Calibrated Model Criticism Using Split Predictive Checks

Jonathan Huggins
Boston University

with Jiawei Li

Model checking: can we have it all?

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient
  3. well-calibrated

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient
  3. well-calibrated
  4. well-powered

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient
  3. well-calibrated
  4. well-powered
- **Problem:** current methods fail to achieve all of these

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient
  3. well-calibrated
  4. well-powered
- **Problem:** current methods fail to achieve all of these
- This talk: ***split predictive checks*** *(SPCs) can be the answer!*

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient
  3. well-calibrated
  4. well-powered
- **Problem:** current methods fail to achieve all of these
- This talk: ***split predictive checks*** *(SPCs) can be the answer!*
  - develop ***single SPCs*** and ***divided SPCs***

Model checking: can we have it all?

- **Model criticism** a crucial part of Bayesian model building:
  - Evaluates current model fit, insights for model elaboration, …
- Want a model criticism method to be…
  1. general-purpose
  2. computationally efficient
  3. well-calibrated
  4. well-powered
- **Problem:** current methods fail to achieve all of these
- This talk: ***<u>split predictive checks</u>*** *(SPCs) can be the answer!*
  - develop ***single SPCs*** and ***divided SPCs***
  - provide asymptotic theory showing good **calibration** & **power**

Bayesian model checking

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$
- Model $\{P_\theta : \theta \in \Theta\}$

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$
- Model $\{P_\theta : \theta \in \Theta\}$
- KL-optimal parameter $\theta_\star := \sup_{\theta\in\Theta} \mathbb{E}\{\log p_\theta(x_1)\}$

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$
- Model $\{P_\theta : \theta \in \Theta\}$
- KL-optimal parameter $\theta_\star := \sup_{\theta \in \Theta} \mathbb{E}\{\log p_\theta(x_1)\}$
- **Goal:** measure the mismatch between $P_\circ$ and $P_{\theta_\star}$

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$
- Model $\{P_\theta : \theta \in \Theta\}$
- KL-optimal parameter $\theta_\star := \sup_{\theta \in \Theta} \mathbb{E}\{\log p_\theta(x_1)\}$
- **Goal:** measure the mismatch between $P_\circ$ and $P_{\theta_\star}$
  - $H_0 : P_{\theta_\star} = P_\circ$ (well-specified)

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$
- Model $\{P_\theta : \theta \in \Theta\}$
- KL-optimal parameter $\theta_\star := \sup_{\theta \in \Theta} \mathbb{E}\{\log p_\theta(x_1)\}$
- **Goal:** measure the mismatch between $P_\circ$ and $P_{\theta_\star}$
    - $H_0 : P_{\theta_\star} = P_\circ$ (well-specified)
    - $H_1 : P_\theta \neq P_\circ \; \forall \theta \in \Theta$ (misspecified)

Bayesian model checking

- Data $\mathbf{x} = (x_1, \ldots, x_N)$ where $x_n \overset{\text{i.i.d.}}{\sim} P_\circ$
- Model $\{P_\theta : \theta \in \Theta\}$
- KL-optimal parameter $\theta_\star := \sup_{\theta \in \Theta} \mathbb{E}\{\log p_\theta(x_1)\}$
- **Goal:** measure the mismatch between $P_\circ$ and $P_{\theta_\star}$
  - $H_0 : P_{\theta_\star} = P_\circ$ (well-specified)
  - $H_1 : P_\theta \neq P_\circ \; \forall \theta \in \Theta$ (misspecified)
- p-values provide measure of fit only if they are well-calibrated!

Predictive checks

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x}_{\mathbf{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x}_{\mathbf{pred}})\}$$

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x}_{\mathbf{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x}_{\mathbf{pred}})\}$$

user-specified statistic

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x}_{\mathbf{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x}_{\mathbf{pred}})\}$$

user-specified statistic

- Standard choices of m

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

user-specified statistic

- Standard choices of m
  - ▸ Prior PCs: $m_{prior}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi_0(d\theta)$ [Box (1980)]

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

user-specified statistic

- Standard choices of m
  - Prior PCs: $m_{prior}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi_0(d\theta)$ [Box (1980)]
  - Posterior PCs: $m_{post}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi(d\theta \mid \mathbf{x})$ [Rubin (1984)]

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x}_{\mathbf{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x}_{\mathbf{pred}})\}$$

user-specified statistic

- Standard choices of m
  - Prior PCs: $m_{prior}(\mathbf{x}_{\mathbf{pred}}) := \int p_\theta(\mathbf{x}_{\mathbf{pred}})\Pi_0(d\theta)$ [Box (1980)]
  - Posterior PCs: $m_{post}(\mathbf{x}_{\mathbf{pred}}) := \int p_\theta(\mathbf{x}_{\mathbf{pred}})\Pi(d\theta \,|\, \mathbf{x})$ [Rubin (1984)]

- Posterior predictive checks (PPCs) p-value is defined as

$$p_{PPC}(\mathbf{x}) = \mathbb{P}_{\mathbf{x}_{\mathbf{pred}} \sim m_{post}(\cdot|\mathbf{x})}\{T(\mathbf{x}) > T(\mathbf{x}_{\mathbf{pred}})\}$$

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

user-specified statistic

- Standard choices of m
  - Prior PCs: $m_{prior}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi_0(d\theta)$ [Box (1980)]
  - Posterior PCs: $m_{post}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi(d\theta \,|\, \mathbf{x})$ [Rubin (1984)]

- Posterior predictive checks (PPCs) p-value is defined as

$$p_{PPC}(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m_{post}(\cdot|\mathbf{x})}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

✓ Flexible, general-purpose, and computationally efficient

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

user-specified statistic

- Standard choices of m
  - Prior PCs: $m_{prior}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi_0(d\theta)$ [Box (1980)]
  - Posterior PCs: $m_{post}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi(d\theta \,|\, \mathbf{x})$ [Rubin (1984)]

- Posterior predictive checks (PPCs) p-value is defined as

$$p_{PPC}(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m_{post}(\cdot|\mathbf{x})}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

✓ Flexible, general-purpose, and computationally efficient

✗ not well-calibrated, conservative p-values [Robins et al (2000), Bayarri & Burger (2000), Moran et al (2022)]

Predictive checks

- A predictive p-value is given by

$$p(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

user-specified statistic

- Standard choices of m
  - Prior PCs: $m_{prior}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi_0(d\theta)$ [Box (1980)]
  - Posterior PCs: $m_{post}(\mathbf{x_{pred}}) := \int p_\theta(\mathbf{x_{pred}})\Pi(d\theta \mid \mathbf{x})$ [Rubin (1984)]

- Posterior predictive checks (PPCs) p-value is defined as

$$p_{PPC}(\mathbf{x}) = \mathbb{P}_{\mathbf{x_{pred}} \sim m_{post}(\cdot \mid \mathbf{x})}\{T(\mathbf{x}) > T(\mathbf{x_{pred}})\}$$

double use of data

✓ Flexible, general-purpose, and computationally efficient

✗ not well-calibrated, conservative p-values [Robins et al (2000), Bayarri & Burger (2000), Moran et al (2022)]

PPCs are not well-calibrated

PPCs are not well-calibrated

- Model: $P_\theta = \text{Poiss}(\theta)$

PPCs are not well-calibrated

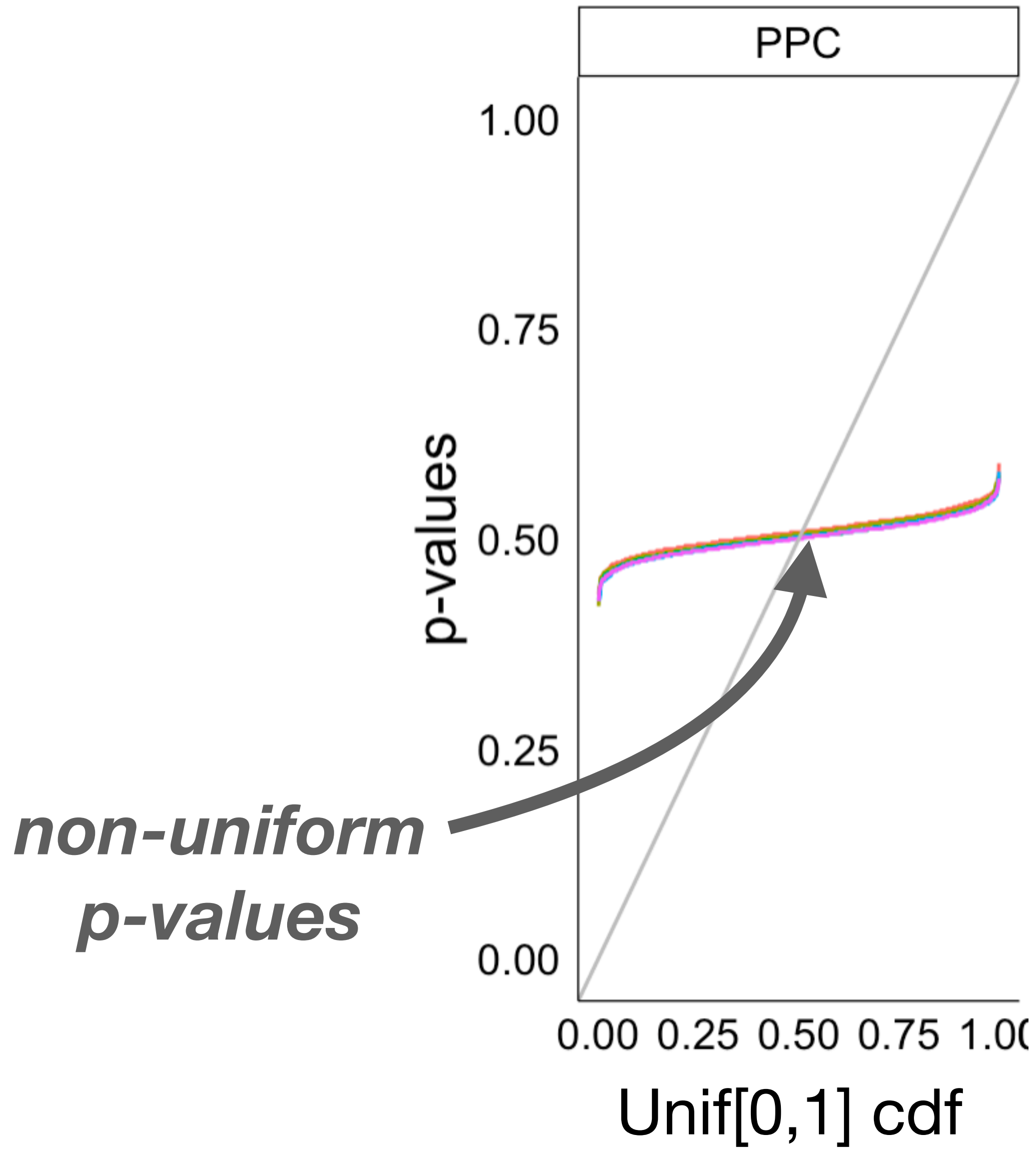
- Model: $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{Poiss}(2)$

PPCs are not well-calibrated

- Model: $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{Poiss}(2)$
- So, model is well-specified [H_0 holds]

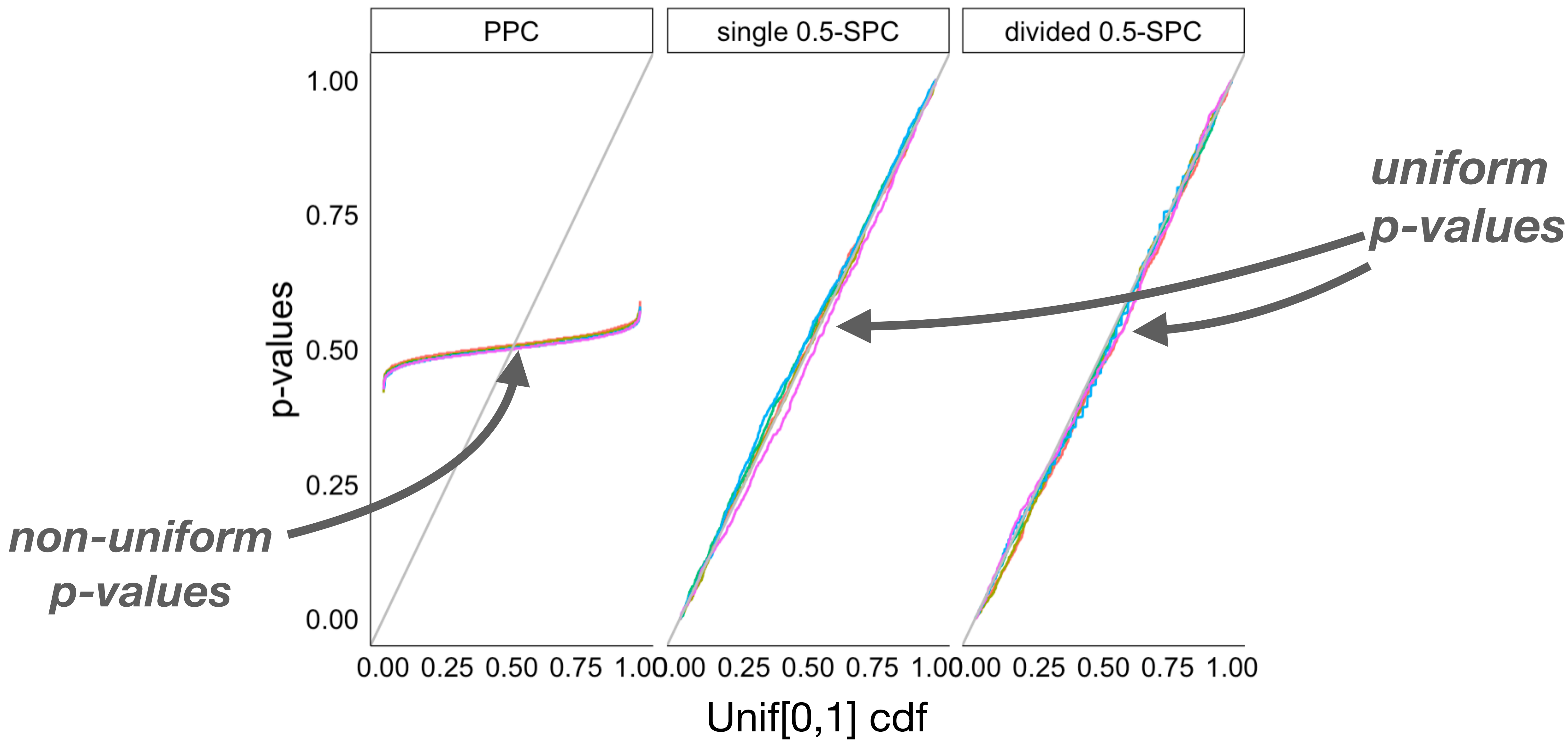
PPCs are not well-calibrated

- Model: $P_\theta =$ Poiss(θ)
- Generate data from $P_\circ =$ Poiss(2)
- So, model is well-specified [H_0 holds]



PPCs are not well-calibrated

- Model: $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{Poiss}(2)$
- So, model is well-specified [H_0 holds]



PPCs produce conservative p-values

PPCs produce conservative p-values

- Geometric model for airline delay data for 2013 NYC departures

$$y \sim \text{Geom}(\theta), \qquad \theta \sim \text{Beta}(0.1, 0.2)$$

PPCs produce conservative p-values

- Geometric model for airline delay data for 2013 NYC departures

$$y \sim \mathrm{Geom}(\theta), \qquad \theta \sim \mathrm{Beta}(0.1, 0.2)$$

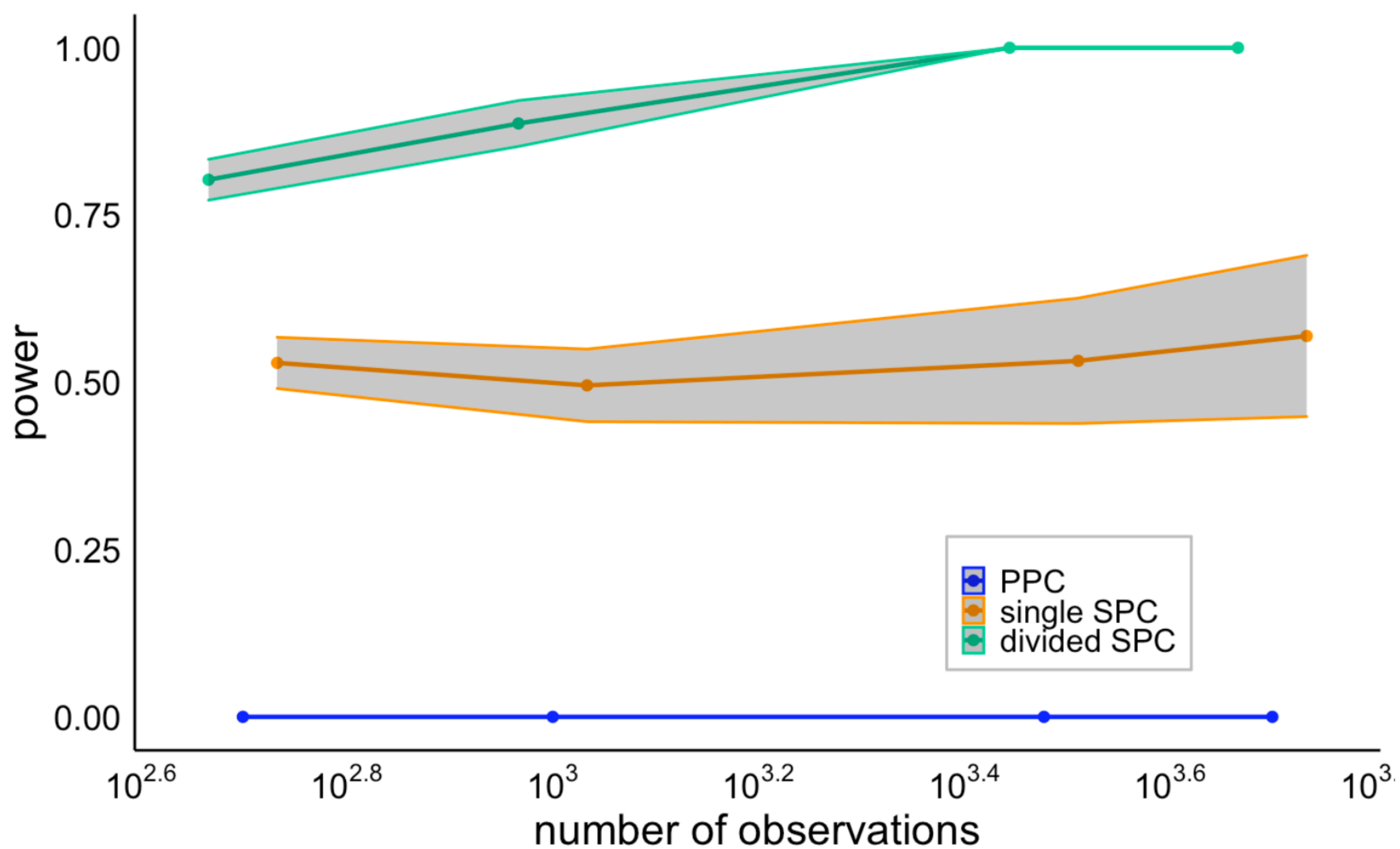
- Model is misspecified [H_1 holds]

