QUANTIFYING THE EVIDENCE FOR A PLANET IN RADIAL VELOCITY DATA

Ben Nelson
@exobenelson
Northwestern University → Insight Data Science Fellow

With: Johannes Buchner (PUC/Chile), Ryan Cloutier (Toronto), Rodrigo Diaz (IAFE), João Faria (Porto), Fabo Feng (Hertfordshire), Eric Ford (Penn State), Nathan Hara (IMCEE), Vinesh Rajpaul (Oxford), Suri Rukdee (PUC/Chile)

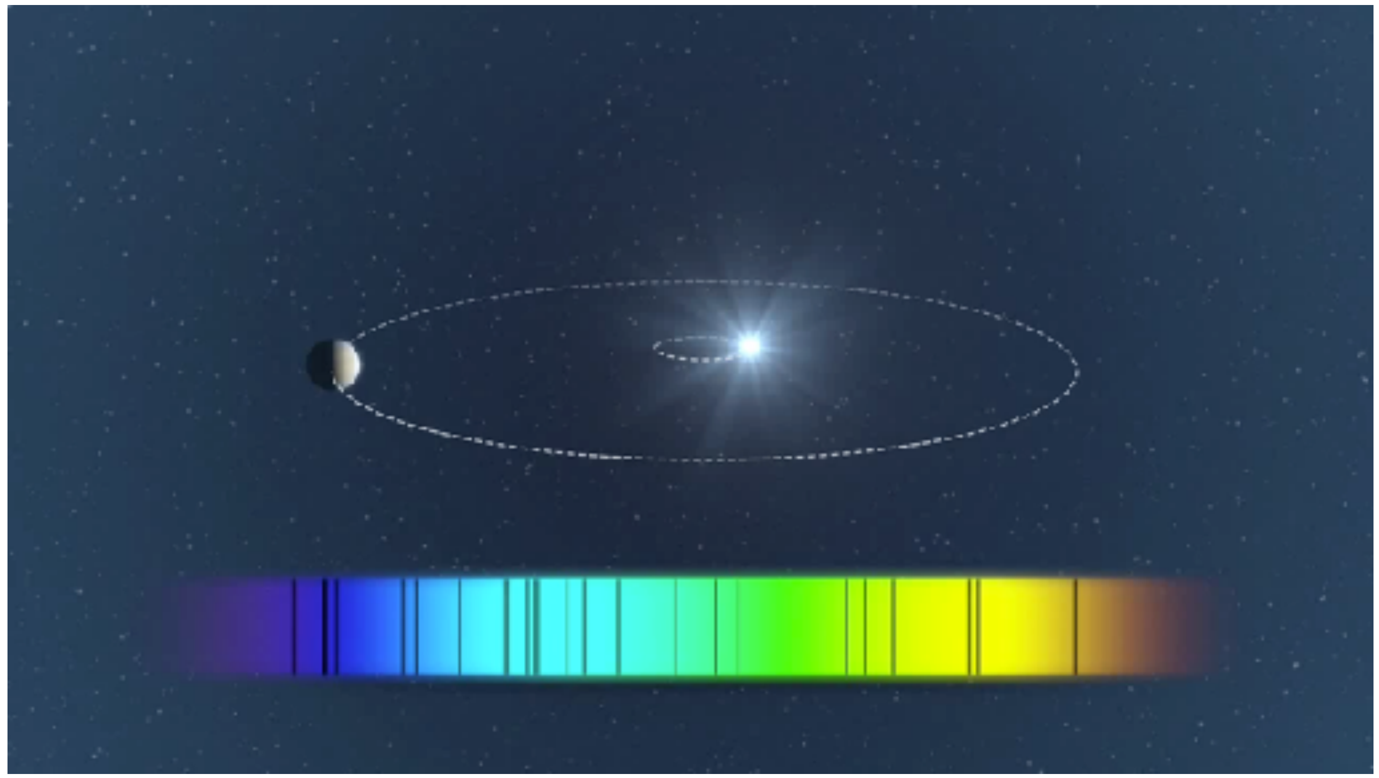
arXiv:1806.04683
github.com/EPRV3EvidenceChallenge

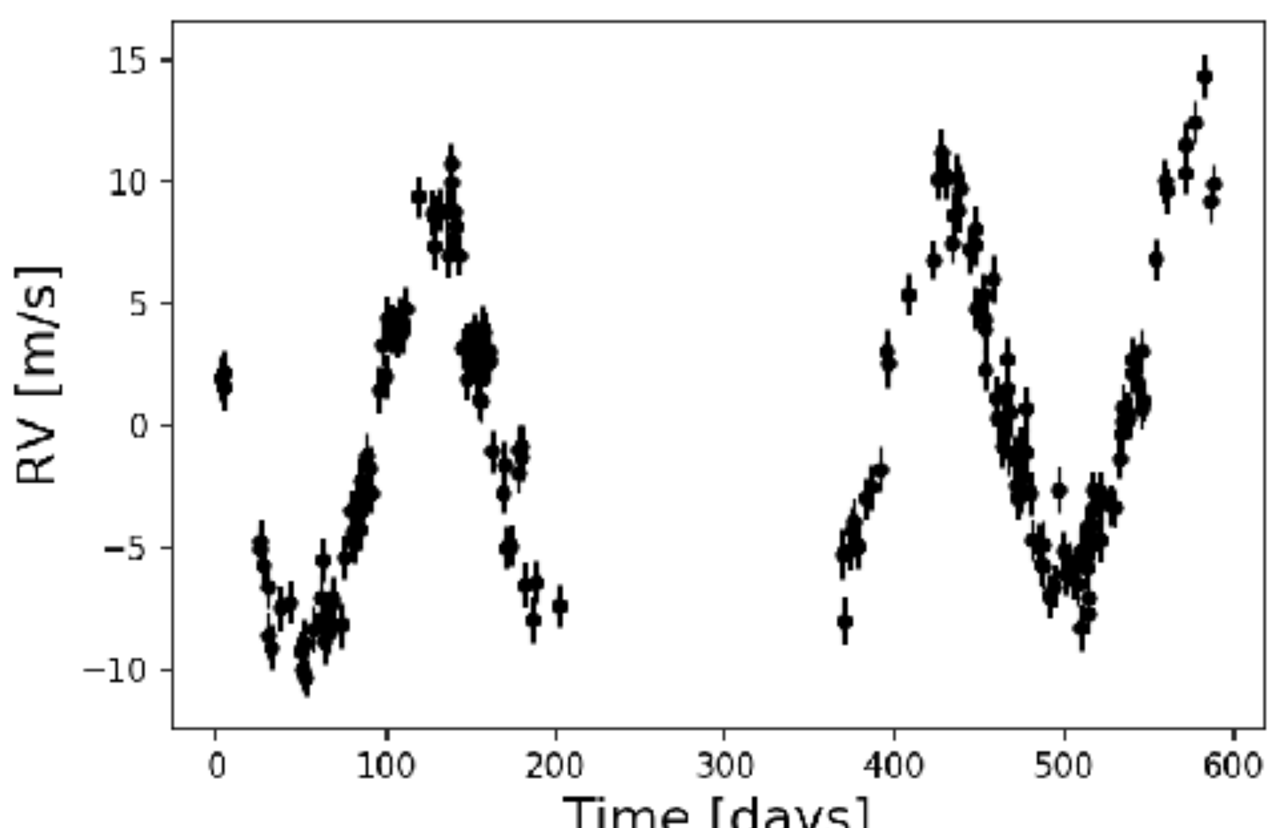
CIERA
Center for Interdisciplinary Exploration and Research in Astrophysics

NICO

July 24, 2018
Sagan Workshop

Radial Velocity Observations





5 parameters describe a planet's RV signature

orbital period
orbital eccentricity
argument of pericenter
orbital phase
mass

What does it mean to "discover" a planet?

Frequentist Approach

Reject the null hypothesis that a model without a planet could reasonably explain the data

conceptually difficult, computationally easy

Bayesian Approach

Evidence (i.e., marginalized likelihood) for a model with the planet is much greater than alternative models without the planet

conceptually easy, computationally difficult

Further reading

jakevdp.github.io

Frequentism & Bayesianism Part 5: Model Selection

Methods for dealing with new RV challenges

Dumusque 2016
Dumusque+ 2017

Computing the "evidence"

$$\mathcal{Z} \equiv p(d|\mathcal{M}) = \int p(\theta|\mathcal{M})p(d|\theta,\mathcal{M})d\theta$$

Thermodynamic integration
(HD208487, Gregory 2007)

Nested sampling / MultiNest
(GJ667C, Feroz & Hobson 2014)

Geometric path Monte Carlo
(GJ581, Hou+ 2014)

Transdimensional MCMC w/ nested sampling
(ν Oph, Brewer & Donovan 2015)

Importance sampling
(GJ876, Nelson+ 2016; HD9174, Jenkins+ 2017)

THE THIRD WORKSHOP ON

EPRV

EXTREMELY PRECISE RADIAL VELOCITIES

Evidence Challenge

How accurately/precisely can one compute the "evidence" for {0, 1, 2, 3} planets in RV data, given a set of priors and likelihood function?

$$\mathcal{Z} \equiv p(d|\mathcal{M}) = \int p(\theta|\mathcal{M})p(d|\theta,\mathcal{M})d\theta$$

July 24, 2018
Sagan Workshop

How is this NOT Xavier Dumusque's RV Fitting Challenge?

THE THIRD WORKSHOP ON EPRV
EXTREMELY PRECISE RADIAL VELOCITIES

Evidence Challenge

How is this NOT Xavier Dumusque's RV Fitting Challenge?

THE THIRD WORKSHOP ON EPRV
EXTREMELY PRECISE RADIAL VELOCITIES

Evidence Challenge

what methods are good at finding "real" planets in data	what methods are good at computing an accurate "evidence"
data: RVs and activity indicators (FWHM, logR'{hk}, BIS span)	data: just RVs
prize: 30 year old Tawny Port wine	prize: knowledge

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

4. What methods should be recommended/avoided/further improved?

EPRV3 Evidence Challenge

We generate 6 RV datasets. Each dataset contained **two** planets with varying levels of detectability.

Data are drawn from a multivariate Gaussian with correlated observations, measured uncertainties, and an additional unknown white noise term (i.e., jitter).

We use a quasi-periodic kernel (Rajpaul+ 2015)...

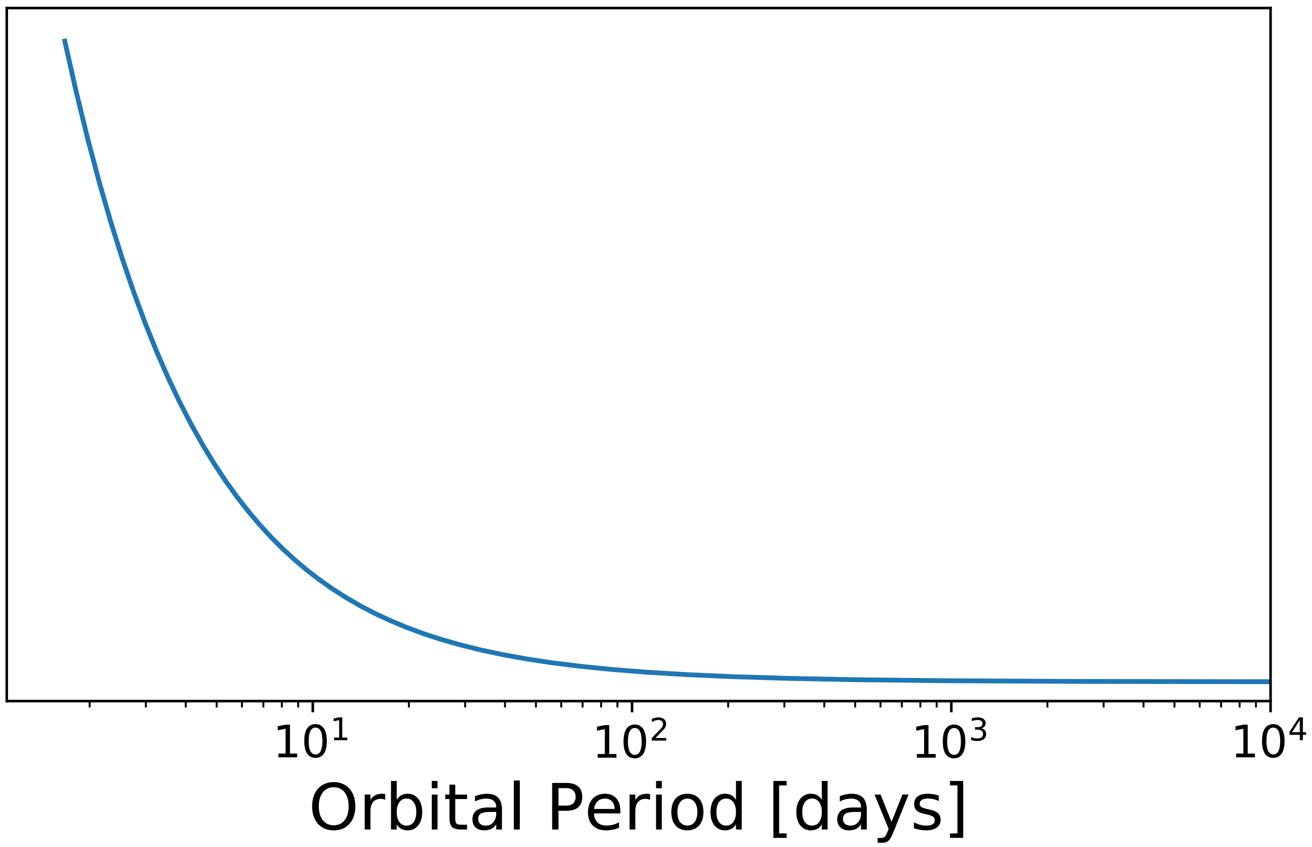
$$K_{i,j} = \alpha^2 exp\left[-\frac{1}{2}\left\{\frac{sin^2[\pi(t_i - t_j)/\tau]}{\lambda_p^2} + \frac{(t_i - t_j)^2}{\lambda_e^2}\right\}\right]$$

...with known hyperparameters α, λ_p, λ_e, and τ.

