SAND2016-1345C

Towards a Model Selection Rule for Quantum State Tomography

Travis L Scholten

@Travis_Sch

Center for Quantum Information and Control, UNM
Center for Computing Research, Sandia National Labs

SQuInT Workshop

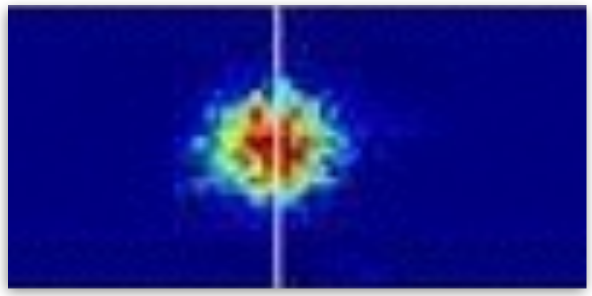
2016 February 19

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
Sandia National Laboratories

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Characterization of quantum devices gets hard as we scale them up.

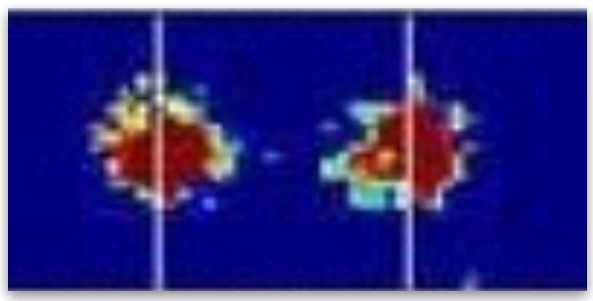
One qubit



$n = 3$ parameters

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}$$

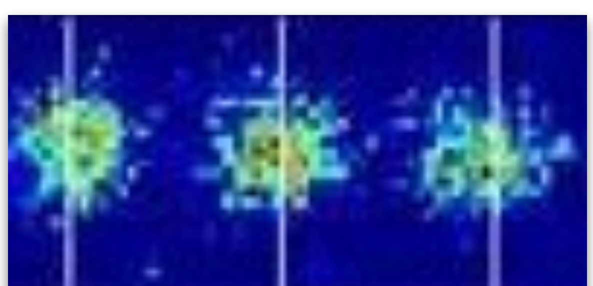
Two qubits



$n = 15$ parameters

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} & \rho_{03} \\ \rho_{10} & \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{20} & \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{30} & \rho_{31} & \rho_{32} & \rho_{33} \end{pmatrix}$$

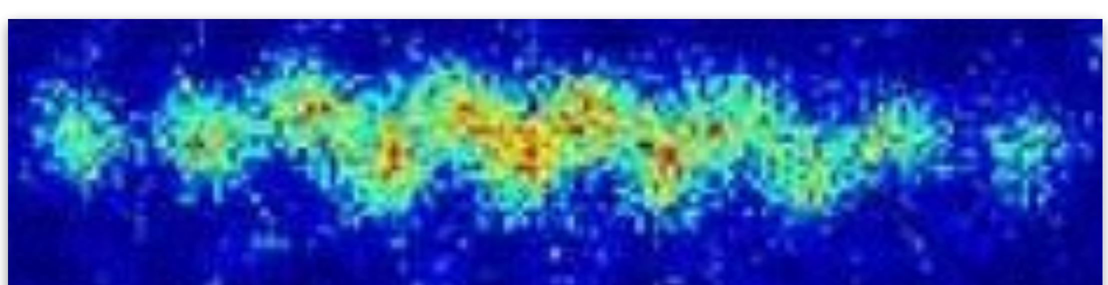
Three qubits



$n = 63$ parameters

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} & \rho_{03} & \rho_{04} & \rho_{05} & \rho_{06} & \rho_{07} \\ \rho_{10} & \rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} & \rho_{16} & \rho_{17} \\ \rho_{20} & \rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} & \rho_{25} & \rho_{26} & \rho_{27} \\ \rho_{30} & \rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} & \rho_{35} & \rho_{36} & \rho_{37} \\ \rho_{40} & \rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} & \rho_{45} & \rho_{46} & \rho_{47} \\ \rho_{50} & \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & \rho_{55} & \rho_{56} & \rho_{57} \\ \rho_{60} & \rho_{61} & \rho_{62} & \rho_{63} & \rho_{64} & \rho_{65} & \rho_{66} & \rho_{67} \\ \rho_{70} & \rho_{71} & \rho_{72} & \rho_{73} & \rho_{74} & \rho_{75} & \rho_{76} & \rho_{77} \end{pmatrix}$$

N qubits



$n = 4^N - 1$ parameters

Model selection can make tomography more tractable.

Model (for tomography) = sets of density matrices

Model selection = find best model

$$\hat{\rho} = \left(\quad \right)$$

What *sets* of density matrices will we consider??

What does "best" mean?

Tomographers *have been doing model selection* all along.

Model (for tomography) = sets of density matrices

Trivial way:

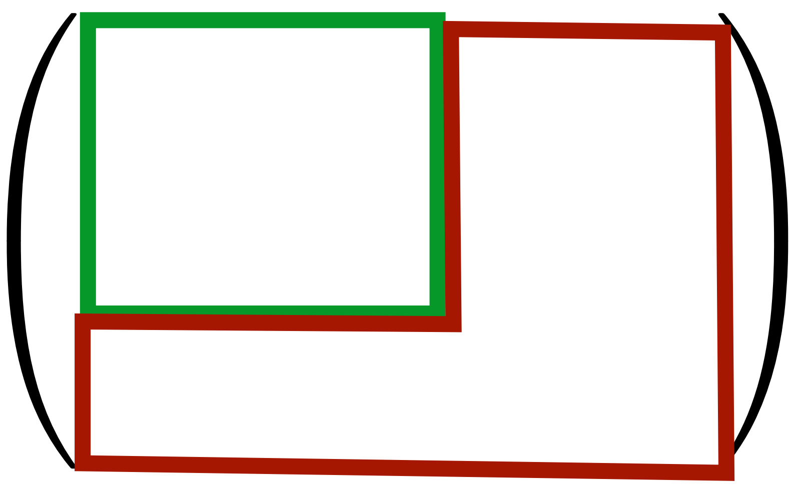
$$\hat{\rho} = \left(\boxed{} \right)$$

Pick Hilbert space by fiat
("Of course it's a qubit!")

Tomographers *have been doing model selection* all along.

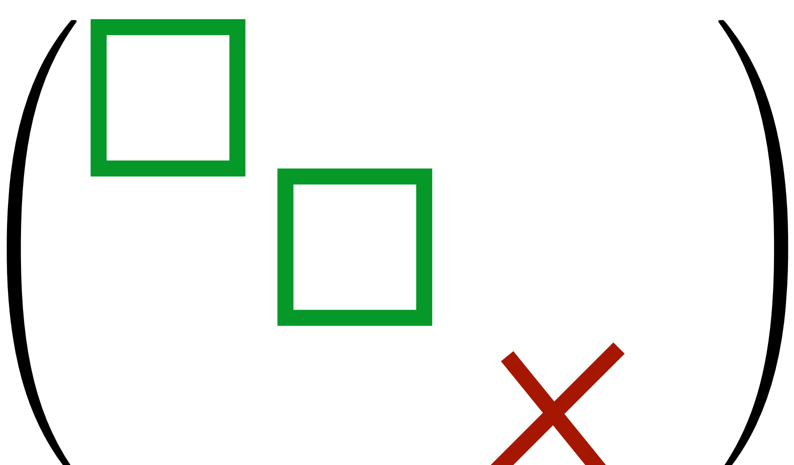
Model (for tomography) = sets of density matrices

Nontrival ways:

$\hat{\rho} =$ 

Restrict estimate to a subspace

$$\hat{\rho} = \sum_{j,k=0}^{N-1} \rho_{jk} |j\rangle\langle k|$$

$\hat{\rho} =$ 

Restrict rank of estimate

$$\hat{\rho} = \sum_{j=0}^{r-1} \lambda_j |\lambda_j\rangle\langle \lambda_j|$$

Finding the best model *seemed* straightforward.

Applying Model Selection to Quantum State Tomography: Choosing Hilbert Space Dimension

Travis L Scholten

APS March Meeting
5 March 2015

Tomography is hard

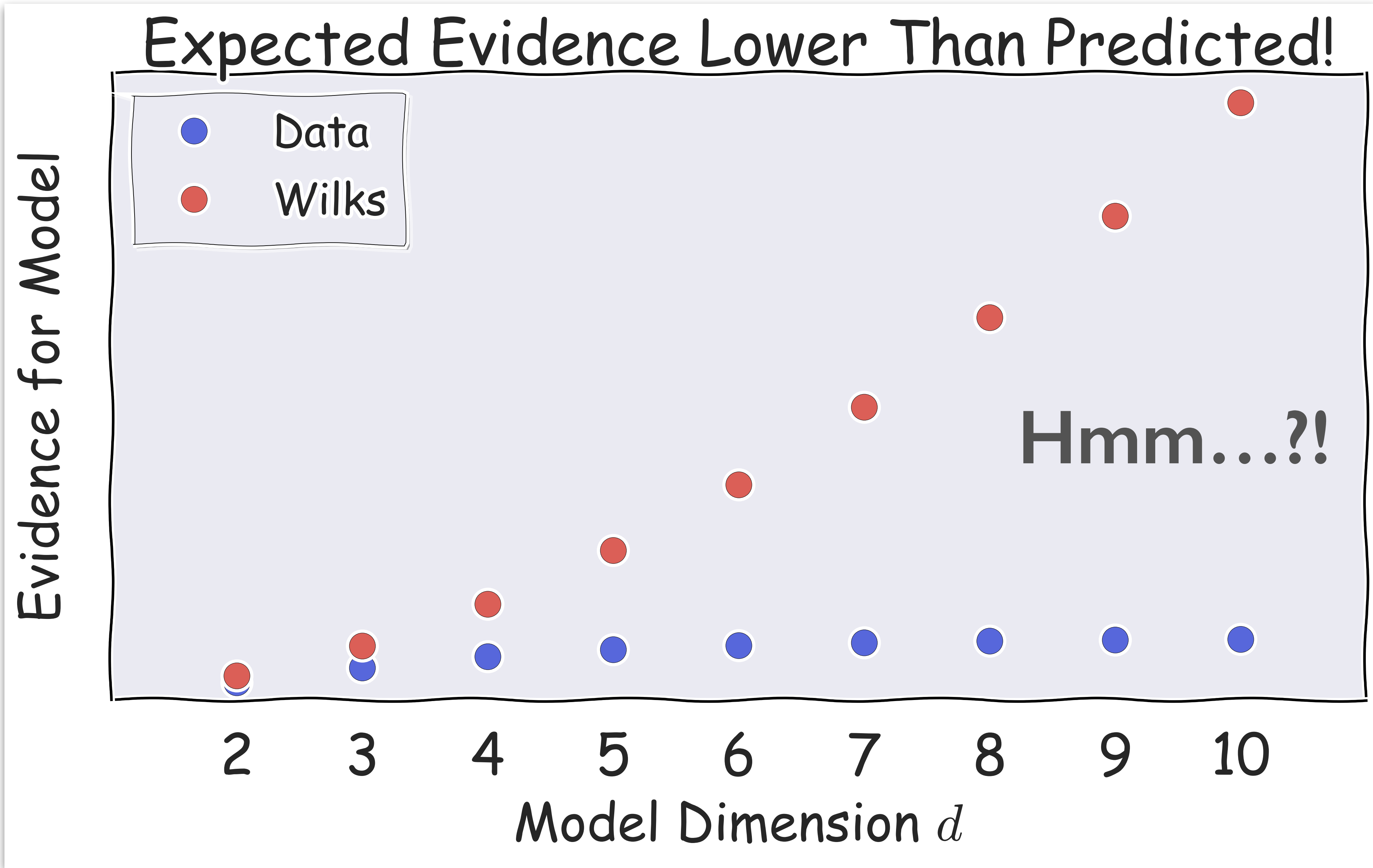
Doing so in infinite dimensional Hilbert space is harder

Let's make it easier…

"Just use loglikelihood ratios and the Wilks Theorem"

"Or information criteria?"

A key tool used for model selection — the *Wilks Theorem* — failed dramatically!



Wilks Theorem: predicts expected evidence when model is not actually better

How did this happen?!

The foundations of model selection are well-studied for classical inference…

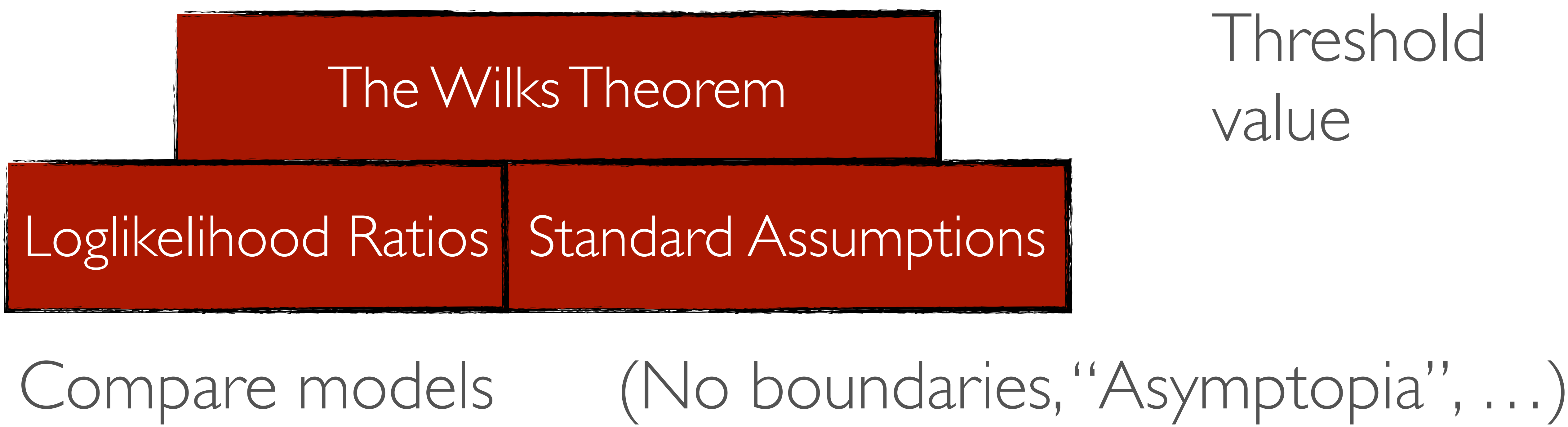
Standard Assumptions

(No boundaries, "Asymptopia", …)