PPCs produce conservative p-values

- Geometric model for airline delay data for 2013 NYC departures

$$y \sim \text{Geom}(\theta), \qquad \theta \sim \text{Beta}(0.1, 0.2)$$

- Model is misspecified [H_1 holds]



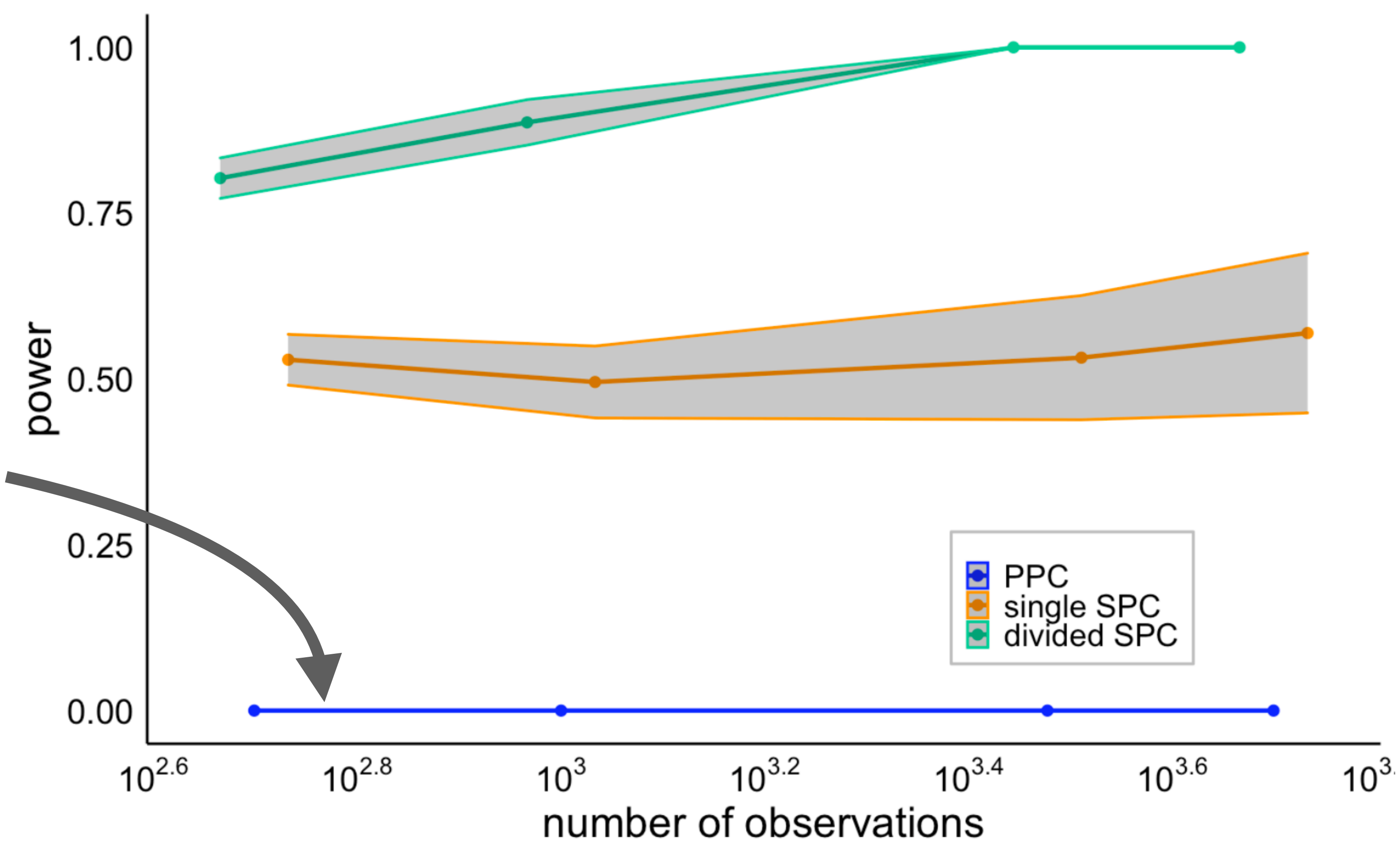
PPCs produce conservative p-values

- Geometric model for airline delay data for 2013 NYC departures

$$y \sim \mathrm{Geom}(\theta), \qquad \theta \sim \mathrm{Beta}(0.1, 0.2)$$

- Model is misspecified [H_1 holds]

PPCs extremely conservative



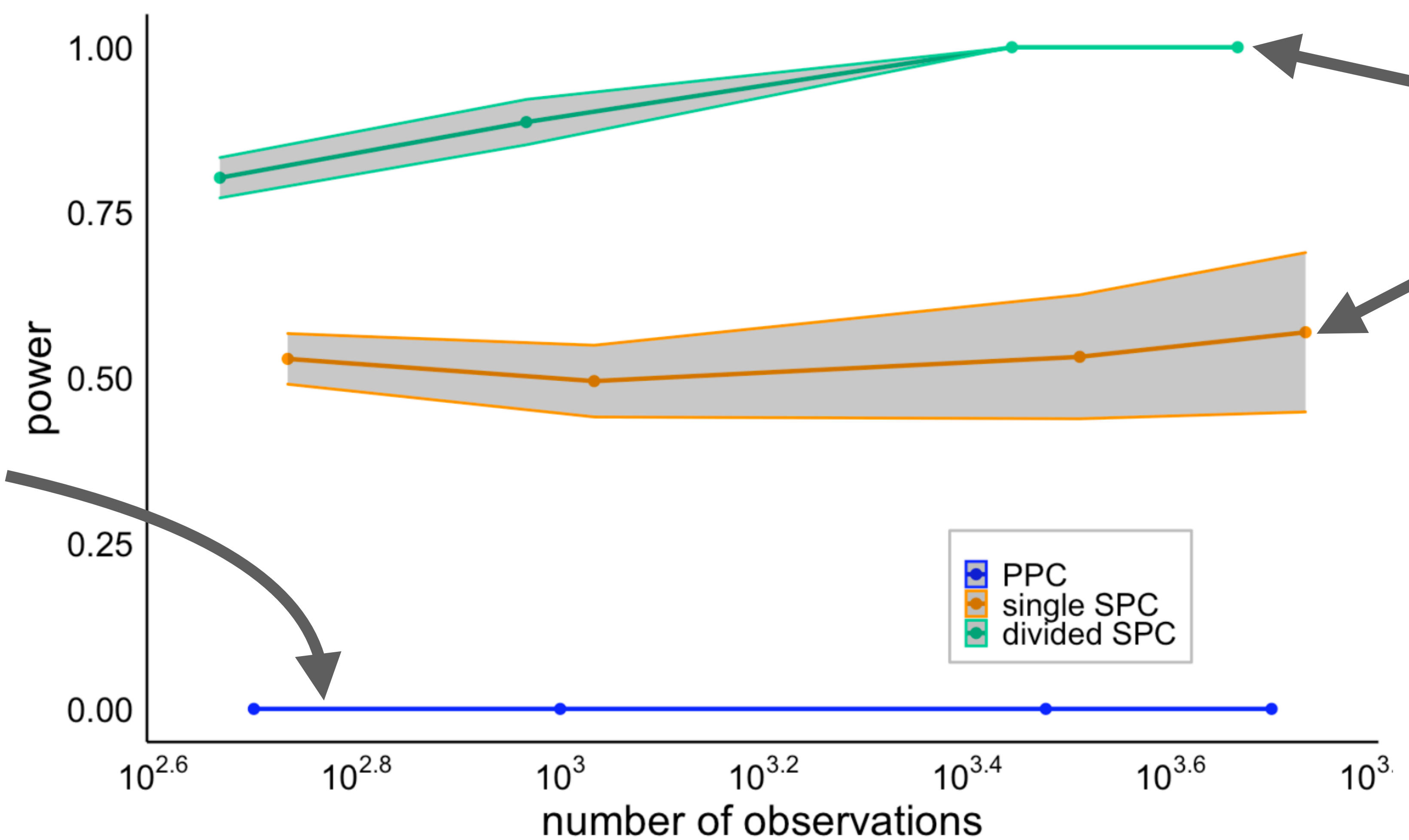
PPCs produce conservative p-values

- Geometric model for airline delay data for 2013 NYC departures

$$y \sim \mathrm{Geom}(\theta), \qquad \theta \sim \mathrm{Beta}(0.1,0.2)$$

- Model is misspecified [H_1 holds]

PPCs extremely conservative



both SPCs have good power

Single split predictive checks

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data

$\mathbf{x}$

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data

split ***disjointly***

$N_{obs} = \lfloor qN \rfloor$

$\mathbf{x}$

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data

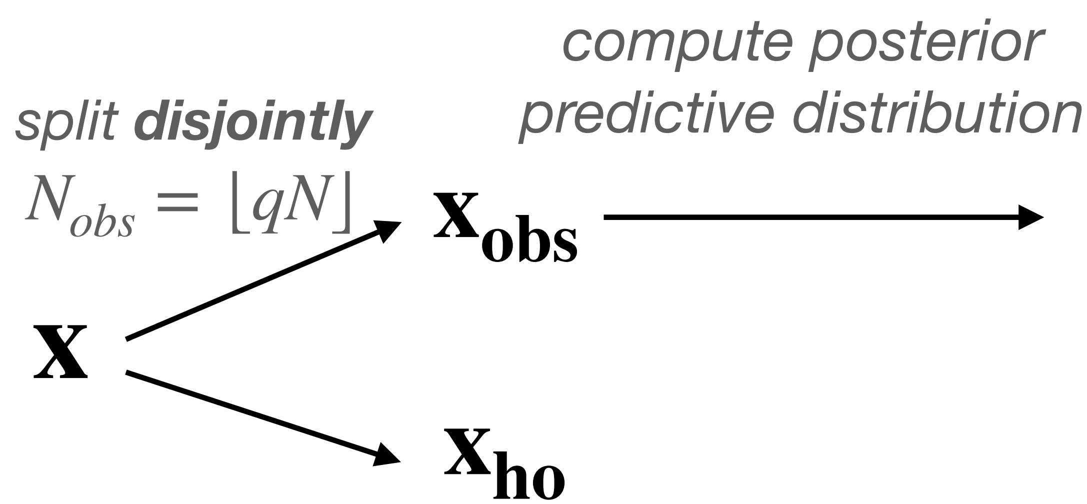
split ***disjointly***

$N_{obs} = \lfloor qN \rfloor$

$$\begin{tikzcd}
 & \mathbf{x_{obs}} \\
\mathbf{x} \arrow[ur] \arrow[dr] & \\
 & \mathbf{x_{ho}}
\end{tikzcd}$$

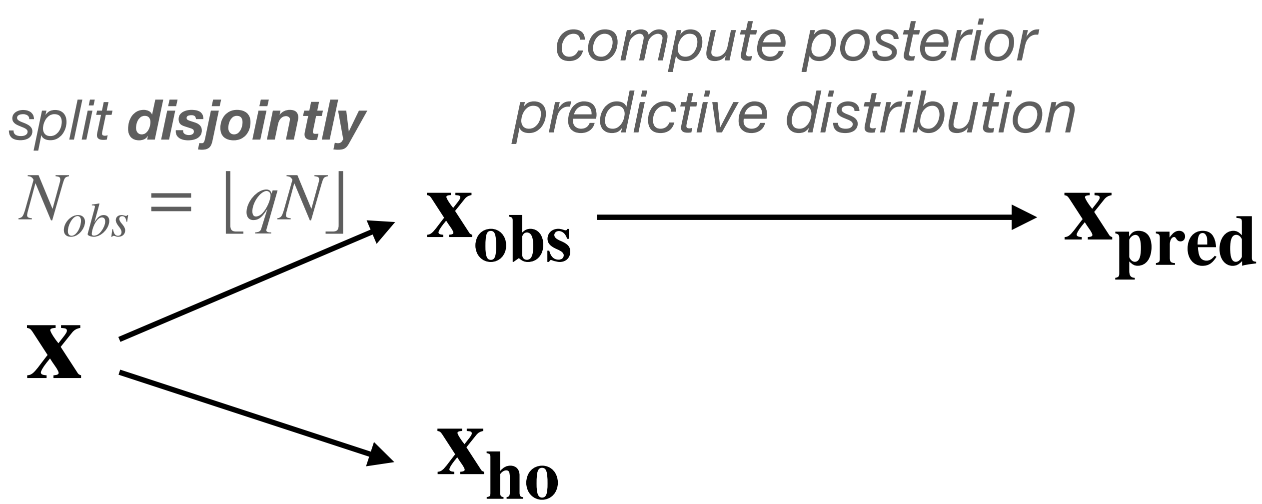
Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data



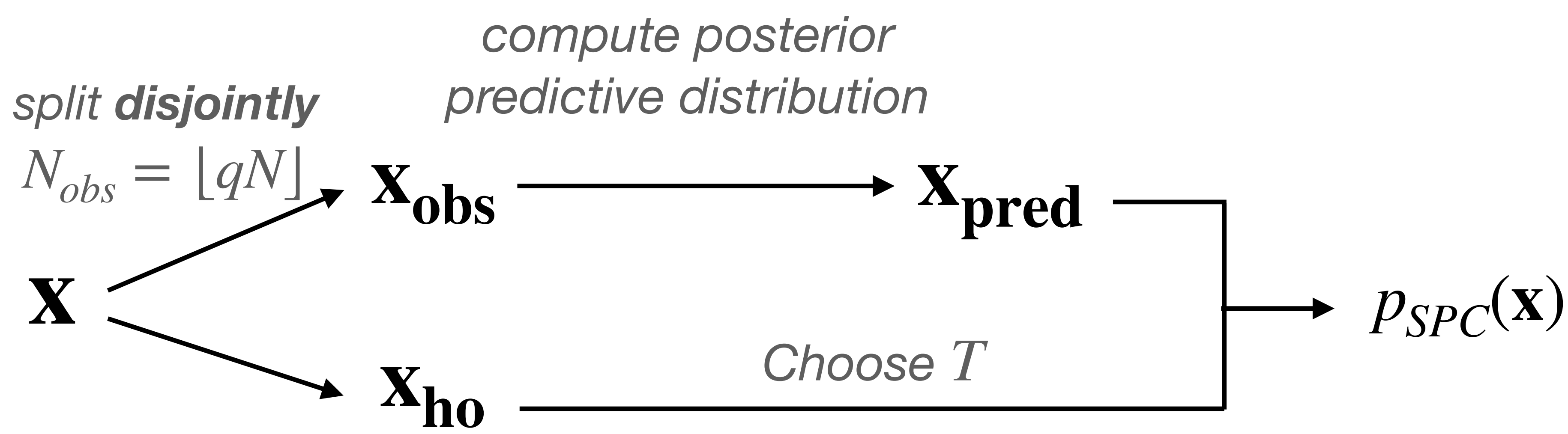
Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data



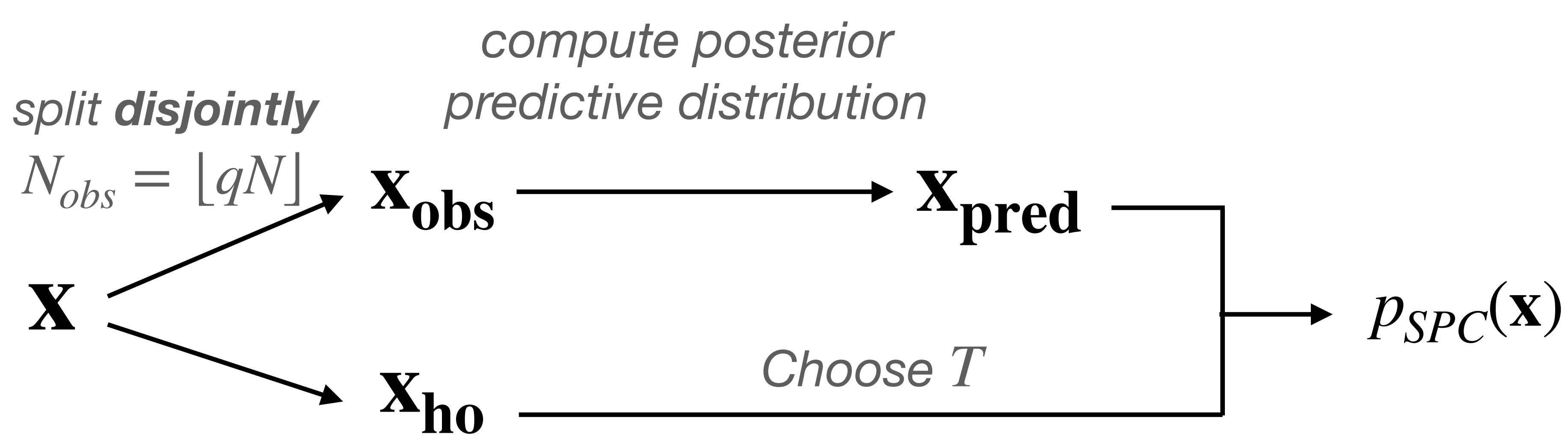
Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data



