The Pennsylvania State University
The Graduate School
Eberly College of Science

IMPLICATIONS FOR PLANET FORMATION FROM POPULATION INFERENCE OF *KEPLER*-PLANET-CANDIDATES AND ECLIPSING BINARIES

A Dissertation in
Astronomy & Astrophysics
by
Megan Iris Shabram

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2015

The dissertation of Megan Iris Shabram was reviewed and approved* by the following:

Eric B. Ford
Professor of Astronomy & Astrophysics
Chair of Committee, Dissertation Advisor

Steinn Sigurdsson
Professor of Astronomy & Astrophysics

Ronald L. Gilliland
Adjunct Professor of Astronomy & Astrophysics

Kevin Luhman
Professor of Astronomy & Astrophysics

Christopher House
Associate Professor of Geosciences

Donald Schneider
Department Head and Distinguished Professor

*Signatures are on file in the Graduate School.

Abstract

The Kepler Space Science Mission has revolutionized our understanding of planetary system architectures, and the diversity of planet bulk densities. From *Kepler*, we now have a population of ∼4,700 planet candidates and ∼ 3000 eclipsing binaries with measured light curves, from which we can begin to characterize the distribution of stars and planets to tease out relationships between planet properties and host star properties in a robust statistical manner. The results of these investigations constrain proposed planet formation theories. This dissertation analyzes three particular sub-populations observed by *Kepler* that are well suited for hierarchical inference to characterize their population properties. First, we investigate the eccentricity distribution for a sample of short-period planet candidates from *Kepler*, where both the transit and occultation are observed for each system. This sub-sample lends a rare opportunity for tractable inference of its eccentricity distribution, exposing at least two populations within the eccentricity distribution and potential correlations of the eccentricity with host star metallicity and planet radius. Secondly, we investigate the mass-radius-eccentricity relation for a sample of near-resonant planet-pairs from *Kepler*. This study greatly improves upon previous research of constraining the mass-radius relation for small planets. Furthermore, we explore the period-eccentricity distribution of eclipsing binary stars from *Kepler*. We find that ∼ 72% of EBs below ∼ 11 days are very circularized, where as ∼ 87% of EBs above ∼ 11 days can take on a wide range in eccentricity values including some with significant eccentricities.

Table of Contents

List of Figures **vii**

List of Tables **xviii**

Acknowledgments **xx**

Chapter 1
Introduction and Background **1**
1.1 Planet Formation 1
1.2 The Kepler Space Science Mission 6
1.3 Hierarchical Bayesian Modeling 10
1.4 Outline 13

Chapter 2
The Eccentricity Distribution of Short-Period Planet Candidates Detected by *Kepler* in Occultation **16**
2.1 Introduction 16
2.2 Observations 18
2.2.1 Derivation of the planet physical and orbital properties . . . 19
2.2.2 Properties of Planet Candidates Analyzed 23
2.3 Method 24
2.3.1 The Hierarchical Bayesian Model 24
2.3.2 Applying the Hierarchical Model to Eccentricity Measurements 26
2.3.3 Evaluating the Hierarchical Model 27
2.4 Results 27
2.4.1 Prior specification 27
2.4.2 Validating the Hierarchical Model 28
2.4.2.1 Generation and Analysis of Simulated Data 29
2.4.2.2 Results for Synthetic Data 31
2.5 Results for *Kepler* planet candidates with occultations 34

2.5.1 Correlation of The Eccentricity Distribution with Star and Planet Properties . . . 38
2.6 Summary of Results . . . 44
2.7 Discussion . . . 46
2.7.1 Potential Biases . . . 48
2.7.2 Future Research . . . 51

Chapter 3
The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs Detected by *Kepler* 53
3.1 Introduction . . . 53
3.2 Methods Overview . . . 55
3.2.1 *Kepler* Data . . . 55
3.2.2 TTV Ephemerides . . . 56
3.2.3 TTV Amplitudes . . . 57
3.2.4 Physical Model . . . 58
3.2.5 Statistical Model for Population . . . 61
3.3 Applying the Hierarchical Model to *Kepler* Data . . . 63
3.3.1 Variables . . . 65
3.3.2 The Population Model . . . 67
3.3.3 Hyperpriors . . . 68
3.3.4 Mid-Level Model Specification . . . 69
3.3.5 Base-Level Model Specification . . . 71
3.3.6 Evaluating the Hierarchical Model . . . 72
3.4 Results . . . 72
3.4.1 Validating the Hierarchical Model: Testing Analysis Models with Corresponding Simulated Data . . . 73
3.4.1.1 Effect of Measurement Uncertainties . . . 73
3.4.1.2 Highly Correlated Uncertainties in Mass and Eccentricity . . . 77
3.4.2 The Mass-Radius-Eccentricity Distribution for *Kepler* Planet Candidates . . . 79
3.5 Summary of Results . . . 81
3.6 Discussion . . . 81
3.6.1 Biases in this Study . . . 84

Chapter 4
The Period-Eccentricity Distribution for Eclipsing Binaries from *Kepler* 114
4.1 Introduction . . . 114

4.2 Observations . . . 116
4.3 Method . . . 117
4.3.1 The Hierarchical Bayesian Model . . . 119
4.3.1.1 The Population Model . . . 119
4.3.1.2 Hyperpriors . . . 120
4.3.1.3 Mid-Level Model Specification . . . 121
4.3.1.4 Base-Level Model Specification . . . 122
4.3.2 Evaluating the Hierarchical Model . . . 122
4.4 Results for *Kepler* Eclipsing Binaries . . . 123
4.4.1 The Eccentricity Distribution . . . 123
4.4.2 The Period-Eccentricity Distribution . . . 126
4.5 Discussion . . . 129

Chapter 5
Conclusions and Future Work **156**
5.1 Conclusions . . . 156
5.2 Future Work . . . 159

Bibliography **167**

List of Figures

1.1 Scatter in planet orbital period and planet mass for the confirmed exoplanets as of August 27th, 2015 from the NASA Exoplanet Archive (Akeson et al., 2013), colored based on the method of detection. Multiple exoplanet detection methods such as transits and radial velocity (RV) are critical for expanding the parameter space of detected planets. The radial velocity detected planets (red) show a cluster for larger planet masses and longer orbital periods, where as the planets detected via the transit method (green) fill in the region of parameter space at shorter orbital periods (less than 1 year). In this figure, timing variations refer to planets detected with a method involving timing such as pulsar timing or transit timing variations. 2

1.2 Planet orbital period versus planet radius for *Kepler* planets. Confirmed planets are shown in blue, and planet candidates are shown in red. Detection of an Earth-radii planet around a solar-like star on a year long orbit is still to be determined within the baseline *Kepler* data products. 7

2.1 Cumulative distributions of $|h|$ and $|k|$. The solid black curve is the true distribution from which the simulated planet's h and k values are drawn. The dashed black curve is the cumulative distribution for one simulated dataset ("R2", "good"; $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, see §4.2.1) that includes simulated observational uncertainties. The gray shaded region is the 68.3% credible interval for the CDFs of the posterior samples for the population parameters of the intrinsic distribution of $|h|$ and $|k|$ (i.e., without intrinsic uncertainties). This is calculated once the simulated observations have been analyzed using the same two-component Gaussian mixture model as was used to generate the data. The two-component Gaussian HB model does a good job of capturing the true distribution for datasets generated with a two-component Gaussian mixture. 32

2.2 Results of an HB model that parameterizes the eccentricity as a standard Beta distribution. We investigate using a Beta distribution analysis model on an "R2", "good" (see Table 2) simulated dataset. A cumulative distribution of the simulated eccentricity data are shown in grey. The distribution generated using a two-component Gaussian mixture model for h and k values (e.g. $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, as described in §4.2.1) is shown in red. The dashed green curve is a cumulative Beta distribution, Beta(α, β), plotted using the posterior modes for α and β, ($\alpha = 0.11\pm^{0.04}_{0.02}$, $\beta = 1.73\pm^{0.85}_{0.24}$) for this HB model. The Beta distribution erroneously predicts a strong peak near $e = 0$ and under predicts the frequency of larger eccentricities. 35

2.3 A histogram of the h and k dataset are shown in grey. Shown in black is a one-component Gaussian distribution using the posterior mode for the dispersion obtained from an HB model that uses a one-component Gaussian mixture model. Shown in red is a two-component Gaussian mixture model using posterior modes for the mixture fractions and dispersions obtained from an HB model that uses a two-component Gaussian mixture model. The black model does a poor job at capturing the shape of the distribution. The red model captures the peaked nature of the true distribution while also allowing for a smaller number of measurements far from the central peak. 36

2.4 Posterior distribution for the dispersion of a Gaussian model for h and k applied to our full *Kepler* short-period planet candidate transit and occultation dataset. The 68.3% credible intervals about the mode are shown as dotted black lines, and the mode is shown as a vertical solid black line. 37

2.5 Joint posterior distributions for a two-component Gaussian mixture model applied to our *Kepler* short-period planet candidate transit and occultation dataset. In each panel, the data are plotted with the horizontal axis representing σ_{low} the lesser value of σ_1 and σ_2, σ_{high} the greater value of σ_1 and σ_2, and f_{low} the corresponding weight for the low mixture component, on a logarithmic scale. Since $f_{high} = 1 - f_{low}$, we only show f_{low} here. The vertical axis shows these same variables, and each panel is the corresponding two-dimensional marginal posteriors for each parameter pair. The contour region plotted over the sampled posterior represents the 68.3% credible interval. The one-dimensional histograms are plotted as log density, with the 68.3% credible intervals shown as dotted black lines, and the mode is shown as a vertical solid black line. The two-component Gaussian mixture model characterizes the eccentricity distribution of our sample of planet candidates better than a one-component analysis model. 39

2.6 One-component Gaussian analysis model applied to subsets of the *Kepler* transit and occultation data. We apply an HB model to small-valued (blue) and large-valued (red) halves of the *Kepler* short-period planet candidate transit and occultation data, sorted by (a) stellar effective temperature, (b) planet radius, (c) orbital period, and (d) stellar metallicity. The dotted lines correspond to the 68.3% credible intervals and the solid vertical lines correspond to the mode for each posterior distribution. for panel (b) planet radius, and panel (d) stellar metallicity, differences in the small-values and large-valued data subsets merit further investigation. In order to explore these results further, we analyze these subset using a two-component Gaussian mixture model (see Figure 2.7). 41

2.7 Figure 2.7: Continued on following page. 42

2.8 Figure 2.7: Two-component Gaussian mixture model applied to the *Kepler* transit and occultation data. We apply an HB model to small- and large-valued halves of the short period *Kepler* candidate occultation data, sorted by stellar effective temperature (a), planet radius (b), orbital period (c), and stellar metallicity (d). The full sample is divided into two equally sized small- and large-value subsets before being processed through our HB model. The small-valued subset of data corresponds to the blue and green clusters, and the large-valued subset corresponds to the red and orange clusters. The two groups of clusters represent samples of the posterior distribution for the hyperparameter vector, in this case for σ_{low} and f_{low} (top left group of clusters), and σ_{high} and f_{high} (bottom right group of clusters). The data are plotted with the vertical axis representing the low value of the mixture fraction, f_{low}, in green and orange, and, f_{high}, in blue and red for the two subsets of sorted data shown. The contours represent 68.3% credible intervals. Interestingly, for planet radius (b), and stellar metallicity (d), the posteriors of the mixture fractions for the planet candidates with large-valued planet radii and for small-valued host star metallicities are consistent with 0 and 1, indicating only one population is required to accurately model the eccentricity distribution for these subsets of planet candidates. For planet radius and stellar metallicity, we also see that the two-component population models are somewhat different for small and large value subsets. 43

2.9 Figure 2.8: Marginal posterior distribution for the critical period break point from a joint period-eccentricity distribution HB model. We analyze the full dataset using an HB model that allows for the eccentricity distribution to differ depending on whether the orbital period is greater or less than the critical period break point. We infer that the present data does not allow us to empirically identify a period cutoff that depicts two populations. 45

3.1 Transit timing variation amplitudes with size proportional to the planet-to-star radius ratio and color indicating the fractional uncertainty. The y axis represents the amplitude of the sine component of the transit timing variation amplitudes, and the x-axis represents the amplitude of the cosine component. The circle sizes are proportional to the planet to star radius ratio, and the fractional uncertainty (ratio of 68% credible interval to mode) is show as the color bar where the blue represents small fractional uncertainty, and the red represents large fractional uncertainty. 59
3.2 Transit timing variation amplitudes with size proportional to the planet-to-star radius ratio and color indicating the fractional uncertainty. The y axis represents the amplitude of the sine component of the transit timing variation amplitudes, and the x-axis represents the amplitude of the cosine component. The circle sizes are proportional to the planet to star radius ratio, and the fractional uncertainty (ratio of 68% credible interval to mode) is show as the color bar where the blue represents small fractional uncertainty, and the red represents large fractional uncertainty. 15 outliers, out of 240 planets total, are not shown for clarity. 60
3.3 Graphical model of Bayesian network. Here we show the relationship between the hyperparameters, latent variables and observables in our mass-radius-eccentricity joint hierarchical Bayesian model. Observables are high level measurements that have been provided by others. The observables are shown in gray. The latent variables are shown in white within the boxed region. The hyperparameters that describe the features of the population model are shown in white above the boxed region. The solid arrows represent probabilistic relationships and can be read as "distributed as", where the dashed arrows represent definitions or physical models. The physical models used in this study (see Equations 3.44 to 3.48) link observables to un-observed quantities (Planet radius, planet mass, eccentricity), and we are able to get posteriors for these un-observed quantities. 64
3.4 Just Another Gibbs Sampler mass-radius-eccentricity HB model, with R syntax highlighting. 73

3.5 Posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs. Here, the eccentricity distribution is fixed. In this simulated data case, we set the uncertainties of our observables artificially small (all set to 0.001) to test the robustness of this model to the quantity of planet pairs in our real data set. Shown in red are the true values for the mass constant $M_o = 1.0\,M_\oplus$, the power-law index $\gamma_{MR} = 0.9$ and the scatter in mass for a given radii $\sigma_{MR} = 1.0\,M_\oplus$. All true values fall within the 68% credible interval (shown above as black dashed lines) of the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. 75

3.6 Posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs, with a fixed eccentricity distribution. In this simulated data case, we set the uncertainties of our observables to the true observational uncertainties from the *Kepler* data, to test how the real measurement uncertainties when combined with the quantity of data available, will impact our results. Shown in red are the true values for the mass constant $M_o = 1.0\,M_\oplus$, the power-law index $\gamma_{MR} = 0.9$ and the scatter in mass for a given radii $\sigma_{MR} = 1.0\,M_\oplus$. All true values fall within the 68% credible interval, shown above as black dashed lines, of the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. 76

3.7 Posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs. In this simulated data case, we assign a Uniform prior distribution to the dispersion for the projected eccentricity distribution, and test the ability of the data to constrain a model on increased complexity. Shown in red are the true values for the mass constant $M_o = 1.0\,M_\oplus$, the power-law index $\gamma_{MR} = 0.9$ and the scatter in mass for a given radii $\sigma_{MR} = 1.0\,M_\oplus$, and the dispersion for the projected eccentricity $\sigma_{hk} = 0.21$. All true values fall with the posterior distribution for the marginals shown above. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. 86

3.8 Hyperparameters for simulated *Kepler* data. Here, the hyperparameters for this HB model are $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, $\sigma_{MR} = 1.0\,M_\oplus$, and the two-component Gaussian mixture model hyperparameters are now $\sigma_{low} = 0.03$, $\sigma_{high} = 0.65$, $f_{low} = 0.75$, and $f_{high} = 0.25$. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. 87

3.9 Posterior modes and credible intervals for planet radius vs. planet mass from simulated mass-radius-eccentricity HB model. The gray bars represent the 68% credible interval of the marginal posteriors for planet mass and radii. The red points represent the posterior modes for the interior planets and the blue points represent the posterior modes for the exterior planets. The black points represent the true planet masses and planet radii generated for this particular model. The thick solid black curve is the true mass-radius relation that the simulated data are drawn from: $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, for $M_p = M_o R_p^{\gamma_{MR}}$. The region of scatter in mass for a given radii, $\sigma_{MR} = 1.0\,M_\oplus$, is shown as the region between the thick dotted black lines. 88

3.10 Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 115^{th} simulated system in the population. In this case, the simulated h and k values are both very close to zero. We can see that the histograms of the marginal posteriors for the interior and exterior planet mass constrain the true generated value to within the 68% credible intervals shown as the dotted black lines. Posterior modes for the marginals are shown as the solid black lines. The 68% credible region for the 2D posteriors is shown as the solid black curve. 89

3.11 Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 43^{th} simulated system in the population. The 68% credible intervals are shown as the dotted black lines on the histograms of the marginal posteriors. Posterior modes for the marginals are shown as the solid black lines. The 68% credible regions for the 2D posteriors are shown as the solid black curve. In this particular case, the measurements of the TTV amplitudes were unavailable, so we included injected TTV measurement uncertainties set to 2 into the data set for this planet. In this case the posteriors for the projected eccentricity do not encompass the truths, but the simulation does will for recovering the true planet masses. 90

3.12 Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 40^{th} simulated system in the population. The 68% credible intervals are shown as the dotted black lines on the marginal posteriors. Posterior modes for the marginals are shown as the solid black lines. The 68% credible regions for the 2D posteriors are shown as the solid black curve. In this case we see a bimodal solution for k, for both the interior and exterior planet. Solutions where k is closer to zero produce larger mass planets, and cases where k deviates from zero produces smaller mass planets. In this case, the HB model is able to recover the true generated values well. 91

3.13 Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 41^{st} simulated system in the population. The 68% credible intervals are shown as the dotted black lines on the marginal posteriors. Posterior modes for the marginals are shown as the solid black lines. The 68% credible regions for the 2D posteriors are shown as the solid black curve. In this case, k deviates a small amount from zero and you can see two modes start to break out. The masses and projected eccentricities are still well constrained in this case. 92

3.14 Hyperparameters for simulated *Kepler* data. Here, the hyperparameters for this HB model are $M_1 = M_2 = 1.0\,M_\oplus$, $\gamma_{MR_1} = \gamma_{MR_2} = 0.9$, $\sigma_{MR_1} = \sigma_{MR_2} = 1.0\,M_\oplus$, and the two-component Gaussian mixture model hyperparameters are now $\sigma_{low} = 0.03$, $\sigma_{high} = 0.65$, $f_{low} = 0.75$, and $f_{high} = 0.25$. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. 93

3.15 Hyperparameters for real *Kepler* data. Here we use a one-component Gaussian mixture model for the eccentricity distribution in our joint mass-radius-eccentricity HB model. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. The posterior modes and credible intervals for the hyperparameters are reported in Table 3.2. 94

3.16 Hyperparameters for real *Kepler* data. Here we use a two-component Gaussian mixture model for the eccentricity distribution in our joint mass-radius-eccentricity HB model. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. The posterior modes and credible intervals for the hyperparameters are reported in Table 3.2. 95

3.17 Planet mass vs. planet radius. Shown are the means from the latent variable posteriors for planet mass and planet radius when using a mass-radius-eccentricity HB model that parameterizes the projected eccentricity as a two-component Gaussian mixture model. The 68% credible intervals of the marginal posteriors for mass and radius are show as gray bars. The mass-radius relation using the posterior means for the mass constant M_o and the exponent γ_{MR} is shown as the solid back curve. The dotted black curve represents the posterior mean for the scatter in planet mass for a given radius σ_{MR}. 96

3.18 10 draws of the eccentricity PDF from the joint posterior, from a mass-radius-eccentricity HB model that uses a two-component Gaussian mixture model to parameterize the eccentricity distribution. The top panel shows a linear scale and the bottom panel shows a log scale in order to see the larger dispersion component more clearly. 97

4.1 Projected eccentricity ($h = e\cos\omega$ and $k = e\sin\omega$) for *Kepler* EBs. The blue dots indicate the scatter in h and k for our sample of 795 EBs. The black bars show the measurement uncertainties, and are only visible in some cases, as the measurement uncertainties for this sample are small. The 68% credible region of the data is shown as the black curve in the bottom left panel. Histograms of the h and k measurements are shown in blue. 117
4.2 *Kepler* EB orbital periods. The top panel shows the histogram of the full range in period for our sample of EBs, ranging from 0.00325 to 623 days. The frequency of orbital period values for the top panel is shown on a log scale for clarity. The bottom panel zooms in on orbital period for clarity in the region of interest, showing the frequency of orbital period between 0 and 20 days. The red dotted line shows the posterior mode of the inferred period break point at which the EBs below this threshold have a different eccentricity distribution than those above this threshold. See §4.3 and 4.4 for details regarding this result. We use the full sample of EBs in our analysis. 118
4.3 Graphical model of Bayesian network. Here we show the relationship between the hyperparameters, latent variables and observables in the joint period-eccentricity hierarchical Bayesian model. The observables are shown in gray. The latent variables are shown in white within the boxed region. The hyperparameters that describe the features of the population model are shown in white above the boxed region. The solid arrows represent probabilistic relationships and can be read as "distributed as", where the dashed arrows represent definitions or physical models and can be read "defined as". 119
4.4 JAGS model for EB period-eccentricity distribution, with R syntax highlighting. 123

4.5 Posteriors for the hyperparameters in the two-component Gaussian mixture model described in Chapter 2, applied to *Kepler* EB projected eccentricity measurements. Approximately half the population comes from an eccentricity distribution with a very small dispersion ($\sigma_{low} = 0.0067$). The other half comes from a distribution with a larger dispersion ($\sigma_{high} = 0.21$), meaning some EBs will have significant eccentricities in this half of the population. The black contour for the 2D joint posteriors represent the 68% credible region. For the marginal distributions shown as histograms, the solid black line represents the posterior mode, and the dotted black lines represents the 68% credible interval. The posterior modes and the 68% credible intervals of the marginal posteriors can be found in Table 4.1. 125
4.6 Cumulative distributions for the EB projected eccentricity distribution. The *Kepler* EB projected eccentricity data are shown as the dotted black curve. The light gray shaded region shows the 68% credible region for the projected eccentricity distribution using posterior draws of the hyperparameters. The dark grey region represents the 95% credible region. The solid black curve is the cumulative distribution for the projected eccentricity using the posterior mode for the hyperparameters. 126
4.7 The black contour for the 2D joint posteriors represent the 68% credible region. For the marginal distributions shown as histograms, the solid black line represents the posterior mode, and the dotted black lines represents the 68% credible interval. The posterior modes and the 68% credible intervals of the marginal posteriors can be found in Table 4.2. 128
4.8 The black contour for the 2D joint posteriors represent the 68% credible region. For the marginal distributions shown as histograms, the solid black line represents the posterior mode, and the dotted black lines represents the 68% credible interval. The posterior modes and the 68% credible intervals of the marginal posteriors can be found in Table 4.3. 130

5.1 Probabilistic detection efficiency toy model. The red curve shows a detection efficiency function, the black dashed line shows a point estimate of the SNR, and the blue curve shows the full posterior measurement of the SNR for this example observation. Using the posterior for the measured SNR instead of a point estimate avoids biases in the inference models. 165

List of Tables

2.1 Result for $e\cos\omega$ and $e\sin\omega$ from MCMC. See §2.1 for details on how these values are calculated. 21
2.2 Values indicating the quantity and quality of the suite of simulated observations used in our analysis. Datasets labeled “good" (“half") consist of 50 (25) planets with measurement uncertainties of 0.04 and 0.08 for h and k respectively. These datasets are designed to be similar to our actual transit and occultation dataset for both h and k. Datasets labeled “better" (“best") contain 50 (500) planets with measurement uncertainties of 0.001, and are designed to forecast the power of this method and model setup when the quantity of real data grows and the quality of data is improved upon (better measurement uncertainty). 30
2.3 Results of validation and sensitivity analysis of three hierarchical models for eccentricities. Shown in this table are Kolmogorov-Smirnov (K-S) statistics comparing datasets of simulated observations with datasets generated using the posteriors of the hyperparameters from an HB model that analyzed the same set of simulated observations (comparing input to output to test model). Here R1, R2, and R3 represent a one-, two- and three-component Gaussian mixture model, respectively. Table 2.2 summarizes the different quantity and quality of simulated observations used in this analysis. A two-component Gaussian mixture model does well across the majority of simulated datasets. See §4.2.2 for a detailed interpretation of these results. 30

3.1 Minimum, median, and maximum values for *Kepler* data used in this study. Tables 3.3-3.5 show the full dataset. 56
3.2 Posterior modes and 68% credible intervals of hyperparameters for mass-radius-eccentricity HB model, using a one-component Gaussian mixture model for eccentricity. 79

3.3 Posterior modes and 68% credible intervals of hyperparameters for mass-radius-eccentricity HB model, using a two-component Gaussian mixture model for eccentricity. 81
3.4 Transit Timing Variation (TTV) amplitudes for 120 near-first-order mean motion resonance transiting exoplanet pairs. 98
3.5 Planet-to-star radius ratio and orbital period for 120 near-first-order mean motion resonance transiting exoplanet pairs. 104
3.6 Host star mass and radius for 120 near-first-order mean motion resonance transiting exoplanet pairs. 110

4.1 Posterior modes and 68% credible intervals of hyperparameters for the two-component Gaussian mixture model for h and k. 124
4.2 Posterior modes and 68% credible intervals of hyperparameters for a one-component (low orbital periods) and a two-component (high orbital periods) Gaussian mixture model for h and k with period break point. 127
4.3 Posterior modes and 68% credible intervals of hyperparameters for a two-component (low orbital periods) and a two-component (high orbital periods) Gaussian mixture model for h and k with period break point. 129
4.4 Orbital period measurements and projected eccentricity measurements with uncertainties for *Kepler* EBs. 131

Acknowledgments

I would like to thank my advisor Professor Eric Ford for advising my dissertation research, and helping me gain confidence. I would like to thank my committee for very useful feedback helping me to develop my research and dissertation during my committee meetings over the last couple years. I would like to thank my peers and office mates at University of Florida, Dan Gettings, Jason Ybarra, Krista Romita, Wenli Mo, Kelsey Jorgenson, and Ben Wu among others. In particular Dan Gettings has been an extremely helpful colleague and friend, often being available to answer questions I have about coding or writing. I would like to thank my peers and office mates at Penn State, Sam Halverson, Arptia Roy, Henry Gebhardt, Sharon Xeusong Wang, and Joanna Bridge among others. In particular I would like to thank Ben Nelson who was my peer at UCSC, UF and PSU. I would like to thank my parents, Paul and Lynda Shabram. I would like to thank Lissa Friedman, my therapist for the last 4.5 years. Furthermore, I would like to thank my best girlfriends over the past 5 years: Rachel Rampy, Amanda Ramcharan, Carrie Vath, Natalie Williams, and Ana Amaya.

