Probabilistic Index Models

Jan De Neve
Department of Data Analysis
Ghent University

M3 Storrs, Connecticut, USA
May 23, 2017

Jan.DeNeve@UGent.be
Introduction
Introduction to Probabilistic Index Models (PIMs)

- Class of (semiparametric) regression models.
- Different than Generalized Linear Models (GLMs).
- Connection with rank-tests.
- Connection with Cox Proportional Hazards models.

Content largely based on 2 publications

PIMs can be used for a variety of applications.

**Current status:** focus mainly on applications in biostatistics.

- Time-to-event data (survival analysis)
- Gene expression studies.

See e.g.


**Goal of this talk:**

Illustrate that PIMs might be useful for analyzing behavioral data
Question: What is a Probabilistic Index Model?
Question: What is a Probabilistic Index Model?

Answer: A regression model for the Probabilistic Index (PI).
Question: What is a Probabilistic Index Model?

Answer: A regression model for the Probabilistic Index (PI).

Question: What is the Probabilistic Index?
Question: What is a Probabilistic Index Model?

Answer: A regression model for the Probabilistic Index (PI).

Question: What is the Probabilistic Index?

Answer: The probability $P(Y_i < Y_j \mid \mathbf{X}_i, \mathbf{X}_j)$ with $(Y_i, \mathbf{X}_i^T)$ and $(Y_j, \mathbf{X}_j)$ i.i.d.
The Probabilistic Index
We will use the BtheB-study (R package HSAUR) for motivation and illustration.

**Beat the Blues Study (BtheB):**

- Clinical trial of an interactive multimedia program called Beat the Blues.
- **BtheB**: designed to deliver cognitive behavioural therapy to depressed patients via a computer terminal.
- Patients with depression recruited in primary care.
We will use the **BtheB**-study (R package HSAUR) for motivation and illustration.

**Beat the Blues Study** (BtheB):

- Clinical trial of an interactive multimedia program called Beat the Blues.
- **BtheB**: designed to deliver cognitive behavioural therapy to depressed patients via a **computer** terminal.
- Patients with **depression** recruited in primary care.
- Randomised to **BtheB program** or to ‘treatment as usual’ (TAU), i.e. **face-to-face counselling**.
- Depression is quantified via **Beck Depression Inventory II** (21 questions, range 0-63)
- 100 subjects in dataset (original study: 167 subjects)
- Longitudinal study, but we only consider a cross-sectional part.

Beat the Blues Study

- Beck Depression Inventory II after 3 months (higher score = more depressed).
- Beck Depression II is also measured at baseline.
- Treatment: BtheB versus TAU, randomized.
- Drugs: did the patient take anti-depressant drugs? - not randomized.
- Complete case analysis: 37 (BtheB) and 36 (TAU).
Beat the Blues Study

- Beck Depression Inventory II after 3 months (higher score = more depressed).
- Beck Depression II is also measured at baseline.
- Treatment: BtheB versus TAU, randomized.
- Drugs: did the patient take anti-depressant drugs? - not randomized.
- Complete case analysis: 37 (BtheB) and 36 (TAU).

Question 1

Is there a difference between the treatments in terms of depression?
Beat the Blues Study

- Beck Depression Inventory II after 3 months (higher score = more depressed).
- Beck Depression II is also measured at baseline.
- Treatment: BtheB versus TAU, randomized.
- Drugs: did the patient take anti-depressant drugs? - not randomized.
- Complete case analysis: 37 (BtheB) and 36 (TAU).

Question 1

Is there a difference between the treatments in terms of depression?

Question 2

Does anti-depressant drug have an effect on depression?
Both treatments seem to have a positive effect.
BtheB does a slightly better job
Modest deviation from normality

Theoretical Quantiles
Sample Quantiles

Normal QQ TAU

Normal QQ BtheB

Theoretical Quantiles
Sample Quantiles

-2 -1 0 1 2
0 10 20 30 40 50

-2 -1 0 1 2
0 10 20 30 40 50
Is there a difference between the treatments (X) in terms of depression (Y)?

\[ H_0 : \mathbb{E}(Y \mid X = TAU) = \mathbb{E}(Y \mid X = BtheB), \quad H_A : \text{not } H_0. \]

Two-sample t-test p-value:

- Welch: 0.041
- Permutation: 0.042.

