Practical Hierarchical Bayesian Modeling

Megan Shabram

September 21, 2017

LSSTC DSFP Session 4

What is your Science Question? Things to Keep in mind

- Iterative approach
- Awareness of probability distributions and When they are applicable.
- What has nature told you about the physical process you want to investigate?
- What evidence are you looking for?
- Pick a model that performs well if nature is more complex (concerns model misspecification)
  - E.g., using Beta instead of Normal for the eccentricity distribution

One- Two and Three- component Gaussian Mixture Model

- Exchangeability
- Separable distributions

Model Complexity vs. Size of Dataset

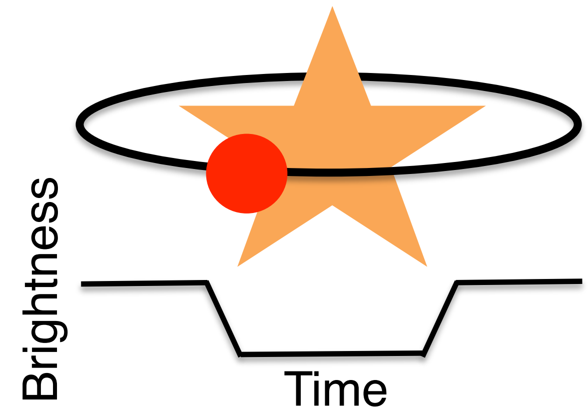
- Statistical significant vs. prediction. Variables that are statistically significant may be bad predictors and vice versa.
- Under- and over- fitting are issues for predictions.
- Solution: Find a model that is relatively insensitive to the generative model.
- Models may converge in less iterations for larger datasets with precise measurements

Measurement Uncertainty

- Is it normally distributed?
- Do you have the full posterior or a summary statistic?

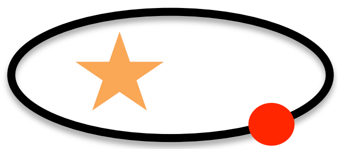
Advancements in **Exoplanet Science**

- Kepler Mission:
  - **~ 5000** exoplanet candidates discovered
  - **~ 3000** Eclipsing Binaries

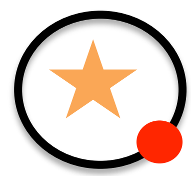


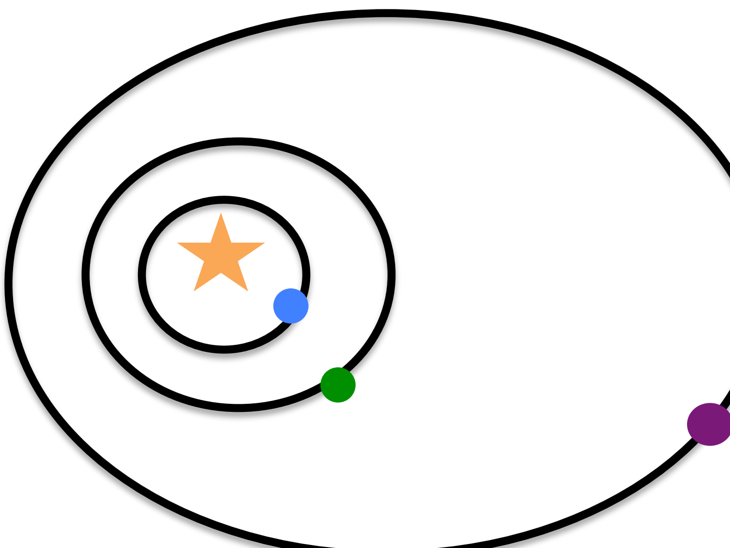
Converging on **Planet Formation** from *Multiple Paradigms*

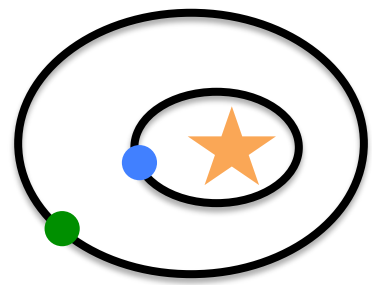
Connecting Short-period Tightly-packed Inner Planetary Systems (STIPS) formation to gas-giant planet formation.



Detection of “hot-Jupiters” demanded generalizing planet formation theories.







“Sub-Neptunes” are dominant result of planet formation. Theories must be expanded further.

What can we learn from population studies of exoplanets?

- **More than one population** in the eccentricity distribution would support more than one mode of planet formation, e.g.,
  - planet-planet scattering
  - tidal circularization
  - smooth disk migration
  - Resonant repulsion
  - Planet-planetesimal disk interactions (Chatterjee & Ford 2014)
- **Correlations** of eccentricity may expose a “tunable” generalized planet formation theory, e.g.,
  - host star metallicity
  - host star effective temperature
  - planet radius
  - orbital period

(Shabram et al. 2015, accepted to ApJ)

What can we learn from population studies?

- The mass-radius relation probes whether planets:
  1. formed **in situ** and may be is composed mostly of iron and silicates with a hydrogen-helium abundant atmosphere.
  2. formed **beyond the snow line** and migrated closer to host star.
  3. has undergone episodes of **mass gain** and/or **mass loss**.

- Complicated since **multiple compositional regimes** can be consistent with a given mass and radius of an exoplanet.

(Rogers & Seager, 2010; Valencia et al., 2013; Lopez & Fortney, 2014)

Limitations of Studying the *Kepler* Exoplanet Population

- Geometrical bias, selection effects
- Planet detection efficiency (completeness)
- Measurement uncertainty for planet and host star properties
- Is sample representative?
- Potential false positives among planet candidates

Challenges for Robust population Inference

- Include **measurement uncertainties** (in population studies).
- Include **non-detections** to minimize **sampling bias** (e.g., case study 2, the mass-radius-eccentricity relationship).
- Include **posterior distributions** in population studies.
- Incorporate false positive rate and detection efficiency.

Hierarchical Bayesian Modeling Addresses these Challenges

- **Incorporate uncertainty** unique to each observation into the inference population parameters.
- Address impact of **selection effects** and **sample bias**.
- Use information about all objects to improve inferences about each object (“pool and muster strength”).
- Test sensitivity to **model misspecification**.

Case Study 1: The Eccentricity Distribution of Short-Period *Kepler* Planet-Candidates and Eclipsing Binaries

- How common are short-period giant planets with eccentric orbits vs. circular orbits?
  - Smooth disk migration
  - High-eccentricity migration
  - Planet-planet scattering
  - Kozai interactions
  - Secular processes followed by tidal circularization
- How about for Eclipsing Binaries?

Case Study 1: The Eccentricity Distribution of Short-Period *Kepler*-Planet-Candidates and Eclipsing Binaries

- Observables:

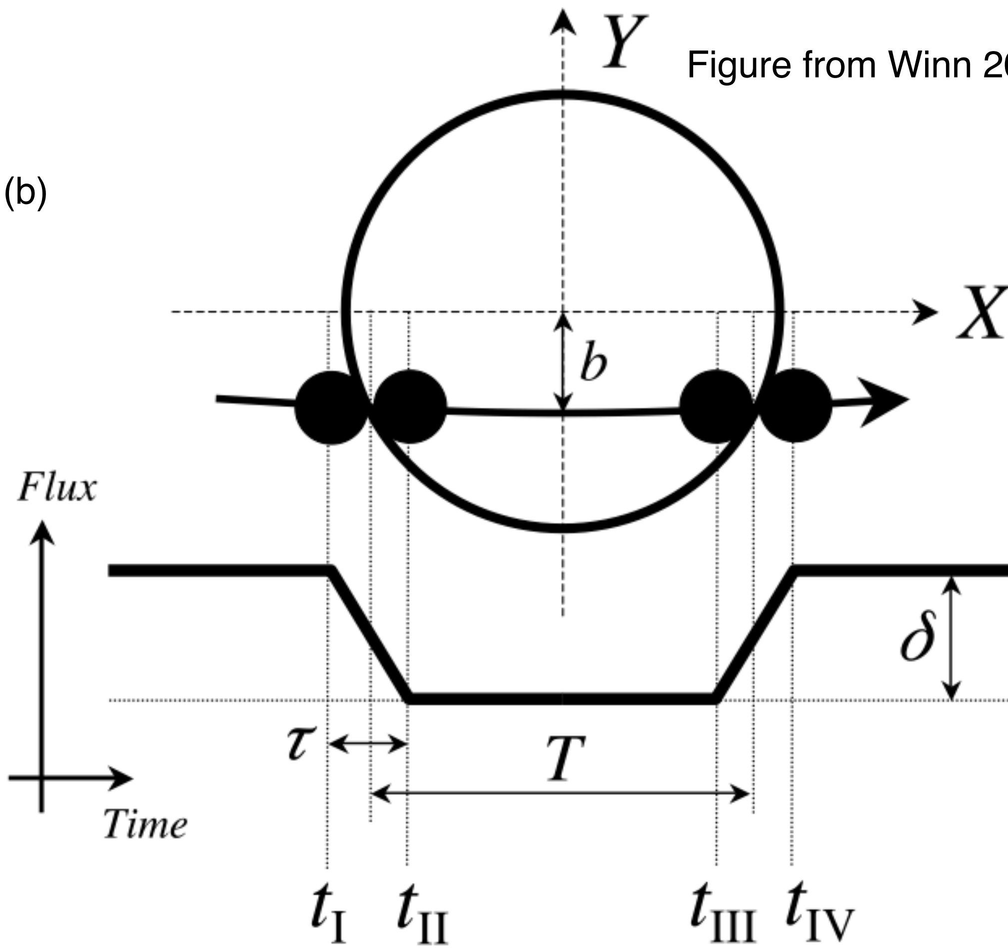
Phase offset (a) and transit duration ratio (b) relate transit observables to "projected eccentricity," *h* and *k*.

