Absolute Penalty and Shrinkage Estimation in Partially Linear Models

S. E. Ahmed

University of Windsor

seahmed@uwindsor.ca

Conference on Nonparametric Statistics and Statistical Learning
The Ohio State University, Columbus, OH
May 19 - 22, 2010

With K. Doksum (Wisconsin) and E. Raheem (Windsor)
Outline

Motivation
Partially linear model and parameter estimation
Shrinkage Estimation
Simulation Study
Application to a real life data
Comparison of Shrinkage with APE
Summary
[Mroz (Econometrica, 1987)] used a sample of 1975 PSID\(^1\) labour supply data to systematically study several theoretic and statistical assumptions used in many empirical models of female labour supply. This data set was used in numerous Econometric studies and textbooks.

- [Wooldridge, 2003] used this data in his textbook to demonstrate many applications of regression models.
- [Long, 1997] fitted parametric logistic regression model to this data.
- [Fox, 2002] used this data for a semiparametric logistic regression.

\(^1\) Panel Study on Income Dynamics (PSID), University of Michigan. The data is in public domain and freely available from http://ideas.repec.org/s/boc/bocins.html
[Fox, 2005] commented that semiparametric model may be used wherever there is reason to believe that one or more covariates enter the regression linearly. This could be known from prior studies, or there are prior reasons to believe so (although rare), or examination of the data might suggest a linear relationship for some covariates. A more general scenario is when some of the covariates are categorical and they enter in the model as dummy variables.

[Engle (JASA, 1986)] considered such a PLM where demand for electricity was the outcome variable and four regions were entered as dummy variable while temperature was modelled nonparametrically.
Motivated by [Engle (JASA, 1986)] and [Fox, 2002], we fit a semiparametric regression model with a continuous outcome variable and demonstrate shrinkage estimation using PSID data.

We first show that semiparametric modeling is appropriate in our case.
The female labour supply data consist of 753 observations on 19 variables. Data were collected from married white women between the ages 30 and 60 in 1975. Of them, 428 were working at some time during the year 1975.

Similar to [Mroz (Econometrica, 1987)], we consider wife’s annual hours of work (hours) as the response variable. This variable is the product of the number of weeks the wife worked for money in 1975 and the average number of hours of work per week during the weeks she worked.
Model Building Strategy

Full Model

\[
\text{hours} = \text{inlf} + \text{age} + \text{huseduc} + \text{exper} + \text{nwifeinc} + \text{mtr} \\
+ \text{k5} + \text{k618} + \text{educ} + \text{husage} + \text{hushrs} + \text{motheduc} \\
+ \text{fatheduc} + \text{unem} + \text{city}
\]  (1)

where, hours = wife’s annual hours of work, inlf = 1 if in labour force in 1975, age = age of wife, huseduc = husband’s education, exper = actual labour market experience, nwifeinc = nonwife income in the family, mtr = marginal tax rate, k5 = kids less than 6 years of age, k618 = kids between 6 and 18 years, husage = husband’s age, hushrs = hours worked by husband in 1975, fatheduc, motheduc = father and mother’s years of schooling, unem = unemployment rate of country of residence, city = 1 if living in a city.
An initial stepwise fitting based on AIC criterion gave us the following model

**Reduced (working) Model**

\[
\text{hours} = \text{inlf} + \text{age} + \text{huseduc} + \text{exper} + \text{nwifeinc} + \text{mtr} + k5
\] (2)
We start with model (2).
Each of the linear terms except inlf, k5 in model (2) is expanded by B-spline basis and tested for nonlinearity with hours.
The deviances and residual degrees of freedom for candidate models are listed Table 1.
Test results for nonlinearity is shown in Table 2.
Table 1: Deviance table for various models fitted with \textit{mroz} data.

<table>
<thead>
<tr>
<th>Model</th>
<th>inlf</th>
<th>k5</th>
<th>age</th>
<th>huseudc</th>
<th>nwifeinc</th>
<th>mtr</th>
<th>exper</th>
<th>Deviance (000,000)</th>
<th>df (res)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>F</td>
<td>F</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>214.53</td>
<td>743</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>F</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>211.51</td>
<td>737</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>F</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>207.55</td>
<td>736</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>F</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>210.81</td>
<td>737</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>F</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>212.47</td>
<td>737</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>F</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>213.02</td>
<td>737</td>
</tr>
</tbody>
</table>

Code: F = factor (dummy), L = linear, S = smoothed term.

- We pick Model 2 and Model 3 based on deviances.
Table 2: Test of nonlinearity of exper, mtr, nwifeinc, huseduc, and age,

<table>
<thead>
<tr>
<th>Model contrasted</th>
<th>Predictor</th>
<th>Difference in deviance (000,000)</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1</td>
<td>exper</td>
<td>3.02</td>
<td>1.75</td>
<td>6</td>
<td>0.1041</td>
</tr>
<tr>
<td>0–2</td>
<td>mtr</td>
<td>6.97</td>
<td>3.53</td>
<td>7</td>
<td>0.0008</td>
</tr>
<tr>
<td>0–3</td>
<td>nwifeinc</td>
<td>3.72</td>
<td>2.17</td>
<td>6</td>
<td>0.0427</td>
</tr>
<tr>
<td>0–4</td>
<td>huseduc</td>
<td>2.06</td>
<td>1.19</td>
<td>6</td>
<td>0.3060</td>
</tr>
<tr>
<td>0–5</td>
<td>age</td>
<td>1.51</td>
<td>0.87</td>
<td>6</td>
<td>0.5141</td>
</tr>
</tbody>
</table>

- mtr and nwifeinc are significant with p-values 0.0008 and 0.04 respectively.
We further test the significance of each of the predictors in Model 2 by fitting additional models:

**Table 3:** Deviance table for additional models (based on Model 2) to test for significance of each of the individual predictors.