I would like to thank the *Kepler* team for the many years of work critical to the research projects outlined in this dissertation. I would like to thank Jessi Cisewski, Jonathan Fortney, Brice Olivier-Demory, Daniel Jontof-Hutter, Angie Wolfgang, and Leslie Rogers among others. I would like to acknowledge the SAMSI Bayesian Characterization of Extrasolar Populations working group for discussions that improved these research projects. In particular, we thank Merlise Clyde, Darin Ragozzine, Thomas Loredo, David Hogg, Thomas Barclay, and Robert Wolpert for their valuable contributions. This work was supported in part by the Pennsylvania State University's Center for Exoplanets and Habitable Worlds, NASA *Kepler Participating* Scientist Program awards NNX12AF73G and NNX14AN76G and NASA Origins of Solar Systems awards NNX13AF61G and NNX14AI76G. This material was based upon work partially supported by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute (SAMSI). Any opinions, findings, and

conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. I would like to acknowledge the Institute for CyberScience at The Pennsylvania State University for providing advanced computing resources and services that have contributed to the research results reported in this paper (http://rcc.its.psu.edu), in particular Hoofar Pourzand and William Brouwer. I would like to acknowledge the University of Florida Research Computing (http://researchcomputing.ufl.edu) for providing computational resources and support that have contributed to the research results reported in this publication. I would like to acknowledge the Center for Exoplanets and Habitable Worlds supported by the Pennsylvania State University, the Eberly College of Science, the Pennsylvania Space Grant Consortium, and the NASA Astrobiology Institute and the Pennsylvania State Astrobiology Research Center. Some/all of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX13AC07G and by other grants and contracts. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

Chapter 1
Introduction and Background

1.1 Planet Formation

Results from ground and space based exoplanet discovery and characterization missions have exposed populations of planets that have orbital configurations and bulk densities quite different from the planets in our solar system. Radial velocity (RV) surveys have uncovered different classes of giant planetary systems, including giant planets with orbits on the order of a few days to giant planets with orbits ∼300 days to 4 years (Ford, 2014; Wright et al., 2011). RV surveys are incomplete for planets at even larger orbital separations, however, direct imaging and microlensing observatories are beginning to probe this parameter space. Planets at these wide separations may be the scattered companions of a class of close-in giant planets, sometimes with orbital periods of hours to days, that were primarily detected with the early RV and transit discovery missions. Furthermore, transit surveys such as the Kepler Space Science mission (Borucki et al., 2010; Koch et al., 2010; Borucki et al., 2011a), are most sensitive to planets with large planet radii or planet-to-star radius ratios, with short orbital periods.

As *Kepler* pushed the limits of transit photometry towards smaller planet radii and longer orbital periods to look for an Earth analog, it became clear that super-Earths and mini-Neptunes may be the dominant result of planet formation processes. This has prompted the revision of previous planet formation models that were developed to explain the formation of our solar system's planets. Figure 1.1 from the NASA Exoplanet Archive (Akeson et al., 2013) shows the scatter in planet mass and planet orbital period for the confirmed exoplanets as of August 27th,

2015, colored based on the method of detection. From this figure, it is clear that methods such as RV and transits for detecting exoplanets probe different regions of parameter space. As we push the limits of these detection methods we can begin to reduce selection biases, filling in gaps in parameter space, or discovering empirical gaps that are markers of the features of planet formation mechanisms.

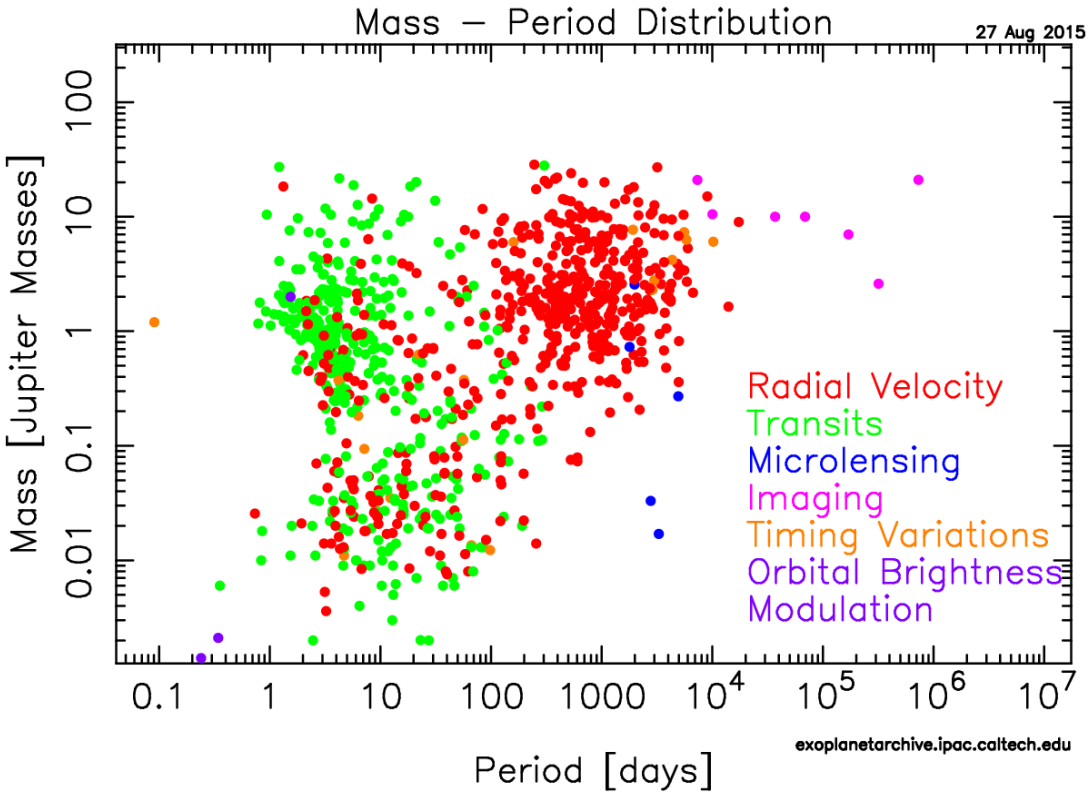

Figure 1.1: Scatter in planet orbital period and planet mass for the confirmed exoplanets as of August 27th, 2015 from the NASA Exoplanet Archive (Akeson et al., 2013), colored based on the method of detection. Multiple exoplanet detection methods such as transits and radial velocity (RV) are critical for expanding the parameter space of detected planets. The radial velocity detected planets (red) show a cluster for larger planet masses and longer orbital periods, where as the planets detected via the transit method (green) fill in the region of parameter space at shorter orbital periods (less than 1 year). In this figure, timing variations refer to planets detected with a method involving timing such as pulsar timing or transit timing variations.

Initial Planet Formation Theories

Initial theories regarding the formation of planetary systems at large were put forth to explain the formation of the solar system, and are highlighted by two dominating mechanisms. These two mechanisms are referred to as the "core accretion" and the "gravitational instability" scenarios. Before the discovery of extrasolar planets, core accretion models received much attention because simulations could not recreate our solar system configuration via gravitational instability models. This is because gravitational instability operates similar to the star formation process, where turbulent material collapses under its own gravity as enough material occupies a small volume. This is thought to sometimes happen in the outer regions ($\sim$100 AU) of massive protoplanetary disks (Boss, 1997; Boley, 2009), much farther out than the locations of the terrestrial planets. However, it was proposed that giant planets may form this way then migrate inwards as the system dynamically evolves. This could potentially explain giant planet formation in our solar system (Cameron, 1978; Boss, 1997) but complications arise when extending this to terrestrial-like planets that orbit closer to their host star, similar to Mars, Earth, Venus, and Mercury.

Terrestrial-like planets are thought to be rocky and substantially comprised of refractory materials, with a thin envelope of gas created primarily by outgassing of material as the planet cools in the later stages of formation. It is unlikely that the terrestrial planets of our solar system are the cores of migrated giant planets, since it is not clear what happens to a giant planet's core as it evolves, and if the mass-loss required to strip a giant planet down to a terrestrial-like planet is feasible. On the other hand, core accretion, the process by which a planetesimal nucleus grows in mass as it draws in and collects material while orbiting in the protoplanetary disk, could create gas-giants and terrestrial-like planets (Pollack et al., 1996).

After the discovery of the population of planets now widely known as hot-Jupiters during early RV and transit surveys, gravitational instability came back into the picture because core accretion was not able to reproduce these heavily irradiated gas giants at extremely close orbital separations, and in some cases, RV detected planets at farther distances on eccentric orbits. It was then proposed that hot-Jupiters may have formed in the outer cold regions of their circumstellar disk via gravitational instability, and mechanisms such as smooth disk migration then

worked to dissipate orbital energy, causing these hot-Jupiters to migrate inward (Bodenheimer et al., 2000; Ida & Lin, 2004). Moreover, it has been observed that hot-Jupiters typically do not have nearby planets or planets on interior orbits, perhaps because these planets were somehow cleared from the system during epochs of dynamical evolution.

Interestingly, Rossiter-McLaughlin observations have exposed some systems where host star rotations are highly mis-aligned with the planetary orbital disk. This favors a scenario where planet-planet scattering excites the eccentricities of two giant planets causing one planet to be ejected or thrown to the outer regions of the planetary system, and one planet to have an excited eccentricity with a small pericenter distance (Rasio & Ford, 1996). This sets up the inner planet of the scattering event to under go tidal circularization and potentially end up as a hot-Jupiter that has cleared out any planets on interior orbits. Furthermore, the planets produced in this scenario could potentially orbit retrograde, a feature that is sometimes present in the hot-Jupiter population (Sanchis-Ojeda et al., 2013). This paints a picture where smooth disk migration may operate at the initial stages of the dynamical evolution, and planet-planet scattering events and secular perturbations act on the system at later stages. These scenarios were heavily developed as they could reproduce the observed hot-Jupiter population at a time in history when most of the detected planets were hot-Jupiters.

From hot-Jupiters to sub-Neptunes

After the *Kepler* mission commenced in early 2009, we were able to untangle the selection effects for transit surveys that favored the detection of primarily hot-Jupiters and begin to acquire a sample that more accurately represents exoplanetary systems around solar-like stars in our vicinity of the Milky Way. With the *Kepler* mission producing a sample of over 4600 planet candidates to date (Akeson et al., 2013), researchers have learned that the largest population of detected planets may in fact be similar to a version of a mini Neptune, with orbital separations less than 0.5 AU (Fressin et al., 2013). Hot-Jupiters now only comprise a small fraction of the detected exoplanets. Additionally, there seem to be a plethora of systems containing multiple planets that have orbital architectures somewhat similar to the architectures of the gas giant and satellite systems within our solar system, e.g., the Kepler-11 system (Lissauer et al., 2011). Many of these sub-Neptune

exoplanets reside in these multiple-planet systems, on short orbital periods, and closely packed together. These are often referred to as Short-period Tightly-packed Inner Planetary Systems (STIPS). Due to the close proximity of the planets in STIPS, the gravitational interactions between planets can cause mutual planetary perturbations where their orbits deviate from strict periodicity. This feature can be measured by tracking the variations of the time each transit occurs and comparing those times to the calculated orbital period if the planet were to be unaffected by the other planets in its system. These measurements are referred to as transit timing variations (TTVs) (Ford et al., 2012b).

Short-period Tightly-packed Inner Planetary Systems

Many of the multiple transiting planet systems discovered by *Kepler* exhibit TTVs. Most now have randomly distributed period ratios but statistically a significant sample orbit near to mean motion resonances (MMR) with each other. An MMR occurs when a pair of planets have orbital periods that are integer ratios of one another. A first-order MMR, in particular, has an orbital period ratio of $j : j - 1$ where j is known as the *resonant number*. In cases where STIPS are in or near MMR, the TTV signals are large. These TTV measurements have provided us with the means to get precise measurements and perform detailed analysis to characterize the orbital and physical properties of these sub-Neptune planets and STIP systems. For instance, the TTV amplitudes can inform us about the planet masses and eccentricities (Lithwick et al., 2012), and when coupled with the planet radii measurements, can provide much insight. Using this information, we can begin to untangle the planet formation process for STIP systems and how it may be similar to or differ from the formation of hot-Jupiters and the planets in our solar system. As RV surveys became more sensitive to giant planets with longer orbital periods, a dearth of giant planets between 10 and 200 day orbital periods was discovered. However, there is a plethora of sub-Neptune planets in this orbital period range. Why a lack of giant planets has been observed in this parameter space is an open question.

One of the striking features from the *Kepler* sample of planets is a statistically significant pile up of planets in orbital configurations just wide of MMR, and a gap just narrow (Lissauer et al., 2011; Fabrycky et al., 2014; Steffen & Hwang, 2015). This is different than what was observed for the gas giants detected in

radial velocity surveys, where these planets seem to have stayed very close to MMR. Efforts have been put forth to explain this feature, including the theory of resonant repulsion (Lithwick & Wu, 2012). Resonant repulsion requires that after these planets have migrated into MMR with each other, potentially via smooth disk migration, a dissipative mechanism transfers orbital energy into heat and pushes the planets apart. It was initially proposed that tides could act as this dissipative mechanism, but recent work has shown that for the sub-Neptune mass planets involved in these configurations, the lower mass compared to hot-Jupiters, and the large initial eccentricities required for the tidal timescale to be less than the age of the system, are unreasonable (Silburt & Rein, 2015). Recent work to characterize this dissipative mechanism was put forth by Chatterjee & Ford (2015), where after these lower mass planets have migrated into MMR via smooth disk migration, planet-planetesimal disk interactions create many micro-scattering events that can dissipate orbital energy on appropriate timescales to move these planets wide of MMR. Higher planet masses may not be as susceptible to these micro-scattering events, and the planet-planetesimal disk scattering mechanism may explain why we see many giant planet systems detected by radial velocity surveys still in MMR.

1.2 The Kepler Space Science Mission

The *Kepler* Space Science Mission has set a high standard for the future in space exploration and scientific discovery as it carried out near-continuous photometry for 4 years using a 0.95-m aperture space-based telescope to monitor more than 150,000 stars selected to be similar in nature to the Sun (Borucki et al., 2010; Koch et al., 2010). As of August 27th 2015 *Kepler* has discovered 4,696 exoplanet candidates, 472 multiple transiting planet systems, and has confirmed 1,887 exoplanets (Akeson et al., 2013). The *Kepler* mission set out to explore the structure and diversity of planetary systems, with the goal of discovering the frequency of Earth-like planets in the habitable zone of Sun-like stars ($\eta_{\oplus}$). The *Kepler* mission also seeks to characterize the orbits and properties of single and multiple planetary systems, as well as characterize the properties of their host stars. Figure 1.2 shows planet orbital period versus planet radius for confirmed planets (blue) and planet candidates (red) detected by *Kepler*.

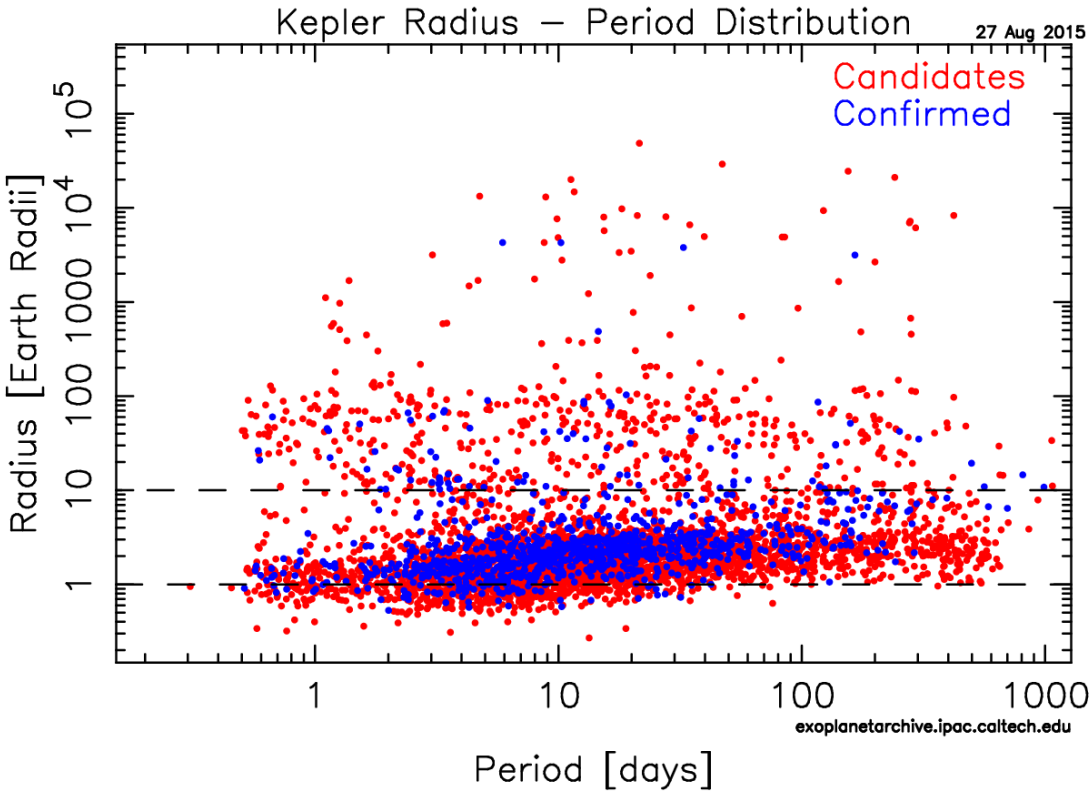

Figure 1.2: Planet orbital period versus planet radius for *Kepler* planets. Confirmed planets are shown in blue, and planet candidates are shown in red. Detection of an Earth-radii planet around a solar-like star on a year long orbit is still to be determined within the baseline *Kepler* data products.

Science with *Kepler* Data

Teams are contributing to innovative, high-performance and reliable exploration technologies while developing the analysis required for the full realizations of the *Kepler* mission objectives. After the launch of the mission in March 2009, the realization that the *Kepler* field stars and the *Kepler* instrument introduce significantly more noise in the photometry data than expected, the primary mission objective of discovering an Earth analog became jeopardized (Gilliland et al., 2011). In 2012, the baseline *Kepler* mission was extended from 3.5 years to 7.5 years to reach the required signal-to-noise ratio (SNR) for detection of potential Earth-like planets on year long orbits within the data. Soon after this extension, one of the remaining three reaction wheels required to keep *Kepler* stable for its remarkable photometric precision, failed. However, the current data is still prolific! The

substantial sample of small planets on short period orbits has revolutionized our understanding of planetary systems architecture. Furthermore, the *Kepler* data has provided the opportunity to employ advanced probabilistic analysis for discovering more from the data than meets the eye.

Initially, the focus of research being done to characterize the *Kepler* sample was largely to constrain the occurrence rates of various planetary mass regimes relative to the planets in our solar system. In particular, many studies examined the occurrence rates of hot-Jupiters, sub-Neptunes, and super-Earths, and how these might correlate with orbital separation (e.g., Fressin et al. (2013); Howard et al. (2012, 2010)). Other work generated histograms of the number of planets in these mass regimes versus semi-major axis, orbital period, planet radius, and stellar effective temperature to learn about parameter distributions for these populations, but mainly as preliminary analysis with the aim to catalog the *Kepler* objects of interests (Borucki et al., 2011a).

Transit timing variations have been detected in dozens of near-resonant systems from *Kepler*, providing access to the mass of small bodies otherwise unattainable with current radial velocity (RV) technology (e.g., Ford et al. (2012b,a, 2011); Steffen et al. (2013); Mazeh et al. (2013); Rowe & Thompson (2015a)). The eccentricity distribution for terrestrial planets discovered by *Kepler* has been explored (Ford et al., 2008; Moorhead et al., 2011; Kane et al., 2012; Plavchan et al., 2014a), however, these studies are limited by uncertainties in host star parameters. In Chapter 2 of this dissertation, we investigate the eccentricity distribution of a sample of short-period planet candidates from *Kepler*, using measurements of the eccentricity that are largely independent of the host star properties.

Selection effects and detection biases are prevalent in the *Kepler* sample. However, much work has been done to catalog the combined differential photometric precision (CDPP) and the *Kepler* pipeline detection efficiency. Accurate estimates of the uncertainties of transit parameters are dependent on the photometric measurement error as a function of time, which is approximated by the CDPP. Furthermore, properly characterizing the detection efficiency is crucial to characterizing the *Kepler* sample as a representative sample of planets (Jenkins et al., 2010b; Christiansen et al., 2012, 2013; Christiansen, 2014). The false positive rate has been shown to be low (Fressin et al., 2013; Batalha et al., 2013), suggesting that preliminary population analysis of the *Kepler* planet candidates is informative

in spite of many of the planet candidates still awaiting confirmation.

Many researchers have begun the efforts to mine the *Kepler* data for existing evidence of Earth analogs by working to improve the pipeline sensitivity (Rowe & Thompson, 2015a; Seader et al., 2015; Tenenbaum et al., 2014; Burke et al., 2014; Tenenbaum et al., 2013, 2012). Some studies extrapolate the occurrence rates to parameter space with smaller planets on longer orbital periods, extracting improved estimates of the number of Earth-like planets in the same period and radius bin as Earth (Foreman-Mackey et al., 2014; Petigura et al., 2013b,a; Dong & Zhu, 2013). These studies, however, require strong assumptions that weaken the validity of their inferences. Other efforts have begun work characterizing populations in the *Kepler* sample but have failed to robustly and comprehensively account for the selection biases in the planet vetting process, the false positive rates, the pipeline detection efficiency, or combinations of these issues (Rogers, 2015a; Wolfgang et al., 2015; Weiss & Marcy, 2014; Wolfgang & Lopez, 2015; Foreman-Mackey et al., 2014; Fressin et al., 2013; Weiss et al., 2013). Without properly accounting for the impact of selection effects, detection-efficiency/completeness and the false-positive-rates/reliability of the *Kepler* sample, the estimates for $\eta_{\oplus}$ and other population level science will remain significantly less accurate.

Kepler's **Observables**

Kepler provides us with high precision broadband photometry at long and short cadences for star-planet occultation events where the light curves can be studied in extreme detail for many planetary systems. The transit depth tells us information about the radius of the planet relative the radius of the host star, allowing us to infer the planet's size. Additionally, we can measure the transit duration, the ingress and egress duration, and the time of reoccurring mid transit and secondary eclipse. Combining these pieces of information using Kepler's laws, we can learn about the planet's impact parameter (the chord that the planet traverses over the face of the star in contrast to the full diameter of the star), the eccentricity of the planetary orbit, and the orbital period/semi-major axis, provided we have information about the stellar mass and radius. This is described in the following relation:

$$D = \rho_* sqrt(1 - b^2)(a/r_+(e\cos\omega, e\sin\omega))^{-1} \tag{1.1}$$

where D is the transit duration, ρ_* is the stellar density, b is the impact parameter, a is the semi-major axis, and r_+ is the distance the planet is from the star at the time of transit and is a function of the orbital eccentricity, and argument of periastron. Constraining these parameters for the *Kepler* sample will provide robust opportunities to explore how orbital parameters are correlated with different planetary sub-populations. For instance, finding how eccentricity changes with planet size, stellar type, stellar metallicity, orbital period, effective temperature, multiplicity and more can shed light on the connection between star formation and the formation of planetary companions, and disentangle degenerate formation scenarios described in §1.1.

The high photometric precision also allows unprecedented stellar characterization through asteroseismology. Eclipsing binary stars can also be studied in detail with *Kepler data*. Stellar evolution is largely coupled with planet formation and dynamical evolution. For example, tidal circularization plays an important role in sculpting of eclipsing binaries and planetary systems orbital properties. The behavior of such systems share common dynamical processes, but operate in different mass and composition regimes, offering an opportunity for comparison. Thus, investigating properties of EBs statistically can shed light onto star and planet formation. The end state of EBs after billions of years of tidal evolution provide a way to constrain the internal structure of stars and planets, as well as the physics of tidal dissipation.

1.3 Hierarchical Bayesian Modeling

The hierarchical Bayesian modeling method has been employed in this dissertation to begin the processes of robustly characterizing populations of exoplanets from the *Kepler* sample. Bayesian modeling accounts for the conditional structure of the results of random processes in nature. Knowing more about the behavior of a system and using that information when developing the population model produces more accurate inferences of the parameters of interest. In a hierarchical Bayesian framework, this philosophy is extended from individual systems to populations of systems for which we have measurements. In this regime, the population-level parameters are simultaneously inferred along with probabilistic parameter estimates for the population constituents.

The population-level parameters being inferred describe the distributions of star

and planet properties while the parameters describing the observables contain the actual properties of one particular star and planetary system, as drawn from an assumed underlying distribution using the population-level parameters. In the case of population inference for the *Kepler* sample, the physical model for a transiting planet light curve relates the observables, (e.g., stellar radius, stellar density, standard deviations of out-of-transit flux over various time scales, transit depth, reference transit time, transit duration) to the planet and host star properties (e.g., the observable parameters plus planet-to-star radius ratio, eccentricity, argument of pericenter, and inclination).

Hierarchical Bayesian inference operates by leveraging the measurement uncertainties and specifying distributions at each level. Information and uncertainty are characterized in the resulting posteriors of the population level parameters that describe the chosen distribution's size and shape (e.g., the population level parameters of a Normal distribution population model are the mean and variance). Robust hierarchical Bayesian models are constructed in such a way that the results for the posteriors for the population level parameters are insensitive to outliers or other small departures from the model assumptions.

Formalism

The posterior probability distribution function for the vector of hyperparameters $\boldsymbol{\phi}$ that describe the population level parameters in an HB model and the latent variables $\boldsymbol{\theta}$, is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{d}, M) = \frac{p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi}, M)p(\boldsymbol{\theta}|\boldsymbol{\phi}, M)p(\boldsymbol{\phi}, M)}{\int d\boldsymbol{\phi}\, p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi}, M)p(\boldsymbol{\theta}, \boldsymbol{\phi}|M)}. \tag{1.2}$$

Here, "|" is read as "given", $\mathbf{d}$ is the vector of data being used to constrain the model. $\boldsymbol{\theta}$ is the vector of latent variables parameterized by the intrinsic distributions for the observables such as star and planet properties surveyed, parameters that describe the signal to noise ratio and other features of the data collection method, and explanatory variables that are given as known or define relationships between observables and latent variables. M represents additional information such as the model specification inherent to the particular HB model. The prior probability distribution function for $\boldsymbol{\phi}$ is $p(\boldsymbol{\phi})$, and is also referred to as the hyperprior. The probability of measuring the data $\mathbf{d}$ given the model parameters $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, and

model specification M is $p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi}, M)p(\boldsymbol{\theta}|\boldsymbol{\phi}, M)$, and is commonly referred to as a likelihood for the hierarchical model with fixed $\boldsymbol{\phi}$. The denominator in Equation 1.2 is the normalization, which is important in model comparison, but not important when evaluating population features using robust model specification. Calculating the normalization for an HB model is non-trivial. Methods such as importance sampling can be used to accomplish this, however new methods are continuing to be developed.

Advantages of HB Models

HB modeling leverages the power of the ensemble of data, where the observed quantities are dimensions in the HB model. As a consequence, the joint posterior for these large dimensional problems provides marginal distributions for the observed variables in addition to the population level parameters. These probabilistic parameter estimates for the observed variables are often more precise than probabilistic parameter estimates that are not from a hierarchical framework, a phenomenon known as “shrinkage" (Loredo, 2007). Moreover, the practice of non-hierarchical models for probabilistic parameter estimation of hierarchical data can lead to further problems. In these cases, models developed for parameter estimation may over-fit data. This is exposed when these models make inferior predictions for new data (Gelman et al., 2004).

Another advantage to HB modeling is the incorporation of each observable's unique measurement uncertainty into the inference of the population parameters, and parameter estimates of the population constituents. This allows for the possibility of robust inference even when the measurement uncertainties are large. Furthermore, HB modeling can produce parameter estimates for quantities that are not observed by incorporating physical models into the framework.