95% CI for \( \mathbb{E}(Y \mid X = TAU) - \mathbb{E}(Y \mid X = BtheB) \):

\[ [0.23, 11.05] \]
The effect of the outlier
Results when the outlier is removed

Two-sample t-test p-value:

- Welch: 0.0083 (with outlier: 0.041)
- Permutation: 0.0087 (with outlier: 0.042).

95% CI for $\mathbb{E}(Y | X = TAU) - \mathbb{E}(Y | X = BtheB)$:

$[1.8, 11.7]$ (with outlier: $[0.23, 11.05]$)
Since the outlier has some effect, we might want to consider a more robust test.

We choose the Wilcoxon–Mann–Whitney (WMW) Rank test

- p-value with outlier: 0.041
- p-value without outlier: 0.022

What is the effect measure associated the WMW test?
Test statistic associated with the WMW test:

\[ T = \frac{U - 0.5}{SE_0(U)} \]

\[ U = \frac{1}{n_B n_T} \sum_i \sum_j I(Y_i^{BtheB} < Y_j^{TAU}) \]

\[ I(TRUE) = 1, \ I(FALSE) = 0, \]

and \( SE_0(U) \) the standard error of \( U \) under \( H_0 : F_{BtheB} = F_{TAU} \).
Test statistic associated with the WMW test:

\[ T = \frac{U - 0.5}{SE_0(U)} \]

\[ U = \frac{1}{n_B n_T} \sum_i \sum_j I(Y_i^{BtheB} < Y_j^{TAU}) \]

and \( SE_0(U) \) the standard error of \( U \) under \( H_0 : F_{BtheB} = F_{TAU} \).

Since

\[ E(U) = P(Y^{BtheB} < Y^{TAU}) \]

it follows that the WMW-test is associated with

\[ H_0 : F_{BtheB} = F_{TAU} \quad H_A : P(Y^{BtheB} < Y^{TAU}) \neq 0.5. \]

Note: under location-shift it also tests for \( H_A : \Delta \neq 0 \) with \( \Delta \) a location parameter (e.g. difference in means or medians).
The effect measure

\[ P\left( Y^{BtheB} < Y^{TAU} \right) \]

has many names:

- Mann–Whitney functional
- The nonparametric treatment effect
- The probability of superiority.
- ...
- The probabilistic index.

*It is the probability that a randomly selected patient receiving BtheB will have a better (here lower) depression score than a randomly selected patient receiving TAU.*

Example: \( \hat{P}(Y^{BtheB} < Y^{TAU}) = 64\% \) (95%CI : [51%, 75%])
The WMW test and the PI have some attractive properties

\[ T = \frac{U - 0.5}{SE_0(U)} \quad \text{and} \quad P\left( Y^{BtheB} < Y^{TAU} \right) \]

The PI

- Applies to ordinal outcomes (discrete or continuous).
- Scale-free.
- Invariant under monotone transformations of the outcome.
- ‘Easy’ to understand.

The WMW test

- Robust to outliers.
- Applies to ordinal outcomes (discrete or continuous).
- Good power properties: ARE.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Uniform</th>
<th>Logistic</th>
<th>( t_3 )</th>
<th>Laplace</th>
<th>( t_5 )</th>
<th>Exp</th>
<th>Cauchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \max(1 - x^2, 0) )</td>
<td>0.86</td>
<td>0.95</td>
<td>1</td>
<td>1.1</td>
<td>1.24</td>
<td>1.5</td>
<td>1.9</td>
<td>( \infty )</td>
</tr>
</tbody>
</table>
Return to the Beat the Blues Study

- Beck Depression Inventory II after 3 months (higher score = more depressed).
- Beck Depression II is also measured at baseline.
- Treatment (BtheB versus TAU, randomized).
- Drugs: did the patient take anti-depressant drugs? - not randomized.
- Complete case analysis: 37 (BtheB) and 36 (TAU).

Question 1

Is there a difference between the treatments in terms of depression?

Question 2

Does anti-depressant drug have an effect on depression?
Does anti-depressant drug have an effect on depression?

