Detecting Compromised Items Using Information from Secure Items

Xi Wang,
Measured Progress, Dover, NH
Yang Liu,
University of Maryland, College Park

Research Background

- Item Preknowledge
- Continuous testing program: item re-use
- Quality control: monitor item statistics to flag problematic items

Secure Section + **Possibly Compromised**

zero or low-exposure items, e.g. field test items

repeatedly-used items

Two-stage Estimation

Secure Section + **Possibly Compromised**

Stage 1: Estimate population ability distribution, $\theta \sim N(\mu, \sigma^2)$ → Stage 2: Compare model-implied responses with observed responses

measured progress

Detection Methods

Predictive checking | **Residual method**

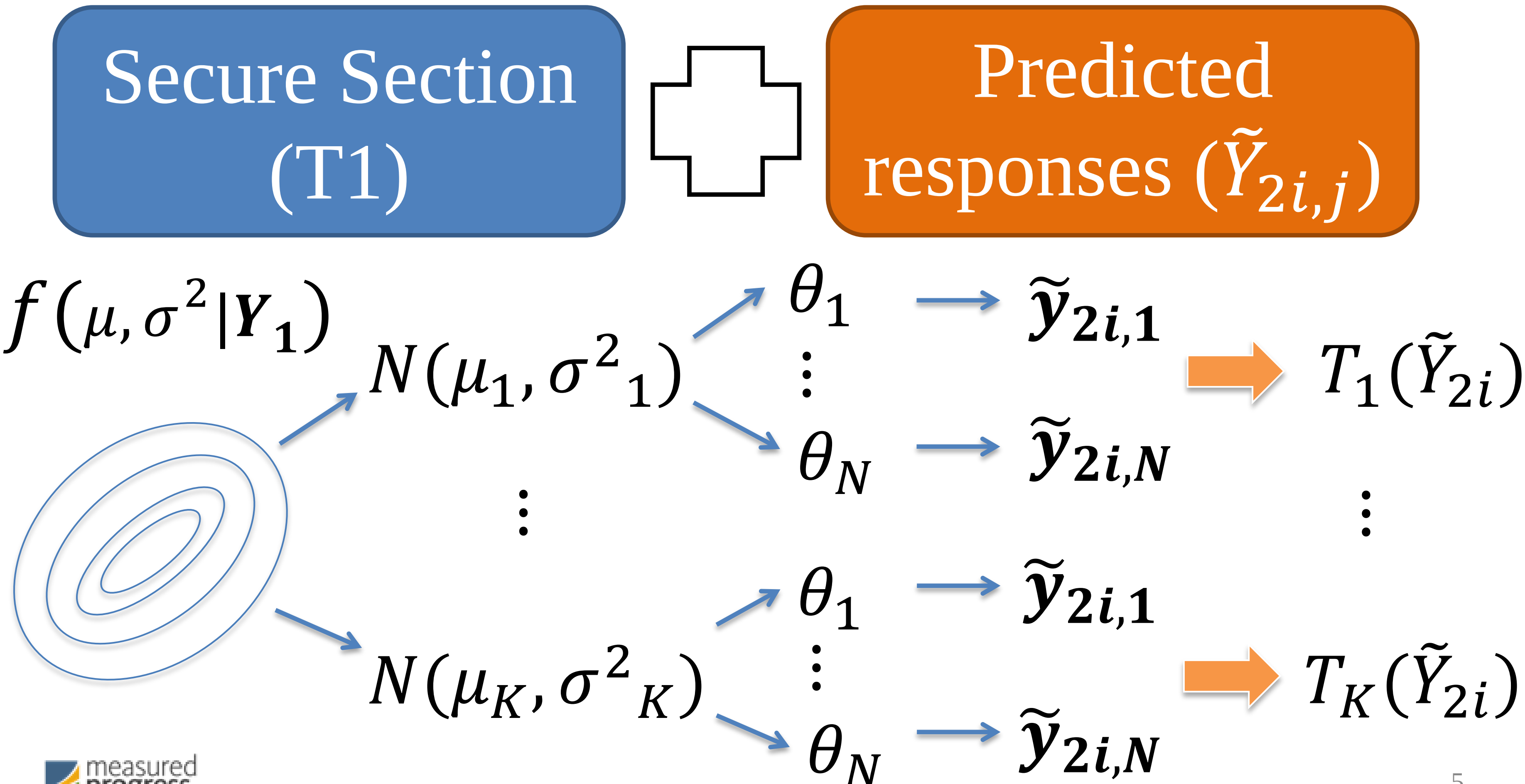
Sampling distribution based on simulation