Single split predictive checks

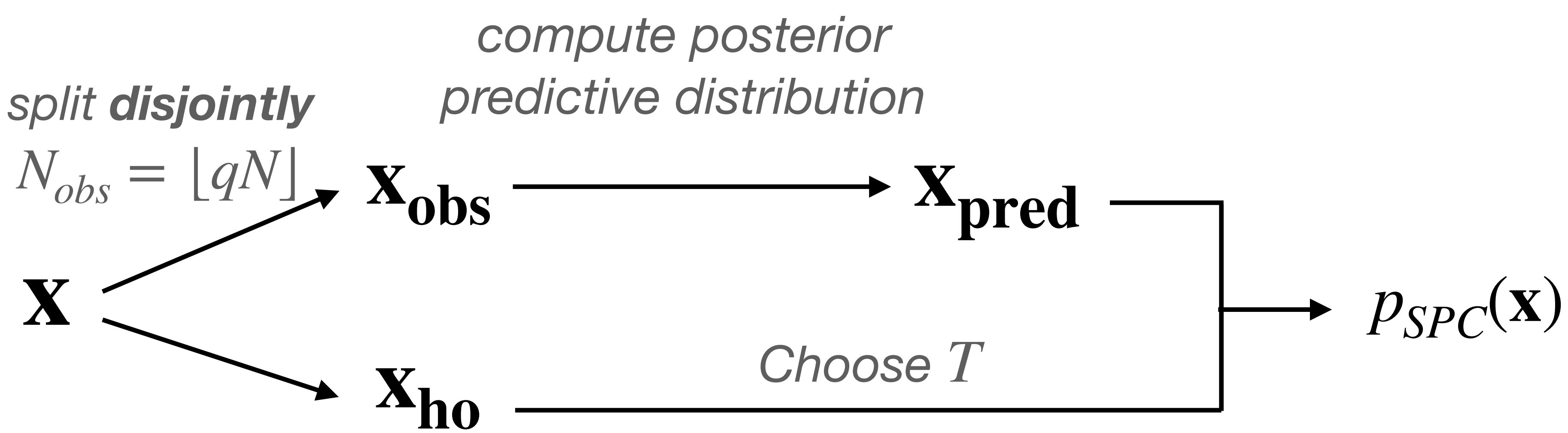
- Single SPCs avoid double use of data by splitting data into “train” and “test” data



- Single q-SPC p-value [see also Moran et al (2022)]:

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data

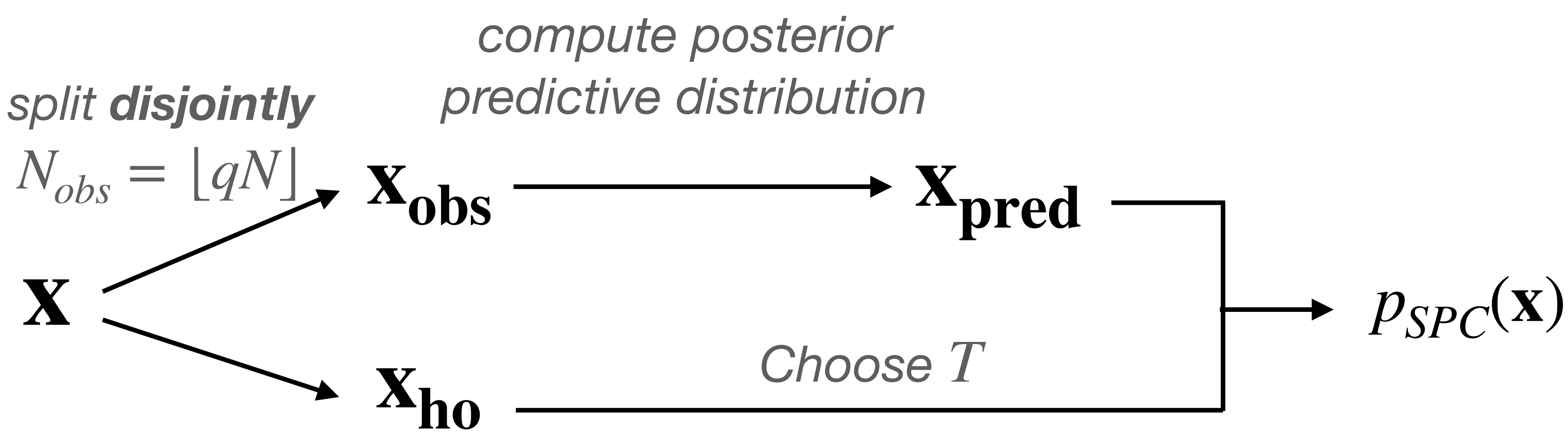


- Single q-SPC p-value [see also Moran et al (2022)]:

$$p_{SPC}(\mathbf{x}) := \mathbb{P}_{\mathbf{x_{pred}} \sim m_{post}^{SPC}(\cdot|\mathbf{x_{obs}})}\{T(\mathbf{x_{ho}}) > T(\mathbf{x_{pred}})\}$$

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into “train” and “test” data



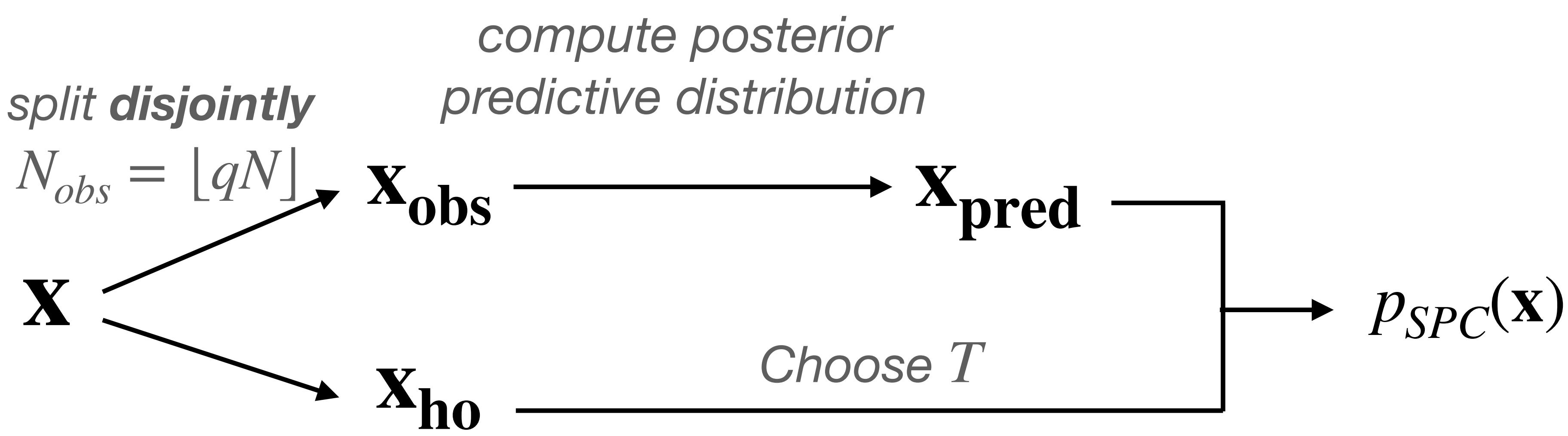
- Single q-SPC p-value [see also Moran et al (2022)]:

$$p_{SPC}(\mathbf{x}) := \mathbb{P}_{\mathbf{x_{pred}} \sim m_{post}^{SPC}(\cdot|\mathbf{x_{obs}})}\{T(\mathbf{x_{ho}}) > T(\mathbf{x_{pred}})\}$$

✓ flexible, computationally efficient, and general-purpose

Single split predictive checks

- Single SPCs avoid double use of data by splitting data into "train" and "test" data



- Single q-SPC p-value [see also Moran et al (2022)]:

$$p_{SPC}(\mathbf{x}) := \mathbb{P}_{\mathbf{x_{pred}} \sim m_{post}^{SPC}(\cdot|\mathbf{x_{obs}})}\{T(\mathbf{x_{ho}}) > T(\mathbf{x_{pred}})\}$$

✓ flexible, computationally efficient, and general-purpose

Question 1: How is calibration and power?

Asymptotic power and calibration of single SPCs

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\xrightarrow{P} 1$

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\xrightarrow{P} 1$

Theorem [LH22]: if $\nu_\circ = \nu(\theta_\star)$, then $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\xrightarrow{P} 1$

Theorem [LH22]: if $\nu_\circ = \nu(\theta_\star)$, then $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

ρ depends on q, Fisher information matrices and asymptotic variances of T

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\xrightarrow{P} 1$

Theorem [LH22]: if $\nu_\circ = \nu(\theta_\star)$, then $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\xrightarrow{P} 1$

Theorem [LH22]: if $\nu_\circ = \nu(\theta_\star)$, then $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$
- if model misspecified, then in general $\rho \neq 1$:

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\xrightarrow{P} 1$

Theorem [LH22]: if $\nu_\circ = \nu(\theta_\star)$, then $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$
- if model misspecified, then in general $\rho \neq 1$:
  - when uncertainty is under-estimated, $\rho \uparrow \infty$ and power $\uparrow$ 1

Asymptotic power and calibration of single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$, then power $\overset{P}{\to} 1$

Theorem [LH22]: if $\nu_\circ = \nu(\theta_\star)$, then $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \overset{P}{\to} \Phi\left(\frac{z_\alpha}{\rho}\right)$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \overset{P}{\to} \alpha$
- if model misspecified, then in general $\rho \neq 1$:
  - when uncertainty is under-estimated, $\rho \uparrow \infty$ and power $\uparrow$ 1
  - when uncertainty is over-estimated, $\rho \downarrow 0$ and power $\downarrow$ 0

Single SPCs have asymptotic power 1 when misspecification is significant

Single SPCs have asymptotic power 1 when misspecification is significant

- Model $P_\theta = \text{Poiss}(\theta)$

Single SPCs have asymptotic power 1 when misspecification is significant

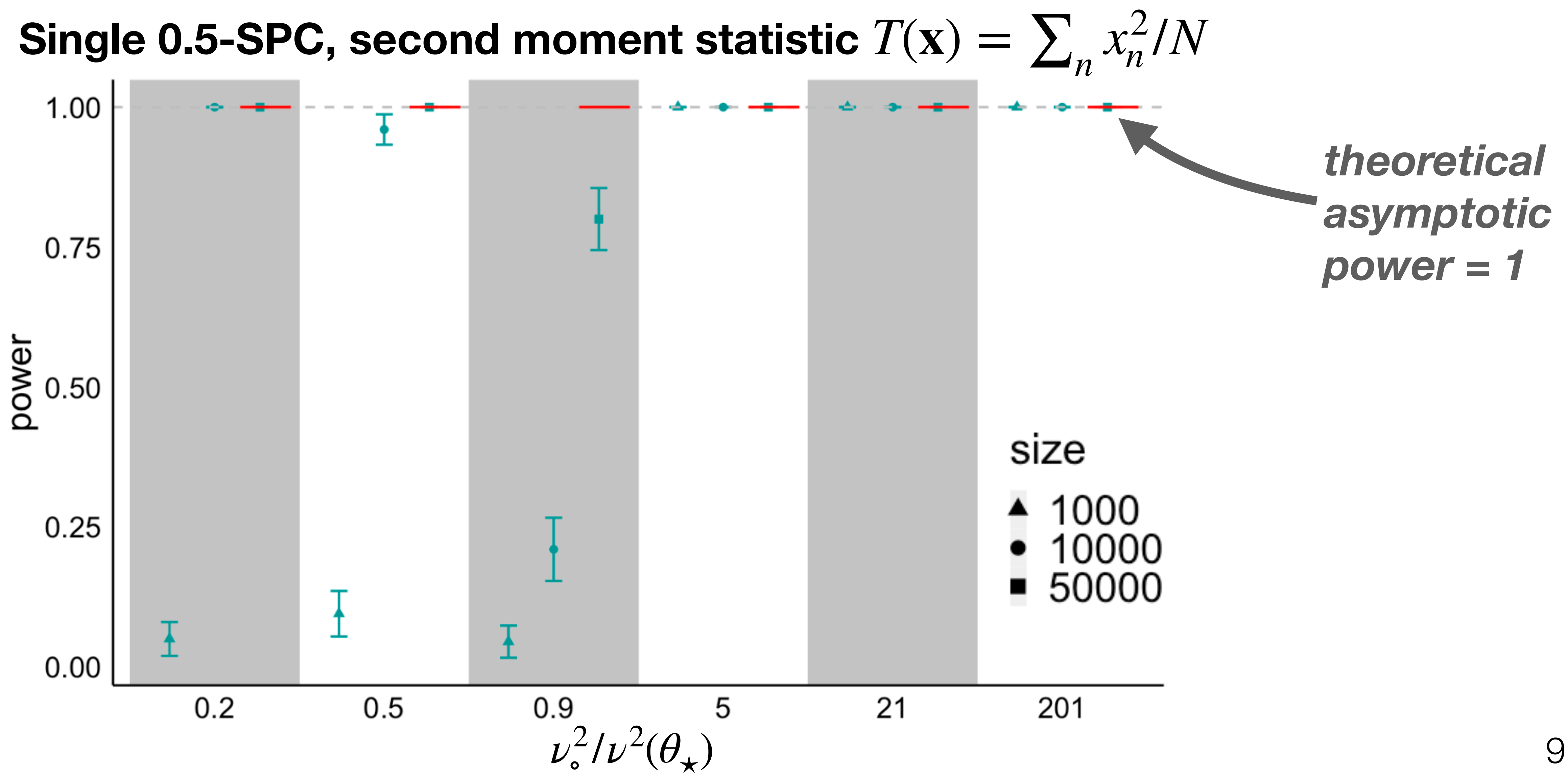
- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$

Single SPCs have asymptotic power 1 when misspecification is significant

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- Recall: when $\nu_\circ / \nu(\theta_\star) \neq 1$, asymptotic power = 1

Single SPCs have asymptotic power 1 when misspecification is significant

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- Recall: when $\nu_\circ / \nu(\theta_\star) \neq 1$, asymptotic power = 1



Possible poor power of single SPCs when misspecification is subtle

Possible poor power of single SPCs when misspecification is subtle

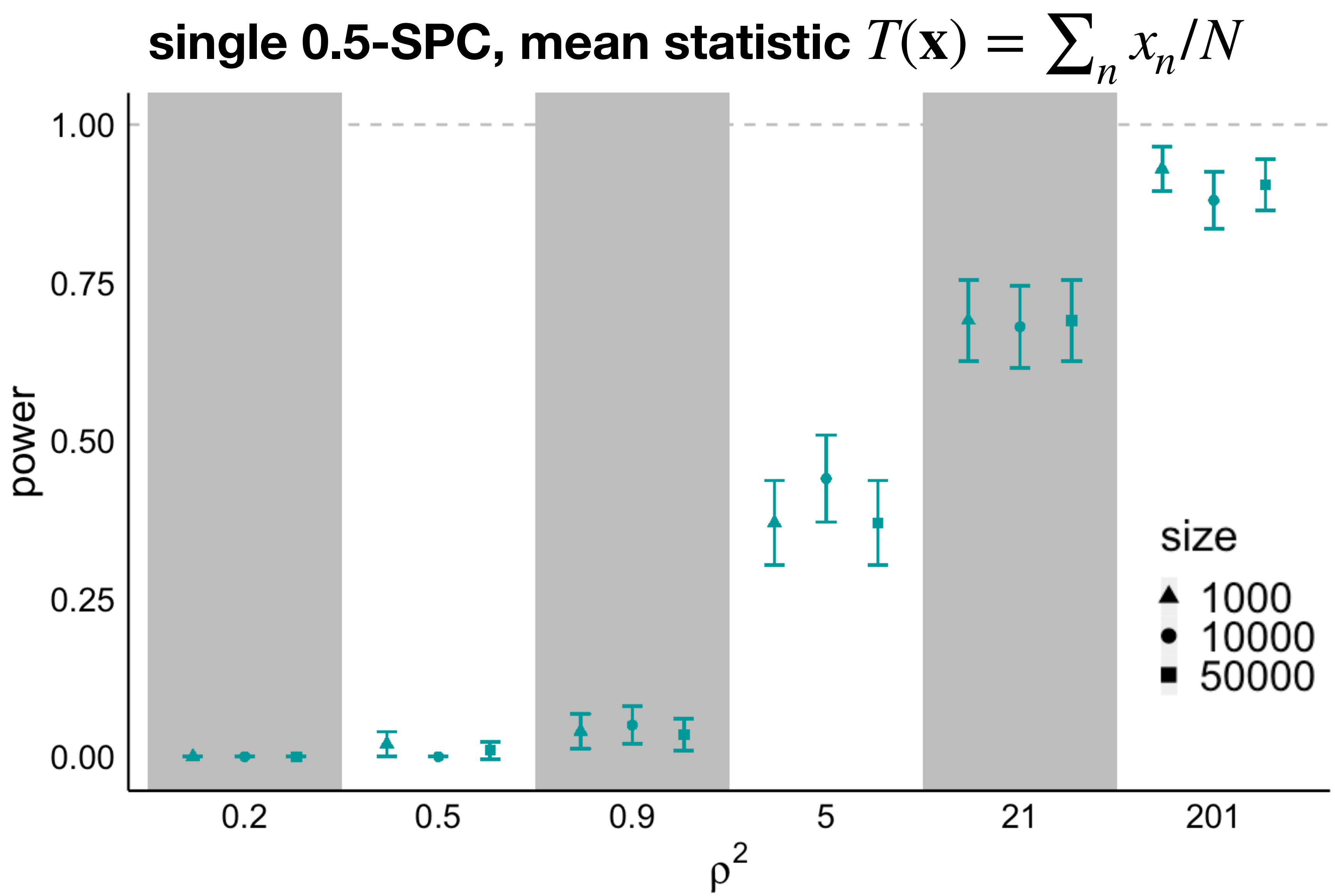
- Recall: if $\nu_\circ = \nu(\theta_\star)$, asymptotic power depends on ρ

