### Probabilistic Index Models

Jan De Neve Department of Data Analysis Ghent University

M3 Storrs, Conneticut, USA May 23, 2017

Jan.DeNeve@UGent.be

Introduction

The PI

PIMs

Illustration

Conclusions

### Introduction

PIMs

Illustration

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

- Thas, O., De Neve, J., Clement, L. and Ottoy, J.P. (2012) Probabilistic Index Models (with Discussion). *JRSS-B*, 74, 623–671.
- De Neve, J. and Thas, O. (2015) A Regression Framework for Rank Tests Based on the Probabilistic Index Model. *JASA*, 110, 1276–1283.

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.

De Neve, J., Thas, O., Ottoy, J.P. and Clement, L. (2013) An extension of the Wilcoxon-Mann-Whitney test for analyzing RT-qPCR data. *SAGMB*, 12, 333–346.

De Neve, J., Meys, J., Ottoy, J..P., Clement, L. and Thas, O. (2014) unifiedWMWqPCR: the unified Wilcoxon–Mann–Whitney test for analyzing RT-qPCR data in R. *Bioinformatics*, 30, 2494–2495.

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  $P(Y_i < Y_j | X_i, X_j)$  with  $(Y_i, X_i^T)$  and  $(Y_j, X_j)$  i.i.d.

Introduction

The PI

PIMs

Illustration

Conclusions

### 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.

Everitt and Hothorn (2015). HSAUR: A Handbook of Statistical Analyses Using R (1st Edition) J. Proudfoot, D. Goldberg and A. Mann (2003). Computerised, interactive, multimedia CBT reduced anxiety and depression in general practice: A RCT. Psychological Medicine, 33, 217227.

### 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?

ntroduction The Pl	PIMs	Illustration	Conclusions
--------------------	------	--------------	-------------

#### Both treatments seem to have a positive effect.



treatment and time

Introduction The PI	PIMs	Illustration	Conclusions
---------------------	------	--------------	-------------

### BtheB does a slightly better job



treatment and time

PIMs

Illustration

#### Modest deviation from normality



PIMs

Illustration

### Is there a difference between the treatments (X) in terms of depression (Y)?

$$H_0: \operatorname{E}(Y \mid X = TAU) = \operatorname{E}(Y \mid X = BtheB), \quad H_A: \operatorname{not} H_0.$$

Two-sample t-test p-value:

- Welch: 0.041
- Permutation: 0.042.

95% CI for E(Y | X = TAU) - E(Y | X = BtheB):

[0.23, 11.05]

ntroduction The PI	PIMs	Illustration	Conclusions
--------------------	------	--------------	-------------

### The effect of the outlier



treatment and time

PIMs

Illustration

Conclusions

#### 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  $E(Y \mid X = TAU) - E(Y \mid X = BtheB)$ :

[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?

PIMs

Illustration

Conclusions

Test statistic associated with the WMW test:

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

 $U = \frac{1}{n_B n_T} \sum_{i} \sum_{j} I\left(Y_i^{BtheB} < Y_j^{TAU}\right), I(TRUE) = 1, I(FALSE) = 0,$ 

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

PIMs

Illustration

Conclusions

Test statistic associated with the WMW test:

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

 $U = \frac{1}{n_B n_T} \sum_{i} \sum_{j} I\left(Y_i^{BtheB} < Y_j^{TAU}\right), I(TRUE) = 1, I(FALSE) = 0,$ 

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

 $\mathbf{E}(U) = \mathbf{P}\left(Y^{BtheB} < Y^{TAU}\right),$ 

it follows that the WMW-test is associated with

$$H_0: F_{BtheB} = F_{TAU}$$
  $H_A: P\left(Y^{BtheB} < Y^{TAU}\right) \neq 0.5.$ 

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:  $\hat{P}(Y^{BtheB} < Y^{TAU}) = 64\% (95\%CI : [51\%, 75\%])$ 

Introduction	The PI	PIMs	Illustration	Conclus
--------------	--------	------	--------------	---------

The WMW test and the PI have some attractive properties

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

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.