The drugs were not randomized.
Assessing the effect of drugs on depression

Ordinary t-test: p-value 0.51, 95% CI: [-3.9, 7.6]

→ ignores the baseline score (confounder)
Assessing the effect of drugs on depression

Ordinary t-test: p-value 0.51, 95% CI: [-3.9, 7.6]

→ ignores the baseline score (confounder)

Solution: write t-test as a regression model and included baseline score as a predictor

\[ \text{lm}(\text{score.3M} \sim \text{drugs} + \text{score.0M}) \]

→ p-value 0.009, 95% CI: [-7.6, -1.1]
(better (lower) score for those receiving drugs)
What if we are interested in the PI:

\[ P(Y^{Drugs} < Y^{No\ Drugs})? \]

**Problem:** Due to the confounder, we cannot trust the WMW test.

**Question:** Can we embed the WMW test in a regression context?
What if we are interested in the PI:

$$P \left( Y_{\text{Drugs}} < Y_{\text{No Drugs}} \right)$$?

**Problem:** Due to the confounder, we cannot trust the WMW test.

**Question:** Can we embed the WMW test in a regression context?

**Answers:** Yes, via a Probabilistic Index Model:

$$P \left( Y_i < Y_j \mid X_i, X_j \right) = m(X_i, X_j; \beta), \quad (Y_i, X_i^T) \text{ i.i.d.}$$

- $$(Y_i, X_i^T) \ i = 1, \ldots, n \text{ i.i.d. sample}$$
- $$X_i \text{ covariate, } p\text{-dimensional, e.g. } X_i^T = (\text{drugs, score.0M})$$
- $$m(\cdot) \text{ a known function}$$
- $$\beta \text{ the regression coefficient.}$$
Probabilistic Index Models
\[ P(Y_i < Y_j | \mathbf{X}_i, \mathbf{X}_j) = m(\mathbf{X}_i, \mathbf{X}_j; \beta), \]

**Question:** how should \( m(\mathbf{X}_i, \mathbf{X}_j; \beta) \) look like?
$$P(Y_i < Y_j \mid X_i, X_j) = m(X_i, X_j; \beta),$$

**Question:** how should $m(X_i, X_j; \beta)$ look like?

Let's have a look at the linear regression model for inspiration

$$E(Y_i \mid X_i) = X_i^T \beta,$$

which implies, exploiting $E(Y_i) - E(Y_j) = E(Y_i - Y_j)$,

$$E(Y_i - Y_j \mid X_i, X_j) = (X_i - X_j)^T \beta.$$
\[ P(Y_i < Y_j \mid X_i, X_j) = m(X_i, X_j; \beta), \]

**Question**: how should \( m(X_i, X_j; \beta) \) look like?

Let's have a look at the linear regression model for inspiration

\[ E(Y_i \mid X_i) = X_i^T \beta, \]

which implies, exploiting \( E(Y_i) - E(Y_j) = E(Y_i - Y_j) \),

\[ E(Y_i - Y_j \mid X_i, X_j) = (X_i - X_j)^T \beta. \]

So maybe the following makes sense

\[ P(Y_i < Y_j \mid X_i, X_j) = g^{-1}[(X_i - X_j)^T \beta], \]

with \( g(\cdot) \) a link-function (e.g. probit or logit) to ensure \( \text{PI} \in [0, 1] \).
PIMs: connection with other models
Connection with other models.

Model 1: the parametric normal linear model:

\[ Y_i = X_i^T \alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \]
Connection with other models.

Model 1: the parametric normal linear model:

\[ Y_i = X_i^T \alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \]

implies

\[
P (Y_i < Y_j \mid X_i, X_j)
= P \left( X_i^T \alpha + \varepsilon_i < X_j^T \alpha + \varepsilon_j \mid X_i, X_j \right)
= P \left( \varepsilon_i - \varepsilon_j < (X_j - X_i)^T \alpha \mid X_i, X_j \right)
= P \left( Z < (X_j - X_i)^T \frac{\alpha}{\sqrt{2\sigma^2}} \right) \quad Z \sim N(0, 1)
= g^{-1}[(X_j - X_i)^T \beta] \quad \text{with} \quad \beta = \frac{\alpha}{\sqrt{2\sigma^2}}, \quad g(\cdot) = \text{probit}(\cdot).\]
Connection with other models.

