A Hierarchical Item Response Model for Cognitive Diagnosis

Mark Hansen & Li Cai
University of California, Los Angeles

International Meeting of the Psychometric Society
July 24, 2013

National Center for Research
on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies | cresst.org

What this presentation is not about

- Diagnostic models for examinees nested within groups

  von Davier, M. (2007). Hierarchical general diagnostic models. ETS Research Report RR-07-19. Educational Testing Service: Princeton, NJ.

  von Davier, M. (2010). Hierarchical mixtures of diagnostic models. *Psychological Test and Assessment Modeling, 52,* 8-28.

- Attribute hierarchies

  Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2004). The attribute hierarchy method (AHM) for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement, 3,* 205-237.

  Templin, J. L., & Bradshaw, L. (in press). Hierarchical Diagnostic Classification Models: A Family of Models for Estimating and Testing Attribute Hierarchies. *Psychometrika.*

What this presentation is about

- Diagnostic classification models that are hierarchical in the factor analytic sense (having independent tiers or layers)

  Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53-61.

  Wherry, R. J. (1959). Hierarchical factor solutions without rotation. *Psychometrika, 24,* 45-51.

  Yung, Y.-F., Thissen, D., & McCleod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64,* 113-128.

- Or, alternatively, hierarchical item factor analysis models in which the latent variables in the "primary" tier are discrete rather than continuous

Motivation

- Cognitive diagnosis models are highly appealing.
- However, the utility of these models may be undermined if they don't fit real data.
- Of concern here are item dependencies that are unrelated to the attributes to be measured by the diagnostic model:
    - Testlets
    - Items with highly redundant/overlapping content
    - Idiosyncratic response style
    - Repeated items (in longitudinal testing contexts)
    - Other design features (mode-of-administration, etc.)
- In this research, we examine the consequences of such dependencies and develop a diagnostic modeling framework that allows them to be modeled explicitly.

Outline

- Local dependence in diagnostic classifications models
  - an illustration
  - evaluation the assumption of local item independence
  - an alternative model
- Is this a problem that deserves attention?
  - we observe evidence of local dependence
  - ignoring dependence may…
    - impair classification decisions
    - lead to mischaracterizations of classification certainty
    - obscure other sorts of misspecification
- Final Comments

Outline

- Local dependence in diagnostic classifications models
  - an illustration
  - evaluation the assumption of local item independence
  - an alternative model
- Is this a problem that deserves attention?
  - we observe evidence of local dependence
  - ignoring dependence may…
    - impair classification decisions
    - lead to mischaracterizations of classification certainty
    - obscure other sorts of misspecification
- Final Comments

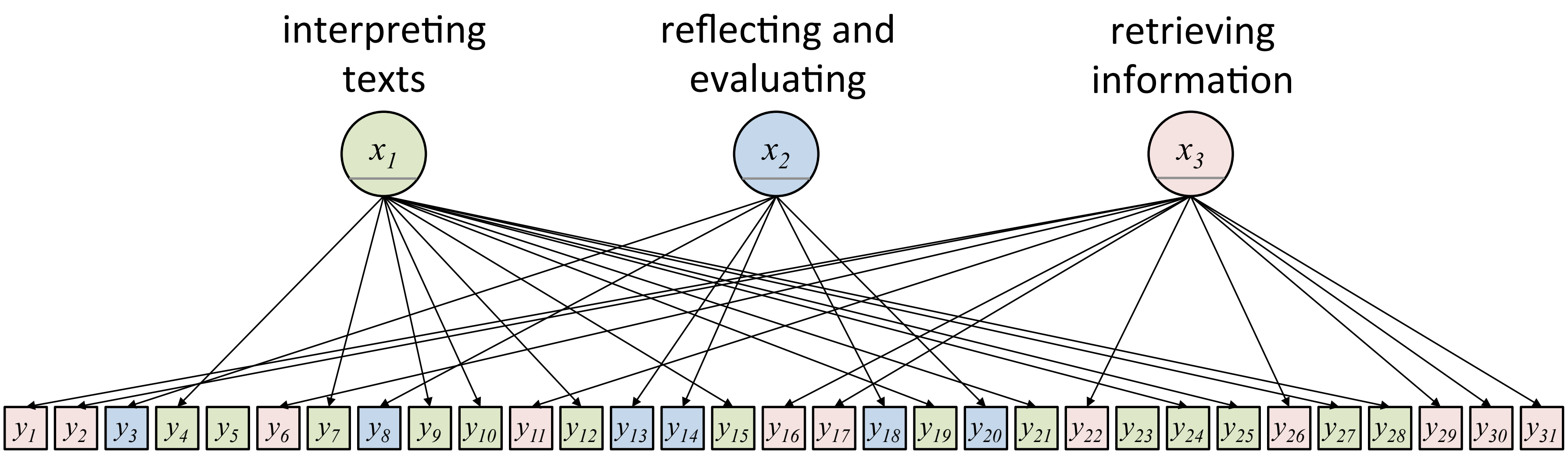
An Illustration

2000 PISA reading, test booklet 8 (31 items), 3000 randomly selected US respondents

- Item responses may be related to the targeted attributes using a standard model for cognitive diagnosis (e.g., von Davier, 2005; Henson, Templin, & Willse, 2009):

$$P(y_i = 1 | \mathbf{x}) = \pi_i(\mathbf{x}) = \frac{1}{1+\exp[-(\alpha_i + h(\boldsymbol{\gamma}_i, \mathbf{q}_i, \mathbf{x}))]} \tag{1}$$

3 reading processes (let these be the discrete latent attributes)



$$\mathbf{q}' = \left[\begin{array}{ccccccccccccccccccccccccccccccc}
0&0&0&1&1&0&1&0&1&1&0&1&0&0&1&0&0&0&1&0&1&0&1&1&1&0&1&1&0&0&0\\
0&0&1&0&0&0&0&1&0&0&0&0&1&1&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&0&1&0&0&0&0&1&0&0&0&0&1&1&0&0&0&0&1&0&0&0&1&0&0&1&1&1
\end{array}\right]$$

An Illustration

2000 PISA reading, test booklet 8 (31 items), 3000 randomly selected US respondents

- Item responses may be related to the targeted attributes using a standard model for cognitive diagnosis (e.g., von Davier, 2005; Henson, Templin, & Willse, 2009):

$$P(y_i = 1 | \mathbf{x}) = \pi_i(\mathbf{x}) = \frac{1}{1 + \exp[-(\alpha_i + h(\boldsymbol{\gamma}_i, \mathbf{q}_i, \mathbf{x}))]} \tag{1}$$

- Assuming conditional independence of the item responses, given the attributes, the conditional response pattern probability is

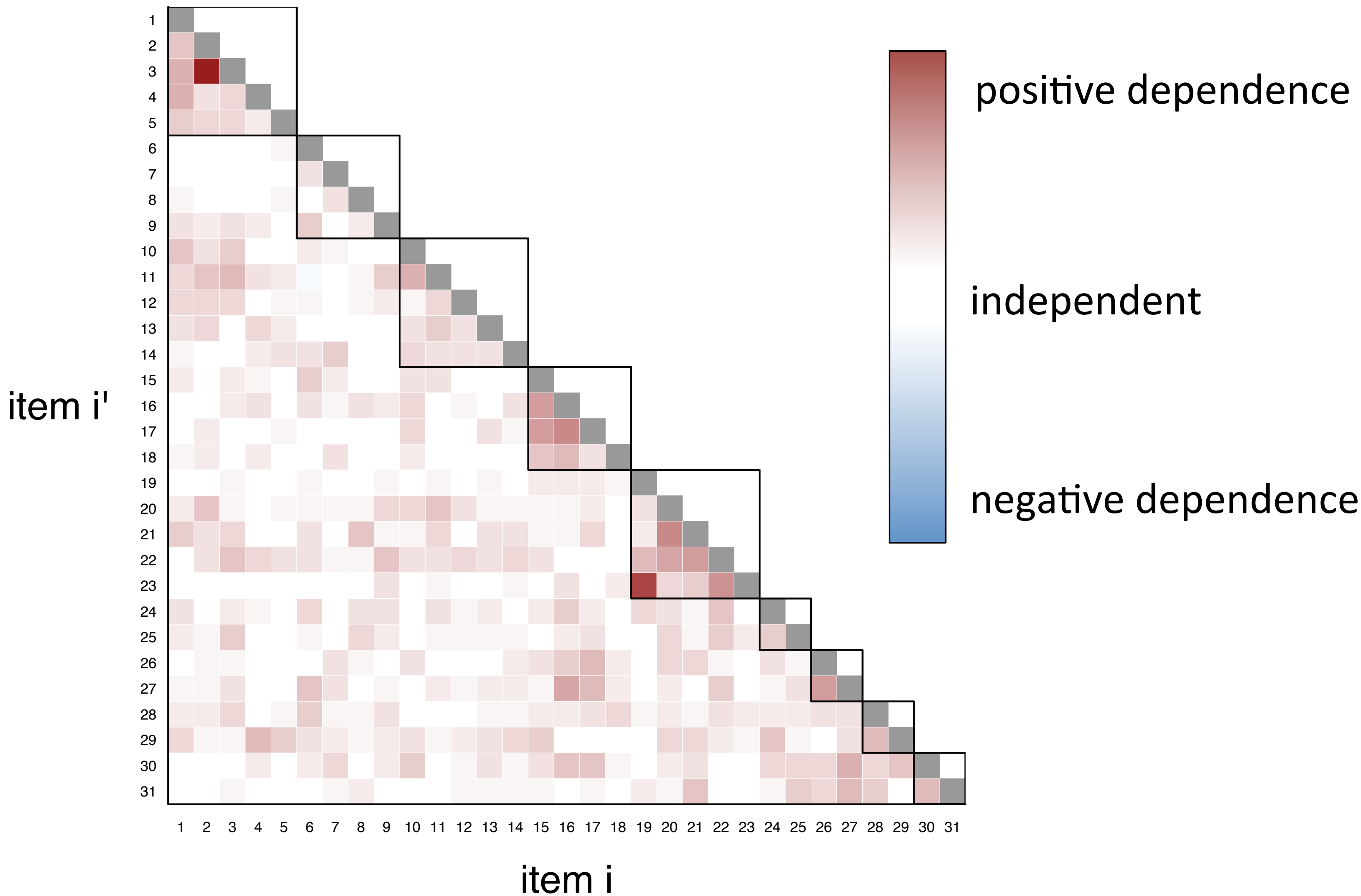
$$P(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^{I} [\pi_i(\mathbf{x})]^{y_i} [1 - \pi_i(\mathbf{x})]^{1-y_i} \tag{2}$$

- However, this assumption of conditional independence may be unrealistic if items fall into clusters based on content subdomains or method-related influences (e.g., testlets).

- We can test the assumption by fitting the model, then comparing the model-implied and observed bivariate response probabilities.