Evaluating HB Models

HB models are fundamentally high-dimensional problems, often without analytical solutions. In this dissertation, we focus on numerically calculating the posteriors for our HB models using Monte Carlo methods. These methods impose some limitations in that Monte Carlo error is convolved in the posterior distributions for the parameter estimates. Monte Carlo error goes as $1/\sqrt{N}$, where N is the number

of iterations or draws from the HB model. Therefore, it is important that we test our models for convergence. We do a first pass analysis to test for convergence by looking at trace plots, which show the parameter estimate as a function of iteration for each chain in the MCMC simulations. Chains that are well-mixed sample parameter space effectively. We also evaluate the autocorrelation of each the Markov Chains in our Markov Chain Monte Carlo simulations as a diagnostic for convergence. The autocorrelation is the correlation or similarity between the parameter estimate for iterations as a function of the distance between iterations, or lag time. Furthermore, we evaluate the Gelman-Rubin statistic $\hat{R}$, which evaluates the across chain variance to test whether all parallel chains converge to the same posterior distribution (Gelman et al., 2004).

1.4 Outline

Chapter 2 examines a sample of short-period planet candidates from *Kepler* for which both a transit and an occultation were observed. Here, we characterize the eccentricity distribution of a sample of 50 short-period planet candidates using transit and occultation measurements from NASA's *Kepler* Mission. First, we evaluate the sensitivity of our hierarchical Bayesian modeling and test its robustness to model misspecification using simulated data. When analyzing actual data assuming a Rayleigh distribution for eccentricity, we find that the posterior mode for the dispersion parameter (the Rayleigh parameter) is $\sigma = 0.081 \pm^{0.014}_{0.003}$. We find that a two-component Gaussian mixture model for $e\cos\omega$ and $e\sin\omega$ provides a better model than either a Rayleigh or Beta distribution. Based on our favored model, we find that $\sim 90\%$ of planet candidates in our sample come from a population with an eccentricity distribution characterized by a small dispersion (~ 0.01), and $\sim 10\%$ come from a population with a larger dispersion (~ 0.22). Finally, we investigate how the eccentricity distribution correlates with selected planet and host star parameters. We find evidence that suggests systems around higher metallicity stars and planet candidates with smaller radii come from a more complex eccentricity distribution.

Chapter 3 describes an analysis of a sample of planet-pairs from *Kepler* in near first-order mean motion resonances with significant transit timing variation signals. We characterize the mass-radius-eccentricity distribution of this sub-population of

transiting planets using Transit Timing Variations (TTV) observations from NASA's *Kepler* mission. *Kepler's* precise measurements of transit times (Mazeh et al., 2013; Rowe & Thompson, 2015a) constrain the planet-star mass ratio, eccentricity and pericenter directions for hundreds of planets. In particularly favorable cases, strongly-interacting planetary systems allow TTVs to provide precise measurements of masses and orbital eccentricities separately (e.g., Kepler-36, Carter et al. (2012)). For each of these precisely characterized systems, there are several planetary systems harboring at least two planets near a mean motion resonance (MMR) for which TTVs provide a joint constraint on planet masses, eccentricities and pericenter directions (Hadden & Lithwick, 2014). Unfortunately, a near-degeneracy between these parameters leads to a posterior probability density with highly correlated uncertainties. Nevertheless, the population encodes valuable information about the distribution of planet masses, orbital eccentricities and the planet mass-radius relationship. We characterize the distribution of masses and eccentricities for near-resonant transiting planets by combining a hierarchical Bayesian model with an analytic model for the TTV signatures of near-first-order-resonant planet-pairs (Lithwick et al., 2012) to accelerate exploration of this complex parameter space. By developing a rigorous statistical framework for analyzing the TTV signatures of a population of planetary systems, we significantly improve upon previous analyses. For example, our analysis includes transit timing measurements of near-resonant transiting planet-pairs regardless of whether there is a significant detection of TTVs, thereby avoiding biases due to only including TTV detections.

Chapter 4 applies the methods from our eccentricity distribution analysis presented in Chapter 2 on a sample of 795 eclipsing binary (EB) systems, for which we have primary and secondary eclipses projected eccentricity measurements. First, we apply a two-component Gaussian mixture HB model to the sample of EB projected eccentricities. In this HB model regime, we find that the mixture fractions and dispersions are $f_{low} = 53 \pm^{2.2}_{1.6}$ %, $f_{high} = 47 \pm^{1.8}_{1.9}$ %, and $\sigma_{low} = 0.0067 \pm^{0.0003}_{0.0003}$, $\sigma_{high} = 0.21 \pm^{0.007}_{0.005}$, respectively. Here, "low" refers to the smaller dispersion component and "high" refers to the component where EBs may have significant eccentricities. Next, we apply an HB model using a piece-wise function, where the parameters for the eccentricity distribution depend on whether the orbital period are below or above a threshold in orbital period. In this scenario, the threshold, which we call the period break point, is simultaneously inferred along with the

eccentricity distribution for each sub-population below and above the break point. In the case where we allow the eccentricity distribution on either side of the period break point to be modeled as a mixture of two Gaussian distributions, we find that the inferred period break point is at $10.74\pm^{0.26}_{0.027}$ days. This is consistent with previous findings from Latham et al. (2002); Mayor et al. (2001), where they observed a period break point at around 10 days for EBs from disk and halo parent populations, indicative of tidal circularization operating on this population of short-period EBs. Furthermore, the eccentricity distribution of EBs below ~ 10.74 days, has population level parameters $f_{low} = 0.72\pm^{0.024}_{0.022}$, $f_{high} = 0.28 \pm^{0.025}_{0.02}$ %, and $\sigma_{low} = 0.0055\pm^{0.0004}_{0.0002}$, $\sigma_{high} = 0.1\pm^{0.0047}_{0.0048}$ for the mixture fractions and dispersions, respectively. The eccentricity distribution for EBs with periods above ~ 10.74 days has population level parameters $f_{low} = 0.13\pm^{0.024}_{0.023}$, $f_{high} = 0.87 \pm^{0.027}_{0.02}$ %, and $\sigma_{low} = 0.018\pm^{0.0035}_{0.0018}$, $\sigma_{high} = 0.26\pm^{0.0074}_{0.011}$ for the mixture fractions and dispersions, respectively. This suggests that $\sim 72\%$ of EBs below the inferred period break point are very circularized, where as $\sim 87\%$ of EBs above the inferred period break point can take on a wide range in eccentricity values including some with significant eccentricities.

Chapter 5 summarizes our results and discusses our conclusions from these projects, as well as discusses future work. As high-fidelity data products from the *Kepler* Mission near completion, timely development of probabilistic population inference to match these advanced data products can compensate for the abbreviated mission and characterize the exoplanet population more accurately than any previous analysis to date. The probabilistic nature of the analysis presented in this dissertation, and applicable for future population characterization studies, lends itself well to predictive studies such as exoplanet population synthesis.

Chapter 2
The Eccentricity Distribution of Short-Period Planet Candidates Detected by *Kepler* in Occultation

2.1 Introduction

The *Kepler* mission has identified a sample of planet candidates detected both in transit and occultation, providing detailed orbital information, including orbital eccentricity, for a subset of systems with a wide variety of stellar host properties. However, early works on the eccentricity distribution of all *Kepler* objects of interest (KOIs), including those in this subset, are limited due to uncertainties in host star properties. Recent studies have focused on applying Bayesian data analysis for robust error estimation (e.g., Parviainen et al., 2013), and other studies have investigated the eccentricity distribution of planets discovered with the radial velocity technique and the role that tidal interactions play in shaping eccentricity distributions (Wang & Ford, 2011; Matsumura et al., 2008; Hansen & Murray, 2015). While some studies have attempted to constrain the eccentricity distribution of planets via transit durations identified by *Kepler*, these studies have been limited by uncertainties in stellar densities (Moorhead et al., 2011; Kane et al., 2012; Plavchan et al., 2014a; Van Eylen & Albrecht, 2015). Lucy (2013) used a Bayesian approach to explore the eccentricity distribution of eclipsing binaries. Kipping

(2014a) explored biases in an eccentricity distribution using a Beta distribution prior, but little else has been done to explore the eccentricity distribution of exoplanets via similar methods and with the goal of quantifying population-level parameters. Hogg et al. (2010) proposed using an hierarchical Bayesian (HB) model to constrain the eccentricity distribution of hot Jupiters, but applied their model to simulated radial velocity observations, only.

Bayesian inference has made its way into exoplanet studies as computing facilities have evolved to accommodate the required calculations. The application of HB modeling is highly relevant for studying the *Kepler* planet sample (e.g. Demory, 2014; Wolfgang & Lopez, 2015; Rogers, 2015b; Foreman-Mackey et al., 2014). This framework allows us to obtain population-level posterior distributions, such as the distribution function for planets, while accounting for measurement uncertainties and potentially, selection effects. HB is particularly well suited for characterizing a population's eccentricity distribution largely because of its ability to accommodate samples where each measurement has a large measurement uncertainty.

As a first step in studying the exoplanet population in general, we use HB modeling to investigate the eccentricity distribution of the subset *Kepler* planet candidates that are detected in both transit and occultation, which provides measurements of projected eccentricity via transit duration ratios and phase offsets. Even for this subset of planet candidates, individual eccentricity measurements often have large uncertainties. Fittingly, HB is designed to account for individual measurement uncertainties. Thus, we approach this problem from both sides: we will apply modern statistical methods that incorporate uncertainties into our eccentricity study (e.g., HB modeling), while also working with a subset of planet candidates with enough information to help bypass some of the uncertainty in their host star parameters. The projected eccentricity measurements ($e\cos\omega$ and $e\sin\omega$) are presumed independent of the stellar host star density and radius, which mitigates the problem of uncertainties in stellar parameters. Applying HB to the eccentricity distribution is a logical starting point while working to construct a comprehensive hierarchical model (i.e., a joint population distribution that includes planet parameters in addition to orbital eccentricity) in which, measurement uncertainties are naturally incorporated into the analysis.

Furthermore, we can investigate various sub-populations of planets from the *Kepler* sample and look for correlations of planet and host star properties within

these subpopulations. In particular, we explore mixture models where the eccentricity distribution can be interpreted as a combination of two sub-populations. With this analysis, more than one population in the eccentricity distribution could arise, for example, due to different formation mechanisms at work. Characterizing the eccentricity distribution in this way provides insight into postulated planet formation theories such as planet orbital migration and planet scattering. In principle, these mechanisms could form two populations that make up the eccentricity distribution: one population that evolved via slow disk migration and another population that evolved via excitation of a large eccentricity (e.g., planet-planet scattering or secular perturbations) proceeded by tidal circularization. With this in mind, the population of planets that came from planet-planet scattering might have a larger dispersion as it would include planets with large eccentricities, while the population of planets that came from disk migration might have a smaller dispersion and contain fewer eccentric planets. These populations might also correlate with host star properties, which would allow for a framework to test physical models of the origin of each population thus shedding light onto planet formation.

Here, we focus on inferring the eccentricity distribution of an interesting subset of planets using HB modeling applied to both simulated and real transit and occultation measurements from the *Kepler* mission. This sample contains predominantly short-period planet candidates, most of which are likely to be hot Jupiters, identified by *Kepler*. We look for correlations between the eccentricity distribution and other properties, such as stellar effective temperature, planet radius, orbital period, and stellar metallicity to begin to synthesize a global understanding of planet formation. This manuscript is organized as follows. In §2.2, we describe our observational data. In §2.3, we describe the method behind the HB analysis calculations, and the priors selected for the study. In §2.4, we present the results of our HB analysis. In §2.5, we investigate potential correlations between the eccentricity distribution and planet or host star properties. In §2.6, we summarize our results, and in §2.7, we discuss our conclusions, potential biases and future work.

2.2 Observations

When a planet both transits and occults its host star, we are able to obtain detailed information about the planetary orbit, including information about the projected

orbital eccentricity, $h = e\cos\omega$ and $k = e\sin\omega$. The relationship between orbital eccentricity and transit observables is outlined in Winn (2010). h can be derived from

$$\Delta t_c \approx \frac{P}{2}\left[1 + \frac{4}{\pi}h\right], \tag{2.1}$$

where Δt_c is the time between the center of the transit and the center of the occultation and P is the orbital period. k can be derived from

$$\frac{T_{\rm occ}}{T_{\rm tra}} \approx \frac{1+k}{1-k}, \tag{2.2}$$

where T_{occ} is the occultation duration and T_{tra} is the transit duration. More precise expressions for h and k are listed in Ragozzine & Wolf (2009), section 2.5 and elsewhere. The original full derivations of these expressions can be found in Sterne (1940) and de Kort (1954). When analyzing *Kepler* observations in §2.5, we calculate the transit and occultation times and durations numerically using Keplerian orbits.

We have measured the offsets and durations of transits and occultations for a sample of planet candidates observed by *Kepler*. This study is based on quarters Q0 through Q12 *Kepler* data (see Burke et al., 2014, for the Q0-Q10 data release). In total, the datasets encompass about ∼1100 days of quasi-continuous photometric monitoring between May 2009 and March 2012. We retrieved the Q0-Q12 FITS files from MAST[1] and extracted the PDCSAP $_{FLUX}$, commonly known as "calibrated light curves", long-cadence photometry (Jenkins et al., 2010a) for each target. Using the calibrated data eliminates the potential for instrumental corrections or cotrending basis vectors to introduce short-timescale correlated noise.

2.2.1 Derivation of the planet physical and orbital properties

We focus on a sample of planet candidates for which an occultation is detected. We address the biases that the selection effects introduce into our sample in §2.3.2 and §2.7. We use the *Kepler* planet candidate list to keep all planets larger than 8 Earth radii and with orbital periods less than 10 days. This initial selection of planet candidates was based on early planet candidate lists from NExSci. Note that most planets with a detectable occultation have very high SNR transits, so we do not

[1] http://archive.stsci.edu/kepler/

expect that many additional planets would be found in the full Q0-Q17 datasets. The preliminary parameters were derived using the KIC stellar values, and updated later in our analysis. We employ a Markov Chain Monte Carlo (MCMC) framework to compute the posterior distribution of the system's orbital parameters using these initial values. When performing MCMC analysis, we used an empirical main sequence mass-radius relationship (Torres et al., 2009) to derive more accurate planetary parameters. After the MCMC analysis was performed, some planet candidate radii changed to be outside the initial range stated above. Our MCMC implementation (described in Gillon et al. (2012)) uses the Gibbs sampler and the Metropolis-Hastings algorithm to estimate the posterior distribution function of all unknown parameters. Our nominal model is based on a star and a single transiting planet on a Keplerian orbit about their center of mass.

The input data provided to each MCMC run consist of the Q0-Q12 *Kepler* photometry and the stellar parameters (effective temperature T_{eff}, metallicity $[Fe/H]$ and spectroscopic $\log g$) extracted from the *Kepler* Input Catalog (KIC) (Brown et al., 2011). We correct for the photometric dilution induced by neighboring stellar sources using a quarter-dependent dilution factor based on the dilution values presented in the literature and on the contamination values reported in the FITS files headers (Bryson et al., 2013).

We divide the total light curve in segments of duration $\sim$24 to 48 hrs. The smooth photometric variations due to stellar variability or instrumental systematic effects in each segment are fit with a time-dependent quadratic polynomial. Baseline polynomial coefficients are determined at each step of the MCMC for each light curve with the singular value decomposition method. The resulting coefficients are then used to correct the calibrated photometric light curves. We assume a quadratic law for the limb-darkening (LD) and use $c_1 = 2u_1 + u_2$ and $c_2 = u_1 - 2u_2$ as jump parameters, where u_1 and u_2 are the quadratic coefficients (Mandel & Agol, 2002). We integrate over the 29.4 minute long cadence integration time when modeling long cadence light curves.

The MCMC has the following set of jump parameters (i.e., parameters that are not fixed in our model and are used as a basis for proposal steps): the planet/star flux ratio, the impact parameter b, the transit duration from first to fourth contact, the time of minimum light T_0, the orbital period, the occultation depth, the two LD combinations c_1 and c_2 and the two parameters $\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$. At

each step of the MCMC, the Keplerian model is constructed based on the e and ω values derived from the $\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$ jump parameters. A uniform prior distribution is assumed for all jump parameters except c_1 and c_2. This corresponds to a prior that is uniform in $e \in [0, 1)$ and $\omega \in [0, 2\pi)$. For the limb-darkening parameters, we assume normal priors which are centered on values of c_1 and c_2 that correspond to the values of u_1 and u_2 from the theoretical tables of Claret & Bloemen (2011) for the stellar parameters obtained from the KIC. The standard deviation of the priors for c_1 and c_2 were set by the corresponding standard deviations propagated from u_1 and u_2's uncertainties. We run two Markov chains of 100,000 steps for each planet candidate. The mixing and convergence of the Markov chains are assessed using the Gelman-Rubin statistic criterion (Gelman & Rubin, 1992). Results for $e\cos\omega$ and $e\sin\omega$ are shown in Table 2.1.

Table 2.1: Result for $e\cos\omega$ and $e\sin\omega$ from MCMC. See §2.1 for details on how these values are calculated.

KOI	$e\cos\omega$	$\sigma_{+e\cos\omega}$	$\sigma_{-e\cos\omega}$	$e\sin\omega$	$\sigma_{+e\sin\omega}$	$\sigma_{-e\sin\omega}$
13.01	0.00379	0.00073	0.00073	0.32343	0.01569	0.01559
17.01	−0.00038	0.03414	0.02824	−0.00144	0.04517	0.04379
18.01	0.02051	0.03321	0.03515	−0.00965	0.04659	0.05250
20.01	−0.01868	0.05117	0.04820	0.00010	0.05112	0.05023
22.01	−0.01123	0.05055	0.05328	−0.00137	0.04962	0.05033
97.01	−0.00353	0.02024	0.02003	−0.00336	0.03877	0.04284
98.01	0.00993	0.04816	0.04114	0.00040	0.04973	0.04662
127.01	0.02353	0.05076	0.05854	0.00120	0.05051	0.05272
128.01	−0.01801	0.03372	0.03033	0.00044	0.04655	0.04697
131.01	−0.01178	0.04109	0.04290	−0.00036	0.04971	0.04784
135.01	−0.04691	0.04427	0.03340	0.00148	0.05046	0.05497
183.01	0.01357	0.05300	0.04929	0.00246	0.05187	0.05141
186.01	−0.01337	0.04675	0.04576	0.00049	0.05070	0.04854
188.01	0.00430	0.03649	0.03731	−0.00252	0.04584	0.04577
200.01	0.01366	0.05739	0.06048	−0.00142	0.05184	0.05280
202.01	0.00563	0.04669	0.04344	−0.00377	0.04464	0.04659
203.01	0.05014	0.02115	0.02894	0.00126	0.05447	0.05276
204.01	0.01586	0.04639	0.05328	0.00028	0.05031	0.04986

206.01	−0.01727	0.05781	0.06547	−0.00069	0.05229	0.05331
254.01	−0.03065	0.05882	0.05747	−0.00141	0.05432	0.05094
421.01	0.00351	0.05653	0.05293	−0.00169	0.05077	0.05120
607.01	−0.0027	0.04021	0.03915	−0.00011	0.04710	0.04763
611.01	0.03344	0.05965	0.05111	0.01742	0.05413	0.05283
728.01	0.00330	0.04798	0.05457	0.00366	0.05367	0.05392
760.01	0.01367	0.03489	0.03363	0.00139	0.04686	0.04647
767.01	−0.00490	0.04486	0.04563	−0.00099	0.05053	0.04772
774.01	−0.16220	0.00638	0.00332	−0.00802	0.06739	0.05652
791.01	0.01588	0.02676	0.02860	0.00223	0.04661	0.04651
797.01	0.04823	0.03688	0.05262	0.01784	0.06601	0.05972
801.01	0.02502	0.05106	0.04972	−0.00221	0.05348	0.05067
805.01	0.38761	0.00080	0.00115	0.02390	0.02642	0.02648
823.01	−0.00629	0.00274	0.00284	−0.35185	0.01225	0.01219
830.01	0.00997	0.05621	0.05942	−0.00098	0.05115	0.05116
850.01	−0.01543	0.05830	0.06948	0.00094	0.05208	0.05235
883.01	−0.02412	0.06865	0.06509	0.00065	0.05498	0.05499
890.01	0.00621	0.05985	0.03740	0.00096	0.04949	0.04824
895.01	−0.06154	0.01675	0.01243	−0.00275	0.06048	0.05811
897.01	0.00571	0.05530	0.04969	0.00083	0.04771	0.05155
908.01	−0.00760	0.04825	0.04482	−0.00138	0.04992	0.04960
913.01	0.00433	0.05289	0.04532	0.00276	0.04787	0.05067
929.01	0.00446	0.03132	0.03626	-2×10^{-05}	0.04638	0.04677
931.01	−0.01950	0.06144	0.06634	0.00042	0.05485	0.05313
1066.01	−0.05694	0.06242	0.02904	−0.00046	0.05689	0.05470
1176.01	0.01372	0.04534	0.04789	−0.00070	0.04914	0.04876
1227.01	0.00424	0.03187	0.04877	−0.30367	0.05812	0.04006
1391.01	−0.02053	0.02736	0.02283	-3×10^{-05}	0.04203	0.04201
1456.01	0.00524	0.03596	0.03497	−0.00229	0.04655	0.04747
1457.01	−0.00701	0.04232	0.02725	−0.00038	0.04791	0.04625
1781.01	0.07127	0.01272	0.02821	0.00197	0.05888	0.05369
1793.01	0.00685	0.04855	0.04713	0.00578	0.04827	0.04871

2.2.2 Properties of Planet Candidates Analyzed

When selecting the initial planet candidates that we perform MCMC fits for planet properties described above, we vet for eclipsing binaries (EBs) using the procedure outlined in (Demory & Seager, 2011). This leaves us with a sample of 85 planet candidates for which we have calculated posteriors for their orbital and physical properties. From this new list of planet candidates with updated properties from MCMC fitting, we do a second updated sweep for eclipsing binaries referring to Tenenbaum et al. (2014) and Bryson et al. (2013), works that were published after our initial planet candidate list was developed. We also reference the *Kepler* Eclipsing Binary catalog[2] for additional newly reported EBs. From this procedure, we are able to exclude an additional 18 planet candidates. We include KOI 1227 in our sample of planet candidates as it appears in both the *Kepler* eclipsing binary catalog with a period of ~ 4 days, and in the *Kepler* planet catalog as a potential planet with an ~ 2 day period. After the vetting outlined above, we exclude an additional 17 planet candidates for which the occultation signal-to-noise was low, resulting in very poor measurements of h and k. This leaves us with 50 planet candidates that have approximately Gaussian measurement uncertainties for h and k to use for our analysis of the eccentricity distribution in §3.4.2. Working in h and k space instead of eccentricity space greatly simplifies our HB model for the eccentricity distribution (see §4.3.1), since the measurement uncertainties for h and k can be assumed to be roughly normally distributed.

The 50 remaining planet candidates have radii estimates of ~ 1.9 to 30 Earth radii, with a median value of 10.6 Earth radii, host star effective temperature of 3948 K to 8848 K, with a median value of 5728 K, orbital period of 1.03 days to 20.13 days with a median value of 4.24 days, and host star metallicity of -0.518 to 0.440 in [Fe/H] with a median value of 0.023 [Fe/H]. The 30 Earth Radii planet candidate (KOI 1793) is large, and an outlier for typical radii in our sample, but still makes it past our EB vetting procedure outlined above. These values came from the *Kepler* Star Properties Catalog as reported at Exoplanet Archive updated December 2013 and revised February 2014 (Buchhave et al., 2012; Huber et al., 2014). The majority of the planets have stellar metallicity values obtained from photometry, and 10 planets have spectroscopically derived stellar parameters.

[2]http://exoplanetarchive.ipac.caltech.edu/docs/eclbin.html

2.3 Method

We aim to simulate and characterize the eccentricity distribution of a subset of the population of planet candidates in the *Kepler* sample for which both transits and occultations have been observed. First, we describe a general HB model, before specializing it for our application of characterizing the eccentricity distribution in §3.2. Next, we build and test a model using simulated data in order to determine the accuracy of our method and robustness to model misspecification in §2.4.2, then we apply our model to the real dataset in §3.4.2.

2.3.1 The Hierarchical Bayesian Model

Hierarchical Bayesian (HB) modeling is a powerful method to estimate population parameters by propagating the unique uncertainty from each measurement of the population constituents into the inference of the population parameters. An HB model requires an analysis model that parameterizes the functional form of the population distribution, $p(x_p|\boldsymbol{\phi})$, where x_p represents the true value of each quantity being measured (later we adapt this model so that x_p represents h and k). $\boldsymbol{\phi}$ is the set of hyperparameters that determine the features or shape of the prescribed analysis model. To infer these population hyperparameters, we must specify the priors for the hyperparameters or the hyperpriors, $p(\boldsymbol{\phi})$. Once this multi-level model is applied to a sample of measurements, both the population's parameters and the true parameters for each of the population members can be inferred simultaneously. The measured properties (d_p) are related to the true properties (x_p) and the measurement uncertainties (σ_p) by $p(d_p|x_p, \sigma_p)$. As a result, the HB model allows us to characterize the true parameter values and population level parameters while using the information contained in the measurements and their uncertainties.

The general form for the posterior for the hyperparameter vector ($\boldsymbol{\phi}$), where D represents the number of measurements that make up the dataset (d_p), is given by:

$$p(\boldsymbol{\phi}|\mathbf{x_p}, \sigma_\mathbf{p}) \propto p(\boldsymbol{\phi}) \prod_{p=1}^{D} \int dx_p \, p(x_p|\boldsymbol{\phi}) p(d_p|x_p, \sigma_p) \tag{2.3}$$

Next, let us consider a simplified HB model where each measurement d_p is

drawn from a normal distribution centered on the true value x_p with measurement uncertainty σ_p

$$p(d_p|x_p, \sigma_p) \sim \text{Normal}_{\text{d}_\text{p}}(x_p, \sigma_p^2). \tag{2.4}$$

Here, the " $\sim$ " can be read as "is distributed as", common notation for statisticians. At the "mid-level" of the hierarchical model, we assume that the population of true values, x_p's, can be parameterized by a Gaussian mixture model, where each component of the population model has mean zero and N_m is the number of mixture components.

$$p(x_p|\boldsymbol{\phi}) = \sum_i^{N_m} f_i \, \text{Normal}_{\text{x}_\text{p}}(0, \sigma_i^2), \tag{2.5}$$

Each component contributes a fraction f_i of the population, so

$$\sum_i^{N_m} f_i = 1. \tag{2.6}$$

$\boldsymbol{\phi}$ then represents all of the f_i and σ_i values[3]. If we assume a common Gaussian mixture model prior for each x_p as shown in Equation (2.5) and Gaussian measurement error as shown in Equation (2.4), then our hierarchical model can be mathematically described as Equation (2.3) adapted to our specific analysis:

$$\begin{aligned} &p(f_i, \sigma_i|d_p, \sigma_p) \propto p(f_i, \sigma_i) \\ &\times \prod_{p=1}^{D} \int dx_p \sum_{i=1}^{N_m} f_i \, \text{Normal}_{\text{x}_\text{p}}(0, \sigma_i^2) \, \text{Normal}_{\text{d}_\text{p}}(x_p, \sigma_p^2). \end{aligned} \tag{2.7}$$

Moving the integral inside the summation and exploiting the symmetry of the Gaussian distribution, we get

$$\begin{aligned} &p(f_i, \sigma_i|d_p, \sigma_p) \propto p(f_i, \sigma_i) \\ &\times \prod_{p=1}^{D} \left[\sum_{i=1}^{N_m} f_i \int dx_p \, \text{Normal}_{\text{x}_\text{p}}(0, \sigma_i^2) \, \text{Normal}_{\text{x}_\text{p}}(d_p, \sigma_p^2) \right]. \end{aligned} \tag{2.8}$$

In Equation (2.9), we extend the limits of the integral to infinity in order to develop an analytic approximation to our hierarchical model that is accurate when $\sigma_i < 1$,

[3]Note that σ_i is a hyperparameter that partly describes the underlying population distribution along with f_i, where σ_p is the measurement uncertainties of the observable quantity.