Model 2: semiparametric linear transformation model (part 1)

\[
h(Y_i) = X_i^T \alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2),
\]

with \( h(\cdot) \) strict monotone and unknown function.
Connection with other models.

Model 2: semiparametric linear transformation model (part 1)

\[ h(Y_i) = X_i^T \alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \]

with \( h(\cdot) \) strict monotone and unknown function. Since

\[ P(Y_i < Y_j \mid X_i, X_j) = P(h(Y_i) < h(Y_j) \mid X_i, X_j), \]

if follows that

\[ P(Y_i < Y_j \mid X_i, X_j) = g^{-1}[(X_j - X_i)^T \beta], \]

with \( \beta = \frac{\alpha}{\sqrt{2\sigma^2}} \) and \( g(\cdot) = \text{probit}(\cdot). \)
Connection with other models.

Model 2: semiparametric linear transformation model (part 2)

Since the difference between two extreme value variables follows a logistic distribution, one can show that

\[
h(Y_i) = X_i^T \alpha + \varepsilon_i, \quad \varepsilon_i \sim F(e) = 1 - \exp[-\exp(e)],
\]

implies the PIM

\[
P(Y_i < Y_j | X_i, X_j) = g^{-1}[(X_j - X_i)^T \beta],
\]

with \( \beta = \alpha \) and \( g(\cdot) = \text{logit}(\cdot) \).

Note: this is related to the Cox proportional hazards model.
PIMs: estimation theory
\[
P (Y_i < Y_j \mid X_i, X_j) = g^{-1}[(X_j - X_i)^T \beta],
\]

How can we semiparametrically estimate \( \beta \) only assuming the PIM (no further distributional assumptions)?
\[
P(Y_i < Y_j \mid X_i, X_j) = g^{-1}[(X_j - X_i)^T \beta],
\]

How can we semiparametrically estimate \( \beta \) only assuming the PIM (no further distributional assumptions)?

**Trick:**

\[
P(Y_i < Y_j \mid X_i, X_j) = \mathbb{E}(l_{ij} \mid X_i, X_j), \quad l_{ij} = I(Y_i < Y_j)
\]

\[
\Rightarrow \mathbb{E}(l_{ij} \mid X_i, X_j) = g^{-1}(X_{ij}^T \beta), \quad X_{ij} = X_j - X_i.
\]

Use `glm()` on transformed outcomes \( l_{ij} \) and predictors \( X_{ij} \) to estimate \( \beta \)!
Challenges in the estimation

cross-correlation:

\[ l_{ij} = I(Y_i < Y_j) \rightarrow I(Y_i < Y_l) \]
\[ \rightarrow I(Y_j < Y_l) \]
\[ \rightarrow I(Y_k < Y_i) \]
\[ \rightarrow I(Y_k < Y_j) \]

Consequences:

- you have to prove that `glm()` gives consistent estimators.
- provide consistent sandwich estimator for \( \text{Var} (\hat{\beta}) \) that takes the cross-correlation into account.
- Both are solved by writing out the influence function upon using Hajek-projections.
- Nice side result: `glm()` does not give the efficient estimator in theory, but in practice it is very close.
PIMs: connection with rank tests
Two-sample design

- $Y_i$: depression score at 3 months.
- $X_i$: anti-depressant drugs (no = 0, yes = 1).

Consider the PIM

$$P(Y_i < Y_j \mid X_i, X_j) = \text{expit}[(X_j - X_i)\beta].$$
Two-sample design

- $Y_i$: depression score at 3 months.
- $X_i$: anti-depressant drugs (no = 0, yes = 1).

Consider the PIM

$$P (Y_i < Y_j \mid X_i, X_j) = \expit[(X_j - X_i)\beta].$$

$$\rightarrow \expit(\beta) = P (Y_i < Y_j \mid X_i = 0, X_j = 1) = P (Y^{no} < Y^{yes})$$

$$\rightarrow \expit(\hat{\beta}) = \frac{1}{n_{no}n_{yes}} \sum_i \sum_j I (Y_i^{no} < Y_j^{yes}) = U.$$ 

- Wilcoxon–Mann–Whitney test is a special case of a PIM.
- PIM sandwich estimator for $\text{Var}(\hat{\beta})$ allows for Wald-type tests and the construction of confidence intervals.
- Similar results hold for the Kruskal-Wallis, Friedman, Jonckheere-Terpstra, ... rank tests.
Return to the BtheB study
- $Y_i$: depression score at 3 months.
- $X_i$: anti-depressant drugs (no = 0, yes = 1).
- $Z_i$: depression score at baseline.