(a) $$\Delta t_c \approx \frac{P}{2}\left[1 + \frac{4}{\pi}h\right]$$

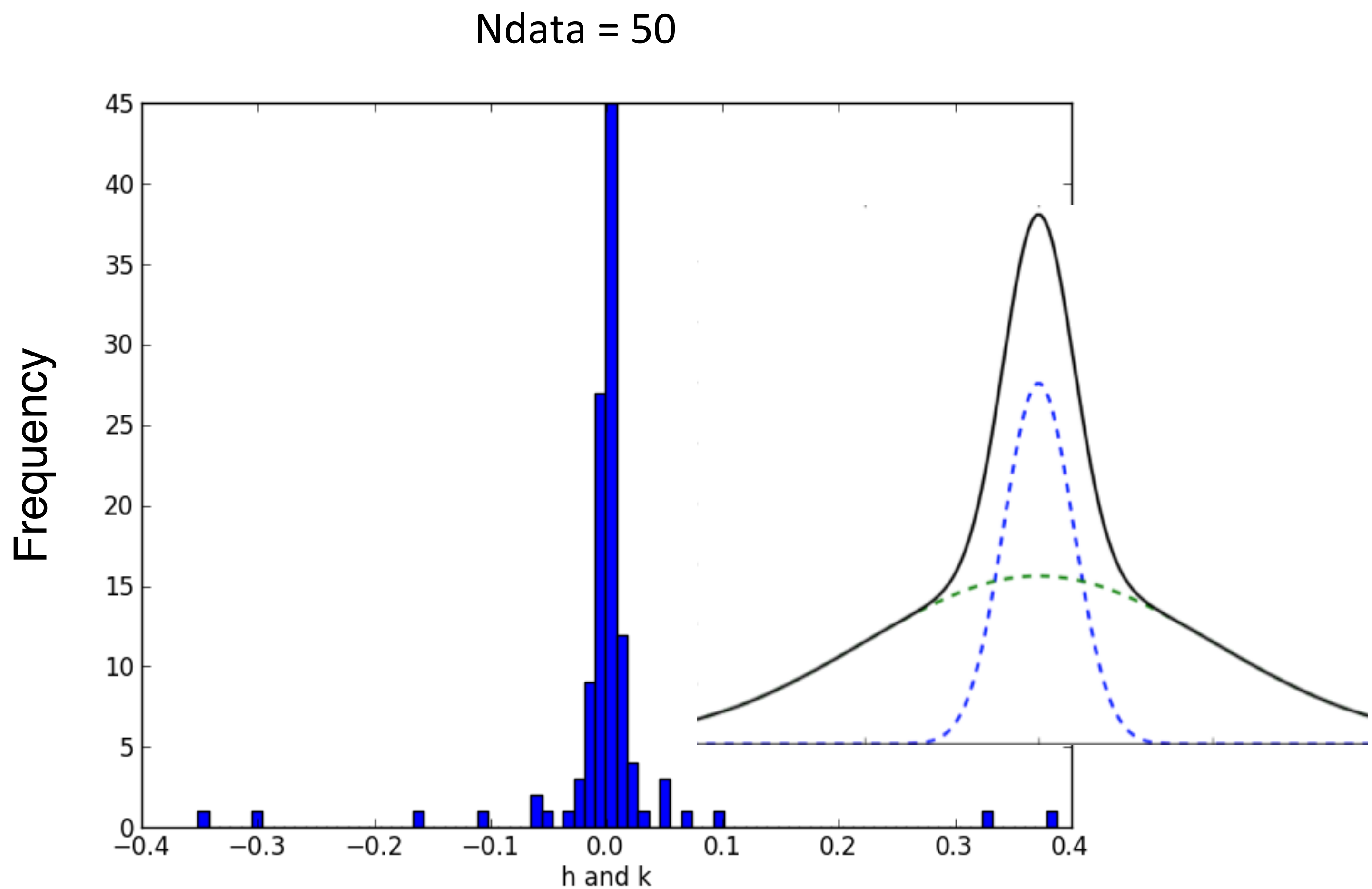
(b) $$\frac{T_{\rm occ}}{T_{\rm tra}} \approx \frac{1+k}{1-k}$$

P = orbital period
h = *e cos* ω
k = *e sin* ω

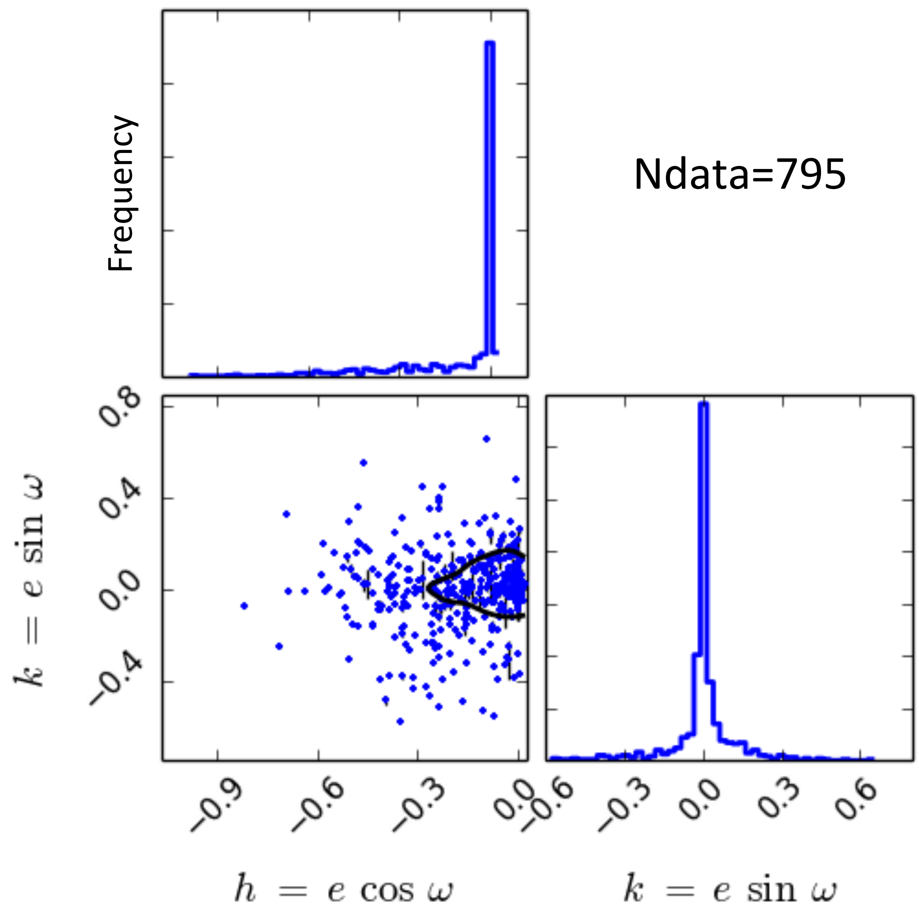
Figure from Winn 2010



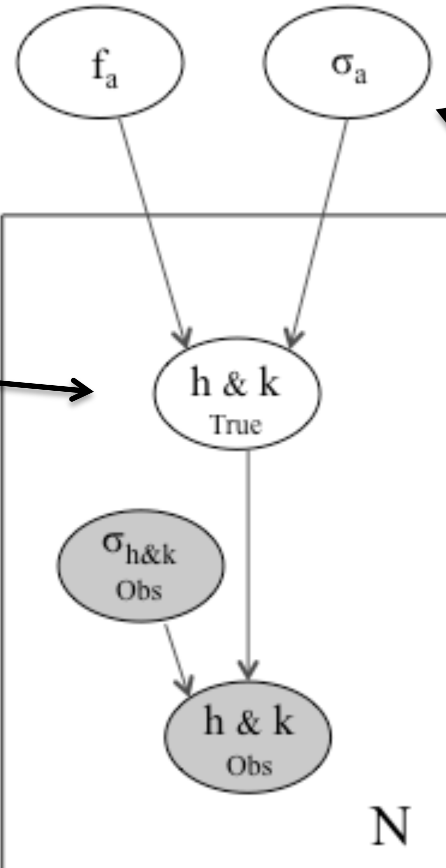
Input: Projected Eccentricity Measurements of Short-Period *Kepler*-Planet-Candidates



Input: Projected Eccentricity Measurements of Eclipsing Binaries from *Kepler*



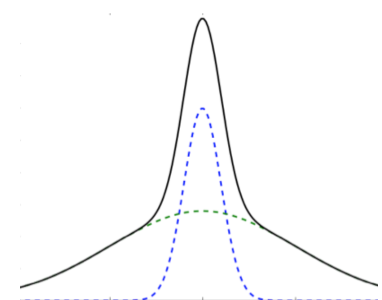
Case Study: The Eccentricity Distribution of Short-Period *Kepler*-Planet-Candidates and Eclipsing Binaries: **Graphical Model**



Population parameters (e.g., hyperparameters)