- Assume known item parameters (e.g., obtained through prior administration)
- No need for re-calibration, so that can be used in small sample size

Sampling distribution based on normal approxi-mation

Detection Method

Predictive checking



measured progress.

Detection Method

Residual Method

$$Z_{2i} = \frac{(p_{2i} - \hat{\pi}_{2i})}{SE(p_{2i} - \hat{\pi}_{2i})}$$

$$\hat{\pi}_{2i} = \int P(Y_{2i} = 1|\theta) f(\theta|\hat{\mu}, \hat{\sigma}^2) d\theta$$

- p_{2i} is the observed proportion correct on item *i* in T2
- $\hat{\pi}_{2i}$ is the estimate of the model implied marginal probability of a correct response on item *i*.
- $Z_{2i} \dot{\sim} N(0,1)$. Large positive values lead to rejection of H_0

Detection Method

SE Estimation in Residual Method

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{e}}} = diag(\boldsymbol{\pi}_2) - \boldsymbol{\pi}_2\boldsymbol{\pi}_2' \\ +\boldsymbol{\Delta}_2\boldsymbol{I}_1^{-1}\boldsymbol{\Delta}_2' \\ - \boldsymbol{\Delta}_2\boldsymbol{I}_1^{-1}\boldsymbol{M} - (\boldsymbol{\Delta}_2\boldsymbol{I}_1^{-1}\boldsymbol{M})'$$

- $\boldsymbol{\pi}_2 = (P(Y_{2i} = 0|\mu, \sigma^2), P(Y_{2i} = 1|\mu, \sigma^2))'$
- $\boldsymbol{\Delta}_2$: Jacobian matrix of $\boldsymbol{\pi}_2$
- $\boldsymbol{I}_1^{-1}$: inverse Fisher information for $(\hat{\mu}, \hat{\sigma}^2)$
- $\boldsymbol{M}$: matrix containing multinomial expectations
- Population estimate vs Sample estimate: $\boldsymbol{M}$ vs $\widehat{\boldsymbol{M}}$

measured progress

Simulation Design

Simulation factors

- Examinee sample size: 50, 200, 1000
- Proportion of examinees with preknowedge: 0%, 10%, and 20%.
- Population ability: $\theta \sim N(0,1)$
- T1 length: 10; T2 item parameter:

	***a*=0.75**			***a*=1**			***a*=1.25**		
b	-0.85	0	0.85	-0.85	0	0.85	-0.85	0	0.85
c	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