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>Deviance (000,000)</th>
<th>df (res)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-2 (base)</td>
<td>inlf F k5 F educ L husage L nwifeinc L mtr S exper L</td>
<td>207.68</td>
<td>737</td>
</tr>
<tr>
<td>6</td>
<td>F – F L L L S L</td>
<td>337.23</td>
<td>738</td>
</tr>
<tr>
<td>7</td>
<td>F F L L L S L</td>
<td>209.12</td>
<td>740</td>
</tr>
<tr>
<td>8</td>
<td>F F – L L S L</td>
<td>210.32</td>
<td>738</td>
</tr>
<tr>
<td>9</td>
<td>F F L – L S L</td>
<td>213.14</td>
<td>738</td>
</tr>
<tr>
<td>10</td>
<td>F F L L – S L</td>
<td>223.05</td>
<td>738</td>
</tr>
<tr>
<td>11</td>
<td>F F L L – L –</td>
<td>237.79</td>
<td>744</td>
</tr>
<tr>
<td>6</td>
<td>F F – L L S –</td>
<td>239.55</td>
<td>739</td>
</tr>
</tbody>
</table>

- Model 2 has the smallest deviance.
- Next Table shows the tests for significance.
Table 4: Test for the significance of inlf, k5, age, huseduc, nwifeinc, mtr, and exper based on Model 2.

<table>
<thead>
<tr>
<th>Model contrasted</th>
<th>Predictor</th>
<th>Difference in deviance (000,000)</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2–6</td>
<td>inlf</td>
<td>129.68</td>
<td>229.93</td>
<td>2</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>2–7</td>
<td>k5</td>
<td>1.57</td>
<td>1.39</td>
<td>4</td>
<td>0.2333</td>
</tr>
<tr>
<td>2–8</td>
<td>age</td>
<td>2.76</td>
<td>4.91</td>
<td>2</td>
<td>0.0074</td>
</tr>
<tr>
<td>2–9</td>
<td>huseduc</td>
<td>5.59</td>
<td>9.91</td>
<td>2</td>
<td>4.97e-5</td>
</tr>
<tr>
<td>2–10</td>
<td>nwifeinc</td>
<td>15.49</td>
<td>27.48</td>
<td>2</td>
<td>1.16e-12</td>
</tr>
<tr>
<td>2–11</td>
<td>mtr</td>
<td>30.23</td>
<td>13.40</td>
<td>8</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>2–12</td>
<td>exper</td>
<td>31.99</td>
<td>37.82</td>
<td>3</td>
<td>&lt; 2.2e-16</td>
</tr>
</tbody>
</table>

- All the predictors are significant at the nominal level except k5, which is insignificant (p−value = 0.2333)
- We used `anova.gam` in `mgcv` package in R to conduct the tests.
Figure 1: (a) nonlinear relationship; (b) contour plot; (c, d) smoothed curve for nwifeinc and mtr by uniform B-spline basis of order 9 and 8 respectively. Shaded regions in (c) and (d) are 95% confidence envelopes.
Therefore, we think, it is sensible to estimate \( \text{nwifeinc} \) and \( \text{mtr} \) using smoothing method while the other variables may enter linearly. Thus, the best fitting models are:

**mtr entering nonlinearly**

\[
\text{hours} = \beta_0 + \beta_1 \text{inlf} + \beta_2 \text{age} + \beta_3 \text{huseduc} + \beta_4 \text{exper} + \beta_5 \text{nwifeinc} + \theta \ g(\text{mtr})
\] (3)

**nwifeinc entering nonlinearly**

\[
\text{hours} = \beta_0 + \beta_1 \text{inlf} + \beta_2 \text{age} + \beta_3 \text{huseduc} + \beta_4 \text{exper} + \beta_5 \text{mtr} + \theta \ g(\text{nwifeinc})
\] (4)
A partially linear regression model:

\[ y_i = x_i'\beta + g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \]  

where \( y_i \)'s are responses, 
\( x_i = (x_{i1}, \ldots, x_{ip})' \) and \( t_i \in [0, 1] \) are design points, 
\( \beta = (\beta_1, \ldots, \beta_p)' \) is an unknown parameter vector, 
\( g(\cdot) \) is an unknown real-valued function defined on \([0, 1]\), and 
\( \varepsilon_i \)'s are unobservable random errors.
Related earlier work

Ahmed et al. (2007) considered kernel estimates of $g(\cdot)$ to construct absolute penalty, shrinkage and pretest estimators of $\beta$ in the case where $\beta = (\beta_1', \beta_2')'$,

- where $\beta_1'$ is a vector of principle parameter and $\beta_2'$ is a vector of nuisance parameters.
Objective

- We consider estimating $\beta$ based on $g(\cdot)$ approximated by a B-spline series.
Let $k$ be an integer larger than or equal to $\nu$ where $\nu$ will be defined in the Assumption (2)

Further, let $S_{m_n,k}$ be the class of functions $s(\cdot)$ on $[0, 1]$ with the following properties:

(i) $s(\cdot)$ is a polynomial of degree $k$ on each of the sub-intervals $\left[\frac{(i-1)}{m_n}, \frac{i}{m_n}\right]$, $i = 1, \ldots, m_n$, where $m_n$ is a positive integer which depends on $n$.

(ii) $s(\cdot)$ is $(k - 1)$ times differentiable.

Then, $S_{m_n,k}$ is called the class of all splines of degree $k$ with $m_n$-equispaced knots.
Consequently,

- $S_{m_n,k}$ has a basis of $m_n + k$ normalized B-spline
  \[ \{ B_{mnj}(\cdot) : j = 1, \ldots, m_n + k \} , \text{ and} \]

- $g(\cdot)$ can be approximated by a linear combination $\theta' B_{mn}(\cdot)$ of
  the basis, where $\theta \in \mathcal{R}^{m_n+k}$ and
  \[ B_{mn}(\cdot) = (B_{m_{n1}}(\cdot), \ldots, B_{m_{m_n},m_{n+k}}(\cdot))' . \]

Therefore, replacing $g(\cdot)$ by $B_{mn}(\cdot)\theta$, in model (5) we get

\[
y = x\beta + B_{mn}(t)\theta + \varepsilon . \quad (6)
\]
Full Model Estimation

\[ (\hat{\beta}, \hat{\theta}) \] is obtained by minimizing

\[ S_n(\beta, \theta) = n^{-1} \sum_{i=1}^{n} [y_i - x_i'\beta - \theta'B_{mn}(t_i)]^2, \quad (7) \]

which gives

\[ \hat{\beta} = (X'M_{B_mn}X)^{-1}X'M_{B_mn}Y \]

and

\[ \hat{\theta} = (B'_{m_n}B_{m_n})^{-1}B'_{m_n}(Y - X\hat{\beta}), \]

where, \( Y = (y_1, \ldots, y_n)' \), \( X = (x_1, \ldots, x_p) \),
\( x_s = (x_{1s}, \ldots, x_{ns})' \), \( s = 1, \ldots, p \),
\( M_{B_mn} = I - B_{m_n}(B'_{m_n}B_{m_n})^{-1}B'_{m_n} \) and
\( B_{m_n} = (B_{m_n}(t_1), \ldots, B_{m_n}(t_n))' \).

