenditure-shares are

\[
u = V(p, x) = x - p'\bar{\beta}(u) - \frac{1}{2}p'\bar{A}(u)p
\]

\[
\omega(p, u) = \bar{\beta}(u) + \bar{A}(u)p
\]

Again, at base prices one has

\[
\bar{\beta}(u) = \beta(x^R), \quad \bar{A}(x^R) = \bar{A}(u) = \bar{A}(V(p, x)) = \bar{A}(V(\bar{p}, x^R))
\]

Therefore, we get

\[
w(p, x^R) = \beta(x^R) + A(x^R)p
\]

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)

\[
\sum_{i=1}^{n} \left[ W_i^j - \beta_0^j - \beta_1^j (\hat{x}_i^R - x_0^R) - \{ a_0^j + a_1^j (\hat{x}_i^R - x_0^R) \} \right] \cdot P_i \right] \cdot 2 \cdot K_h (\hat{x}_i^R - x_0^R) \\
\]

\[ \hat{\beta}^j (x_0^R) := \beta_0^j , \quad \hat{a}^j (x_0^R) = (\hat{a}_1^j, \ldots, \hat{a}_M^j)' (x_0^R) := a_0^j \]

**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 \( A(x^R) \) are cont. and bounded

Set \( a_0^j (x^R) := \beta^j (x^R) , \quad P_i^0 \equiv 1 \) for all  \( i \)
Set $\alpha_k := (a_{0}^{k}, a_{1}^{k}, \ldots, a_{M}^{k})'$ for $k = 1, \ldots, M$. Then it holds

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

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 $\Omega^{-1} := E[(P^0, P^1, \ldots, P^M)'(P^0, P^1, \ldots, P^M)|x^R]$ and $\gamma_j = (a_j^1, a_j^2, \ldots, a_j^M), j = 0, \ldots, M$
Consistent initial estimator for $\hat{\chi}_i^R$

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

- Again, is implicitly defined by

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

$$\iff \quad x^N = N(p, x) = x + p'\beta(x) + \frac{1}{2}p'A p$$

Further, note $x^R = R(p, x) = N^{-1}(p, x)$.

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

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

so initial estimate for function $N$ would do
Initial Estimators for $\beta$ and $A$

- Recall that for $\bar{p} = 0_M$ we have $x^R = R(\bar{p}, x) = x$, s.th.
  \[ E[W_i | X_i = x, P_i = \bar{p}] = \beta(x) = \beta(x^R) \]

  Use smoother for people facing $\bar{p}$ (or including neighbors)

- $A$ 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:
  \[ \Upsilon(p, x) = \nabla_{pp} w(p, x) + \nabla_x w(p, x) w(p, x)' \]

  Therefore, a consistent estimator for $A$ is given by
  \[ \hat{A}_0 = \frac{1}{n} \sum_{i=1}^{n} \Upsilon(P_i, X_i) \]

  with estimating $w(p, x)$ and its derivatives nonparametrically.
Inference and Restrictions

- Easy to impose homogeneity,
- straightforward 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