measured progress

Simulation Design

Type-I error simulation

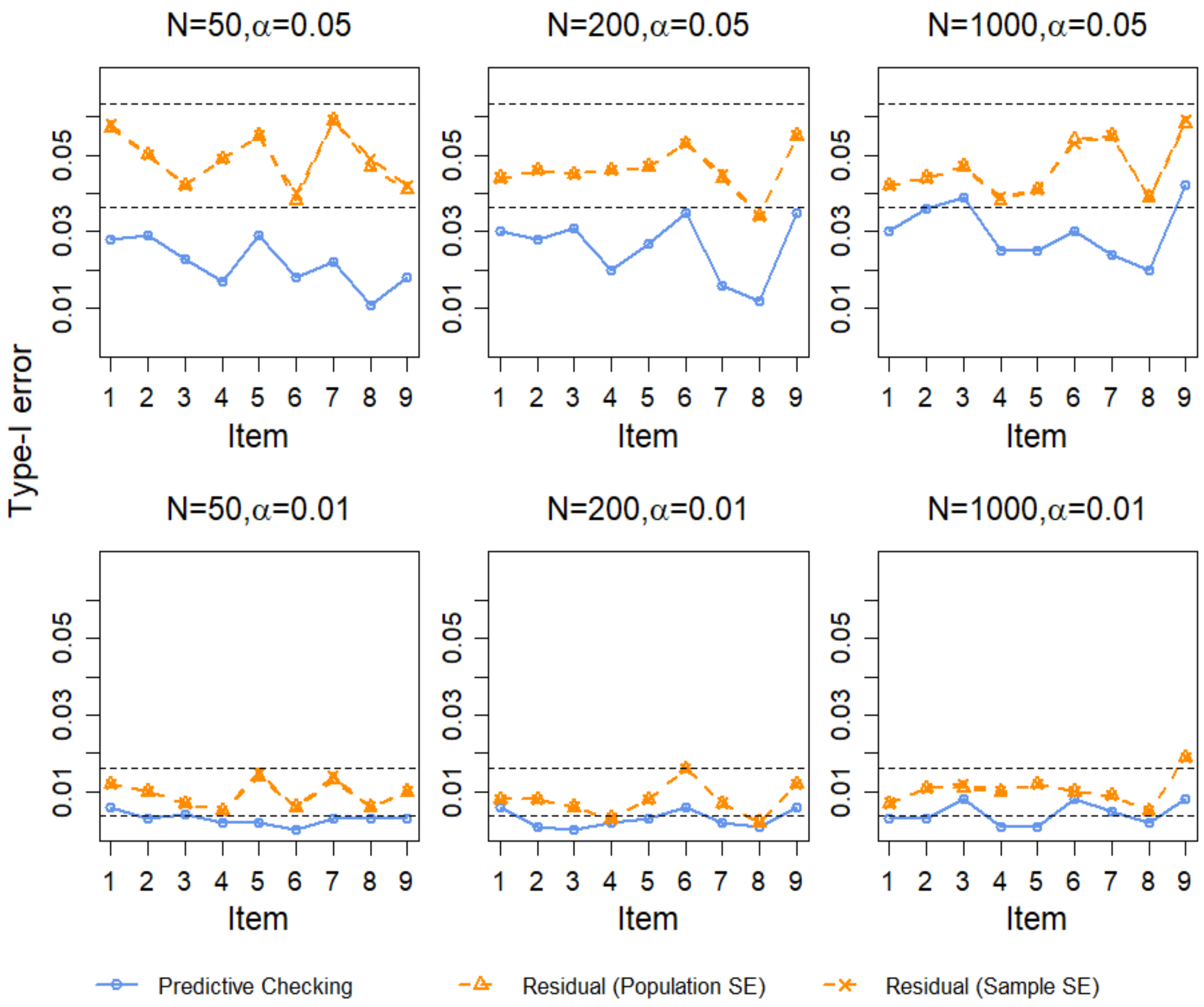
- Response generating model: 3PLM

Power simulation

- Preknowledge introduced to a random examinee subgroup
- Compromised responses generated from Bernoulli(p_c), $p_c = \max\left(P(Y = 1|\theta), 0.9\right)$
- 1000 replications
- α=0.01, 0.05