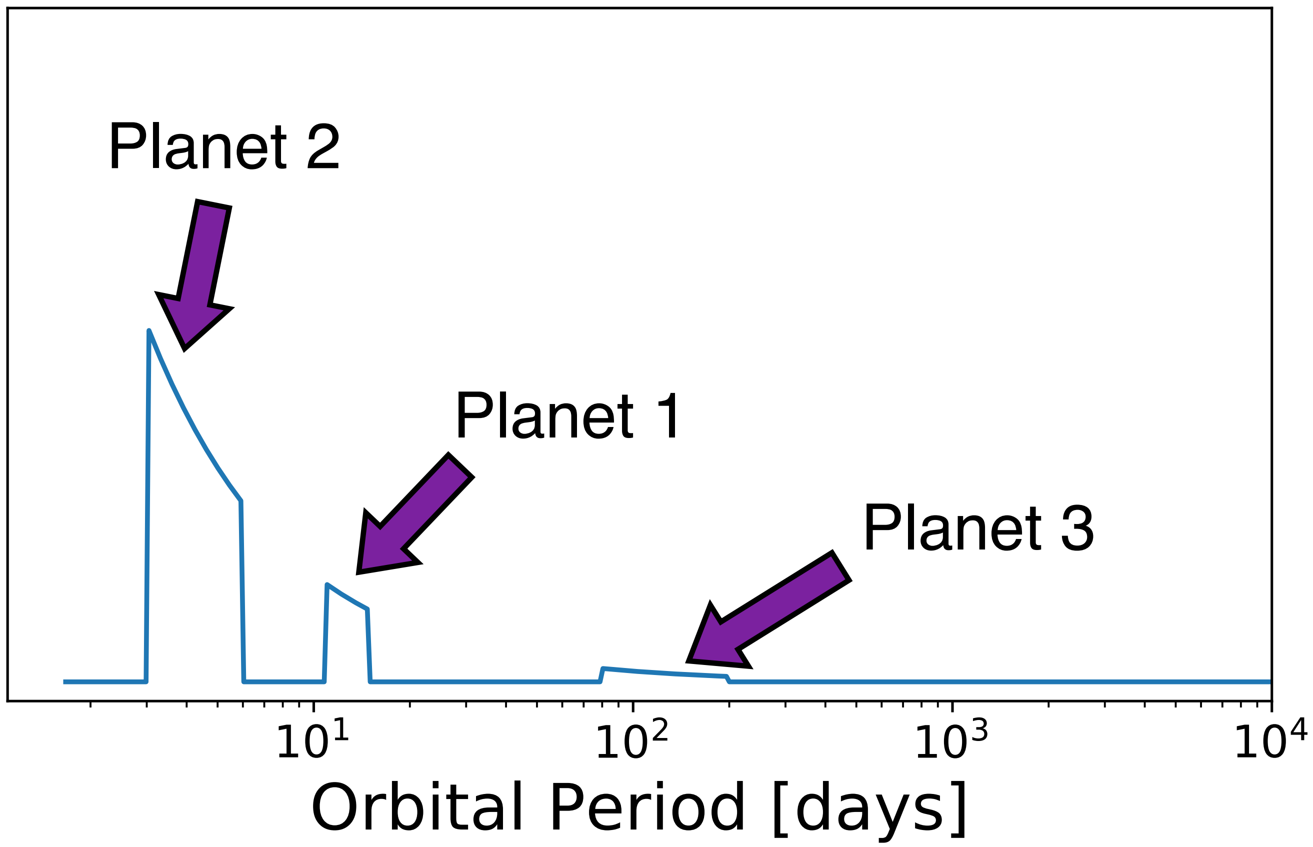
EPRV3 Evidence Challenge

Two sets of priors

Broad



Narrow



EPRV3 Evidence Challenge

More details and results at:
github.com/EPRV3EvidenceChallenge/

Methods teams submitted:

<u>Frequentist</u>
BIC
leave-one-out cross-validation
time-series cross-validation

<u>Bayesian</u>
Chib's approximation
Laplace approximation
Laplace approximation + l1 periodogram
Perrakis estimator
importance sampling + MCMC
importance sampling + variational Bayes
nested sampling (MultiNest)
nested sampling + MCMC
diffusive nested sampling (DNest4)

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

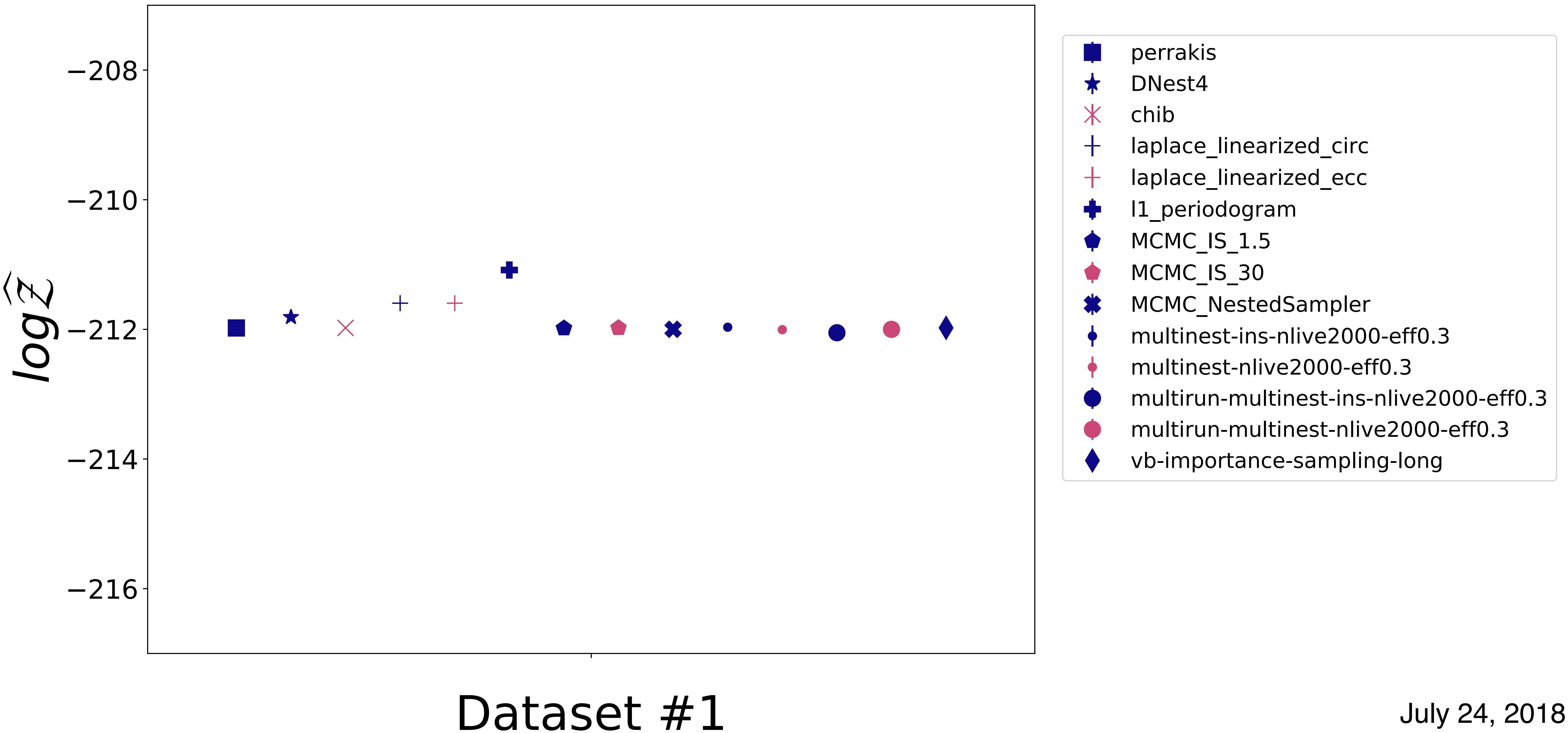
2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

4. What methods should be recommended/avoided/further improved?

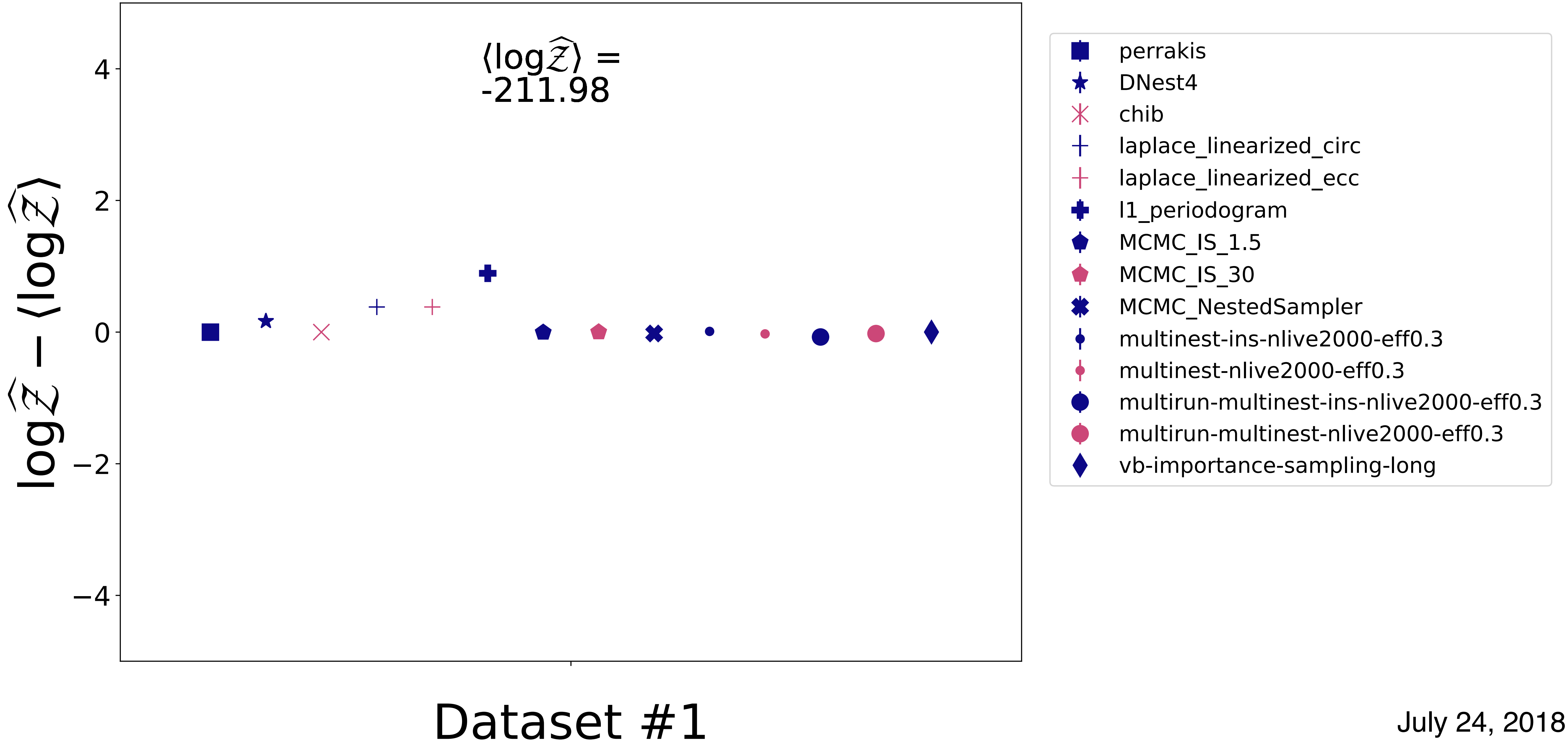
Evidence Challenge Results

0-planet model (2 parameters)



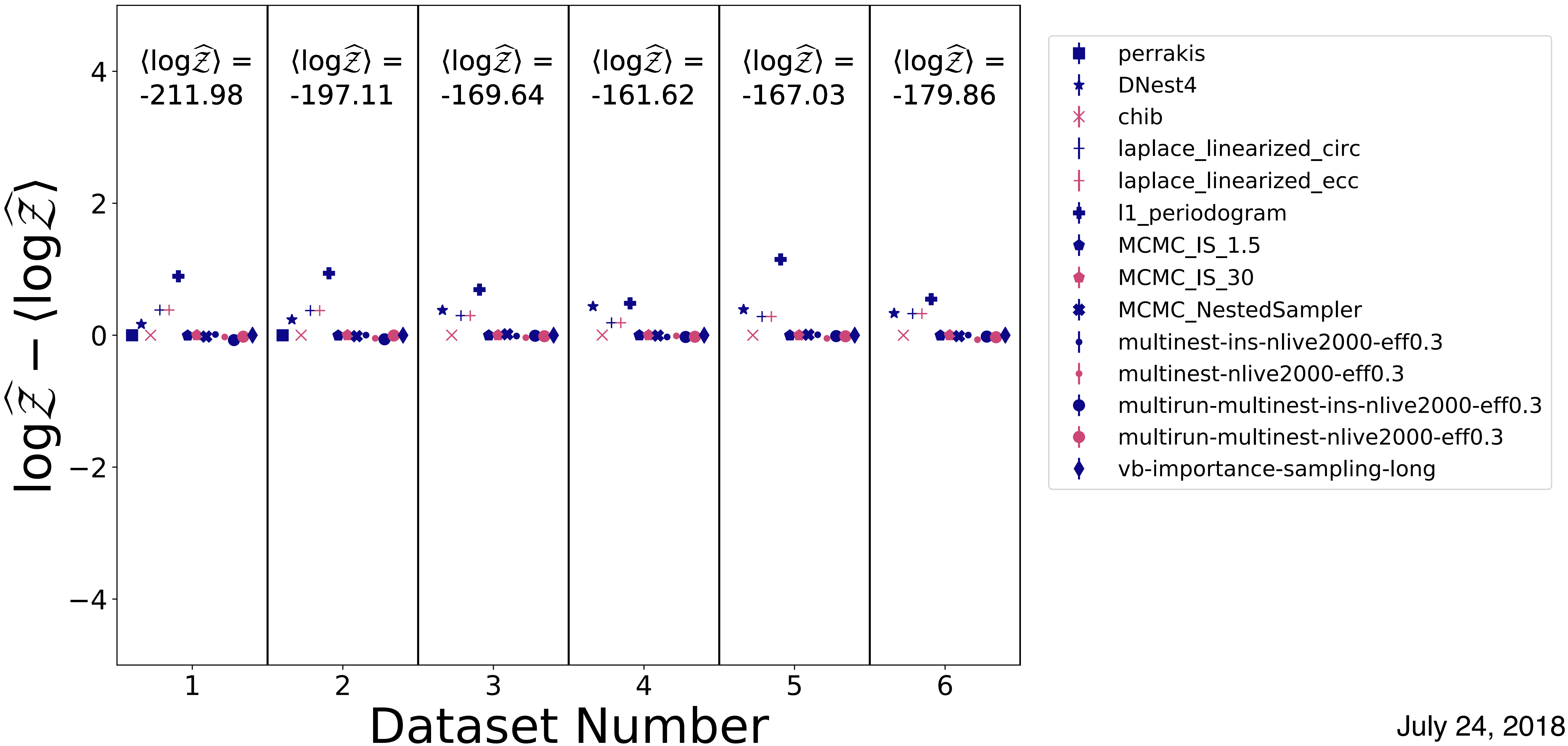
Evidence Challenge Results

0-planet model (2 parameters)