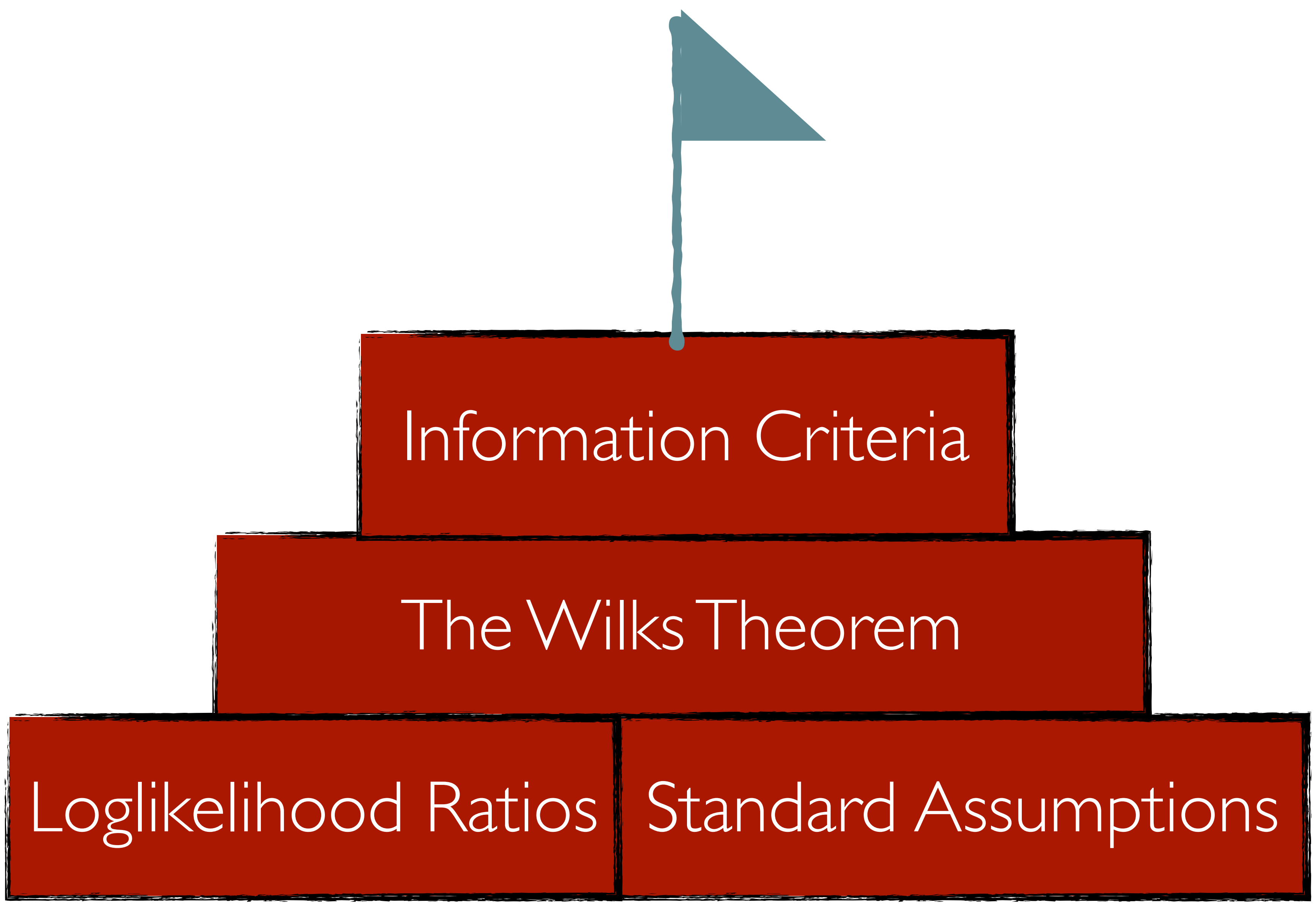
The foundations of model selection are well-studied for classical inference…

Loglikelihood Ratios	Standard Assumptions
Compare models	(No boundaries, "Asymptopia", …)

The foundations of model selection are well-studied for classical inference…



The foundations of model selection are well-studied for classical inference…



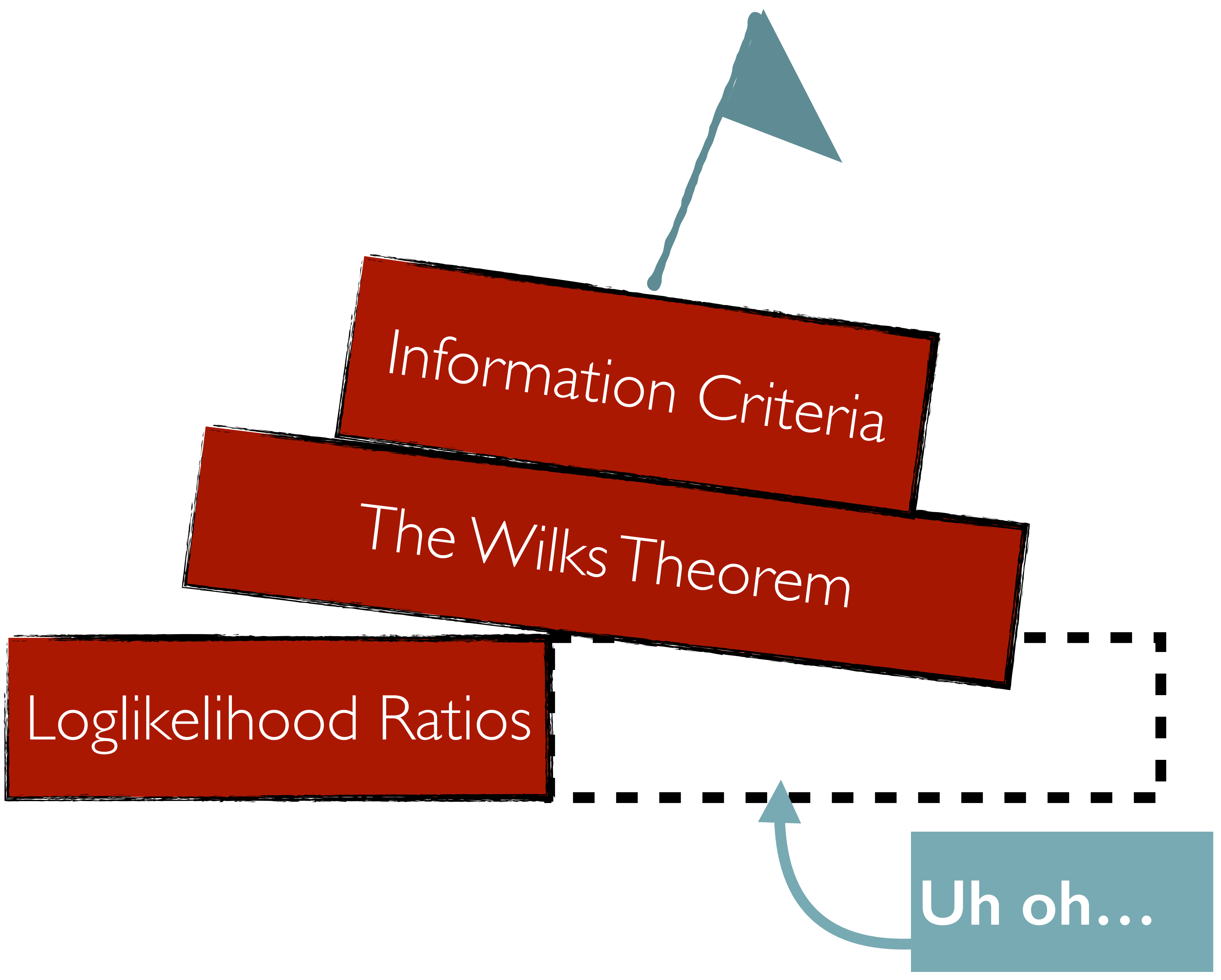
Decision rule

Threshold value

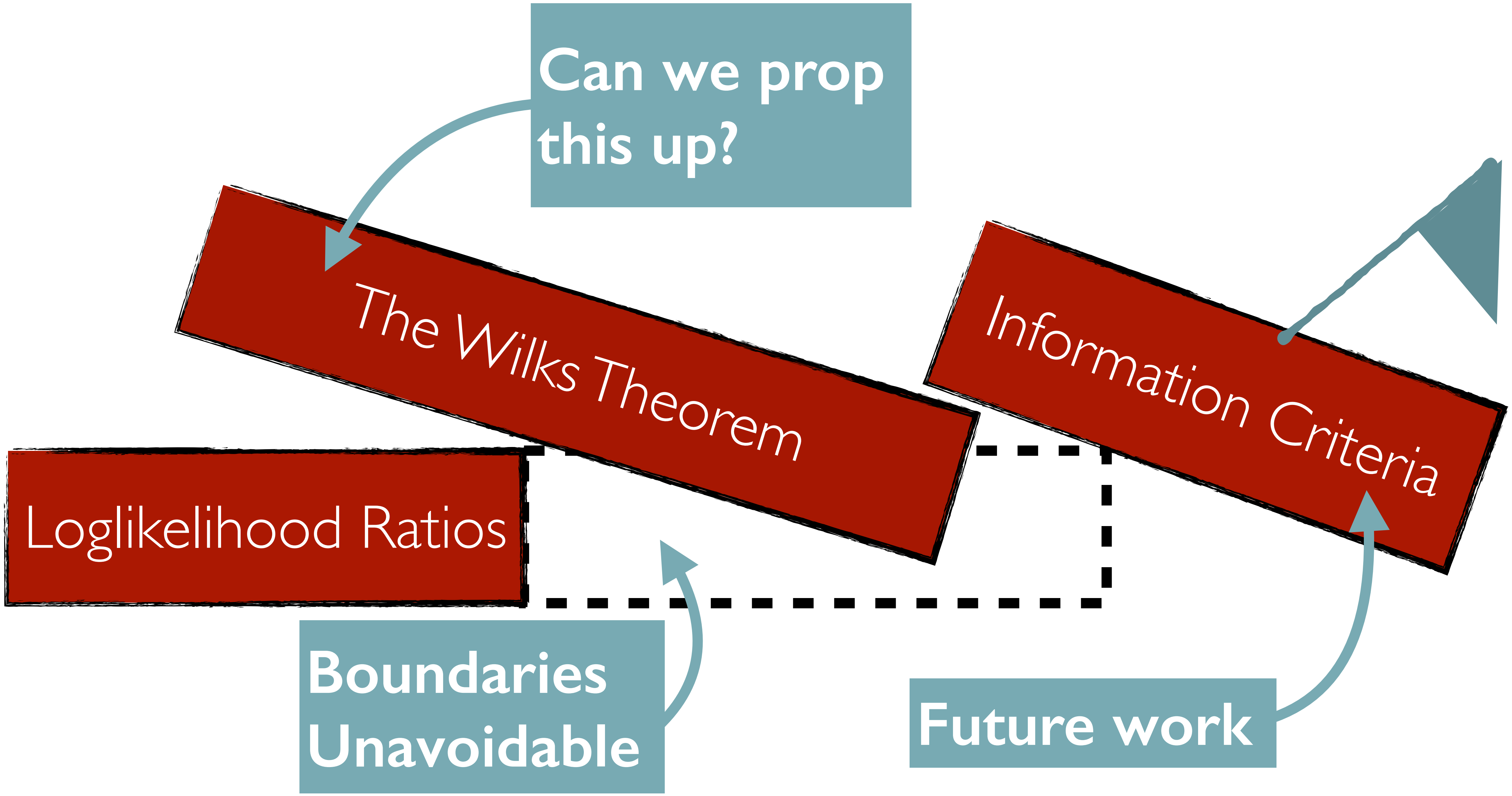
Compare models

(No boundaries, “Asymptopia”, …)

…but start to break down for tomography of quantum systems…



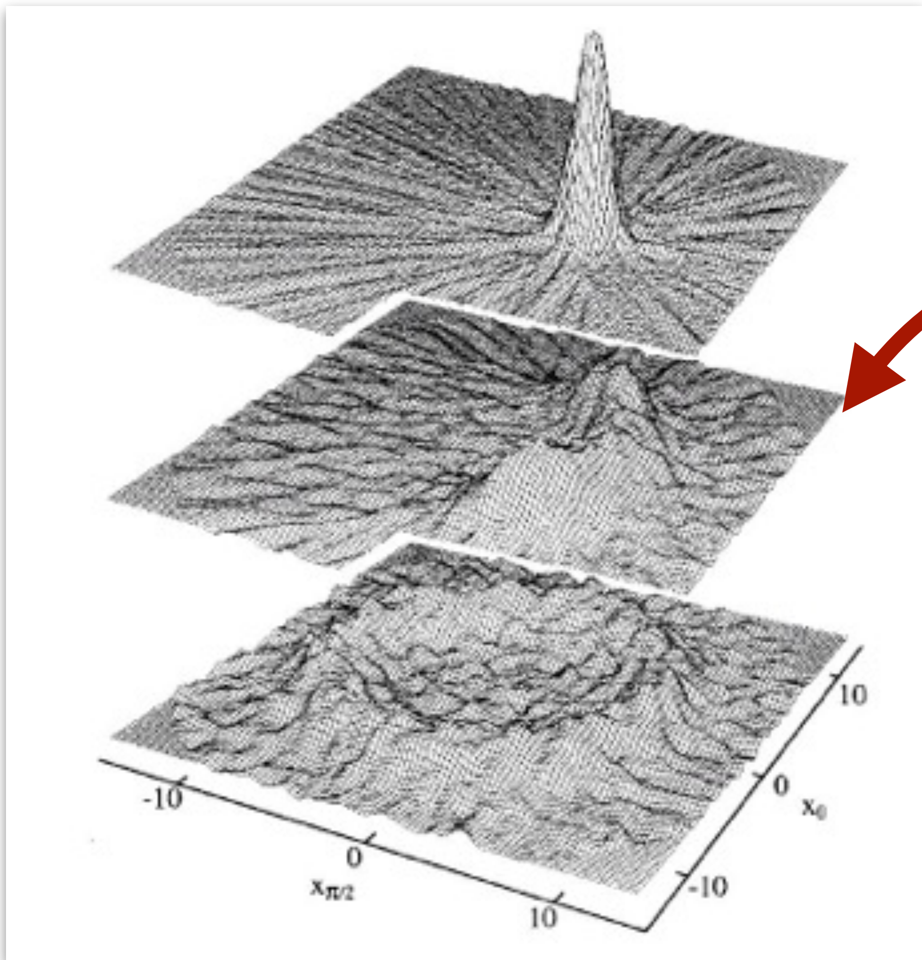
…because quantum state spaces have boundaries!



How did this happen?!

We studied selecting
Hilbert space dimension
for **CV (optical) tomography**
to see why.

Continuous-variable (CV) systems are a nice sandbox for model selection.



$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & \cdots \\ \rho_{10} & \rho_{11} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Finite data…infinite parameters!

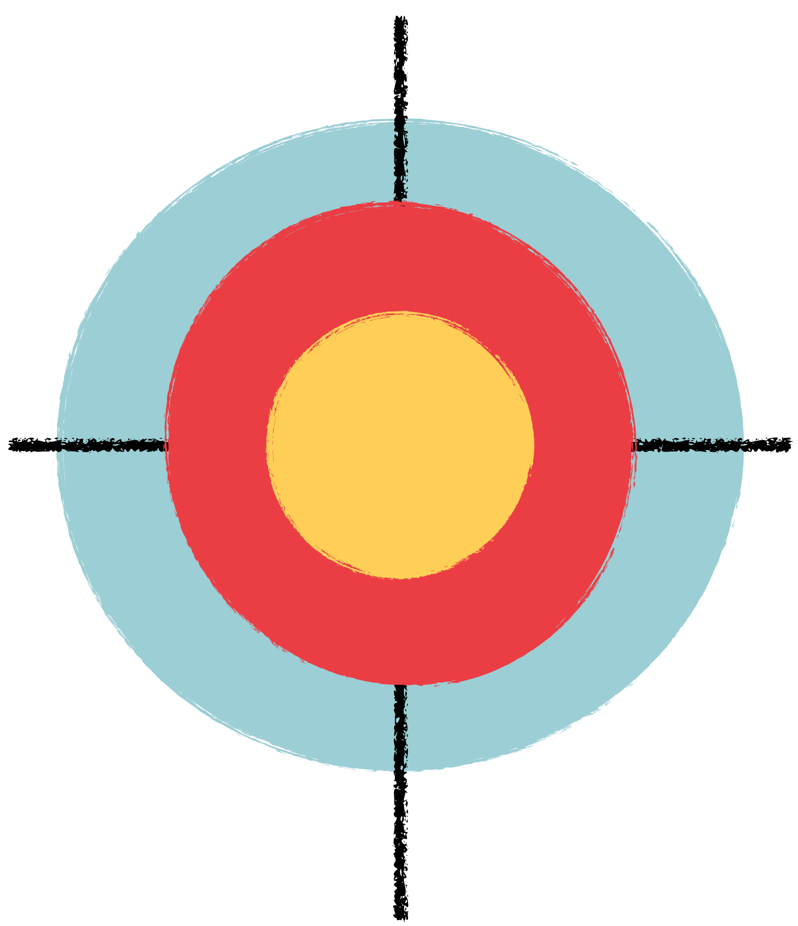
How do we find a small,
yet *good* model?