The estimator \( \hat{\beta} \) is called a semiparametric least squares estimator (SLSE) of \( \beta \).
\( \beta \) in the linear part can be partitioned as \((\beta_1, \beta_2)\).

- \( \beta_1 \) is the coefficient vector for main effects (e.g., treatment effect, genetic effects) and \( \beta_2 \) is a vector for “nuisance” effects (e.g., age, laboratory).
SURE $\hat{\beta}_1$ of $\beta_1$ is

$$
\hat{\beta}_1 = (X_1' M_{B_{mn}} M_{B_{mn}} x_2 M_{B_{mn}} X_1)^{-1} X_1' M_{B_{mn}} M_{B_{mn}} x_2 M_{B_{mn}} Y,
$$

where

- $X_1$ is composed of the first $p_1$ column vectors of $X$,
- $X_2$ is composed of the last $p_2$ column vectors of $X$, and
- $M_{B_{mn}} x_2 = I - B_{mn} X_2 (X_2' B_{mn} B_{mn} X_2)^{-1} X_2' B_{mn}$. 
SRE $\tilde{\beta}_1$ of $\beta_1$ has the form

$$\tilde{\beta}_1 = (X_1'M_{B_{mn}}X_1)^{-1}X_1'M_{B_{mn}}Y.$$
Inference about $\beta_1$ may benefit from moving the least squares estimate for the full model in the direction of the least squares estimate without the nuisance variables (Steinian shrinkage).

Or to drop the nuisance variables if there is evidence that they do not provide useful information (Bancroft: pretesting).
A semiparametric Stein-type estimator (SSTE) \( \hat{\beta}_1^S \) of \( \beta_1 \), can be defined as

\[
\hat{\beta}_1^S = \hat{\beta}_1 - (\hat{\beta}_1 - \tilde{\beta}_1)(p_2 - 2) T_n^{-1}, \quad p_2 \geq 3,
\]

where

\[
T_n = \frac{n}{\hat{\sigma}_n^2} \hat{\beta}_2' X'_2 B_{mn}' M_{B_{mn}} x_2 B_{mn} X'_2 \hat{\beta}_2,
\]

with

\[
\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - x'_i \hat{\beta} - \hat{\theta}' B d_{mn}(t_i))^2.
\]

Here, the test statistic \( T_n \sim \chi^2_{p_2, \alpha} \).
Positive-rule Stein-type semiparametric estimator (PSSTE) by using positive-part of the SSTE which will control the possible over shrinking in SSTE. A PSSTE has the form

\[
\hat{\beta}_1^{S+} = \hat{\beta}_1^S - [1 - (p_2 - 1)T_n^{-1}] \mathbb{I}(T_n \leq p_2 - 2)(\hat{\beta}_1 - \tilde{\beta}_1), \quad p_2 \geq 3.
\]
Bancroft (1944) introduced the idea of the preliminary test estimation as a basis for identifying model-estimator uncertainty.

The semiparametric preliminary test estimator (SPTE) $\hat{\beta}_1^{PT}$ of $\beta_1$ can be defined as

$$\hat{\beta}_1^{PT} = \hat{\beta}_1 I(T_n > \chi^2_{p_2, \alpha}) + \tilde{\beta}_1 I(T_n \leq \chi^2_{p_2, \alpha}), \quad p_2 \geq 1,$$

Thus, $\hat{\beta}_1^{PT}$ chooses $\tilde{\beta}_1$ when $H_0 : \beta_2 = 0$ is tenable, otherwise $\hat{\beta}_1$. 
The improved preliminary test semiparametric estimator (ISPTE) is defined as

\[ \hat{\beta}_1^{IPT} = \hat{\beta}_1^S I(T_n > \chi^2_{p_2,\alpha}) + \tilde{\beta}_1 I(T_n \leq \chi^2_{p_2,\alpha}), \quad p_2 \geq 3, \]
First Order Asymptotics

Assumption 1

There exist bounded functions $h_s(\cdot)$ over $[0, 1]$, $s = 1, \ldots, p$, such that

$$x_{is} = h_s(t_i) + u_{is}, \ i = 1, \ldots, n, s = 1, \ldots, p,$$

where $u_i = (u_{i1}, \ldots, u_{ip})'$ are real vectors satisfying

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} u_{ik} u_{ij}}{n} = b_{kj}, \ \text{for} \ k = 1, \ldots, p, \ j = 1, \ldots, p,$$

and the matrix $B = (b_{kj})$ is nonsingular. Moreover,

$$\max_{1 \leq k \leq p} \|Au_k^*\| = O\left(\left[\text{tr}(A'A)\right]^{\frac{1}{2}}\right), \ \text{for any matrix} \ A,$$

where $u_k^* = (u_{1k}, \ldots, u_{nk})'$ and $\| \cdot \|$ denotes the Euclidean norm.
Assumption 2

The functions $g(\cdot)$ and $h_j(\cdot)$ satisfy the Lipschitz condition of order $\nu$, i.e., there exists a constant $c$ such that

$$|f_j(s) - f_j(t)| \leq c|s - t|^{(\nu)}, \quad \text{for any } s, t \in [0, 1], \quad j = 0, 1, \ldots, p,$$

where $f_0(\cdot) = g(\cdot)$ and $f_j(\cdot) = h_j(\cdot)$, $j = 1, \ldots, p$. 
Lemma

If Assumptions 1 and 2 are satisfied, and \( \varepsilon_i \) are independent with mean zero and constant variance \( \sigma^2 \) and \( \mu_{3i} = E\varepsilon_i^3 \) being uniformly bounded, then

\[
\sqrt{n}(\hat{\beta} - \beta) \to_D N(0, \sigma^2 B^{-1}) \quad \text{and} \quad B'_m(t)\hat{\theta} - g(t) = O_p(n^{-\frac{\nu}{2\nu+1}}),
\]

where “\( \to_D \)” denotes convergence in distribution and \( B \) is defined in Assumption 1.
To study the asymptotic quadratic risks of $\hat{\beta}_1$, $\tilde{\beta}_1$, $\hat{\beta}_1^{PT}$, $\hat{\beta}_1^S$ and $\hat{\beta}_1^{S+}$, we consider a sequence of local alternatives.

Under fixed alternatives all the estimators are asymptotically equivalent to $\hat{\beta}_1$, while $\tilde{\beta}_1$ has unbounded risk.