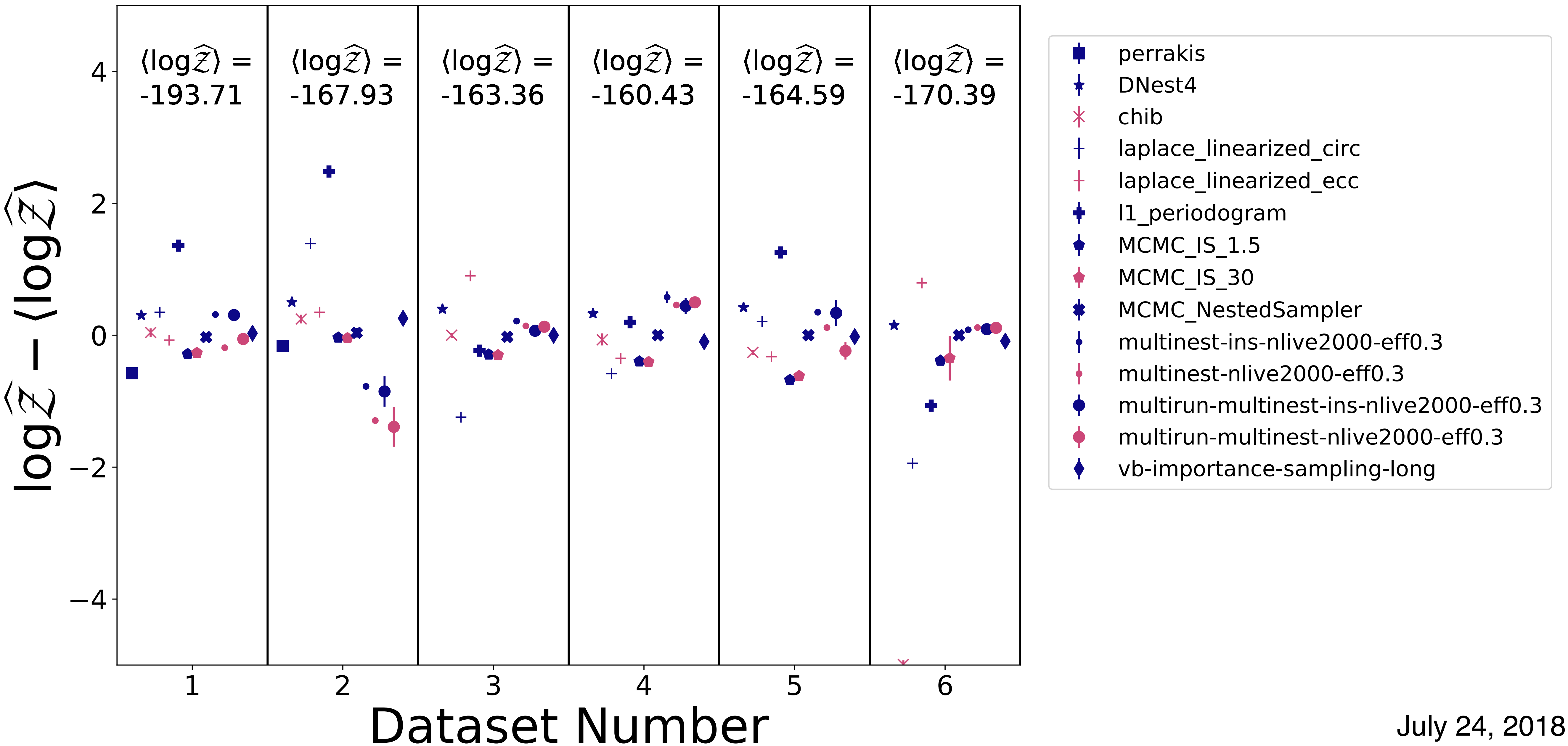
Evidence Challenge Results

0-planet model (2 parameters)



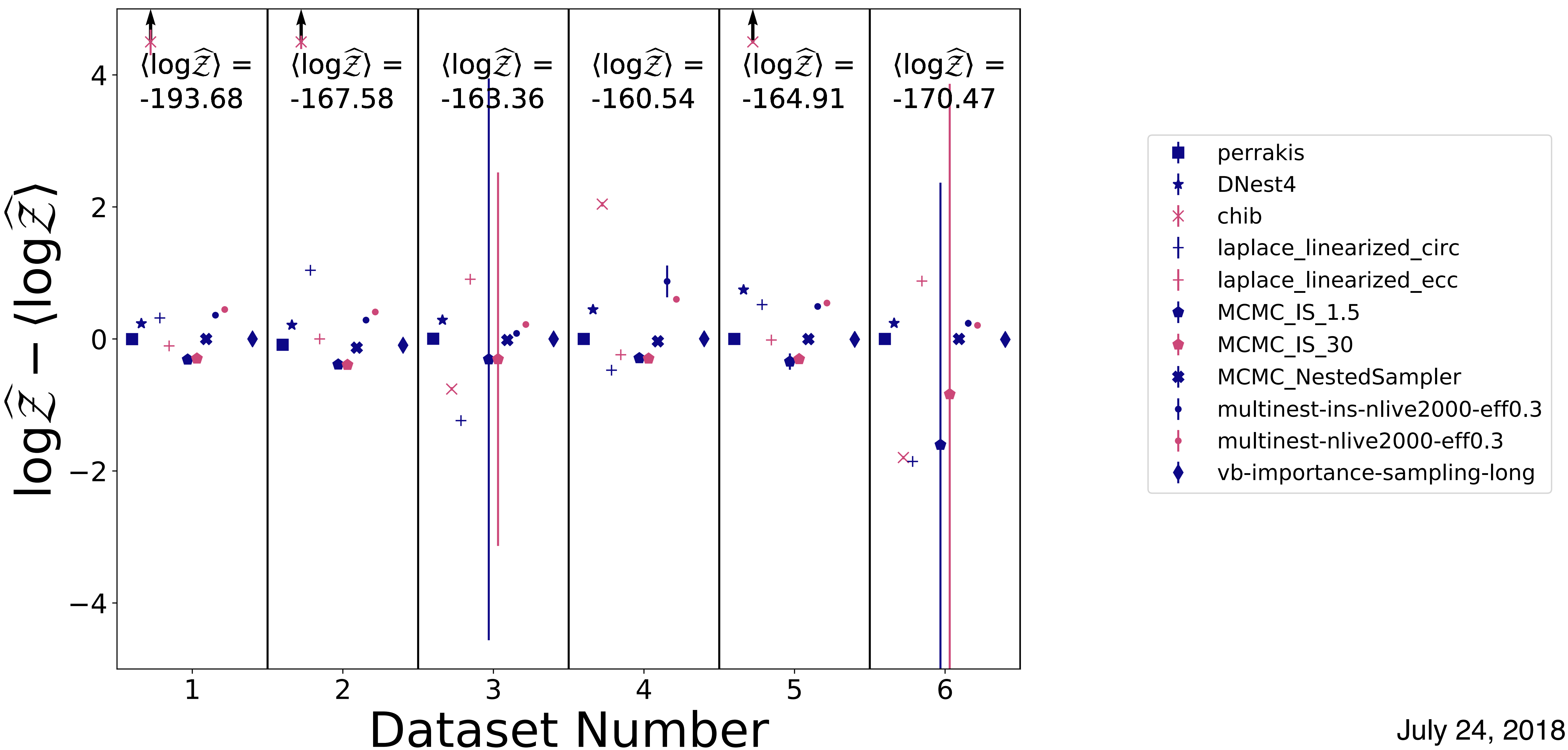
Evidence Challenge Results

1-planet model (7 parameters)



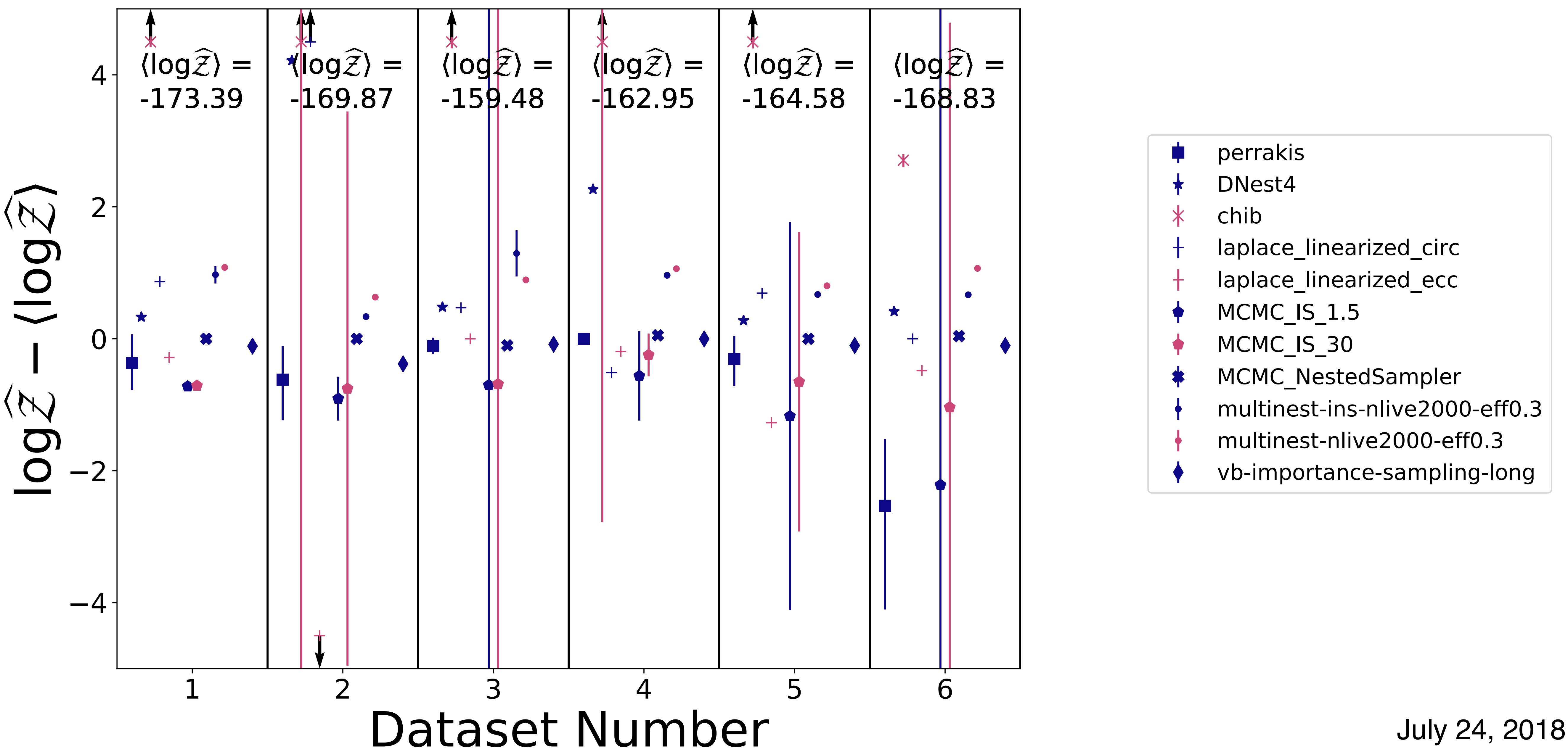
Evidence Challenge Results

1-planet model (7 parameters)   narrow priors



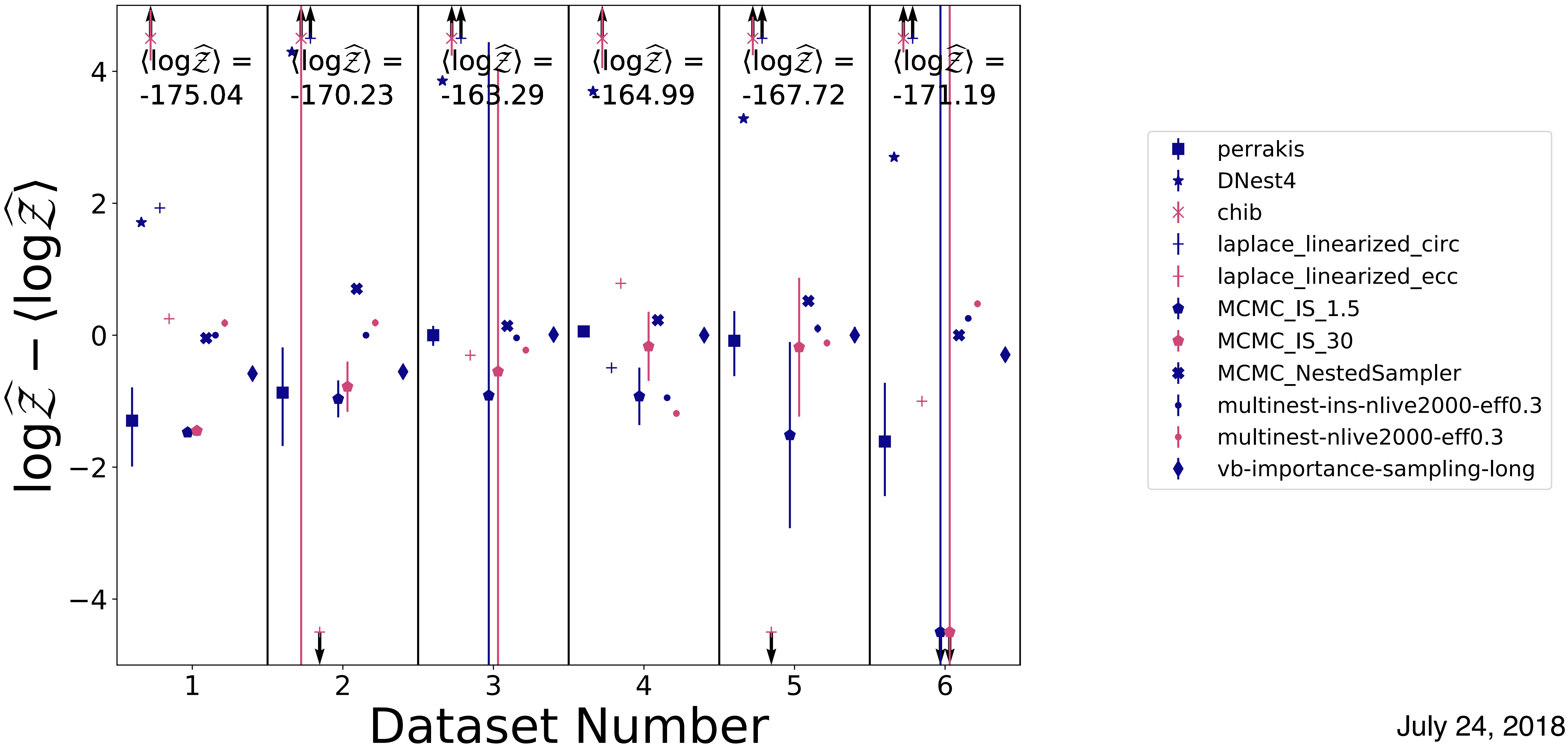
Evidence Challenge Results

2-planet model (12 parameters) narrow priors



Evidence Challenge Results

3-planet model (17 parameters) narrow priors



July 24, 2018
Sagan Workshop

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

model: 0 planets ----> 1 planet ----> 2 planets ----> 3 planets
dispersion: ~few ~OOM ~few to several OOMs

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

Internal estimates of one evidence calculation underestimate the uncertainty.
Monte Carlo methods seem to provide reasonable uncertainties.

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

4. What methods should be recommended/avoided/further improved?

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

model: 0 planets ----> 1 planet ----> 2 planets ----> 3 planets
dispersion: ~few ~OOM ~few to several OOMs

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

Internal estimates of one evidence calculation underestimate the uncertainty.
Monte Carlo methods seem to provide reasonable uncertainties.