We used *nested subspace* models.

Model = sets of density matrices
spanned by number states

$$\mathcal{M}_d = \{\rho \mid \rho \in \mathcal{B}(\mathcal{H}_d),\ \mathcal{H}_d = \mathrm{Span}(|0\rangle, \cdots, |d-1\rangle)\}$$

Based on assuming states have low energy



$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} & \cdots \\ \rho_{10} & \rho_{11} & \rho_{21} & \cdots \\ \rho_{20} & \rho_{21} & \rho_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

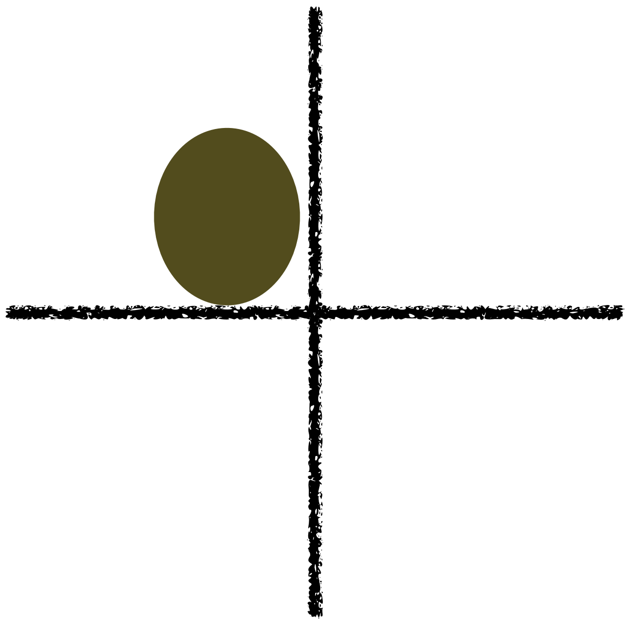
We studied *heterodyne tomography*.

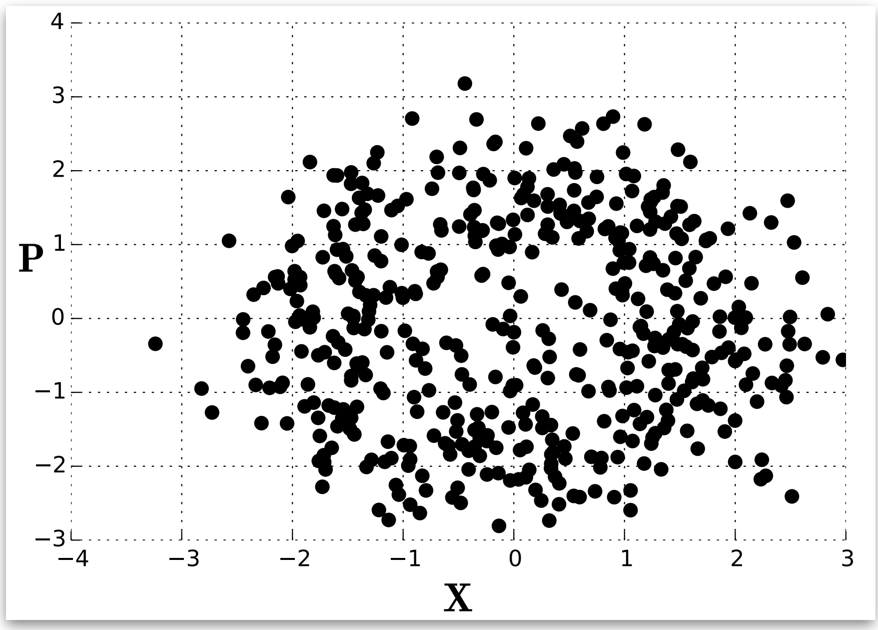
$$\rho_0 \to \{\alpha_1, \alpha_2, \cdots\} \to \hat{\rho} \in \mathcal{M}_d$$

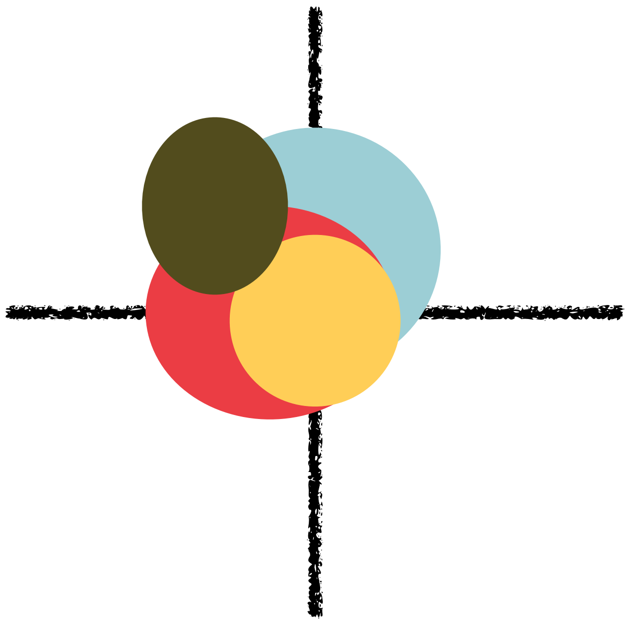
Pick true state (arbitrary)

Simulate POVM (rejection sampling)

Find ML estimate (within model)







Loglikelihood ratio statistics give us *evidence* for choosing between models.

Compare model to truth with *loglikelihood ratios*

$$\lambda(\rho_0, \mathcal{M}_d) = -2\log\left(\frac{\mathcal{L}(\rho_0)}{\max\limits_{\rho\in\mathcal{M}_d}\mathcal{L}(\rho)}\right)$$

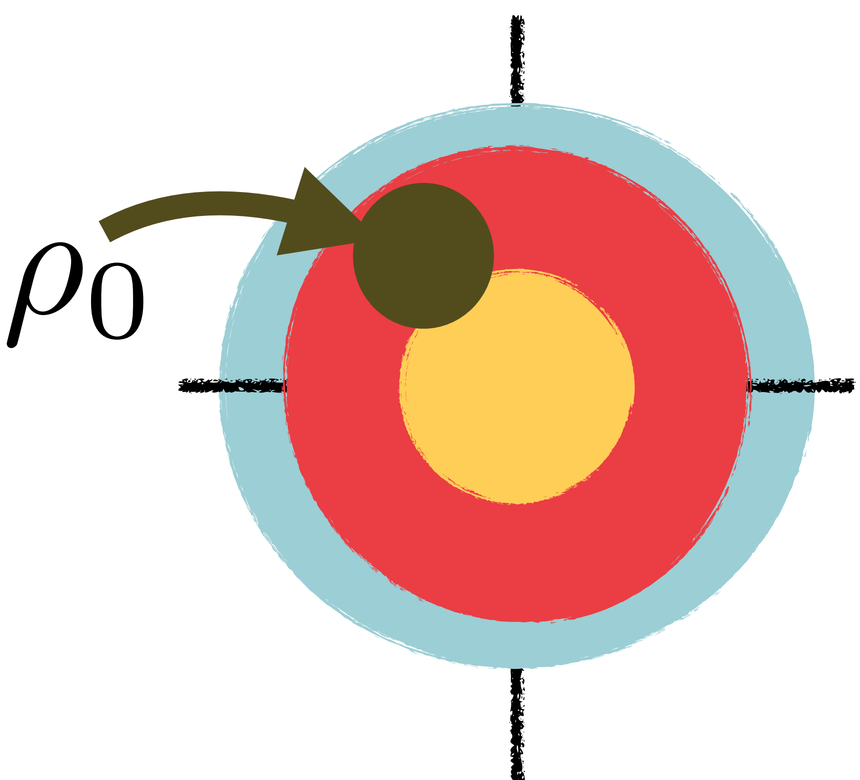
Loglikelihood ratio statistics give us *evidence* for choosing between models.

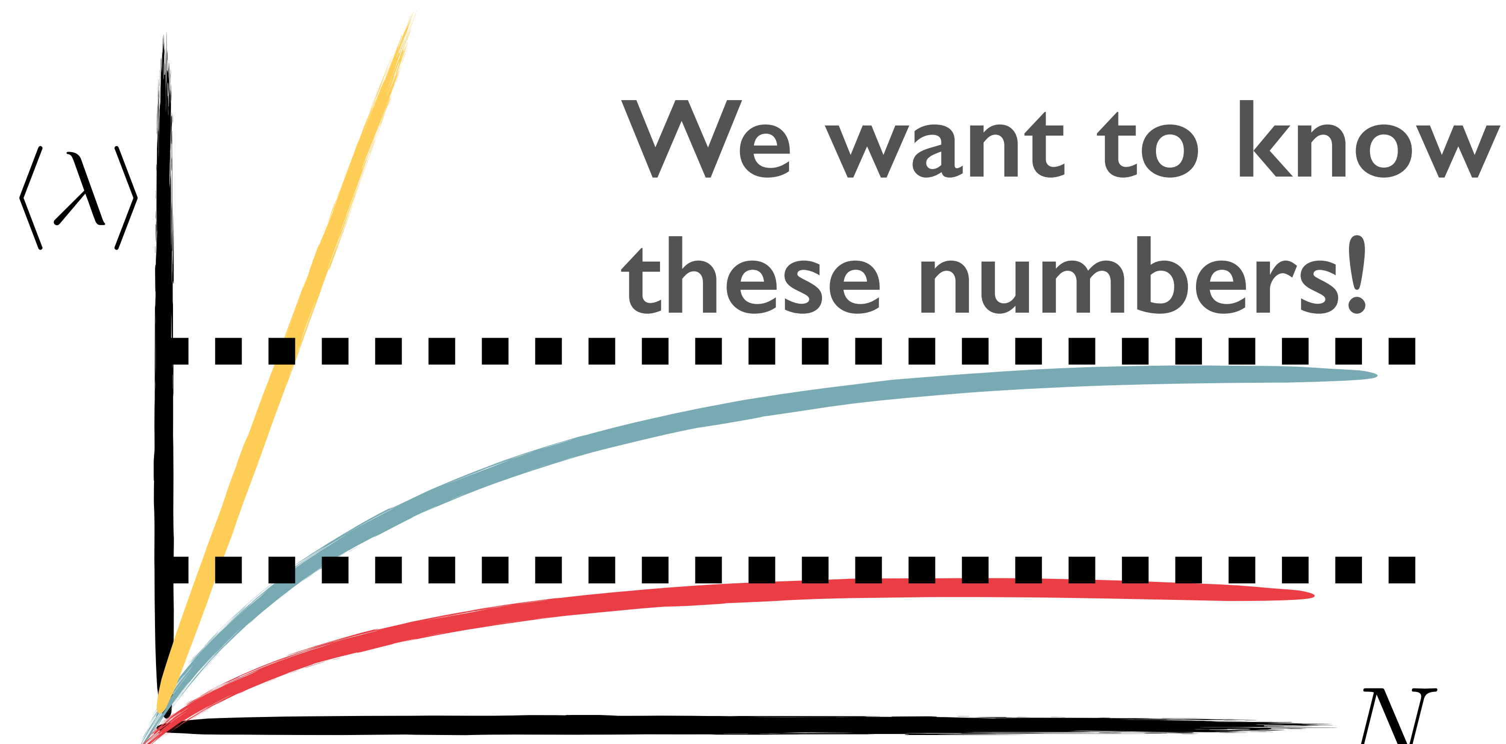
Compare model to truth with *loglikelihood ratios*

$$\lambda(\rho_0, \mathcal{M}_d) = -2 \log \left(\frac{\mathcal{L}(\rho_0)}{\max\limits_{\rho \in \mathcal{M}_d} \mathcal{L}(\rho)} \right)$$

The Wilks Theorem:

$$\rho_0 \in \mathcal{M}_d \implies \lambda \sim \chi^2_{d^2}$$

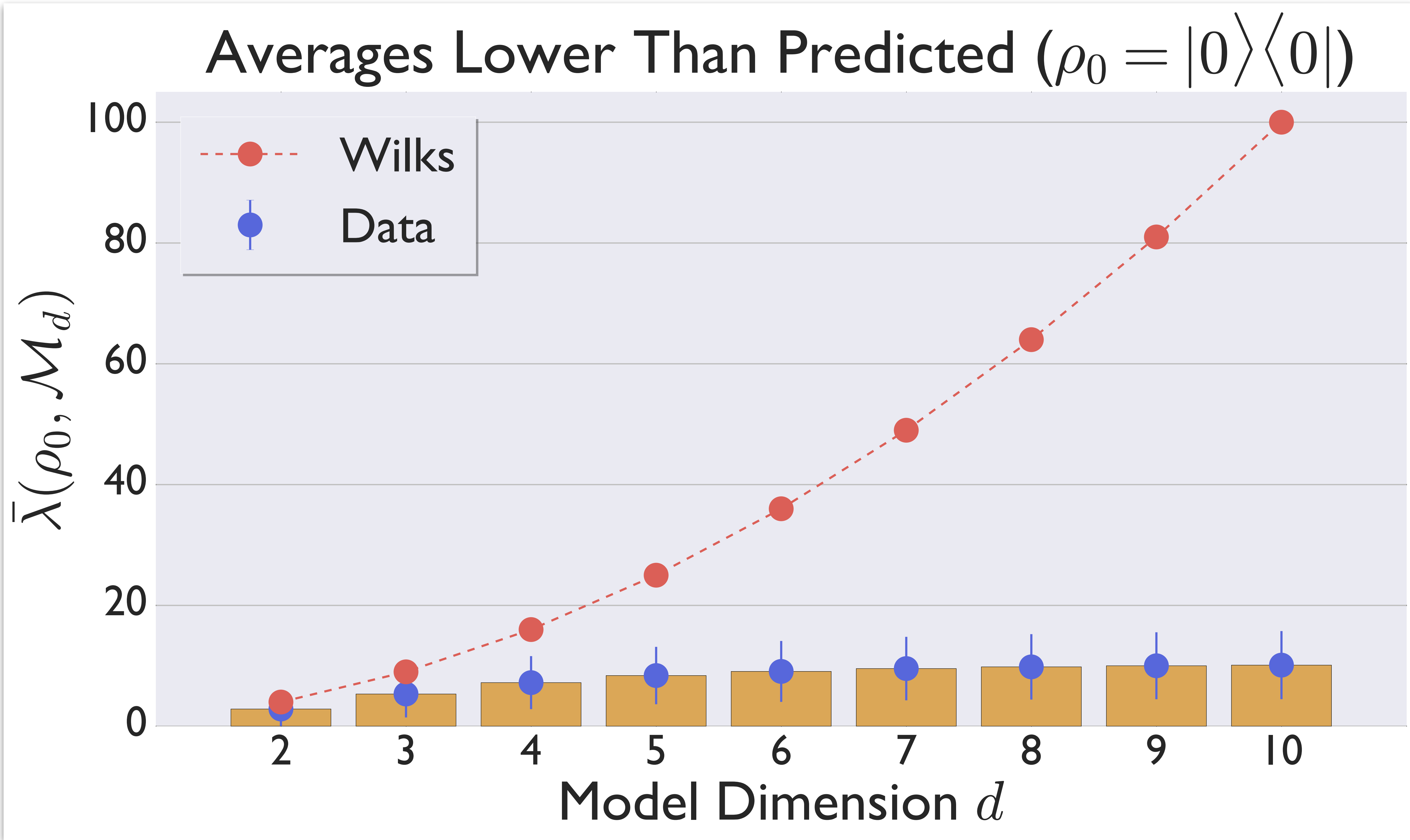




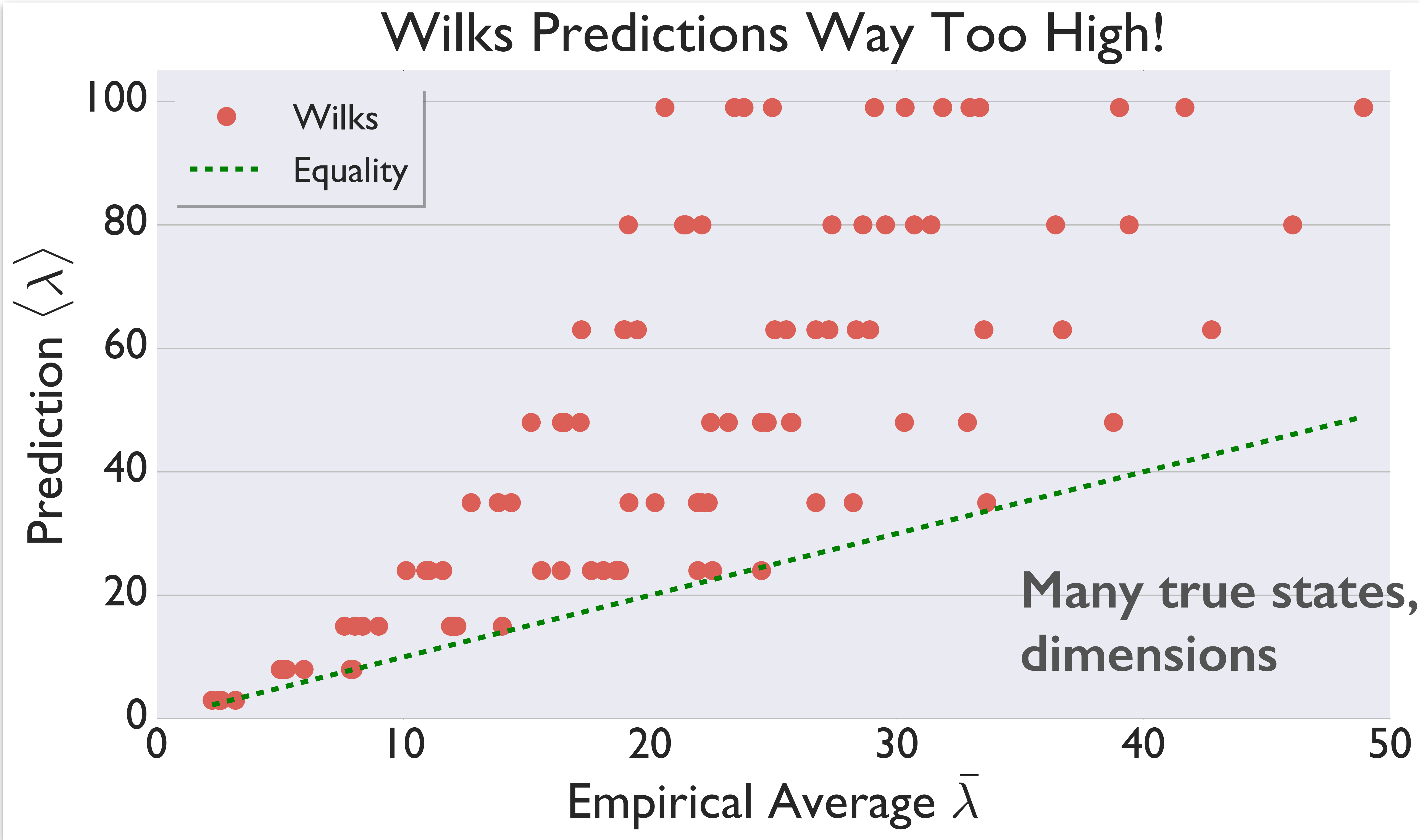
I thought you said **Wilks does not work**?

Yes. **Let's see the evidence**.

The Wilks Theorem incorrectly predicts the behavior in CV tomography.



The Wilks Theorem incorrectly predicts the behavior in CV tomography.



Wilks Theorem gives wrong behavior
for the loglikelihood ratio statistic.

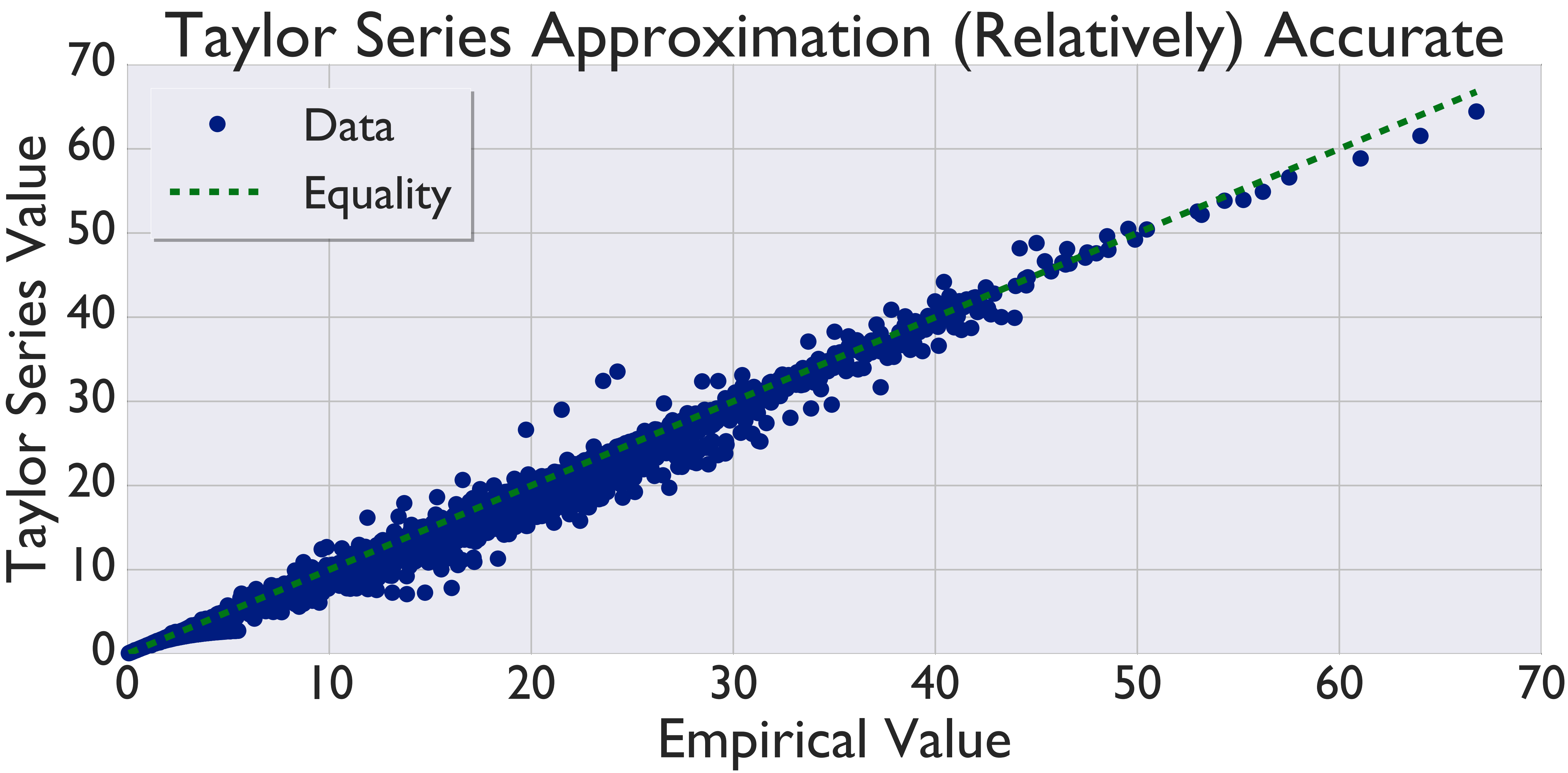
Wilks Theorem should not be used*
for choosing Hilbert space dimension.

Let's fix this.

*Nor AIC, etc, because they rely on the Wilks Theorem!

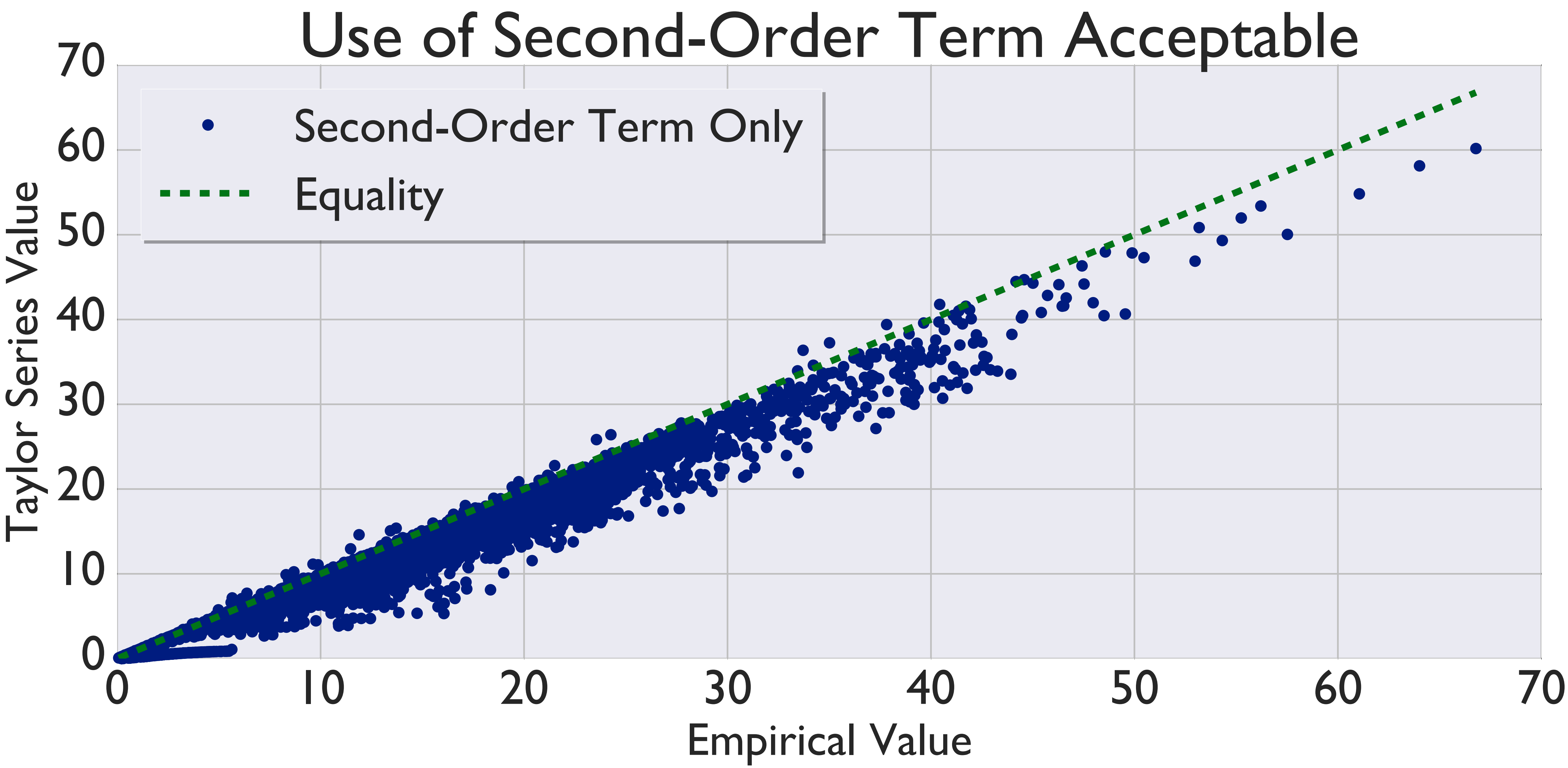
Wilks relies on a Taylor series. Let's start our investigation there.

$$\lambda(\rho_0, \mathcal{M}_d) \approx \nabla\lambda \cdot (\rho_0 - \hat{\rho}_d) + \frac{1}{2}(\rho_0 - \hat{\rho}_d)\frac{\partial^2 \lambda}{\partial^2 \rho}(\rho_0 - \hat{\rho}_d)$$



We ignore the first-order term and get an accurate enough approximation.

$$\lambda(\rho_0, \mathcal{M}_d) \approx \frac{1}{2}(\rho_0 - \hat{\rho}_d)\frac{\partial^2 \lambda}{\partial \rho^2}(\rho_0 - \hat{\rho}_d)$$



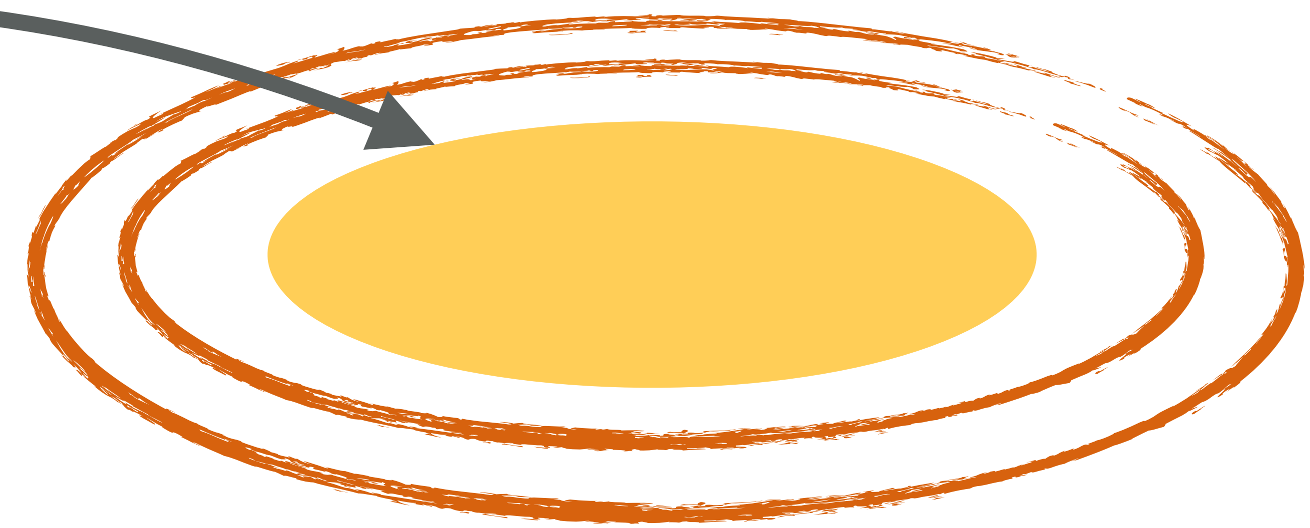
Let's cast this equation in a more sensible form.

$$\begin{aligned}\lambda(\rho_0, \mathcal{M}_d) &\approx \frac{1}{2}(\rho_0 - \hat{\rho}_d)\frac{\partial^2 \lambda}{\partial \rho^2}(\rho_0 - \hat{\rho}_d) \\ &\approx \langle\langle \rho_0 - \hat{\rho}_d | H | \rho_0 - \hat{\rho}_d \rangle\rangle \quad \text{(Superoperators!)} \\ &\approx \mathrm{Tr}(HF)\end{aligned}$$

Statistic depends on *observed information* (H) and *fluctuations* of the estimate (F).