Possible poor power of single SPCs when misspecification is subtle

- Recall: if $\nu_\circ = \nu(\theta_\star)$, asymptotic power depends on ρ
- For mean statistic $T(\mathbf{x}) = \sum_n x_n/N$, have $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ/\sigma(\theta_\star)$

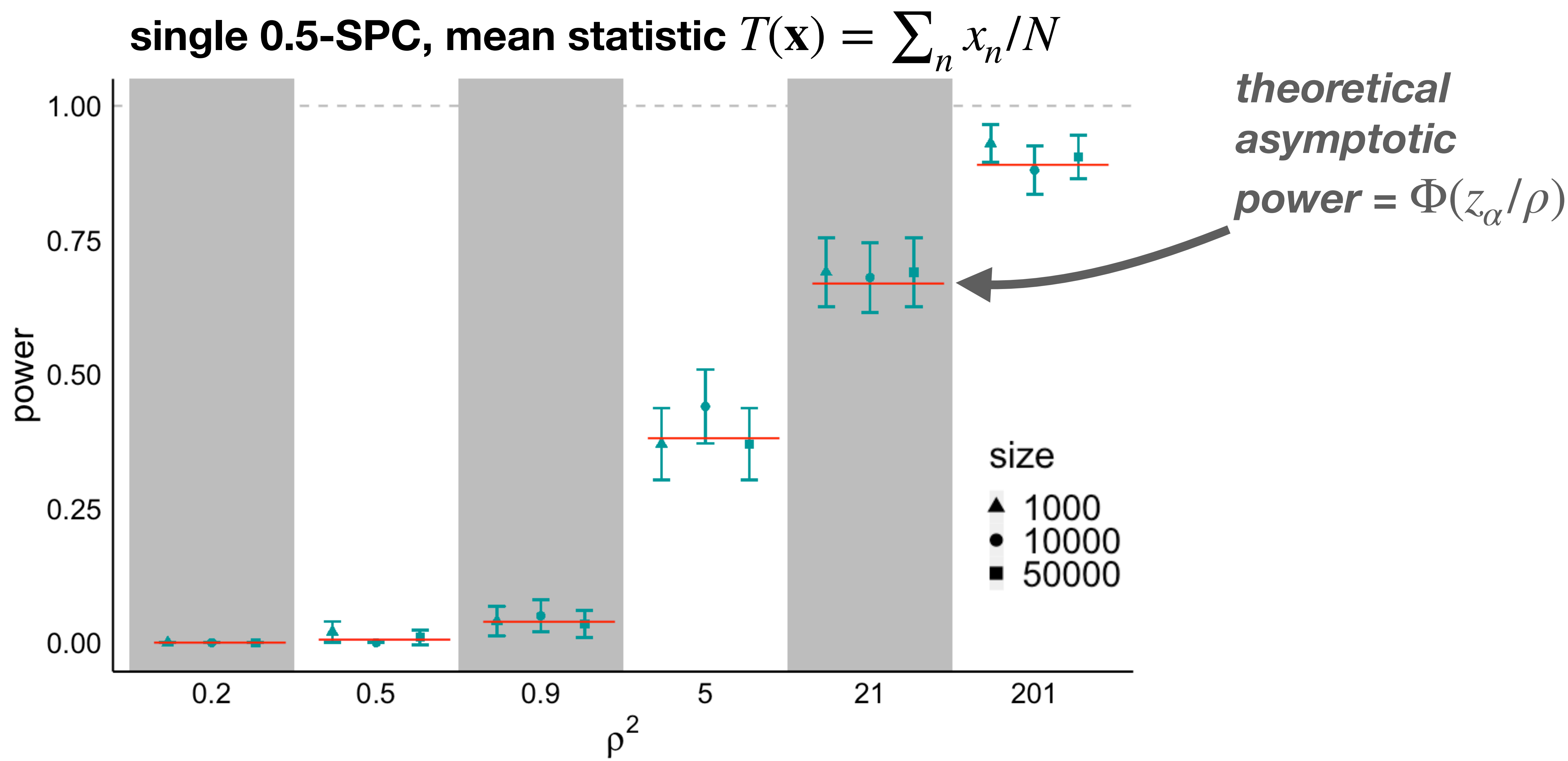
Possible poor power of single SPCs when misspecification is subtle

- Recall: if $\nu_\circ = \nu(\theta_\star)$, asymptotic power depends on ρ
- For mean statistic $T(\mathbf{x}) = \sum_n x_n/N$, have $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ/\sigma(\theta_\star)$



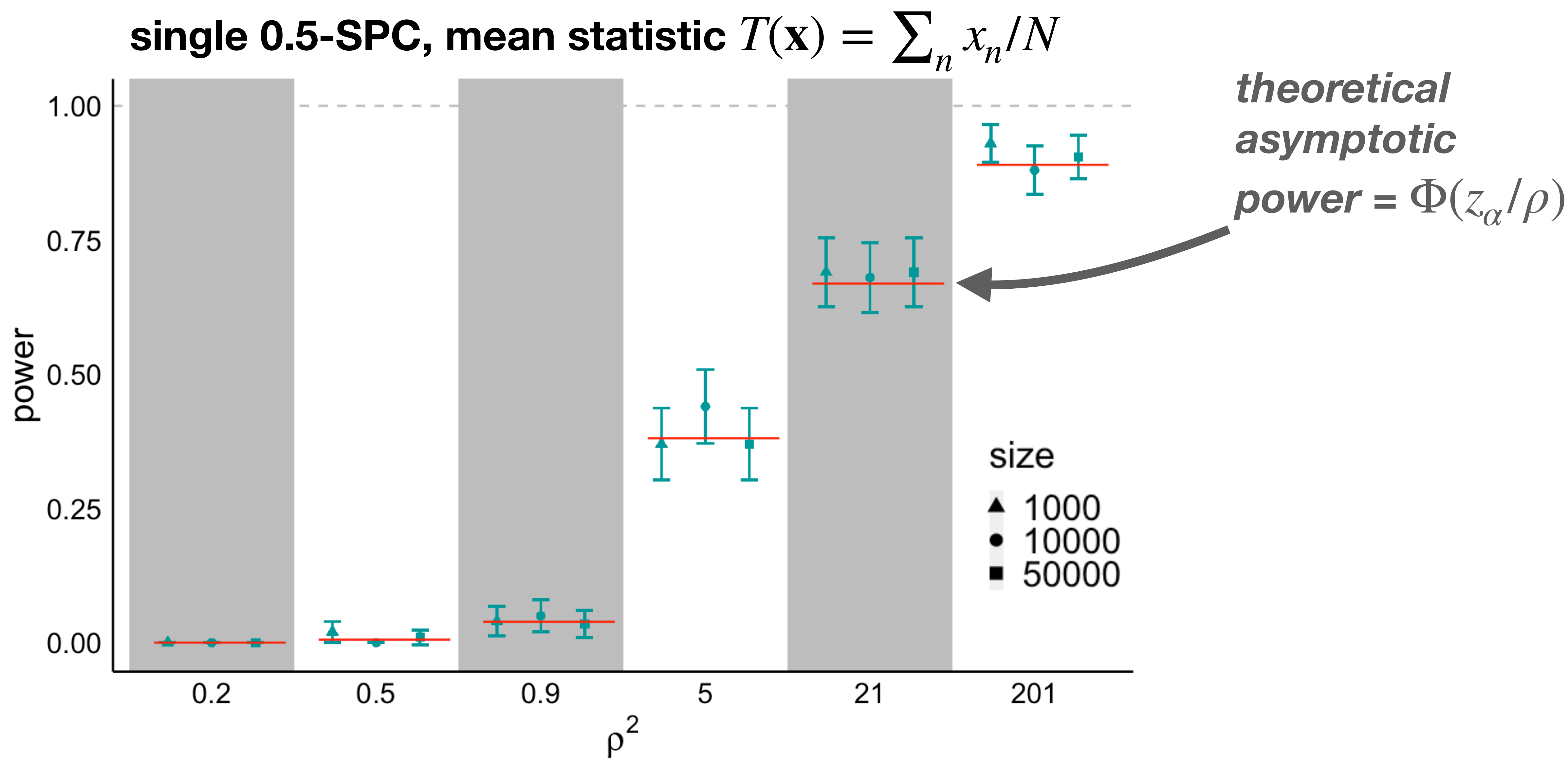
Possible poor power of single SPCs when misspecification is subtle

- Recall: if $\nu_\circ = \nu(\theta_\star)$, asymptotic power depends on ρ
- For mean statistic $T(\mathbf{x}) = \sum_n x_n/N$, have $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ/\sigma(\theta_\star)$



Possible poor power of single SPCs when misspecification is subtle

- Recall: if $\nu_\circ = \nu(\theta_\star)$, asymptotic power depends on ρ
- For mean statistic $T(\mathbf{x}) = \sum_n x_n/N$, have $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ/\sigma(\theta_\star)$



Question 2: Can we improve power to detect subtle misspecification?

An SPC with better power

An SPC with better power

- **Goal:** propose an SPC that yields larger power under ***subtle misspecification*** – while maintaining calibration properties

An SPC with better power

- **Goal:** propose an SPC that yields larger power under ***subtle misspecification*** – while maintaining calibration properties
- From previous theorems, we expect:

An SPC with better power

- **Goal:** propose an SPC that yields larger power under ***subtle misspecification*** – while maintaining calibration properties

- From previous theorems, we expect:

  - when model is **misspecified**, single SPC p-values are ***non-uniform*** (also true when ρ is small)

An SPC with better power

- **Goal:** propose an SPC that yields larger power under ***subtle misspecification*** – while maintaining calibration properties
- From previous theorems, we expect:
  - when model is **misspecified**, single SPC p-values are ***non-uniform*** (also true when ρ is small)
  - when model is **correctly-specified**, single SPC p-values are ***uniform***

An SPC with better power

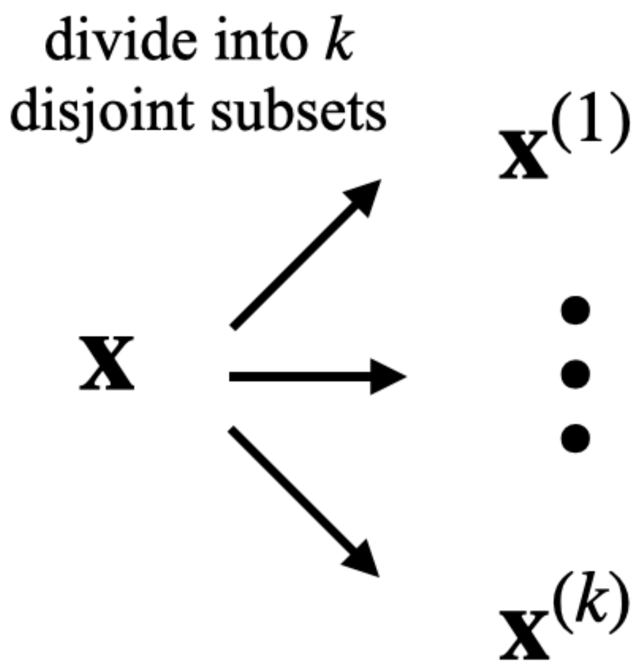
- **Goal:** propose an SPC that yields larger power under ***subtle misspecification*** – while maintaining calibration properties
- From previous theorems, we expect:
  - when model is **misspecified**, single SPC p-values are ***non-uniform*** (also true when ρ is small)
  - when model is **correctly-specified**, single SPC p-values are ***uniform***
- **Idea:** Test for ***uniformity*** of single SPC p-values