Evaluating the Assumption of Local Item Independence

RMSEA values based on Chen & Thissen (1997) local dependence X^2 for a 3-attribute diagnostic model fit to the PISA dataset; model estimated and LD indices obtained using flexMIRT 2.0 (Cai, 2013)

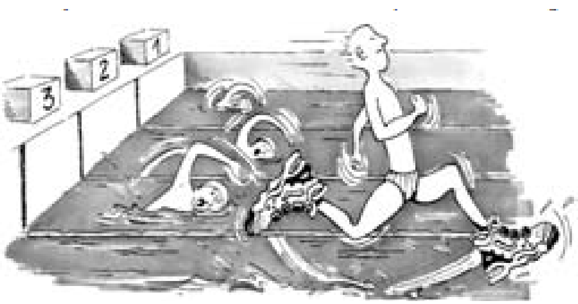


Testlets as a Possible Sources of Inter-Item Dependence

a reading passage/article

Feel good in your runners

For 14 years the Sports Medicine Centre of Lyon (France) has been studying the injuries of young sports players and sports professionals. The study has established that the best course is prevention … and good shoes.



Knocks, falls, wear and tear...

Eighteen per cent of sports players aged 8 to 12 already have heel injuries. The cartilage of a footballer's ankle does not respond well to shocks, and 25% of professionals have discovered for themselves that it is an especially weak point. The cartilage of the delicate knee joint can also be irreparably damaged and if care is not taken right from childhood (10–12 years of age), this can cause premature osteoarthritis. The hip does not escape damage either and, particularly when tired, players run the risk of fractures as a result of falls or collisions.

According to the study, footballers who have been playing for more than ten years have bony outgrowths either on the tibia or on the heel. This is what is known as "footballer's foot", a deformity caused by shoes with soles and ankle parts that are too flexible.

Protect, support, stabilise, absorb

If a shoe is too rigid, it restricts movement. If it is too flexible, it increases the risk of injuries and sprains. A good sports shoe should meet four criteria:

Firstly, it must **provide exterior protection:** resisting knocks from the ball or another player, coping with unevenness in the ground, and keeping the foot warm and dry even when it is freezing cold and raining.

It must **support the foot**, and in particular the ankle joint, to avoid sprains, swelling and other problems, which may even affect the knee.

It must also provide players with good **stability so that they** do not slip on a wet ground or skid on a surface that is too dry.

Finally, it must **absorb shocks,** especially those suffered by volleyball and basketball players who are constantly jumping.

Dry feet

To avoid minor but painful conditions such as blisters or even splits or athlete's foot (fungal infections), the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in. The ideal material for this is leather, which can be waterproofed to prevent the shoe from getting soaked the first time it rains.

4 items related to the passage

Use the article on the previous page to answer the questions below.

QUESTION 7.1

What does the author intend to show in this text?

A. That the quality of many sports shoes has greatly improved.
B. That it is best not to play football if you are under 12 years of age.
C. That young people are suffering more and more injuries due to their poor physical condition.
D. That it is very important for young sports players to wear good sports shoes.

QUESTION 7.2

According to the article, why should sports shoes not be too rigid?

QUESTION 7.3

One part of the article says, "A good sports shoe should meet four criteria."
What are these criteria?

QUESTION 7.4

Look at this sentence from near the end of the article. It is presented here in two parts:

"To avoid minor but painful conditions such as blisters or even splits or athlete's foot (fungal infections),..."	(first part)
"…the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in."	(second part)

What is the relationship between the first and second parts of the sentence?
The second part

A. contradicts the first part.
B. repeats the first part.
C. illustrates the problem described in the first part.
D. gives the solution to the problem described in the first part.

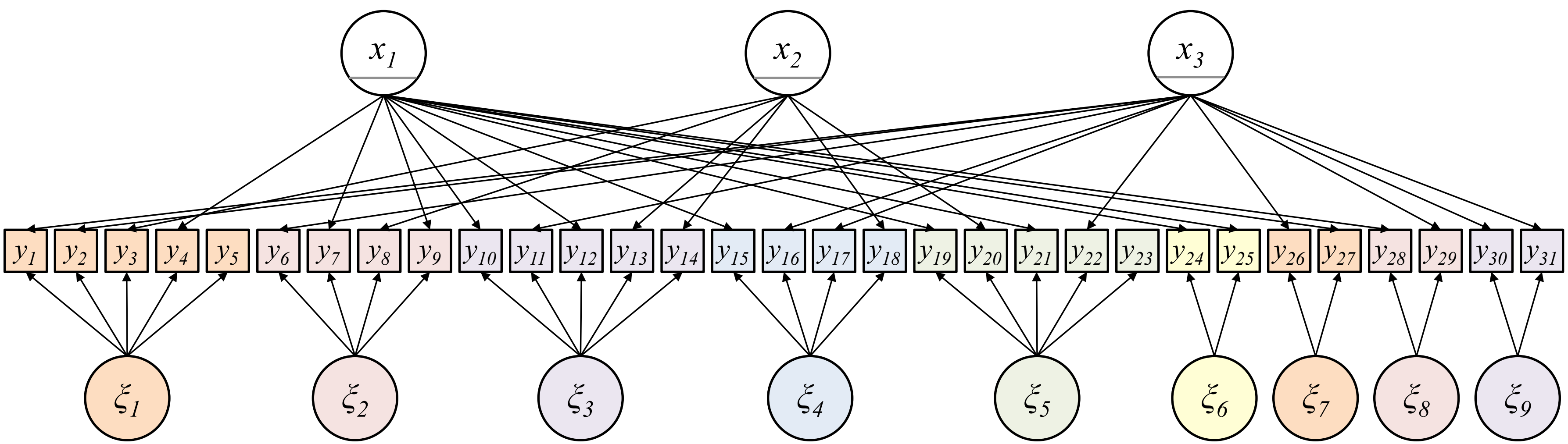
OECD (2009). *Take the Test: Sample Questions from the OECD's PISA Assessment.* (http://www.oecd.org/pisa/pisaproducts/pisa2006/41943106.pdf)

An Alternative Model: Accounting for Sources of Dependence

- A strategy used in item factor analysis is to include random effects to account for potential residual dependence due to the common sources of variation shared by clusters of items (see, e.g., Gibbons & Hedeker, 1992; Wainer, Bradlow, & Wang, 2007; Maydeu-Olivares & Coffman, 2006; Cai, 2010).

- Let there be S mutually exclusive item clusters. We may then specify a diagnosis model with specific dimension (random effect) ξ_s:

$$P(y_i = 1 | \mathbf{x}, \xi_s) = \pi_i(\mathbf{x}, \xi_s) = \frac{1}{1 + \exp[-(\alpha_i + h(\boldsymbol{\gamma}_i, \mathbf{q}_i, \mathbf{x}) + \underline{\beta_{is}\xi_s})]} \quad (3)$$



Estimation of the Alternative Model

- It may now be more reasonable to assume conditional independence of item responses (given attributes and the specific dimensions), such that

$$P(\mathbf{y}|\mathbf{x},\boldsymbol{\xi}) = \prod_{i=1}^{I}[\pi_i(\mathbf{x},\boldsymbol{\xi})]^{y_i}[1-\pi_i(\mathbf{x},\boldsymbol{\xi})]^{1-y_i} \tag{4}$$

- In order to estimate the model, we will need to integrate out the latent dimensions. This could involve rather high-dimensional integration.

$$P(\mathbf{y}|\mathbf{x}) = \underbrace{\int\cdots\int}_{S\text{-fold integration}} \prod_{i=1}^{I}[\pi_i(\mathbf{x},\xi_1,\ldots,\xi_S)]^{y_i}[1-\pi_i(\mathbf{x},\xi_1,\ldots,\xi_S)]^{1-y_i}\, g(\xi_1)\cdots g(\xi_S)\, d\xi_1\cdots d\xi_S \tag{5}$$

- However, we may integrate over the specific dimensions without having to perform the full S-dimensional integration by using analytical dimension reduction (see Gibbons & Hedeker, 1992; Rijmen, 2009).

Estimation of the Alternative Model

- This reduction is possible because each item is influenced by at most one specific dimension and those specific dimensions are conditionally independent.

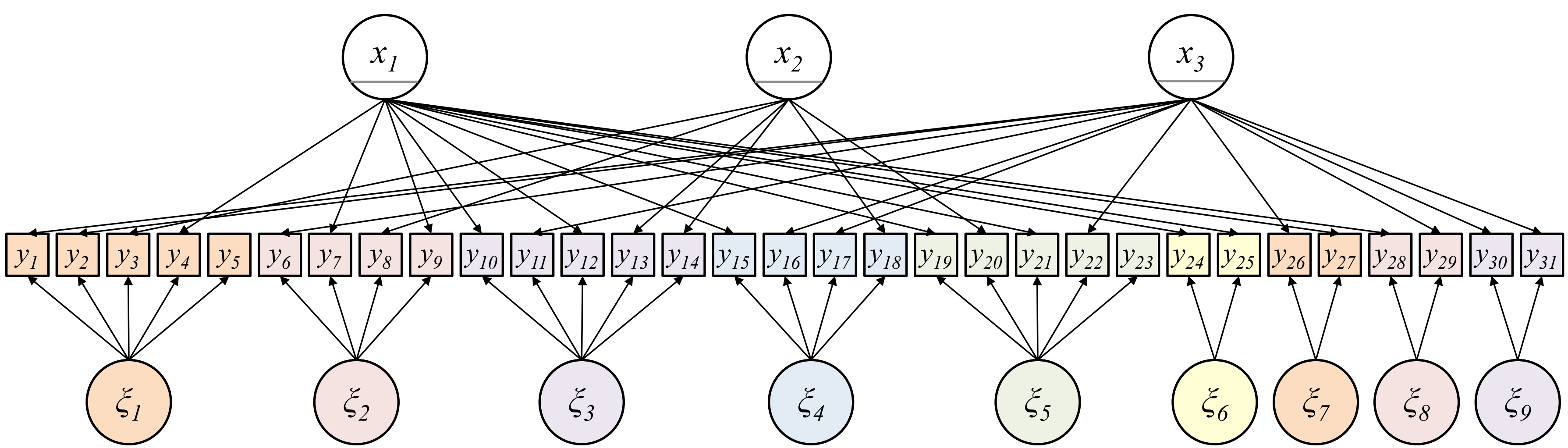
- As a result, we can rearrange the terms of the conditional response pattern probability, from

$$P(\mathbf{y}|\mathbf{x}) = \int \cdots \int \prod_{i=1}^{I} [\pi_i(\mathbf{x}, \xi_1, \dots, \xi_S)]^{y_i} [1 - \pi_i(\mathbf{x}, \xi_1, \dots, \xi_S)]^{1-y_i} g(\xi_1) \cdots g(\xi_S) d\xi_1 \cdots d\xi_S$$

to

$$P(\mathbf{y}|\mathbf{x}) = \prod_{s=1}^{S} \int \prod_{i \in \zeta_s} [\pi_i(\mathbf{x}, \xi_s)]^{y_i} [1 - \pi_i(\mathbf{x}, \xi_s)]^{1-y_i} g(\xi_s) d\xi_s \qquad (6)$$

which is a series of products of one-dimensional integration.



Estimation of the Alternative Model

- As a consequence of dimension reduction, computation time increases linearly in the number of dimensions (rather than exponentially).

- Standard numerical procedures such as the EM algorithm (Dempster, Laird, & Rubin, 1977) may be used to maximize the marginal likelihood.

- This estimation strategy has been implemented in flexMIRT version 2 (Cai, 2013), which was used in the analyses reported here.