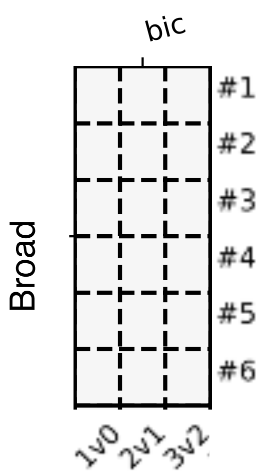
3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

4. What methods should be recommended/avoided/further improved?

...and can I somehow cram all this **information** into a single figure?

datasets
models
priors
methods
logZ or logOddsRatio

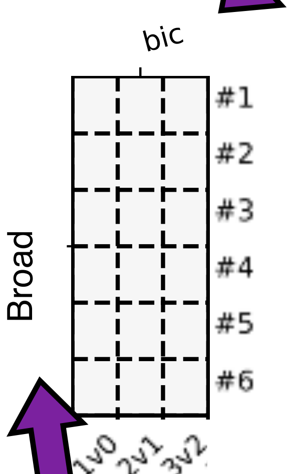
What different methods say about n vs n-1 planets



July 24, 2018
Sagan Workshop

What different methods say about n vs n-1 planets

method

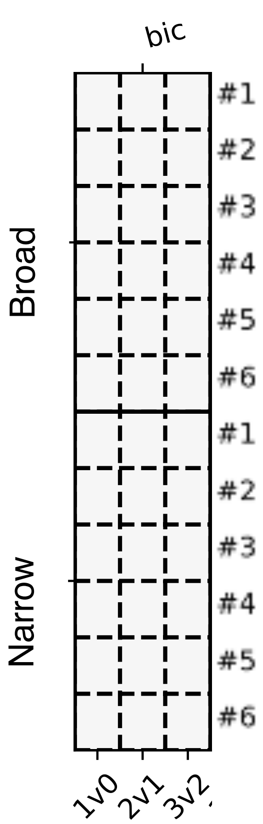


dataset numbers

(n)v(n-1), i.e., $\frac{p(\mathcal{M}_n|d)}{p(\mathcal{M}_{n-1}|d)}$

choice of prior

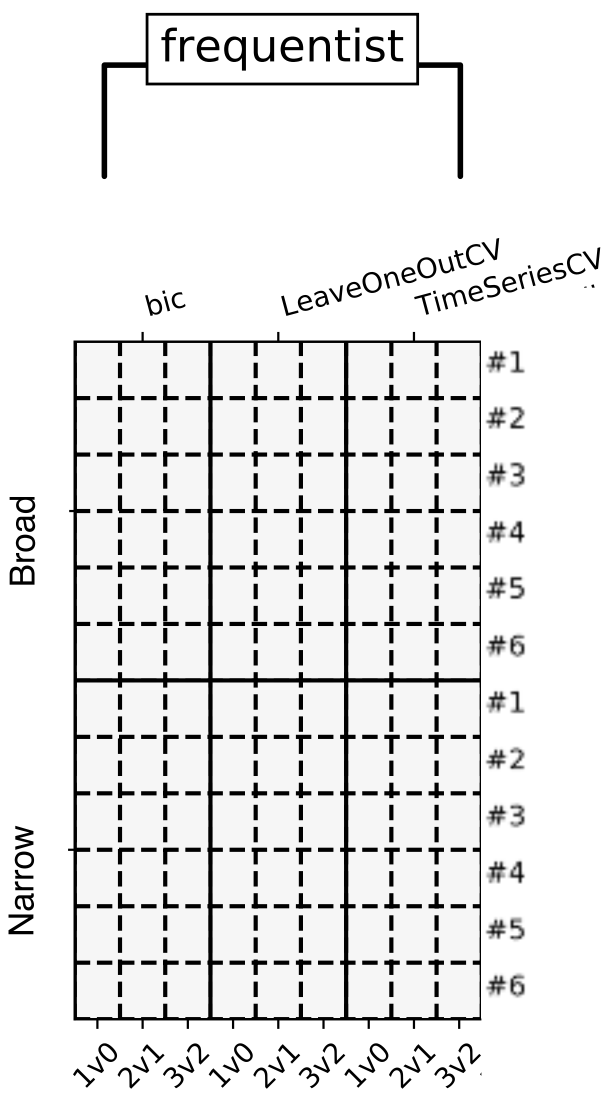
What different methods say about n vs n-1 planets



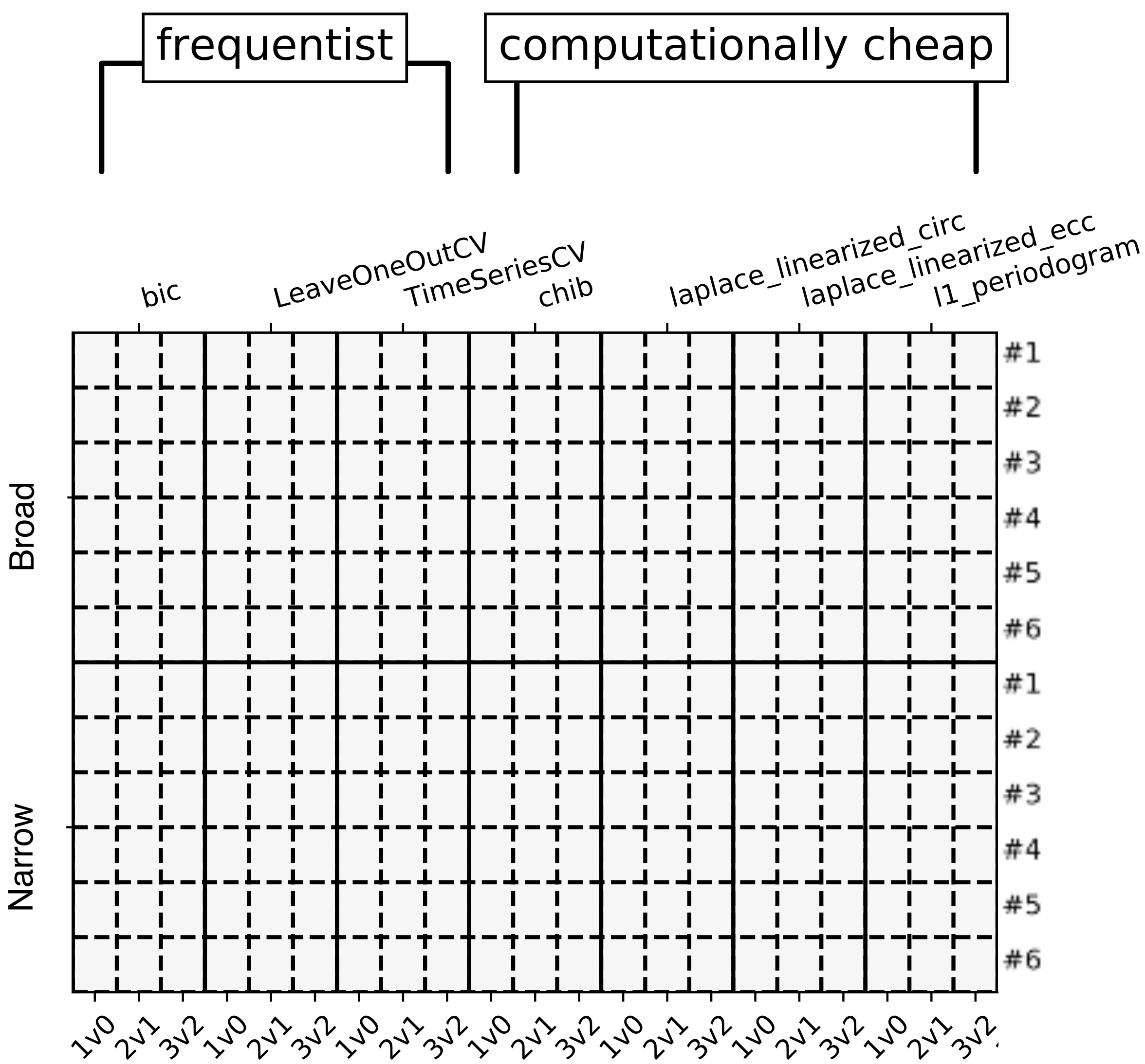
datasets with broad prior

datasets with narrow prior

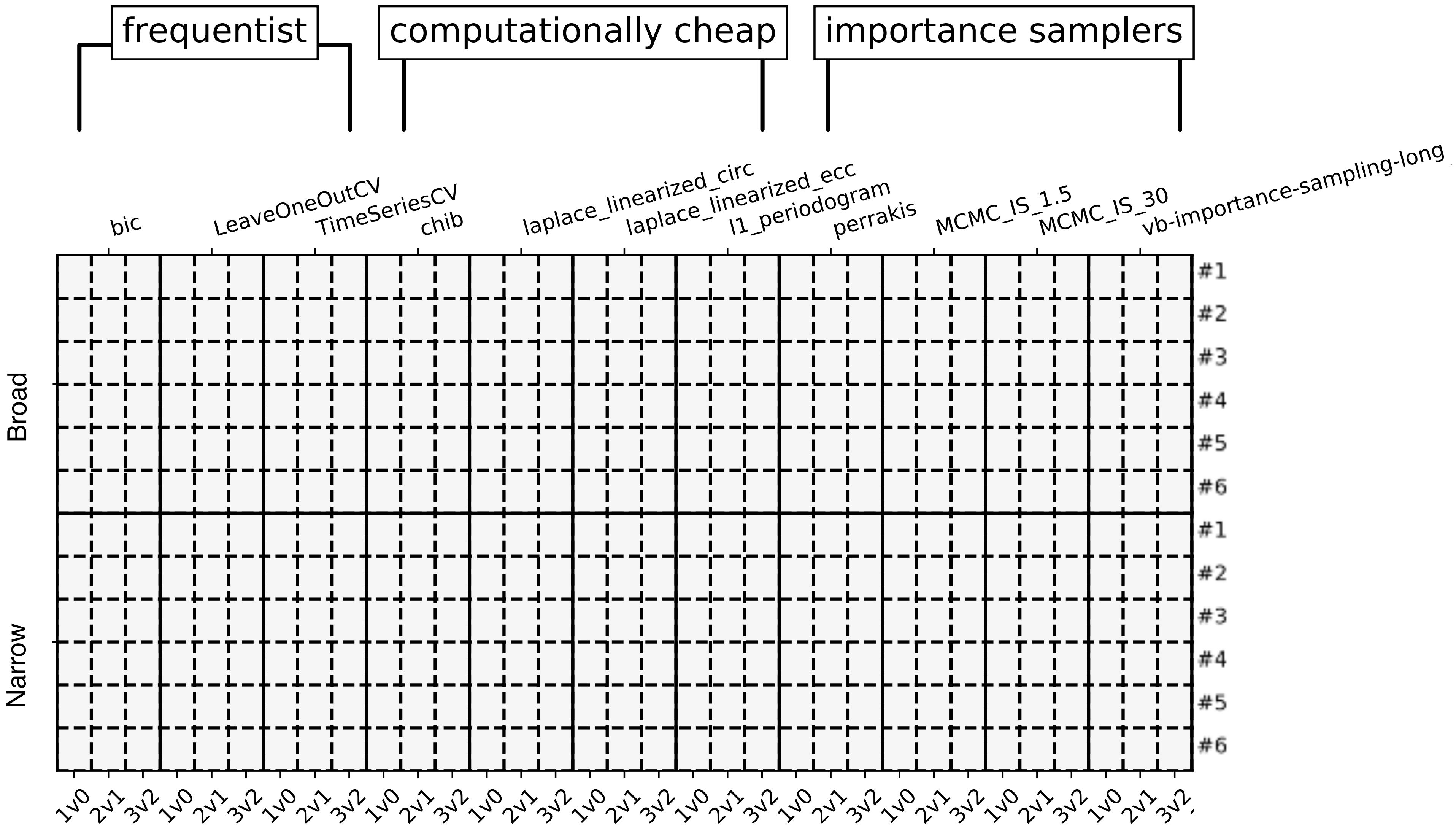
What different methods say about n vs n-1 planets



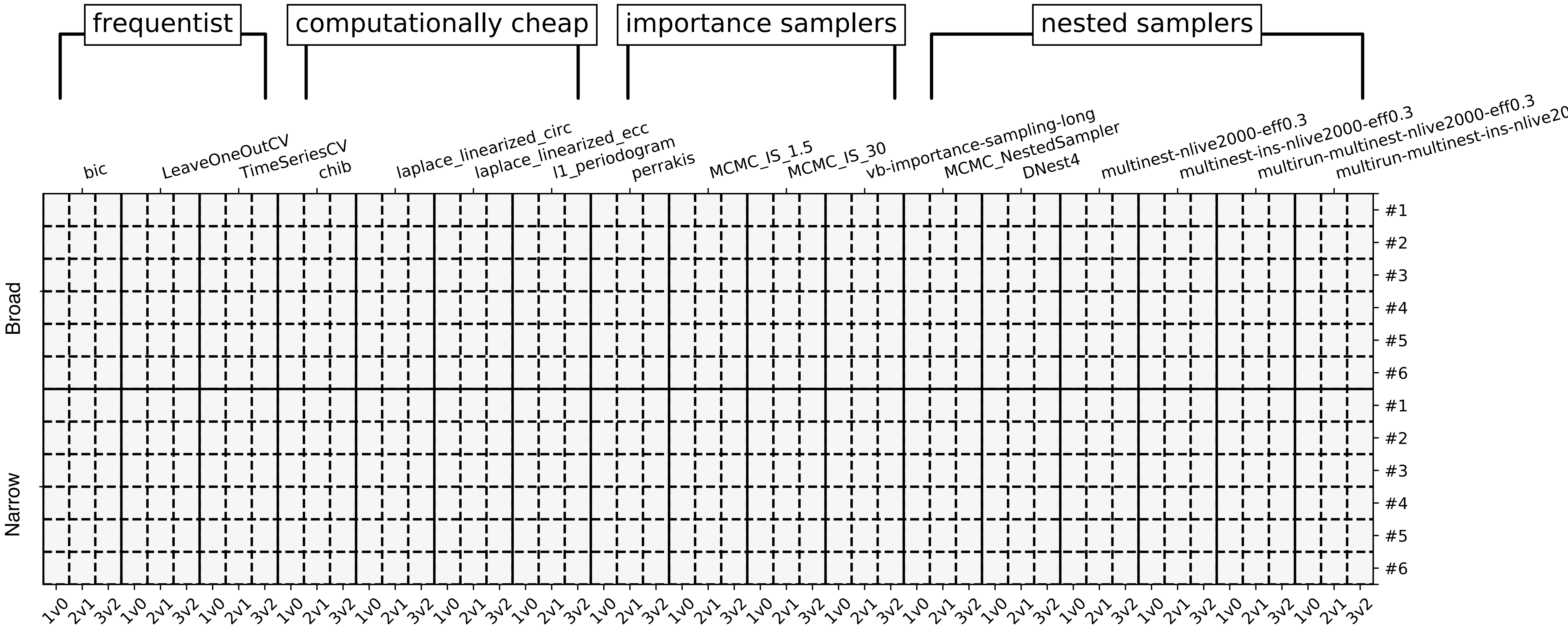
What different methods say about n vs n-1 planets