Divided SPCs

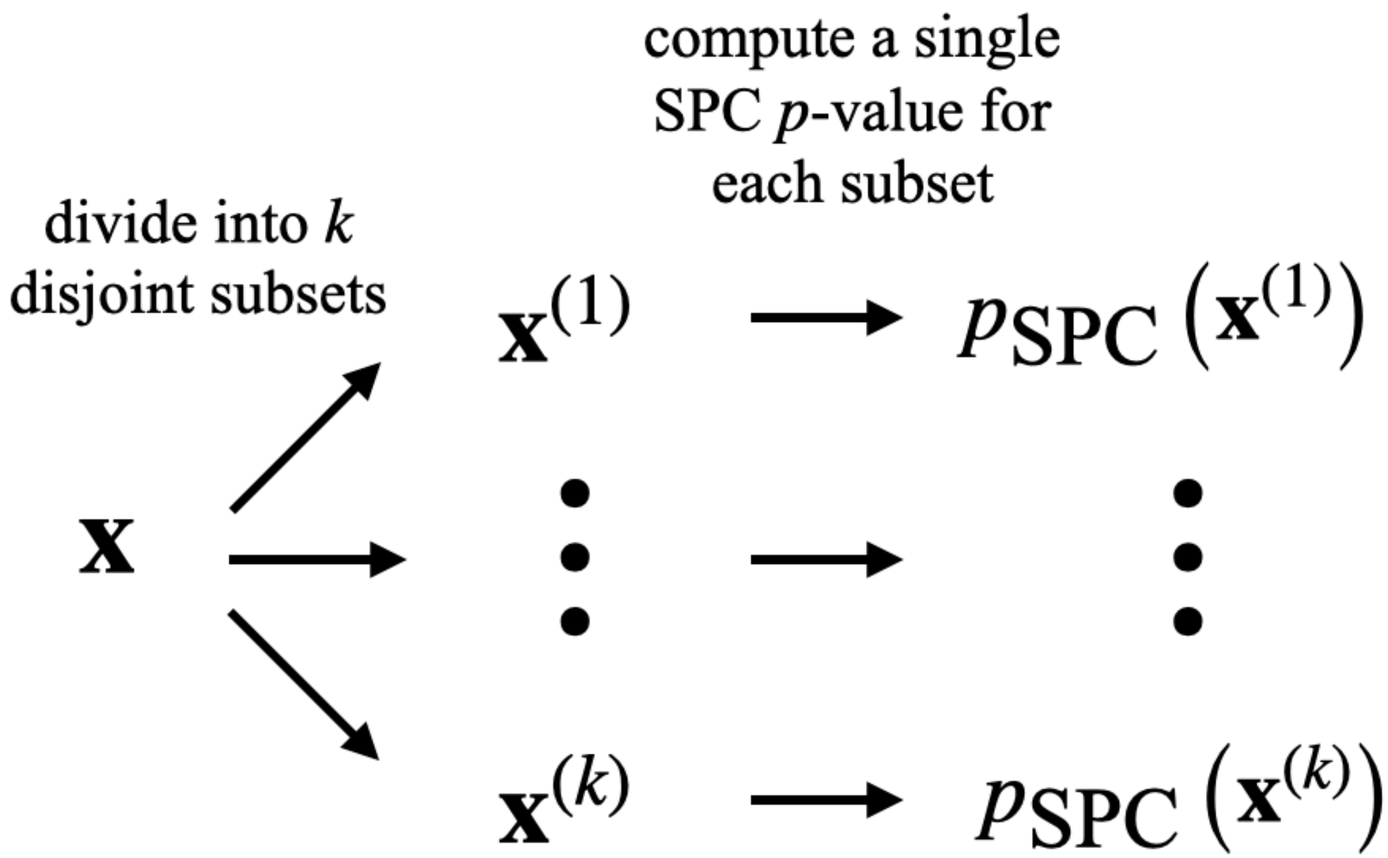
Divided SPCs

$\mathbf{x}$

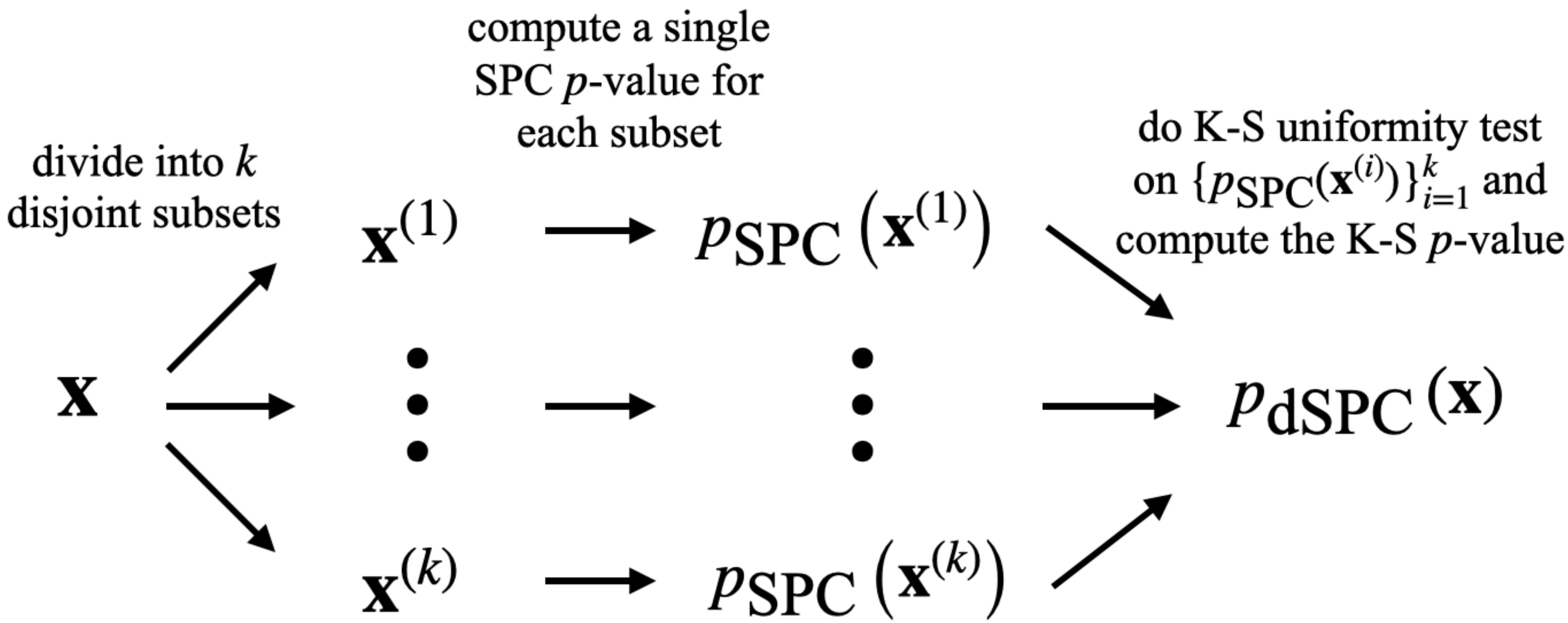
Divided SPCs



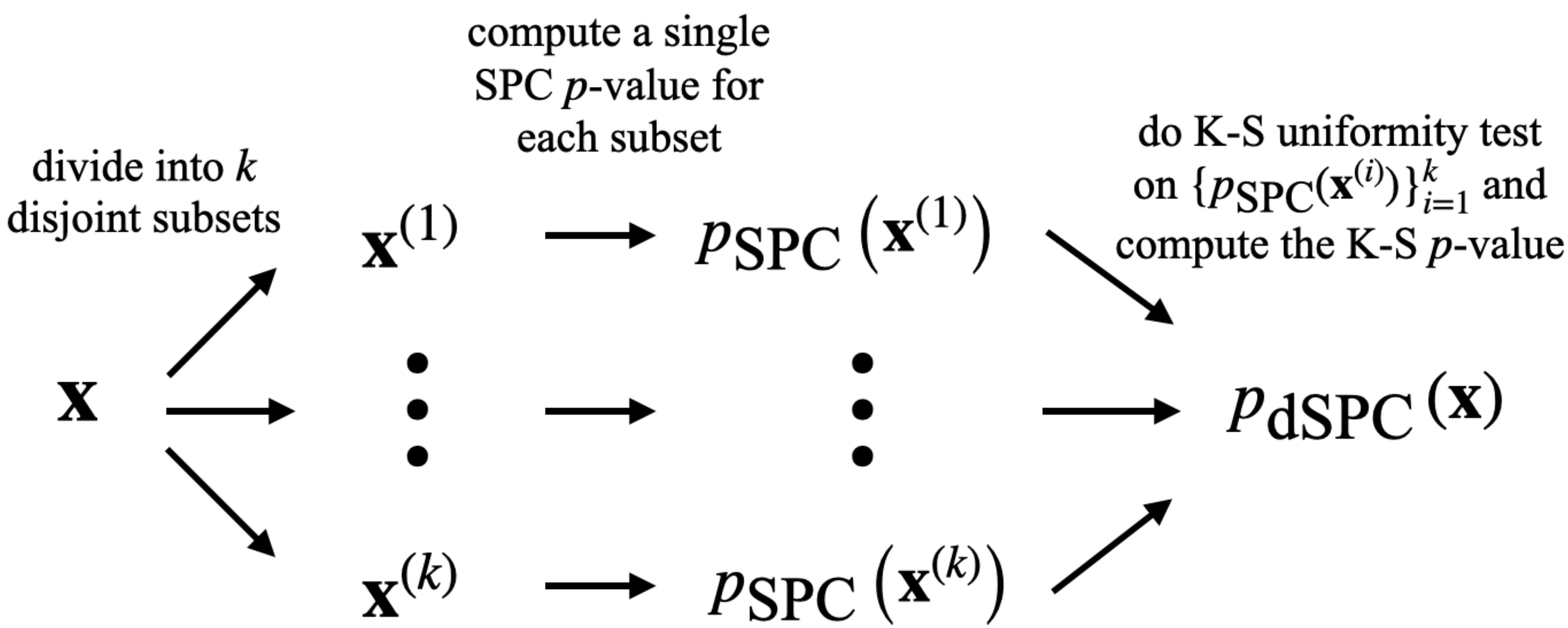
Divided SPCs



Divided SPCs



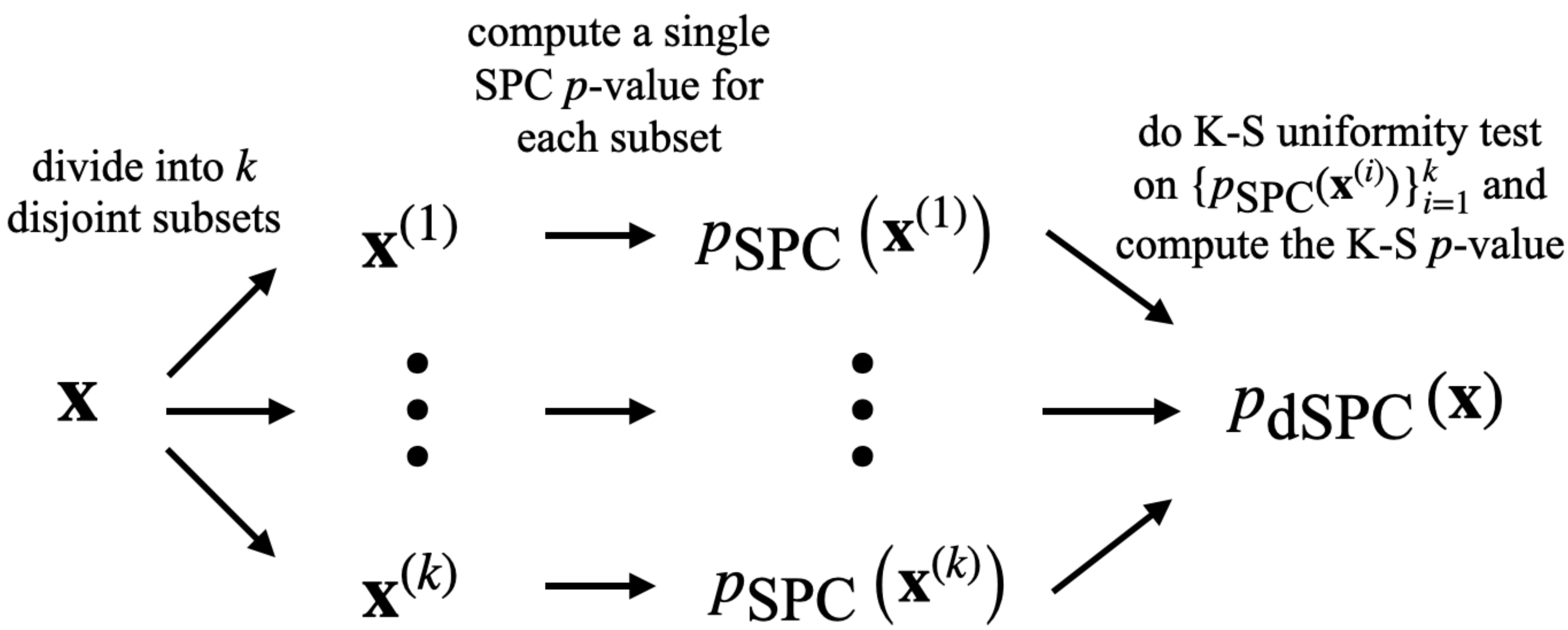
Divided SPCs



- Divided SPC p-value:

$$p_{dSPC}(\mathbf{x}) := p_{KS}(p_{SPC}(\mathbf{x}^{(1)}), \ldots, p_{SPC}(\mathbf{x}^{(k)}))$$

Divided SPCs

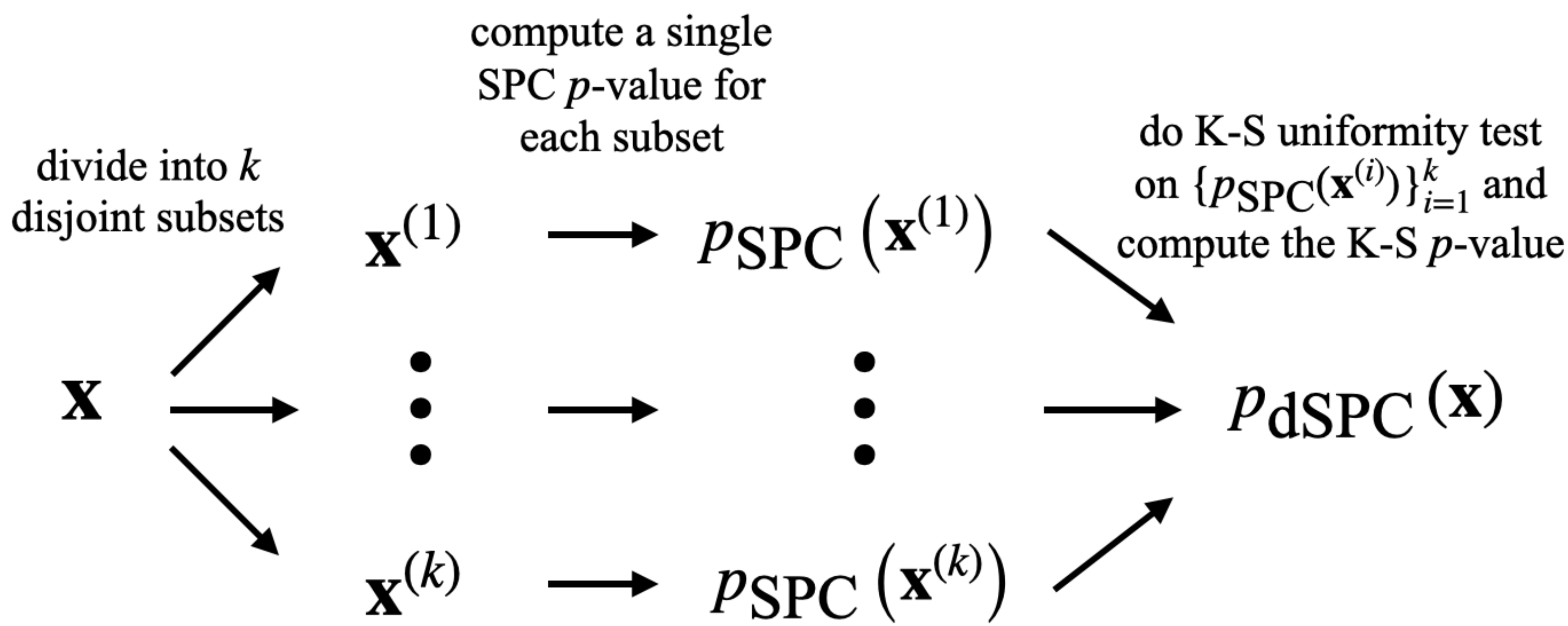


- Divided SPC p-value:

Kolmogorov-Smirnov uniformity test p-value

$$p_{dSPC}(\mathbf{x}) := p_{KS}(p_{SPC}(\mathbf{x}^{(1)}), \ldots, p_{SPC}(\mathbf{x}^{(k)}))$$

Divided SPCs



- Divided SPC p-value:

Kolmogorov-Smirnov uniformity test p-value

$$p_{dSPC}(\mathbf{x}) := p_{KS}(p_{SPC}(\mathbf{x}^{(1)}), \ldots, p_{SPC}(\mathbf{x}^{(k)}))$$

Question 3: How is calibration and power of divided SPCs?

Divided SPCs are well-calibrated and have asymptotic power 1

Divided SPCs are well-calibrated and have asymptotic power 1

Theorem [LH22]: if H_0 holds, then $\mathbb{P}\{p_{dSPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Divided SPCs are well-calibrated and have asymptotic power 1

Theorem [LH22]: if H_0 holds, then $\mathbb{P}\{p_{dSPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$ or $\rho \neq 1$ (so H_1 holds), then power $\xrightarrow{P} 1$

Divided SPCs are well-calibrated and have asymptotic power 1

Theorem [LH22]: if H_0 holds, then $\mathbb{P}\{p_{dSPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$ or $\rho \neq 1$ (so H_1 holds),
then power $\xrightarrow{P} 1$

Question 4: How to scale k in divided SPCs?

Divided SPCs are well-calibrated and have asymptotic power 1

Theorem [LH22]: if H_0 holds, then $\mathbb{P}\{p_{dSPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$ or $\rho \neq 1$ (so H_1 holds),
then power $\xrightarrow{P} 1$

Question 4: How to scale k in divided SPCs?

Condition: theorems hold only if $k = O(N^{\beta - \varepsilon})$ [for some $\varepsilon > 0$]

Divided SPCs are well-calibrated and have asymptotic power 1

Theorem [LH22]: if H_0 holds, then $\mathbb{P}\{p_{dSPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$ or $\rho \neq 1$ (so H_1 holds), then power $\xrightarrow{P} 1$

Question 4: How to scale k in divided SPCs?

*$\beta = 1/2$ **if model is smooth enough and higher-order moments exist***

Condition: theorems hold only if $k = O(N^{\beta - \varepsilon})$ [for some $\varepsilon > 0$]

Divided SPCs are well-calibrated and have asymptotic power 1

Theorem [LH22]: if H_0 holds, then $\mathbb{P}\{p_{dSPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Theorem [LH22]: if $\nu_\circ \neq \nu(\theta_\star)$ or $\rho \neq 1$ (so H_1 holds), then power $\xrightarrow{P} 1$

Question 4: How to scale k in divided SPCs?

$\beta = 1/2$ if model is smooth enough and higher-order moments exist

Condition: theorems hold only if $k = O(N^{\beta - \varepsilon})$ [for some $\varepsilon > 0$]

- In practice, we set $k = N^{0.49}$

Numerical investigation: subtle-to-mild misspecification

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

- Model $P_\theta = \text{Poiss}(\theta)$

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x})P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x})P_{\theta_\star}(d\mathbf{x})$$

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

over-dispersed

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

over-dispersed *under-dispersed*

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

over-dispersed ***under-dispersed***

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- For mean statistic $T(\mathbf{x}) = \sum_n x_n / N$, we have

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

over-dispersed *under-dispersed*

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- For mean statistic $T(\mathbf{x}) = \sum_n x_n / N$, we have
  - $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ / \sigma(\theta_\star)$

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

over-dispersed ***under-dispersed***

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- For mean statistic $T(\mathbf{x}) = \sum_n x_n / N$, we have
  - $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ / \sigma(\theta_\star)$
- Recall:

Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

over-dispersed ***under-dispersed***

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- For mean statistic $T(\mathbf{x}) = \sum_n x_n / N$, we have
  - $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ / \sigma(\theta_\star)$
- Recall:
  - Single SPCs: if $\nu_\circ = \nu(\theta_\star)$ asymptotic power depends on ρ

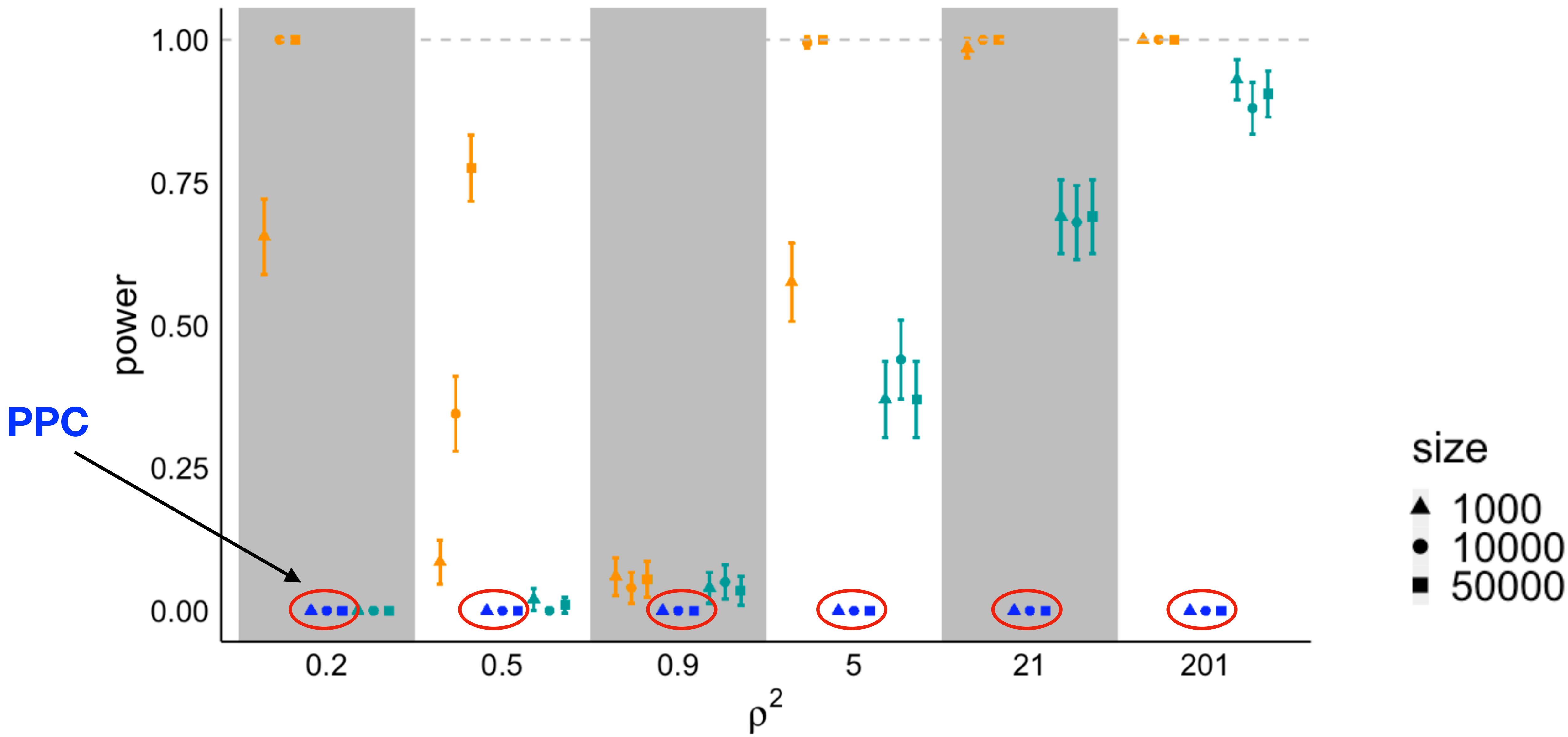
Numerical investigation: subtle-to-mild misspecification