Outline

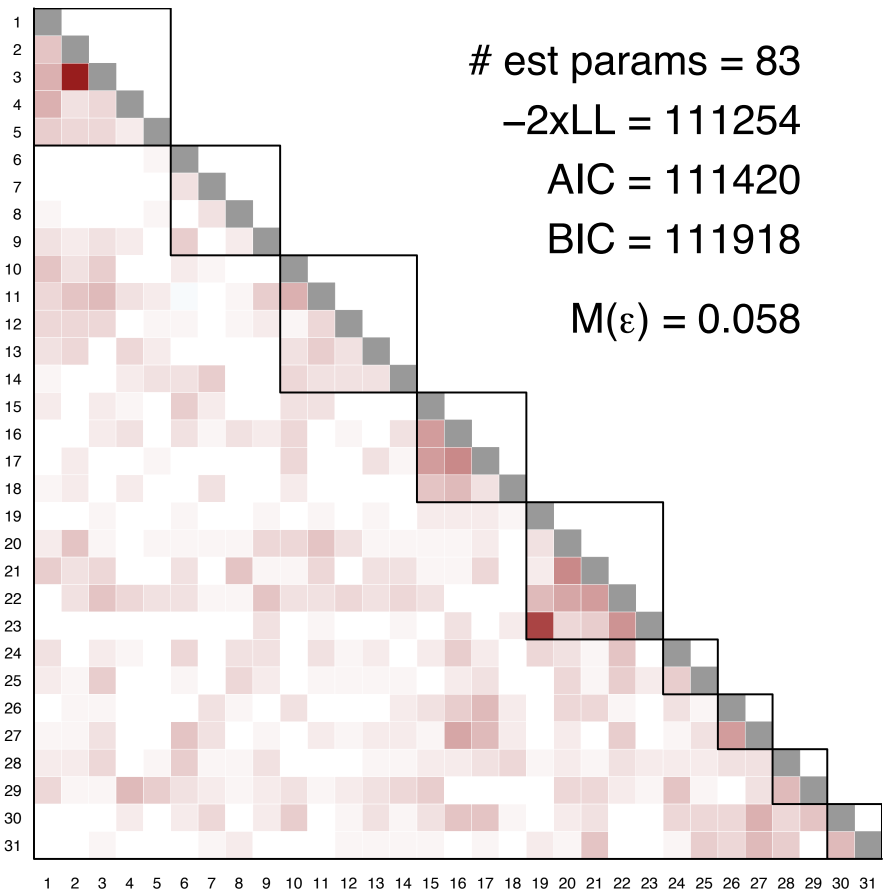
- Local dependence in diagnostic classifications models
  - an illustration
  - evaluation the assumption of local item independence
  - an alternative model
- Is this a problem that deserves attention?
  - we observe evidence of local dependence
  - ignoring dependence may…
    - impair classification decisions
    - lead to mischaracterizations of classification certainty
    - obscure other sorts of misspecification
- Final Comments

Is this a problem that deserves attention?

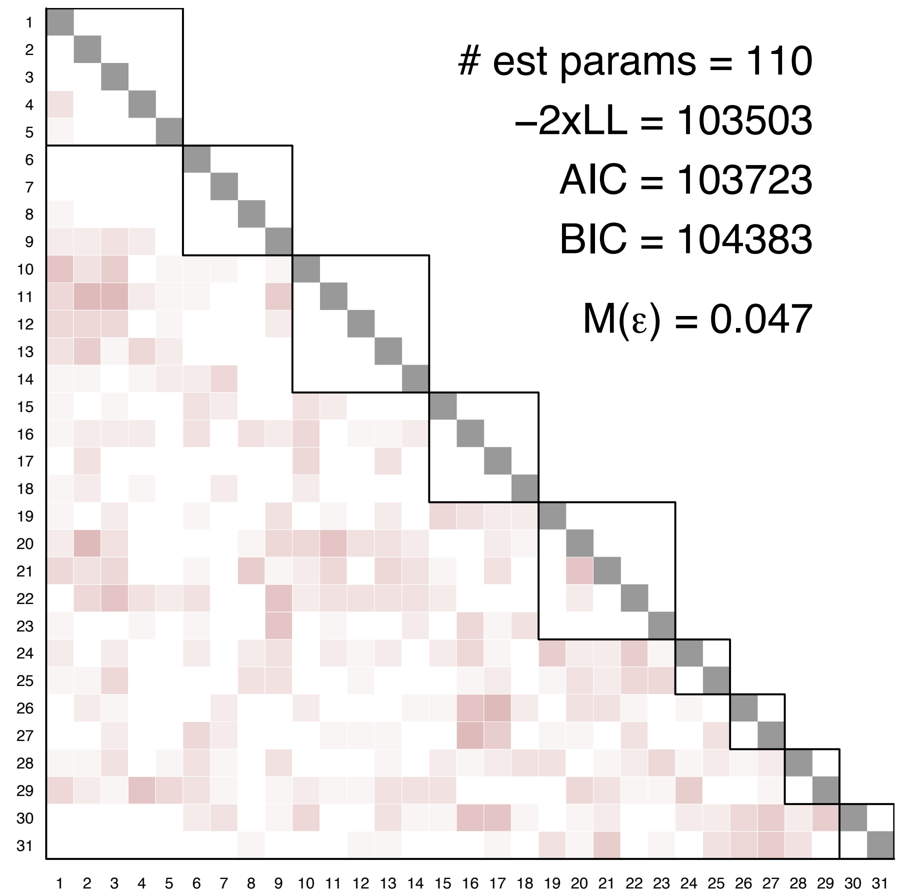
We observe evidence of local dependence

Ex. 1: a testlet-based assessment (PISA Reading - 3 Reading Processes)

standard model



hierarchical model

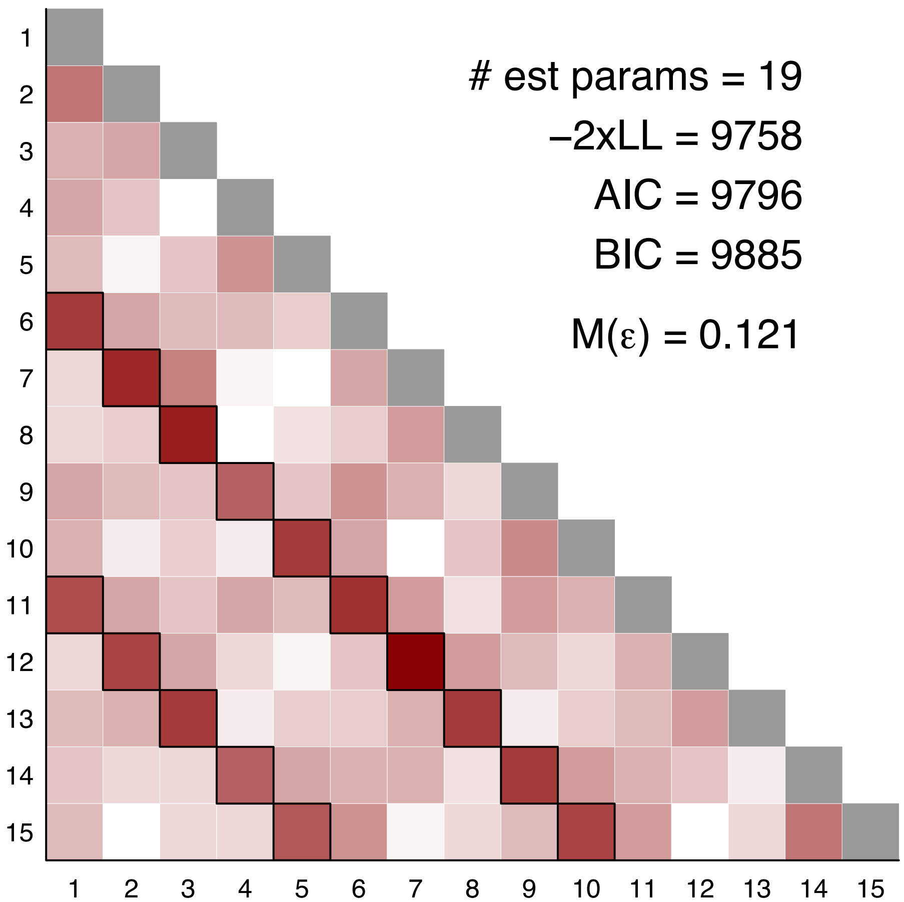


Is this a problem that deserves attention?

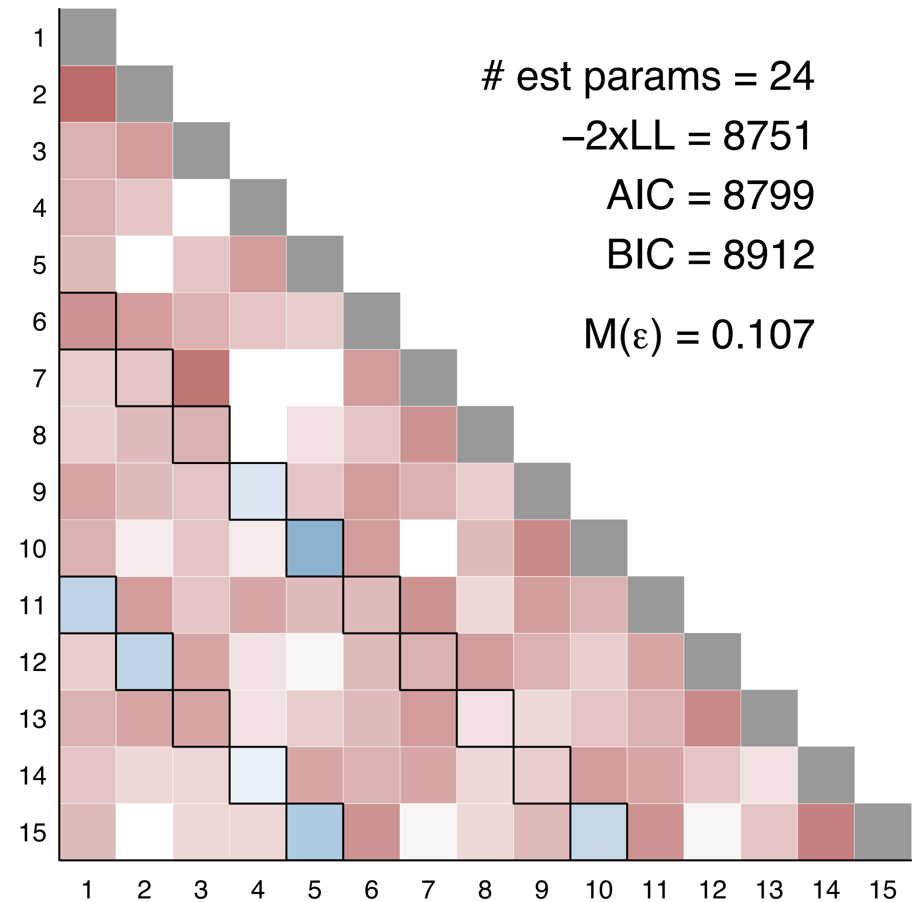
We observe evidence of local dependence

Ex. 2: physical functioning measured at 3 points in time (5 repeated items)