$\forall; \in [1, N_m]$ (i.e. allowing the underlying model to assign eccentricities > 1).

$$p(f_i, \sigma_i | d_p, \sigma_p) \propto p(f_i, \sigma_i) \times \prod_{p=1}^{D} \left[\sum_{i=1}^{N_m} \frac{f_i \exp\left[-d_p^2/(\sigma_p^2 + \sigma_i^2)\right]}{\sqrt{2\pi(\sigma_p^2 + \sigma_i^2)}} \right]. \quad (2.9)$$

We discuss how we modify this derivation when evaluating our model numerically, applied to the eccentricity distribution in §2.4.1. The posterior distribution for the hyperparameter vector is conditional on all observations. The posterior modes and credible intervals can be calculated from Equation (2.8) using MCMC or estimated analytically based on Equation (2.9). Recent applications of HB modeling applied to other *Kepler* observations include Morton & Winn (2014), Rogers (2015b), Wolfgang & Lopez (2015), and Foreman-Mackey et al. (2014).

2.3.2 Applying the Hierarchical Model to Eccentricity Measurements

Next, we tailor the above model to the eccentricity distribution. The set of projected eccentricity measurements h and k for each planet candidate become the x_p's described in §2.3.1. We assume that each true value of h and k is drawn from a distribution that is a mixture of N_m normal distributions (the analysis model), where each mixture component contributes a fraction f_i, is centered on zero, and has a standard deviation σ_i. Thus, the hyperparameters $\phi = \left\{f_1, ..., f_{N_{m-1}}, \sigma_1, ..., \sigma_{N_m}\right\}$ describe the underlying population's distribution of h and k's. Since fractions sum to one, $f_{N_m} = 1 - \sum_{i=1}^{N_{m-1}} f_i$.

The values of h and k provide an alternate parameterization for the eccentricity (e), and the argument of periastron (ω). We assume that the orientations of planetary systems' pericenter directions (ω) will be randomly distributed with respect to the direction towards Earth, i.e., ω is uniform random $[0, 2\pi]$. Thus, the prior probability distribution for each planet's h and k has radial symmetry. While this is an excellent general model for planets, it is an approximation for our sample of planet candidates since 1) the geometric transit probability and occultation probability depend somewhat on ω for eccentric orbits and 2) the detection probability of both the transit and occultation depends on the transit

and occultation durations and thus the eccentricity and pericenter direction, and the occultation duration also depends on the orbital period. We will discus these issues further in §4.5. Results of this analysis can be found in §2.4 and §3.4.2.

2.3.3 Evaluating the Hierarchical Model

We sample from the posterior using MCMC. To calculate Markov chains we use the publicly-available code Just Another Gibbs Sampler (JAGS; Plummer, 2003). JAGS uses Gibbs sampling when possible, and otherwise reverts to standard random walk Metropolis–Hastings. We simultaneously sample from both the posterior distributions for the population parameters and the posterior predictive distributions for each observable. We compare the within-chain variance to the between-chain variance and evaluate the Gelman-Rubin ($\hat{R}$) ratio to test for non–convergence, and accept chains with an $\hat{R} < 1.01$. We also look at the autocorrelation function for the Markov chains and accept cases that have a zero crossing at a lag of ≤ 5. The exact JAGS input model used in our study can be found online[4].

2.4 Results

2.4.1 Prior specification

We consider three different analysis models and calculate posteriors for each using simulated data to test the accuracy and robustness of our method. The three analysis models used for x_p's in our calculation are (i) a single Gaussian ($N_m = 1$), (ii) a two-component Gaussian mixture ($N_m = 2$), and (iii) a three-component Gaussian mixture ($N_m = 3$).

We use these same three models both to analyze the data and to generate simulated observations. In each model, the population parameters, also known as hyperparameters, (ϕ) are a union of the set of dispersions for each mixture component (σ_i's) and the set of fractions of planets associated with each of the mixture components (f_i's). In each case, each mixture component is a Gaussian centered at zero and represents a unique population.

When evaluating the eccentricity parameter space, we take our priors for h and

[4]http://www.astro.ufl.edu/∼mshabram/jags_model/eccmodel.txt

k to be a mixture of Gaussian distributions, each with zero mean but truncated such that $e = \sqrt{h^2 + k^2} < 1$. Following truncation, the prior for h and k is renormalized, so that the total probability integrates to unity. The truncation accounts for the selection effect of not detecting planets on hyperbolic orbits ($e > 1$) as any such planets are not bound to their host systems and do not transit more than once. A Rayleigh distribution can also be parameterized as the square root of the sum of squared Normal distributions with zero mean, where the variance of each component is equivalent to the Rayleigh parameter. Thus, the prior population distribution for e is a truncated Rayleigh distribution for $N_m = 1$ and can be visualized as a mixture of truncated Rayleigh distributions of $N_m \geq 2$. Choosing a Rayleigh distribution for the eccentricity distribution is physically motivated by the fact that it naturally arises for exoplanets on circular orbits and subjected to a series of many normally distributed small random perturbations to its orbit. It is therefore a common distribution "shape" used for eccentricity distributions (e.g., Moorhead et al. (2011), Fabrycky et al. (2014)). We justify its superiority over a Beta distribution (Kipping, 2014a) in §2.4.2.2.

Calculating a posterior probability distribution function (PDF) from a HB model also requires specifying a prior probability distribution for the population parameters (ϕ). This is known as the *hyperprior*. Our hyperparameters are the dispersions σ_i's (for each mixture component) and the associated mixture fractions. We assume a uniform prior for the dispersions of each mixture component between 0 and 1. The mixture component fractions follow a Dirichlet distribution with the concentration parameter set to 1 (e.g., no component is given special weight). This is the multidimensional generalization of the Beta distribution. The Dirichlet distribution forces the sum of the mixture component fractions to equal one.

2.4.2 Validating the Hierarchical Model

Since the true distribution parameters for the synthetically generated datasets are known, analyzing these simulated observations with our hierarchical model allows us to directly compare the output population parameters and the input population parameters. We are also able to test the sensitivity of the posterior to the chosen analysis model.

We expect to see variations in the ability of a given analysis model to recover

the input model's parameters. For instance, if the analysis model is the same as the model used to generate the simulated observations, then we expect to be able to recover the input population parameters, within the limits of measurement uncertainties and Monte Carlo error. However, if the analysis model is different than the model used to generate the data, there could be larger differences between the posterior predictive distributions for the eccentricity distribution and the actual distributions used to generate the data. If we can identify an analysis model that is relatively insensitive to the model that was used to generate the simulated observations, then we can increase our confidence in the robustness of the procedure when applying our hierarchical model to a real dataset.

2.4.2.1 Generation and Analysis of Simulated Data

We are also interested in understanding the effect that the quantity and quality of the data has on our inference. It is important that we choose an analysis model that is relatively robust to model misspecification, so we can be confident when applying the HB model to real data. To accomplish this, we generate several simulated datasets varying the number of planets in the sample, the simulated measurement uncertainties, or both. For each pair of generative model and analysis model, we analyze four datasets of different qualities. We summarize each in Table 2.2. Datasets labeled "good" ("half") consist of 50 (25) planets with measurement uncertainties of 0.04 and 0.08 for h and k respectively.[5] These datasets are designed to be similar to our actual transit and occultation dataset for both h and k. Datasets labeled "better" ("best") contain 50 (500) planets with measurement uncertainties of 0.001. The mixture fractions and dispersions used to generate the synthetic datasets are the following: for a single Gaussian distribution, labeled as "R1" in Table **??**, $f = 1.0$ and $\sigma = 0.3$, for a two-component Gaussian mixture model ("R2"), $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, and for a three-component Gaussian mixture model ("R3"), $f_1 = 0.6$, $f_2 = 0.3$, $f_3 = 0.1$, $\sigma_1 = 0.05$, $\sigma_2 = 0.2$, and $\sigma_3 = 0.5$. We generate 20 datasets for each pair of data quality and generative model in order to quantify Monte Carlo error. The goal of this particular experiment is to identify an analysis model for the x_p's that performs well for a variety of plausible distributions

[5]The uncertainty in the phase offset of the transit is typically smaller than that of the occultation and transit duration ratio, thus the eclipse data constrain h with more precision than k.

used to generate simulated data.

Table 2.2: Values indicating the quantity and quality of the suite of simulated observations used in our analysis. Datasets labeled “good" (“half") consist of 50 (25) planets with measurement uncertainties of 0.04 and 0.08 for h and k respectively. These datasets are designed to be similar to our actual transit and occultation dataset for both h and k. Datasets labeled “better" (“best") contain 50 (500) planets with measurement uncertainties of 0.001, and are designed to forecast the power of this method and model setup when the quantity of real data grows and the quality of data is improved upon (better measurement uncertainty).

Model Name	Np	σ_h	σ_k
half	25	0.040	0.080
good	50	0.040	0.080
better	50	0.001	0.001
best	500	0.001	0.001

Table 2.3: Results of validation and sensitivity analysis of three hierarchical models for eccentricities. Shown in this table are Kolmogorov-Smirnov (K-S) statistics comparing datasets of simulated observations with datasets generated using the posteriors of the hyperparameters from an HB model that analyzed the same set of simulated observations (comparing input to output to test model). Here R1, R2, and R3 represent a one-, two- and three-component Gaussian mixture model, respectively. Table 2.2 summarizes the different quantity and quality of simulated observations used in this analysis. A two-component Gaussian mixture model does well across the majority of simulated datasets. See §4.2.2 for a detailed interpretation of these results.

	Analysis Model		
Generalized Model Name	$N_m = 1$	$N_m = 1$	$N_m = 1$
R1 half	0.1310	0.1615	0.1740
R1 good	0.1550	0.1650	0.1700
R1 better	0.0990	0.1150	0.1275
R1 best	0.0400	0.0500	0.0550
R2 half	0.2365	0.2100	0.2715
R2 good	0.2545	0.1875	0.2330
R2 better	0.1890	0.0825	0.1590
R2 best	0.1705	0.0660	0.1485

R3 half	0.1535	0.1215	0.1415
R3 good	0.2560	0.2085	0.2285
R3 better	0.1270	0.1035	0.1350
R3 best	0.1726	0.0835	0.2142

2.4.2.2 Results for Synthetic Data

First, we validate our HB model using the same model for the analysis as used to generate a simulated dataset. Next, we consider the results of applying an analysis model that differs from the model used to generate the data. The purpose of making these comparisons is to identify an appropriate analysis model, balancing the need for flexibility with the desire to minimize model parameters. By analyzing a variety of simulated datasets, we develop intuition for how different models perform, prior to analyzing the actual data. Since the true distribution of exoplanet eccentricities likely differs from any of our analysis models, it is important to analyze data sets generated under alternative models, so as to test the robustness of our approach.

When we use the same analysis model and generative model, it would be possible to compare the true model parameters to the posterior distribution for the model parameters. However, most of our comparisons involve different analysis and generative models. In these cases, it is not possible to compare the true model parameters to the posterior distribution for the model parameters. Instead, we compare the predictive posterior distribution for the eccentricity distribution (i.e., the distribution of interest). We use the K-S distance to measure how the predictive posterior distribution for eccentricities under each analysis model compares to the true eccentricity distribution used to generate the simulated data. Table 3 shows the median Kolmogorov-Smirnov (K-S) distance between each simulated dataset's true h and k distribution and the posterior predictive distribution for h and k based on 20 simulations of a particular hierarchical model from Table 2 (see §2.4.2.1 for a list of the chosen "true" eccentricity distribution values used in our study).

We illustrate an example case in Figure 2.1, by showing cumulative distributions of $|h|$ and $|k|$. The solid black curve is the true distribution from which the simulated planet's h and k values are drawn. The dashed black curve is the cumulative distribution for one simulated dataset ("R2", "good"; $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, see §2.4.2.1) that includes simulated observational uncertainties. The

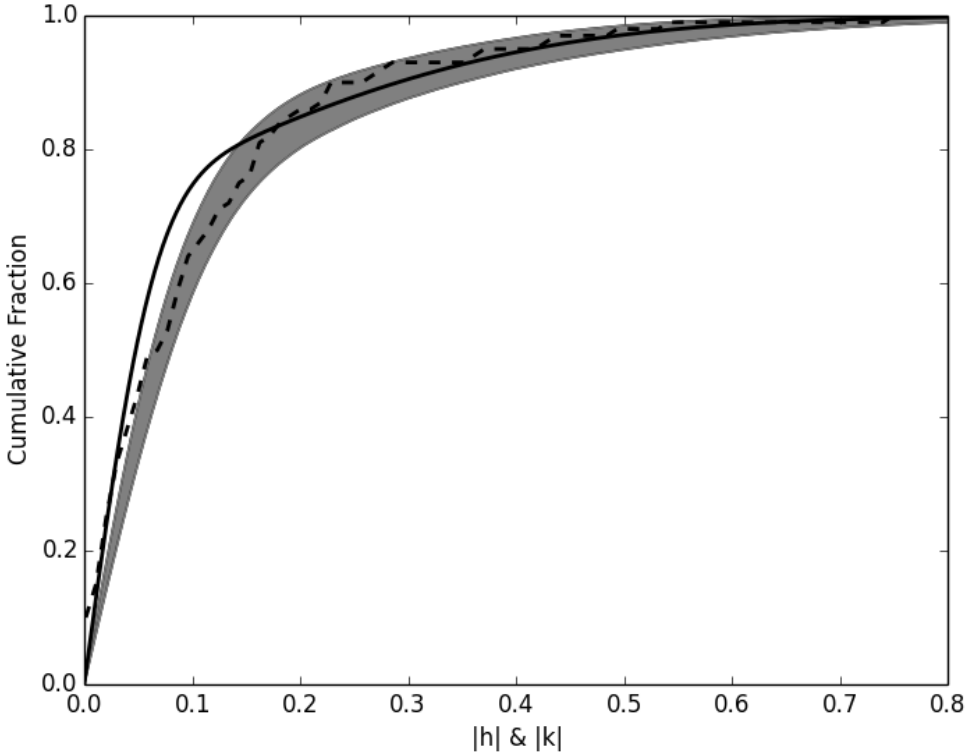

Figure 2.1: Cumulative distributions of $|h|$ and $|k|$. The solid black curve is the true distribution from which the simulated planet's h and k values are drawn. The dashed black curve is the cumulative distribution for one simulated dataset ("R2", "good"; $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, see §4.2.1) that includes simulated observational uncertainties. The gray shaded region is the 68.3% credible interval for the CDFs of the posterior samples for the population parameters of the intrinsic distribution of $|h|$ and $|k|$ (i.e., without intrinsic uncertainties). This is calculated once the simulated observations have been analyzed using the same two-component Gaussian mixture model as was used to generate the data. The two-component Gaussian HB model does a good job of capturing the true distribution for datasets generated with a two-component Gaussian mixture.

gray shaded region is the 68.3% credible interval for the CDFs of the posterior distribution for the population parameters of the intrinsic distribution of $|h|$ and $|k|$ (i.e., without intrinsic uncertainties). This is calculated once the simulated observations have been analyzed using the same two-component Gaussian mixture model as was used to generate the data.

The posterior predictive distributions are generated from the posteriors for the

hyperparameters obtained from applying the hierarchical model to each simulated dataset. Each column of Table 3 represents comparison results for posteriors calculated using an analysis model with one-, two-, or three-components in the Gaussian mixture model respectively, for each eccentricity distribution. The analysis model names are described in Table 2. Each row of Table 3 gives results for a specific generative model and data quality. R1, R2, and R3, indicate one, two and three component models for generating the simulated observations, respectively.

If the model is working properly, we expect to get posterior distributions for the population's parameters that are consistent with values used to generate the data. Indeed, we find K-S distances are between ~ 0.05 to 0.1 for these cases. Since models have different parameters (even when they are represented by the same variable names), the most appropriate way to compare the performance is based on the posterior predictive distribution for the population of measurements. In this case, the K-S distances between the posterior predictive distributions for the HB model and actual model are ~ 0.1 to 0.2.

For simulated datasets with smaller measurement uncertainties, we find that the K-S distance between the posterior predictive distribution and associated simulated data for h and k is similar for analysis models that have at least the same number for mixture components or more. Additionally, for several combinations of analysis and generated models, we note that the $N_m = 2$ analysis model results in a smaller K-S distance to the R3 data than the $N_m = 3$ analysis model. This is likely due to the greater flexibility of the $N_m = 3$ model and finite number of measurements, i.e., the three-component model "over-fits" the discrete dataset. We found that the $N_m = 2$ analysis model did a better job overall at recovering the predictive distribution for the simulated datasets across all versions for simulated data.

Some authors advocate parameterizing the eccentricity distribution as a Beta distribution, $e \sim \text{Beta}(\alpha, \beta)$ (e.g., Kipping, 2014a). Therefore, we also investigate using a Beta distribution analysis model using one ("R2", "good") simulated dataset (see Table 2). In this model set up, α and β become the hyperparameters (population level parameters) that we wish to infer. We use a Gamma distribution with $k = 2$ and $\theta = 1$ as the prior probability distribution for α and β. Figure 2.2 shows the results of this HB model as eccentricity vs. cumulative fraction. The simulated eccentricity data are shown as the dotted black curve. The true eccentricity distribution generated using a two-component Gaussian mixture model

for h and k (e.g. $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, as described in §2.4.2.1) is shown in red. The dashed green curve is plotted using the posterior modes for α and β, ($\alpha = 0.11\pm^{0.04}_{0.02}$, $\beta = 1.73\pm^{0.85}_{0.24}$) for this HB model. The K-S distance between the "R2," "good" distribution (red) used and the distribution using the posteriors modes of α and β (dashed-green) is 0.5, which is in support of these being two distinct distributions. Our results indicate that the standard Beta distribution is a poor choice for an analysis model to parameterize the eccentricity distribution as it erroneously predicts a strong peak near $e = 0$ and under predicts the frequency of larger eccentricities.

In principle, we could model the transit and occultation times and durations directly, instead of h and k. However, this would increase the computational complexity. Even taking advantage of our approximations, the calculations presented represent a significant computational investment ($\sim$2.4 CPU months). Modeling h and k also facilitates deriving analytical expressions for testing the algorithms. By modeling the projected eccentricity, h and k, we were able to thoroughly test both the code and algorithm on real and simulated data sets.

2.5 Results for *Kepler* planet candidates with occultations

We calculated posteriors of one-, two-, and three-component Gaussian mixture models applied to real *Kepler* transit and occultation data (see §2.2 for description of dataset). Figure 2.3 shows a histogram of the observed h and k values from Table 1 (shown in grey). Since we are assuming the argument of periastron (ω) is random, h and k are equivalent, or drawn from the same distribution. A Gaussian distribution using population parameters from the posterior mode for the dispersion for a one-component model is overplotted (shown as the dotted black curve). This one-component model does a poor job at capturing the shape of the distribution because it struggles to match the moderate eccentricity outliers. A two-component Gaussian mixture model using population parameters from the mode of the 2D marginal posterior for the mixture fraction and dispersions is shown in red. This model captures the peaked nature of the observed distribution as well as the small number of measurements away from the peak. This suggests that two populations

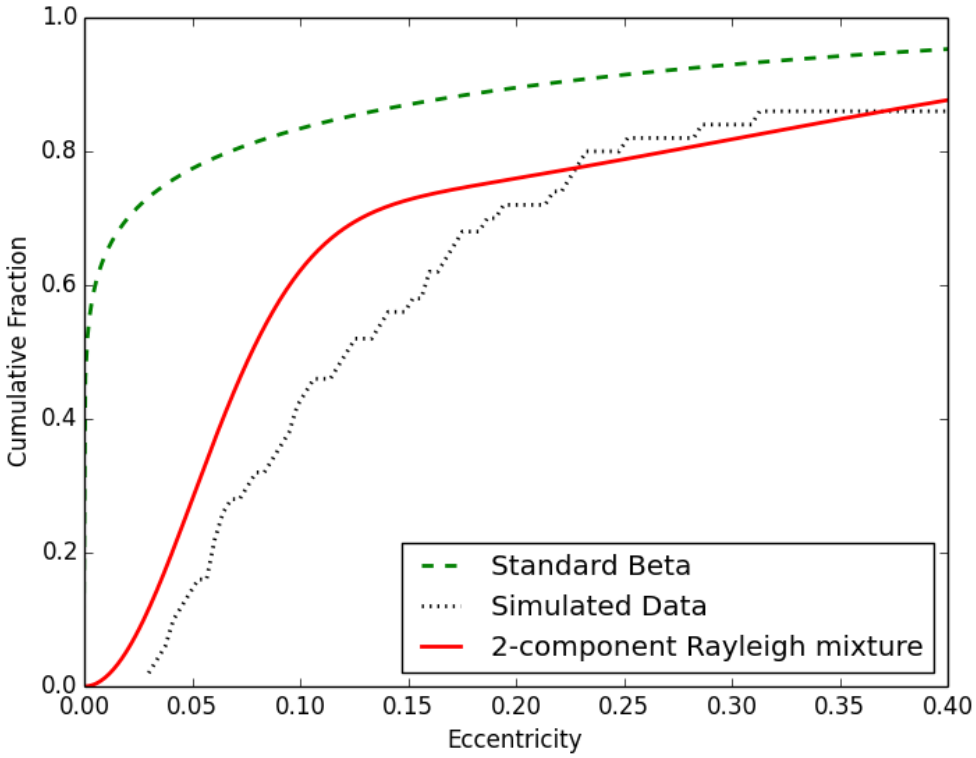

Figure 2.2: Results of an HB model that parameterizes the eccentricity as a standard Beta distribution. We investigate using a Beta distribution analysis model on an "R2", "good" (see Table 2) simulated dataset. A cumulative distribution of the simulated eccentricity data are shown in grey. The distribution generated using a two-component Gaussian mixture model for h and k values (e.g. $f_1 = 0.7$, $f_2 = 0.3$, $\sigma_1 = 0.05$, and $\sigma_2 = 0.3$, as described in §4.2.1) is shown in red. The dashed green curve is a cumulative Beta distribution, Beta(α, β), plotted using the posterior modes for α and β, ($\alpha = 0.11\pm^{0.04}_{0.02}$, $\beta = 1.73\pm^{0.85}_{0.24}$) for this HB model. The Beta distribution erroneously predicts a strong peak near $e = 0$ and under predicts the frequency of larger eccentricities.

can explain the eccentricity distribution of our sample, although in §2.4.4.2 we show with synthetic data that using a two-component Gaussian mixture model is optimal for the present dataset. We also consider using a three component Gaussian mixture model and find that only two of the three components can be constrained given the quantity and quality of the *Kepler* transit and occultation dataset, suggesting that the available data are not able to indicate the presence of a third population, or that a third population may not exist.

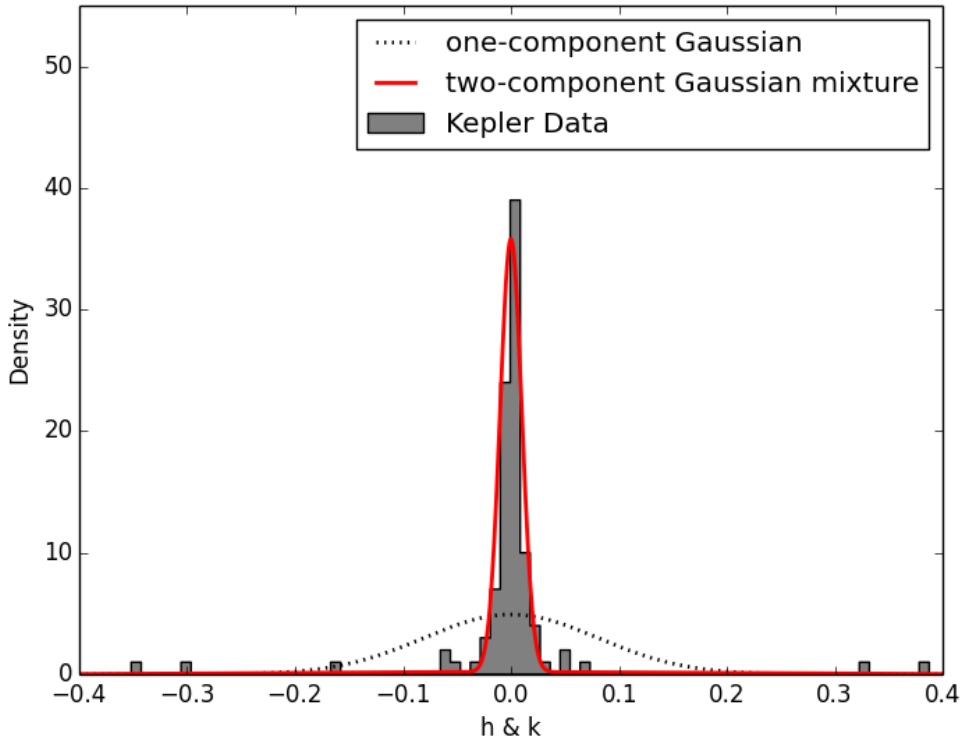

Figure 2.3: A histogram of the h and k dataset are shown in grey. Shown in black is a one-component Gaussian distribution using the posterior mode for the dispersion obtained from an HB model that uses a one-component Gaussian mixture model. Shown in red is a two-component Gaussian mixture model using posterior modes for the mixture fractions and dispersions obtained from an HB model that uses a two-component Gaussian mixture model. The black model does a poor job at capturing the shape of the distribution. The red model captures the peaked nature of the true distribution while also allowing for a smaller number of measurements far from the central peak.

The posterior distribution for the dispersion of true values of h and k assuming a one-component Gaussian model is displayed in Figure 2.4, which is based on our full dataset for planet candidates with both *Kepler* transit and occultation measurements. The posterior mode for the dispersion is $\sigma = 0.081 \pm^{0.014}_{0.003}$. We use this value as the dispersion for our one-component Gaussian population model shown as the dotted black curve in Figure 2.3.