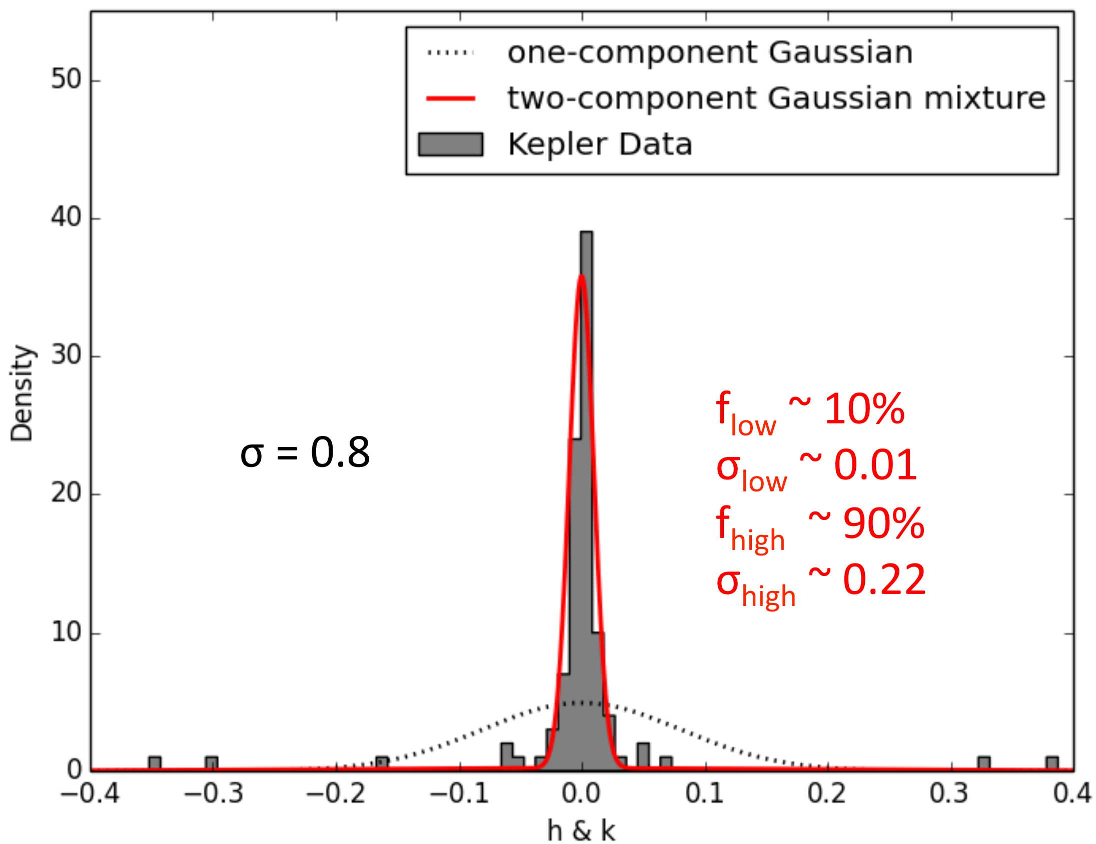
Latent variables

Measurements and measurement uncertainties



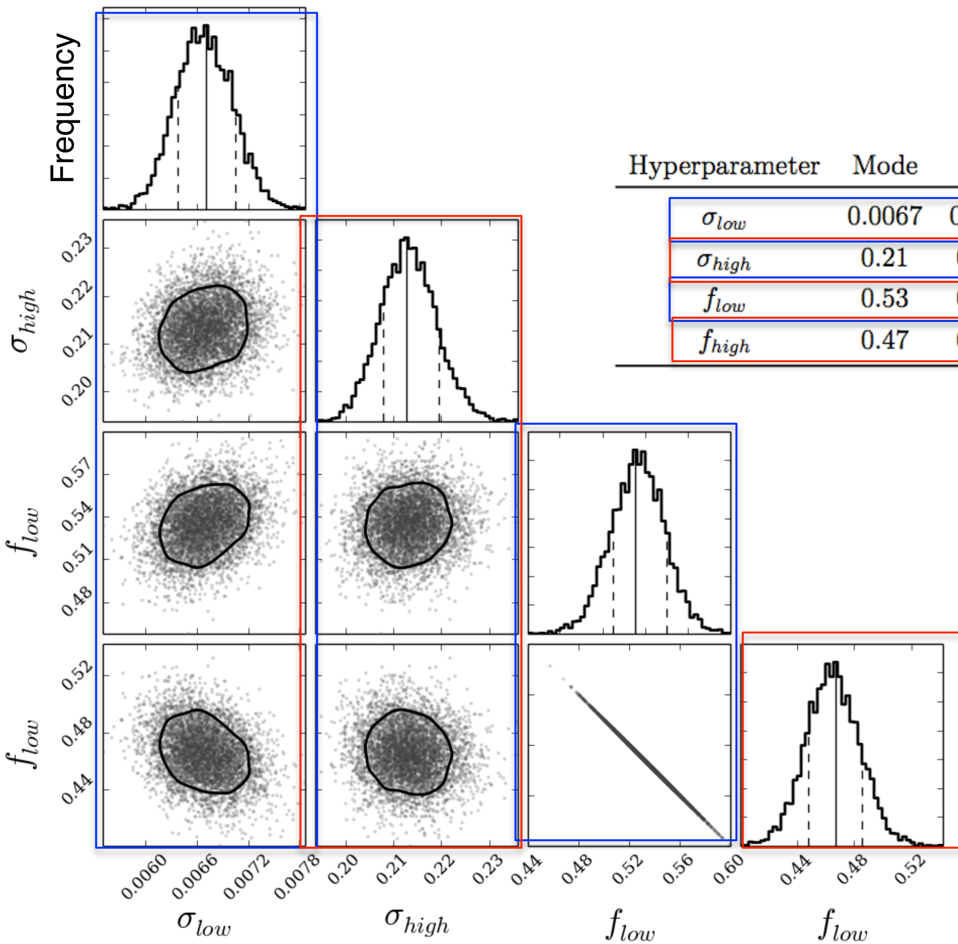
f_a => vector of mixture fractions
σ_a => vector of dispersions
$h\&k_{true}$ => true projected eccentricities
$h\&k_{obs}$ => measurements
$\sigma_{h\&k}$ => measurement uncertainties

Case Study 1: Short-Period *Kepler*-Planet-Candidates show evidence of at least two populations in the eccentricity distribution:



Shabram et al. 2016

Case Study 1: Eclipsing Binaries from *Kepler* show evidence of at least two populations in the eccentricity distribution:



Hyperparameter	Mode	68% credible interval +	68% credible interval −
σ_{low}	0.0067	0.0003	0.0003
σ_{high}	0.21	0.007	0.005
f_{low}	0.53	0.022	0.016
f_{high}	0.47	0.018	0.019

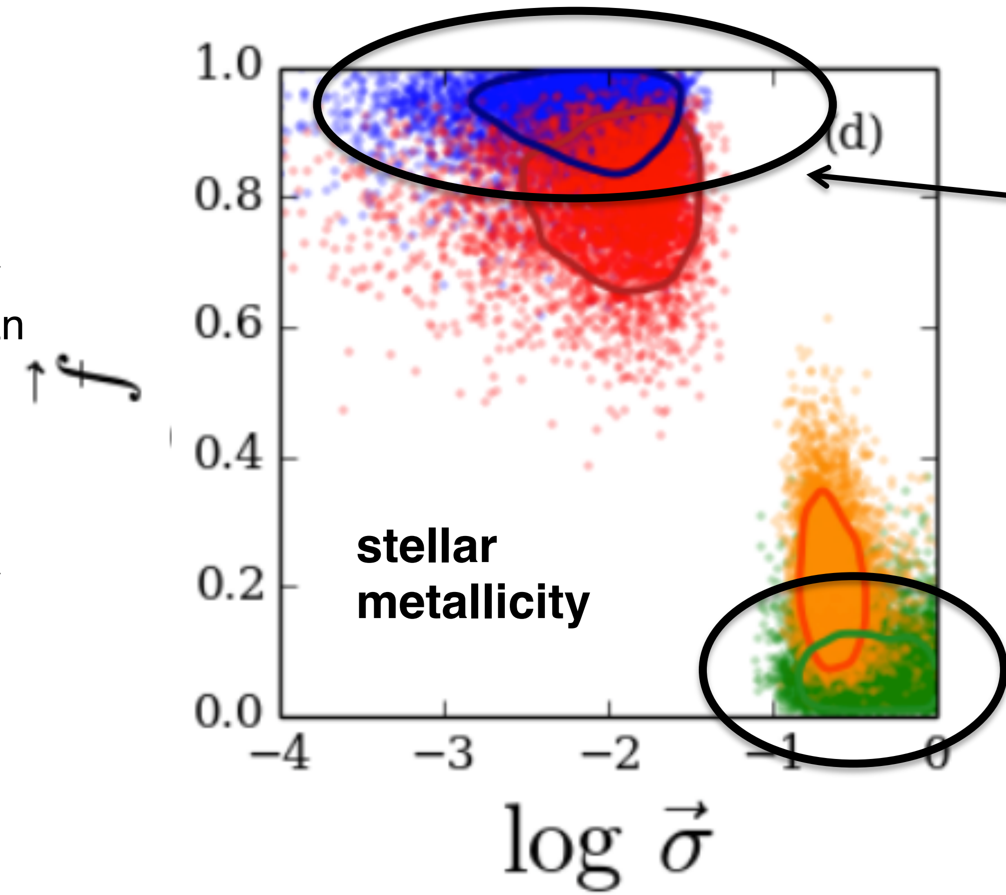
Precise measurements results in precise constraints on small eccentricity component

Shabram, Prsa et al. *in prep*

Case Study 1: Short-Period *Kepler*-Planet-Candidates: Potential correlation of eccentricity with **stellar metallicity** and **planet radius**

Blue and green represent posterior with stellar metallicity **less than** median

Red and orange represent posterior with stellar metallicity **greater than** median