standard model



hierarchical model

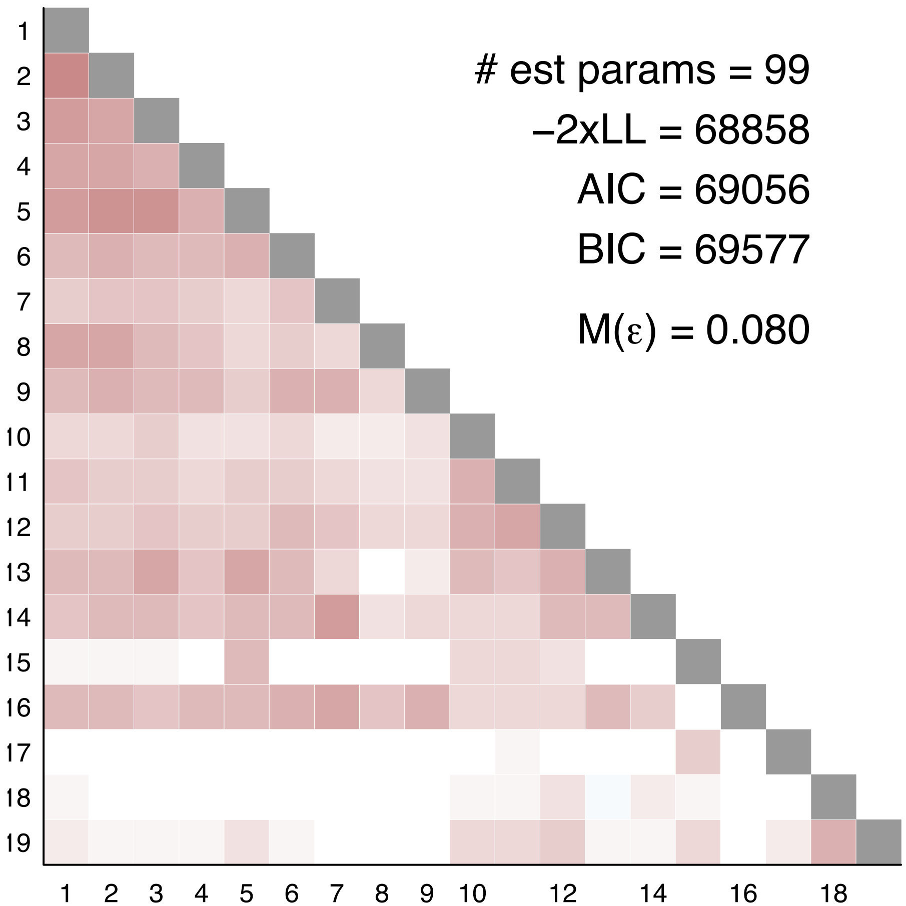


Is this a problem that deserves attention?

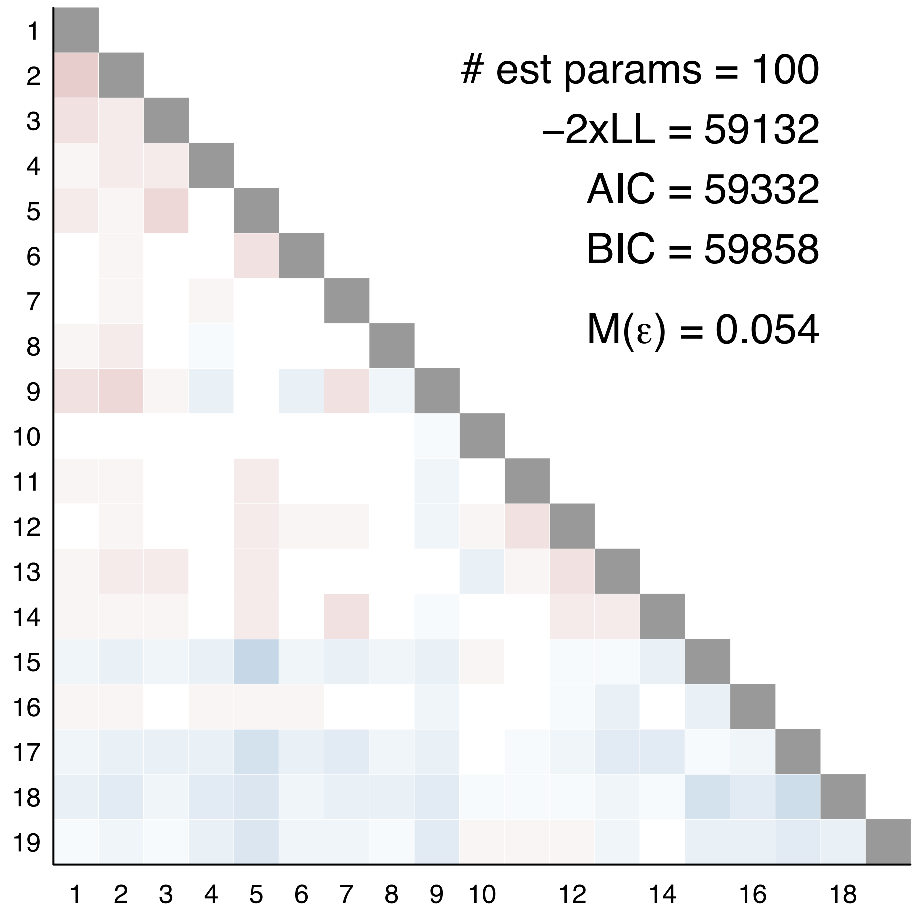
We observe evidence of local dependence

Ex. 3: cigarette dependence (3 attributes: abuse, craving, control)

standard model



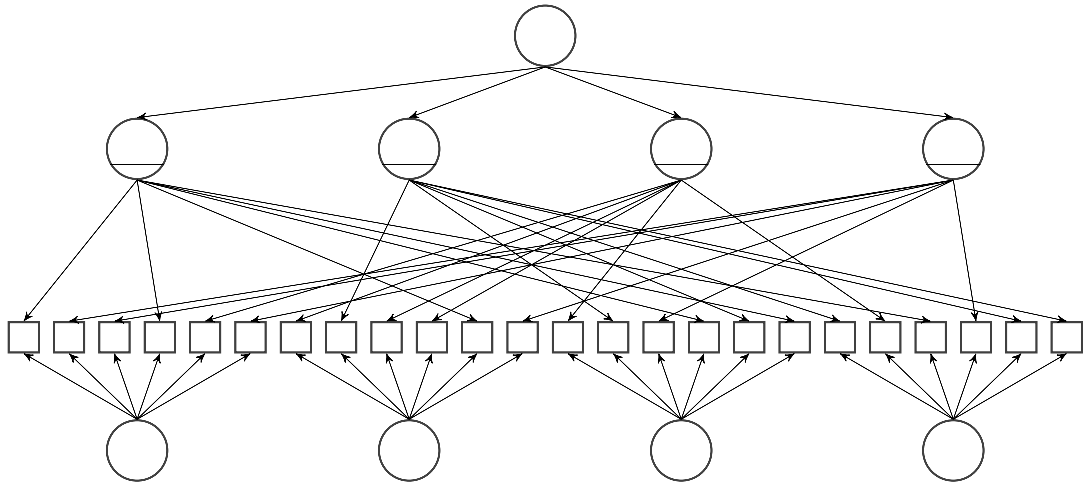
hierarchical model



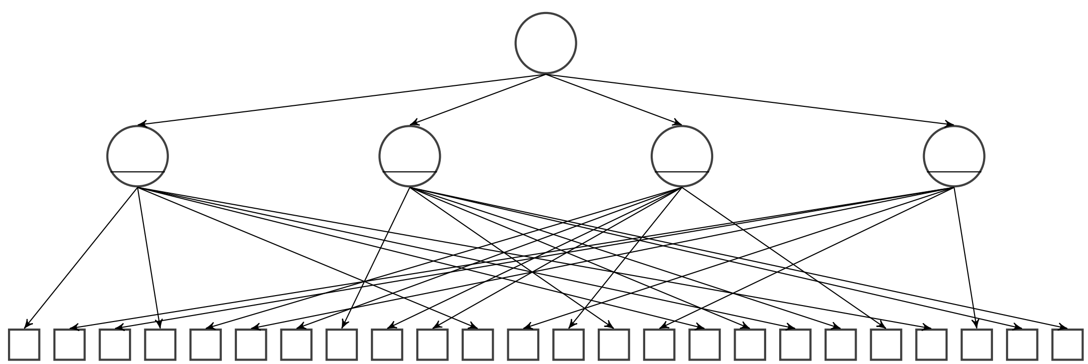
Is this a problem that deserves attention?

Examining Impact of Local Dependence (Simulation Results)

If the data generating model is something like this…



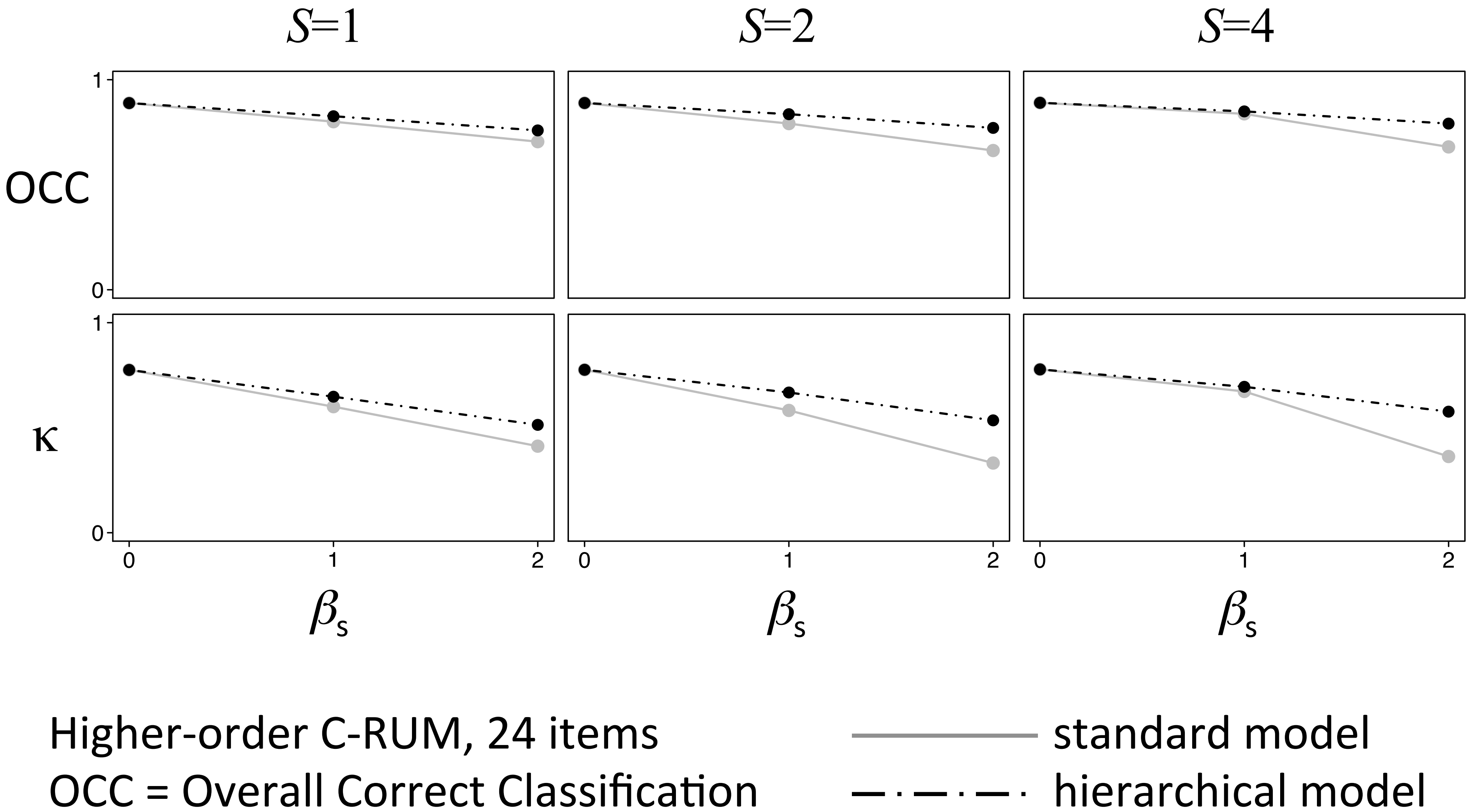
…but the fitted/estimated model is something like this…



…does it really matter?