A sequence $\{K_n\}$ of local alternatives defined by

$$K_n : \beta_{2(n)} = n^{-\frac{1}{2}} \omega, \; \omega \neq 0 \text{ fixed}$$

(8)
Asymptotic Properties

Asymptotic Distributional Bias (ADB)

The asymptotic distributional bias (ADB) of an estimator $\delta$ is defined as

$$\text{ADB}(\delta) = \lim_{n \to \infty} E \left\{ n^{\frac{1}{2}} (\delta - \beta_2) \right\}.$$
Suppose that Assumptions 1 and 2 hold. Under \( \{ K_n \} \), the ADB of the estimators are as follows:

- \( \text{ADB}(\hat{\beta}_1) = 0 \),
- \( \text{ADB}(\tilde{\beta}_1) = -B_{11}^{-1}B_{12}\omega \),
- \( \text{ADB}(\hat{\beta}_1^{PT}) = -B_{11}^{-1}B_{12}\omega H_{p_2+2}(\chi^2_{p_2,\alpha}; \Delta) \),
- \( \text{ADB}(\hat{\beta}_1^S) = -(p_2 - 2)B_{11}^{-1}B_{12}\omega E(\chi_{p_2,\alpha}^{-2}; \Delta) \),
Under Assumptions 1 and 2, and \( \{K_n\} \), the ADB of the estimators are as follows:

\[
ADB(\hat{\beta}_1^{S^+}) = -B_{11}^{-1}B_{12}\omega H_{p_2+2}(p_2 - 2; \Delta) - (p_2 - 2)B_{11}^{-1}B_{12}\omega \\
\left\{ E \left[ \chi_{p_2+2}^{-2}(\Delta) I(\chi_{p_2+2}^2(\Delta) > p_2 - 2) \right] \right\},
\]

\[
ADB(\hat{\beta}_1^{IPT}) = -B_{11}^{-1}B_{12}\omega H_{p_2+2}(p_2 - 2; \Delta) - (p_2 - 2)B_{11}^{-1}B_{12}\omega \\
\left\{ E \left[ \chi_{p_2+2}^{-2}(\Delta) I(\chi_{p_2+2}^2(\Delta) > \chi_{p_2,\alpha}^2) \right] \right\},
\]

where \( B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \) with \( B \) defined in Assumption 1, \( \Delta = (\omega' B_{22.1} \omega)\sigma^{-2} \), \( B_{22.1} = B_{22} - B_{21}B_{11}^{-1}B_{12} \), \( H_V(x; \Delta) \) denotes the noncentral chi-square distribution function with noncentrality parameter \( \Delta \) and \( v \) degrees of freedom.
Asymptotic Distributional Risk (ADR)

- Define a quadratic loss function using a positive definite matrix (p.d.m.) $\mathbf{M}$, namely,

$$
\mathcal{L}(\delta, \beta_1) = n(\delta - \beta_1)' \mathbf{M}(\delta - \beta_1),
$$

$\delta$ can be any one of estimator of $\hat{\beta}_1$,

- We assume that the asymptotic distribution function of $\delta$ under $\{K_n\}$ exists and is given by

$$
F(x) = \lim_{n \to \infty} P\{ \sqrt{n}(\delta - \beta_1) \leq x | K_n \}
$$

- Then, the AQDR of $\delta$ is defined as

$$
R(\delta, \mathbf{M}) = tr \left\{ \mathbf{M} \int_{\mathbb{R}^p} \int xx' dF(x) \right\} = tr(\mathbf{MV}),
$$

where $\mathbf{V}$ is the dispersion matrix for the asymptotic distribution $F(x)$. 

Asymptotic Quadratic Distributional Risk (AQDR)

Suppose that assumptions 1 and 2 hold, then under \( \{K_n\} \) the AQDR of the estimators are:

\[
R(\hat{\beta}_1; M) = \sigma^2 \text{tr}(MB_{11.2}^{-1}),
\]

\[
R(\tilde{\beta}_1; M) = \sigma^2 \text{tr}(MB_{11}^{-1}) + \omega' B_{21} B_{11}^{-1} MB_{11}^{-1} B_{12} \omega,
\]
Asymptotic Quadratic Distributional Risk (AQDR) ...

\[ R(\hat{\beta}_1^{PT}; \mathbf{M}) = \sigma^2 \left[ \text{tr}(\mathbf{M} \mathbf{B}_{11.2}^{-1}) \left\{ 1 - H_{p_2+2}(\chi_{p_2, \alpha}^2; \Delta) \right\} \right. \]
\[ + \left. \text{tr}(\mathbf{M} \mathbf{B}_{11}^{-1}) H_{p_2+2}(\chi_{p_2, \alpha}^2; \Delta) \right] \]
\[ + \mathbf{\omega}' \mathbf{B}_{21} \mathbf{B}_{11}^{-1} \mathbf{M} \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \]
\[ \left[ 2H_{p_2+2}(\chi_{p_2, \alpha}^2; \Delta) - H_{p_2+4}(\chi_{p_2, \alpha}^2; \Delta) \right], \]
Asymptotic Quadratic Distributional Risk (AQDR) ...

\[
R(\hat{\beta}_{1n}^S; \mathbf{M}) = \sigma^2 \left[ \text{tr}(\mathbf{M}\mathbf{B}_{11.2}^{-1}) - (p_2 - 2)\text{tr}(\mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{M}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22.1}^{-1}) \right]
\cdot \left\{ 2E(\chi_{p_2,\alpha}^2(\Delta)) - (p_2 - 2)E(\chi_{p_2+2}^{-4}(\Delta)) \right\}
+ (p_2^2 - 4)\omega'\mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{M}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\omega E(\chi_{p_2+4}^{-4}(\Delta)),
\]
Asymptotic Quadratic Distributional Risk (AQDR) ...

\[
R(\hat{\beta}^{S^+}; M) = R(\hat{\beta}^{S}; M) + (p_2 - 2) \text{tr}(B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}B_{22.1}^{-1}) \left\{ 2E \left[ \chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2) \right] \right.
\]
\[
- (p_2 - 2)E \left[ \chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2) \right] \right\} - \sigma^2 \text{tr}(B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}B_{22.1}^{-1})
\]
\[
\cdot H_{p_2+2}(p_2 - 2; \Delta) + \omega' B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}\omega \left[ 2H_{p_2+2}(p_2 - 2; \Delta) - H_{p_2+4}(p_2 - 2; \Delta) \right]
\]
\[
- (p_2 - 2)\omega B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}\omega' \left\{ 2E \left[ \chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2) \right] \right.
\]
\[
- 2E \left[ \chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2) \right] + (p_2 - 2)E \left[ \chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2) \right] \right\} ,
\]
Asymptotic Quadratic Distributional Risk (AQDR) ...