Next, we investigate joint posterior distributions for a two-component Gaussian mixture model applied to our full *Kepler* transit and occultation dataset. These

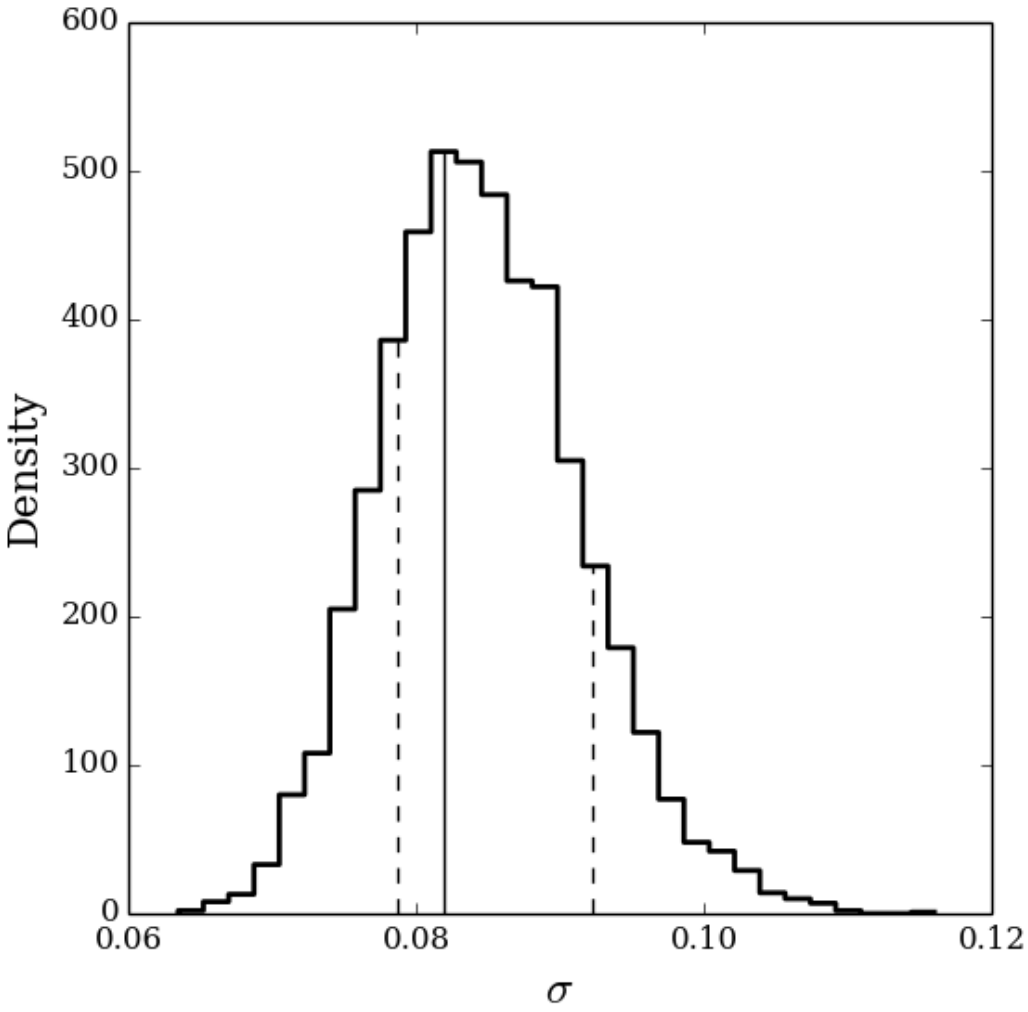

Figure 2.4: Posterior distribution for the dispersion of a Gaussian model for h and k applied to our full *Kepler* short-period planet candidate transit and occultation dataset. The 68.3% credible intervals about the mode are shown as dotted black lines, and the mode is shown as a vertical solid black line.

results are shown in Figure 2.5, where the panels on the diagonal show the marginalized posterior distribution for the population parameters: σ_{low} the lesser value of σ_1 and σ_2, σ_{high} the greater value of σ_1 and σ_2, and f_{low}, the weights for the mixture component ($f_{high} = 1 - f_{low}$). The use of σ_{low} and σ_{high} instead of σ_1 and σ_2 (and corresponding fractions) is helpful for visualizing the results, since our model has symmetry under exchanging (σ_1, f_1) and (σ_2, f_2). The off-diagonal

panels show posterior samples and contours for the 68.3% credible interval of the two-dimensional marginal posteriors for each parameter pair. The fact that (σ_{low}, f_{low}) and (σ_{high}, f_{high}) form two distinct clusters demonstrates the value of a two-component (two population) model for the eccentricity distribution of our sample of *Kepler* planet candidates.

As expected, the uncertainties in measurements of $k = e \sin \omega$ are much greater than the uncertainties in the measurements of $h = e \cos \omega$. We note that in our sample, the k values are more tightly clustered around zero than h. Therefore, we investigated if excluding these values significantly impacts our results. By doing this we are decreasing our effective sample size, but maintaining the number of measurements with small uncertainties. When running our simulations without k values, we get the following results: using a single Gaussian model $\sigma = 0.074 \pm^{0.016}_{0.003}$, and when using a two-component Gaussian mixture model, the marginal posterior modes for the mixture fractions and dispersions are $f_{low} = 0.93 \pm^{0.029}_{0.051}$, $\sigma_{low} = 0.003 \pm^{0.010}_{0.001}$, and $f_{high} = 0.07 \pm^{0.062}_{0.019}$, $\sigma_{high} = 0.187 \pm^{0.547}_{0.028}$, respectively. These values differ from the values obtained using the full h and k planet candidate dataset by $\sim$ 9.0%, 4.4%, 107.7%, 44.4%, and 16.2% for σ, f_{low}, σ_{low}, f_{high}, and σ_{high} respectively. Each of these overlaps the 68.3% credible interval for the same parameters with the full dataset. The most notable difference is for σ_{low}, which we estimate to be 0.03 when using k alone, but 0.01 when including both h and k observations.

2.5.1 Correlation of The Eccentricity Distribution with Star and Planet Properties

Another goal of this study is to investigate potential correlations between the eccentricity distribution and planet or host star properties. Specifically, we consider whether the planet candidate eccentricity distribution is correlated with stellar metallicity, host star effective temperature, planet radius, or orbital period. Values for each planet candidate are obtained from the *Kepler* Star Properties Catalog as reported at Exoplanet Archive updated December 2013 and revised February 2014 (Buchhave et al., 2012; Huber et al., 2014). The majority of the planets have effective temperatures and stellar metallicity values obtained via KIC photometry, however 10 planets have spectroscopically derived values. Given the relatively small sample size, we focus on comparing the distribution of planet candidates with large

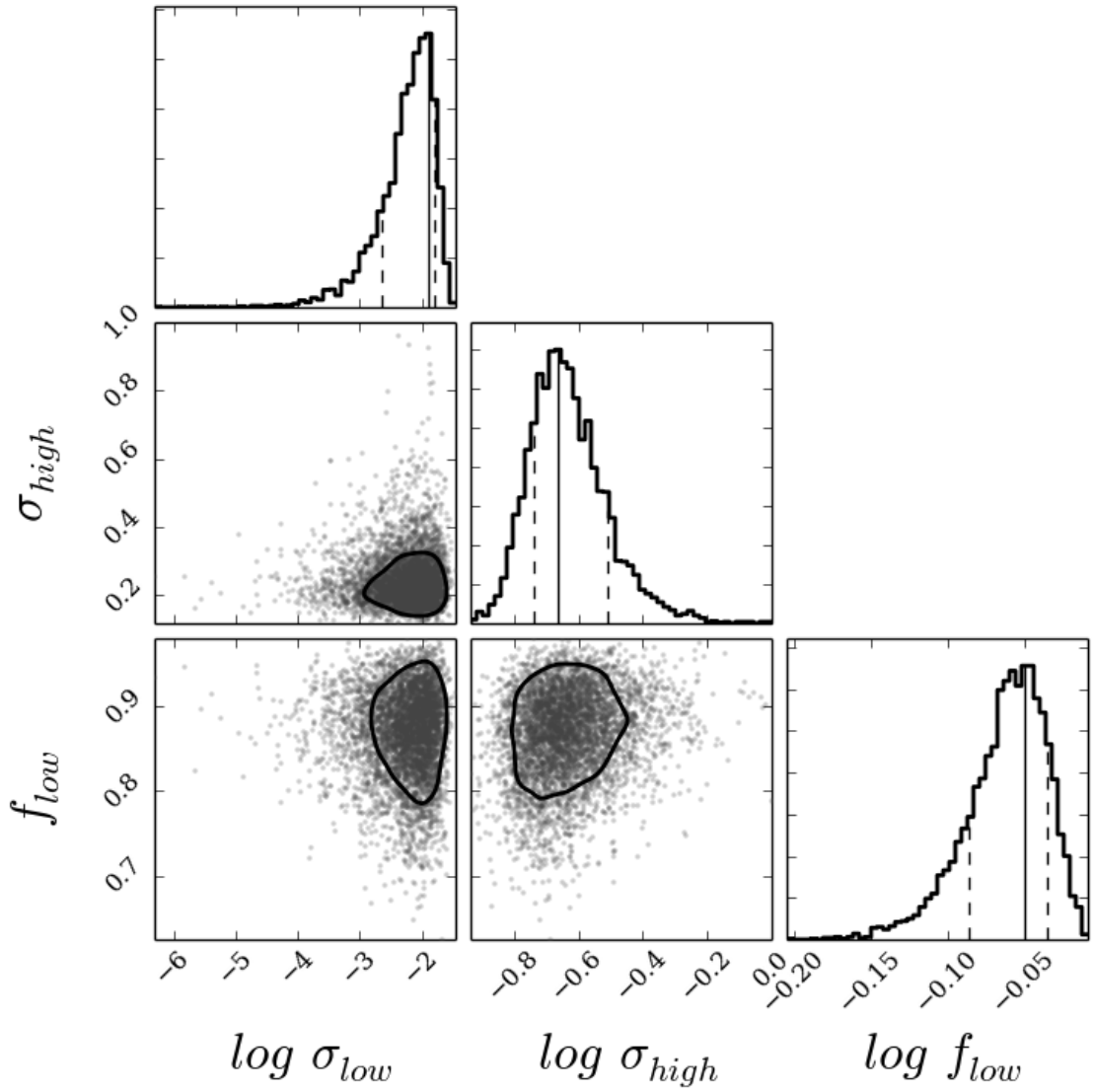

Figure 2.5: Joint posterior distributions for a two-component Gaussian mixture model applied to our *Kepler* short-period planet candidate transit and occultation dataset. In each panel, the data are plotted with the horizontal axis representing σ_{low} the lesser value of σ_1 and σ_2, σ_{high} the greater value of σ_1 and σ_2, and f_{low} the corresponding weight for the low mixture component, on a logarithmic scale. Since $f_{high} = 1 - f_{low}$, we only show f_{low} here. The vertical axis shows these same variables, and each panel is the corresponding two-dimensional marginal posteriors for each parameter pair. The contour region plotted over the sampled posterior represents the 68.3% credible interval. The one-dimensional histograms are plotted as log density, with the 68.3% credible intervals shown as dotted black lines, and the mode is shown as a vertical solid black line. The two-component Gaussian mixture model characterizes the eccentricity distribution of our sample of planet candidates better than a one-component analysis model.

and small values for each parameter. We sort the planet candidates in our dataset from largest to smallest values of a given property, and then create two sub-samples of the original population. Unless otherwise specified, we divide the data in half to maximize the statistical power when comparing the two samples and to avoid introducing an additional parameter specifying the dividing point between the high and low subsets. We analyze each subset as described in section §2.3.

Initially, we evaluate each subset of data using an HB model with a one-component Gaussian distribution for h and k, as we did before for the full dataset. This is shown in Figure 2.6, where we have applied the HB model to the small-valued (blue) and large-valued (red) halves of the *Kepler* occultation data, sorted by (a) stellar effective temperature, (b) planet radius, (c) orbital period, and (d) stellar metallicity. The histograms of the posterior distribution for the dispersions for stellar effective temperature (a) and orbital period (c) suggest that the two subsets do not come from significantly different distributions if we assume the eccentricity distribution is described by a simple Rayleigh distribution. However, for planet radius (b), and stellar metallicity (d), the differences in the posteriors for the dispersion suggests that the two subsets may have different eccentricity distributions. This provides motivation to consider more complex models for correlations between the eccentricity distribution and stellar and planet properties.

Next, we look at posterior distributions based on applying the HB model using a two-component Gaussian mixture for the analysis model applied to the small-valued half or large-valued half of the data subsets, again based on sorting by (a) stellar effective temperature, (b) planet radius, (c) orbital period, and (d) stellar metallicity. Figure 2.7 shows posterior distributions for the small-valued half of data (blue and green clusters), and large-valued half (red and orange clusters). The two groups of clusters represent samples of the posterior distribution for the hyperparameter vector, in this case for σ_{low} and f_{low} (top left group of clusters in each sub-plot), and σ_{high} and f_{high} (bottom right group of clusters in each sub-plot). The data are plotted with the vertical axis representing the low value of the mixture fraction, f_{low}, in green and orange, and, f_{high}, in blue and red for the two subsets of sorted data shown (small and large). The contours correspond to 68.3% credible intervals. In this plot we can compare the two high and low value subsets to the full sample to check for correlations with each parameter. Interestingly, the

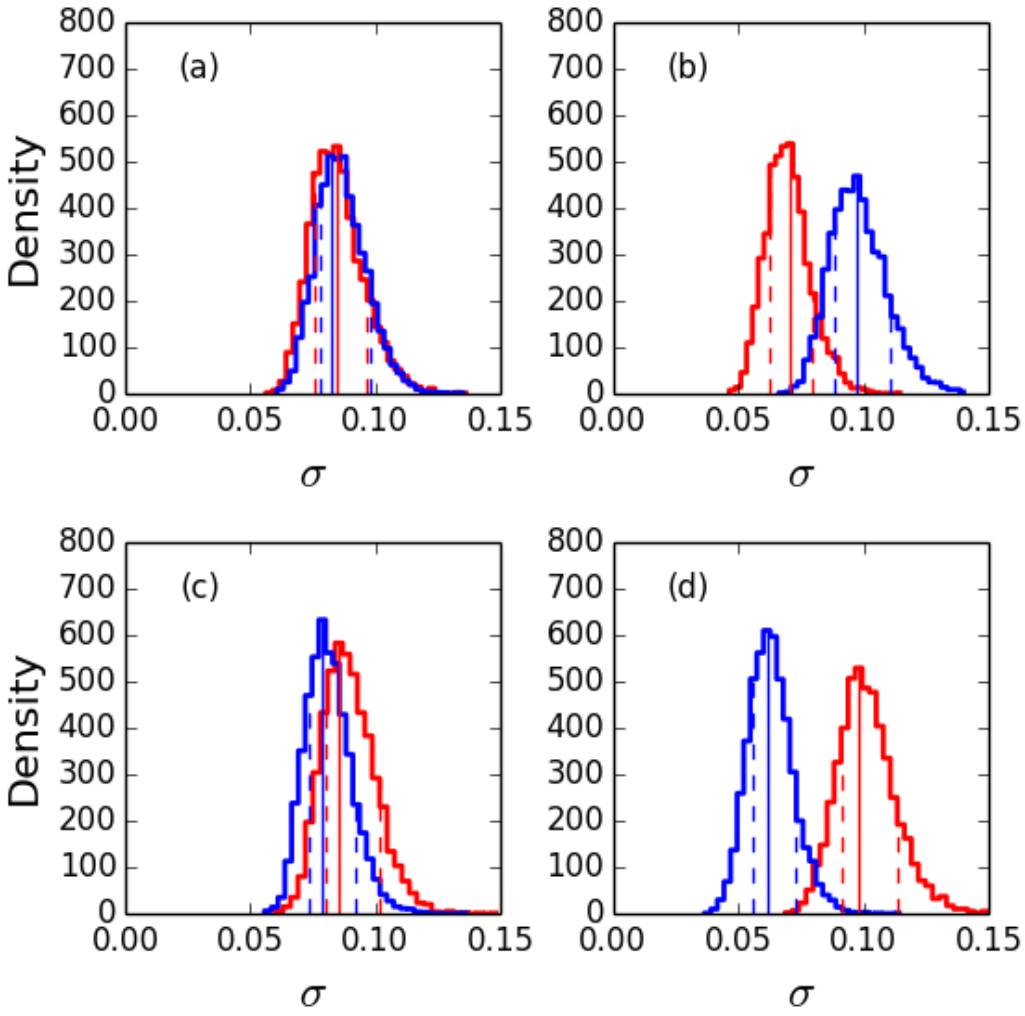

Figure 2.6: One-component Gaussian analysis model applied to subsets of the *Kepler* transit and occultation data. We apply an HB model to small-valued (blue) and large-valued (red) halves of the *Kepler* short-period planet candidate transit and occultation data, sorted by (a) stellar effective temperature, (b) planet radius, (c) orbital period, and (d) stellar metallicity. The dotted lines correspond to the 68.3% credible intervals and the solid vertical lines correspond to the mode for each posterior distribution. for panel (b) planet radius, and panel (d) stellar metallicity, differences in the small-values and large-valued data subsets merit further investigation. In order to explore these results further, we analyze these subset using a two-component Gaussian mixture model (see Figure 2.7).

posteriors of the mixture fractions for the planet candidates with larger planet

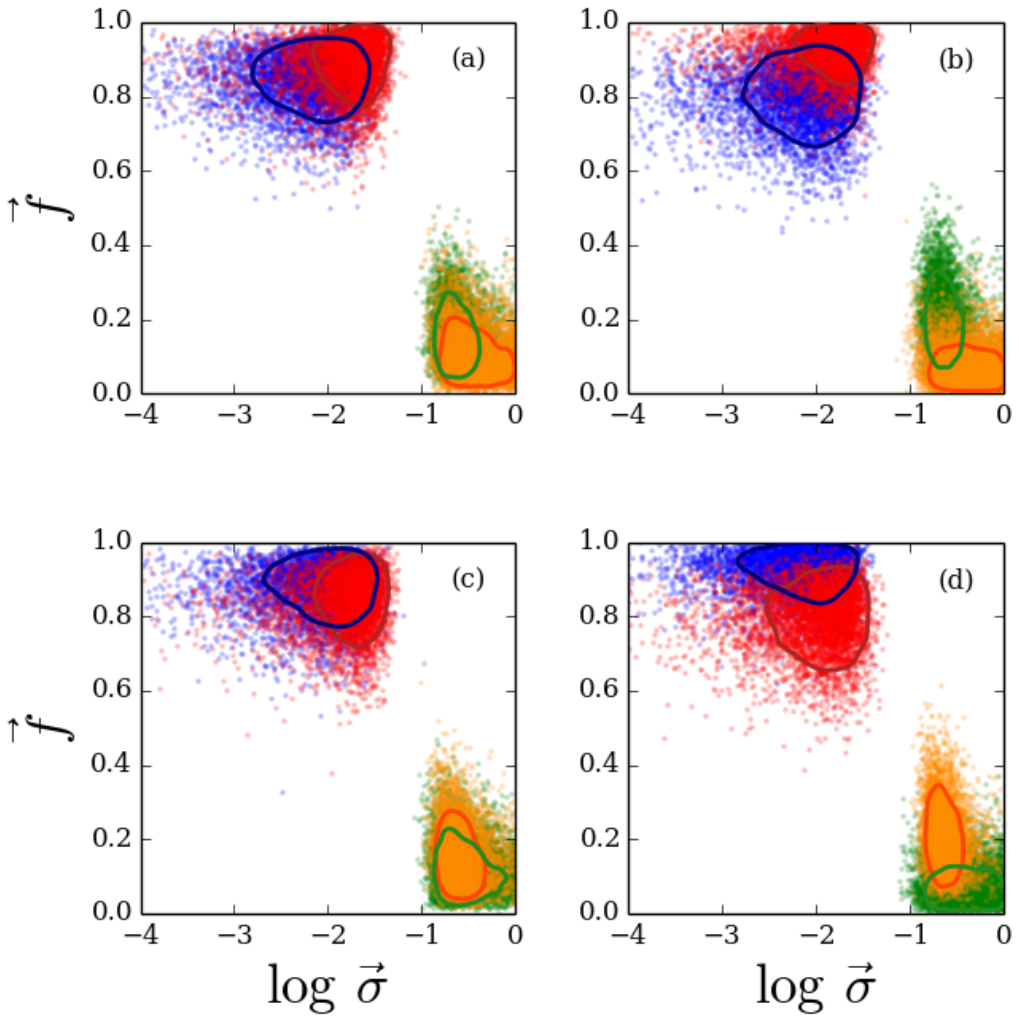

Figure 2.7: Continued on following page.

radii are consistent with 0 and 1 for planet radius (b) and host star metallicity (d), indicating only one population is required to accurately model the eccentricity distribution for this subset of planet candidates. When modeling the eccentricity distribution of a sample of planet candidates with host star metallicities less than 0.023 dex (median [Fe/H] of our full sample), we find a one-component mixture model is sufficient. On the other hand, planet candidates of host star metallicities above 0.023 dex are better modeled with a two-component mixture model for h and k. We find a similar, but weaker correlation between eccentricity distribution

Figure 2.7: Two-component Gaussian mixture model applied to the *Kepler* transit and occultation data. We apply an HB model to small- and large-valued halves of the short period *Kepler* candidate occultation data, sorted by stellar effective temperature (a), planet radius (b), orbital period (c), and stellar metallicity (d). The full sample is divided into two equally sized small- and large-value subsets before being processed through our HB model. The small-valued subset of data corresponds to the blue and green clusters, and the large-valued subset corresponds to the red and orange clusters. The two groups of clusters represent samples of the posterior distribution for the hyperparameter vector, in this case for σ_{low} and f_{low} (top left group of clusters), and σ_{high} and f_{high} (bottom right group of clusters). The data are plotted with the vertical axis representing the low value of the mixture fraction, f_{low}, in green and orange, and, f_{high}, in blue and red for the two subsets of sorted data shown. The contours represent 68.3% credible intervals. Interestingly, for planet radius (b), and stellar metallicity (d), the posteriors of the mixture fractions for the planet candidates with large-valued planet radii and for small-valued host star metallicities are consistent with 0 and 1, indicating only one population is required to accurately model the eccentricity distribution for these subsets of planet candidates. For planet radius and stellar metallicity, we also see that the two-component population models are somewhat different for small and large value subsets.

and planet radius. Planet candidates with radii smaller than 10.6 $R_\oplus$ are better modeled with a two-component mixture model, while a one component mixture model is favored for planet candidates with radii above 10.6 $R_\oplus$. There is not a strong correlation between planet radius and metallicity in our sample. Further, we verified that the subsamples based on metallicity and planet radius are distinct from each other.

Next, we consider whether the data can constrain more complex models that allow for a flexible choice of the break point between the two subsets of planet candidates, rather than fixing the break point to divide the dataset in half. We choose to investigate a more flexible model for orbital period first, because the measurement uncertainties for orbital period are negligible. We allow the period break point to be a free parameter in a single HB model where each subset of planet candidates are modeled with a two-component mixture model as before. Instead of fixing the break point near the median and dividing the data into equal sized subsets, we place a uniform prior on the period break point. Figure 2.8 shows the marginal posterior distribution for the period break point from analyzing the

actual dataset. The marginal posterior for the period break has peak values that are clustered near the minimum and maximum of the period values in our sample. This indicates that the model favors period breaks causing one data subset to have so few observations that the population parameters from one subset are minimally constrained. We conclude that the present dataset is not able to usefully constrain this more complex model. Therefore, we do not attempt to apply a similar model allowing for two eccentricity distributions with an unknown break in terms of the host star metallicity or planet radius, since their measurement uncertainties are much larger.

2.6 Summary of Results

We investigated the eccentricity distribution for a sample of short-period single-planet candidate systems from *Kepler* that are detected in both transit and occultation. We demonstrated that HB models are well-suited for characterizing the eccentricity distribution using transit and occultation data. We modeled the distribution of h and k as coming from either a single Gaussian distribution with zero mean or a mixture of Normal distributions. After testing our hierarchical model on a suite of simulated datasets and analysis models with one, two, or three mixture components, we find that a two-component mixture model ($N_m = 2$) performed well in all cases considered, including simulated datasets generated using a three-component mixture model. Thus, the two component mixture model is a robust analysis model for our hierarchical model applied to *Kepler* transit and occultation data. Additionally, we investigate the usage of a standard Beta distribution analysis model in our HB model. Our results indicate that the standard Beta distribution is a poor choice for an analysis model to parameterize the eccentricity distribution.

Next, we applied HB modeling to analyze a real dataset of $h = e\cos\omega$ and $k = e\sin\omega$ measurements, derived from transit and occultation measurements. If we model the population distribution of h and k with a single Gaussian, then we infer a dispersion of $\sigma = 0.081\pm^{0.014}_{0.003}$. When we applied the two-component mixture model to the full dataset, we found $f_{low} = 0.89\pm^{0.045}_{0.057}$, $\sigma_{low} = 0.01\pm^{0.014}_{0.002}$, and $f_{high} = 0.11\pm^{0.057}_{0.045}$, $\sigma_{high} = 0.22\pm^{0.100}_{0.026}$. These results suggest the presence of a small population of planet candidates ($\sim 11\%$) that contain planets with a broad

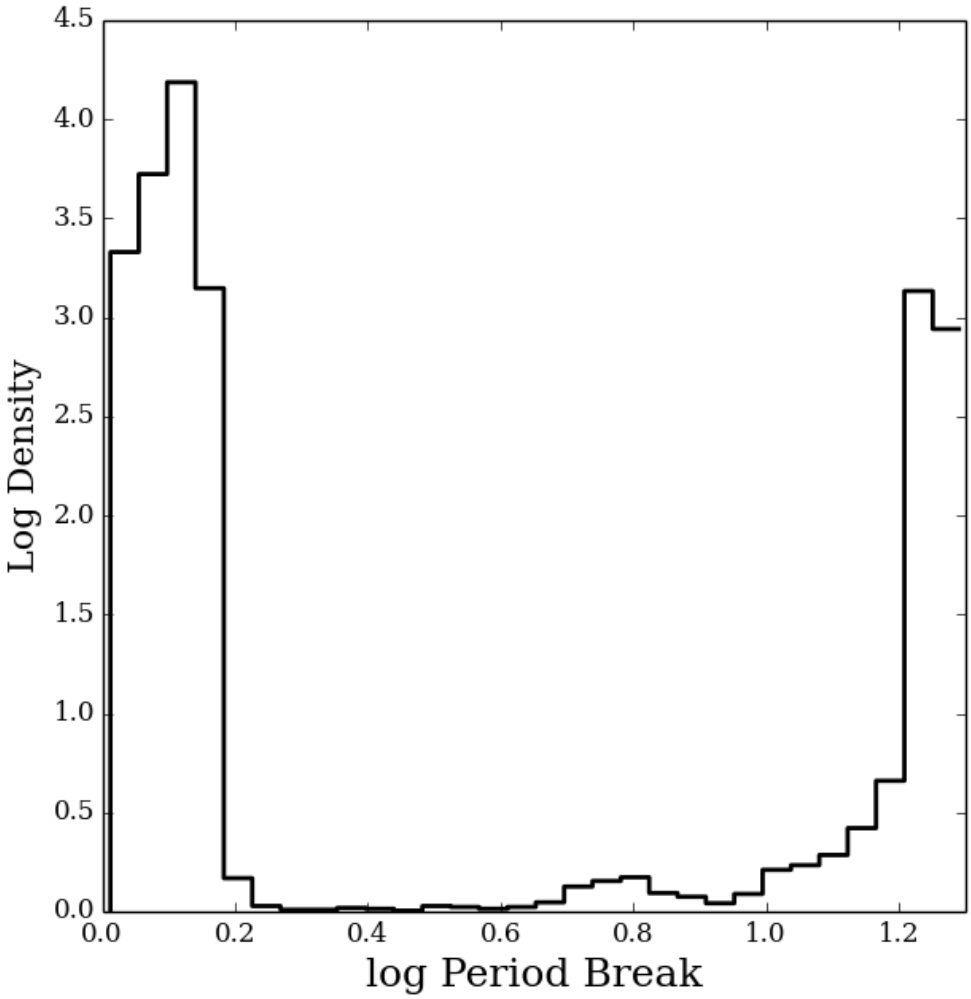

Figure 2.8: Marginal posterior distribution for the critical period break point from a joint period-eccentricity distribution HB model. We analyze the full dataset using an HB model that allows for the eccentricity distribution to differ depending on whether the orbital period is greater or less than the critical period break point. We infer that the present data does not allow us to empirically identify a period cutoff that depicts two populations.

range of orbital eccentricities and a larger population of planet candidates ($\sim 89\%$) that contain planets on nearly circular orbits.