What different methods say about n vs n-1 planets

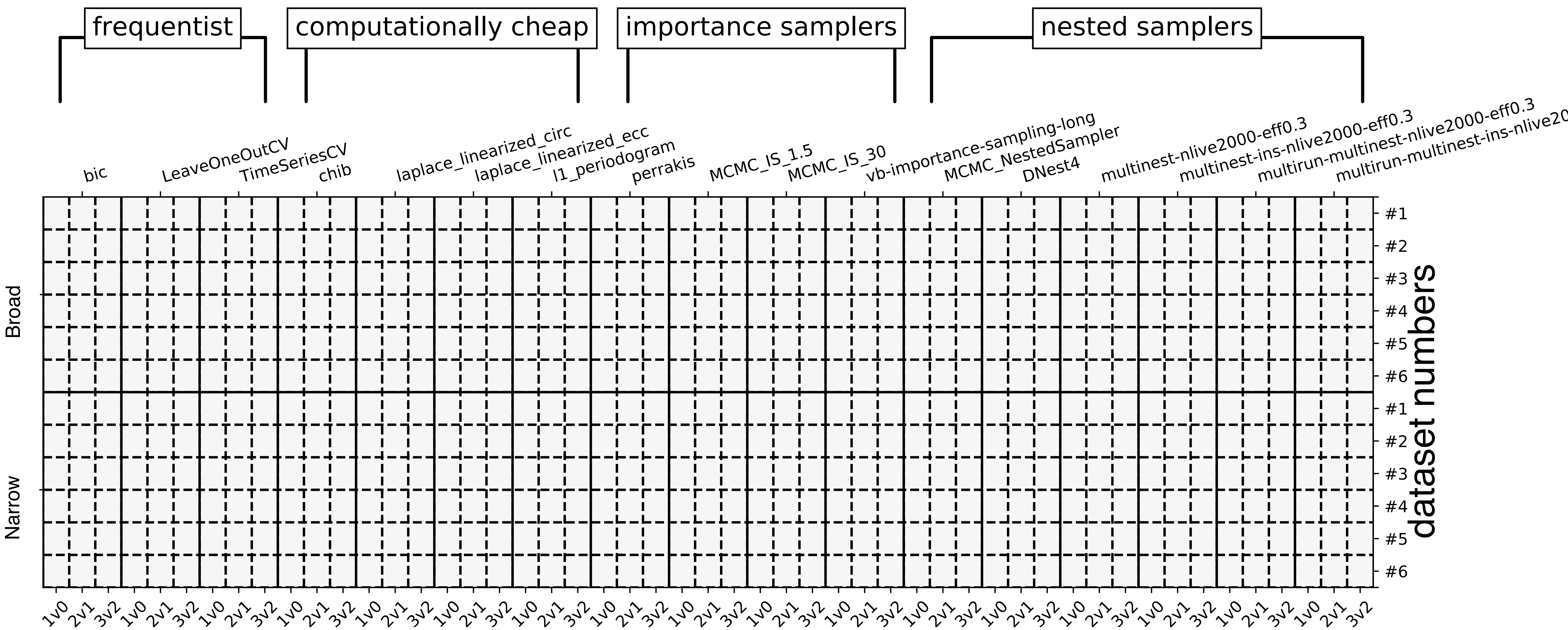


What different methods say about n vs n-1 planets



July 24, 2018
Sagan Workshop

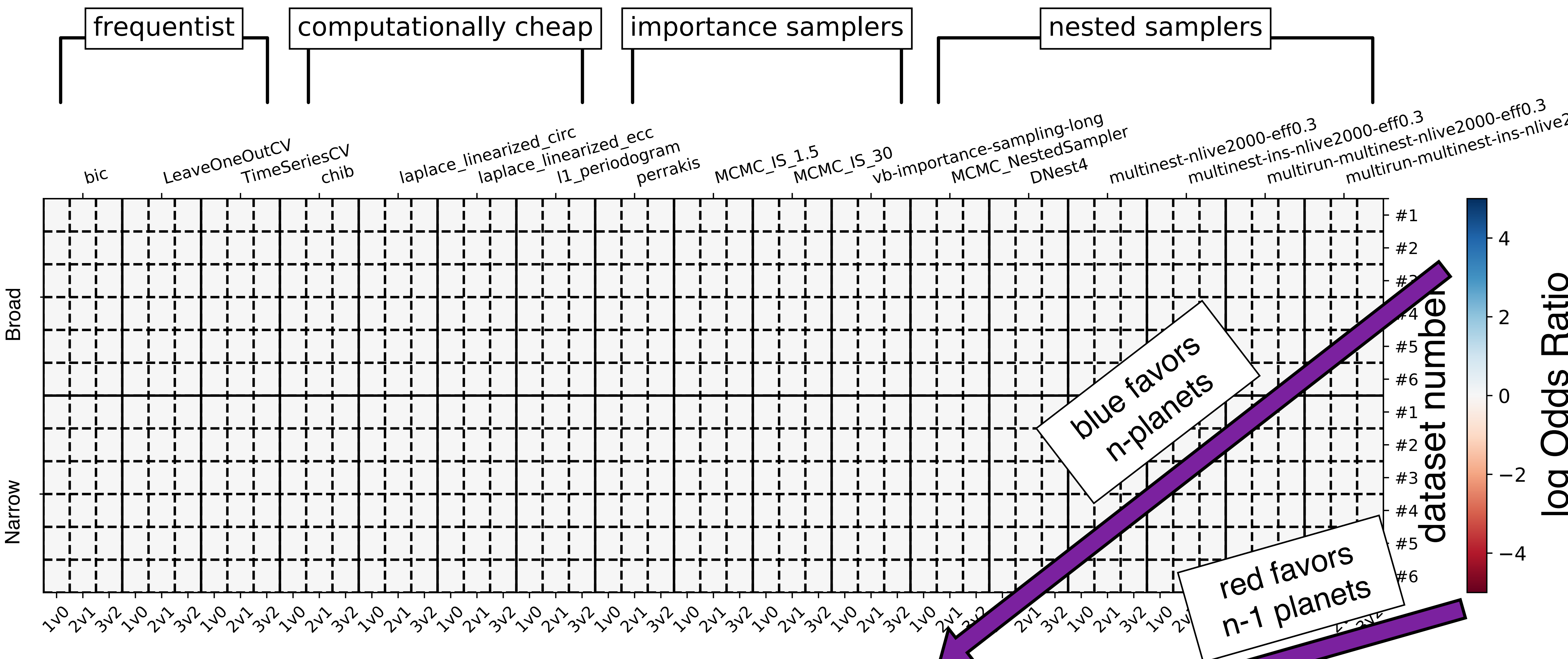
What different methods say about n vs n-1 planets



(n)v(n-1), i.e., $\frac{p(\mathcal{M}_n|d)}{p(\mathcal{M}_{n-1}|d)}$

July 24, 2018
Sagan Workshop

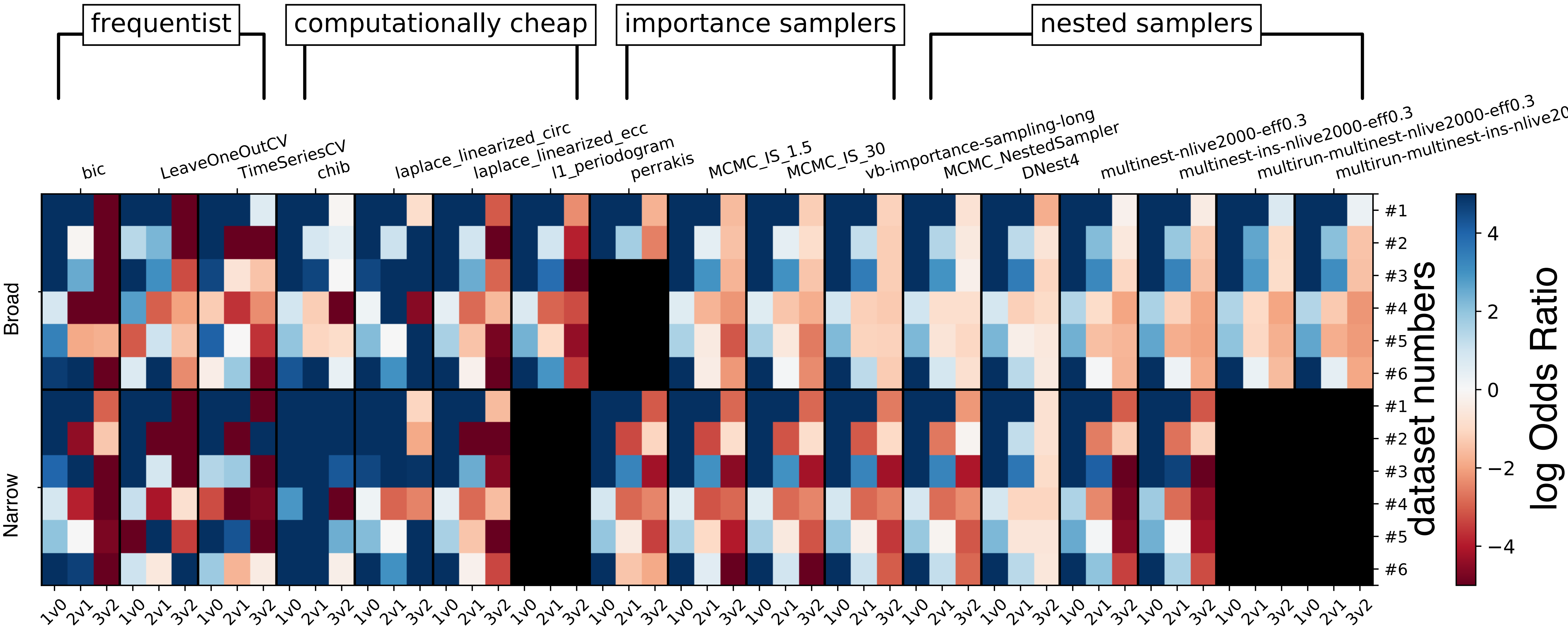
What different methods say about n vs n-1 planets



(n)v(n-1), i.e., $\frac{p(\mathcal{M}_n|d)}{p(\mathcal{M}_{n-1}|d)}$

July 24, 2018
Sagan Workshop

What different methods say about n vs n-1 planets



(n)v(n-1), i.e., $\frac{p(\mathcal{M}_n|d)}{p(\mathcal{M}_{n-1}|d)}$

July 24, 2018
Sagan Workshop

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

model: 0 planets ----> 1 planet ----> 2 planets ----> 3 planets
dispersion: ~few ~OOM ~few to several OOMs

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

Internal estimates of one evidence calculation underestimate the uncertainty.
Monte Carlo methods seem to provide reasonable uncertainties.

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

Importance and nested sampling methods mostly arrive at the same conclusions.
Cheaper methods are relatively overconfident in estimated odds ratios.

4. What methods should be recommended/avoided/further improved?

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

   model: 0 planets ----> 1 planet ----> 2 planets ----> 3 planets
   dispersion: ~few ~OOM ~few to several OOMs

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

   Internal estimates of one evidence calculation underestimate the uncertainty.
   Monte Carlo methods seem to provide reasonable uncertainties.

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

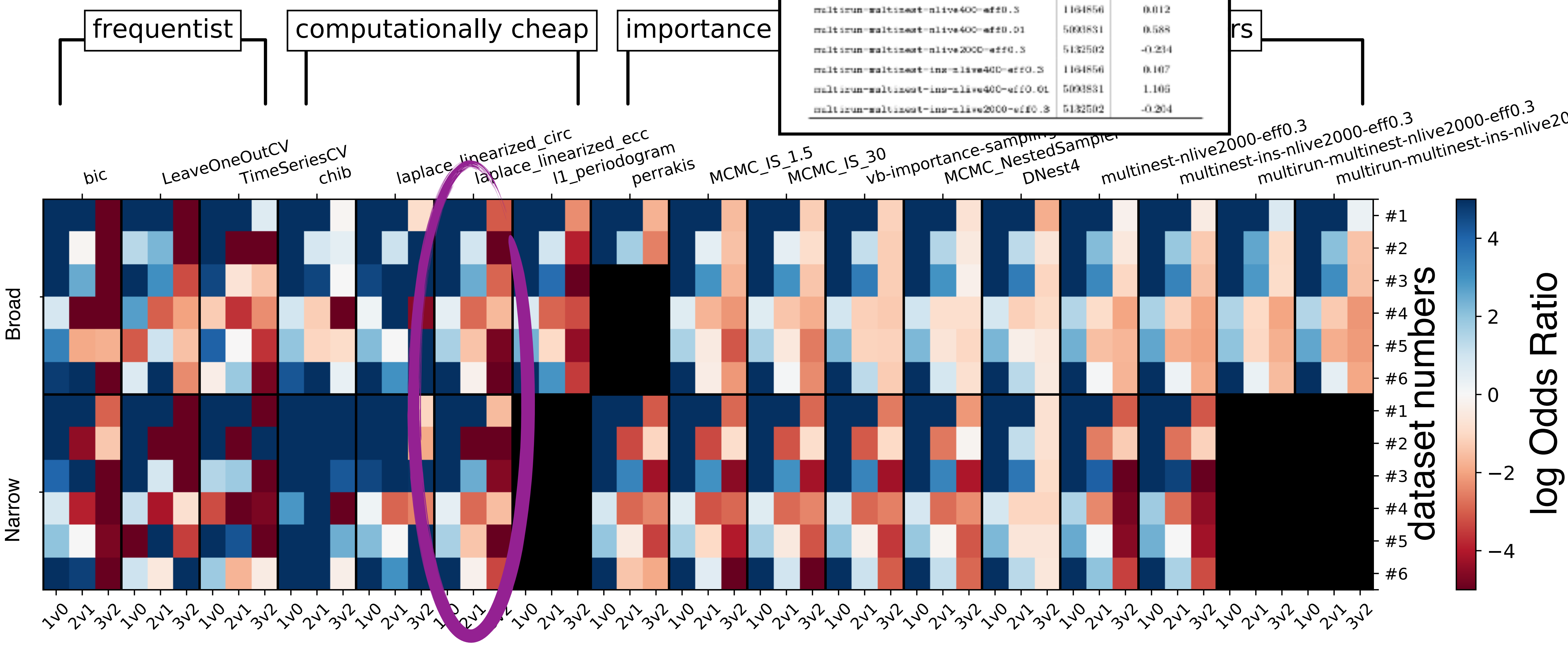
   Importance and nested sampling methods mostly arrive at the same conclusions.
   Cheaper methods are relatively overconfident in estimated odds ratios.

4. What methods should be recommended/avoided/further improved?

What different met … out n vs n-1 p…

Table 4. Number of likelihood evaluations ($n_{\mathcal{L}}$) reported to calculate $\log \hat{\mathcal{Z}}$ for Dataset 2 and $\mathcal{M}_2$, assuming broad period priors. Similar methods with different tuning parameters or simplifying assumptions are grouped together. The median $\log \hat{\mathcal{Z}}$ for this set of methods is −195.005.

Method (directory name)	$n_{\mathcal{L}}$	$\log \hat{\mathcal{Z}} - \langle \log \hat{\mathcal{Z}} \rangle$
chib	1000000	-0.342
laplace_linearized_circ	264	1.012
laplace_linearized_ecc	319	-0.128
vb-importance-sampling	261979	-0.449
vb-importance-sampling-long	2883983	-0.012
MCMC_NestedSampler	8814939	0.062
multinest-nlive400-eff0.3	173460	-0.516
multinest-nlive400-eff0.01	768868	0.551
multinest-nlive2000-eff0.3	1017587	-0.578
multinest-ins-nlive400-eff0.3	173460	0.018
multinest-ins-nlive400-eff0.01	768868	0.984
multinest-ins-nlive2000-eff0.3	1017587	-0.34
multirun-multinest-nlive400-eff0.3	1164856	0.012
multirun-multinest-nlive400-eff0.01	5093831	0.588
multirun-multinest-nlive2000-eff0.3	5132502	-0.234
multirun-multinest-ins-nlive400-eff0.3	1164856	0.107
multirun-multinest-ins-nlive400-eff0.01	5093831	1.106
multirun-multinest-ins-nlive2000-eff0.3	5132502	-0.204



(n)v(n-1), i.e., $\frac{p(\mathcal{M}_n|d)}{p(\mathcal{M}_{n-1}|d)}$

Four Questions of the Evidence Challenge

1. What is the dispersion in $\mathcal{Z}$?

   model: 0 planets ----> 1 planet ----> 2 planets ----> 3 planets
   dispersion: ~few ~OOM ~few to several OOMs

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

   Internal estimates of one evidence calculation underestimate the uncertainty.
   Monte Carlo methods seem to provide reasonable uncertainties.

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

   Importance and nested sampling methods mostly arrive at the same conclusions.
   Cheaper methods are relatively overconfident in estimated odds ratios.

4. What methods should be recommended/avoided/further improved?

   Recommended: Laplace approximation (for large datasets or complex likelihoods)
   Recommended/further improved: most of the numerical methods

Evidence Challenge Conclusions and Links

1. What is the dispersion in $\mathcal{Z}$?

   model: 0 planets ----> 1 planet ----> 2 planets ----> 3 planets
   dispersion: ~few ~OOM ~few to several OOMs

2. Does the uncertainty in $\mathcal{Z}$ accurately reflect the observed dispersion?

   Internal estimates of one evidence calculation underestimate the uncertainty.
   Monte Carlo methods seem to provide reasonable uncertainties.

3. How does (1) and (2) affect our ability to favor n vs n-1 planets?

   Importance and nested sampling methods mostly arrive at the same conclusions.
   Cheaper methods are relatively overconfident in estimated odds ratios.