\[ R(\hat{\beta}^{\text{IPT}}; M) = R(\hat{\beta}^{S}; M) + (p_2 - 2)\text{tr}(B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}B_{22}^{-1}) \left\{ 2E \left[ \chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq \chi_{p_2,\alpha}^2) \right] \right. \\
- (p_2 - 2)E \left[ \chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq \chi_{p_2,\alpha}^2) \right] \right\} - \sigma^2\text{tr}(B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}B_{22}^{-1}) \\
\cdot H_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta) + B_{11}^{-1}B_{12}\omega'B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}\omega \left[ 2H_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta) - H_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta) \right] \\
- (p_2 - 2)\omega'B_{21}B_{11}^{-1}MB_{11}^{-1}B_{12}\omega \left\{ 2E \left[ \chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq \chi_{p_2,\alpha}^2) \right] \right. \\
- 2E \left[ \chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq \chi_{p_2,\alpha}^2) \right] + (p_2 - 2)E \left[ \chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq \chi_{p_2,\alpha}^2) \right] \} \right. \]
THEORETICAL SUMMARY

Dominance of Shrinkage Estimator

By comparing $R(\hat{\beta}_1^S)$ and $R(\hat{\beta}_1)$ following dominance condition holds. If $\mathbf{M} \in \mathbf{M}^D$, $\hat{\beta}_1^S$ dominates $\hat{\beta}_1$ for any $\omega$ in the sense of AQDR, where

$$\mathbf{M}^D = \left\{ \mathbf{M} : \frac{\text{tr}(\mathbf{B}_{21} \mathbf{B}_{11}^{-1} \mathbf{M} \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \mathbf{B}_{22.1}^{-1})}{\text{ch}_{\text{max}}(\mathbf{B}_{21} \mathbf{B}_{11}^{-1} \mathbf{M} \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \mathbf{B}_{22.1}^{-1})} \geq \frac{p_2 + 2}{2} \right\}.$$
THEORETICAL SUMMARY

Dominance of Shrinkage Estimator

- $\hat{\beta}^{S+}$ dominates $\hat{\beta}^S$ for all the values of $\omega$, with strict inequality holds for some $\omega$.
- Risk of $\hat{\beta}^{S+}$ is also smaller than the risk of $\hat{\beta}_1$ in the entire parameter space and the upper limit is attained when $\Delta$ approaches $\infty$.
- This implies that

$$R(\hat{\beta}_1^{S+}) \leq R(\hat{\beta}_1^S) \leq R(\hat{\beta}_1), \text{ for any } M \in M^D \text{ and } \omega,$$

with strict inequality holds for some $\omega$.

Thus, we conclude that $\hat{\beta}_1^S$ and $\hat{\beta}_1^{S+}$ perform better than $\hat{\beta}_1$ in the entire parameter space induced by $\Delta$. The gain in risk over $\hat{\beta}_1$ is substantial when $\Delta = 0$ or near.
Similar results hold for the PT and IPT estimators as well.

- When $\Delta = 0$,

$$R(\hat{\beta}_1^{IPT}) < R(\hat{\beta}_1^{PT}) < R(\hat{\beta}_1).$$

- Interestingly, the preliminary test estimators outperform Stein-type estimators when either $\Delta = 0$ or close to zero.

More importantly, Stein-type estimators dominate $\hat{\beta}_1$. 
MONTE CARLO STUDY

Setup

We simulate the response from the following model:

\[ y_i = x_{1i} \beta_1 + x_{2i} \beta_2 + \ldots + x_{pi} \beta_p + g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \]

where \( t_i = (i - 0.5)/n \), \( x_{1i} = (\zeta_{1i}^{(1)})^2 + \zeta_{1i}^{(1)} + \xi_{1i} \), 
\( x_{2i} = (\zeta_{2i}^{(1)})^2 + \zeta_{1i}^{(1)} + 2\xi_{2i} \), 
\( x_{si} = (\zeta_{si}^{(1)})^2 + \zeta_{si}^{(1)} \) with 
\( \zeta_{si}^{(1)} \) i.i.d. \( \sim N(0, 1) \), \( \zeta_{1i}^{(1)} \) i.i.d. \( \sim N(0, 1) \),
\( \xi_{1i} \sim \text{Bernoulli (0.45)} \) and \( \xi_{2i} \sim \text{Bernoulli (0.45)} \) for all 
\( s = 3, \ldots, p \) and \( i = 1, \ldots, n \).
Moreover, \( \varepsilon_i \) are i.i.d. \( N(0, 1) \), and \( g(t) = \sin(4\pi t) \).
Hypotheses that we are testing

- \( H_0 : \beta_j = 0, \) for \( j = p_1 + 1, p_1 + 2, \ldots, p_1 + p_2, \) with \( p = p_1 + p_2. \)
- We are interested in estimating \( \beta_1, \beta_2, \beta_3 \) and \( \beta_4 \) when the remaining regression parameters may not be useful.
- We partition the regression coefficients as \( \beta = (\beta_1, \beta_2) = (\beta_1, 0) \) with \( \beta_1 = (2, 1.5, 1, 0.6), \)
- We defined the parameter \( \Delta^* = ||\beta - \beta^{(0)}||, \) where \( \beta^{(0)} = (\beta_1, 0) \) and \( || \cdot || \) is the Euclidean norm.
- To determine the behavior of the estimators for \( \Delta^* > 0, \) further datasets were generated from those distributions under local alternative hypotheses. We considered \( \Delta^* = 0, .1, .2, .3, .4, .5, .8, 1, 2, \) and 4.
Numerically calculated the relative MSEs of the proposed estimators $\hat{\beta}_1, \hat{\beta}_1^S, \hat{\beta}_1^{S+}, \hat{\beta}_1^{PT}, \hat{\beta}_1^{IPT}$, relative to the unrestricted estimator $\hat{\beta}_1$, using:

$$RMSE(\hat{\beta}_1 : \hat{\beta}_1^{\diamond}) = \frac{MSE(\hat{\beta}_1)}{MSE(\hat{\beta}_1^{\diamond})}.$$ 

The amount by which an RMSE is larger than unity indicates the degree of superiority of the estimator $\hat{\beta}_1^{\diamond}$ over $\hat{\beta}_1$. 
Table 5: Simulated relative MSE with respect to $\hat{\beta}_1$ for $n = 40$, $p_2 = 5$.