Examining Impact of Local Dependence (Simulation Results)

Ignoring nuisance dimensions impairs classification decisions



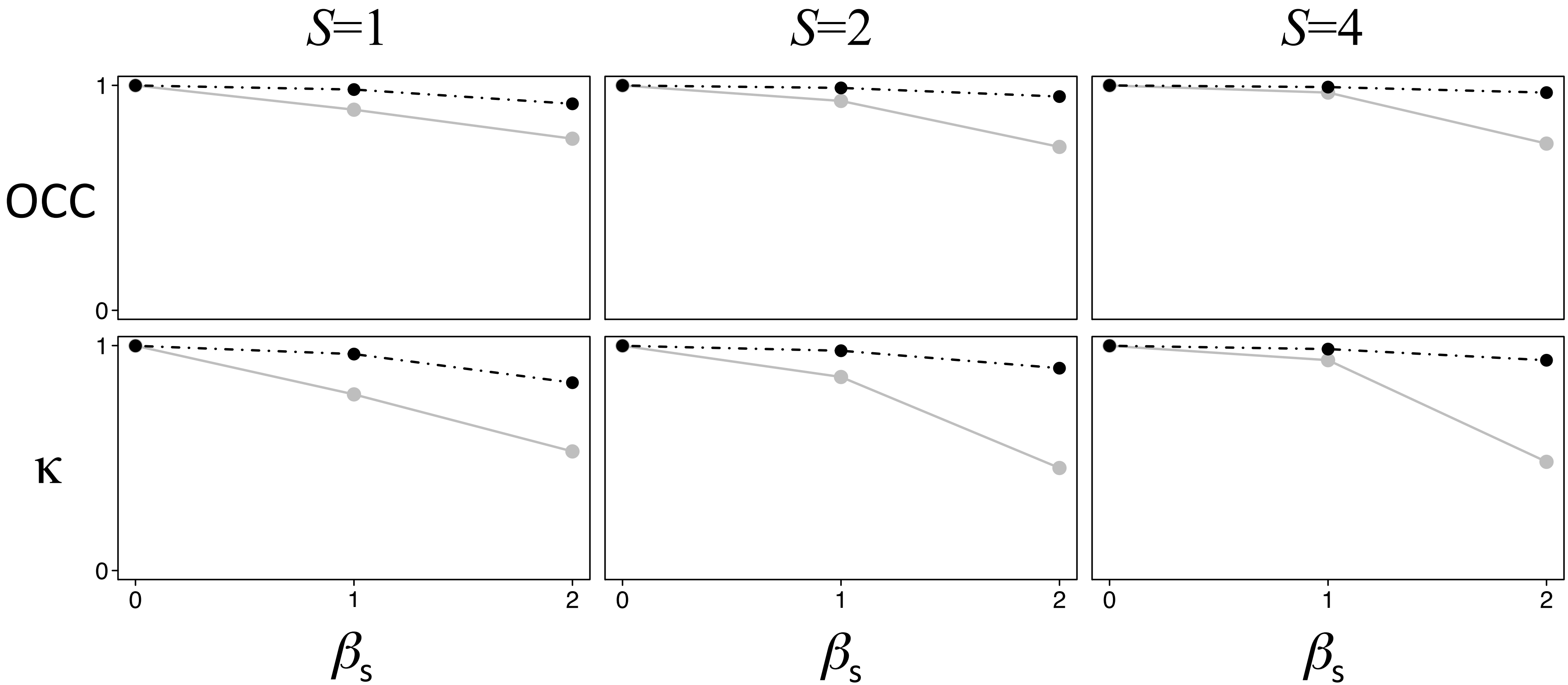
Higher-order C-RUM, 24 items

OCC = Overall Correct Classification

κ = Cohen's kappa

Examining Impact of Local Dependence (Simulation Results)

Ignoring nuisance dimensions impairs classification decisions



Higher-order C-RUM, 120 items
OCC = Overall Correct Classification
κ = Cohen's kappa

standard model
hierarchical model

Examining Impact of Local Dependence (Simulation Results)

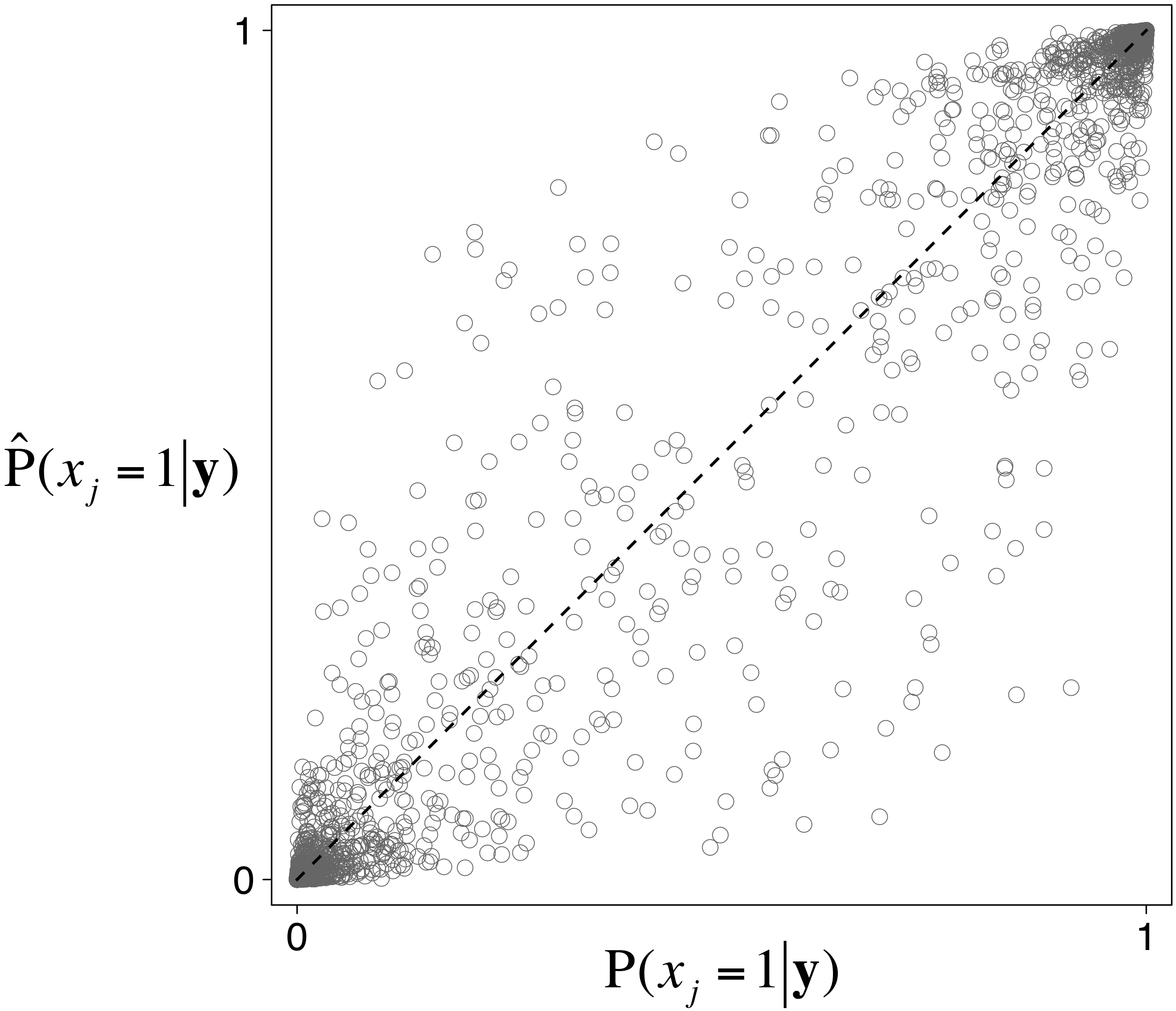
Ignoring nuisance dimensions *leads to mischaracterization of classification certainty*

For each observed response pattern $\mathbf{y}$, compare estimated posterior probability,

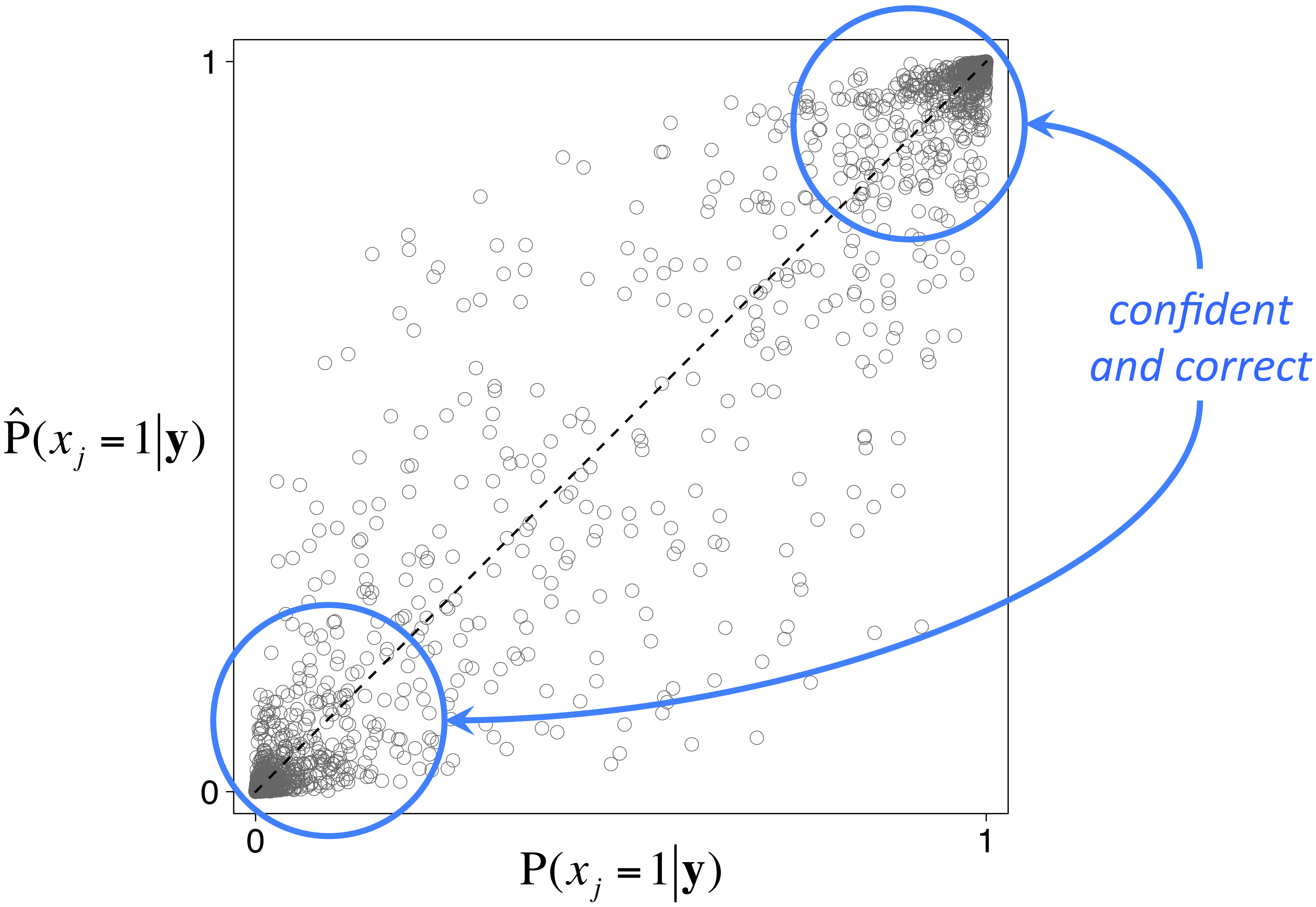
$$\hat{\mathrm{P}}(x_j = 1 | \mathbf{y})$$

with the model-implied proportion possessing attribute j...