Next, we assessed whether there is evidence for more complexity in the eccentricity distribution by considering analysis models that allow correlations between

the eccentricity distribution and other planet or host star parameters. For the current sample of *Kepler* planet candidates seen in both transit and occultation, we find interesting correlations of the eccentricity distribution with either the planet radius or the host star metallicity, but not with stellar effective temperature or orbital period.

We present evidence that host stars in our sample with higher metallicity and planet candidates with smaller radii have a more complex eccentricity distribution than stars with low metallicity and planet candidates with larger radii. The eccentricity distribution of these more complex populations are well described by a two-component Gaussian mixture model with a zero mean, suggesting a potential physical explanation in terms of proposed planet formation models, which we will describe in more detail in §4.5.

2.7 Discussion

Previous studies of the period-size distribution of *Kepler* planet candidates have identified two common architectures of planetary systems: 1) Systems with Tightly-packed Inner Planets (STIPS; Lissauer et al. (2011); Payne et al. (2013); Boley et al. (2014)) and 2) systems with a single short-period planet (often a giant planet) with either no additional planets detected or a large gap between the short-period planet and the next detectable planet (Steffen et al., 2012a, 2013; Dawson & Murray-Clay, 2013). This work focuses on a sample containing primarily isolated giant planet candidates. Of course, these systems may have undetected companions, particularly at larger orbital separations where the geometric transit probability is small.

Two broad classes of mechanisms have been proposed to explain the formation of hot Jupiters. In both models, planets form at greater distances from their host star. In one model, a giant planet experiences a gradual inspiral through a gas or planetesimal disk until they halt near their present orbit (e.g., Kley & Nelson 2012). In the other model, gravitational perturbations from another massive body (potentially another planet or a stellar companion) excite a giant planet's orbital eccentricity until its periastron distance is so small that tidal forces begin to circularize the orbit (e.g., Rasio & Ford 1996; Fabrycky & Tremaine 2007; Naoz et al. 2011). Several studies have assessed the relative merit of these two classes of models, making use of the observed orbital period distribution (Ford &

Rasio, 2006; Valsecchi & Rasio, 2014), spin-orbit obliquity distribution (Morton & Johnson, 2011; Albrecht et al., 2012; Naoz et al., 2012; Dawson, 2014) and orbital architectures (Steffen et al., 2012b; Dawson et al., 2012; Dawson & Murray-Clay, 2013). Collectively, these studies suggest that multiple mechanisms likely contribute to the formation of hot-Jupiters. In this case, the two populations would likely have different distributions of orbital eccentricities, with disk migration leading to the smaller dispersion of eccentricities. This motivates us to consider interpreting evidence of a two-component mixture model for the eccentricity distribution in terms of two formation models.

Our analysis of *Kepler's* short-period planet candidates with occultation measurements suggests that the eccentricity distribution can be well described by a two-component mixture model, where the less abundant population of planets has a broader dispersion of eccentricities. If the mixture components indeed translate into formation mechanisms of hot-Jupiters, then this could suggest that disk migration could be the more common formation mechanism. Alternatively, if tidal circularization of highly eccentric proto-hot Jupiters is sufficiently rapid, then the current eccentricities could reflect late-stage excitation of orbital eccentricities due to undetected planets. Of course, a complete formation theory would need to explain all observations, including the low abundance of additional planets near hot-Jupiters (Steffen et al., 2012a), the final semi-major axis of hot-Jupiters (e.g., Ford & Rasio 2008; Valsecchi & Rasio 2014), the distribution of orbital obliquities (e.g., Fabrycky & Winn 2009; Morton & Johnson 2011; Albrecht et al. 2012) and correlations between obliquity and other star and planet properties (e.g., Winn 2010; Morton & Winn 2014).

It is particularly interesting to compare the eccentricity distribution of our sample to that of other subsamples of the *Kepler* planet candidate list, particularly subsamples dominated by smaller planets. Since the eccentricity affects the transit duration (Barnes, 2007; Burke, 2008), the distribution of transit durations can constrain the eccentricity distribution for arbitrary sub-samples of *Kepler* planet candidates (Ford et al., 2008). Early studies of the eccentricity distribution of *Kepler's* planet candidates (Moorhead et al., 2011; Kane et al., 2012; Plavchan et al., 2014b) were limited due to the uncertainty in stellar parameters. Transit durations combined with stellar properties from photometry (Moorhead et al., 2011), high-resolution spectroscopy (Buchhave et al., 2012; Dawson & Johnson,

2012) and/or flicker (Kipping et al., 2014) can effectively recognize high eccentricity planets. However, further research is needed to obtain stellar properties precise and accurate enough to enable population studies of the more typical low eccentricity planets. Fortunately, one can characterize the eccentricity distribution of substantial subsets of *Kepler* planet candidates, either by using stars with high quality stellar characterization (e.g., asteroseismology, Huber et al. 2013; Ford et al. in prep.) or by using ratios that eliminate the dependence on stellar properties (e.g., Kipping 2011; Fabrycky et al. 2014; Morehead et al. in prep.).

Previous studies suggest that the typical eccentricity of planet candidates in systems with multiple transiting planet candidate systems ($\simeq 0.00 - 0.06$) is likely smaller than in our sample (e.g., Fang & Margot 2012; Fabrycky et al. 2014). Similarly, Wu & Lithwick (2013) and Hadden & Lithwick (2014) analyze transit timing variations (TTVs) in systems with near-resonant planet candidates. Hadden & Lithwick (2014) report a maximum likelihood estimate of the dispersion of eccentricities of $\sigma_e = 0.018^{+0.005}_{-0.004}$. This is significantly smaller than the $\sigma = 0.08^{+0.014}_{-0.003}$ of our one-component model for the eccentricity distribution. Their result is comparable to the $\sigma_{\rm low} = 0.01^{+0.014}_{-0.002}$ that describes nearly 90% of planets when using our two-component model. Hadden & Lithwick (2014) explicitly consider dividing their sample into subsets based on the estimated planet size being larger or smaller than $2.5R_\oplus$. Interestingly, they find an even smaller eccentricity dispersion ($\sigma = 0.008^{+0.003}_{-0.002}$) for subset of larger planets with measurable TTVs. This is opposite of what would be expected based on a simple comparison to either our results or planets discovered by radial velocity surveys, both of which are dominated by significantly larger and/or more massive planets. Future observations and analyses with improved statistical methodology will be important for understanding the underlying nature of these differences.

2.7.1 Potential Biases

We note several potential sources of bias in our characterization of the eccentricity distribution. First, there is a purely geometrical effect due to the fact that we analyze only planet candidates observed to both transit and occult their host stars. Since the geometrical transit and occultation probabilities are both functions of the orbital eccentricity and direction of pericenter, the eccentricity distribution of planets in

our sample is different than the eccentricity distribution of all planets (even if we controlled for orbital separation relative to the star radius). Mathematically, we assume a uniform distribution for the argument of periastron, ω, which is true for all planets in nature, but not true for our sample. Previous studies have suggested that the difference between the eccentricity distribution of transiting planets and all short-period planets is modest (Burke, 2008; Kipping, 2014a). For our sample, the effect will be even weaker, since we require both transit and occultation to be included in our sample and the two have opposite dependence on $e \sin \omega$. We confirm this by computing the distribution of ω for a simulated population generated by starting with a uniform distribution and rejecting planets that do not both transit and occult. We find ω to be very nearly uniformly distributed for these cases, where there is a non-zero effect but it is less that 4%. Therefore, we do not account for this effect in our study. We quantify the significance of this effect in an upcoming study by employing Approximate Bayesian Computing which can naturally model complex selection criteria such as this (Cisewski et al. in prep).

A second potential bias in our characterization of the eccentricity distribution is simply that the population of planet candidates we study may not be representative of all planet candidates. The decreasing geometric transit probability as a function of orbital period or semi-major axis inevitably leads to our sample being dominated by planet candidates with short orbital periods, similar to all population studies based on known transiting planets. Detection probability is also a function of the signal-to-noise of the transit and occultation. Since the transit is typically much deeper than the occultation, the detection probability for the occultation (rather than the transit) is the dominant effect for this study. The occultation signal-to-noise depends on the occultation depth and duration. In principle, the duration depends on the density of the host star, impact parameter, eccentricity and pericenter. In practice, the occultation duration is most sensitive to the impact parameter. Thus, our sample may have excluded some planets with large impact parameters, resulting in the occultation going undetected. Similarly, the occultation depth depends on the effective temperature of the planet, and thus indirectly on the effective temperature of the star, the orbital distance and the stellar radius. Therefore, our sample is likely enriched in planets with larger radii, planets with higher effective temperatures, host stars with higher effective temperatures and smaller radii, and planets orbiting even more closely to their host star. If planet

formation proceeds differently around more massive stars, then the eccentricity distribution for planets in our sample could deviate from the eccentricity distribution of the overall planet population.

Furthermore, our sample could be enhanced with planets with significant eccentricity if tidal forces on planets in eccentric orbits led to substantial heating and increased thermal emission. Since the sample analyzed in this study consists of mostly giant planets and host stars with a single detected transiting planet, the eccentricity distribution for our sample may differ from the eccentricity distribution of smaller planets and/or planets in systems with multiple closely-spaced planets. These potential biases can also be viewed as opportunities to constrain planet formation and tidal theories. By comparing the eccentricity distribution of different planet populations, future studies can quantify how the eccentricity distribution changes with planet size, multiplicity and stellar properties.

We anticipate several ways that future observations will allow for improvements to our analysis. First, we analyzed a subsample of the *Kepler* planet candidates that had already been evaluated for any indication that the *Kepler* Object of Interest (KOI) was actually due to an eclipsing binary star, rather than a planet (e.g., Tenenbaum et al. 2014; Bryson et al. 2013). Both the transit shape and comparison of the target centroid location during and out-of-transit provide powerful diagnostics for recognizing likely false positives. Estimates of the false positive rate are sufficiently low ($\sim 10 - 20\%$; Fressin et al. 2013; Burke et al. 2014), that we can interpret our results in terms of the eccentricity distribution of planets. Nevertheless, one should be cognizant that the sample of planet candidates we analyze may include one or more false positives, such as diluted eclipsing binaries. In particular, our study necessarily selects planet candidates for which an occultation is measured, which may lead to an increased rate of diluted eclipsing binary false positives. Properly accounting for a non-zero fraction of false positives would require a significantly more complex model. Therefore, we leave such work for future studies. Alternatively, future observations of these very interesting planet candidates may identify any remaining false positives and characterize the false positive rate sufficiently well that adding further complexities to the model is not necessary.

2.7.2 Future Research

Transit and occultation observations from future missions could lead to improved understanding of the eccentricity distribution of short-period planets as a function of host star mass and temperature. In particular, there could be differences between a volume-limited sample of target stars and our sample due to the target selection algorithm for the *Kepler* planet search targets as well as variations in *Kepler's* detection sensitivity. Given the *Kepler* target selection criteria and detection sensitivity, most of the planet candidates we analyze are orbiting F and G stars. Future missions such as NASA's Transiting Exoplanet Survey Satellite (TESS) (Ricker et al., 2014) and ESA's PLATO (Rauer et al., 2014) are expected to survey a broader range of target stars and to have a simpler target selection function.

Finally, future observations and analysis of host star properties could also result in improved characterization of correlations between the eccentricity distribution and host star properties. In particular, a large fraction of host star metallicities used in our analysis were derived from photometric observations as opposed to higher quality spectroscopic observations (Huber et al., 2013). As metallicities derived from high-resolution spectroscopy are published for more stars in our sample, we will be able to make more robust conclusions about the potential correlation of the eccentricity distribution with metallicity.

This paper demonstrates that it is practical to apply rigorous HB models to evaluate key dynamical properties of exoplanet populations. In principle, these methods can be readily generalized to provide a more accurate characterization of other aspects of the exoplanet population, such as the frequency of planets as a function of size and orbital period (Foreman-Mackey et al., 2014), the planet radius-mass relationship (Rogers, 2015b; Wolfgang & Lopez, 2015), the distribution of mutual orbital inclinations and multiplicity, and the frequency of small planets in the habitable zone of solar-type stars (Foreman-Mackey et al., 2014). A challenge for future HB analysis will be to develop rigorous model comparison techniques. Each unique problem is often limited by the statistical power of the data, where there is no universal technique applicable in all cases. In practice, the high-dimensional integration required can be computationally challenging. Therefore, careful thought and problem specification is needed, so as to render the necessary calculations tractable. Fortunately, recent collaborations between astronomers and statisticians,

such as the 2013 program on Modern Statistical and Computational Methods for Analysis of *Kepler* Data (SAMSI) at the Statistical and Applied Mathematical Sciences Institute have significantly enhanced the level of sophistication among exoplanet researchers. Forthcoming publications will describe recent efforts application of importance sampling and Approximate Bayesian computing (ABC) to enable application of HB models to more complex problems (e.g., Rogers 2015b; Morton & Winn 2014; Cisewski et al. in prep).

Chapter 3
The Mass-Radius-Eccentricity Distribution of Near-Resonant Transiting Exoplanet Pairs Detected by *Kepler*

3.1 Introduction

With the onslaught of planet candidate data from *Kepler*, scientists are working towards understanding the physical and orbital properties of exoplanet systems on a broad scale, incorporating formation theories that can explain both the configuration of our solar system and planetary systems that are very different. The *Kepler* Mission has provided us with an ensemble of planetary systems (e.g., Burke et al., 2015) for which we are beginning to harness the statistical power offered (e.g., Rowe & Thompson, 2015b). Among the plethora of information to be gained from these exoplanet systems is information about the mass-radius relationship for these systems (Wu & Lithwick, 2013; Weiss et al., 2013; Weiss & Marcy, 2014; Marcy et al., 2014; Wolfgang et al., 2015).

Theoretical studies to explore planet composition beyond measuring bulk density have revealed that multiple compositional regimes can describe the same mass and radius of an exoplanet (Rogers & Seager, 2010; Valencia et al., 2013; Lopez & Fortney, 2014). This feature complicates our ability to deduce whether: 1) a planet of a given mass and radius formed in situ or formed beyond the snow line

and dynamically evolved migrating inward to locations closer to its host where it is more easily detected by *Kepler*, if a planet, 2) if it has undergone episodes of mass gain and/or mass loss, and 3) if it is composed mostly of iron and silicates with a hydrogen-helium abundant atmosphere (Lopez & Fortney, 2014). Moreover, complex internal structure of these planets can yield different planet radii for a given mass (Rogers & Seager, 2010). With this in mind, obtaining empirical constraints on composition distribution from the *Kepler* sample of exoplanets is non-trivial.

Efforts to characterize the population of planetary system compositions for small planets has been in the lime light, as scientists work to constrain the joint mass-radius distribution for planets that span a large range of masses and radii. Studies to constrain the features of a power law mass-radius joint distribution for small planets have provided insight into the broad features of this relation (Weiss et al., 2013; Weiss & Marcy, 2014; Lissauer et al., 2011), but there remains considerable room for improvement.requires scientist One concern arrises when models assume that planets of the same radii must have the same masses and/or compositions (or planets of the same mass must have the same radius). A few studies have allowed a mass-radius relation that is not a one-to-one function. This requires scientists to approach the mass-radius data in a novel way, to gain accuracy, even if it comes at the expense of precision in some cases. Furthermore, large uncertainties in stellar properties translate into large uncertainties in planet properties. With such an understanding we are able to develop models that incorporate the large measurement uncertainty. This has sparked the development of models that can empirically measure the scatter in mass for a given planet radius (Wolfgang & Lopez, 2015; Wolfgang et al., 2015; Rogers, 2015a).

A few early analyses have used mass constraints from transit timing variations (TTVs) to investigate the mass-radius relation for small planets. Most previous studies have used mass constraints from radial velocity (RV) observations (Wolfgang & Lopez, 2015; Wolfgang et al., 2015; Rogers, 2015a; Weiss et al., 2013; Weiss & Marcy, 2014). Here, we improve upon these previous analyses in three ways: 1) we use a more uniform sample of planet candidates that come from the transit timing variations of planet pairs in or near mean-motion resonance (Lithwick et al., 2012), 2) we greatly increase the sample size of small planets in our population study compared to previous studies of the same nature, and 3) by using analytical relations that relate the transit timing variation amplitude to the mass and eccentricity for

planet-pairs. These analytical relations approximate the dominant TTV signature of near resonant planets including the effects of planet periods, masses, eccentricities, and pericenter distances. Constraining the masses of TTV systems using Newton's laws would require much larger computing investments.

One complication is that the TTV signature often results in highly correlated uncertainties for the mass, eccentricity, and pericenter of these planet pairs near a first-order mean motion resonance (MMR). We employ a hierarchical Bayesian (HB) analysis to place probabilistic constraints on masses and eccentricities irrespective of their highly correlated uncertainties, while simultaneously learning about the populations mass-radius-eccentricity features. Thanks to the small fractional uncertainty in the transit timing variation amplitudes and the large sample size, we constrain the mass-radius model in a robust manner. We also model measurements of the planet-to-star radius ratios and stellar radii, in lieu of planet radii, so as to minimize effects of uncertainties in stellar radii. Furthermore, we include systems with weak detections or non-detections to minimize biases against measuring the mass of low mass planets.

This chapter is organized as follows. In §3.2, we describe our physical model relating our observables to our approximations for the TTV amplitudes. In §3.3, we describe the HB statistical model, and the priors selected for the study. In §3.4, we present the results of our HB analysis and investigate potential correlations between the mass-radius-eccentricity distribution and planet or host star properties. In §3.5, we summarize our results, and in §3.6, we discuss our conclusions, potential biases and future work.

3.2 Methods Overview

3.2.1 *Kepler* Data

Of the plethora of multiple transiting planet systems discovered by *Kepler*, many have features resulting from their gravitational interactions with one another that are detectable within the current data. The gravitational tug exhibited by these planets on each other causes the orbits to deviate from strict periodicity. These deviations are known as transit timing variations (TTVs). Within the sample of multiple transiting planet systems, we identify every known planet-pair near a

first-order mean motion resonance (MMR), using the planet catalogs based on data from Q1 through Q16 (Rowe & Thompson, 2015b). These systems are identified reliably since the measurement uncertainties of orbital period for *Kepler* planet candidates are very small. This results in a sample of 120 planet-pairs (See Tables 3.3-3.5). The minimum, median, and maximum values for the observables we use in our study are summarized in Table 3.1 below.

Table 3.1: Minimum, median, and maximum values for *Kepler* data used in this study. Tables 3.3-3.5 show the full dataset.

Variable Name	Minimum	Median	maximum
Planet-to-star radius ratio	0.0042	0.014	0.0718
|TTV amplitude| (days)	6×10^{-06}	0.0018	0.67
Orbital Period (days)	1.68	12.55	284.06
Stellar mass ($M_{\odot}$)	0.189	0.965	4.23
Stellar radius ($R_{\odot}$)	0.169	0.987	2.046

3.2.2 TTV Ephemerides

First, we fit a linear ephemeris to the observed transit times to estimate the orbital periods. For each planet we measured transit timing variations, or deviations from a linear fit to the transit times. We included weak and non-detections of TTVs to minimize detection biases. Many of the planets appeared to have more outlying transit times than one would expect for Gaussian measurement uncertainties. To characterize the measurement errors on the dataset, we extracted a subsample of planet pairs far enough from resonance such that $|\Delta| > 0.08$, where

$$\Delta = \frac{P'(j-1)}{Pj} - 1, \tag{3.1}$$

is the distance from resonance, and j is an integer such that the planet-pair has a period ratio near j:j-1. For $|\Delta| > 0.08$, planets are unlikely to have any detectable TTVs, and their distribution was broader than a Gaussian. Next, we fit a linear and sinusoidal model to the observed transit times:

$$t_n = t_o + nP + V_{Re} \cos\left(\frac{2\pi(t_n - t_{conj})}{P_{TTV}}\right) + V_{Im} \sin\left(\frac{2\pi(t_n - t_{conj})}{P_{TTV}}\right) \tag{3.2}$$

where t_n is the time on the n^{th} transit, t_{conj} is the time of conjunction, V_{Re} and V_{Im} are the TTV amplitudes of the cosine and sine components, respectively, where we adapt the nomenclature from Lithwick et al. (2012), where their derivations are in terms of the real and imaginary components. P_{TTV} is the TTV period, also known as the super period.

$$P_{TTV} = \frac{P'}{j|\Delta|} \tag{3.3}$$

3.2.3 TTV Amplitudes

We assume a Student's-t distribution for transit time uncertainty in order to remain robust to outliers. The likelihood function for computing the observed TTV amplitudes is thus:

$$\mathcal{L} = \prod_{n=1}^{N} T_{\hat{t_n}}(t_n, \sigma_n, \nu) \tag{3.4}$$

The student-t distribution will not penalize the likelihood function as severely as would a Normal distribution, in cases where there are outliers. This is because the tails of the distribution fall gradually, and the density in the tails remains larger. For the sample as a whole, we fitted sinusoidal models to the TTVs with the functional form:

$$M = a_0 + a_1 t + a_2 \sin\left(\frac{2\pi}{P_{TTV}}(t - t_c)\right) + a_3 \cos\left(\frac{2\pi}{P_{TTV}}(t - t_c)\right) \tag{3.5}$$

where M is the TTV model, P_{TTV} is the periodicity of the TTVs The coefficients a_{theta} were all free parameters, and we explored their posteriors for each planet with fits to Equation 3.5 using a Markov Chain Monte Carlo algorithm, and assuming uncertainties were drawn from a student-t distribution. The phase of the TTVs is determined by the conjunctions of the planets, and is zero if conjunction occurs at the transit location and there is no free eccentricity.

We obtained measurements of the planet-to-star radius ratio, host star radius, host star mass, and orbital periods from the *Kepler* planet candidate list based on data from Q1 through Q12 of the NASA Exoplanet Archive[1] (Akeson et al., 2013) for our sample of *Kepler* planets. In cases where there are missing data we use the

[1]http://exoplanetarchive.ipac.caltech.edu, July 28^{th}, 2015

value from the Q1 through Q16 planet candidate catalog[1]. Table 3.1 shows the resultant TTV amplitudes for our sample of planet pairs, Table 3.2 shows the radius ratios and orbital periods for these planets pairs, and Table 3.3 shows the stellar masses and radii for each system. Furthermore, we show in Figure 3.1 the TTV amplitudes, the planet-to-star radius ratio and the fractional uncertainty in R_p/R_* used in our study. The y-axis represents the amplitude of the sine component of the transit timing variation amplitudes, and the x-axis represents the amplitude of the cosine component. The circle sizes are proportional to the planet-to-star radius ratio, and the fractional uncertainty (ratio of 68% credible interval to mode) is shown as the color bar where the blue represents small fractional uncertainty, and the red represents large fractional uncertainty. Figure 3.2 is the same figure zoomed in, with 15 outliers, out of 240 planets total, are not shown for clarity.

3.2.4 Physical Model

Most planet-pairs detected by *Kepler* that orbit near first-order MMR have transit timing variations that can be described analytically. A first-order MMR occurs when two planets have period ratios that are integer ratios, specifically, $j : j - 1$, where j is the resonant number. In systems of two planets near-first order MMR that are not packed together extremely closely, the transit timing variations are approximately sinusoidal. TTV signals that are well-fit by a superposition of two sinusoids could be indicative of a three planet system, but are not included in this study. Fitting N-body simulations to the transit timing variations can potentially help break this degeneracy in mass and eccentricity. N-body simulations are computationally costly, but are perhaps necessary for tightly-packed information-rich TTV systems. In cases where the systems are well approximated by a sinusoid, these analytical approximations are advantageous.

The analytical relations for these TTVs described in Lithwick et al. (2012) relate the amplitude of the TTV signal including sin and cos components (for both the interior and exterior planet) to the other planet's mass and a linear combination of both planets' free eccentricity (the component of the eccentricity not caused by resonance) vectors. The forced eccentricity vector is one component of the eccentricity that is caused by the proximity of the planet-pair to MMR, where the free eccentricity component is not caused by the proximity to resonance, with

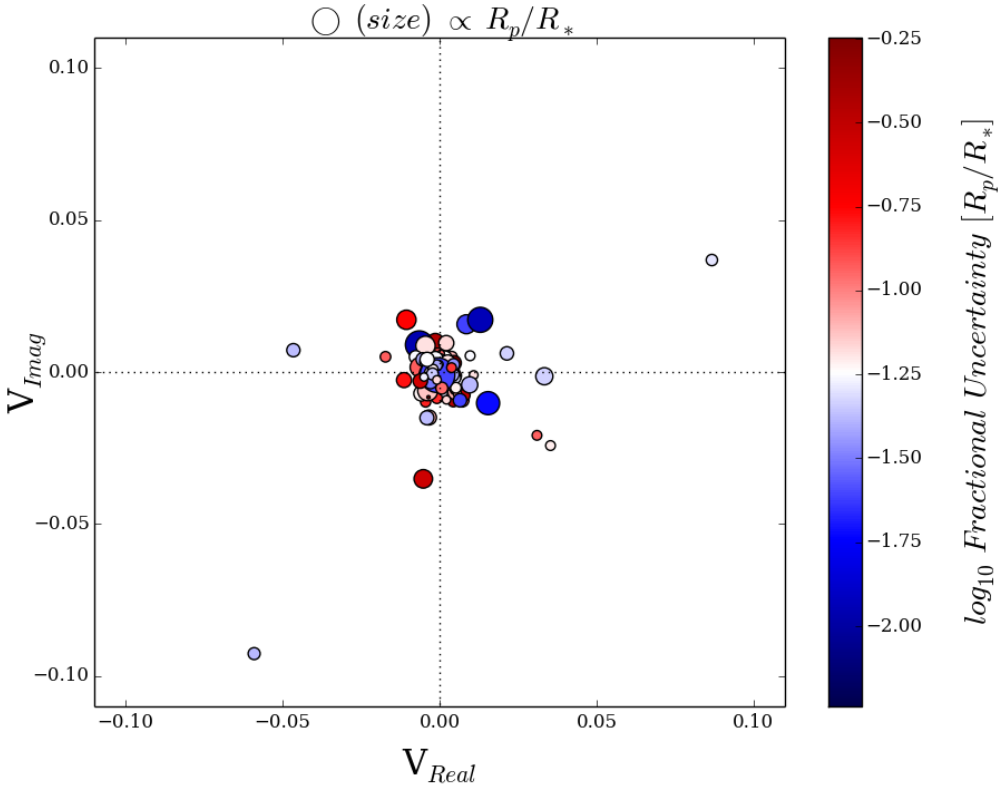

Figure 3.1: Transit timing variation amplitudes with size proportional to the planet-to-star radius ratio and color indicating the fractional uncertainty. The y axis represents the amplitude of the sine component of the transit timing variation amplitudes, and the x-axis represents the amplitude of the cosine component. The circle sizes are proportional to the planet to star radius ratio, and the fractional uncertainty (ratio of 68% credible interval to mode) is show as the color bar where the blue represents small fractional uncertainty, and the red represents large fractional uncertainty.

timescales much less than the secular time. However, these analytical relations make it clear that there will be a near degeneracy between the mass and free eccentricity (from here on referred to simply as eccentricity) when they are constrained by the TTV amplitudes. An advantage of using these analytical relations is that in the limit that these approximations are exact, we can foresee this mass-eccentricity degeneracy. Lithwick et al. (2012) show that in cases where there is non-zero free eccentricity, the phase of the longitude of conjunction of the planet pairs evolves in time. However, if the eccentricity is zero, this phase will remain constant. This can

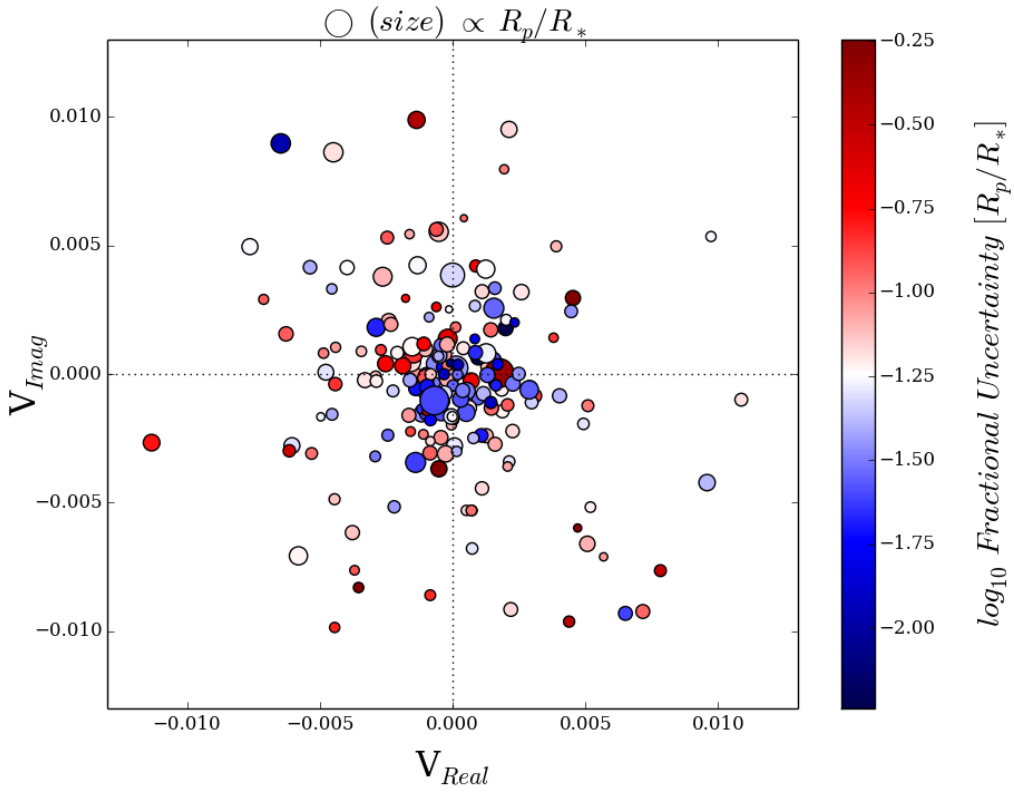

Figure 3.2: Transit timing variation amplitudes with size proportional to the planet-to-star radius ratio and color indicating the fractional uncertainty. The y axis represents the amplitude of the sine component of the transit timing variation amplitudes, and the x-axis represents the amplitude of the cosine component. The circle sizes are proportional to the planet to star radius ratio, and the fractional uncertainty (ratio of 68% credible interval to mode) is show as the color bar where the blue represents small fractional uncertainty, and the red represents large fractional uncertainty. 15 outliers, out of 240 planets total, are not shown for clarity.

help to determine if there is a non-zero eccentricity for a planet.