 $\max(1-x^2, 0)$ Normal Uniform Logistic t3 Laplace Exp Cauchy t5 0.86 0.95 1 11 1 24 15 19 3  $\infty$ 

ntroduction The PI	PIMs	Illustration	Conclusions
--------------------	------	--------------	-------------

### 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.



Depression score at baseline

PIMs

Illustration

Conclusions

### Assessing the effect of drugs on depression

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

 $\rightarrow$  ignores the baseline score (confounder)

PIMs

Illustration

Conclusions

Assessing the effect of drugs on depression

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

 $\rightarrow$  ignores the baseline score (confounder)

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

 $Im(score.3M \sim drugs + score.0M)$ 

 $\rightarrow$  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:

 $P\left(Y^{Drugs} < Y^{No Drugs}\right)?$ 

Problem: Due to the confounder, we cannot trust the WMW test. Question: Can we embed the WMW test in a regression context? What if we are interested in the PI:

 $P\left(Y^{Drugs} < Y^{No Drugs}\right)?$ 

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:

 $P(Y_i < Y_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = m(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\beta}), \quad (Y_i, \boldsymbol{X}_i^{\mathsf{T}}) \text{ i.i.d.}$ 

- $(Y_i, \boldsymbol{X}_i^T)$   $i = 1, \ldots, n$  i.i.d. sample
- $\boldsymbol{X}_i$  covariate, *p*-dimensional, e.g.  $\boldsymbol{X}_i^T = (\text{drugs, score.0M})$
- $m(\cdot)$  a known function
- $\beta$  the regression coefficient.

Introduction

The PI

PIMs

Illustration

Conclusions

### Probabilistic Index Models

### $P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = m(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\beta}),$

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

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = m(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\beta}),$$

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

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

$$\mathrm{E}(Y_i \mid \boldsymbol{X}_i) = \boldsymbol{X}_i^T \boldsymbol{\beta},$$

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

 $\mathbb{E}(Y_i - Y_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = (\boldsymbol{X}_i - \boldsymbol{X}_j)^T \boldsymbol{\beta}.$ 

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = m(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\beta}),$$

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

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

$$\mathrm{E}(Y_i \mid \boldsymbol{X}_i) = \boldsymbol{X}_i^T \boldsymbol{\beta},$$

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

$$\mathrm{E}(Y_i - Y_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = (\boldsymbol{X}_i - \boldsymbol{X}_j)^T \boldsymbol{\beta}.$$

So maybe the following makes sense

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = g^{-1}[(\boldsymbol{X}_i - \boldsymbol{X}_j)^T \boldsymbol{\beta}],$$

with  $g(\cdot)$  a link-function (e.g. probit or logit) to ensure PI  $\in [0, 1]$ .

## PIMs: connection with other models

PIMs

Illustration

Conclusions

### Connection with other models.

Model 1: the parametric normal linear model:

$$Y_i = \boldsymbol{X}_i^{\mathsf{T}} \boldsymbol{\alpha} + \varepsilon_i, \quad \varepsilon_i \sim \mathsf{N}(0, \sigma^2),$$

PIMs

Illustration

Conclusions

### Connection with other models.

Model 1: the parametric normal linear model:

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\alpha} + \varepsilon_i, \quad \varepsilon_i \sim \mathsf{N}(0, \sigma^2),$$

implies

$$P(Y_{i} < Y_{j} | X_{i}, X_{j})$$

$$= P\left(X_{i}^{T}\alpha + \varepsilon_{i} < X_{j}^{T}\alpha + \varepsilon_{j} | X_{i}, X_{j}\right)$$

$$= P\left(\varepsilon_{i} - \varepsilon_{j} < (X_{j} - X_{i})^{T}\alpha | X_{i}, X_{j}\right) \quad \varepsilon_{i} - \varepsilon_{j} \sim N(0, 2\sigma^{2})$$

$$= P\left(Z < (X_{j} - X_{i})^{T}\frac{\alpha}{\sqrt{2\sigma^{2}}}\right) \quad Z \sim N(0, 1)$$

$$= g^{-1}[(X_{j} - X_{i})^{T}\beta] \quad \text{with} \quad \beta = \frac{\alpha}{\sqrt{2\sigma^{2}}}, \quad g(\cdot) = \text{probit}(\cdot).$$

PIMs

Illustration

### Connection with other models.

Model 2: semiparametric linear transformation model (part 1)

$$h(Y_i) = \boldsymbol{X}_i^T \boldsymbol{\alpha} + \varepsilon_i, \quad \varepsilon_i \sim \mathsf{N}(0, \sigma^2),$$