<table>
<thead>
<tr>
<th>$\Delta^*$</th>
<th>$\tilde{\beta}_1$</th>
<th>$\hat{\beta}_1^S$</th>
<th>$\hat{\beta}_1^{S+}$</th>
<th>$\hat{\beta}_1^{PT}$</th>
<th>$\hat{\beta}_1^{IPT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.90</td>
<td>1.66</td>
<td>1.85</td>
<td>2.23</td>
<td>2.45</td>
</tr>
<tr>
<td>0.1</td>
<td>2.50</td>
<td>1.56</td>
<td>1.69</td>
<td>1.83</td>
<td>2.05</td>
</tr>
<tr>
<td>0.2</td>
<td>1.90</td>
<td>1.37</td>
<td>1.43</td>
<td>1.26</td>
<td>1.47</td>
</tr>
<tr>
<td>0.3</td>
<td>1.35</td>
<td>1.24</td>
<td>1.25</td>
<td>0.96</td>
<td>1.14</td>
</tr>
<tr>
<td>0.4</td>
<td>0.97</td>
<td>1.15</td>
<td>1.15</td>
<td>0.88</td>
<td>1.02</td>
</tr>
<tr>
<td>0.5</td>
<td>0.68</td>
<td>1.09</td>
<td>1.09</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>0.8</td>
<td>0.32</td>
<td>1.04</td>
<td>1.04</td>
<td>0.99</td>
<td>1.03</td>
</tr>
<tr>
<td>1.0</td>
<td>0.21</td>
<td>1.03</td>
<td>1.03</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>2.0</td>
<td>0.06</td>
<td>1.01</td>
<td>1.01</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>4.0</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
MONTE CARLO RESULTS

n=30, p1=4, p2=5

n=40, p1=4, p2=5

n=30, p1=4, p2=9

n=40, p1=4, p2=9
MONTE CARLO RESULTS

**n=50, p1=4, p2=5**

![Graph showing relative efficiency for n=50, p1=4, p2=5]

**n=80, p1=4, p2=5**

![Graph showing relative efficiency for n=80, p1=4, p2=5]

**n=50, p1=4, p2=9**

![Graph showing relative efficiency for n=50, p1=4, p2=9]

**n=80, p1=4, p2=9**

![Graph showing relative efficiency for n=80, p1=4, p2=9]
From Figures we see that the restricted estimator outperforms all other estimators for all the cases considered when the restriction is at or near $\Delta^* = 0$.

As the restriction moves away from $\Delta^* = 0$, the restricted estimator becomes unbounded.

At or near the restriction, $\Delta^* = 0$, preliminary test (PT) estimators outperform Stein-type estimators.

The proposed estimators are bounded.

In summary, simulation results are in agreement with our asymptotic results and the general theory of these estimators available in literature.
[Mroz (Econometrica, 1987)] used a simple model to explain woman’s labour supply in 1975.

The response variable (hours) was modelled with age of woman (age), years of schooling (educ), other family income (nwifeinc), number of children under 6 years (k5), number of children between ages 6 and 18 (k618) and the marginal tax rate (mtr).

Modification to Mroz’s model

Although the linearity in the parameters allows for relatively simple estimation schemes, we modified the model by incorporating nwifeinc and mtr non-linearly. Moreover, in order to utilize the entire data set, we include inlf in Mroz’s model as a dummy variable.
Thus our labour supply functions become

**Mroz Model 1:**

\[
\text{hours} = \beta_0 + \beta_1 \text{inlf} + \beta_2 \ln(\text{wage}) + \beta_3 \text{age} + \beta_4 \text{educ} \\
+ \beta_5 \text{k5} + \beta_6 \text{k618} + \beta_7 \text{mtr} + g(\text{nwifeinc})
\]  \hspace{1cm} (9)

**Mroz Model 2:**

\[
\text{hours} = \beta_0 + \beta_1 \text{inlf} + \beta_2 \ln(\text{wage}) + \beta_3 \text{age} + \beta_4 \text{educ} \\
+ \beta_5 \text{k5} + \beta_6 \text{k618} + \beta_7 \text{nwifeinc} + g(\text{mtr})
\]  \hspace{1cm} (10)
Uncertain Prior Information

Since the variables that are not in the models are conjectured to be insignificant, we take it as our UPI. In our case, $p_1 = 7$ and $p_2 = 8$.

The results of the shrinkage estimates along with bootstrapped biases and standard errors and relative risks (MSEs) are presented in the following slides.
Relative Risks when nwifeinc entered nonparametrically

Table 6: Mroz Model with s(nwifeinc): Estimates (row 1), biases (row 2) and standard errors (row 3) of intercept ($\beta_0$), inlf ($\beta_1$), lwage ($\beta_2$), age ($\beta_3$), educ ($\beta_4$), k5 ($\beta_5$), k618 ($\beta_6$) and mtr ($\beta_7$). RR is the relative risk compared to the unrestricted estimator. Results are based on 9,999 bootstrap samples.