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \quad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

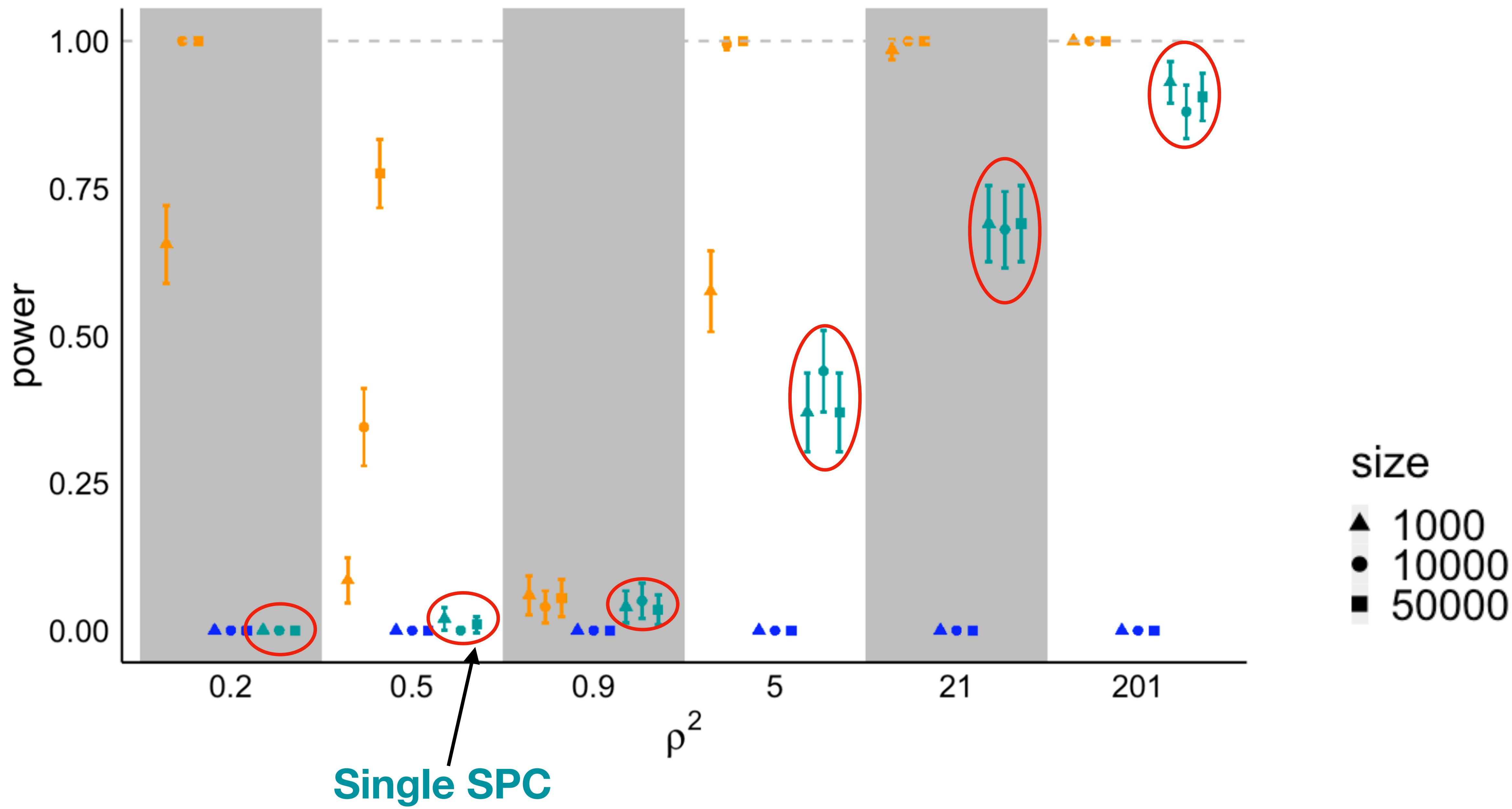
over-dispersed ***under-dispersed***

- Model $P_\theta = \text{Poiss}(\theta)$
- Generate data from $P_\circ = \text{NegBin}(2, \tau)$ and $P_\circ = \text{Binom}(30, p)$
- For mean statistic $T(\mathbf{x}) = \sum_n x_n / N$, we have
    - $\nu_\circ = \nu(\theta_\star)$ and $\rho = \sigma_\circ / \sigma(\theta_\star)$
- Recall:
    - Single SPCs: if $\nu_\circ = \nu(\theta_\star)$ asymptotic power depends on ρ
    - Divided SPCs: asymptotic power 1

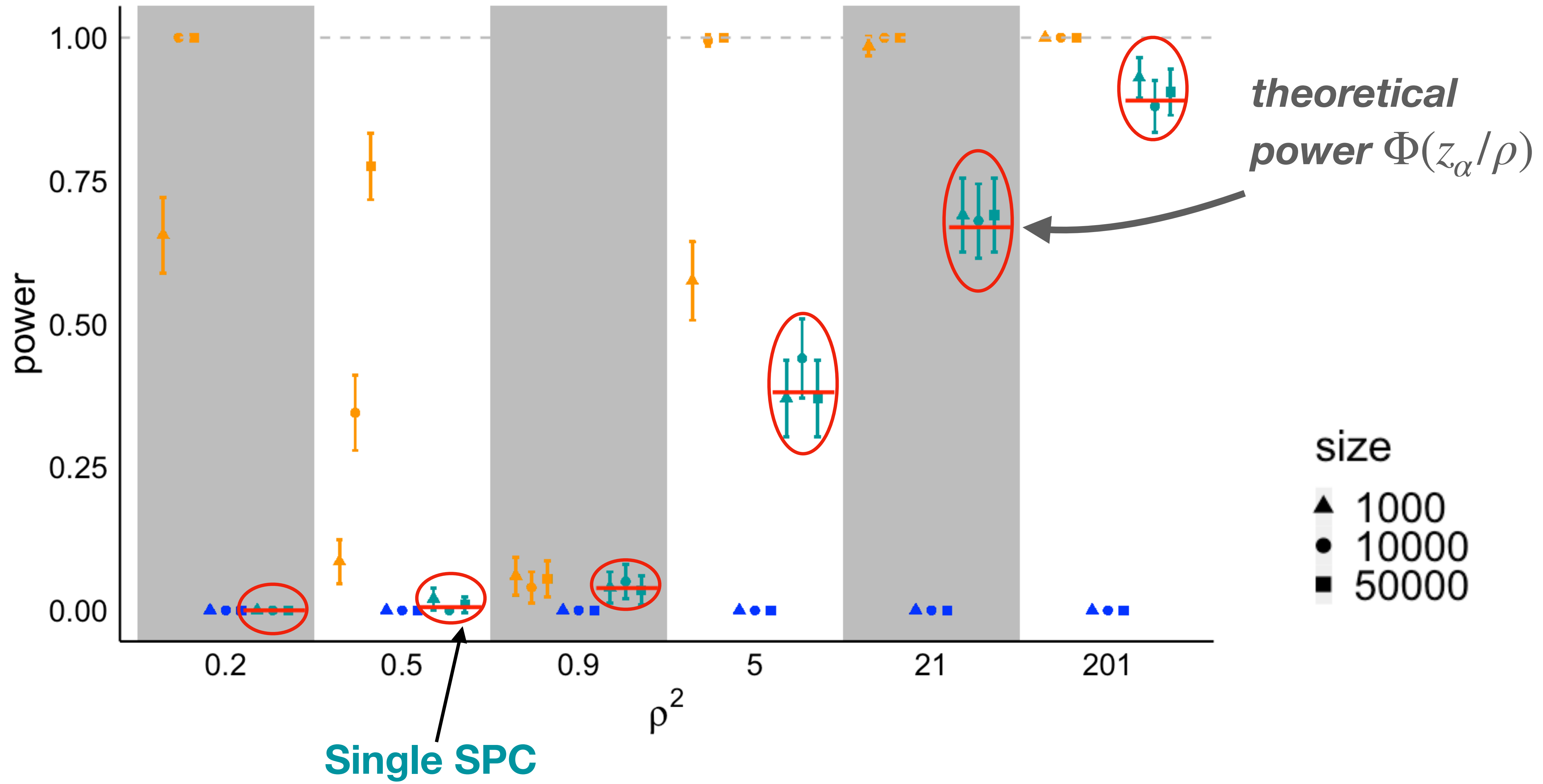
Divided SPCs have superior power



Divided SPCs have superior power

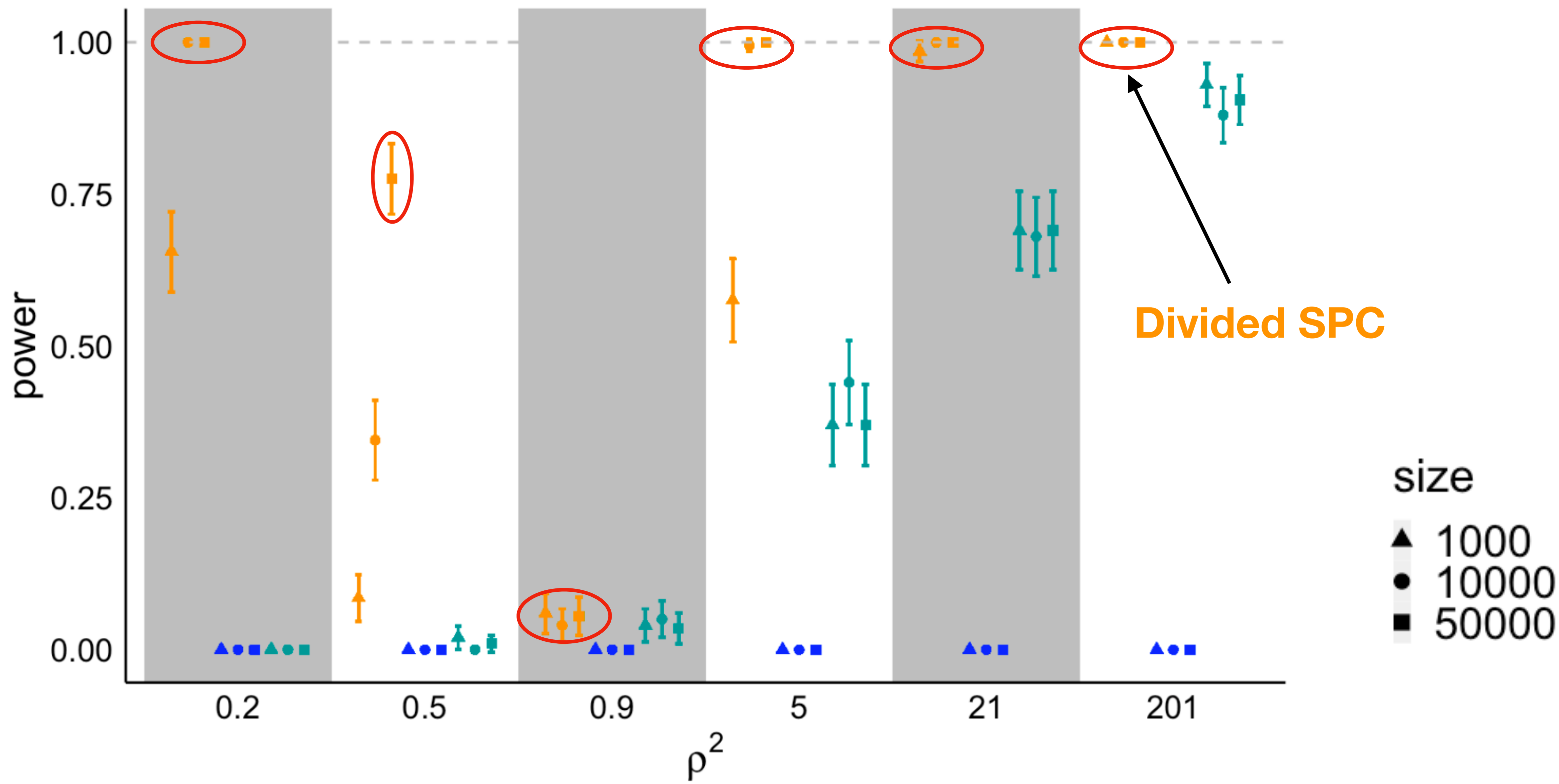


Divided SPCs have superior power



✓ If $\nu_\circ = \nu(\theta_\star)$ asymptotic power of single SPC depends on ρ

Divided SPCs have superior power



✔ If $\nu_\circ = \nu(\theta_\star)$ asymptotic power of single SPC depends on ρ

✔ Divided SPCs have asymptotic power 1

Comparison between single and divided SPCs

Comparison between single and divided SPCs

Single SPCs	Divided SPCs

Comparison between single and divided SPCs

Single SPCs	Divided SPCs
Compute posterior once	

Comparison between single and divided SPCs

Single SPCs	Divided SPCs
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*

Comparison between single and divided SPCs

Single SPCs	Divided SPCs
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*
Asymptotic power could be small (if subtle misspecification)	

Comparison between single and divided SPCs

Single SPCs	Divided SPCs
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*
Asymptotic power could be small (if subtle misspecification)	Asymptotic power = 1

Comparison between single and divided SPCs

Single SPCs	Divided SPCs
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*
Asymptotic power could be small (if subtle misspecification)	Asymptotic power = 1
Doesn’t requires much data *(single “train-test” split)*	

Comparison between single and divided SPCs

Single SPCs	**Divided SPCs**
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*
Asymptotic power could be small (if subtle misspecification)	Asymptotic power = 1
Doesn't requires much data *(single "train-test" split)*	Requires more data *(compute k single SPC p-values)*

Comparison between single and divided SPCs

Single SPCs	Divided SPCs
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*
Asymptotic power could be small (if subtle misspecification)	Asymptotic power = 1
Doesn’t requires much data *(single “train-test” split)*	Requires more data *(compute k single SPC p-values)*

- Single SPCs are always a good place to start

Comparison between single and divided SPCs

Single SPCs	**Divided SPCs**
Compute posterior once	Compute posterior k times *(with k times smaller datasets)*
Asymptotic power could be small (if subtle misspecification)	Asymptotic power = 1
Doesn’t requires much data *(single “train-test” split)*	Requires more data *(compute k single SPC p-values)*

- Single SPCs are always a good place to start
- Consider divided SPCs if single SPCs inconclusive

How to “split” in single SPCs with time-series data?

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

How to "split" in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

$$x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor} \quad x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N$$

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

$$x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor} \;\vdots\; x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N$$

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

$$\underbrace{x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\mathbf{obs}}} \vdots \; x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N$$

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

$$\underbrace{x_1 \quad \ldots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x_{obs}}} \vdots \underbrace{x_{\lfloor N/2 \rfloor + 1} \quad \ldots \quad x_N}_{\mathbf{x_{ho}}}$$

How to "split" in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

↓

***predict future** well?*

$$\underbrace{x_1 \quad \cdots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\mathrm{obs}}} \vdots \underbrace{x_{\lfloor N/2 \rfloor + 1} \quad \cdots \quad x_N}_{\mathbf{x}_{\mathrm{ho}}}$$

How to "split" in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

↓

predict future well?

$$\underbrace{x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\mathbf{obs}}} \vdots \underbrace{x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N}_{\mathbf{x}_{\mathbf{ho}}}$$

interpolated single SPC

How to "split" in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

predict future well?

$$\underbrace{x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\mathrm{obs}}} \,\vdots\, \underbrace{x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N}_{\mathbf{x}_{\mathrm{ho}}}$$

interpolated single SPC

$$x_1 \quad x_2 \quad \dots \quad x_i \quad x_{i+1} \quad \dots \quad x_{N-1} \quad x_N$$

How to “split” in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

↓

***predict future** well?*

$$\underbrace{x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\mathbf{obs}}} \vdots \underbrace{x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N}_{\mathbf{x}_{\mathbf{ho}}}$$

interpolated single SPC

$$x_1 \quad x_2 \quad \dots \quad x_i \quad x_{i+1} \quad \dots \quad x_{N-1} \quad x_N$$

How to "split" in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

$$\underbrace{x_1 \quad \dots \quad x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\mathrm{obs}}} \,\vdots\, \underbrace{x_{\lfloor N/2 \rfloor + 1} \quad \dots \quad x_N}_{\mathbf{x}_{\mathrm{ho}}}$$

***predict future** well?*

interpolated single SPC

$$x_1 \quad x_2 \quad \dots \quad x_i \quad x_{i+1} \quad \dots \quad x_{N-1} \quad x_N$$

$$\underbrace{x_1 \;\; x_3 \;\; \dots \;\; x_i \;\; x_{i+2} \;\; \dots \;\; x_{N-1}}_{\mathbf{x}_{\mathrm{obs}}} \quad \underbrace{x_2 \;\; x_4 \;\; \dots \;\; x_{i+1} \;\; x_{i+3} \;\; \dots \;\; x_N}_{\mathbf{x}_{\mathrm{ho}}}$$

How to "split" in single SPCs with time-series data?