The second-order term shows how *fluctuations* and *information* contribute.

$$\langle\lambda\rangle \approx \mathrm{Tr}(\langle HF\rangle)$$
$$\approx \mathrm{Tr}(\langle H\rangle\langle F\rangle)$$
$$\approx \mathrm{Tr}(\langle H\rangle\langle H\rangle^{-1})$$
$$\approx d^2$$



Wilks: Information and fluctuations align
& saturate (classical) Cramer-Rao bound

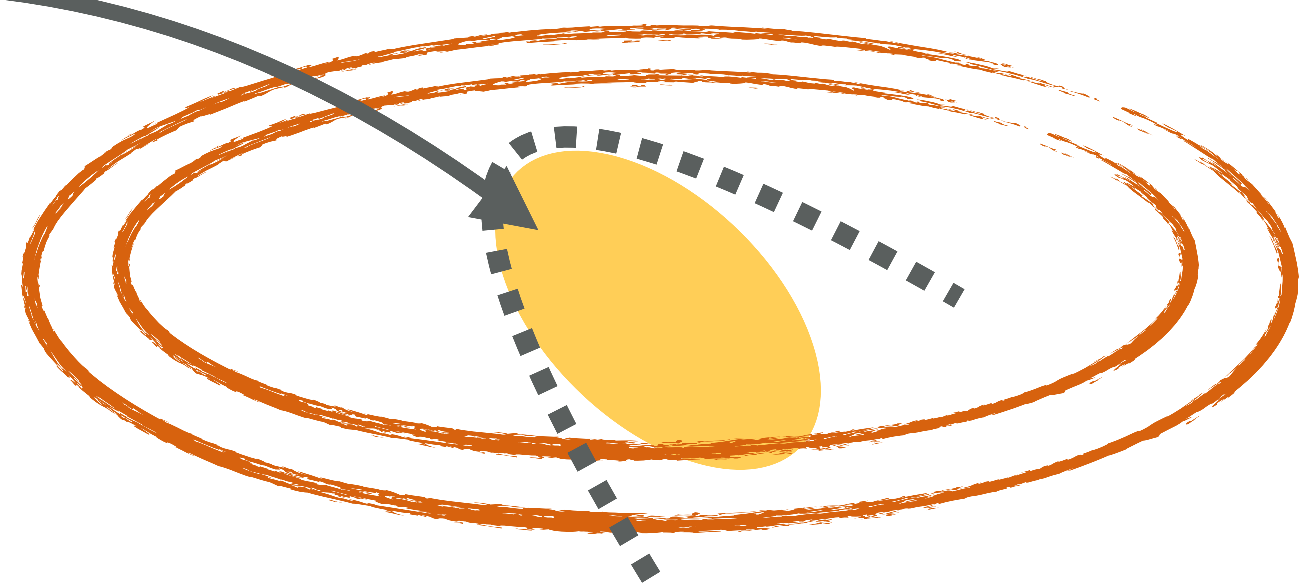
Where does this go wrong in tomography?

State-space boundaries distort fluctuations.

$$\langle \lambda \rangle \approx \mathrm{Tr}(\langle HF \rangle)$$

$$\approx \mathrm{Tr}(\langle H \rangle \langle F \rangle)$$

$$\approx \ ???$$



Reality: Information and fluctuations **do not** align
& **do not** saturate (classical) Cramer-Rao bound

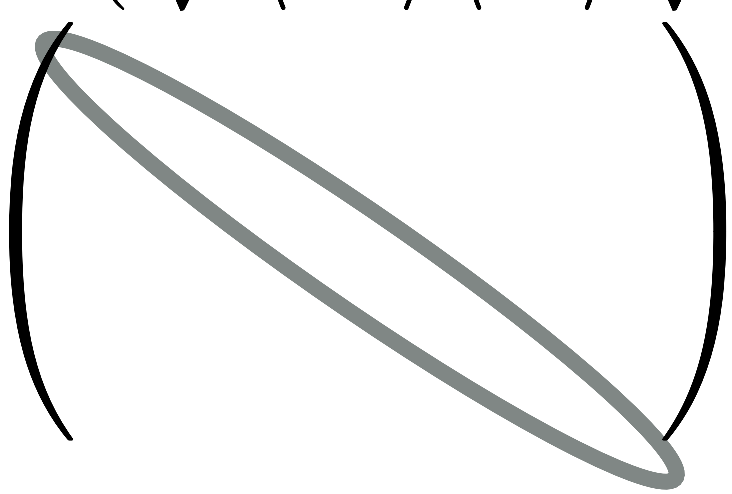
We have to respect state space boundaries!

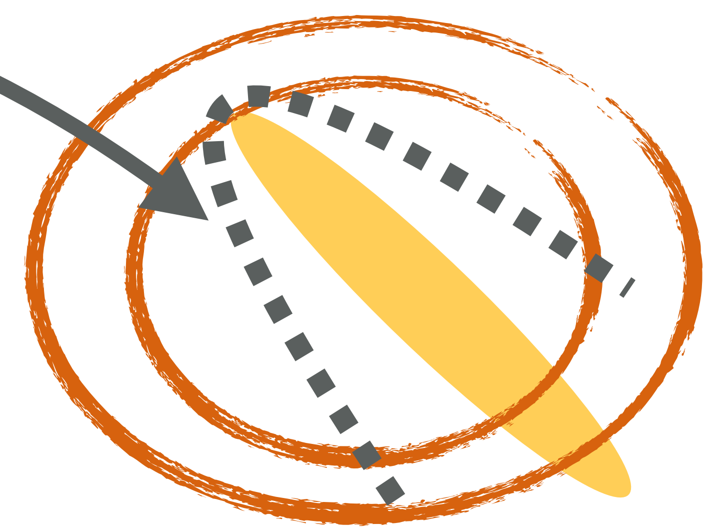
Let's rescale state space so the Fisher information is isotropic.

$$\langle \lambda \rangle \approx \mathrm{Tr}(\langle HF \rangle)$$

$$\approx \mathrm{Tr}(\langle H \rangle \langle F \rangle)$$

$$= \mathrm{Tr}(\sqrt{\langle H \rangle} \langle F \rangle \sqrt{\langle H \rangle})$$

$\approx$ 



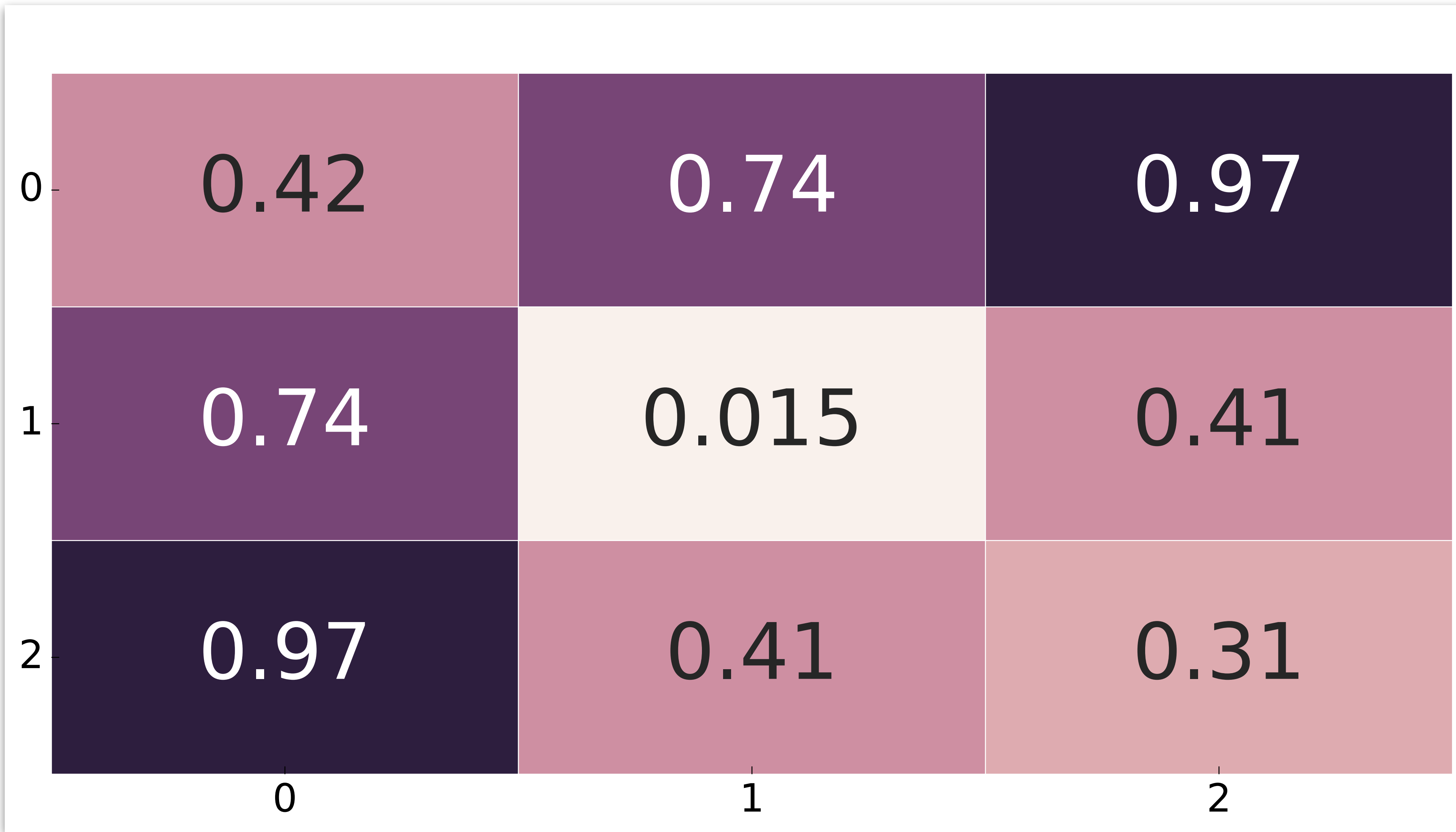
What are these numbers?
Can we make sense of them?

Different matrix units have different contributions.

$$\rho_0 = |0\rangle\langle 0|$$
$$d = 2$$

Wilks: Each parameter contributes one unit!

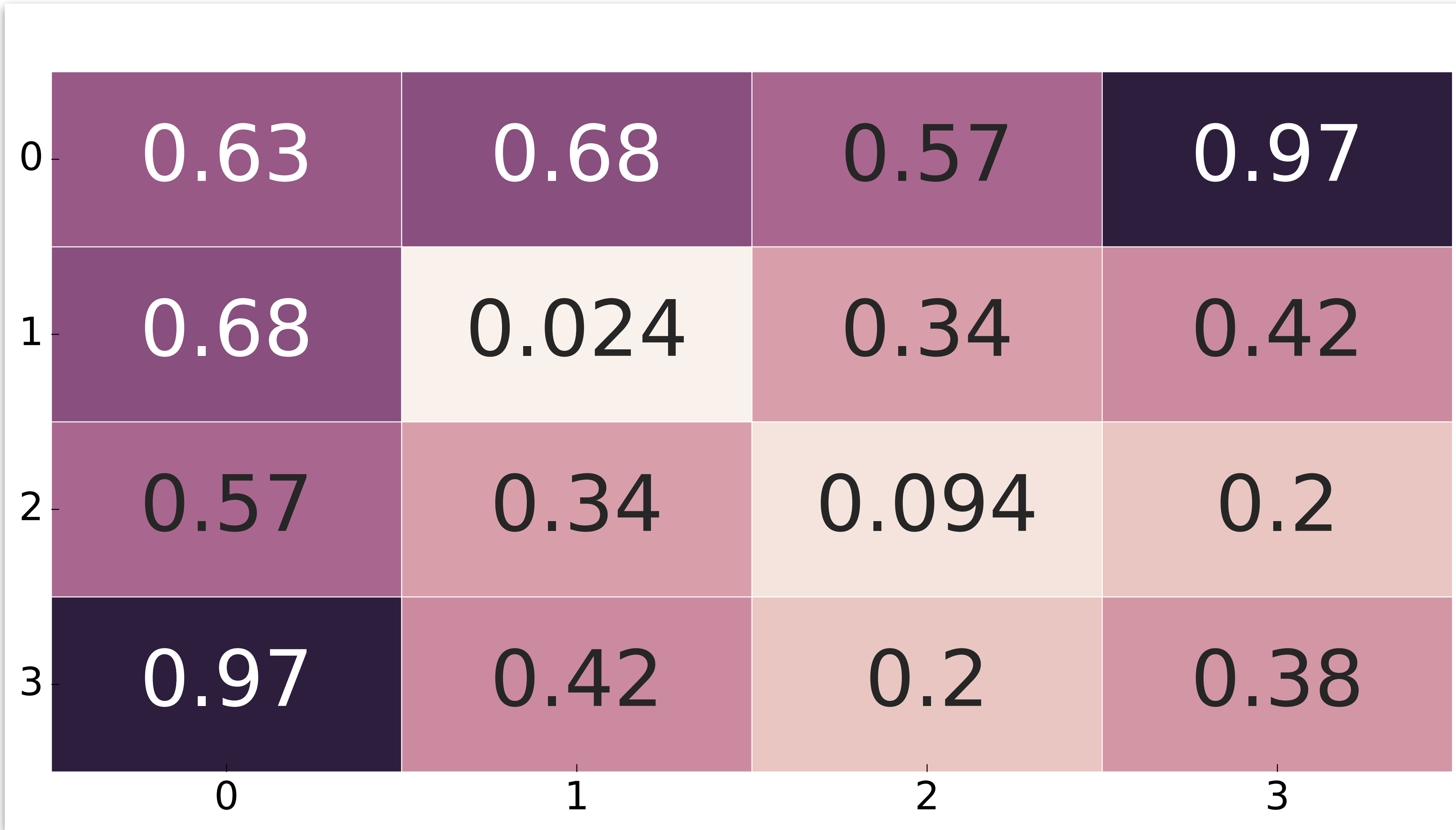
Different matrix units have different contributions.



$$\rho_0 = |0\rangle\langle 0|$$
$$d = 3$$

Wilks: Each parameter contributes one unit!