<table>
<thead>
<tr>
<th>Est</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>4741.33</td>
<td>1085.60</td>
<td>-0.04</td>
<td>-1.00</td>
<td>-29.15</td>
<td>-122.09</td>
<td>25.71</td>
<td>-5348.04</td>
<td>1.186</td>
</tr>
<tr>
<td></td>
<td>9.09e-13</td>
<td>0.00e+00</td>
<td>2.01e-16</td>
<td>2.22e-16</td>
<td>-7.10e-15</td>
<td>-1.84e-13</td>
<td>1.17e-13</td>
<td>-9.09e-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.32e-12</td>
<td>2.09e-13</td>
<td>5.15e-16</td>
<td>5.58e-15</td>
<td>2.50e-14</td>
<td>8.45e-14</td>
<td>6.76e-14</td>
<td>2.41e-12</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>4988.89</td>
<td>1084.06</td>
<td>-0.083</td>
<td>0.88</td>
<td>-18.87</td>
<td>-121.43</td>
<td>20.61</td>
<td>-5430.78</td>
<td>1.045</td>
</tr>
<tr>
<td></td>
<td>-9.09e-13</td>
<td>0.00e+00</td>
<td>8.32e-17</td>
<td>-4.66e-15</td>
<td>1.77e-14</td>
<td>-1.98e-13</td>
<td>1.21e-13</td>
<td>0.00e+00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.66e-12</td>
<td>1.83e-13</td>
<td>5.61e-16</td>
<td>1.01e-14</td>
<td>2.51e-14</td>
<td>9.23e-14</td>
<td>7.34e-14</td>
<td>2.46e-12</td>
<td></td>
</tr>
<tr>
<td>S+</td>
<td>4988.89</td>
<td>1084.06</td>
<td>-0.083</td>
<td>0.88</td>
<td>-18.87</td>
<td>-121.43</td>
<td>20.61</td>
<td>-5430.78</td>
<td>1.045</td>
</tr>
<tr>
<td></td>
<td>-9.09e-13</td>
<td>0.00e+00</td>
<td>8.33e-17</td>
<td>-4.55e-15</td>
<td>1.77e-14</td>
<td>-1.98e-13</td>
<td>1.21e-13</td>
<td>0.00e+00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.67e-12</td>
<td>1.82e-13</td>
<td>5.60e-16</td>
<td>9.94e-15</td>
<td>2.47e-14</td>
<td>9.15e-14</td>
<td>7.25e-14</td>
<td>2.47e-12</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>4988.89</td>
<td>1084.07</td>
<td>-0.083</td>
<td>0.88</td>
<td>-18.87</td>
<td>-121.43</td>
<td>20.61</td>
<td>-5430.78</td>
<td>1.044</td>
</tr>
<tr>
<td></td>
<td>-9.09e-13</td>
<td>0.00e+00</td>
<td>8.32e-17</td>
<td>-4.44e-15</td>
<td>1.77e-14</td>
<td>-1.98e-13</td>
<td>1.24e-13</td>
<td>0.00e+00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.66e-12</td>
<td>1.82e-13</td>
<td>5.58e-16</td>
<td>9.90e-15</td>
<td>2.49e-14</td>
<td>9.12e-14</td>
<td>7.22e-14</td>
<td>2.47e-12</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>4988.89</td>
<td>1084.07</td>
<td>-0.083</td>
<td>0.88</td>
<td>-18.88</td>
<td>-121.43</td>
<td>20.61</td>
<td>-5430.78</td>
<td>1.044</td>
</tr>
<tr>
<td></td>
<td>-9.09e-13</td>
<td>0.00e+00</td>
<td>8.33e-17</td>
<td>-4.66e-15</td>
<td>1.77e-14</td>
<td>-1.98e-13</td>
<td>1.21e-13</td>
<td>0.00e+00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.63e-12</td>
<td>1.82e-13</td>
<td>5.55e-16</td>
<td>9.85e-15</td>
<td>2.45e-14</td>
<td>9.06e-14</td>
<td>7.33e-14</td>
<td>2.43e-12</td>
<td></td>
</tr>
</tbody>
</table>
Table 7: Mroz Model with \( s(mtr) \): Estimates (row 1), biases (row 2) and standard errors (row 3) of intercept (\( \beta_0 \)), inlf (\( \beta_1 \)), lwage (\( \beta_2 \)), age (\( \beta_3 \)), educ (\( \beta_4 \)), k5 (\( \beta_5 \)), k618 (\( \beta_6 \)) and nwifeinc (\( \beta_7 \)). RR is the relative risk compared to the unrestricted estimator. Results are based on 9,999 bootstrap samples.

<table>
<thead>
<tr>
<th>Est</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_6 )</th>
<th>( \beta_7 )</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2944.74</td>
<td>1116.842</td>
<td>-0.109</td>
<td>-1.644</td>
<td>-36.919</td>
<td>-101.413</td>
<td>15.867</td>
<td>-35.620</td>
<td>1.181</td>
</tr>
<tr>
<td></td>
<td>9.09e-13</td>
<td>0.00e+00</td>
<td>-8.32e-17</td>
<td>-1.33e-15</td>
<td>-1.42e-14</td>
<td>-1.56e-13</td>
<td>7.63e-14</td>
<td>-7.10e-15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.19e-12</td>
<td>2.07e-13</td>
<td>3.94e-16</td>
<td>4.99e-15</td>
<td>2.145e-14</td>
<td>6.79e-14</td>
<td>5.68e-14</td>
<td>1.02e-14</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>3256.79</td>
<td>1111.35</td>
<td>-0.20</td>
<td>0.33</td>
<td>-19.38</td>
<td>-100.77</td>
<td>13.49</td>
<td>-34.53</td>
<td>1.059</td>
</tr>
<tr>
<td></td>
<td>0.00e+00</td>
<td>2.27e-13</td>
<td>-2.77e-17</td>
<td>-4.49e-15</td>
<td>-1.42e-14</td>
<td>-1.42e-13</td>
<td>1.15e-13</td>
<td>-1.42e-14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.18e-12</td>
<td>1.85e-13</td>
<td>4.72e-16</td>
<td>8.34e-15</td>
<td>2.48e-14</td>
<td>7.46e-14</td>
<td>6.46e-14</td>
<td>1.09e-14</td>
<td></td>
</tr>
<tr>
<td>S+</td>
<td>3256.79</td>
<td>1111.35</td>
<td>-0.20</td>
<td>0.33</td>
<td>-19.38</td>
<td>-100.77</td>
<td>13.49</td>
<td>-34.53</td>
<td>1.059</td>
</tr>
<tr>
<td></td>
<td>0.00e+00</td>
<td>2.27e-13</td>
<td>-2.77e-17</td>
<td>-4.61e-15</td>
<td>-1.42e-14</td>
<td>-1.42e-13</td>
<td>1.15e-13</td>
<td>-1.42e-14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.18e-12</td>
<td>1.83e-13</td>
<td>4.66e-16</td>
<td>8.18e-15</td>
<td>2.51e-14</td>
<td>7.48e-14</td>
<td>6.37e-14</td>
<td>1.09e-14</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>3256.79</td>
<td>1111.35</td>
<td>-0.20</td>
<td>0.33</td>
<td>-19.38</td>
<td>-100.77</td>
<td>13.49</td>
<td>-34.52</td>
<td>1.057</td>
</tr>
<tr>
<td></td>
<td>0.00e+00</td>
<td>2.27e-13</td>
<td>-2.77e-17</td>
<td>-4.49e-15</td>
<td>-1.42e-14</td>
<td>-1.42e-13</td>
<td>1.15e-13</td>
<td>-1.42e-14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.20e-12</td>
<td>1.84e-13</td>
<td>4.63e-16</td>
<td>8.29e-15</td>
<td>2.53e-14</td>
<td>7.53e-14</td>
<td>6.46e-14</td>
<td>1.10e-14</td>
<td></td>
</tr>
<tr>
<td>IPT</td>
<td>3256.78</td>
<td>1111.35</td>
<td>-0.20</td>
<td>0.33</td>
<td>-19.38</td>
<td>-100.77</td>
<td>13.49</td>
<td>-34.53</td>
<td>1.058</td>
</tr>
<tr>
<td></td>
<td>0.00e+00</td>
<td>2.27e-13</td>
<td>0.00e-17</td>
<td>-4.55e-15</td>
<td>-1.42e-14</td>
<td>-1.42e-13</td>
<td>1.15e-13</td>
<td>-1.42e-14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.19e-12</td>
<td>1.86e-13</td>
<td>4.66e-16</td>
<td>8.29e-15</td>
<td>2.53e-14</td>
<td>7.56e-14</td>
<td>6.42e-14</td>
<td>1.10e-14</td>
<td></td>
</tr>
</tbody>
</table>
In general, RR is higher for the model where mtr was entered nonlinearly compared to the one where nwifeinc entered nonlinearly.