- If data are structured: must decide how to split the data
- Example: single 0.5-SPC with time-series data

extrapolated single SPC

↓

predict future well?

$$\underbrace{x_1 \;\; \ldots \;\; x_{\lfloor N/2 \rfloor}}_{\mathbf{x}_{\text{obs}}} \; \vdots \; \underbrace{x_{\lfloor N/2 \rfloor + 1} \;\; \ldots \;\; x_N}_{\mathbf{x}_{\text{ho}}}$$

interpolated single SPC

↓

explain data well?

$$x_1 \quad x_2 \quad \ldots \quad x_i \quad x_{i+1} \quad \ldots \quad x_{N-1} \quad x_N$$

$$\underbrace{x_1 \;\; x_3 \;\; \ldots \;\; x_i \;\; x_{i+2} \;\; \ldots \;\; x_{N-1}}_{\mathbf{x}_{\text{obs}}} \quad \underbrace{x_2 \;\; x_4 \;\; \ldots \;\; x_{i+1} \;\; x_{i+3} \;\; \ldots \; x_N}_{\mathbf{x}_{\text{ho}}}$$

SPCs for time series data

SPCs for time series data

- Airline delays data (N = 327,346)

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance
  - Response: arrival delays in minutes

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance
  - Response: arrival delays in minutes
  - Success rate statistic $T(\mathbf{x}) := N / \sum_n x_n$

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance
  - Response: arrival delays in minutes
  - Success rate statistic $T(\mathbf{x}) := N/\sum_n x_n$

model fits data well

Methods	P-values
PPC	0.132
Interpolated single 0.5-SPC	0.626
Extrapolated single 0.5-SPC	0.000
Double-interpolated divided 0.5-SPC, $k = N^{0.49}$	0.000
Interpolated divided extrapolated 0.5-SPC, $k = N^{0.49}$	0.001

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance
  - Response: arrival delays in minutes
  - Success rate statistic $T(\mathbf{x}) := N/\sum_n x_n$

Methods	P-values
PPC	0.132
Interpolated single 0.5-SPC	0.626
Extrapolated single 0.5-SPC	$\boxed{0.000}$
Double-interpolated divided 0.5-SPC, $k = N^{0.49}$	0.000
Interpolated divided extrapolated 0.5-SPC, $k = N^{0.49}$	0.001

model fits data well

model ***does not predict*** *future well*

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance
  - Response: arrival delays in minutes
  - Success rate statistic $T(\mathbf{x}) := N / \sum_n x_n$

Methods	P-values
PPC	0.132
Interpolated single 0.5-SPC	0.626
Extrapolated single 0.5-SPC	0.000
Double-interpolated divided 0.5-SPC, $k = N^{0.49}$	$\boxed{0.000}$
Interpolated divided extrapolated 0.5-SPC, $k = N^{0.49}$	$\boxed{0.001}$

model fits data well

*model **does not predict** future well*

*both **divided SPCs** pick up misspecification*

SPCs for time series data

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: month, day of week, distance
  - Response: arrival delays in minutes
  - Success rate statistic $T(\mathbf{x}) := N / \sum_n x_n$

Methods	P-values
PPC	0.132
Interpolated single 0.5-SPC	0.626
Extrapolated single 0.5-SPC	0.000
Double-interpolated divided 0.5-SPC, $k = N^{0.49}$	$\boxed{0.000}$
Interpolated divided extrapolated 0.5-SPC, $k = N^{0.49}$	$\boxed{0.001}$

model fits data well

*both **divided SPCs** pick up misspecification*

*model **does not predict** future well*

- Different SPCs offer different (potentially useful!) insights

Concluding Thoughts

Concluding Thoughts

✓ Unlike PPCs, SPCs are **well-calibrated and well-powered…**

Concluding Thoughts

✓ Unlike PPCs, SPCs are **well-calibrated and well-powered…**
while also being general-purpose, computational efficient, and flexible

Concluding Thoughts

- ✓ Unlike PPCs, SPCs are **well-calibrated and well-powered…** while also being general-purpose, computational efficient, and flexible
- ✓ Single SPCs and divided SPCs offer **complimentary strengths**

Concluding Thoughts

✓ Unlike PPCs, SPCs are **well-calibrated and well-powered…**
while also being general-purpose, computational efficient, and flexible

✓ Single SPCs and divided SPCs offer **complimentary strengths**

✓ SPCs introduce extra flexibility with different structured data and complex models such as time series, hierarchical, spatial data…

Concluding Thoughts

- ✓ Unlike PPCs, SPCs are **well-calibrated and well-powered…**
  while also being general-purpose, computational efficient, and flexible
- ✓ Single SPCs and divided SPCs offer **complimentary strengths**
- ✓ SPCs introduce extra flexibility with different structured data and complex models such as time series, hierarchical, spatial data…
- More to do: non-asymptotic results, non-normal limiting distributions, cross-validation versions, …

Concluding Thoughts

✓ Unlike PPCs, SPCs are **well-calibrated and well-powered…**
while also being general-purpose, computational efficient, and flexible

✓ Single SPCs and divided SPCs offer **complimentary strengths**

✓ SPCs introduce extra flexibility with different structured data and complex models such as time series, hierarchical, spatial data…

- More to do: non-asymptotic results, non-normal limiting distributions, cross-validation versions, …

J. Li and J. H. Huggins (2022). Calibrated Model Criticism Using Split Predictive Checks. *arXiv:2203.15897 [stat.ME]*

read the paper

Details of asymptotic theory for single SPCs

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \Sigma_\star J_\star^{-1} \dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \dot{\nu}(\theta_\star)}$$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}}^{\text{🥪}} \dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \dot{\nu}(\theta_\star)}$$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}}^{\text{🥪}} \dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \dot{\nu}(\theta_\star)}$$

where $\Sigma_\star = \mathrm{Cov}\{\, \nabla \ell(\theta_\star; X)\}$ and $J_\star = -\,\mathbb{E}\{\, \nabla^{\otimes 2} \ell(\theta_\star; X)\}$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

q controls the weights of two forms of mismatch

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}}^{\text{🥪}} \dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \dot{\nu}(\theta_\star)}$$

where $\Sigma_\star = \mathrm{Cov}\{\nabla \ell(\theta_\star; X)\}$ and $J_\star = -\mathbb{E}\{\nabla^{\otimes 2} \ell(\theta_\star; X)\}$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\dfrac{z_\alpha}{\rho}\right)$

q controls the weights of two forms of mismatch

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}}^{🥪} \dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \dot{\nu}(\theta_\star)}$$

where $\Sigma_\star = \mathrm{Cov}\{\nabla \ell(\theta_\star; X)\}$ and $J_\star = -\mathbb{E}\{\nabla^{\otimes 2} \ell(\theta_\star; X)\}$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\dfrac{z_\alpha}{\rho}\right)$

q controls the weights of two forms of mismatch

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}}\dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1}\dot{\nu}(\theta_\star)}$$

where $\Sigma_\star = \mathrm{Cov}\{\nabla \ell(\theta_\star; X)\}$ and $J_\star = -\mathbb{E}\{\nabla^{\otimes 2}\ell(\theta_\star; X)\}$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$
- if model misspecified, then in general $\rho \neq 1$:

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\dfrac{z_\alpha}{\rho}\right)$

q controls the weights of two forms of mismatch

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}}\dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1}\dot{\nu}(\theta_\star)}$$

where $\Sigma_\star = \mathrm{Cov}\{\nabla \ell(\theta_\star; X)\}$ and $J_\star = -\mathbb{E}\{\nabla^{\otimes 2}\ell(\theta_\star; X)\}$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$
- if model misspecified, then in general $\rho \neq 1$:
  - when uncertainty is under-estimated, $\rho \uparrow \infty$ and power $\uparrow 1$

Details of asymptotic theory for single SPCs

$$\nu_\circ := \lim_{N\to\infty} \int T(\mathbf{x}) P_\circ(d\mathbf{x}), \qquad \nu(\theta_\star) := \lim_{N\to\infty} \int T(\mathbf{x}) P_{\theta_\star}(d\mathbf{x})$$

Theorem: if $\nu_\circ = \nu(\theta_\star)$, $\mathbb{P}\{p_{SPC}(\mathbf{X}) < \alpha\} \xrightarrow{P} \Phi\left(\frac{z_\alpha}{\rho}\right)$

q controls the weights of two forms of mismatch

$$\rho^2 := \frac{q\sigma_\circ^2 + (1-q)\dot{\nu}(\theta_\star)^T \overbrace{J_\star^{-1}\Sigma_\star J_\star^{-1}} \dot{\nu}(\theta_\star)}{q\sigma(\theta_\star)^2 + (1-q)\dot{\nu}(\theta_\star)^T J_\star^{-1} \dot{\nu}(\theta_\star)}$$

where $\Sigma_\star = \mathrm{Cov}\{\nabla \ell(\theta_\star; X)\}$ and $J_\star = -\mathbb{E}\{\nabla^{\otimes 2} \ell(\theta_\star; X)\}$

- if model correctly specified, then $\rho = 1$ and $\mathbb{P}\{p_{SPC}(\mathbf{x}) < \alpha\} \xrightarrow{P} \alpha$
- if model misspecified, then in general $\rho \neq 1$:
  - when uncertainty is under-estimated, $\rho \uparrow \infty$ and power $\uparrow$ 1
  - when uncertainty is over-estimated, $\rho \downarrow 0$ and power $\downarrow$ 0

Time series data using negative binomial regression model

Time series data using negative binomial regression model

- Airline delays data (N = 327,346)

Time series data using negative binomial regression model

- Airline delays data (N = 327,346)
- Consider negative binomial regression model

Time series data using negative binomial regression model

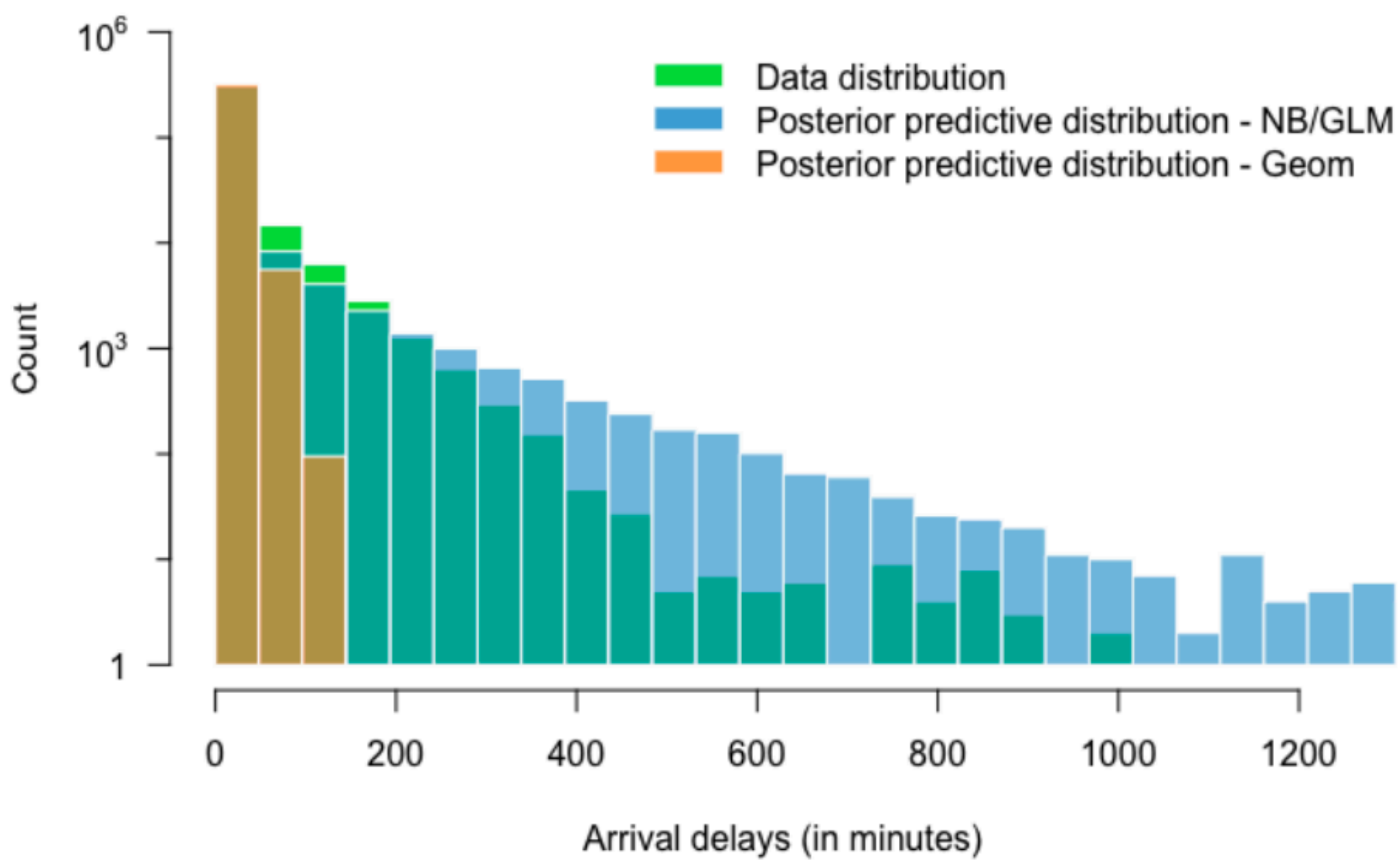
- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: Month, Day of Week and Distance

Time series data using negative binomial regression model

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: Month, Day of Week and Distance
  - Response: Arrival delays in minutes

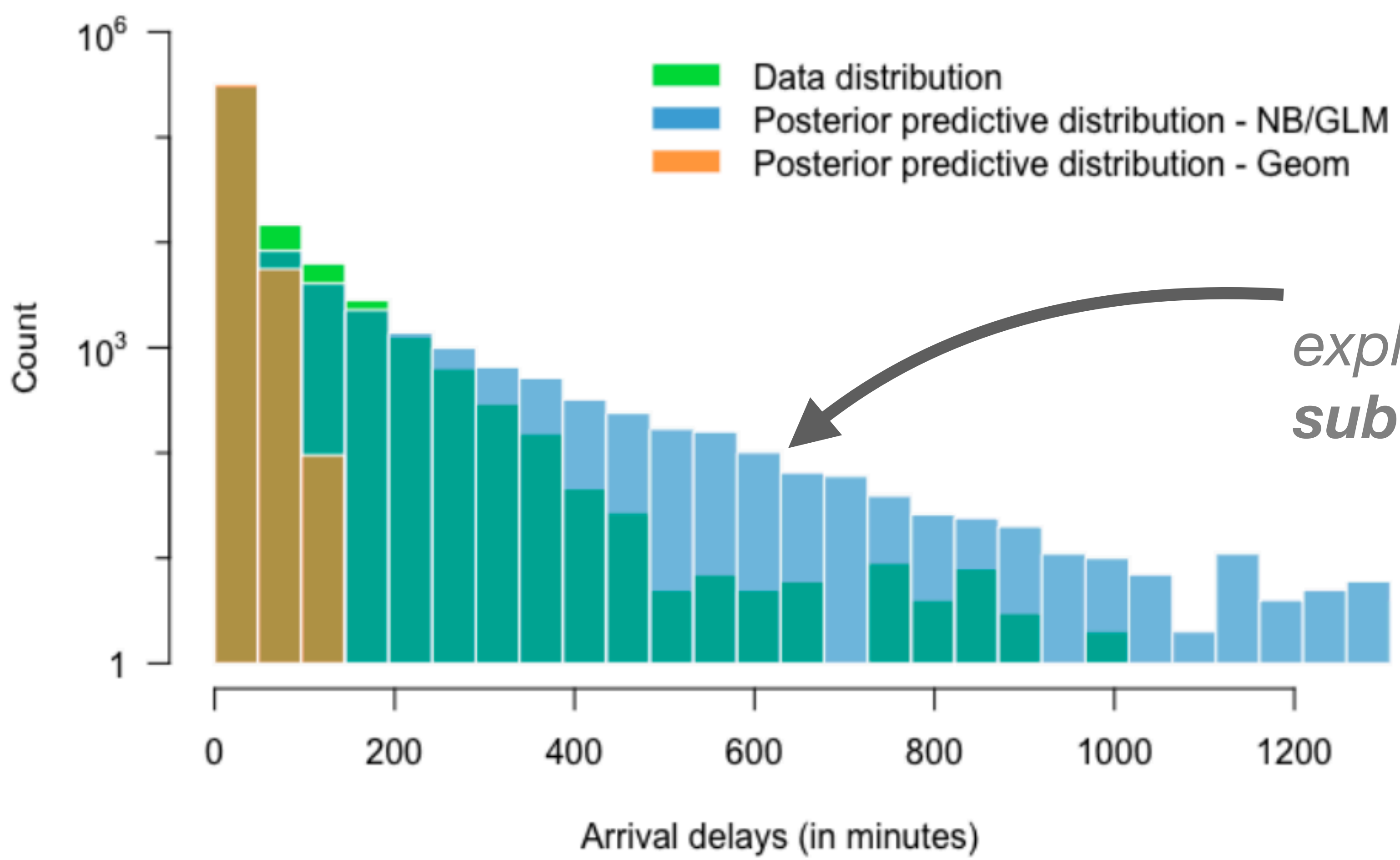
Time series data using negative binomial regression model

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: Month, Day of Week and Distance
  - Response: Arrival delays in minutes



Time series data using negative binomial regression model

- Airline delays data (N = 327,346)
- Consider negative binomial regression model
  - Covariates: Month, Day of Week and Distance
  - Response: Arrival delays in minutes



explains the data fairly well
subtle misspecification!