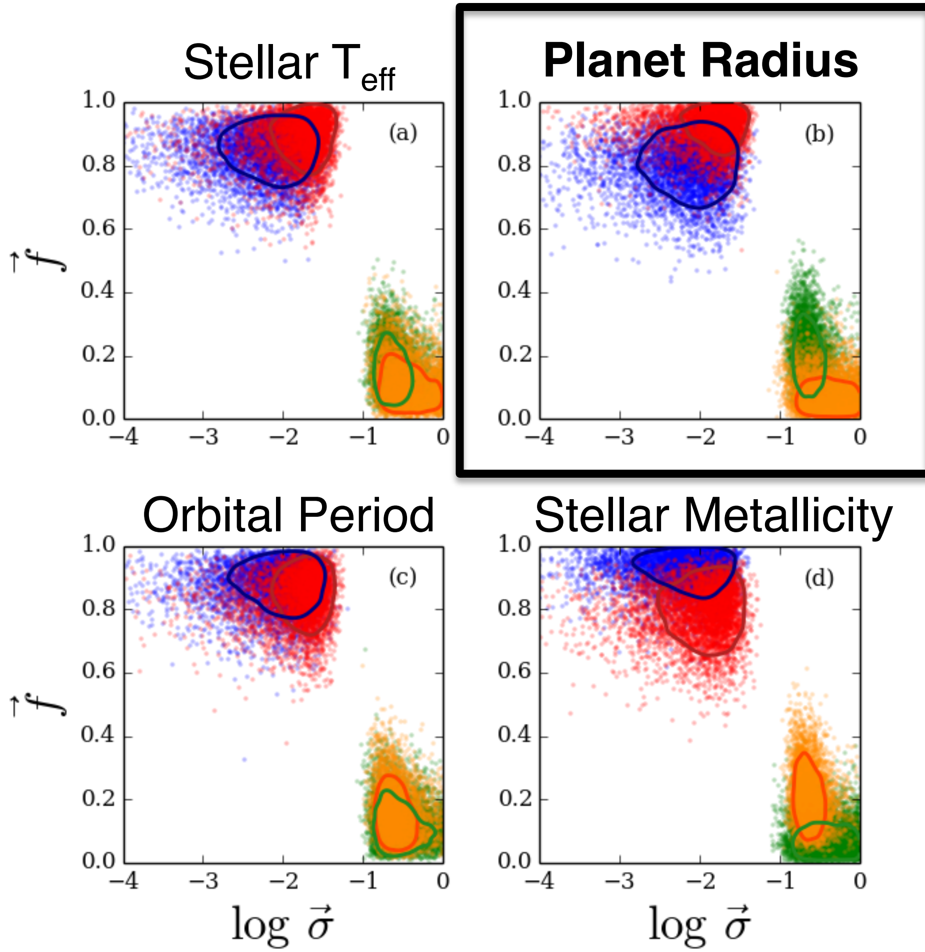
Bottom half of sample (blue and green) consistent with **zero and one**

Shabram et al. 2015 *accepted to ApJ*

Case Study 1: Short-Period *Kepler*-Planet-Candidates: Potential correlation of eccentricity with **stellar metallicity** and **planet radius**

Blue and green represent posterior with stellar metallicity **less than** median

Red and orange represent posterior with stellar metallicity **greater than** median



Similar effect for planet radius where large planet radii may be better characterized by a one-component model

Shabram et al. 2015 *accepted to ApJ*

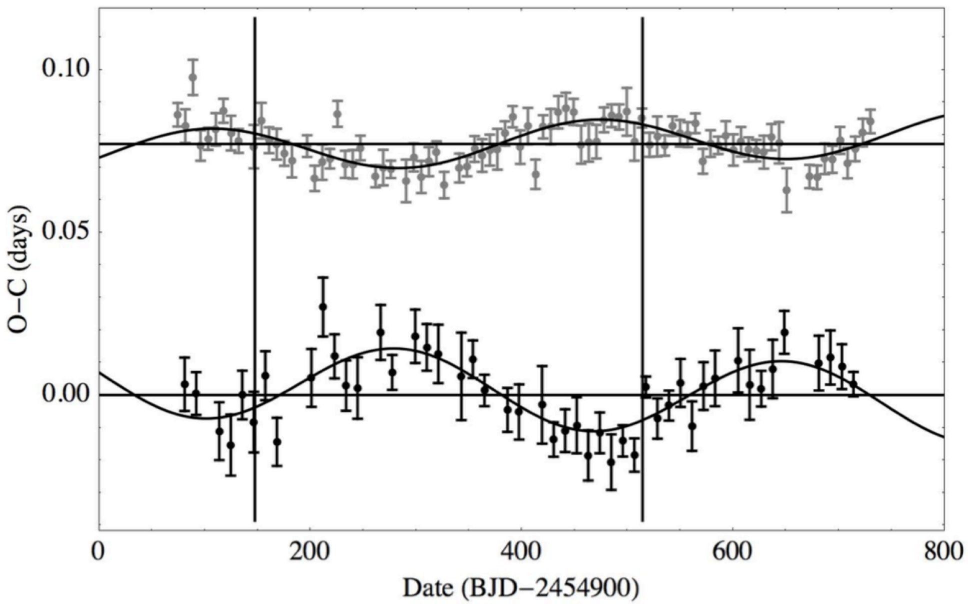
Case Study: The Eccentricity Distribution of Short-Period *Kepler*-Planet-Candidates and Eclipsing Binaries: **Potential Biases**

- Detection probability for occultation of planet-candidates.
- Sample dominated by objects at shorter orbital periods
- Potentially some diluted EBs present in the data set.
- Planet-candidate sample enriched with smaller-radii host stars, higher effective temperature (Teff) host stars, larger planet radii/Teff.

Case Study 2: The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs

- Empirically characterize the mass-radius relation
- Learn about the eccentricity of transiting exoplanet pairs near first order mean motion resonance
- Infer parameter estimates for masses and eccentricities using an analytical model that relates transit timing variation (TTV) amplitude to the the mass and eccentricity
- Allow for a scatter in the planet mass at a given radius.

Case Study 2: The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs - **Observables**



Analytical relation for the transit timing variation amplitude (Lithwick & Wu 2013):

Interior planet

$$TTV'_{amplitude} \sim \frac{M_{Planet}}{M_{Star}}\left(1+f(e,\omega)\right)$$

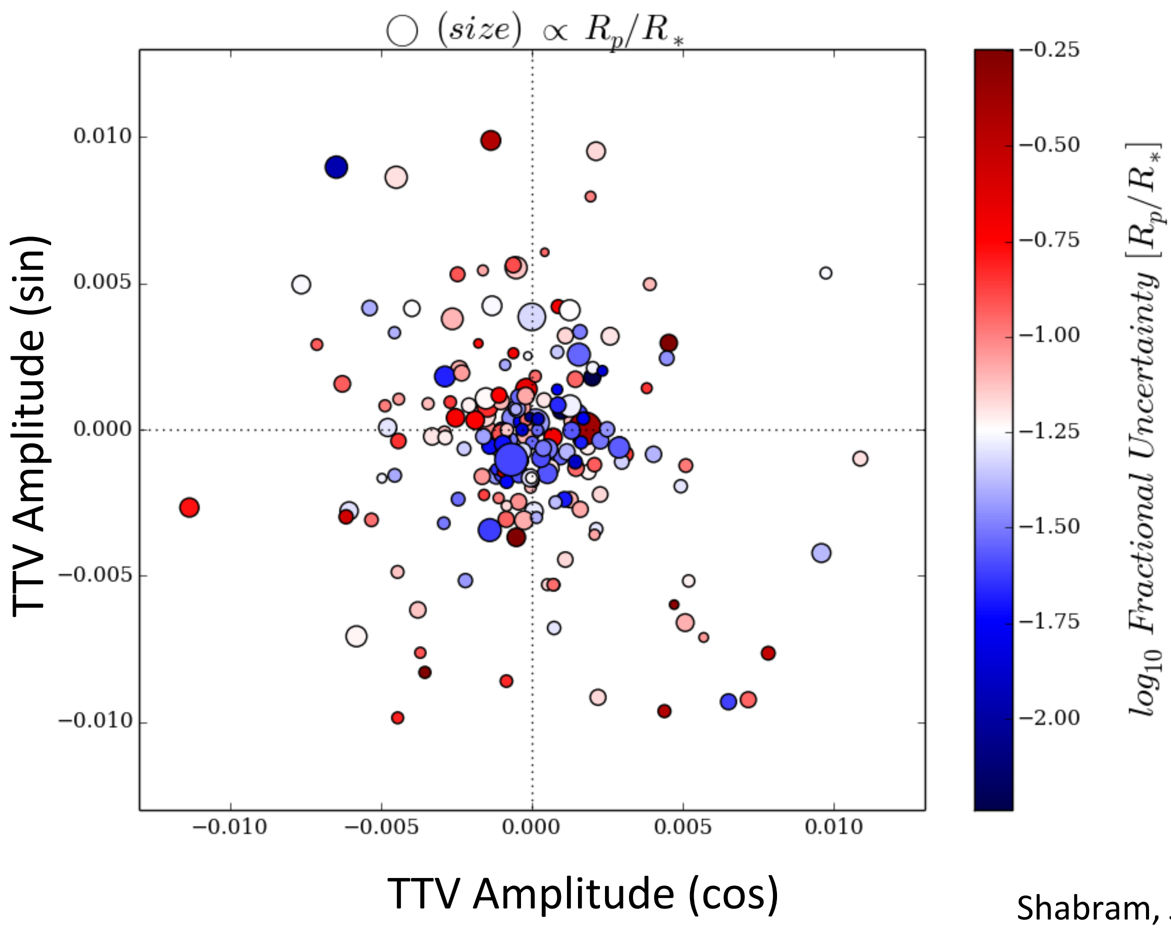
Exterior planet

$$TTV_{amplitude} \sim \frac{M'_{Planet}}{M_{Star}}\left(1+f(e,\omega)\right)$$

Figure from Steffen et. al. 2012

- Transit timing variations for planet pairs near a first order resonance are well approximated by a sinusoid.
- **Near degeneracy between the mass ratio and eccentricity** terms makes it difficult to constrain the mass for cases with non-zero eccentricity.