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

PIMs

Illustration

### Connection with other models.

Model 2: semiparametric linear transformation model (part 1)

$$h(Y_i) = \boldsymbol{X}_i^{\mathsf{T}} \boldsymbol{\alpha} + \varepsilon_i, \quad \varepsilon_i \sim \mathsf{N}(0, \sigma^2),$$

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

$$P(Y_i < Y_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = P(\boldsymbol{h}(Y_i) < \boldsymbol{h}(Y_j) \mid \boldsymbol{X}_i, \boldsymbol{X}_j),$$

if follows that

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = g^{-1}[(\boldsymbol{X}_j - \boldsymbol{X}_i)^T \boldsymbol{\beta}],$$

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

PIMs

Illustration

### 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

$$h(Y_i) = \boldsymbol{X}_i^T \boldsymbol{\alpha} + \varepsilon_i, \quad \varepsilon_i \sim F(e) = 1 - \exp[-\exp(e)],$$

implies the PIM

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = g^{-1}[(\boldsymbol{X}_j - \boldsymbol{X}_i)^T \boldsymbol{\beta}],$$

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

Note: this is related to the Cox proportional hazards model.

### PIMs: estimation theory

PIMs

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = g^{-1}[(\boldsymbol{X}_j - \boldsymbol{X}_i)^T \boldsymbol{\beta}],$$

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

PIMs

Illustration

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = g^{-1}[(\boldsymbol{X}_j - \boldsymbol{X}_i)^T \boldsymbol{\beta}],$$

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

Trick:

$$P(Y_i < Y_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = E(I_{ij} \mid \boldsymbol{X}_i, \boldsymbol{X}_j), \quad I_{ij} = I(Y_i < Y_j)$$

$$\Rightarrow \mathrm{E}\left(I_{ij} \mid \boldsymbol{X}_{i}, \boldsymbol{X}_{j}\right) = g^{-1}(\boldsymbol{X}_{ij}^{T}\boldsymbol{\beta}), \quad \boldsymbol{X}_{ij} = \boldsymbol{X}_{j} - \boldsymbol{X}_{i}.$$

Use glm() on transformed outcomes  $I_{ij}$  and predictors  $\boldsymbol{X}_{ij}$  to estimate  $\beta$ !

### Challenges in the estimation

cross-correlation:

$$\begin{split} I_{ij} &= \mathrm{I}\left(\mathbf{Y}_{i} < \mathbf{Y}_{j}\right) \quad \rightarrow \quad \mathrm{I}\left(\mathbf{Y}_{i} < \mathbf{Y}_{l}\right) \\ & \rightarrow \quad \mathrm{I}\left(\mathbf{Y}_{j} < \mathbf{Y}_{l}\right) \\ & \rightarrow \quad \mathrm{I}\left(\mathbf{Y}_{k} < \mathbf{Y}_{i}\right) \\ & \rightarrow \quad \mathrm{I}\left(\mathbf{Y}_{k} < \mathbf{Y}_{j}\right) \end{split}$$

#### Consequences:

- you have to prove that glm() gives consistent estimators.
- provide consistent sandwich estimator for  $\operatorname{Var}\left(\hat{oldsymbol{eta}}
  ight)$  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<sub>i</sub>: depression score at 3 months.
- $X_i$ : anti-depressant drugs (no = 0, yes = 1).

Consider the PIM

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

### Two-sample design

- Y<sub>i</sub>: depression score at 3 months.
- $X_i$ : anti-depressant drugs (no = 0, yes = 1).

Consider the PIM

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

$$\begin{aligned} & \rightarrow \operatorname{expit}(\beta) = \operatorname{P}\left(Y_i < Y_j \mid X_i = 0, X_j = 1\right) = \operatorname{P}\left(Y^{no} < Y^{yes}\right) \\ & \rightarrow \operatorname{expit}(\hat{\beta}) = \frac{1}{n_{no}n_{yes}} \sum_i \sum_j \operatorname{I}\left(Y_i^{no} < Y_j^{yes}\right) = U. \end{aligned}$$

- Wilcoxon-Mann-Whitney test is a special case of a PIM.
- PIM sandwich estimator for  $Var(\hat{\beta})$  allows for Wald-type tests and the construction of confidence intervals.
- Similar results hold for the Kruskal-Wallis, Friedman, Jonckheere-Terpstra, ... rank tests.