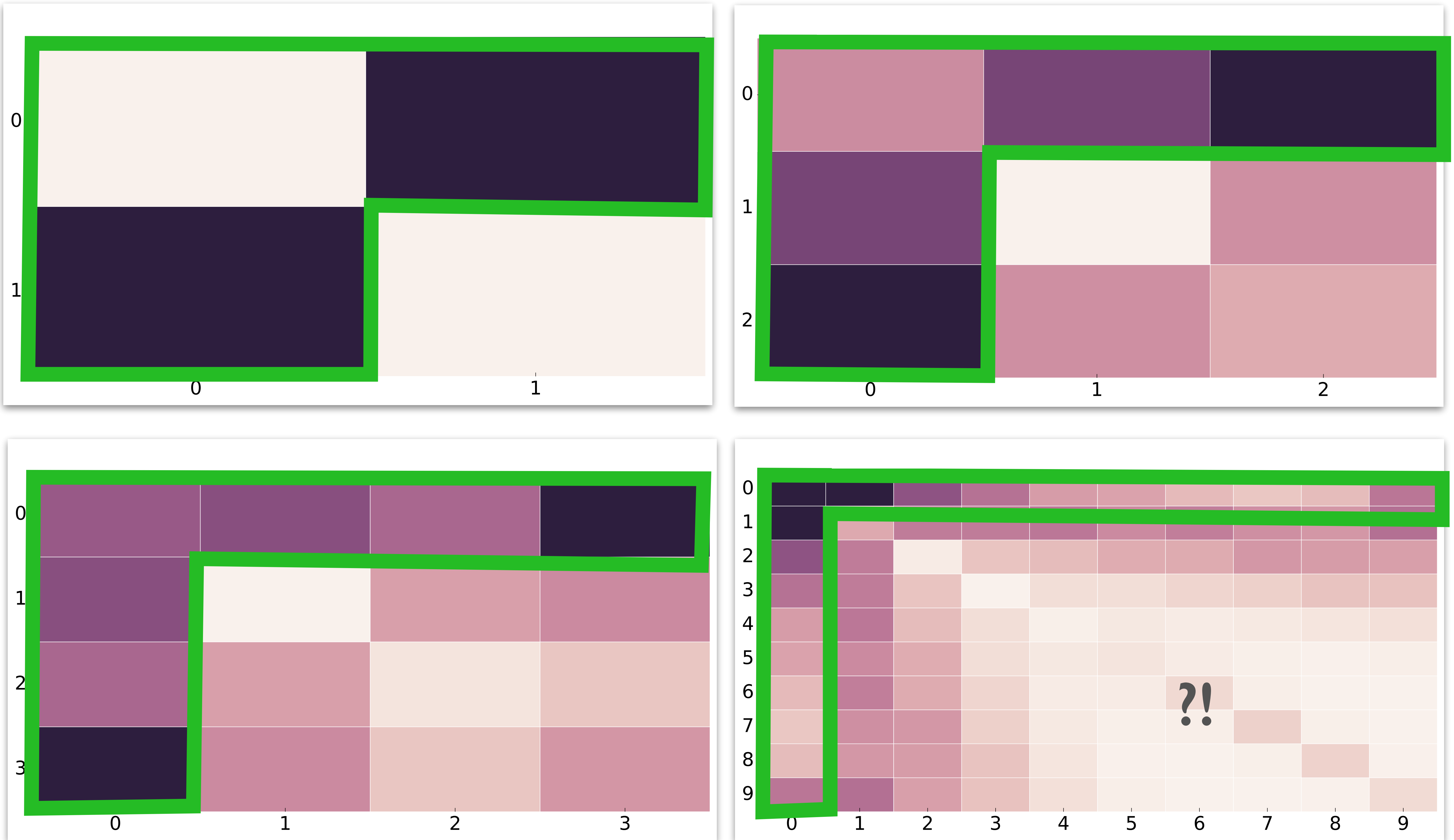
Different matrix units have different contributions.



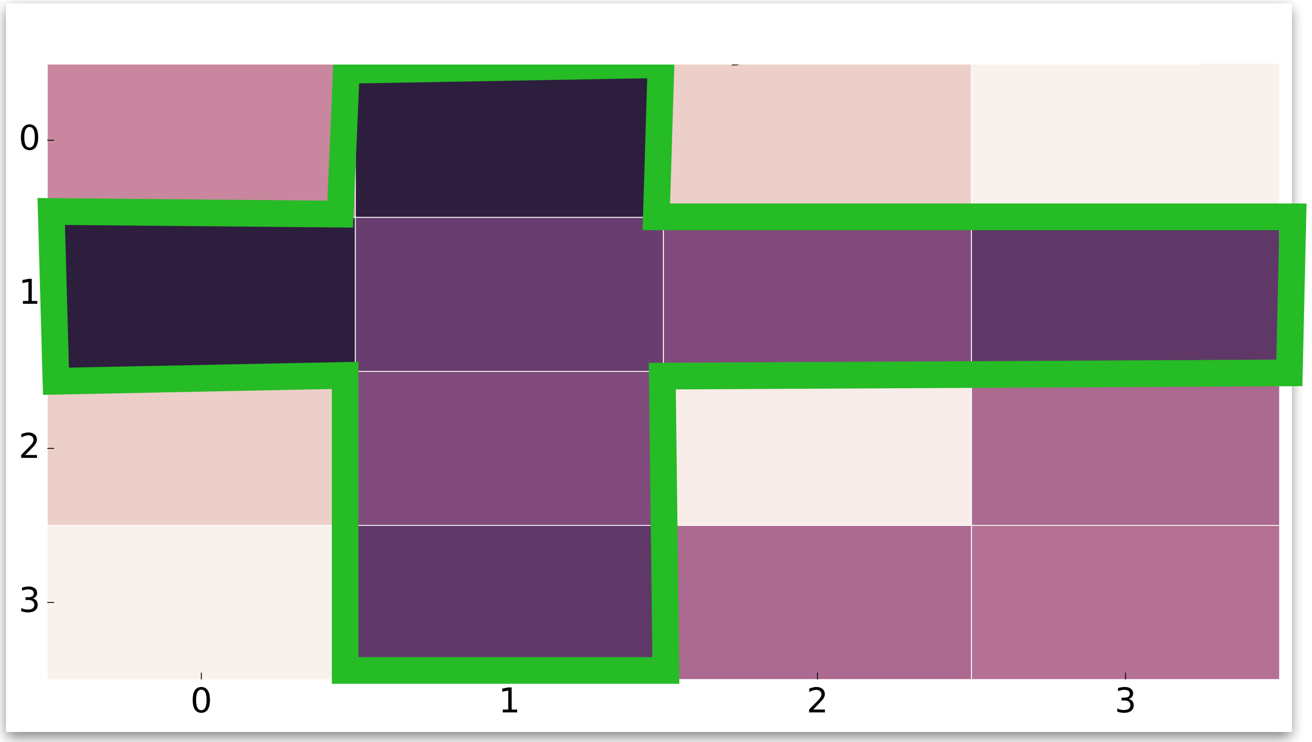
$$\rho_0 = |0\rangle\langle 0|$$
$$d = 4$$

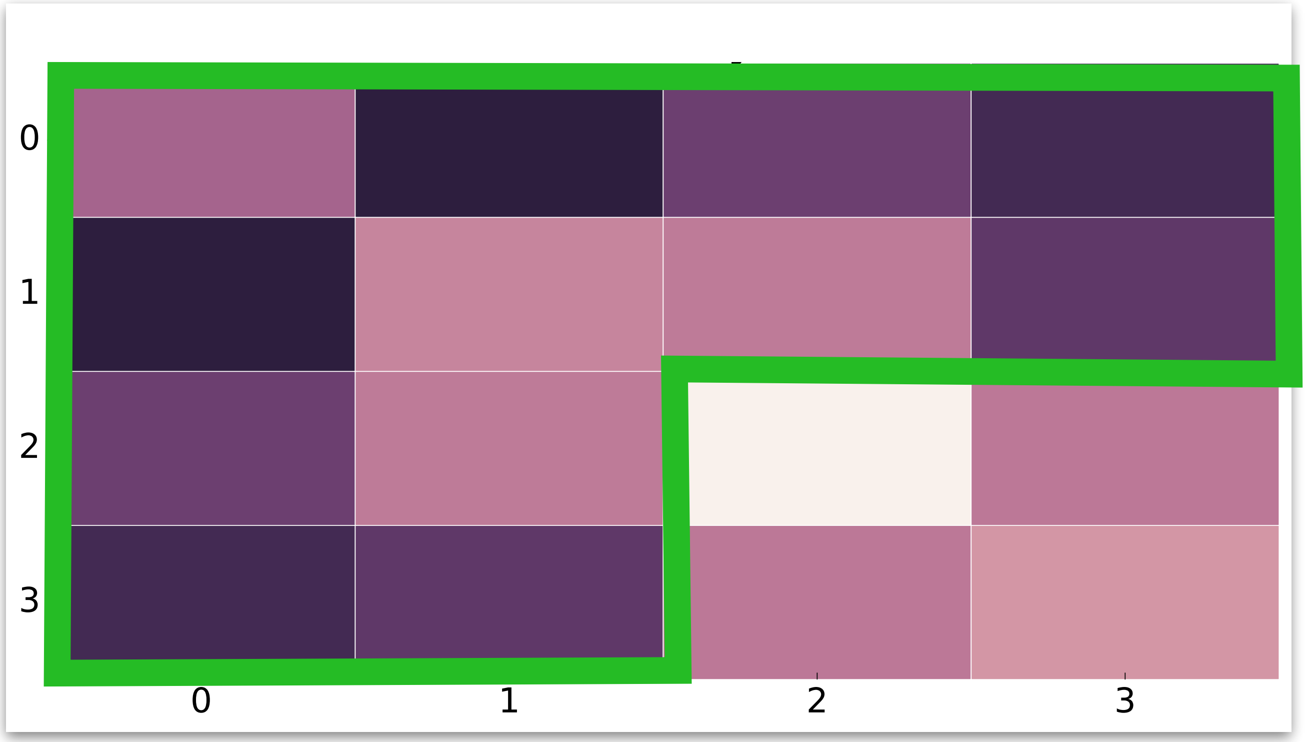
See a pattern?

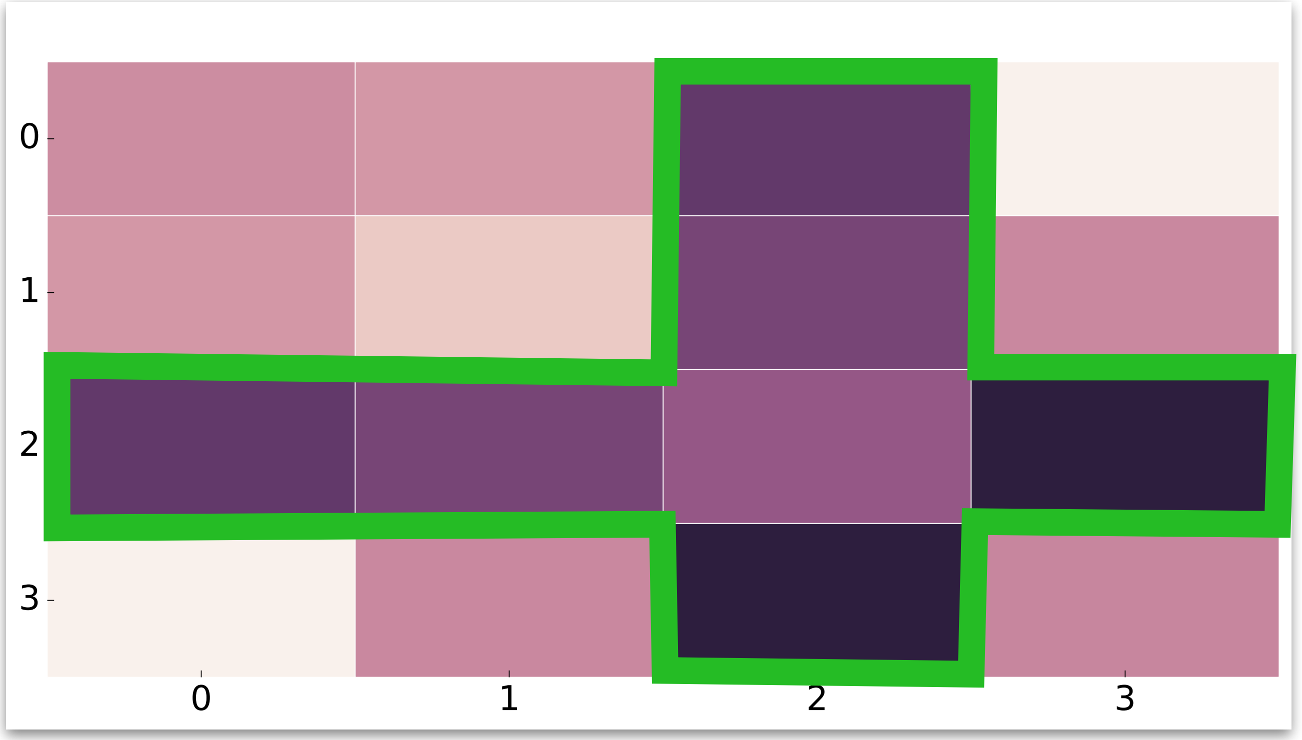
"Coherent" and "incoherent" matrix units contribute differently.



This analysis is borne out in looking at other true states as well.





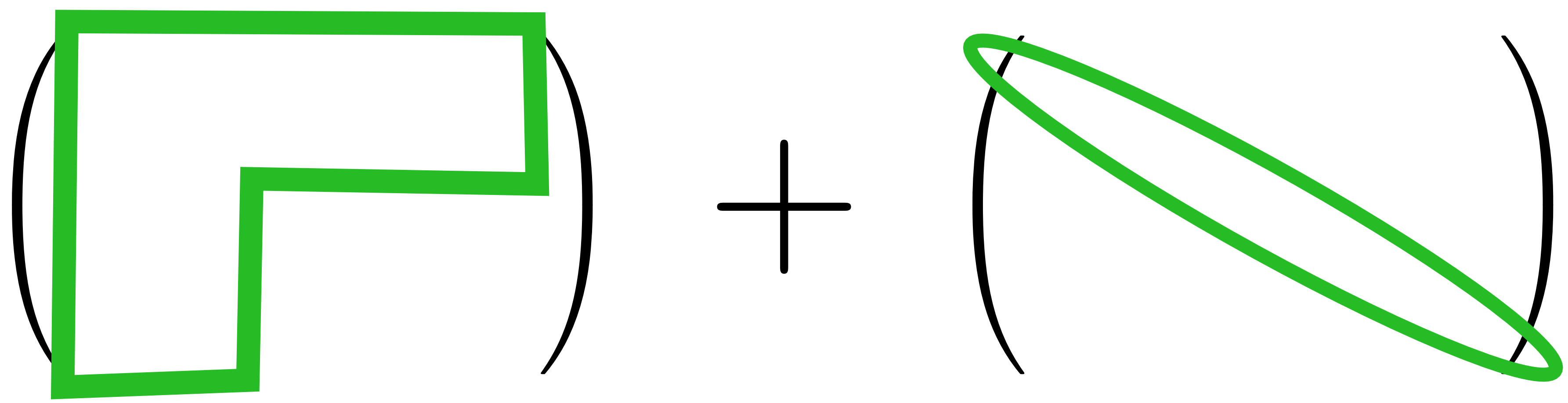


Is there a way
to model this behavior?

We build a simple model to explain these results, and replace the Wilks Theorem.