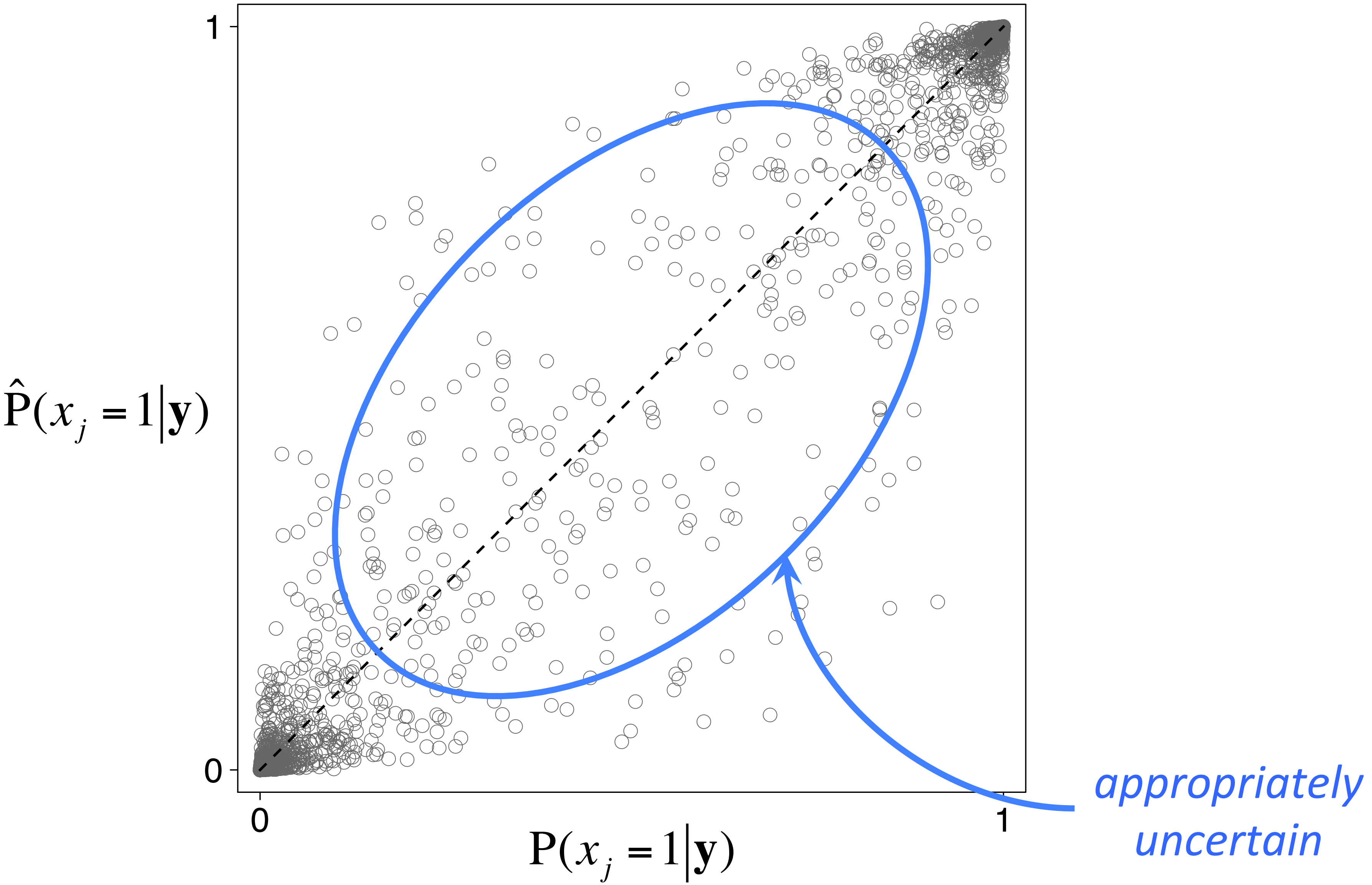
$$\mathrm{P}(x_j = 1 | \mathbf{y})$$



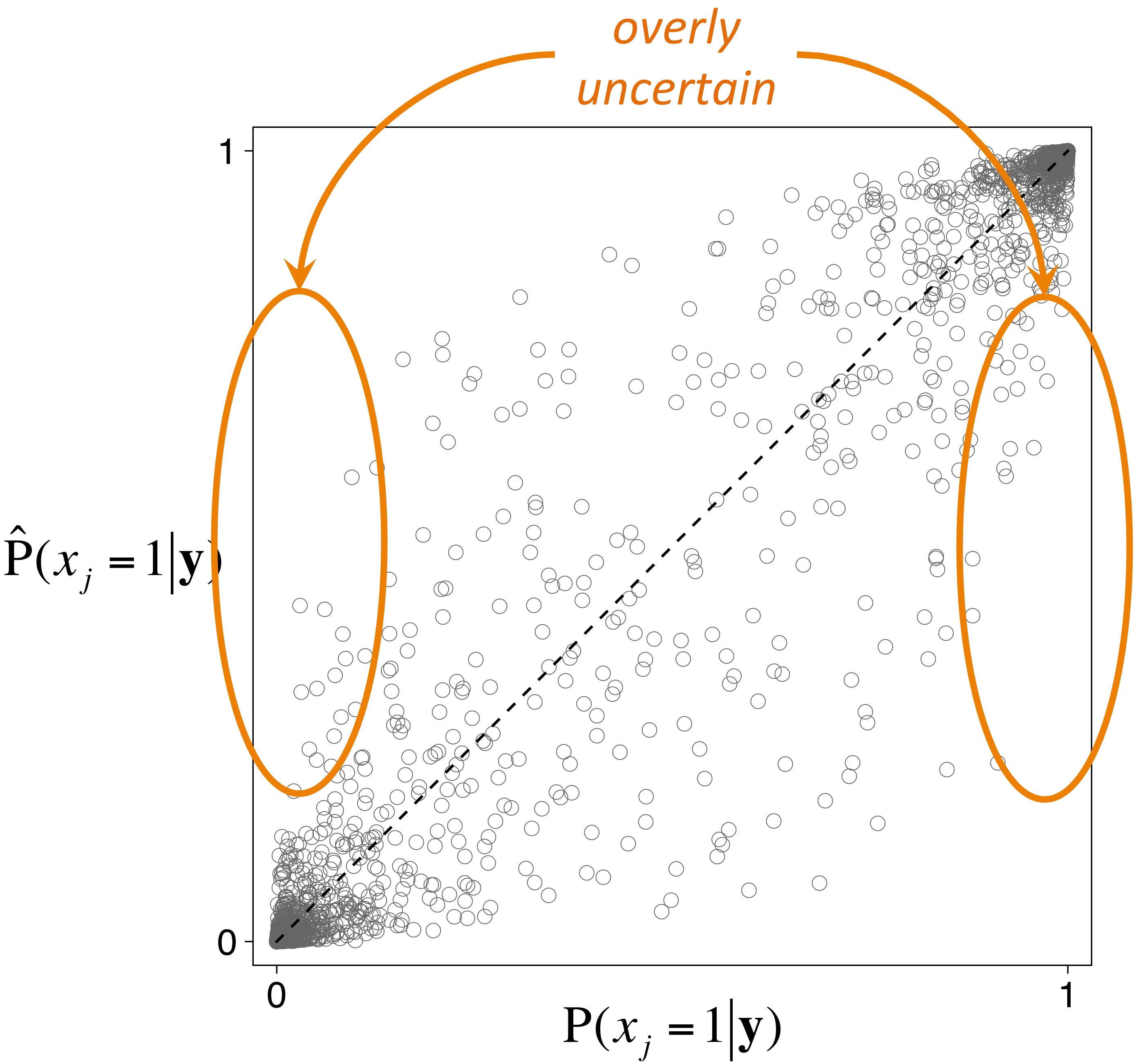
1
0
$\hat{\mathrm{P}}(x_j = 1|\mathbf{y})$
0
1
$\mathrm{P}(x_j = 1|\mathbf{y})$



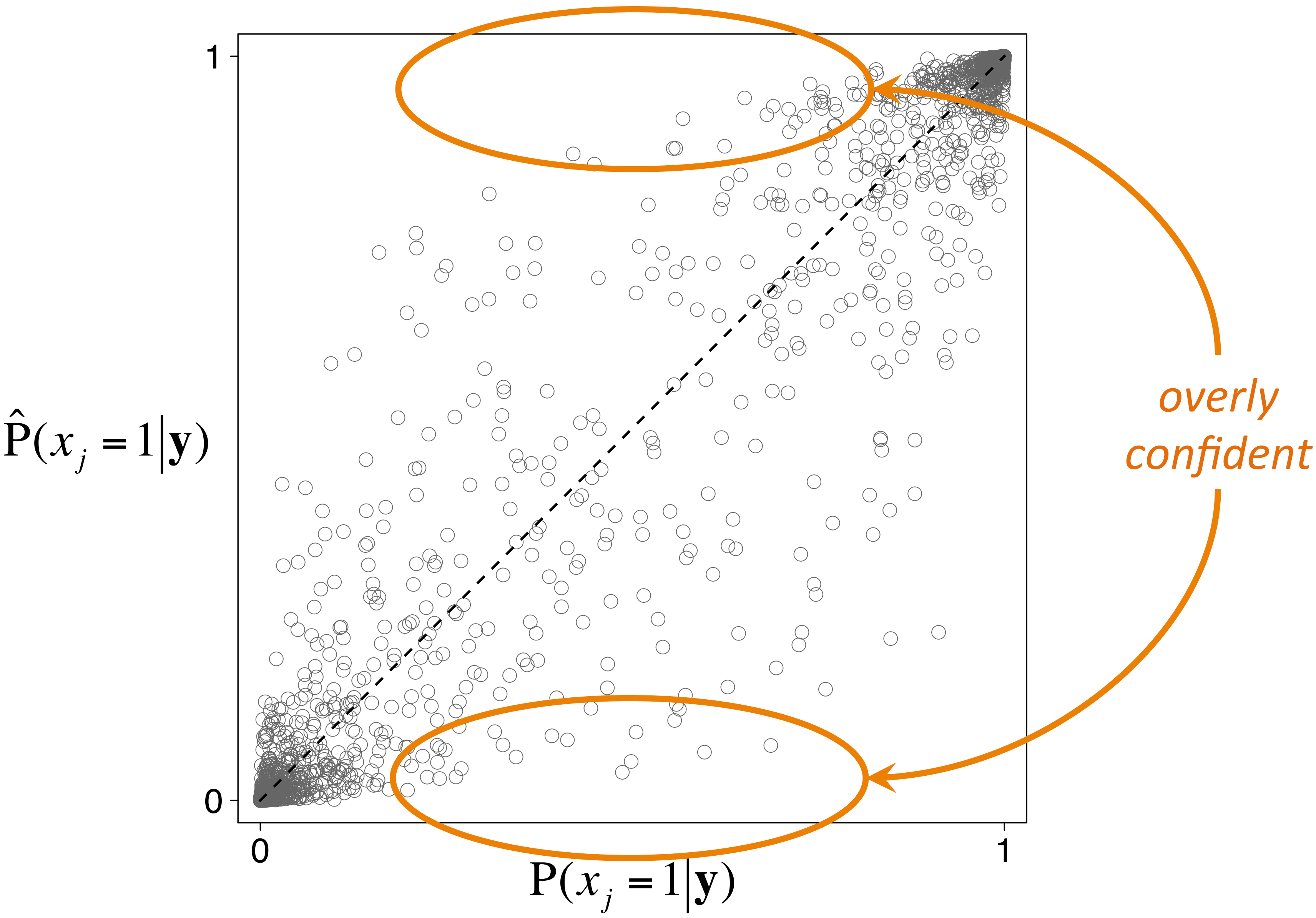
1
0
0
1
$\hat{P}(x_j = 1|\mathbf{y})$
$P(x_j = 1|\mathbf{y})$
confident
and correct