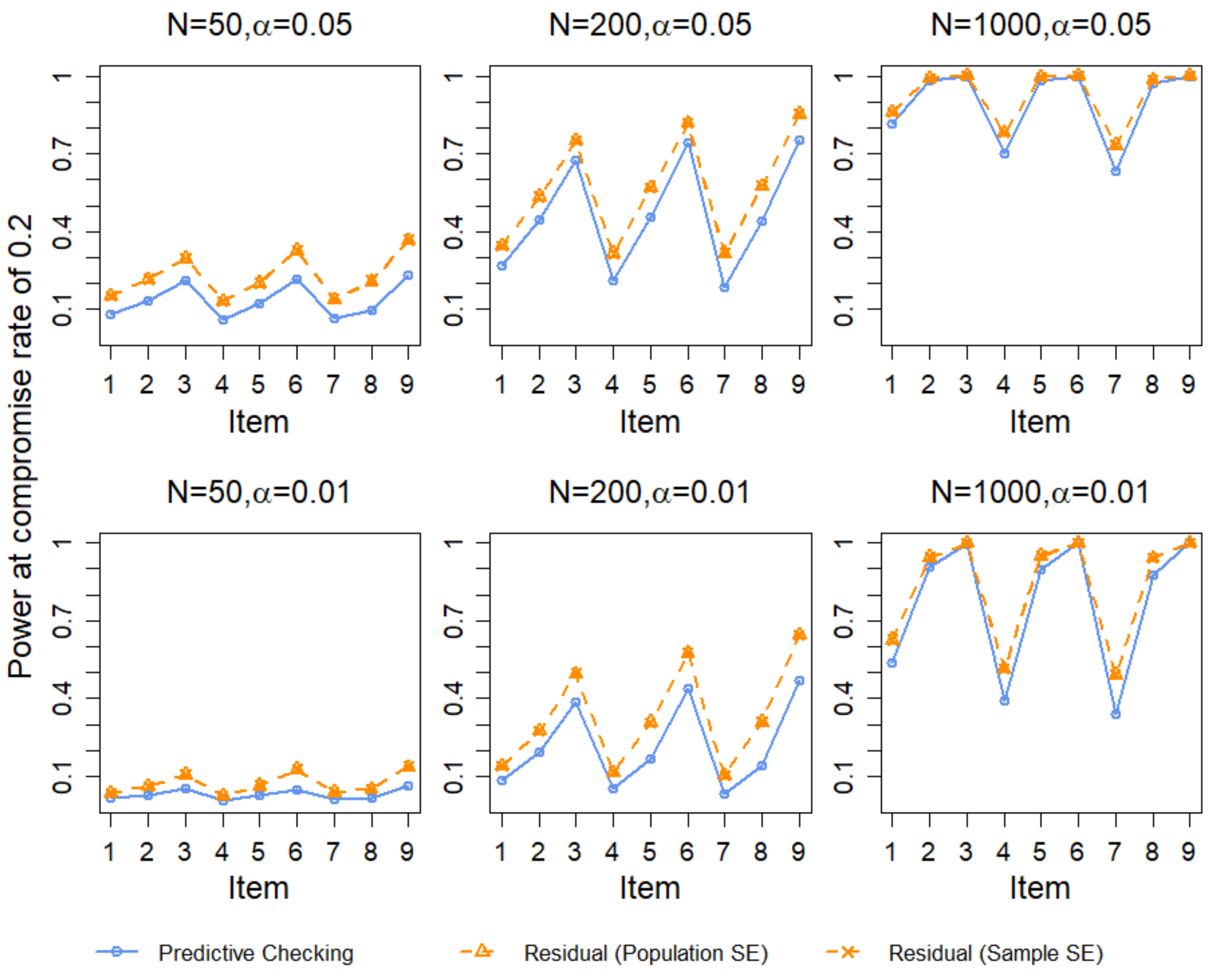
Simulation Results

Type-I error



Simulation Results

Power with 20% preknowledge



Discussions

- **Results summary**
  - Type-I error well controlled.
  - Effective in detecting moderate to large shift in item difficulty, even at small sample sizes.
- **Future directions**
  - Relax assumption of $\theta \sim N(\mu, \sigma^2)$: Consider more flexible methods to estimate non-normal distribution for θ
  - Take item parameter error into account

measured progress

Thank you.

wang.xi@measuredprogress.org