Coherent and **diagonal** matrix elements dominate

$$\langle \lambda(\rho_0, \mathcal{M}_d) \rangle \approx$$



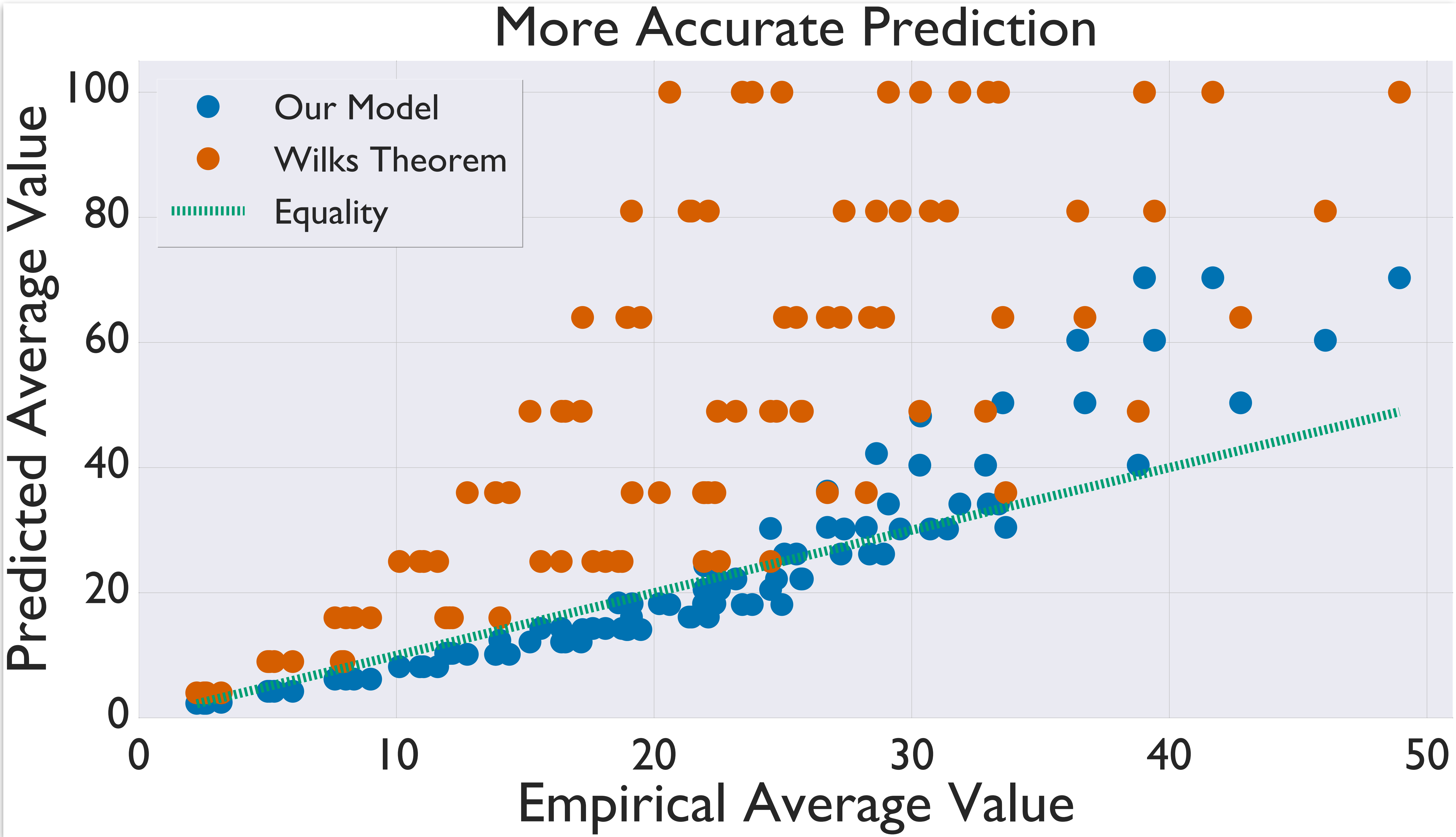
$$\langle \lambda(\rho_0, \mathcal{M}_d) \rangle \approx 2rd - r(r+1) + S(d,r)$$

$$r = \mathrm{rank}(\rho_0)$$

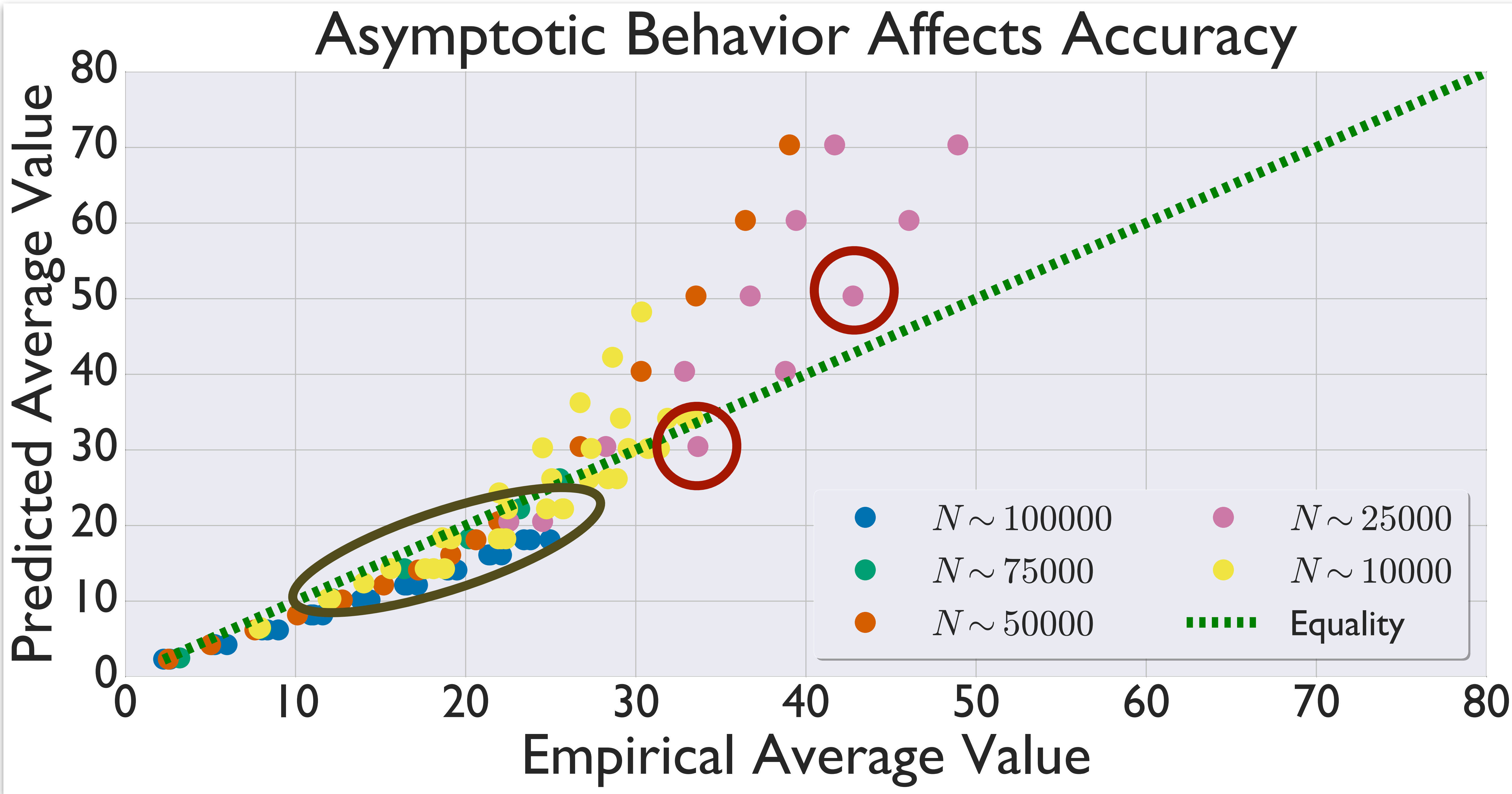
Does depend on rank of true state…
but **unitarily invariant**.

Reduces to Wilks in certain cases.

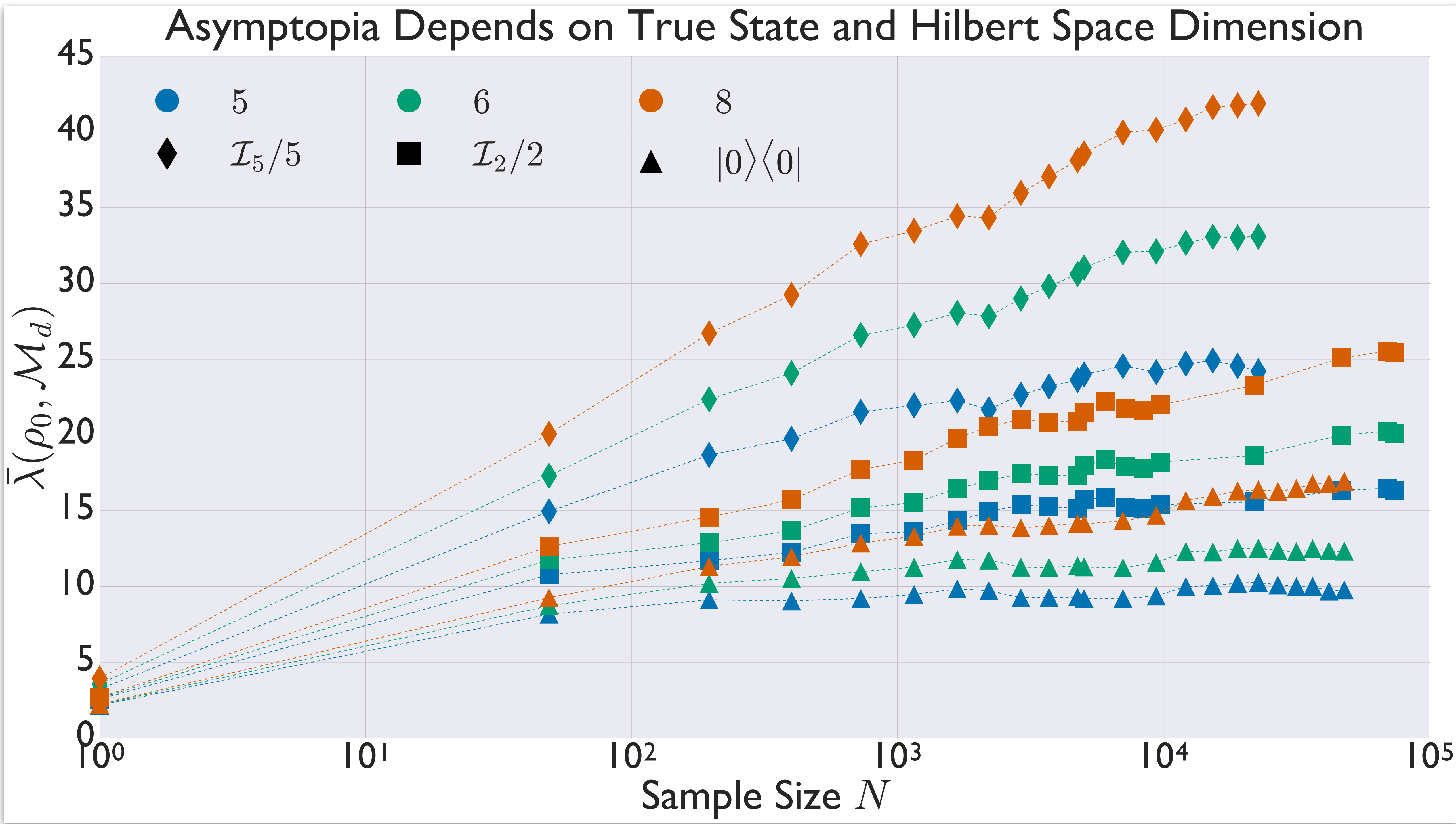
Our model does *better* than Wilks at predicting the expected value.



Let's check why the model might be inaccurate…



…it turns out asymptopia can be really, really far away!

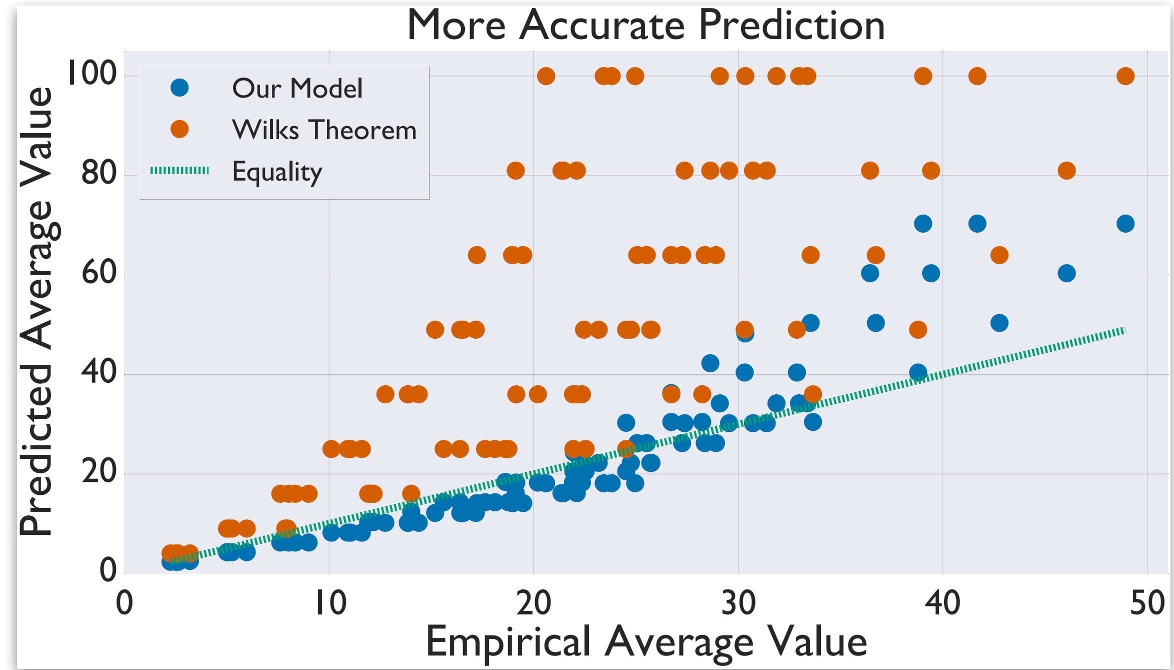


Very large sample sizes "activate" some parameters!

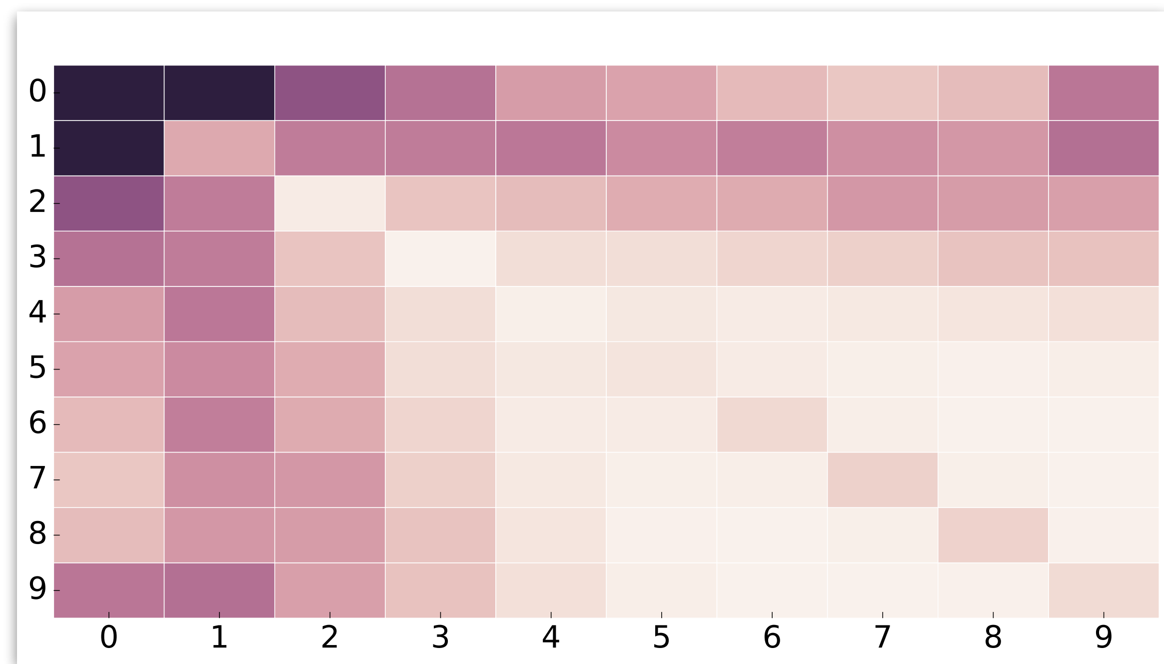
Where does this leave us?