Introduction

The PI

PIMs

Illustration

Conclusions

### Return to the BtheB study

- Y<sub>i</sub>: depression score at 3 months.
- $X_i$ : anti-depressant drugs (no = 0, yes = 1).
- $Z_i$ : depression score at baseline.

Consider the PIM

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

- Y<sub>i</sub>: depression score at 3 months.
- $X_i$ : anti-depressant drugs (no = 0, yes = 1).
- $Z_i$ : depression score at baseline.

Consider the PIM

 $P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = expit[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad \boldsymbol{X}^T = (X, Z).$ In R via library('pim')

```
> 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 ***
```

P 
$$(Y_i < Y_j | \mathbf{X}_i, \mathbf{X}_j)$$
 = expit $[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad \mathbf{X}^T = (X, Z).$   
From pim():  $\hat{\beta}_X = -0.88$  and  $\hat{\beta}_Z = 0.082$   
 $\hat{P}(Y_i < Y_j | X_i = 0, X_j = 1, Z_i = Z_j) = expit(-0.88) = 0.29.$ 

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]).

PIMs

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = \exp[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad \boldsymbol{X}^T = (X, Z).$$
  
From pim():  $\hat{\beta}_X = -0.88$  and  $\hat{\beta}_Z = 0.082$ 

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

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]).  $\rightarrow$  more likely that patients receiving anti-depressant drugs will be better off.

$$P(Y_i < Y_j | \boldsymbol{X}_i, \boldsymbol{X}_j) = expit[(X_j - X_i)\beta_X + (Z_j - Z_i)\beta_Z], \quad \boldsymbol{X}^{T} = (X, Z).$$
  
From pim():  $\hat{\beta}_X = -0.88$  and  $\hat{\beta}_Z = 0.082$ 

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

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]).  $\rightarrow$  more likely that patients receiving anti-depressant drugs will be better off.

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

 $\rightarrow$  more likely that patients with a higher score at baseline will have a higher scare after 3 months.

Introduction

The PI

PIMs

Illustration

Conclusions

# Conclusions and ongoing/future research

Introd	uction
	accion

PIMs

Illustration

Conclusions

### Conclusions:

- PIMs: regression model for the Probabilistic Index  $P(Y_i < Y_j | X_i, X_j)$ .
- 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

Thas, O., De Neve, J., Clement, L. and Ottoy, J.P. (2012) Probabilistic Index Models (with Discussion). JRSS-B, 74, 623–671.

De Neve, J. and Thas, O. (2013) Goodness-of-Fit Methods for Probabilistic Index Models (213). Comm. in Stat. - Theory and Methods, 42, 1193–1207.

De Neve, J., Thas, O., Ottoy, J.P. and Clement, L. (2013) An extension of the Wilcoxon-Mann-Whitney test for analyzing RT-qPCR data. *SAGMB*, 12, 333–346.

De Neve, J., Meys, J., Ottoy, J..P., Clement, L. and Thas, O. (2014) unifiedWMWqPCR: the unified Wilcoxon–Mann–Whitney test for analyzing RT-qPCR data in R. *Bioinformatics*, 30, 2494–2495.

Vermeulen, K., Thas, O., Vansteelandt, S. (2014) Increasing the power of the Mann–Whitney test in randomized experiments through flexible covariate adjustment. *Stat. Med.* DOI: 10.1002/sim.6386

De Neve, J. and Thas, O. (2015) A Regression Framework for Rank Tests Based on the Probabilistic Index Model. JASA, 110, 1276–1283.

Vermeulen, K, Amorim, G, De Neve, J., Thas, O. and Vansteelandt, S. Semiparametric estimation of probabilistic index models: efficiency and bias. Under review.

Amorim, G., Thas, O., Vermeulen, K, Vansteelandt, S., and De Neve, J. Small Sample Inference for Probabilistic Index Models. Under review.

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

Meys, J, De Neve, J. and Sabbe, N. pim: Fit Probabilistic Index Models. R package version 2.0.0.2. https://CRAN.R-project.org/package=pim

> Thank you. Jan.DeNeve@UGent.be