Input: The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs – *Kepler* data



Ndata = 120 planet pairs

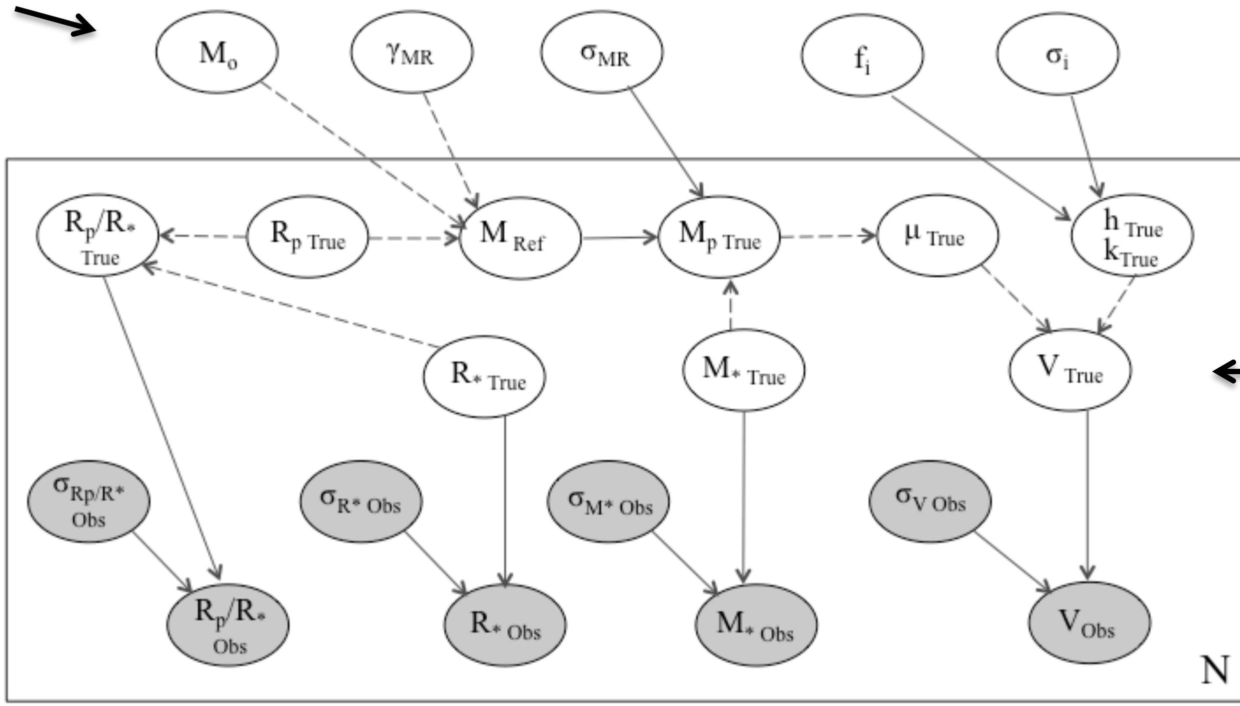
Small fractional uncertainties improve chance of constraining HB model

Shabram, Jontoff-Hutter, Ford *in prep*

Case Study 2: The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs

Graphical Model of Bayesian Network:

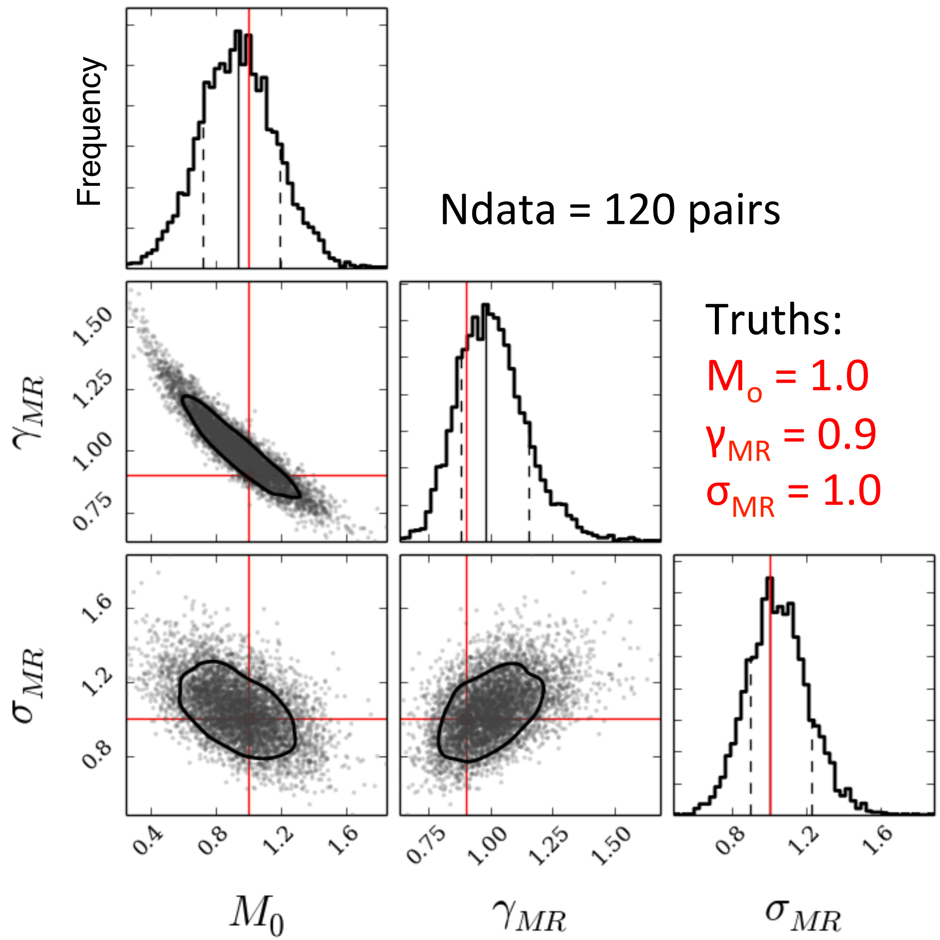
Population level parameters
(e.g., hyperparameters)



Latent variables

Gray = measurements and measurement uncertainties

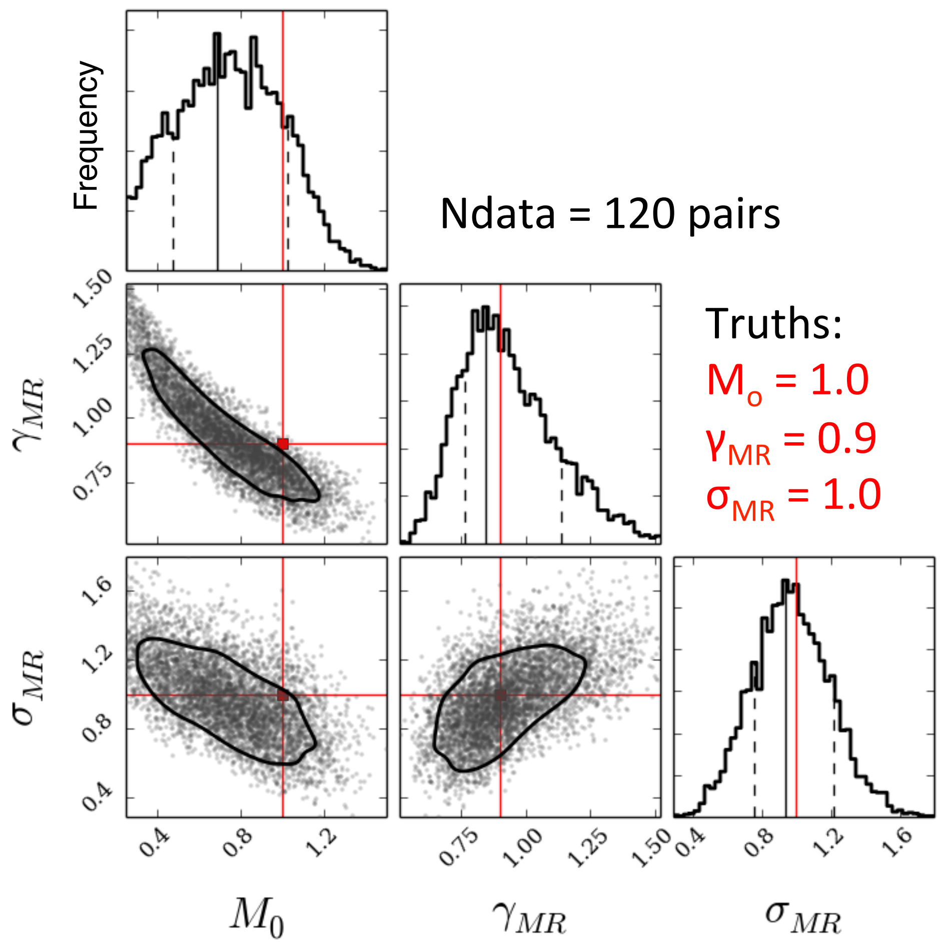
Case Study 2: Model performs well using simulated data with **artificially small measurement uncertainties.**