Three key takeaways:

Don't use Wilks Theorem for this problem!

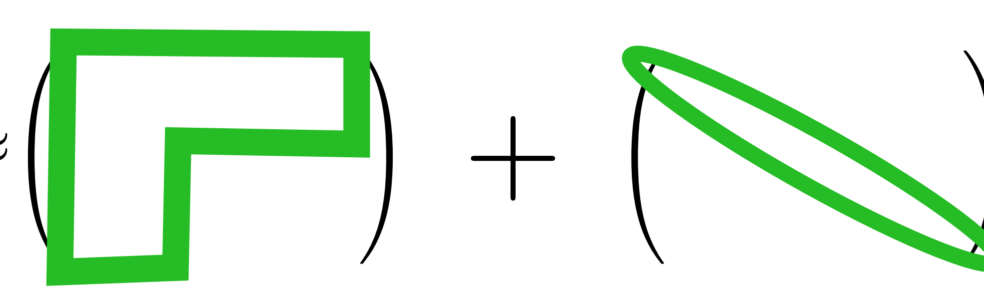


Different parameters = different contributions (Sample-size dependent)

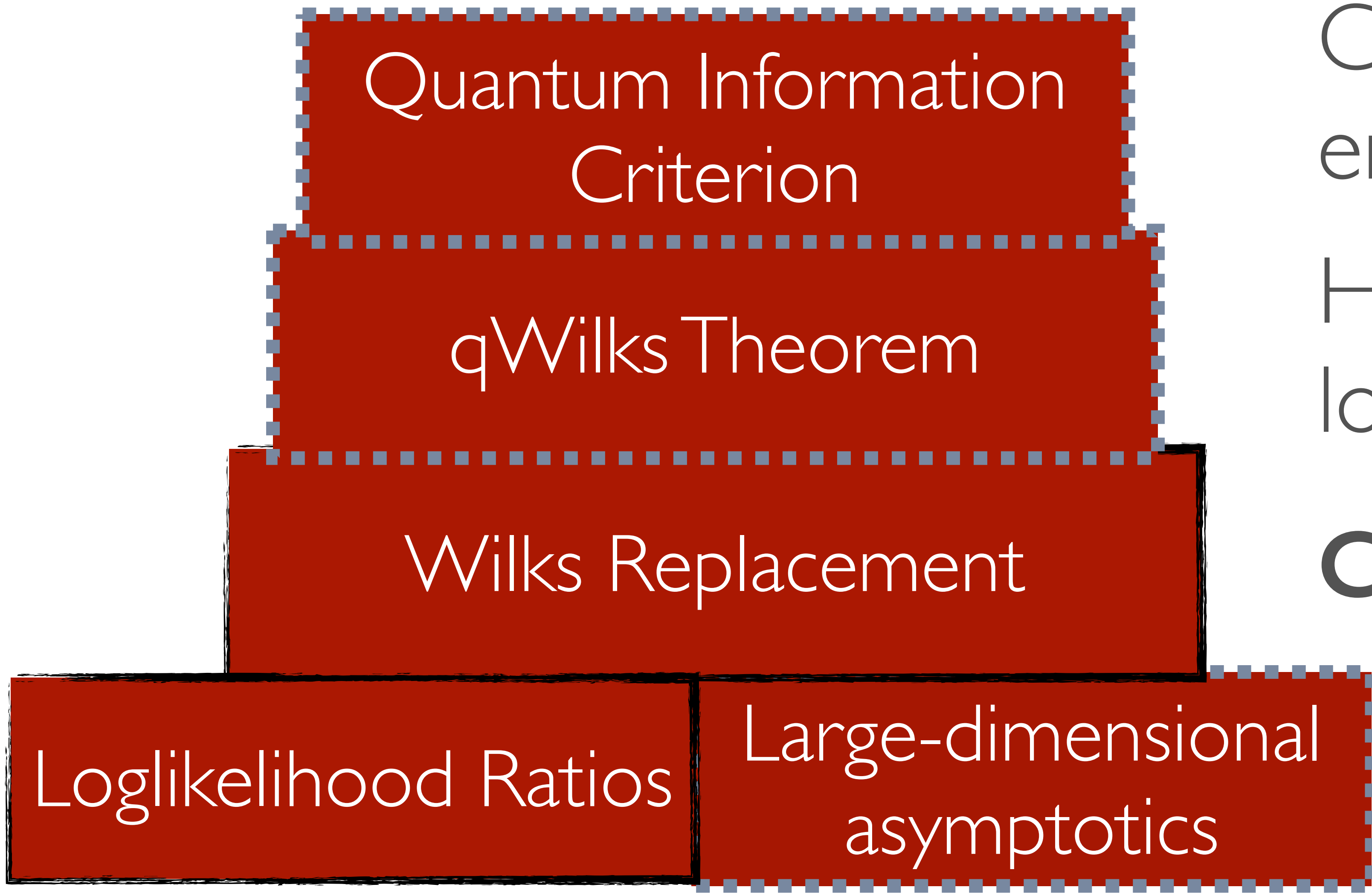


We built a replacement which works better.

$\langle \lambda(\rho_0, \mathcal{M}_d) \rangle \approx$



We're building a foundation on which we can do model selection correctly…



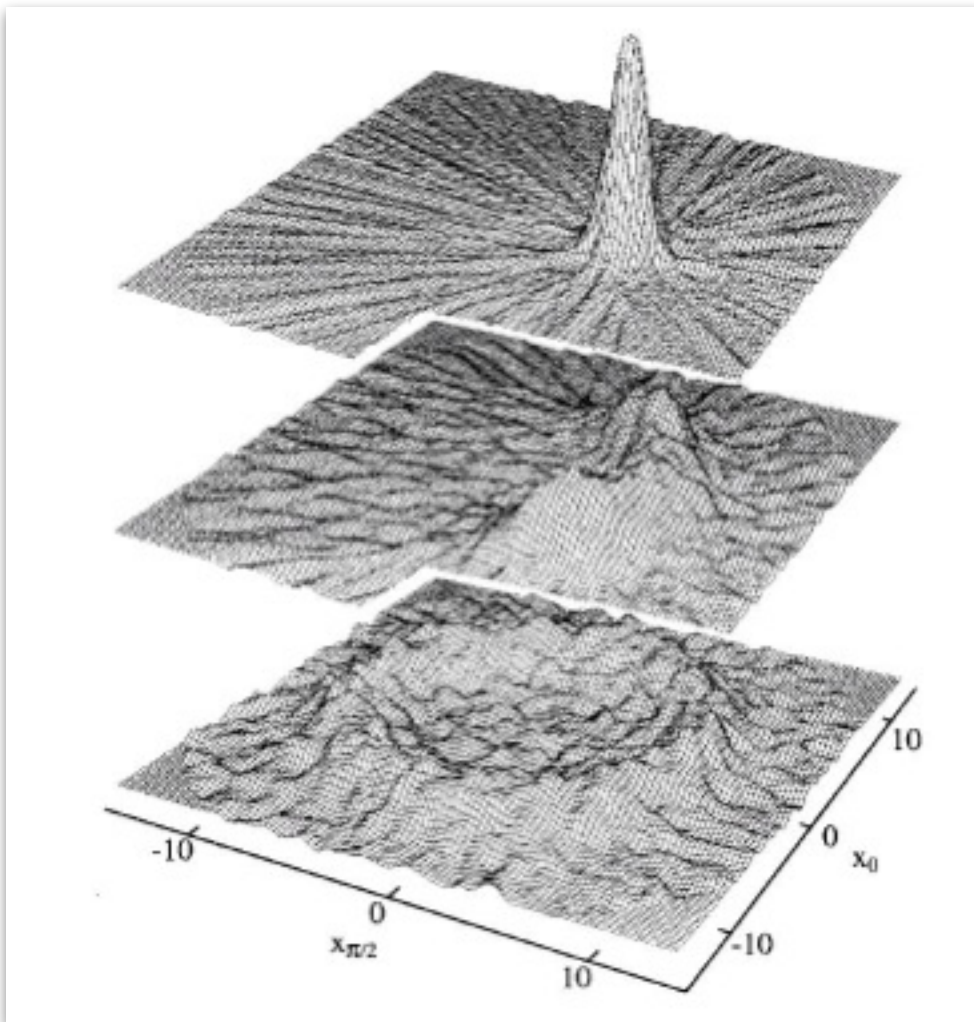
Correct measure of error/inaccuracy?

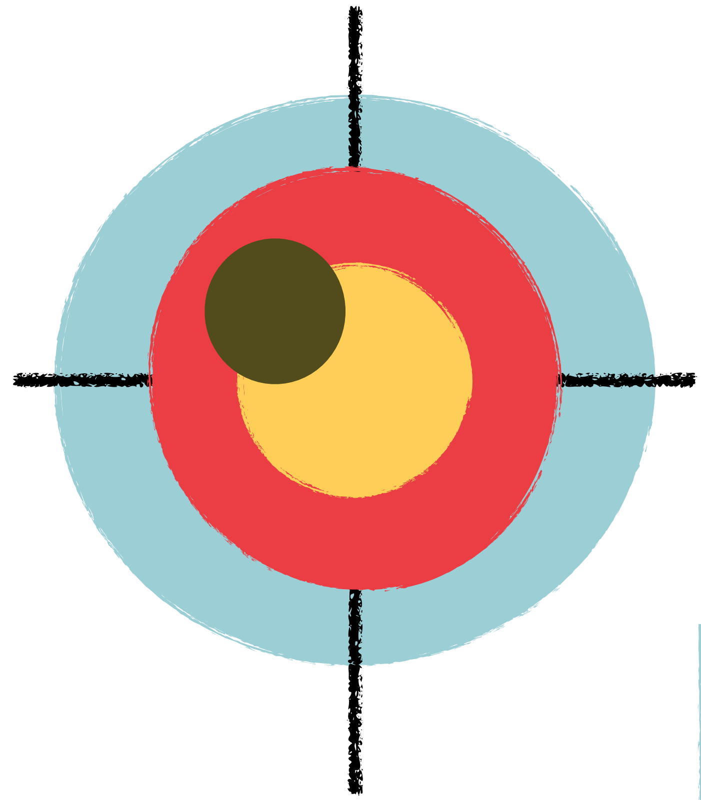
How *do* boundaries affect loglikelihood ratios?

Current work

How to handle big data *and* many parameters?

…and advancing the state-of-the-art in quantum tomography.





$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} & \cdots \\ \rho_{10} & \rho_{11} & \rho_{21} & \cdots \\ \rho_{20} & \rho_{21} & \rho_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Only **so much structure**…
so let us **model that well**!

Thank you!

@Travis_Sch

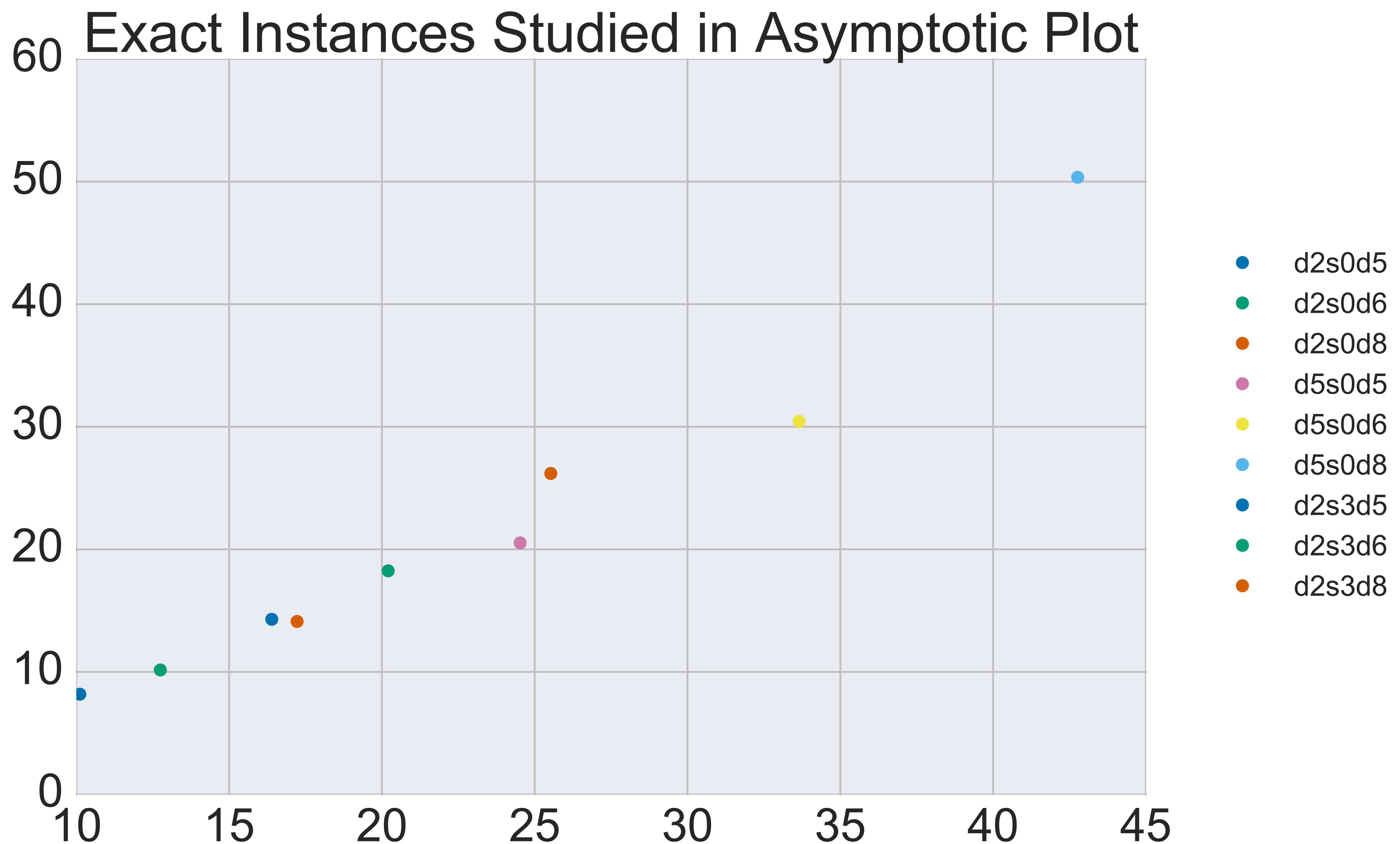
Image credits:

Wigner function: By Gerd Breitenbach (dissertation) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/)], via Wikimedia Commons

gmon qubit: By Michael Fang, John Martinis group. http://web.physics.ucsb.edu/~martinisgroup/photos.shtml

NIST Ion Trap: http://phys.org/news/2006-07-ion-large-quantum.html

Supplemental Material



Exact Instances Studied in Asymptotic Plot
60
50
40
30
20
10
0
10
15
20
25
30
35
40
45
d2s0d5
d2s0d6
d2s0d8
d5s0d5
d5s0d6
d5s0d8
d2s3d5
d2s3d6
d2s3d8