We also bootstrapped the RRs based on the best-fitt models (3), (4) [results not shown here]

However, RRs are lower in Mroz’s models compared to those of best-fit models (3), (4).

This confirms that the UPI is not very reliable in Mroz’s models. In particular, shrinkage and positive shrinkage estimates are better options for estimating $\beta_1$ in such cases.
Figure 4: Cowboy Rope Lasso
They’ve found APE to be better when $p_2 = 3$ (or small).

Shrinkage was better for $p_2 = 11$.

Their results were based on $p_1 = 3$. 
### Table 8: Simulated RMSE with respect to $\hat{\beta}_1^{UR}$ for $p_1 = 3$, $\Delta^* = 0$

<table>
<thead>
<tr>
<th>n</th>
<th>$p_2$</th>
<th>$\beta_1^R$</th>
<th>$\hat{\beta}_1^S$</th>
<th>$\hat{\beta}_1^{S+}$</th>
<th>$\hat{\beta}_1^{PT}$</th>
<th>$\hat{\beta}_1^{IPT}$</th>
<th>L(CV)</th>
<th>L(GCV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>3</td>
<td>2.49</td>
<td>1.26</td>
<td>1.37</td>
<td>1.95</td>
<td>2.03</td>
<td>1.46</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4.84</td>
<td>2.15</td>
<td>2.49</td>
<td>3.04</td>
<td>3.66</td>
<td>2.12</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>11.89</td>
<td>3.98</td>
<td>4.87</td>
<td>4.76</td>
<td>7.61</td>
<td>3.13</td>
<td>4.39</td>
</tr>
<tr>
<td>40</td>
<td>3</td>
<td>2.40</td>
<td>1.25</td>
<td>1.35</td>
<td>1.89</td>
<td>1.97</td>
<td>1.62</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4.10</td>
<td>1.99</td>
<td>2.32</td>
<td>2.76</td>
<td>3.24</td>
<td>2.37</td>
<td>2.55</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>8.40</td>
<td>3.55</td>
<td>4.24</td>
<td>4.23</td>
<td>6.11</td>
<td>3.49</td>
<td>4.05</td>
</tr>
<tr>
<td>80</td>
<td>3</td>
<td>2.23</td>
<td>1.26</td>
<td>1.35</td>
<td>1.85</td>
<td>1.91</td>
<td>1.82</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.63</td>
<td>1.91</td>
<td>2.26</td>
<td>2.67</td>
<td>3.04</td>
<td>2.65</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>6.19</td>
<td>3.17</td>
<td>3.88</td>
<td>3.96</td>
<td>5.13</td>
<td>4.00</td>
<td>4.17</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>2.24</td>
<td>1.23</td>
<td>1.34</td>
<td>1.83</td>
<td>1.89</td>
<td>1.63</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.51</td>
<td>1.92</td>
<td>2.24</td>
<td>2.65</td>
<td>2.99</td>
<td>2.27</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>5.90</td>
<td>3.13</td>
<td>3.81</td>
<td>3.99</td>
<td>5.01</td>
<td>3.31</td>
<td>3.40</td>
</tr>
</tbody>
</table>
Graphical Comparison of RR when $p_1 = 3$ and $p_2$ varies

- $p_1 = 3, n = 30$
- $p_1 = 3, n = 40$
- $p_1 = 3, n = 80$
- $p_1 = 3, n = 100$
Graphical Comparison of RRs when $p_1 = 4$ and $p_2$ varies

$p_1 = 4, n = 30$

$p_1 = 4, n = 40$

$p_1 = 4, n = 80$

$p_1 = 4, n = 100$
Shrinkage method is good for moderate sample and large $p_2$ sizes.

This result is in agreement with [Ahmed et al (ANZ J Stat, 2007)]

However, as $n$ gets larger (> 80, and perhaps smaller), LASSO performs better.
Thank you!

Engle, R.F. and W.J. Granger and J. Rice and A. Weiss (1986) Semiparametric estimates of the relation between weather and electricity sales  
Journal of the American Statistical Association, 80, 310 – 319

Sage Publications, Thousand Oaks

Fox, John (2005) Introduction to Nonparametric Regression  
Regression models for categorical and limited dependent variables,
_SAGE Publications, Thousand Oaks._

The sensitivity of an empirical model of married womens hours of work to
economic and statistical assumptions,
_Econometrica 55_(4), 765799.

Introductory econometrics: a modern approach
_South Western College Publishing._

Regression shrinkage and selection via the lasso.


References

High order approximation for the coverage probability by confident set centered at the positive-part James-Stein estimator.

Positive Shrinkage, Improved Pretest and Absolute Penalty Estimators in Partially Linear Models.
*Linear Algebra and its Applications*, 430, 2749-2761.

Asymptotic simultaneous estimation of Poisson means.
*Linear Algebra and its Applications*, 430, 2734-2748.

Asymptotic expansion of the coverage probability of James-Stein Estimators.
Theore of Preliminary Test and Stein-Type Estimation with Applications Wiley.

Nonparametric Inference Strategies for the Quantile Functions Under Left Truncation and Right Censoring.
Journal of Nonparametric Statistics, 9, 189 - 198.

Positive-part Shrinkage M-estimation in Linear Models.

Asymptotic expansion of the coverage probability of James-Stein Estimators.