How to "split" in divided SPC with time-series data?

How to “split” in divided SPC with time-series data?

- Divided SPC

How to "split" in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

How to "split" in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

How to "split" in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\downarrow$$

$$\mathbf{X}^{(1)}$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\mathbf{X}^{(1)} \qquad \mathbf{X}^{(i)}$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$
\begin{array}{ccccccc}
X_1 & \cdots & \cdots & X_i & \cdots & \cdots & X_N \\
\downarrow & & & \downarrow & & & \downarrow \\
\mathbf{X}^{(1)} & & & \mathbf{X}^{(i)} & & & \mathbf{X}^{(k)}
\end{array}
$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$
\begin{array}{ccccc}
X_1 \quad \cdots & \cdots & X_i \quad \cdots & \cdots & X_N \\
\downarrow & & \downarrow & & \downarrow \\
\mathbf{X}^{(1)} & \cdots & \mathbf{X}^{(i)} & \cdots & \mathbf{X}^{(k)}
\end{array}
$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

extrapolated dividing

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

extrapolated dividing

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \mid \quad \cdots \mid \quad X_i \quad \mid \cdots \quad \cdots \mid \quad X_N$$

extrapolated dividing

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

extrapolated dividing

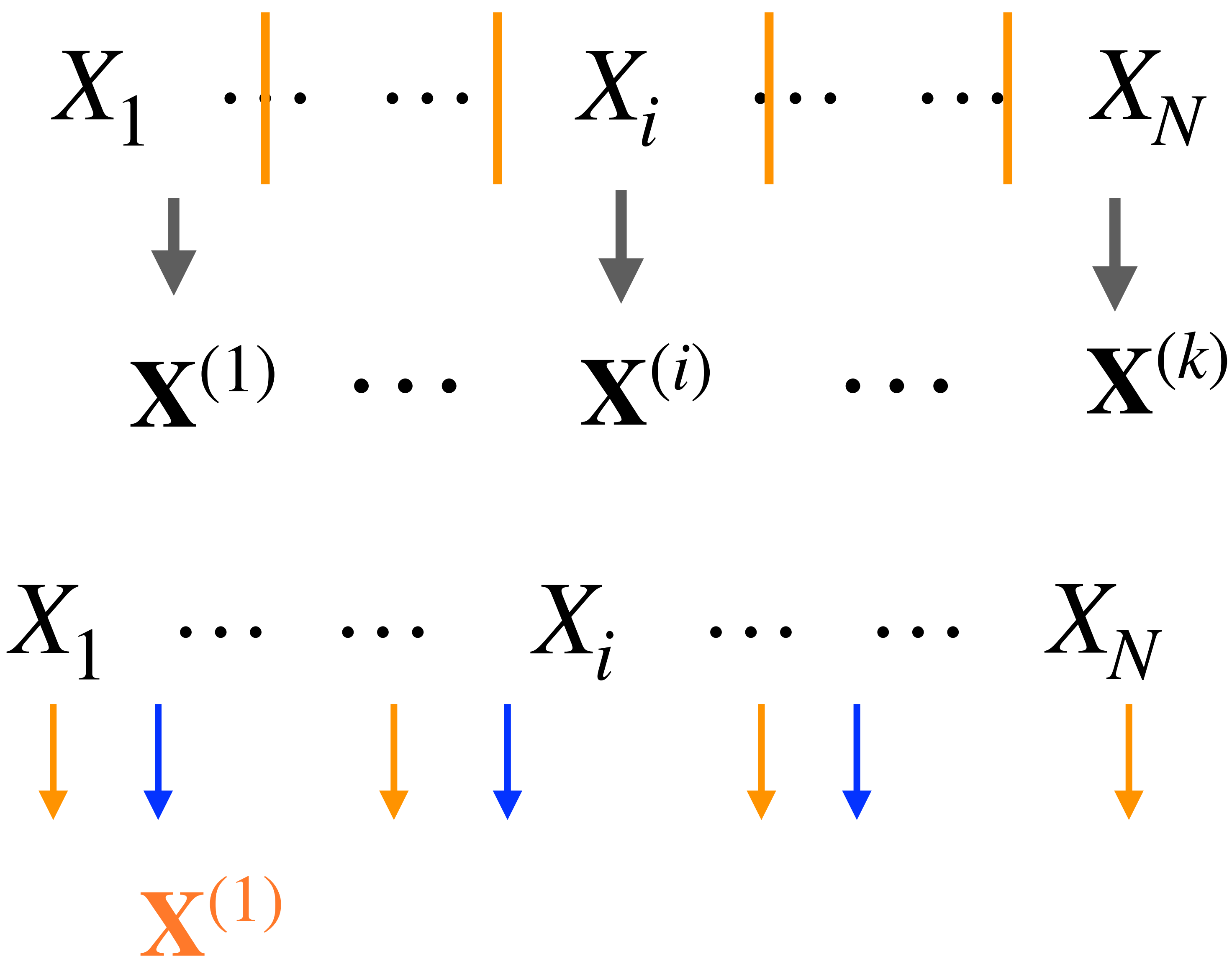
$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\mathbf{X}^{(1)}$$

How to “split” in divided SPC with time-series data?

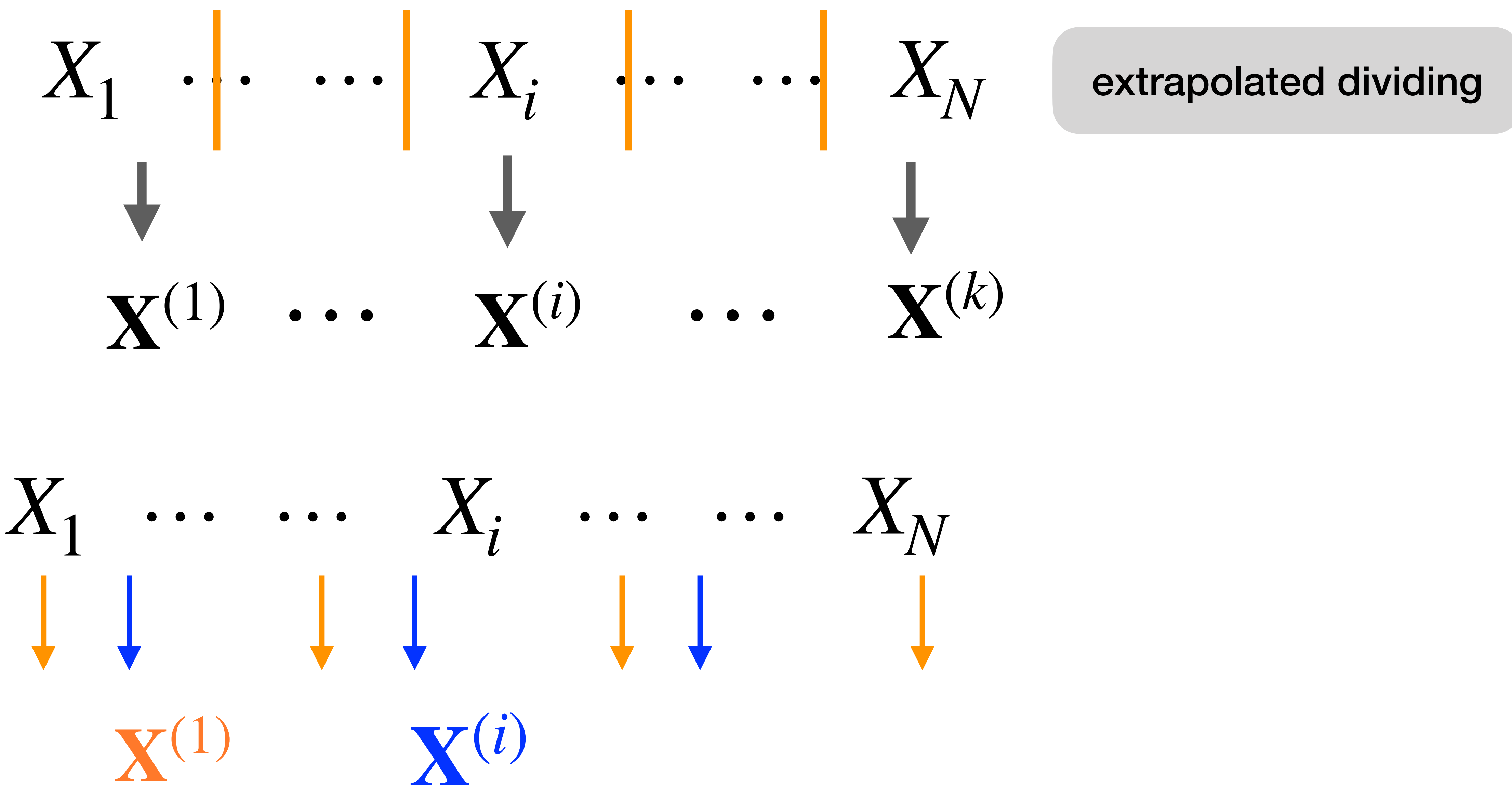
- Divided SPC



extrapolated dividing

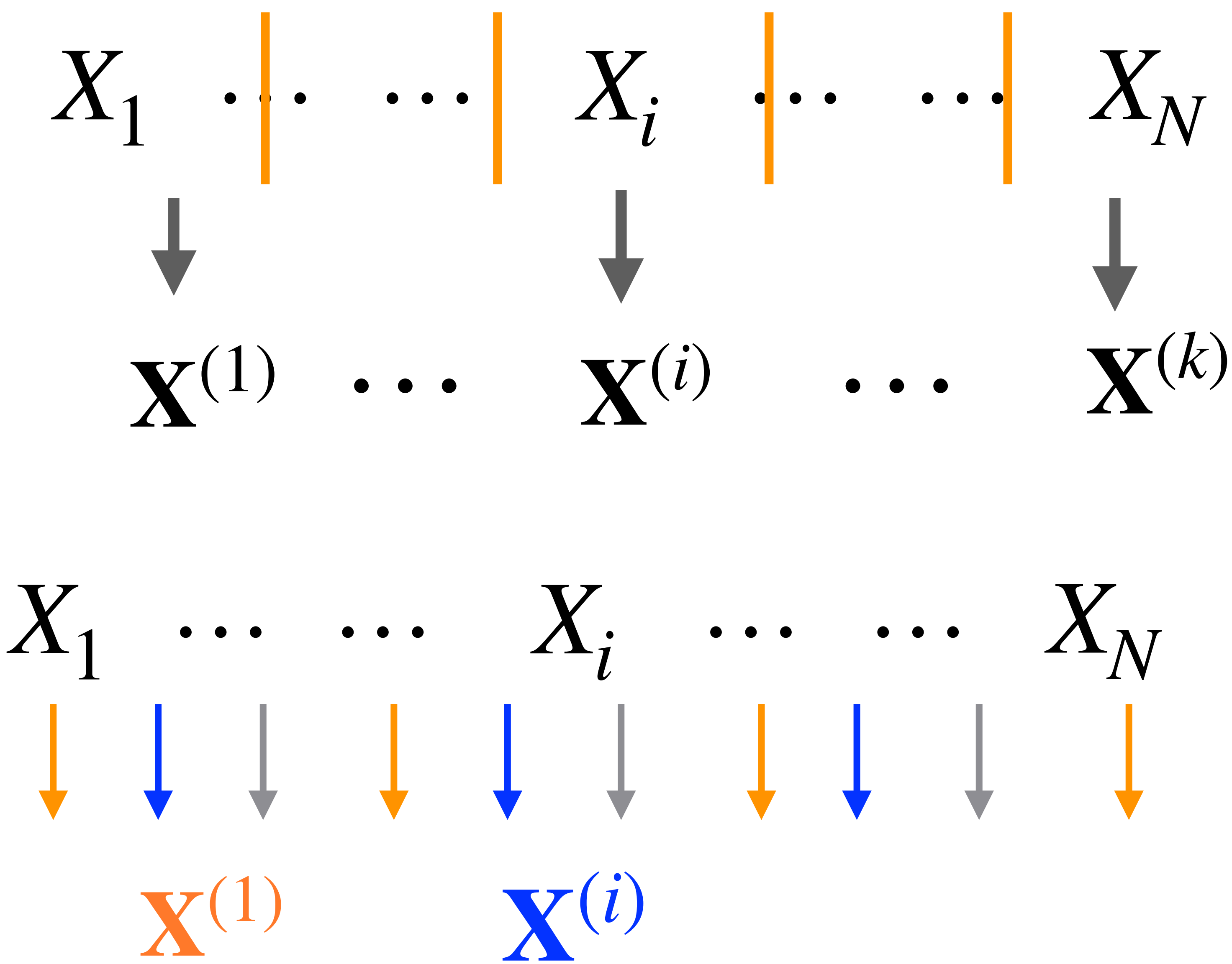
How to “split” in divided SPC with time-series data?

- Divided SPC



How to “split” in divided SPC with time-series data?

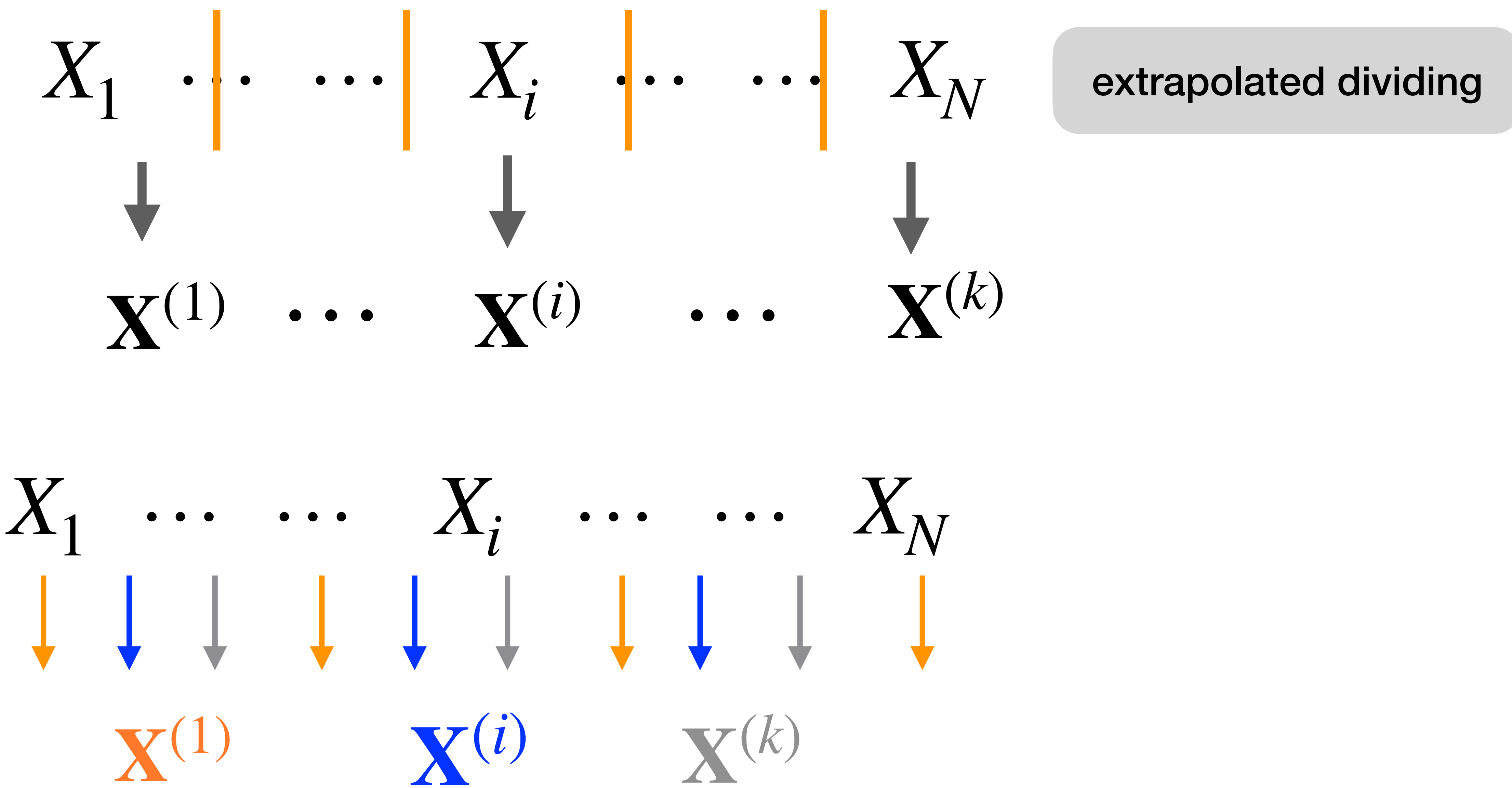
- Divided SPC



extrapolated dividing

How to “split” in divided SPC with time-series data?

- Divided SPC



How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

extrapolated dividing

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

How to “split” in divided SPC with time-series data?

- Divided SPC

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

extrapolated dividing

$$X_1 \quad \cdots \quad \cdots \quad X_i \quad \cdots \quad \cdots \quad X_N$$

$$\mathbf{X}^{(1)} \quad \cdots \quad \mathbf{X}^{(i)} \quad \cdots \quad \mathbf{X}^{(k)}$$

interpolated dividing