Consider the PIM

$$P (Y_i < Y_j \mid X_i, X_j) = \text{expit}[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad X^T = (X, Z).$$
- $Y_i$: depression score at 3 months.
- $X_i$: anti-depressant drugs (no = 0, yes = 1).
- $Z_i$: depression score at baseline.

Consider the PIM

$$P(Y_i < Y_j | X_i, X_j) = \expit[(X_j-X_i)\beta_X + (Z_j-Z_i)\beta_Z], \quad X^T = (X, Z).$$

In R via library('pim')

```r
> m <- pim(bdi.3m ~ drug + bdi.pre, data = Data)
> summary(m)
pim.summary of following model:
bdi.3m ~ drug + bdi.pre
Type:   difference
Link:   logit

    Estimate Std. Error   z value  Pr(>|z|)
drugYes -0.87679   0.31925  -2.746   0.00602 **
bdi.pre  0.08240   0.01775   4.641 3.47e-06 ***
---
```

[33/37]
\[ P(Y_i < Y_j \mid X_i, X_j) = \text{expit}[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad X^T = (X, Z). \]

From `pim()`: \( \hat{\beta}_X = -0.88 \) and \( \hat{\beta}_Z = 0.082 \)

\[ \hat{P}(Y_i < Y_j \mid X_i = 0, X_j = 1, Z_i = Z_j) = \text{expit}(-0.88) = 0.29. \]

The estimated probability that a patient receiving anti-depressant drugs will have a worse score (i.e. higher) as compared to a patient not receiving anti-depressant drugs is 29% (95% CI: [0.18, 0.44]).
$$P(Y_i < Y_j \mid X_i, X_j) = \expit[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad X^T = (X, Z).$$

From `pim()`: $\hat{\beta}_X = -0.88$ and $\hat{\beta}_Z = 0.082$

$$\hat{P}(Y_i < Y_j \mid X_i = 0, X_j = 1, Z_i = Z_j) = \expit(-0.88) = 0.29.$$ 

The estimated probability that a patient receiving anti-depressant drugs will have a worse score (i.e. higher) as compared to a patient not receiving anti-depressant drugs is 29% (95% CI: [0.18, 0.44]). → more likely that patients receiving anti-depressant drugs will be better off.
\[ P(Y_i < Y_j \mid X_i, X_j) = \expit[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad X^T = (X, Z). \]

From \texttt{pim()}: \( \hat{\beta}_X = -0.88 \) and \( \hat{\beta}_Z = 0.082 \)

\[ \hat{P}(Y_i < Y_j \mid X_i = 0, X_j = 1, Z_i = Z_j) = \expit(-0.88) = 0.29. \]

The estimated probability that a patient receiving anti-depressant drugs will have a worse score (i.e. higher) as compared to a patient not receiving anti-depressant drugs is 29\% (95\% CI: [0.18, 0.44]).

→ more likely that patients receiving anti-depressant drugs will be better off.

\[ \hat{P}(Y_i < Y_j \mid X_i = X_j, Z_i = z, Z_j = z+10) = \expit(10 \times 0.082) = 0.70. \]

→ more likely that patients with a higher score at baseline will have a higher score after 3 months.
Conclusions and ongoing/future research
Conclusions:

- PIMs: regression model for the Probabilistic Index 
  \[ P(Y_i < Y_j \mid X_i, X_j) \].
- Extends the Wilcoxon–Mann–Whitney test in a similar fashion as that the linear model extends the two-sample t-test.
- Estimation theory is semiparametric.
- Can be used for a variety of applications.

Ongoing/future research:

- Extend PIMs to deal with latent variables (like SEM extends linear models).
- Study what type of PIMs make sense for discrete ordinal outcomes.
- Assessing goodness-of-fit.
References


De Neve, J., Thas, O. and Gerds, T. Unified effect measures for semiparametric linear transformation models.


Thank you.
Jan.DeNeve@UGent.be