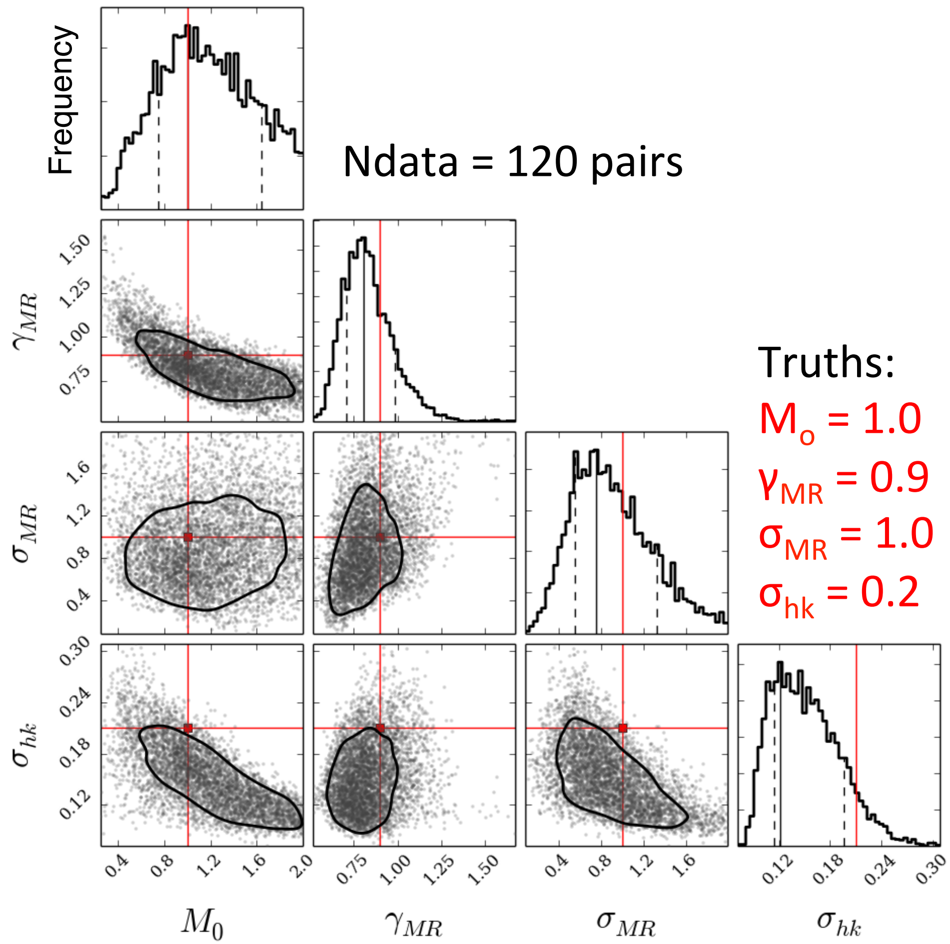
Ndata = 120 pairs

Truths:
$M_o = 1.0$
$\gamma_{MR} = 0.9$
$\sigma_{MR} = 1.0$

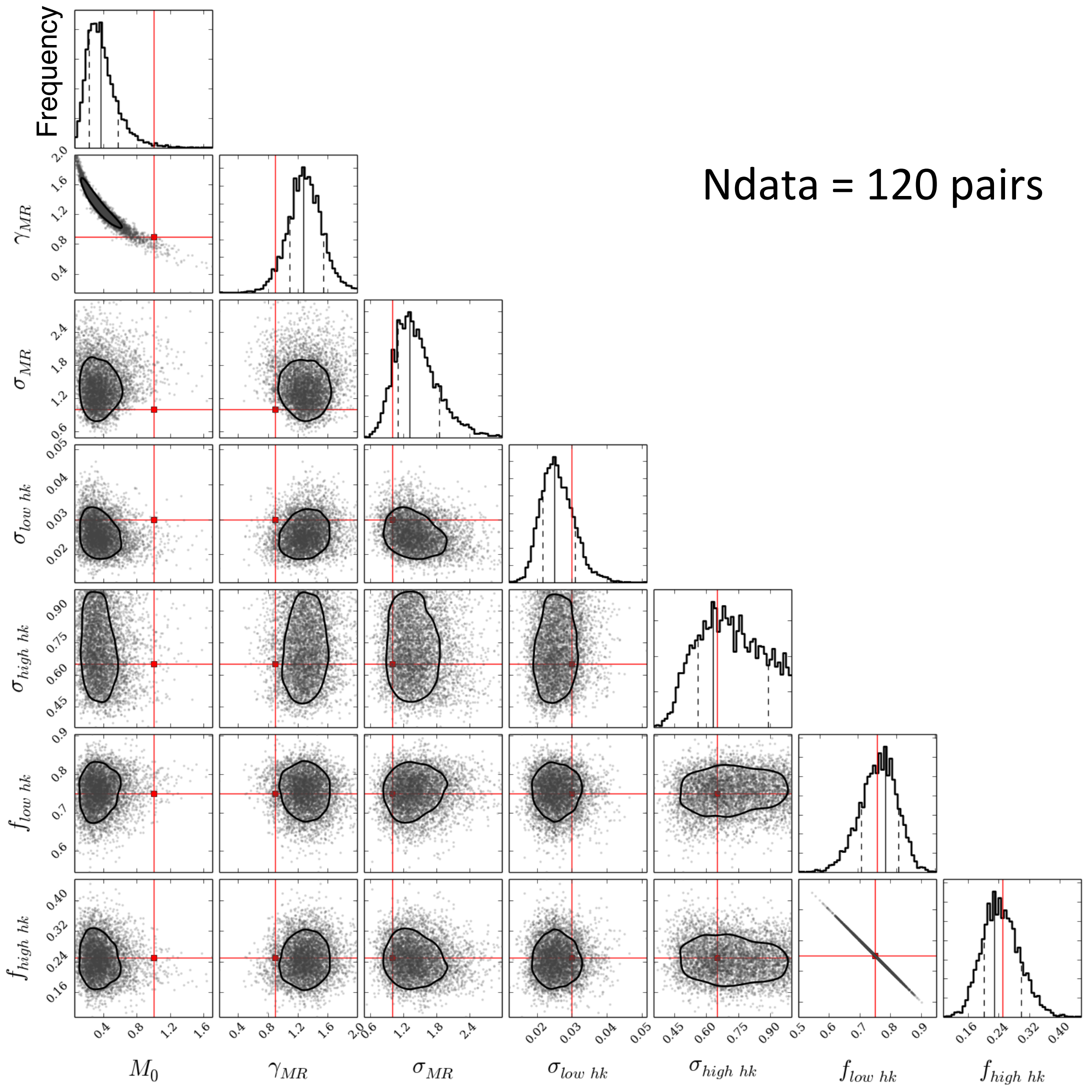
Case Study 2: Model performs well using simulated data with real ***Kepler*** **Measurement Uncertainties.**



Case Study 2: Model performs well using simulated data with **one-component eccentricity distribution.**



Case Study 2: Model performs well using simulated data with with **Eccentricity Distribution of Increased Complexity**



Ndata = 120 pairs

Truths:

$M_o = 1.0$

$\gamma_{MR} = 0.9$

$\sigma_{MR} = 1.0$

$F_{hk\ low} = 0.75$

$f_{hk\ high} = 0.25$

$\sigma_{hk\ low} = 0.03$

$\sigma_{hk\ high} = 0.65$

Two-Component Gaussian Mixture Model for Eccentricity

Case Study 2: The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs: **Potential Biases**

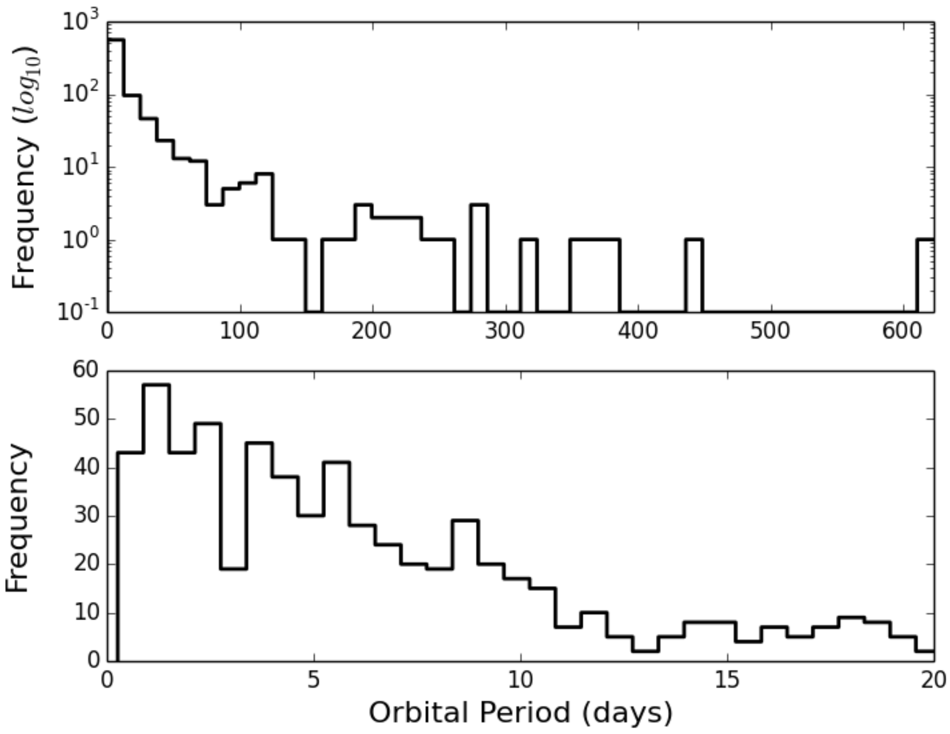
- Inferred mass-radius is suspiciously flat/low mass compared to theoretical models.
- TTV planet pairs near first order MMR may have different formation history than other classes of planets and thus differ in their composition and structure.
- Need to further test model assumptions
  - Gaussianity
  - Stellar properties
  - M-R power law potentially not flexible
  - Mass scatter could depend on mass/radius
  - Currently treating eccentricity distribution as separable from mass-radius distribution