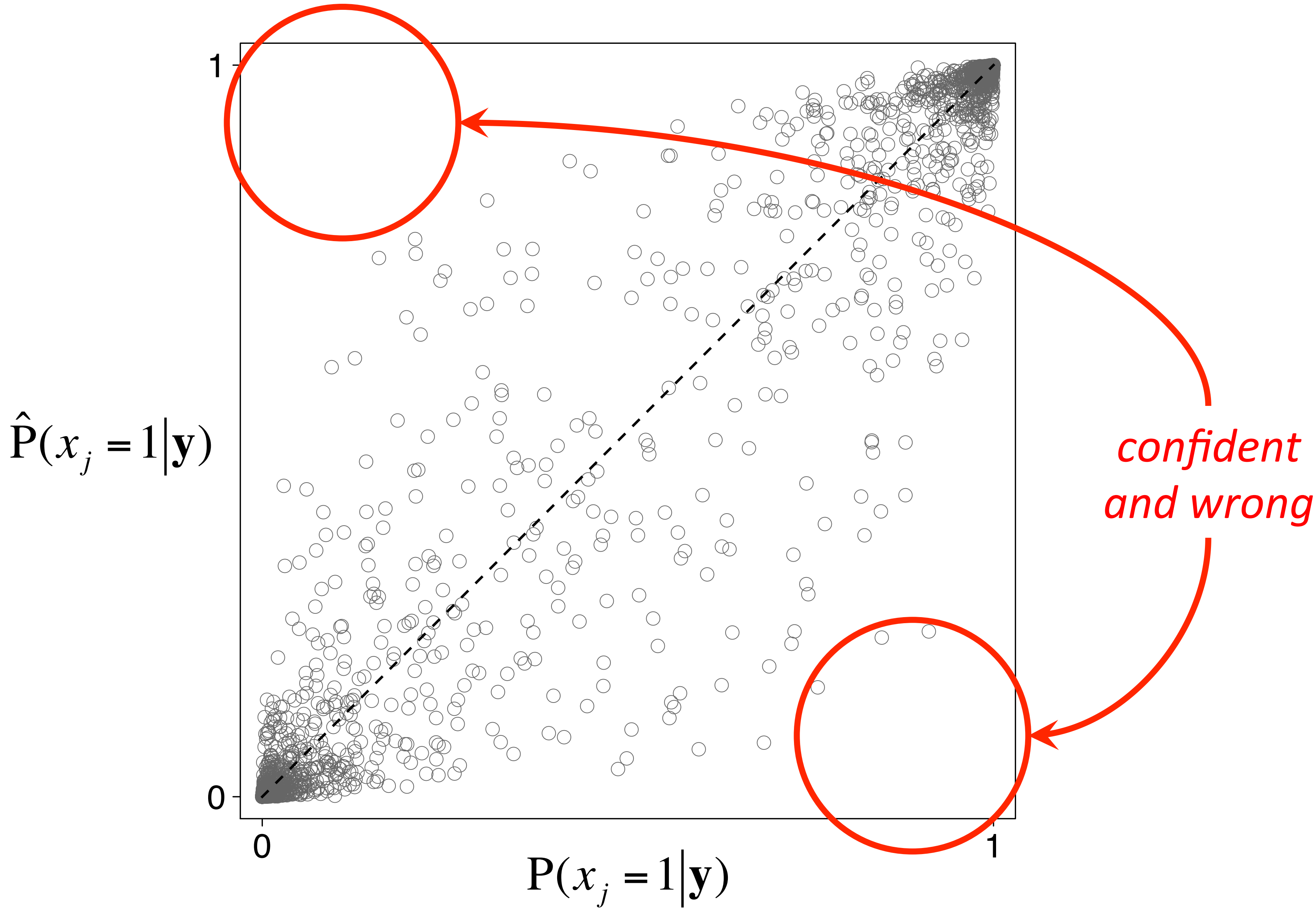
1
0
0
1
$\hat{\mathrm{P}}(x_j = 1|\mathbf{y})$
$\mathrm{P}(x_j = 1|\mathbf{y})$
appropriately
uncertain



overly
uncertain
1
0
0
1
$\hat{P}(x_j = 1|\mathbf{y})$
$P(x_j = 1|\mathbf{y})$

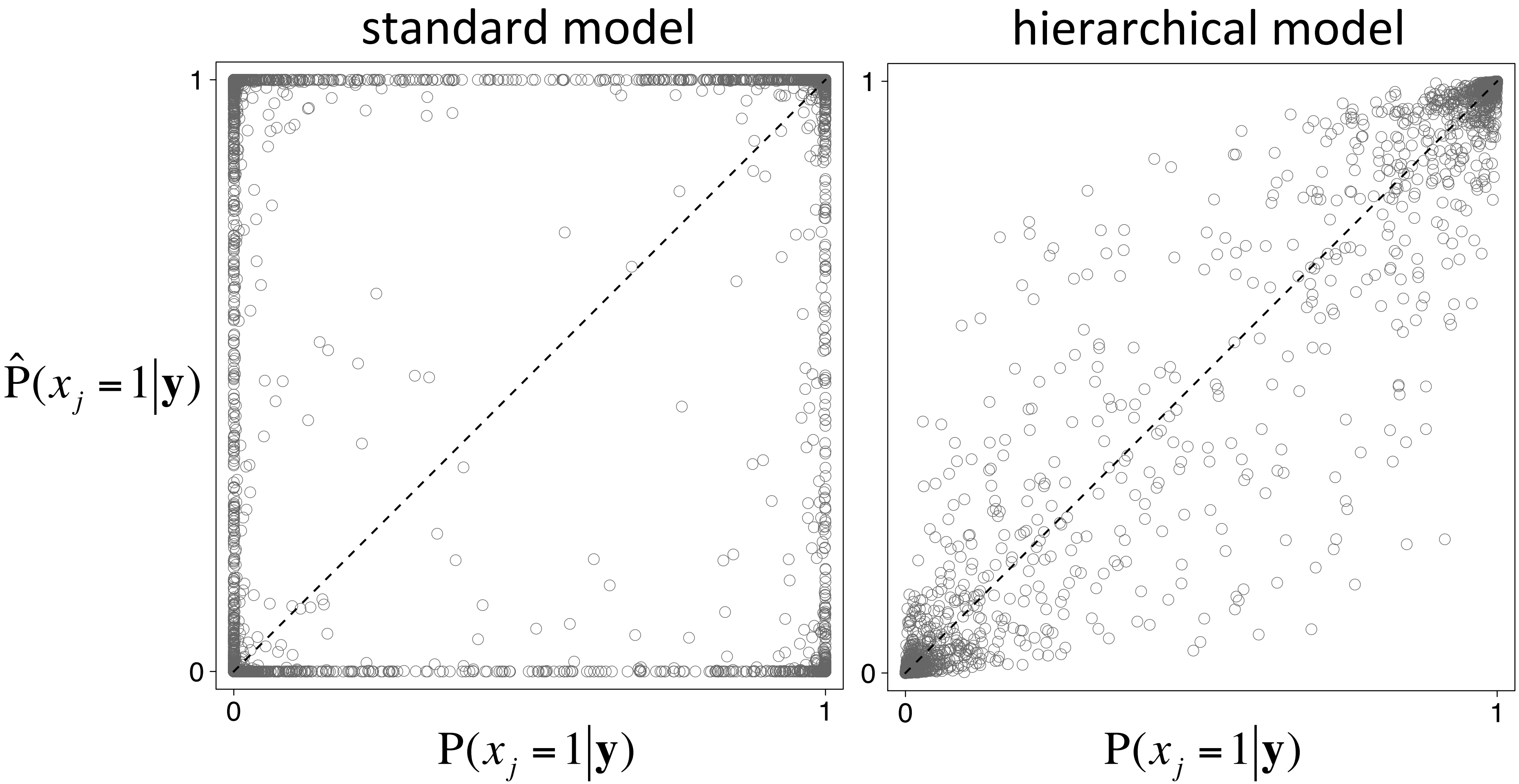


1
0
0
1
$\hat{P}(x_j = 1|\mathbf{y})$
$P(x_j = 1|\mathbf{y})$
overly
confident


1
0
0
1
$\hat{P}(x_j = 1|\mathbf{y})$
$P(x_j = 1|\mathbf{y})$
confident
and wrong

Examining Impact of Local Dependence (Simulation Results)