Following the analytical relations from Lithwick et al. (2012) described above, the real and imaginary parts for the transit timing variation amplitudes from Equations 8, 9 and 10 from Lithwick et al. (2012) correspond to the coefficients of the amplitudes of the sine and cosine terms in Equations 3.6 and 3.7 below.

$$V_{int} = P_{int}\frac{\mu_{ext}}{\pi j^{2/3}(j-1)^{1/3}\Delta}(-f - \frac{3((f\,h_{int} + g\,h_{ext}) - \mathrm{i}(f\,k_{int} + g\,k_{ext}))}{2\Delta}) \quad (3.6)$$

$$V_{ext} = P_{ext}\frac{\mu_{int}}{\pi j\Delta}(-g + \frac{3((f\,h_{int} + g\,h_{ext}) - \mathrm{i}(f\,k_{int} + g\,k_{ext}))}{2\Delta}). \quad (3.7)$$

These equations can then be broken up into their real and imaginary parts (or amplitude of the sine and cosine components) as follows:

$$V_{Re\,int} = -\frac{P_{int}}{\pi j^{2/3(j-1)^{1/3}}\Delta}\,\mu_{ext}\,f \quad (3.8)$$

$$- \frac{3P_{int}}{2\pi j^{2/3(j-1)^{1/3}}\Delta^2}\,\mu_{ext}\,(f\,h_{int} + g\,h_{ext}), \quad (3.9)$$

$$V_{Im\,int} = \frac{3P_{int}}{2\pi j^{2/3(j-1)^{1/3}}\Delta^2}\,\mu_{ext}\,(f\,k_{int} + g\,k_{ext}), \quad (3.10)$$

$$V_{Re\,ext} = -\frac{P_{ext}}{\pi j\Delta}\,\mu_{int}\,g + \frac{3P_{ext}}{2\pi j\Delta^2}\,\mu_{int}\,(f\,h_{int} + g\,h_{ext}), \quad (3.11)$$

$$V_{Im\,ext} = -\frac{3P_{ext}}{2\pi j\Delta^2}\,\mu_{int}\,(f\,k_{int} + g\,k_{ext}). \quad (3.12)$$

In the above equations, P_{int} and P_{ext} are the interior and exterior planet-pair orbital periods, respectively, Δ is the distance from resonance, which is a function of the resonant number j, P_{int}, and P_{ext}, μ_{int} and μ_{ext} are the planet-to-star mass ratios for the interior and exterior planets, respectively, f and g are the Laplace coefficients for the disturbing function derived in Lithwick et al. (2012), and h_{int}, k_{int}, h_{ext}, and k_{ext} are the projected eccentricity values, (where $h = e\cos\omega$ and $k = e\sin\omega$) for the interior and exterior planet, respectively.

3.2.5 Statistical Model for Population

Hierarchical Bayesian (HB) modeling is well suited for studying populations where the measurement uncertainties for individual objects are large because hierarchical models can "pool and muster" strength from the whole population. For example, we combine many weak constraints on planet masses to develop population models for the mass-radius distribution including models that capture the intrinsic scatter in mass for a given radius. In HB modeling, the population-level parameters are

known as hyperparameters, and the mid level parameters that we are trying to infer are known as latent variables. The observables are measurements of the latent variables. The general form for the posterior for the hyperparameter vector $\boldsymbol{\phi}$ and associated latent variables $\boldsymbol{\theta}$ is given as follows:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{d}) \propto p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{\phi}) \tag{3.13}$$

where, $\mathbf{d}$ is the vector of measurements of observables associated with selected latent variables that depends on the specific HB model constructed for a given data set and science questions of interest. The symbol "|" is read as "given". Here, $p(\mathbf{d}, \boldsymbol{\theta}|\boldsymbol{\phi}) = p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})$ is the likelihood function for our HB model, and $p(\boldsymbol{\phi})$ is the prior probability density for the hyperparameters. Here, $p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})$ is the probability of the measurements at the base-level of the hierarchical model given the measurement uncertainties and population model parameters. $p(\boldsymbol{\theta}|\boldsymbol{\phi})$ is the probability of the mid-level model parameters relating the true values of the observed parameters to the population level parameters.

Another useful benefit to the Bayesian approach of multi-level modeling is the ability to obtain probabilistic parameter estimates for variables that are not observable or not yet observed. This is a feature we take advantage of in the case of modeling the joint mass-radius-eccentricity distribution using analytic expressions that relate TTV amplitudes to mass and eccentricity for which both quantities mass and eccentricity we do not have measurements for. Moreover, in the HB framework, we can get more accurate results for parameter estimates by marginalizing over all other parameters in the joint posterior. This becomes important when constructing models in order to remove selection effects or other sources of bias in the parameter estimates. We emphasize careful consideration of biases and selection effects before drawing conclusions about the exoplanet population at large from these population studies.

When developing the population model, the fractional uncertainty of observed variables coupled with the sample size can heavily influence whether or not a data set will be able to constrain a given HB model. For instance, a more complex population model may be a better description of nature, but may be poorly constrained by the given data set. In comparison, a more simplified model would not describe nature perfectly, but might capture the broad features present in the data and provide

useful insight. Therefore, we start with relatively simple physical and statistical models and gradually add complexity.

With this in mind it is important to explore the robustness of HB models for a given science question and data set. The predictions of a robust HB model will be insensitive to small changes in the population model and outliers. Testing the robustness of population models for a given science question and data set is non-trivial. For example, one obvious warning flag would be if the posterior distribution is comparable to the prior distribution. In this case, results are sensitive to the choice of priors. Furthermore, testing an HB model on simulated datasets can expose any mathematical artifacts and identify realistic sample sizes and fractional uncertainties required in order to constrain the variables meaningfully.

3.3 Applying the Hierarchical Model to *Kepler* Data

We aim to characterize the mass-radius-eccentricity relation for a sample of near-first-order MMR planet-pairs from *Kepler* for which we have precisely measured TTV amplitudes (low fractional measurement uncertainties). We do so via constructing a hierarchical Bayesian model and testing our assumptions with simulated data. Once we validate our model and demonstrate it is robust, we apply it to real measurements of the TTV amplitudes described in §3.2.

In our model, we treat the orbital period and resonant number for systems in our sample as known quantities (zero measurement uncertainty). This is a good approximation since these parameters have very small fractional uncertainties. When working with TTV amplitudes, this leaves the mass, eccentricity, and pericenter as unknown variables. This results in a mass-eccentricity degeneracy for TTV amplitude measurements.

We begin by expanding the generalized form for an HB model shown in equation 3.13 above in terms of the variables required for our mass-radius-eccentricity relation using TTV amplitudes for near-first-order MMR planet-pairs. The population level variables are characteristic of the parameterization of the population. We begin constructing our population model by parameterizing the mass-radius joint distribution as a power-law relation with scatter in mass following Wolfgang et al. (2015). We treat the eccentricity distribution as a separable dependency, first modeling it as a one-component Gaussian mixture model, then evolving this to a

two-component Gaussian mixture model following Shabram et al. 2015. Separable in this context means that there is no functional dependance of the given variable on another given variable in the model. We model the occurrence rate of planet radii as a power-law function, which we also treat as separable. We illustrate the dependancies and separability of our model set-up using a graphical model of our Bayesian network shown in Figure 3.3.

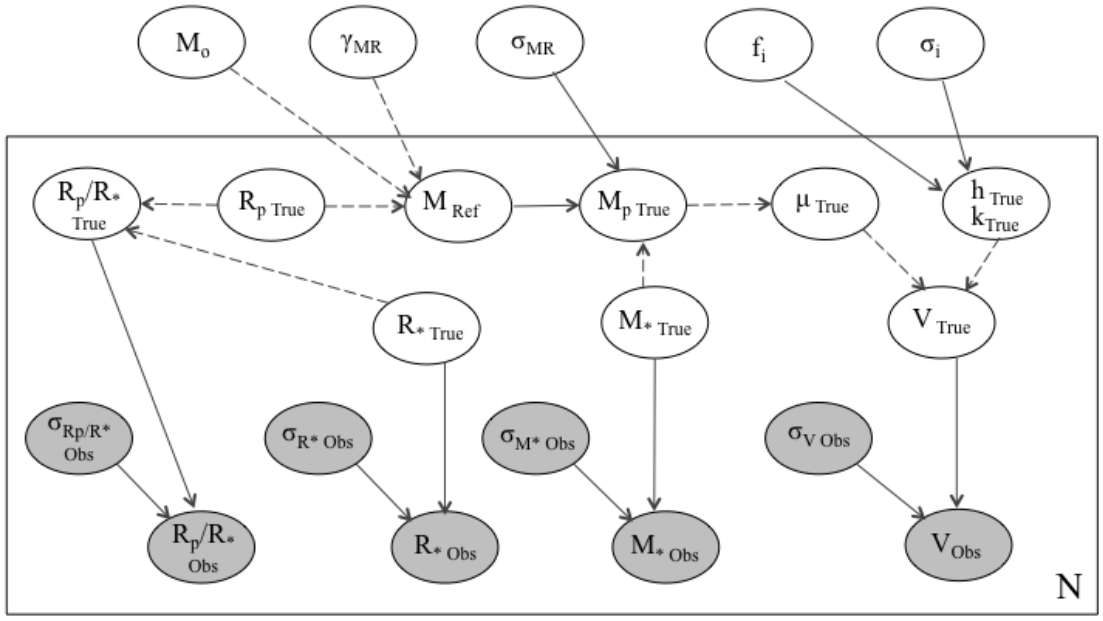

Figure 3.3: Graphical model of Bayesian network. Here we show the relationship between the hyperparameters, latent variables and observables in our mass-radius-eccentricity joint hierarchical Bayesian model. Observables are high level measurements that have been provided by others. The observables are shown in gray. The latent variables are shown in white within the boxed region. The hyperparameters that describe the features of the population model are shown in white above the boxed region. The solid arrows represent probabilistic relationships and can be read as "distributed as", where the dashed arrows represent definitions or physical models. The physical models used in this study (see Equations 3.44 to 3.48) link observables to un-observed quantities (Planet radius, planet mass, eccentricity), and we are able to get posteriors for these un-observed quantities.

3.3.1 Variables

Inherent in the design of our science investigation is making usage of physical models that describe relationships between model parameters and the usage of probability distributions that recognize that noise renders these physical models unable to pin down exact parameter estimates using observables. In our particular investigation, the physical models we are using relate the amplitude of the TTV signal to the planet-to-star mass ratio, eccentricity, orbital period, resonant number, and distance from resonance (see Equations 3.6-3.12 and Tables 3.1 and 3.2) Complicating matters further, these variables are included for both the interior and exterior planets for each observed system. To minimize the effect of uncertainty in the stellar radius, we use the measurements for the planet-to-star radius ratio, and stellar radius in lieu of reported measurement for planet radius. This allows us to use separable probability distributions for planet radius and stellar radius, and link the two variables through a physical model relationship (planet radius divided by stellar radius). In the following expressions, $\mathbf{d}$ from equation 3.13 above is the vector of observations we use in our study:

$$\mathbf{d} = \{\mathbf{V_{obs}}, \boldsymbol{\sigma_{V_{obs}}}, \boldsymbol{R_p/R_{*_{obs}}}, \boldsymbol{\sigma_{R_p/R_{*_{obs}}}}, R_{\star\, obs}, \sigma_{R_{\star\, obs}}, M_{\star\, obs}, \sigma_{M_{\star\, obs}}\} \tag{3.14}$$

$$\mathbf{V_{obs}} = \{V_{obs\, Re\, int}, V_{obs\, Re\, ext}, V_{obs\, Im\, int}, V_{obs\, Im\, ext}\} \tag{3.15}$$

$$\boldsymbol{\sigma_{V_{obs}}} = \{\sigma_{V_{obs\, Re\, int}}, \sigma_{V_{obs\, Re\, ext}}, \sigma_{V_{obs\, Im\, int}}, \sigma_{V_{obs\, Im\, ext}}\} \tag{3.16}$$

$$\boldsymbol{R_p/R_{*_{obs}}} = \{R_p/R_{\star_{obs\, int}}, R_p/R_{\star_{obs\, ext}}\} \tag{3.17}$$

$$\boldsymbol{\sigma_{R_p/R_{*_{obs}}}} = \{\sigma_{R_p/R_{\star_{obs\, int}}}, \sigma_{R_p/R_{\star_{obs\, ext}}}\} \tag{3.18}$$

Here, $\mathbf{V_{obs}}$ is the vector of real and imaginary components of the TTV amplitude for both the interior and exterior planet, and $\boldsymbol{\sigma_{V_{obs}}}$ is the vector of associated amplitude uncertainties. $\boldsymbol{R_p/R_{*_{obs}}}$ is the vector of planet to star radius ratios for the interior and exterior planet, and $\boldsymbol{\sigma_{R_p/R_{*_{obs}}}}$ is the vector of associated measurement uncertainties. $M_{\star\, obs}$ and $R_{\star\, obs}$ are the host star mass and radius for each planet pair and $\sigma_{M_{\star\, obs}}$ and $\sigma_{R_{\star\, obs}}$ are the measurement uncertainties, respectively.

The vector of latent variables in our study, $\boldsymbol{\theta}$ in equation 1, becomes:

$$\boldsymbol{\theta} = \{\mathbf{V_{true}}, \mathbf{R_p/R_{*_{true}}}, \boldsymbol{\mu_{true}}, M_{\star\, true}, R_{\star\, true}, \boldsymbol{R_{p\, true}}, \boldsymbol{M_{p\, true}}, \boldsymbol{h_{true}}, \boldsymbol{k_{true}}\} \tag{3.19}$$

$$\mathbf{V_{true}} = \{V_{true\, Re\, int}, V_{true\, Re\, ext}, V_{true\, Im\, int}, V_{true\, Im\, ext}\} \tag{3.20}$$

$$\boldsymbol{R_p/R_{*_{true}}} = \{R_p/R_{*_{true\, int}}, R_p/R_{*_{true\, ext}}\} \tag{3.21}$$

$$\mathbf{M_{Ref}} = \{\mu_{M\, int}, \mu_{M\, ext}\} \tag{3.22}$$

$$\boldsymbol{R_{p\, true}} = \{R_{p\, true\, int}, R_{p\, true\, ext}\} \tag{3.23}$$

$$\boldsymbol{M_{p\, true}} = \{M_{p\, true\, int}, M_{p\, true\, ext}\} \tag{3.24}$$

$$\boldsymbol{h_{true}} = \{h_{true\, int}, h_{true\, ext}\} \tag{3.25}$$

$$\boldsymbol{k_{true}} = \{k_{true\, int}, k_{true\, ext}\} \tag{3.26}$$

Above, $\mathbf{V_{true}}$ is the vector of true values that occurred in nature for the real and imaginary components of the transit timing variation amplitude for the interior and exterior planet pairs. Furthermore, $\boldsymbol{R_p/R_{*_{true}}}$, $\mathbf{M_{Ref}}$, $\boldsymbol{R_{p\, true}}$, $\boldsymbol{M_{p\, true}}$, $\boldsymbol{h_{true}}$, and $\boldsymbol{k_{true}}$ are the true values for the planet-to-star radius ratio, planet-to-star mass ratio, planet radius, planet mass, and the components of the eccentricity distribution ($h = e\cos\omega$, and $k = e\sin\omega$), respectively.

The vector of hyperparameters that describe the features of the developed population model, $\boldsymbol{\phi}$ are:

$$\boldsymbol{\phi} = \{M_0, \gamma_{MR}, \sigma_{MR}, \boldsymbol{f}, \boldsymbol{\sigma_c}\}, \tag{3.27}$$

where M_0 is the constant for the power law planet mass-radius relationship, γ_{MR} is the exponent for this planet mass-radius relationship, and σ_{MR} is the scatter in planet mass for a given radius. $\boldsymbol{f}$ and $\boldsymbol{\sigma_c}$ are the mixture fractions and dispersions for a two-component Rayleigh mixture eccentricity model (see Chapter 2 for details regarding the eccentricity distribution model we use here), and γ_{R_p} is the power-law index for the planet occurrence rate as a function of of planet radii.

3.3.2 The Population Model

We now expand the hierarchical Bayesian model described in equation 3.13 in terms of our mass-radius-eccentricity population model as follows:

$$\begin{aligned}
&p(\gamma_{MR}, M_0, \sigma_{RM}, \mathbf{f}, \boldsymbol{\sigma_c}, \boldsymbol{R_p/R_{*true}}, \boldsymbol{\mu_{true}}, R_{*\,true}, M_{*\,true}, \boldsymbol{h_{true}}, \boldsymbol{k_{true}}, \\
&\boldsymbol{R_{p\,true}}, \boldsymbol{M_{p\,true}} | \mathbf{V_{obs}}, \boldsymbol{\sigma_{V_{obs}}}, \boldsymbol{R_p/R_{\star\,obs}}, \boldsymbol{\sigma_{R_p/R_{\star\,obs}}}, R_{*\,obs}, \sigma_{R_{*\,obs}}, M_{*\,obs}, \sigma_{M_{*\,obs}}) \\
&\propto p(\gamma_{MR}) p(M_0) p(\sigma_{RM}) p(\boldsymbol{f}) p(\boldsymbol{\sigma_c}) \\
&p(R_{*\,true}) p(M_{*\,true}) p(\boldsymbol{c}) p(\boldsymbol{h_{true}}, \boldsymbol{k_{true}} | \boldsymbol{\sigma_c}) \\
&p(\boldsymbol{R_{p\,true}}) p(\boldsymbol{M_{Ref}} | M_o, \gamma_{MR}, \boldsymbol{R_p/R_{*true}}, \boldsymbol{R_{p\,true}}) \\
&p(\boldsymbol{M_{p\,true}} | \boldsymbol{M_{Ref}}, M_o, \gamma_{MR}, \sigma_{MR}, \boldsymbol{R_p/R_{*true}}, \boldsymbol{R_{p\,true}}) \\
&p(R_{*\,obs} | R_{*\,true}, \sigma_{R_{*\,obs}}) p(M_{*\,obs} | M_{*\,true}, \sigma_{M_{*\,obs}}) \\
&p(\boldsymbol{R_p/R_{*\,obs}} | \boldsymbol{R_p/R_{*\,true}}, \boldsymbol{\sigma_{R_p/R_{*\,obs}}}) p(\boldsymbol{V_{obs}} | \boldsymbol{V_{true}}, \boldsymbol{\sigma_{V_{obs}}})
\end{aligned} \tag{3.28}$$

Here "p()" indicate each separable function in our HB model. We define the dependencies and distributions of each of the separable equations shown above in Equations 3.29-3.52 in §3.3.3 through §3.3.5 below.

3.3.3 Hyperpriors

Equations 3.29 through 3.34 below show the hyperprior distributions for our HB model. These are the variables that describe the shape of our population model as inferred from our data set. In these Equations, the "$\sim$" is read "distributed as" and the "$\leftarrow$" is read "defined as". We assign a Uniform distribution between 0 and 2 for the power-law exponent γ_{MR} of our mass-radius relation within the HB model. Previous studies of planets in a similar size regime have found that the parameters describing a power-law model for the mass-radius relation to be $M_o = 2.69M_\oplus$ and $\gamma_{MR} = 0.93$ (Weiss & Marcy, 2014), and $M_o = 2.7M_\oplus$, $\gamma_{MR} = 1.3$, and $\sigma_{MR} = 1.9M_\oplus$ (Wolfgang et al., 2015). We choose a range that encompasses those results. The mass constant M_o represents the mass of a one Earth radii planet. We choose a Uniform prior distribution for M_o ranging from $0.25M_\oplus$ to $2M_\oplus$. We select this range based on physical intuition for plausible sizes of Earth-size planets. We choose a Uniform prior distribution ranging from $0M_\oplus$ to $2M_\oplus$ for the scatter in our mass-radius relation. Based on results from Chapter 1 exploring the eccentricity distribution of a sample of planets from *Kepler*, we use a mixture of Normal distributions to characterize the distribution for $h = e\cos\omega$ and $k = e\sin\omega$ in our HM model. For the mixture fractions $\boldsymbol{f}$ describing this eccentricity distribution, we choose a Dirichlet prior distribution with $\alpha = 1$, which forces the sum of the fractions to 1. We use a Categorical distribution of the mixture fractions to assign which mixture components will be used for which planet, which is described in the mid-level model specification in §3.3.4. For the dispersion of each component $\boldsymbol{\sigma_c}$, we choose a Uniform prior distribution ranging from 0 to 1.

$$\gamma_{MR} \sim Uniform(0, 2) \tag{3.29}$$

$$M_o \sim Uniform(0.25M_\oplus, 2M_\oplus) \tag{3.30}$$

$$\sigma_{MR} \sim Uniform(0, 2) \tag{3.31}$$

$$\boldsymbol{f} \sim Dirichlet(\alpha = 1) for\, i = 1...N_{comp} \tag{3.32}$$

$$\boldsymbol{\sigma_c} \sim Uniform(0, 1) for\, i = 1...N_{comp} \tag{3.33}$$

$$\tag{3.34}$$

3.3.4 Mid-Level Model Specification

We now move into the description of the mid-level of our HB model, where we describe the relationship of the latent variables to the population level parameters (a.k.a hyperpriors described above). We choose a Uniform distribution for stellar mass M_* and stellar radius R_* ranging from $0.8M_\odot$ to $3M_\odot$ and $0.5R_\odot$ to $2R_\odot$, respectively, shown in Equations 3.35 and 3.36 below. Based of research in Shabram et al. 2015 and described in Chapter 1 of this dissertation, we parameterize the eccentricity distribution of the planets in our sample as a mixture of Normal distributions centered at zero. $Ncomp$ is the number of mixture components in the Gaussian mixture model used to parameterize the projected eccentricity distribution, and in this case, $Ncomp = 2$. We assign a categorical distribution (Equation 3.34) to the variable c, where c takes on values equal to 1 for component-one, or 2 for component-two, of the two-component Gaussian mixture model for projected eccentricity. Equation 3.38 below show the prior distributions for $h = e\cos\omega$ and $k = e\sin\omega$, where, $\boldsymbol{\sigma_c}$ represent the dispersions for each mixture component. We truncate the prior distributions for h and k such that $e = \sqrt{h^2 + k^2} < 1$. The prior distributions are renormalized after truncation. The planet radii are assigned to a Uniform prior distribution ranging from $0.1R_\oplus$ to $15R_\oplus$ in Equation 3.37 We define a reference mass M_{Ref} as a power-law mass-radius relation in Equation 3.40. The true planet mass is distributed Normally around the reference mass, shown in Equation 3.41, with a dispersion of σ_{MR}, the hyperparameter describing the scatter in planet mass for a given planet radius. We define the true planet-to-star mass ratio μ_{true} as the planet mass divided by the stellar mass in Equation 3.42, and the true planet-to-star radius ratio $\boldsymbol{R_p/R_{*true}}$ as the true planet radius divided by the stellar radius in Equation 3.43 below.

$$R_{*\,true} \sim Uniform(0.5R_\odot, 2R_\odot) \tag{3.35}$$

$$M_{*\,true} \sim Uniform(0.8M_\odot, 3M_\odot) \tag{3.36}$$

$$c_j \sim Categorical(\boldsymbol{f})\; for\; j = 1...N_{comp} \tag{3.37}$$

$$\boldsymbol{h_{true}, k_{true}}|\boldsymbol{\sigma_c} \sim Normal(0, \boldsymbol{\sigma_c})\; for\; \sqrt{h^2 + k^2} < 1 \tag{3.38}$$

$$\boldsymbol{R_{p\,true}} \sim Uniform(0.1R_\oplus, 15R_\oplus) \tag{3.39}$$

$$\boldsymbol{M_{Ref}}|M_o, \gamma_{MR}, \boldsymbol{R_{p\,true}} \leftarrow M_o\, \boldsymbol{R_{p\,true}}^{\gamma_{MR}} \tag{3.40}$$

$$\boldsymbol{M_{p\,true}}|M_{Ref}, M_o, \gamma_{MR}, \sigma_{MR}, \boldsymbol{R_{p\,true}} \sim Normal(\boldsymbol{M_{Ref}}, \sigma_{MR}) \tag{3.41}$$

$$\boldsymbol{\mu_{true}} \leftarrow \boldsymbol{M_{p\,true}}/M_{*\,true} \tag{3.42}$$

$$\boldsymbol{R_p/R_{*\,true}} \leftarrow \boldsymbol{R_{p\,true}}/R_{*\,true} \tag{3.43}$$