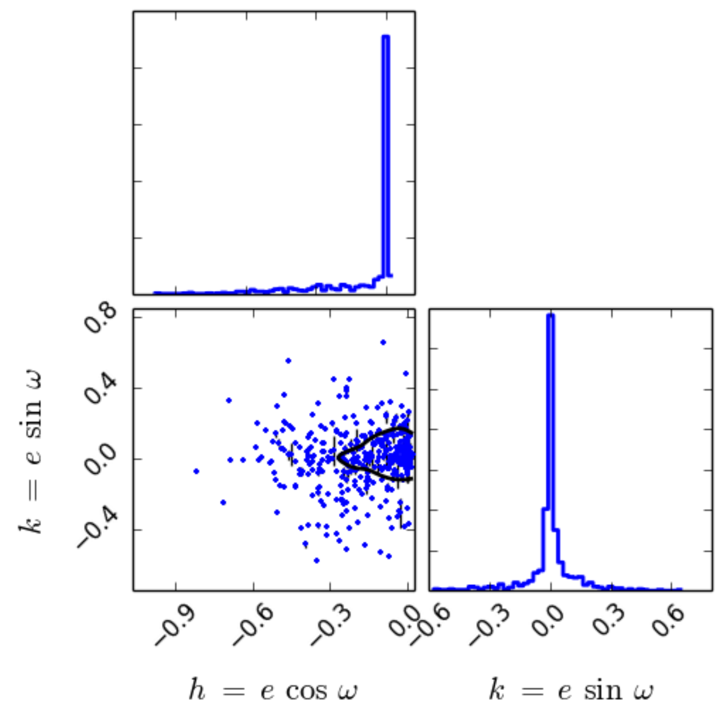
Case study 3: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*

- Probe binary star formation
- Probe efficiency of tidal circularization

Input: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*

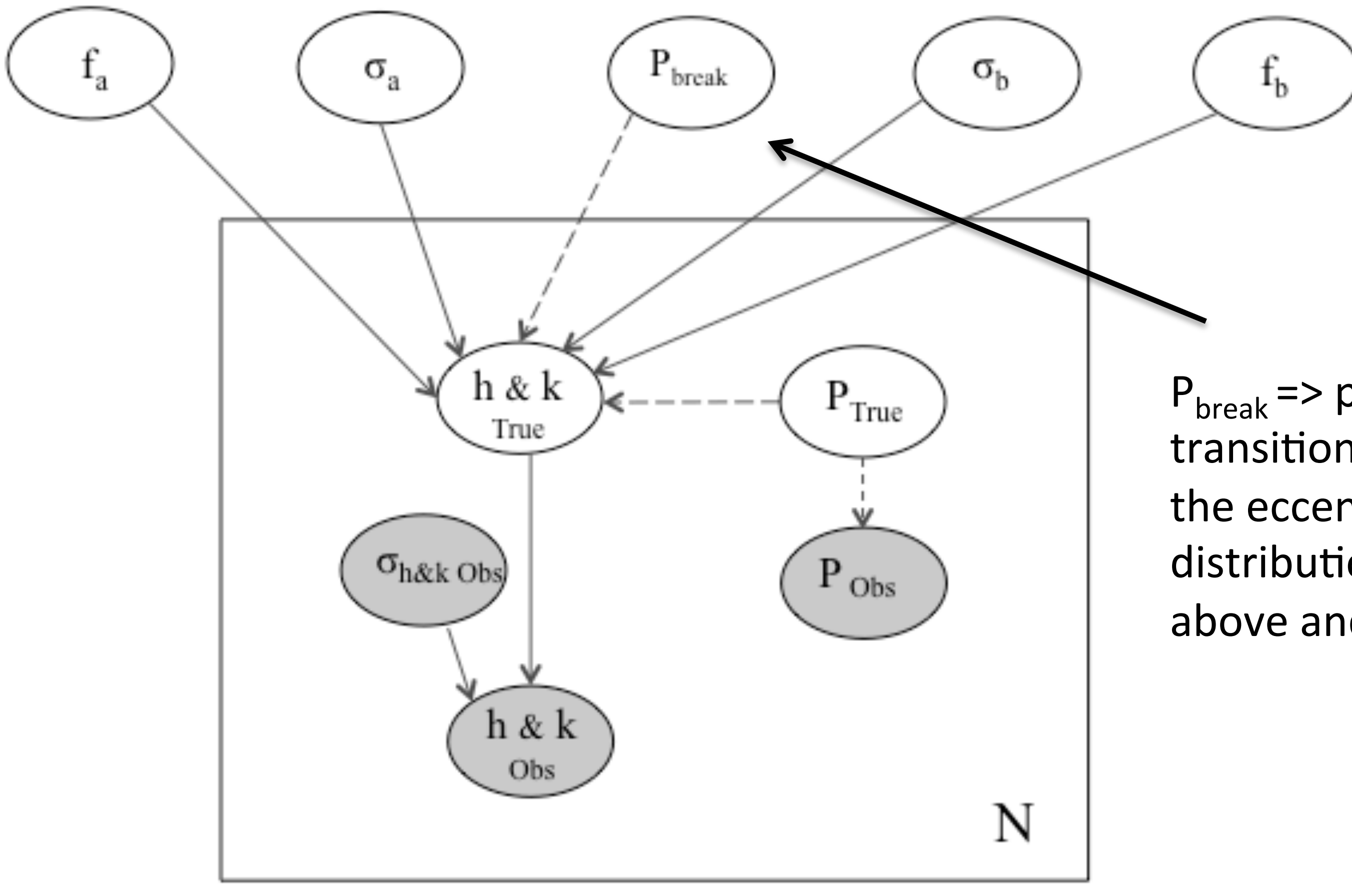
Ndata = 795 EBs





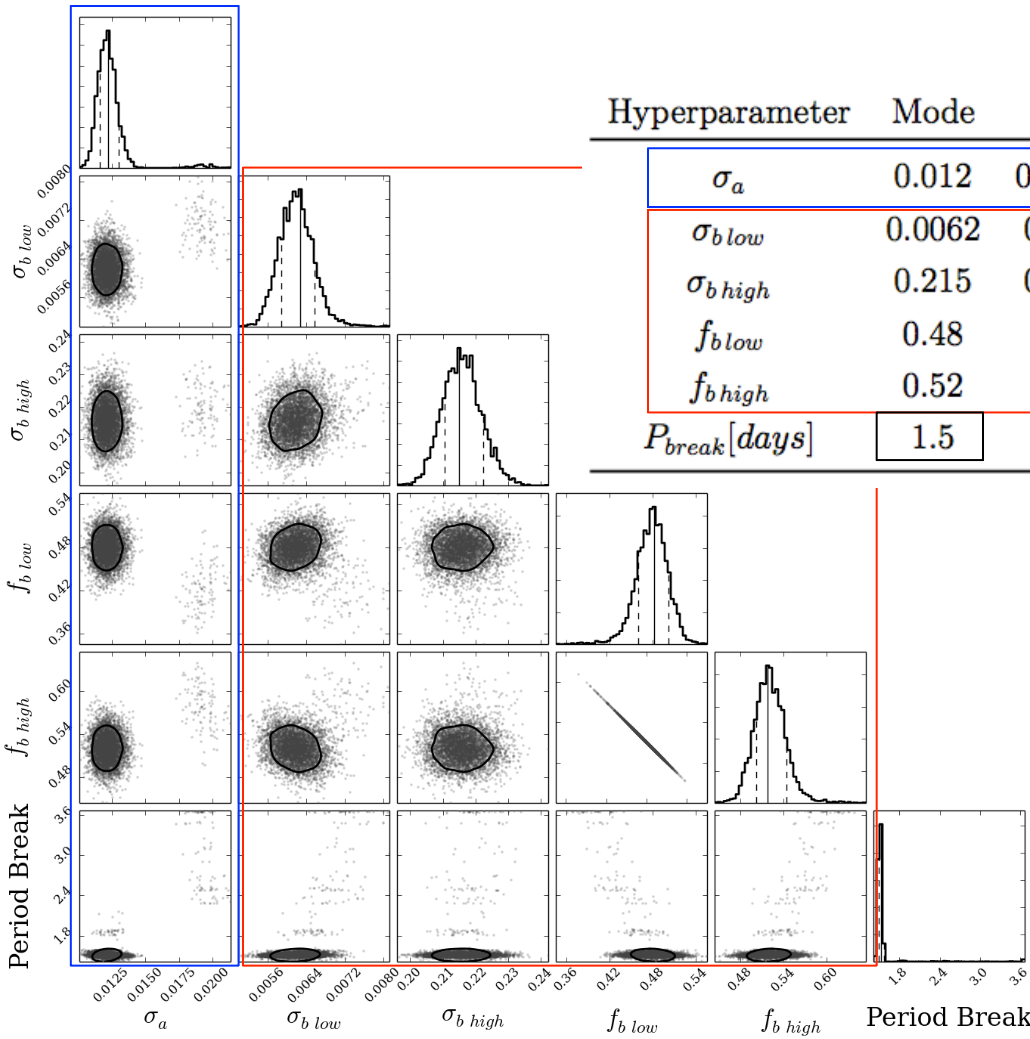
Case study 3: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*

Graphical Model of Bayesian Network:



P_{break} => period transition where the eccentricity distribution differs above and below