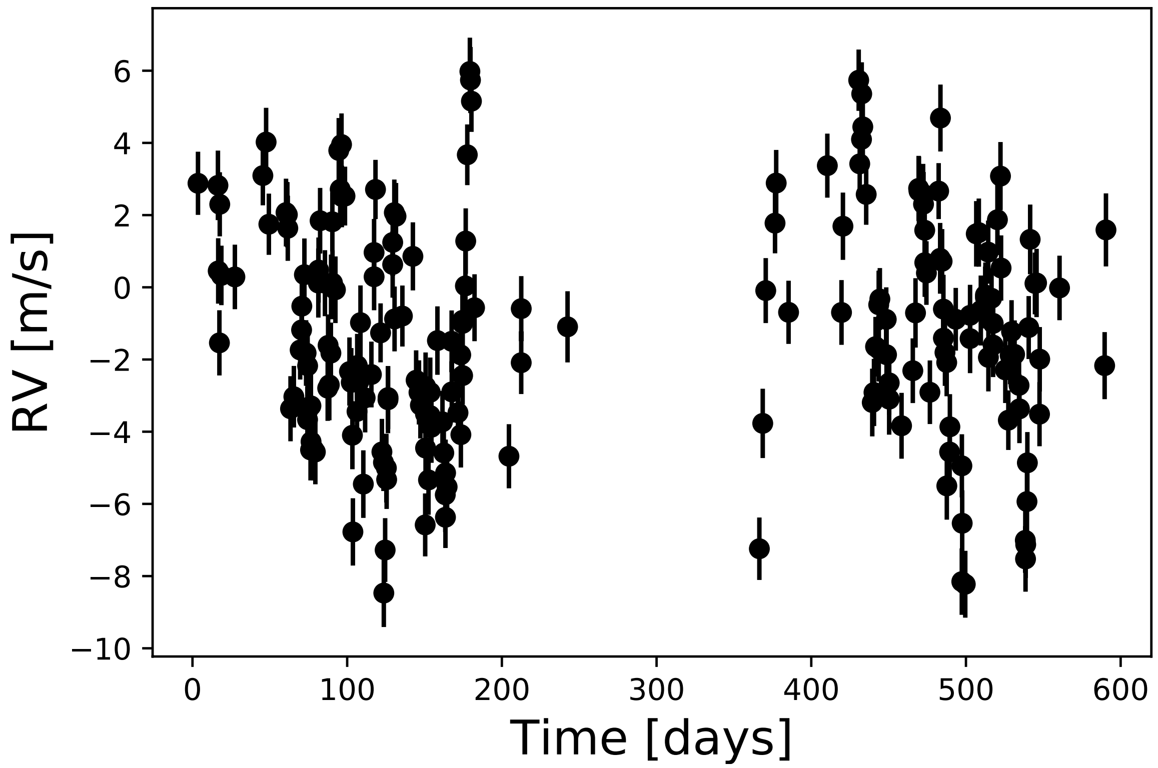
4. What methods should be recommended/avoided/further improved?

   Recommended: Laplace approximation (for large datasets or complex likelihoods)
   Recommended/further improved: most of the numerical methods

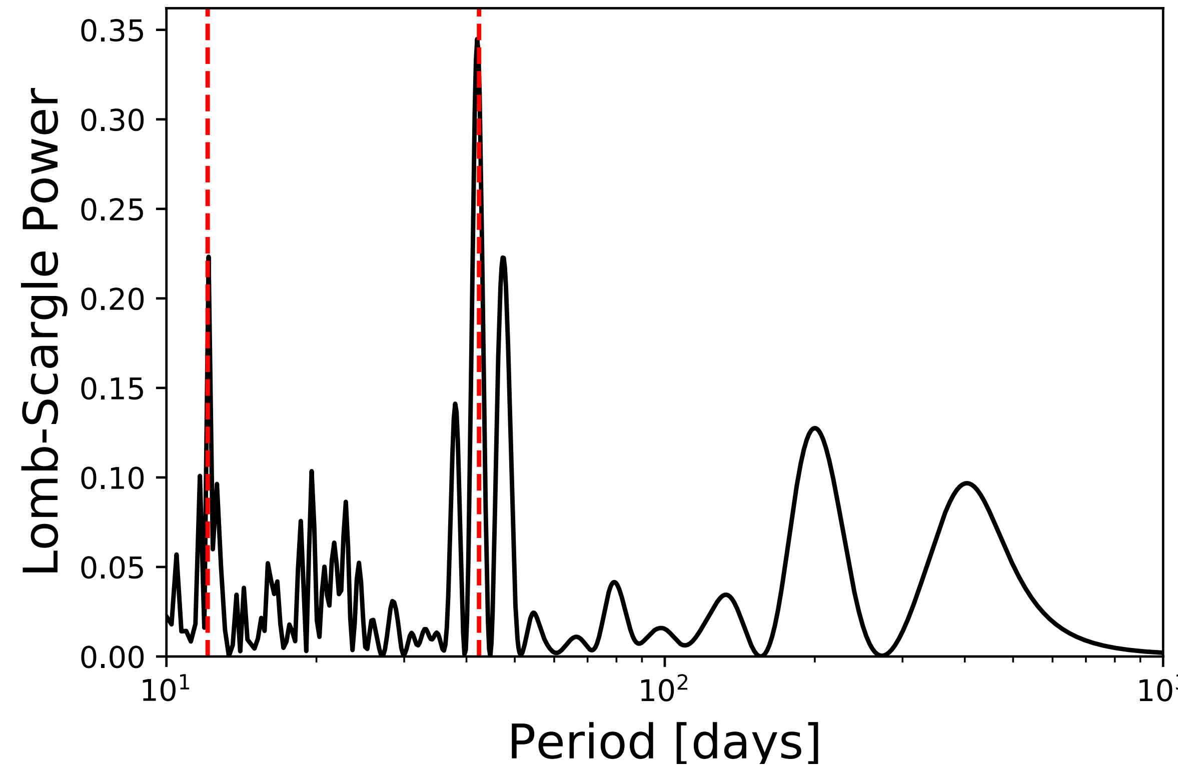
arXiv:1806.04683 github.com/EPRV3EvidenceChallenge

Supplemental Slides

Dataset #1

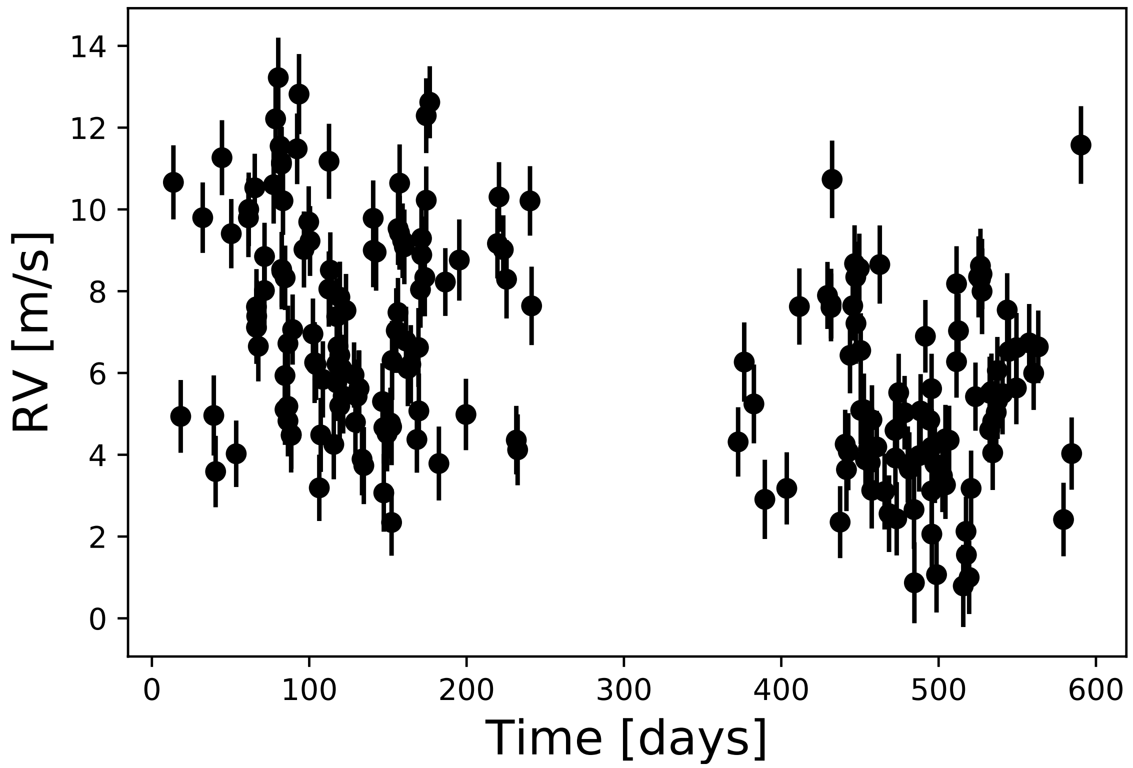


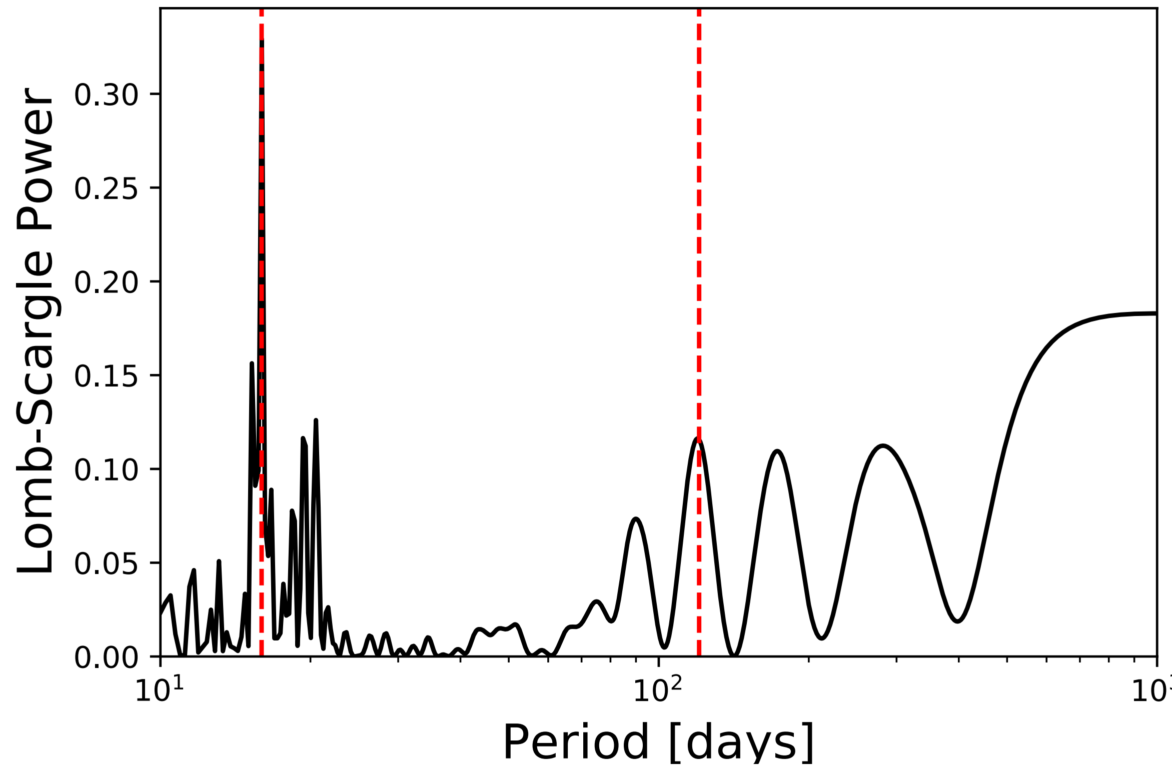
RV [m/s]
6
4
2
0
−2
−4
−6
−8
−10
0
100
200
300
400
500
600
Time [days]



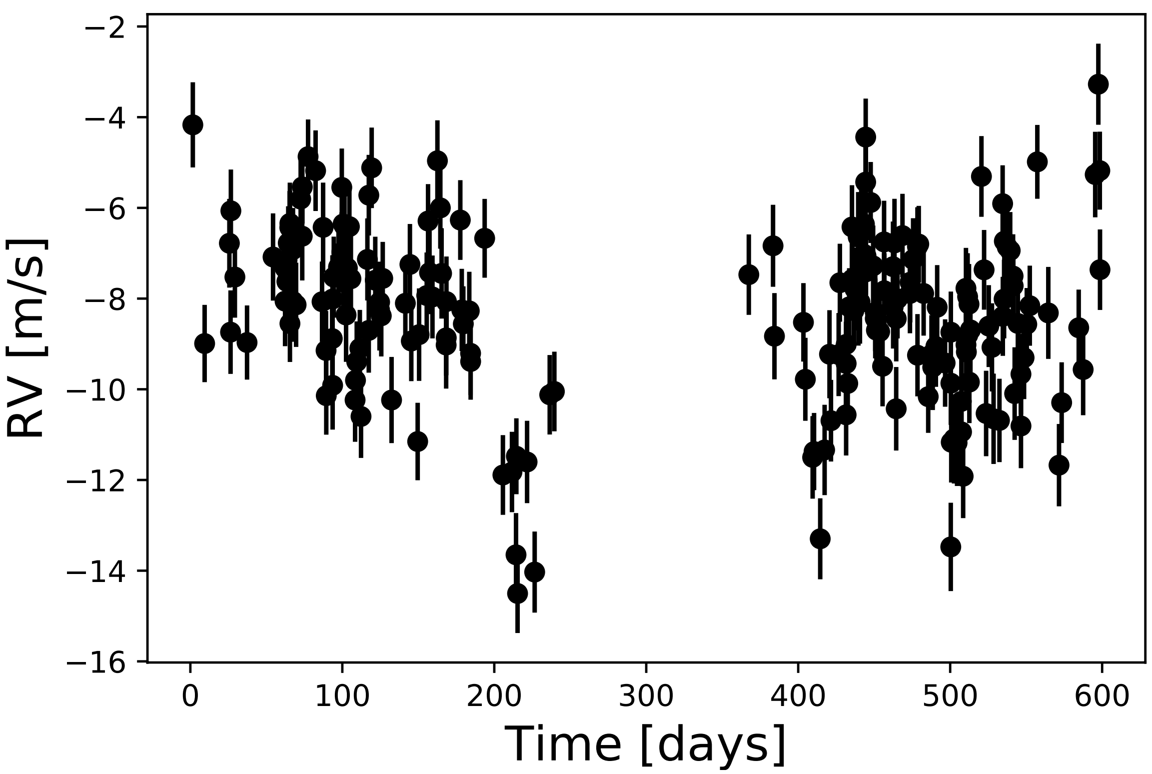
Lomb-Scargle Power
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
10^1
10^2
10^3
Period [days]