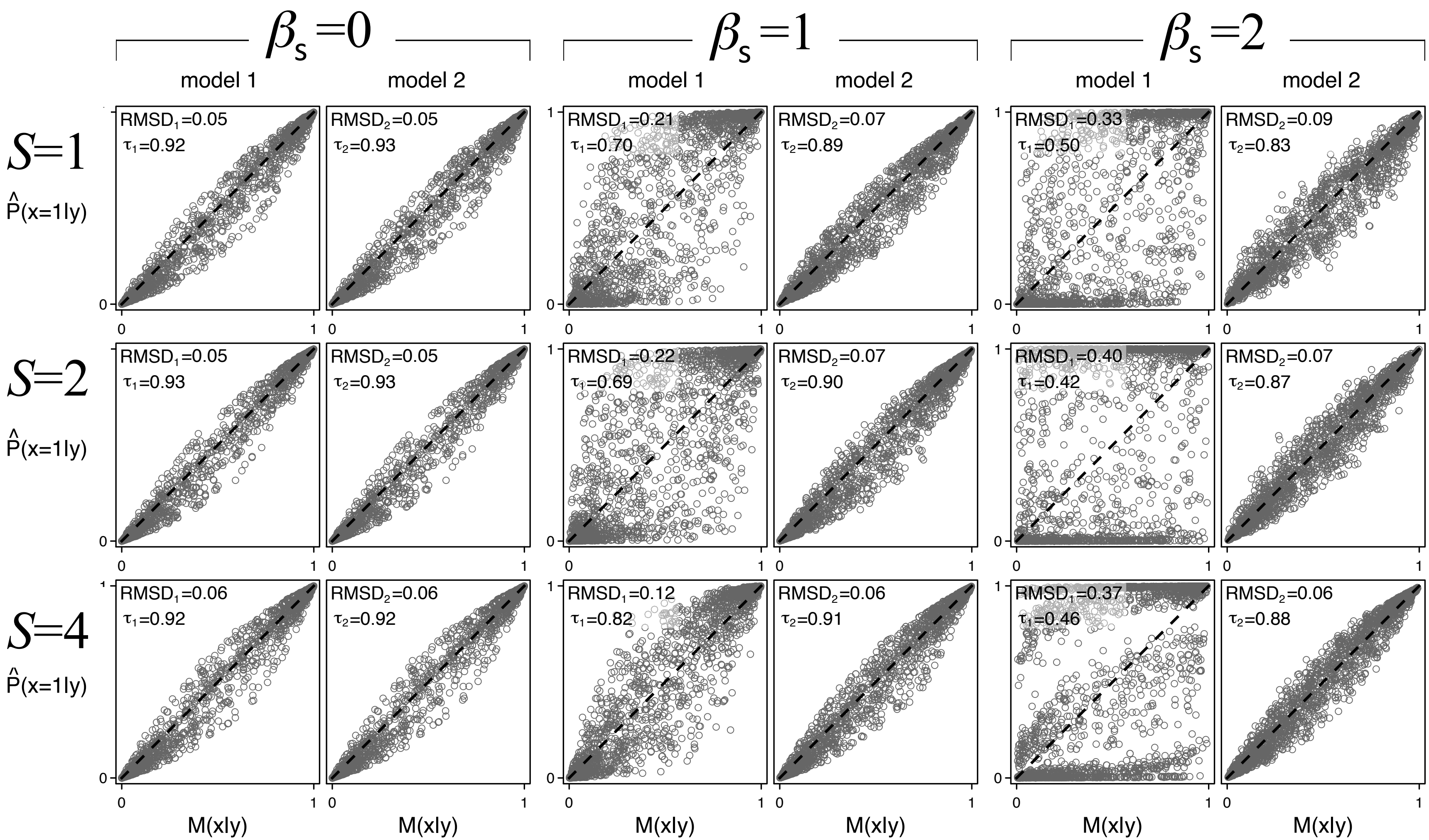
Ignoring nuisance dimensions *leads to mischaracterization of classification certainty*



(higher-order C-RUM, 120 items, S=4, β_s=2)

Examining Impact of Local Dependence (Simulation Results)

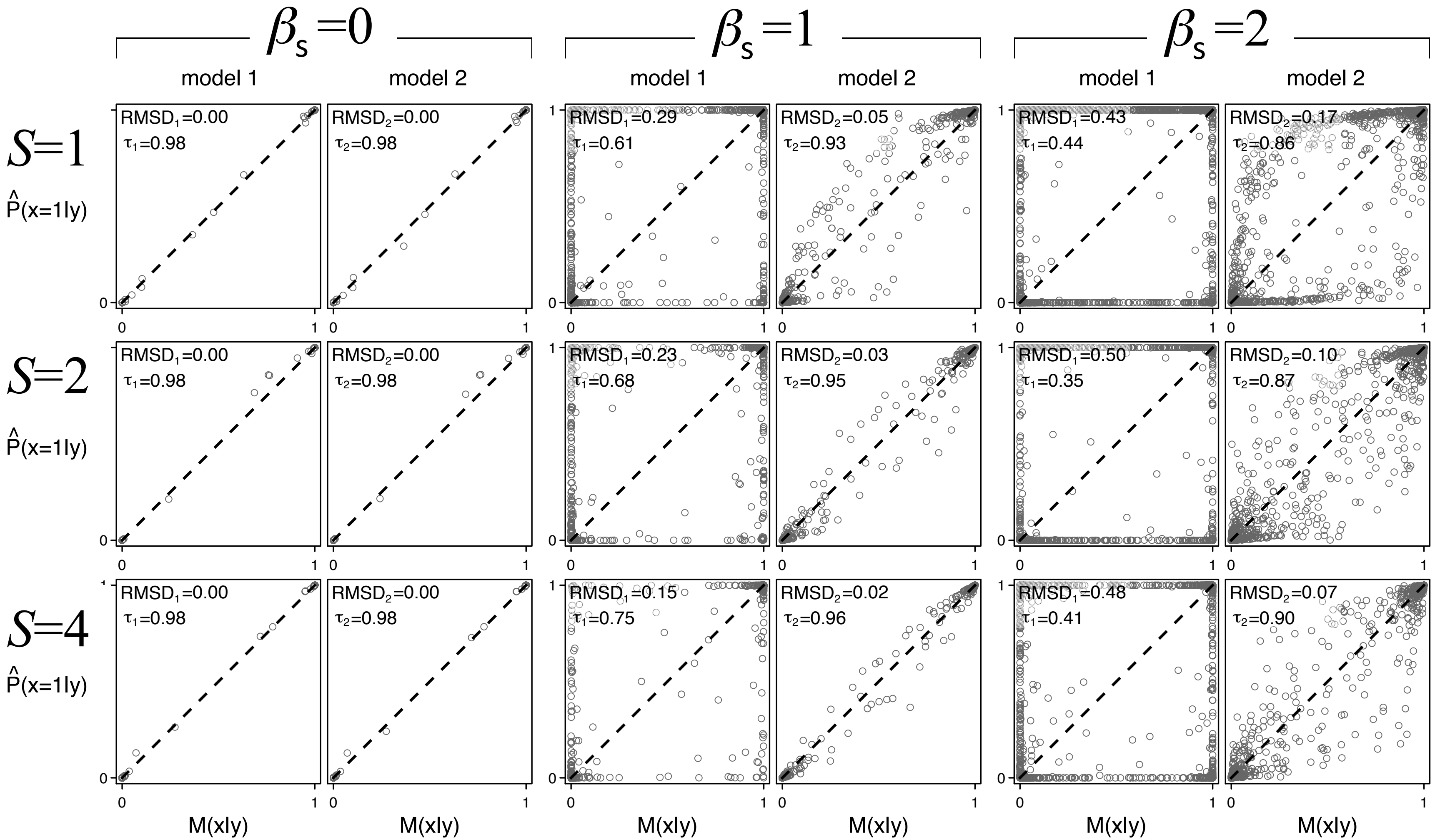
Ignoring nuisance dimensions *leads to mischaracterization of classification certainty*



higher-order C-RUM, 24 items

Examining Impact of Local Dependence (Simulation Results)

Ignoring nuisance dimensions *leads to mischaracterization of classification certainty*



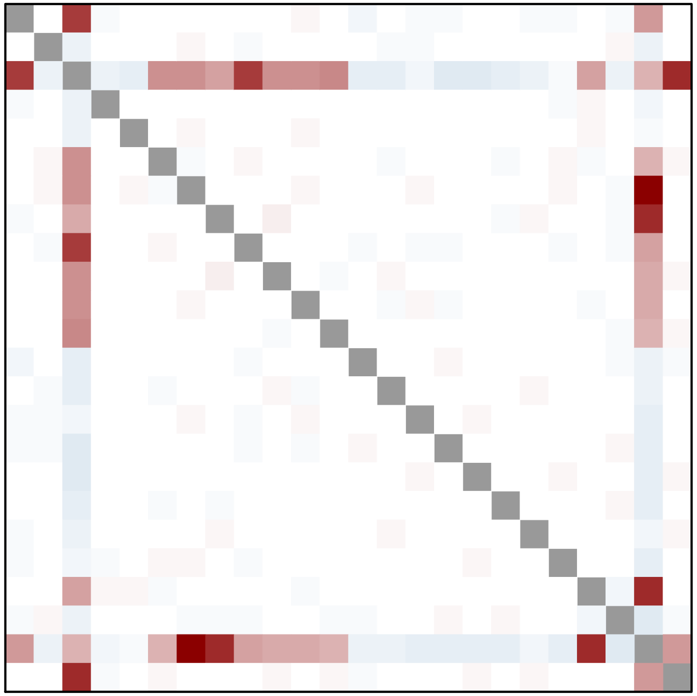
higher-order C-RUM, 120 items

"The trouble with people is not what they don't know but that they know so much that ain't so." -J. Billings, 1874

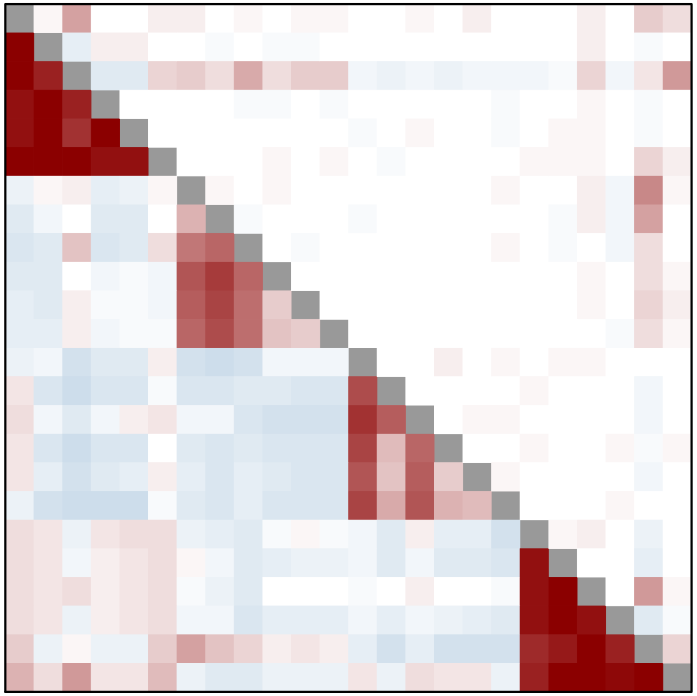
Examining Impact of Local Dependence (Simulation Results)

Ignoring nuisance dimensions *may obscure other model misspecifications*

no nuisance dimensions



$S=4,\ \beta_s=2$



- standard model below diagonal, hierarchical model above
- higher-order DINA with Q-matrix misspecification
  create extraneous paths from x_1 to y_3 and y_{23}

Outline

- Local dependence in diagnostic classifications models
  - an illustration
  - evaluation the assumption of local item independence
  - an alternative model
- Is this a problem that deserves attention?
  - we observe evidence of local dependence
  - ignoring dependence may…
    - impair classification decisions
    - lead to mischaracterizations of classification certainty
    - obscure other sorts of misspecification
- Final Comments

Final Comments

- Diagnostic models gaining in popularity, potentially useful for formative assessment
- Despite interest in these models, relatively little attention paid to model fit
- Model-based inferences may be flawed if the model fails to attend to local independence violations
- More realistic models ***may*** enable more valid inferences (including better classification) and, perhaps, better treatment outcomes (?)

Thank you for listening!

Mark Hansen markhansen@ucla.edu
Li Cai lcai@ucla.edu

Acknowledgements

This research is supported by grants from the Institute of Education Sciences (R305B080016), the National Institute on Drug Abuse (R01DA026943), and the National Science Foundation (1260746). Views expressed here are those of the presenters and do not reflect the views or policies of the funding agencies.