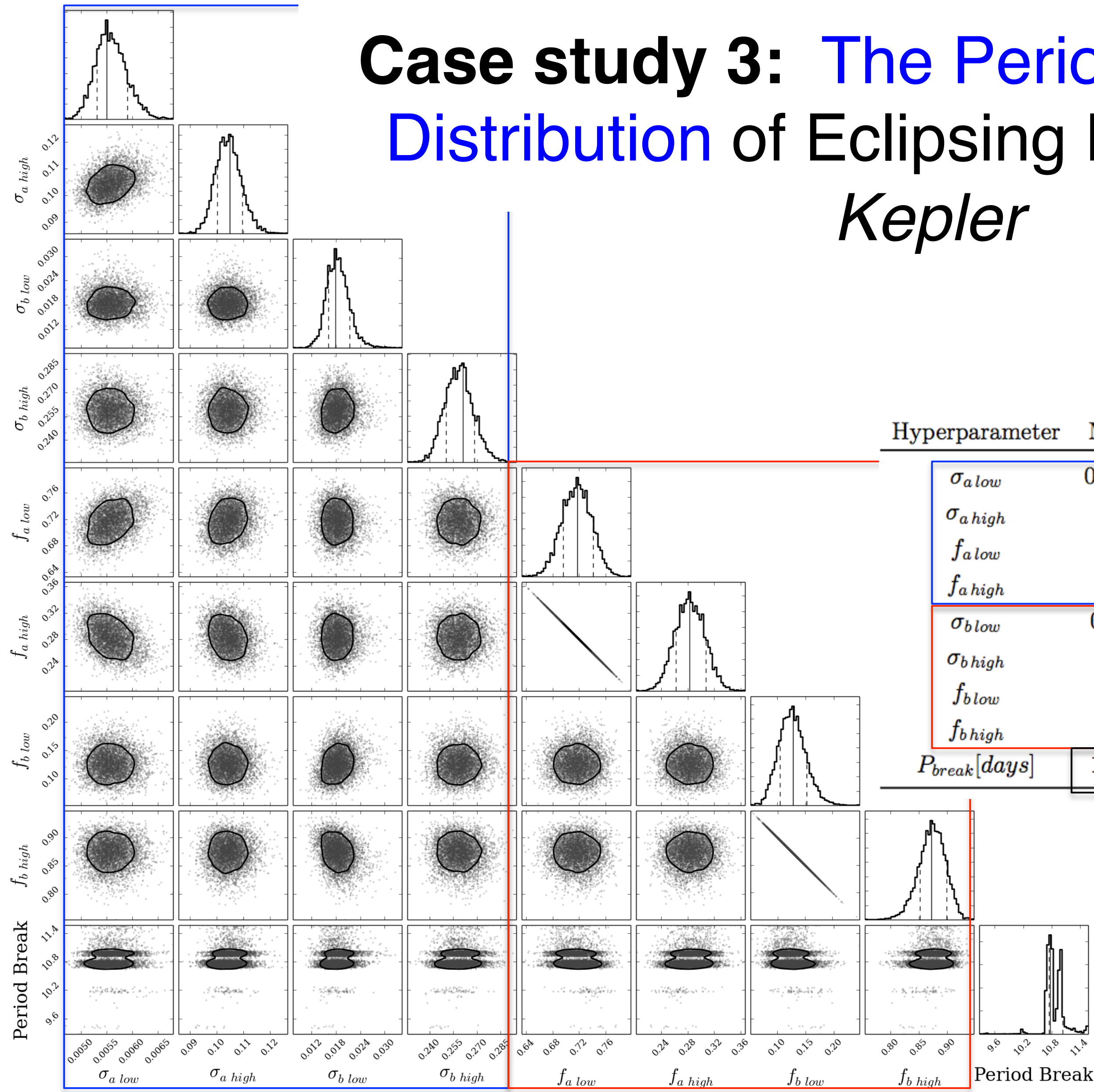
Case study 3: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*



Hyperparameter	Mode	68% credible interval +	−
σ_a	0.012	0.00079	0.00062
$\sigma_{b\,low}$	0.0062	0.0003	0.0004
$\sigma_{b\,high}$	0.215	0.0073	0.0045
$f_{b\,low}$	0.48	0.02	0.022
$f_{b\,high}$	0.52	0.026	0.016
$P_{break}[days]$	$\boxed{1.5}$	0.04	0.03

Shabram, Prsa *in prep*

Case study 3: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*

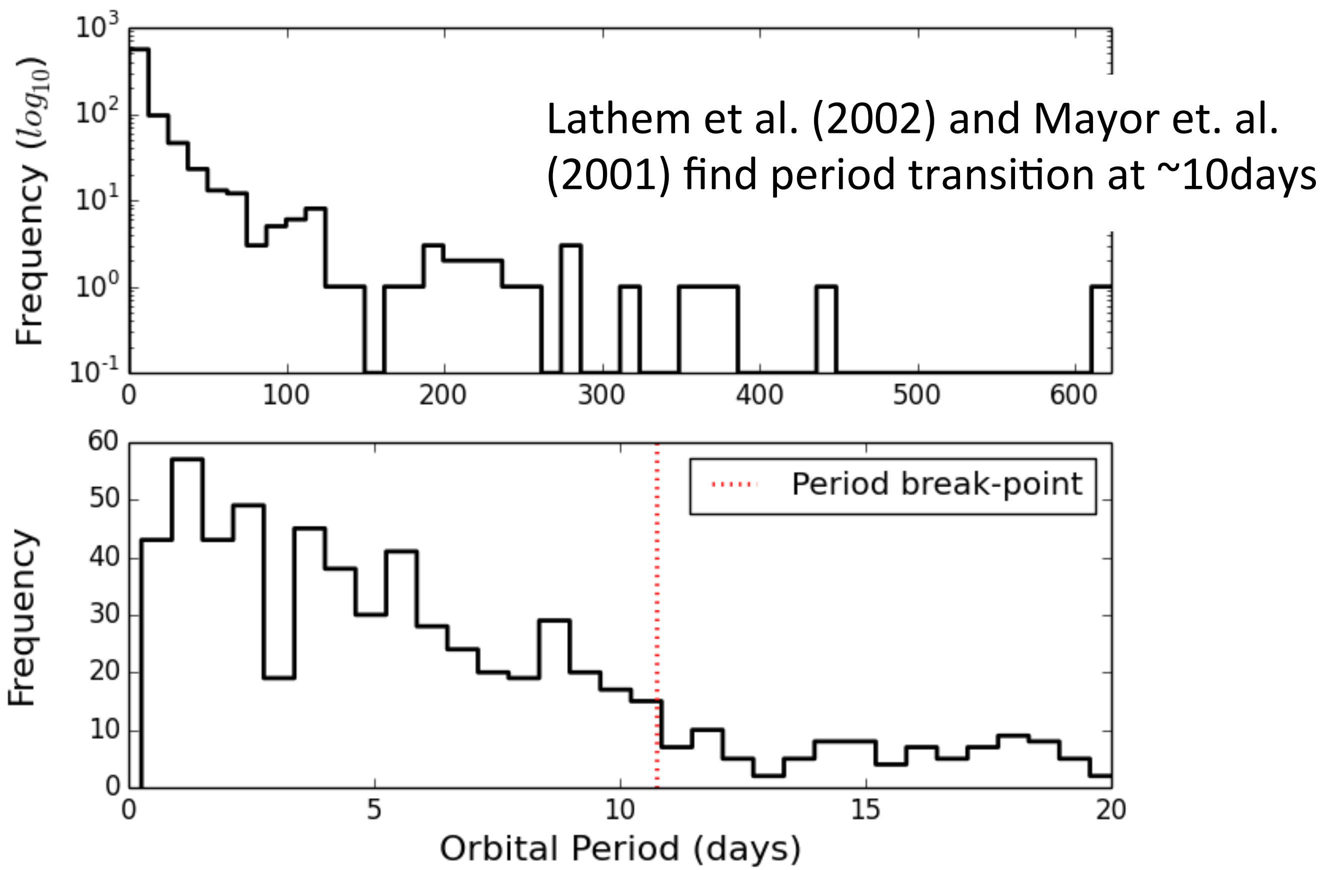


Shabram, Prsa *in prep*

		68% credible interval	
Hyperparameter	Mode	+	−
$\sigma_{a\,low}$	0.0055	0.0004	0.0002
$\sigma_{a\,high}$	0.1	0.0047	0.0048
$f_{a\,low}$	0.72	0.024	0.022
$f_{a\,high}$	0.28	0.025	0.02
$\sigma_{b\,low}$	0.018	0.0035	0.0018
$\sigma_{b\,high}$	0.26	0.0074	0.011
$f_{b\,low}$	0.13	0.024	0.023
$f_{b\,high}$	0.87	0.027	0.02
$P_{break}[days]$	10.74	0.26	0.027

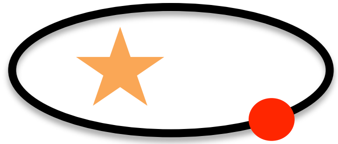
~70% **below** 11 days are very circularized
~90% **above** 11 days exhibit a wide range of eccentricities

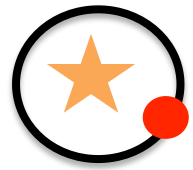
Case study 3: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*

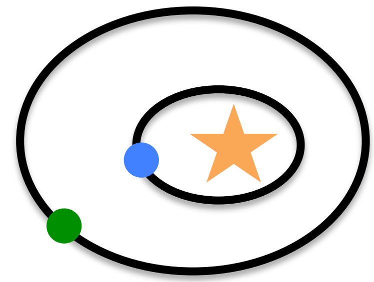


Case Study 3: The Period-Eccentricity Distribution of Eclipsing Binaries from *Kepler*: **Potential Biases**

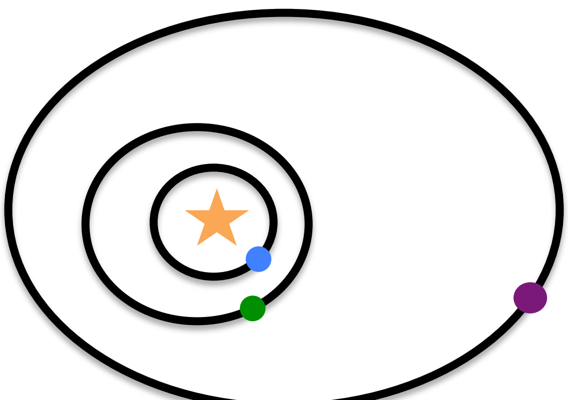
- Sample focuses on short-period EBs
- Sample dominated by field stars from the disk (as apposed to cluster or halo stars)
- Model doesn't include difference in stellar properties
- Model assumes simple dependence on period







Thank You!



Detection bias