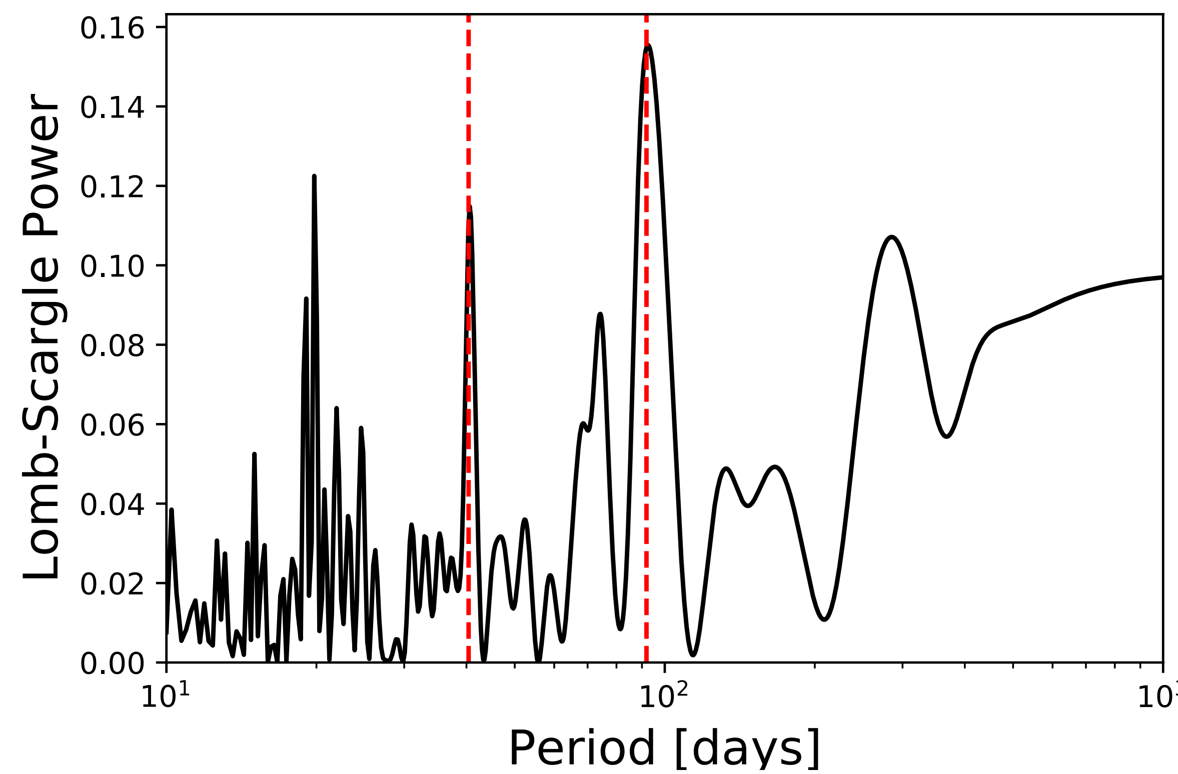
Dataset #2


RV [m/s]
Time [days]
0
2
4
6
8
10
12
14
0
100
200
300
400
500
600

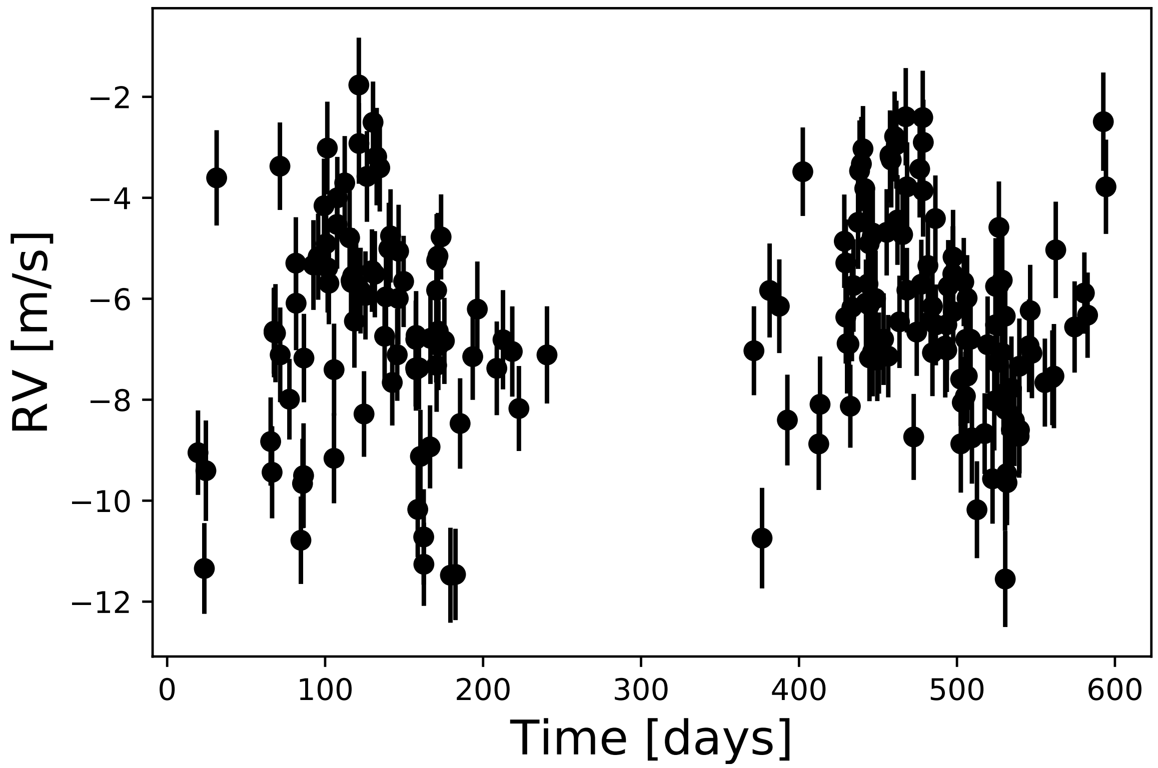

Lomb-Scargle Power
Period [days]
0.00
0.05
0.10
0.15
0.20
0.25
0.30
10^1
10^2
10^3

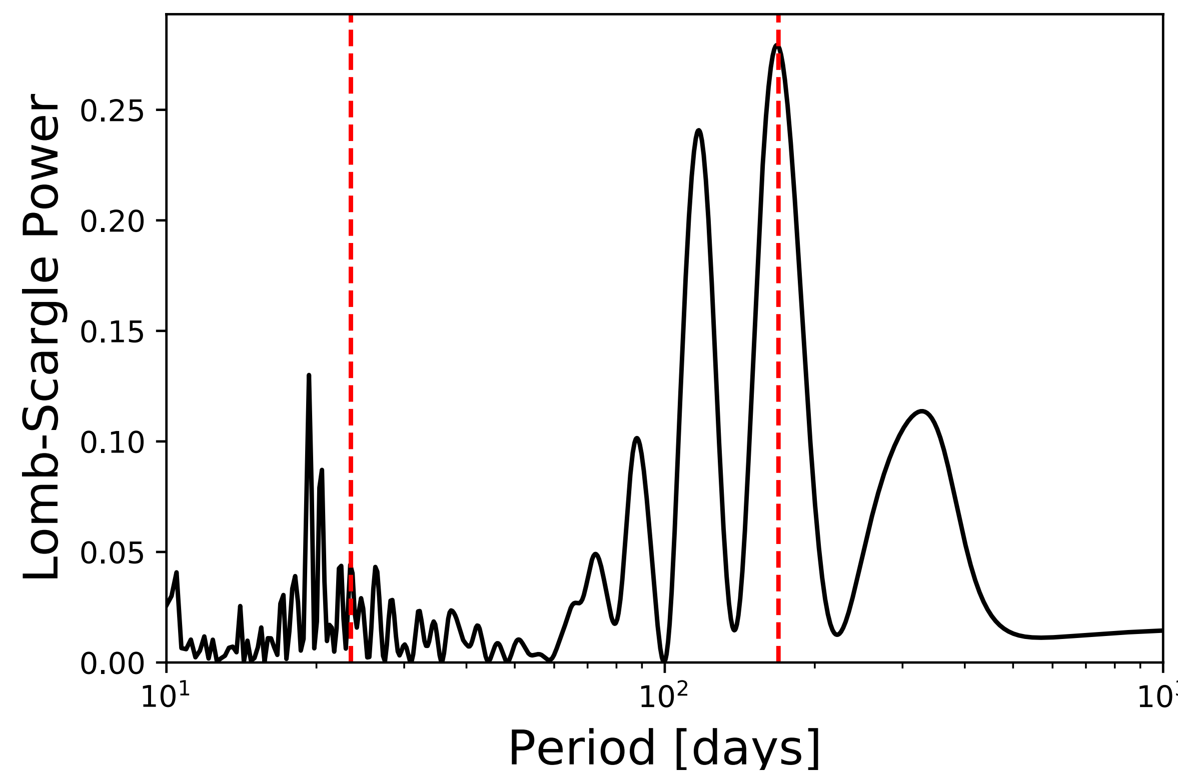
Dataset #3


RV [m/s]
Time [days]
−2
−4
−6
−8
−10
−12
−14
−16
0
100
200
300
400
500
600

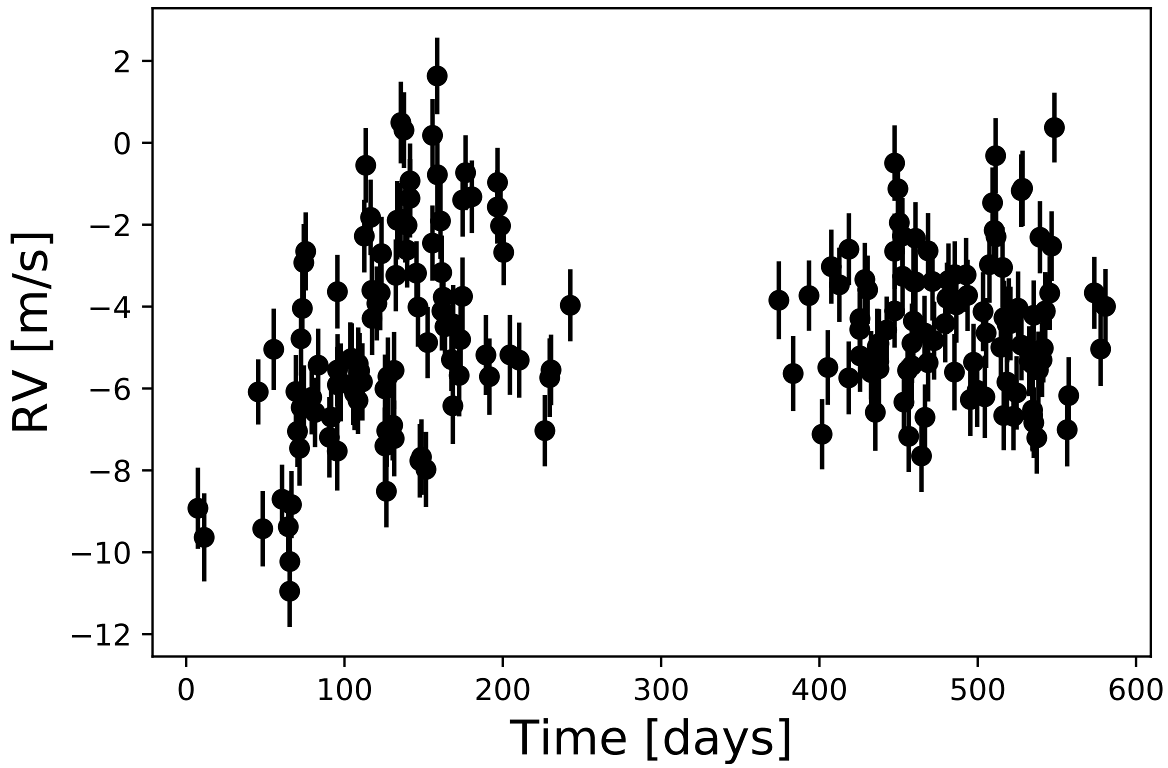

Lomb-Scargle Power
Period [days]
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
10^1
10^2
10^3

Dataset #4

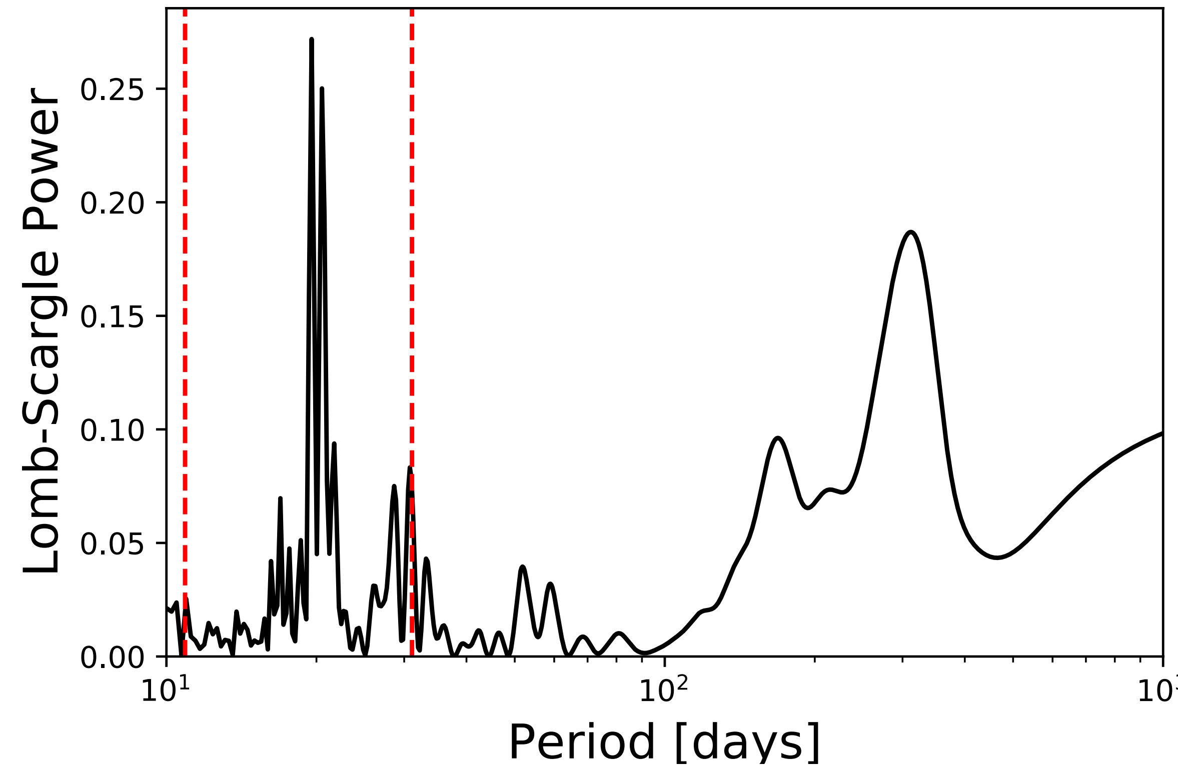

RV [m/s]
−2
−4
−6
−8
−10
−12
0
100
200
300
400
500
600
Time [days]


Lomb-Scargle Power
0.25
0.20
0.15
0.10
0.05
0.00
10^1
10^2
10^3
Period [days]

Dataset #5

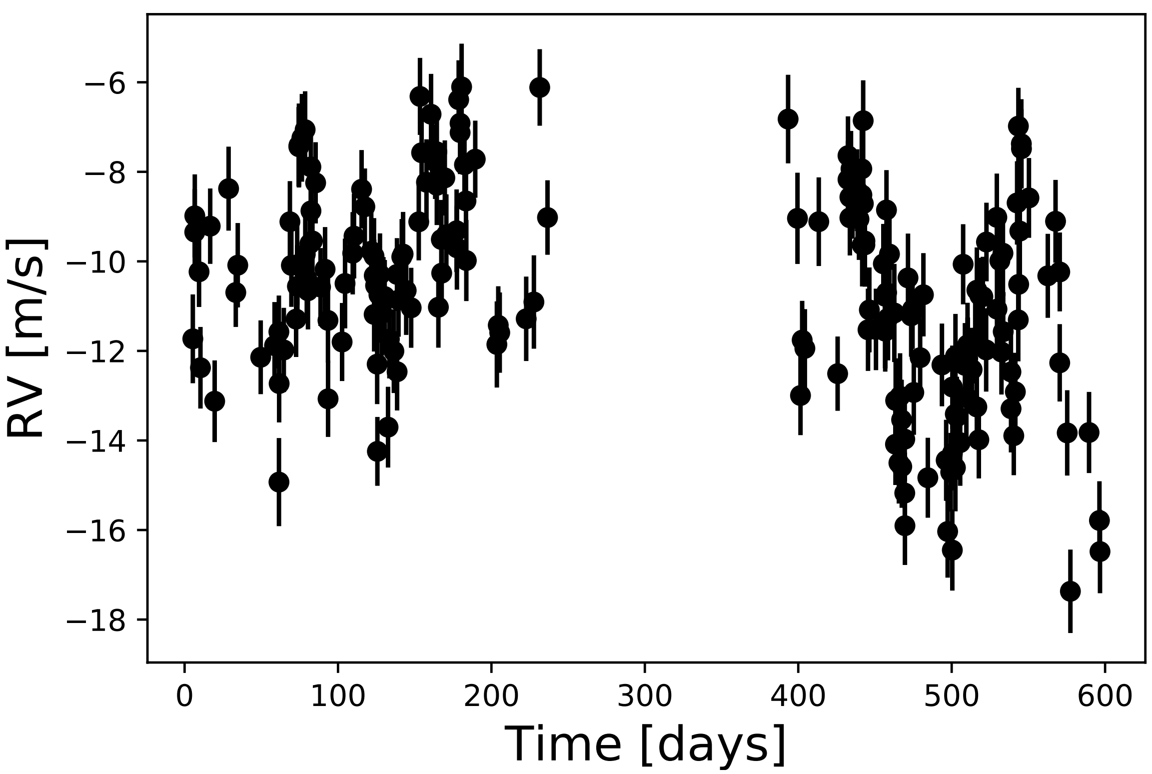


RV [m/s]
2
0
−2
−4
−6
−8
−10
−12
0
100
200
300
400
500
600
Time [days]

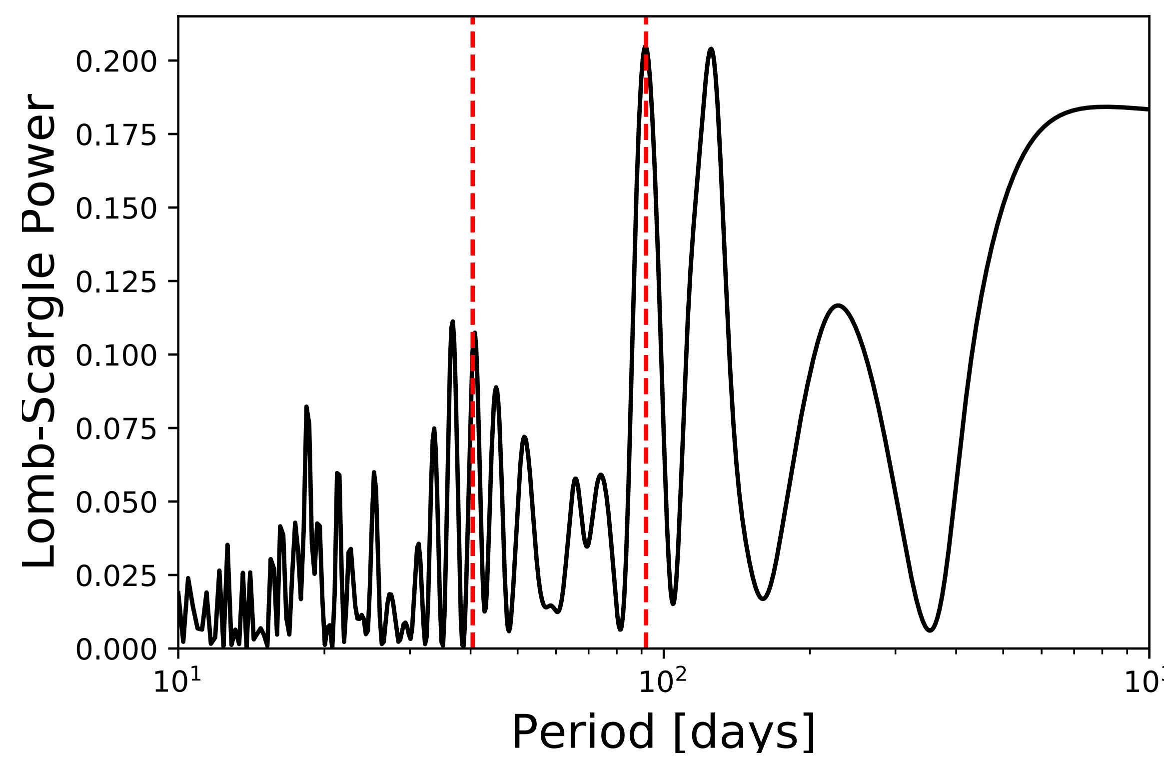


Lomb-Scargle Power
0.25
0.20
0.15
0.10
0.05
0.00
10^1
10^2
10^3
Period [days]

Dataset #6

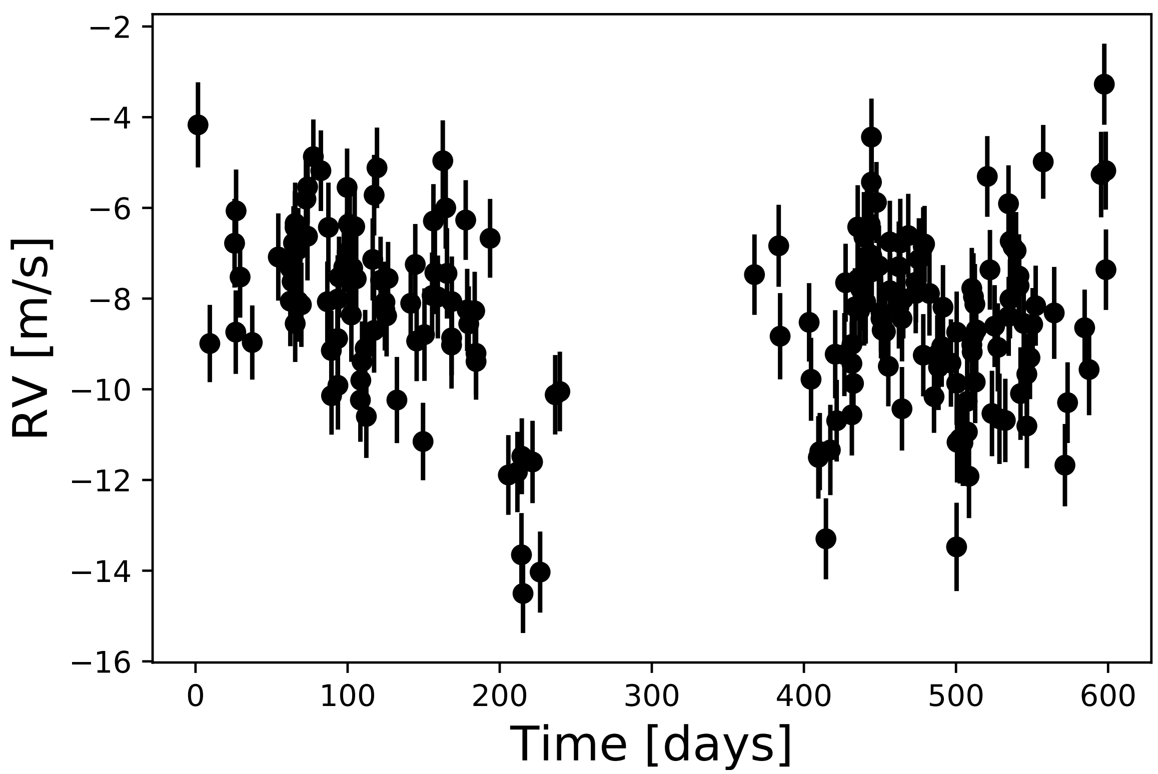


RV [m/s]
Time [days]
−6
−8
−10
−12
−14
−16
−18
0
100
200
300
400
500
600

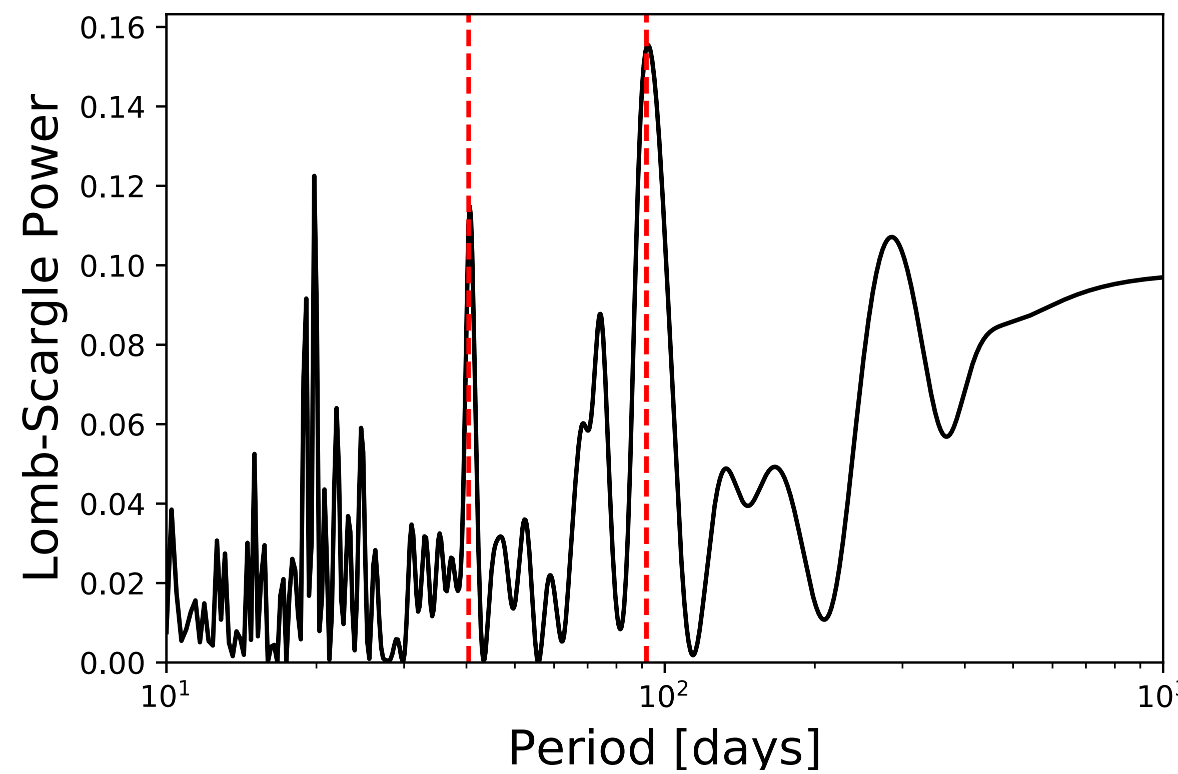


Lomb-Scargle Power
Period [days]
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
10^1
10^2
10^3

Dataset #3

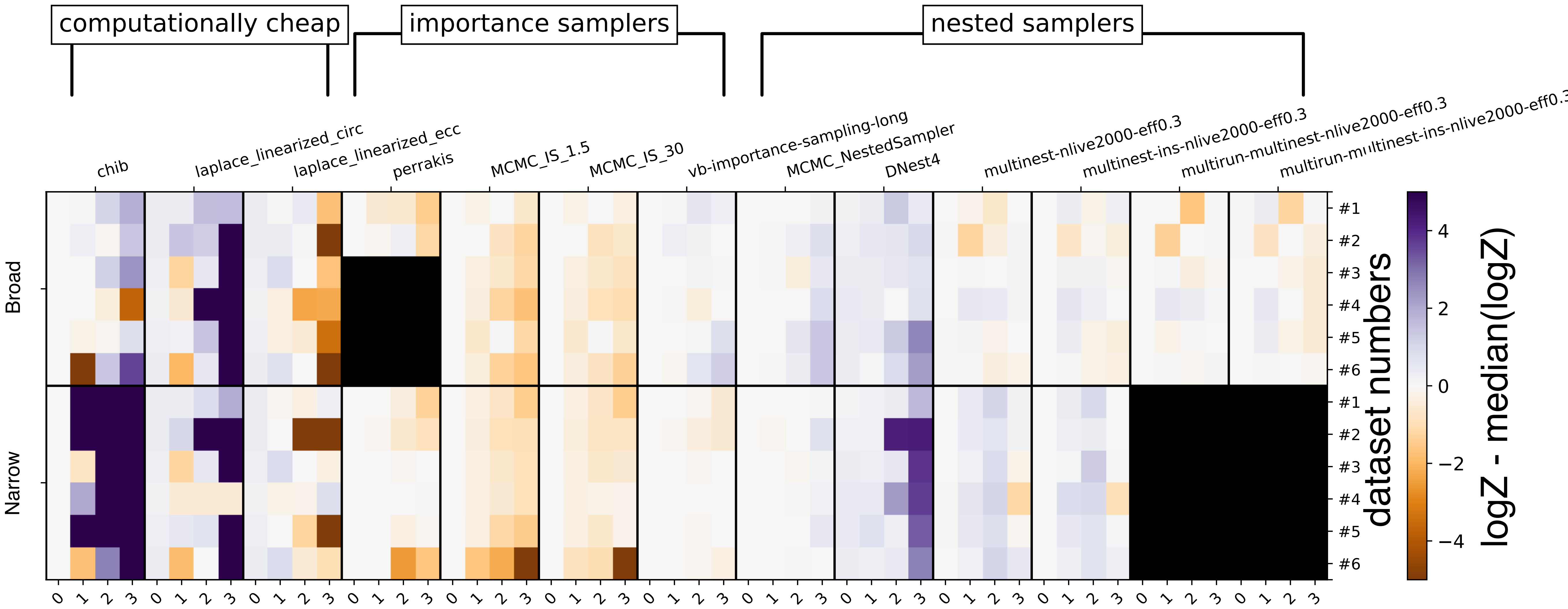


RV [m/s]
−2
−4
−6
−8
−10
−12
−14
−16
0
100
200
300
400
500
600
Time [days]



Lomb-Scargle Power
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
10^1
10^2
10^3
Period [days]

What different methods say about the evidence estimate



July 24, 2018
Sagan Workshop

Efficiently computing the FML for thousands of datasets

Systematic mischaracterization of exoplanetary system dynamical histories from a model degeneracy near mean-motion resonance

John H. Boisvert, Benjamin E. Nelson, Jason H. Steffen

(Submitted on 26 Apr 2018)

There is a degeneracy in the radial velocity exoplanet signal between a single planet on an eccentric orbit and a two-planet system with a period ratio of 2:1. This degeneracy could lead to misunderstandings of the dynamical histories of planetary systems as well as measurements of planetary abundances if the correct architecture is not established. We constrain the rate of mischaracterization by analyzing a sample of 60 non-transiting, radial velocity systems orbiting main sequence stars from the NASA Exoplanet Archive (NASA Archive) using a new Bayesian model comparison pipeline. We find that 15 systems (25% of our sample) show compelling evidence for the two-planet case with a confidence level of 95%.

Comments: 4 pages, 4 figures, and a two-page appendix containing 2 figures
Subjects: **Earth and Planetary Astrophysics (astro-ph.EP)**
Cite as: arXiv:1804.10143 [astro-ph.EP]
(or arXiv:1804.10143v1 [astro-ph.EP] for this version)

arXiv:1804.10143

The Laplace Approximation

Want to solve...

$$\int dx \exp[f(x)]$$

Taylor expand around x_o, the location of the global mode...

$$f(x) \simeq f(x_o) + \frac{1}{2} \sum_{a,b} \frac{\partial^2 f}{\partial x_a \partial x_b} (x - x_o)^2$$

Approximate integral as...

$$\int dx \exp[f(x)] \simeq \left[\frac{(2\pi)^2}{|\det H(x_o)|} \right]^{1/2} \exp[f(x_o)]$$