It is important to note that the "$\leftarrow$" represents model definitions, or physical models, and are used to link variables but are not probabilistic and therefore do not appear in Equation 3.25. Below the $\mathbf{V_{true}}$ vector is expanded into the real (and imaginary parts for the interior and exterior planet pair as derived in Equations 3.6 through 3.12 in §3.2.4:

$$V_{true\,Re\,int} \leftarrow -\frac{P_{int}}{\pi j^{2/3(j-1)^{1/3}}\Delta}\,\mu_{true\,ext}\,f \tag{3.44}$$

$$-\frac{3P_{int}}{2\pi j^{2/3(j-1)^{1/3}}\Delta^2}\,\mu_{true\,ext}\,(f\,h_{true\,int}+g\,h_{true\,ext}) \tag{3.45}$$

$$V_{true\,Im\,int} \leftarrow \frac{3P_{int}}{2\pi j^{2/3(j-1)^{1/3}}\Delta^2}\,\mu_{true\,ext}\,(f\,k_{true\,int}+g\,k_{true\,ext}) \tag{3.46}$$

$$V_{true\,Re\,ext} \leftarrow -\frac{P_{ext}}{\pi j\Delta}\,\mu_{true\,int}\,g+\frac{3P_{ext}}{2\pi j\Delta^2}\,\mu_{true\,int}\,(f\,h_{true\,int}+g\,h_{true\,ext}) \tag{3.47}$$

$$V_{true\,Im\,ext} \leftarrow -\frac{3P_{ext}}{2\pi j\Delta^2}\,\mu_{true\,int}\,(f\,k_{true\,int}+g\,k_{true\,ext}) \tag{3.48}$$

3.3.5 Base-Level Model Specification

The base-level of the HB model links the variables for which we have measurements to their associated latent variables. In this study we have measurements and uncertainties represented as summary statistics for the planet-to-star radius ratios of both the interior and exterior planet-pairs, the stellar mass and radii for each system, and the amplitude of the sine and cosine components of the interior and exterior planet-pair TTV amplitudes. In this HB model we assume the observed value is distributed Normally about the true value of the system with a dispersion equal to the uncertainty reported in the summary statistics. In principle, the observed value is not necessarily Normally distributed. Therefore, it would be useful to use the full posteriors for the measurements in future studies.

$$R_{*\,obs}|R_{*\,true},\sigma_{R_{*\,obs}} \sim Normal(R_{*\,true},\sigma_{R_{*\,obs}}) \tag{3.49}$$

$$M_{*\,obs}|M_{*\,true},\sigma_{M_{*\,obs}} \sim Normal(M_{*\,true},\sigma_{M_{*\,obs}}) \tag{3.50}$$

$$\boldsymbol{R_p/R_{*\,obs}}|\boldsymbol{R_p/R_{*\,true}},\boldsymbol{\sigma_{R_p/R_{*\,obs}}} \sim Normal(\boldsymbol{R_p/R_{*\,true}},\boldsymbol{\sigma_{R_p/R_{*\,obs}}}) \tag{3.51}$$

$$\boldsymbol{V_{obs}}|\boldsymbol{V_{true}},\boldsymbol{\sigma_{V_{obs}}} \sim Normal(\boldsymbol{V_{true}},\boldsymbol{\sigma_{V_{obs}}}) \tag{3.52}$$

3.3.6 Evaluating the Hierarchical Model

We sample from the posterior using MCMC. To calculate Markov chains we use the publicly-available code Just Another Gibbs Sampler (JAGS; Plummer, 2003). JAGS uses Gibbs sampling when possible, and otherwise reverts to standard Metropolis–Hastings. We simultaneously sample from both the posterior distributions for the population parameters and the posterior predictive distributions for each observable. We evaluate the Gelman-Rubin (R) ratio to test for non–convergence, and accept chains with an $\mathrm{R} < 1.01$. We also look at the autocorrelation function for the Markov chains and accept cases that have a zero crossing at a lag of ≤ 5 using the default thinning by a factor of 500, and discarding all but the final 1000 iterations for 5 chains in total. The exact JAGS input model used in our study is shown as Figure 3.4.

3.4 Results

The primary goal of our study is to characterize population level parameters that describe a joint planet mass, radius and eccentricity distribution. Our method simultaneously provides estimates of planet mass and eccentricity for the individual planets in our sample. We start with a simplistic model, testing with simulated data. With simulated data, the population level parameters are known, so we can see how the sample size and measurement uncertainty impacts the model's ability to recover posteriors that accurately characterize the true chosen population level parameters for the simulated data set. This allows us to discover the threshold sample size and data quality needed to constrain the model, and adjust the model accordingly so that it can be constrained by the real data set.

```
model
{
  pi <- 3.14
  mass.const ~ dunif(0.0E+00, 3)
  mass.radius.exp ~ dunif(0.0E+00, 2)
  sigma.mass.radius ~ dunif(0.0E+00, 3)
  m.sun <- 3.32996E+05
  r.sun <- 110
  min.radius.s <- 0.5 * r.sun
  max.radius.s <- 2 * r.sun
  min.radius.p <- 0.1
  max.radius.p <- 10
  min.mass.s <- 0.8 * m.sun
  max.mass.s <- 3 * m.sun
  for (j in 1:2) {
    e.sigma[j] ~ dunif(0.0E+00, 1)
    e.phi[j] <- 1/(e.sigma[j] * e.sigma[j])
    a[j] <- 1
  }
  f ~ ddirch(a[])
  for (i in 1:Ndata) {
    C1[i] <- P.int[i]/((pi * (j[i]^(2/3)) * ((j[i] - 1)^1/3) * delta[i]))
    C2[i] <- 3/(2 * delta[i])
    C3[i] <- P.ext[i]/(pi * j[i] * delta[i])
    c[i] ~ dcat(f[])
    h.true.int[i] ~ dnorm(0.0E+00, e.phi[c[i]]) T(-1, 1)
    k.true.int[i] ~ dnorm(0.0E+00, e.phi[c[i]]) T(-sqrt(1-h.true.int[i] * h.true.int[i]), sqrt(1 - h.true.int[i] * h.true.int[i]))
    h.true.ext[i] ~ dnorm(0.0E+00, e.phi[c[i]]) T(-1, 1)
    k.true.ext[i] ~ dnorm(0.0E+00, e.phi[c[i]]) T(-sqrt(1 - h.true.ext[i] * h.true.ext[i]), sqrt(1 - h.true.ext[i] * h.true.ext[i]))
    radius.s.true[i] ~ dunif(min.radius.s, max.radius.s)
    radius.s.obs[i] ~ dnorm(radius.s.true[i], 1/(sigma.radius.s[i] * sigma.radius.s[i])) T(0.0E+00, )
    mass.s.true[i] ~ dunif(min.mass.s, max.mass.s)
    mass.s.obs[i] ~ dnorm(mass.s.true[i], 1/(sigma.mass.s[i] * sigma.mass.s[i]))
    radius.p.true.int[i] ~ dunif(0.1, 15)
    radius.ratio.true.int[i] <- radius.p.true.int[i]/radius.s.true[i]
    radius.ratio.obs.int[i] ~ dnorm(radius.ratio.true.int[i], 1/(sigma.radius.ratio.obs.int[i] * sigma.radius.ratio.obs.int[i]))
    mass.ref.int[i] <- mass.const * radius.p.true.int[i]^mass.radius.exp
    mass.p.true.int[i] ~ dnorm(mass.ref.int[i], 1/(sigma.mass.radius * sigma.mass.radius)) T(0.0E+00, )
    mass.ratio.true.int[i] <- mass.p.true.int[i]/mass.s.true[i]
    va.true.int[i] <- ((-1) * C1[i] * mass.ratio.true.ext[i] * fc[i]) -
      (C1[i] * C2[i] * mass.ratio.true.ext[i] * ((fc[i] * h.true.int[i]) + (g[i] * h.true.ext[i])))
    va.obs.int[i] ~ dnorm(va.true.int[i], 1/(sigma.va.obs.int[i] * sigma.va.obs.int[i]))
    vb.true.int[i] <- C1[i] * C2[i] * mass.ratio.true.ext[i] * ((fc[i] * k.true.int[i]) * (g[i] * k.true.ext[i]))
    vb.obs.int[i] ~ dnorm(vb.true.int[i], 1/(sigma.vb.obs.int[i] * sigma.vb.obs.int[i]))
    radius.p.true.ext[i] ~ dunif(0.1, 15)
    radius.ratio.true.ext[i] <- radius.p.true.ext[i]/radius.s.true[i]
    radius.ratio.obs.ext[i] ~ dnorm(radius.ratio.true.ext[i], 1/(sigma.radius.ratio.obs.ext[i] * sigma.radius.ratio.obs.ext[i]))
    mass.ref.ext[i] <- mass.const * radius.p.true.ext[i]^mass.radius.exp
    mass.p.true.ext[i] ~ dnorm(mass.ref.ext[i], 1/(sigma.mass.radius * sigma.mass.radius)) T(0.0E+00, )
    mass.ratio.true.ext[i] <- mass.p.true.ext[i]/mass.s.true[i]
    va.true.ext[i] <- ((-1) * C3[i] * mass.ratio.true.int[i] * g[i]) + (C2[i] * C3[i] * mass.ratio.true.int[i] *
      ((fc[i] * h.true.int[i]) * (g[i] * h.true.ext[i])))
    va.obs.ext[i] ~ dnorm(va.true.ext[i], 1/(sigma.va.obs.ext[i] * sigma.va.obs.ext[i]))
    vb.true.ext[i] <- (-1) * C2[i] * C3[i] * mass.ratio.true.int[i] * ((fc[i] * k.true.int[i]) * (g[i] * k.true.ext[i]))
    vb.obs.ext[i] ~ dnorm(vb.true.ext[i], 1/(sigma.vb.obs.ext[i] * sigma.vb.obs.ext[i]))
  }
}
```

Figure 3.4: Just Another Gibbs Sampler mass-radius-eccentricity HB model, with R syntax highlighting.

3.4.1 Validating the Hierarchical Model: Testing Analysis Models with Corresponding Simulated Data

3.4.1.1 Effect of Measurement Uncertainties

In Figure 3.5 we show posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs. Here, we use a single Gaussian distribution to parameterize the distribution of projected eccentricities $h = e\cos\omega$ and $k = e\sin\omega$ with a fixed dispersion of the projected eccentricity distribution $\sigma_{hk} = 0.01$ (eccentricities will be small for all simulated planets). The observable quantities are the sine and cosine components of

the transit timing variation amplitude $\boldsymbol{V_{obs}}$, the stellar mass M_*, the stellar radius R_*, and the planet-to-star radius ratio for the interior and exterior planet pairs $\boldsymbol{R_p/R_{*\,obs}}$, described in §3.3.1. In this simulated data case, we set the uncertainties of all observables artificially small to test the robustness of this model to the quantity of planet pairs in our real data set (120 pairs, 240 planet candidates total). This means that for this simulated data case, $\boldsymbol{\sigma_{V_{obs}}} = \sigma_{M_*} = \sigma_{R_*} = \boldsymbol{\sigma_{R_p/R_{*\,obs}}} = 0.001$. The chosen values for the population level parameters for the power-law mass-radius relation parameterization are $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, and $\sigma_{MR} = 1.0\,M_\oplus$, described in §3.3.4. All true values for the hyperparameters fall within the 68% credible interval of the marginal posteriors for these hyperparameters.

In Figure 3.6, we perform the same analysis as we did above, but this time we set the measurement uncertainties for the observables to be the real uncertainties from the *Kepler* data set. These measurement uncertainties are shown in Tables 3.1 through 3.3. The mean values for these measurement uncertainties are $\boldsymbol{\sigma_{V_{obs}}} = 0.14\,days$, $\sigma_{M_*} = 0.014\,M_\odot$, $\sigma_{R_*} = 0.11\,R_\odot$, and $\boldsymbol{\sigma_{R_p/R_{*\,obs}}} = 0.0005$. In this test case, the true values for the hyperparameters ($M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, and $\sigma_{MR} = 1.0\,M_\oplus$), are all still within the 68% credible interval of the marginal posteriors shown in Figure 3.6. From this we conclude that when using the true measurement uncertainties and the quantity of data available, we will be able to constrain the population level parameters to with the 68% credible interval of the posterior modes.

Next, we try a population model where we choose a prior for the dispersion of a one-component Gaussian distribution model for eccentricity, allowing for flexibility in the eccentricity distribution. In this case, we take our population model described above, but we now have a new population level parameter which we call σ_{hk}, which represents the dispersion for a one-component Gaussian distribution. We use a uniform prior for σ_{hk} ranging from 0 to 3. In our simulated data case, the chosen population parameters are: $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, $\sigma_{MR} = 1.0\,M_\oplus$, and $\sigma_{hk} = 0.21$. We choose a larger dispersion for the eccentricity distribution than the previous simulated cases where σ_{hk} was set to a small value of 0.01, in order to asses the impact of systems where planets may contain cases with larger eccentricities. Figure 3.7 shows the results of this simulation. Here, the mass-radius relation true hyperparameters all fall within the 68% credible interval of the marginal posterior distributions. The hyperparameter for the eccentricity distribution, however, is just

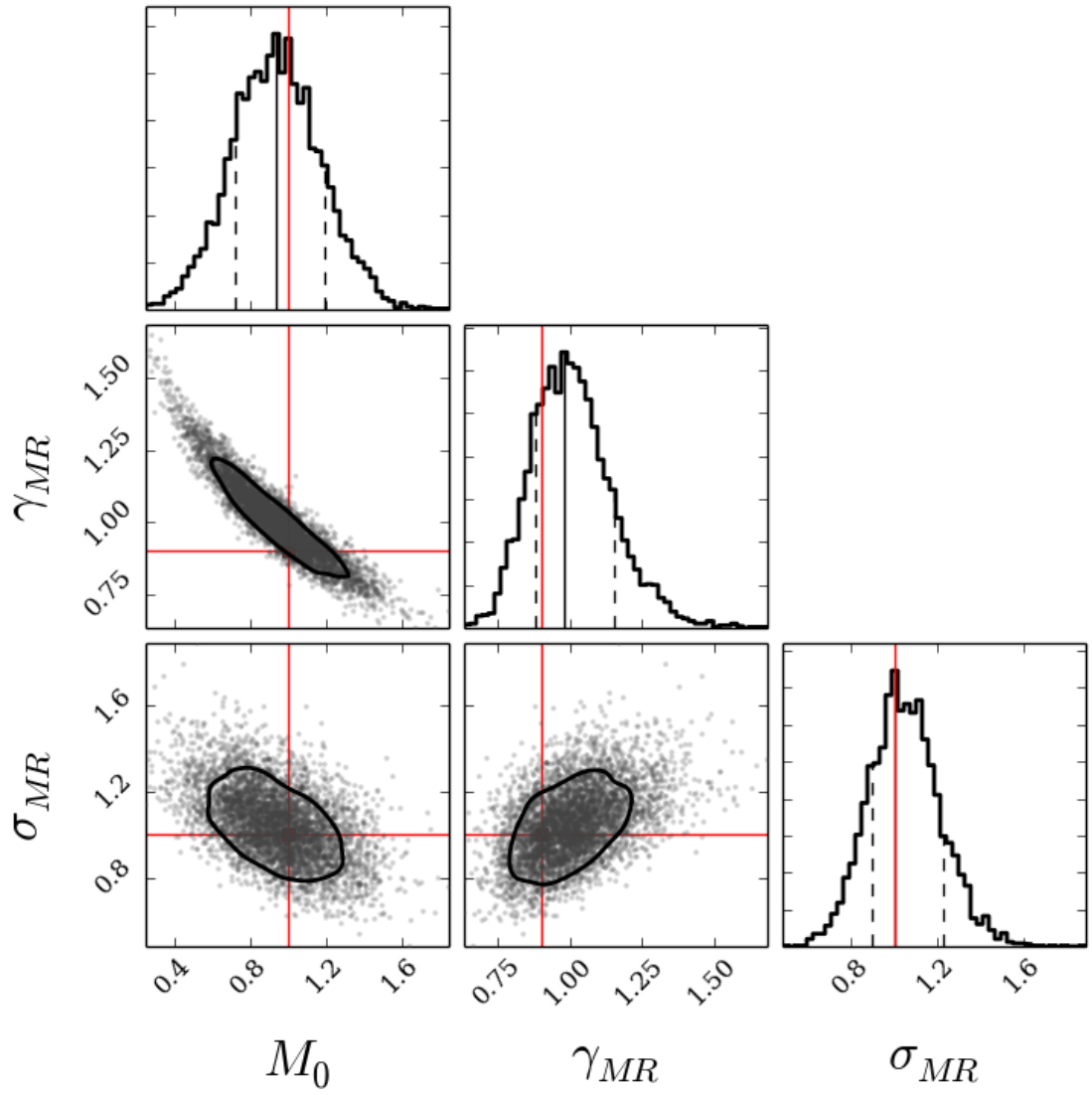

Figure 3.5: Posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs. Here, the eccentricity distribution is fixed. In this simulated data case, we set the uncertainties of our observables artificially small (all set to 0.001) to test the robustness of this model to the quantity of planet pairs in our real data set. Shown in red are the true values for the mass constant $M_o = 1.0\,M_\oplus$, the power-law index $\gamma_{MR} = 0.9$ and the scatter in mass for a given radii $\sigma_{MR} = 1.0\,M_\oplus$. All true values fall within the 68% credible interval (shown above as black dashed lines) of the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves.

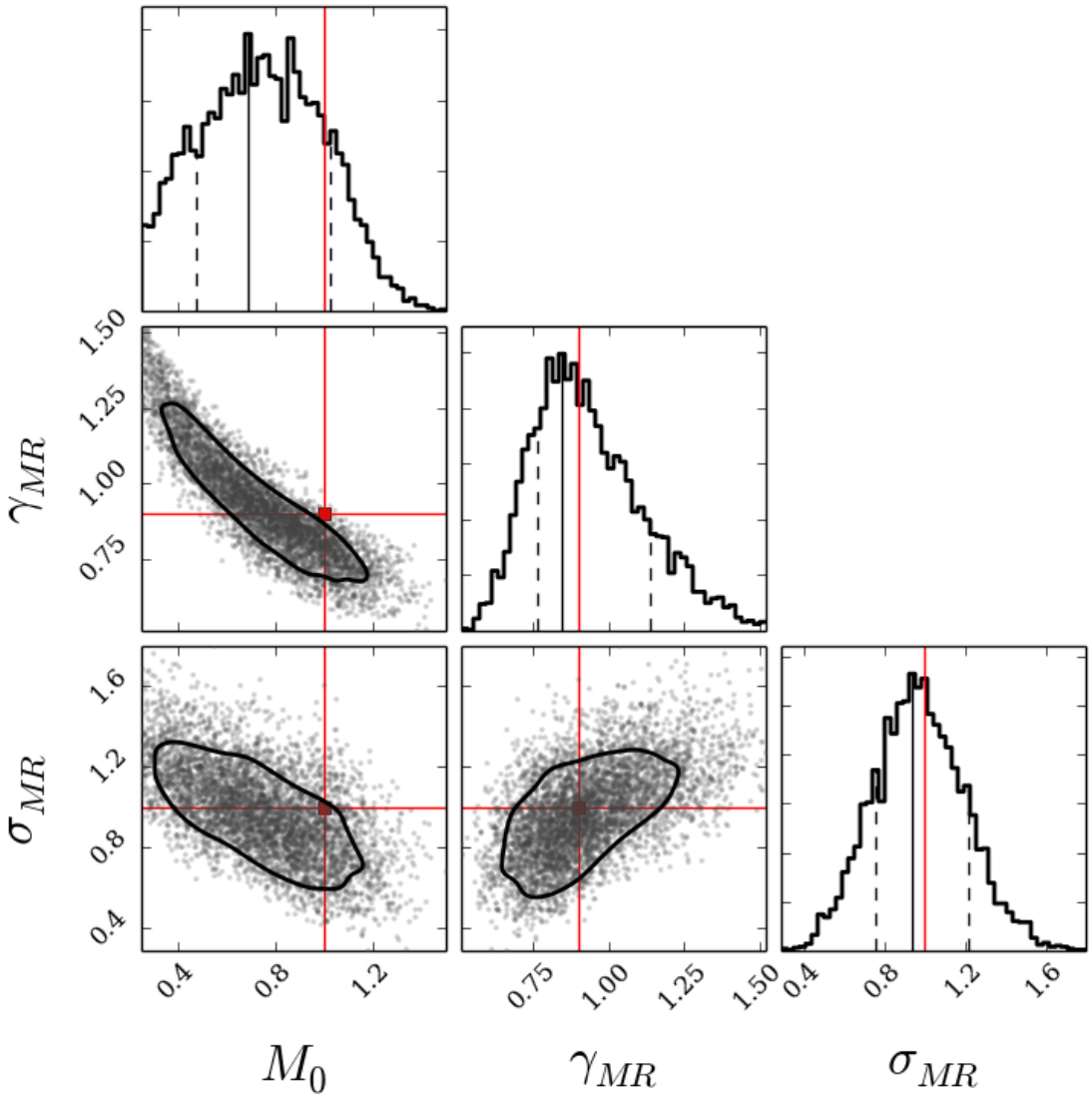

Figure 3.6: Posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs, with a fixed eccentricity distribution. In this simulated data case, we set the uncertainties of our observables to the true observational uncertainties from the *Kepler* data, to test how the real measurement uncertainties when combined with the quantity of data available, will impact our results. Shown in red are the true values for the mass constant $M_o = 1.0\,M_\oplus$, the power-law index $\gamma_{MR} = 0.9$ and the scatter in mass for a given radii $\sigma_{MR} = 1.0\,M_\oplus$. All true values fall within the 68% credible interval, shown above as black dashed lines, of the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves.

outside the 68% credible interval, but still well within the posterior distribution, so we accept this test case.

Next, we try another simulated data scenario, where we increase the complexity of our HB model to include a two-component Gaussian mixture model for the projected eccentricity distribution. Results for this simulated data case are shown in Figure 3.8. Here, the hyperparameters for this HB model are $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, $\sigma_{MR} = 1.0\,M_\oplus$, and the two-component Gaussian mixture model hyperparameters are now $\sigma_{low} = 0.03$, $\sigma_{high} = 0.65$, $f_{low} = 0.75$, and $f_{high} = 0.25$. Here f_{low} and f_{high} are the mixture fraction for the two-components, and σ_{low} and σ_{high} are the two dispersion components. The subscript "$_{low}$" indicates the component of the projected eccentricity distribution with a smaller dispersion (most eccentricities in this population are small), and the subscript "$_{high}$" indicates the larger dispersion component that allows for some cases with higher eccentricity. For this particular case, we investigate the posterior modes and credible intervals for planet radius vs. planet mass. We show these posterior summary statistics in Figure 3.9. The gray bars represent the 68% credible interval of the marginal posteriors for planet mass and radii. The red points represent the posterior modes for the interior planets and the blue points represent the posterior modes for the exterior planets. The black points represent the true planet masses and planet radii generated for this particular model. The thick solid black curve is the true mass-radius relation that the simulated data are drawn from. The region of scatter in mass for a given radius, is shown as the region between the thick dotted black lines.

3.4.1.2 Highly Correlated Uncertainties in Mass and Eccentricity

We inspected the posteriors for latent variables of the mass-radius-eccentricity HB model that uses a two-component Gaussian mixture model. In particular, we inspect the mass and projected eccentricity (h and k) for interior and exterior planet pairs, to investigate the impact of the degeneracy between these variables when using the analytical approximations for TTV amplitude described in §3.2.4. Approximately 10% of the simulated systems did not return the true generated value used to within the 68% credible interval for marginal posteriors of these latent variables. Figures 3.10 through 3.13 show for examples cases for these latent variables of interest. In all four figures, the 68% credible intervals are shown as the

dotted black lines on the histograms of the marginal posteriors, posterior modes for the marginals are shown as the solid black lines, and the 68% credible regions for the 2D posteriors are shown as the solid black curve.

Figure 3.10 shows the 115^{th} simulated system in the population. In this case, the simulated h and k values are both very close to zero. We can see that the histograms of the marginal posteriors for the interior and exterior planet mass constrain the true generated value to within the 68% credible intervals shown as the dotted black lines. Figure 3.11 shows the 43^{th} simulated system in the population. In this particular case, the measurements of the TTV amplitudes were unavailable, so we included injected TTV measurement uncertainties set to 2 into the data set for this planet. In this case the posteriors for the projected eccentricity do not encompass the truths, but the simulation does well for recovering the true planet masses. Figure 3.12 shows the 40^{th} simulated system in the population. In this case we see a bimodal solution for k, for both the interior and exterior planet. Solutions where k is closer to zero produce larger mass planets, and cases where k deviates from zero produces smaller mass planets. In this case, the HB model is able to recover the true generated values well. Furthermore, Figure 3.13 shows the 41^{st} simulated system in the population. In this case, k deviates a small amount from zero and you can see two modes start to appear. The masses and projected eccentricities are still well constrained in this case. By evaluating the parameter space for different generated values for these selected latent variables, we can see that our model is performing well for simulated data.

Finally, we test an HB model where we use a mixture of two power-law distributions to parameterize the mass-radius-eccentricity distribution. In this case, there are now two-components for each hyperparameter describing the power-law mass radius relation. Figure 3.14 shows the marginal posteriors for these hyperparameters when this two-component power-law model is applied to the same simulated data model as used in Figure 3.8. We conclude that this model is performing at a satisfactory level, since the truths for the mass-radius power-law population parameters are within the 68% credible region, and the truths for the eccentricity distribution population parameters are within the posteriors.

3.4.2 The Mass-Radius-Eccentricity Distribution for *Kepler* Planet Candidates

After evaluating the robustness of our mass-radius-eccentricity HB model on simulated data, we now apply this model to the real *Kepler* data. We summarize our findings in Figures 3.15 through 3.18 below.

In Figure 3.15 we show the posteriors for hyperparameters for real *Kepler* data. Here we use a one-component Gaussian mixture model for the eccentricity distribution in our joint mass-radius-eccentricity HB model. The posterior modes and credible intervals for the hyperparameters are reported in Table 3.2. In Figure 3.16 we show again the hyperparameters for real *Kepler* data. In this case, we use a two-component Gaussian mixture model for the eccentricity distribution in our joint mass-radius-eccentricity HB model. The posterior modes and credible intervals for the hyperparameters in this HB model are reported in Table 3.2. The two-component power-law mass-radius HB model applied to simulated data in Figure 3.14 did not perform well when applied to the real data set.

We then investigate the mass-radius relationship by plotting planet mass vs. planet radius using the means from the latent variable posteriors for planet mass and planet radius in Figure 3.17. In this particular case, we investigate the real *Kepler* data using a mass-radius-eccentricity HB model that parameterizes the projected eccentricity as a two-component Gaussian mixture model. The 68% credible intervals of the marginal posteriors for mass and radius are show as gray bars. The mass-radius relation using the posterior means for the mass constant M_o and the exponent γ_{MR} is shown as the solid black curve. The dotted black curve represents the posterior mean for the scatter in planet mass for a given radius σ_{MR}. We question the validity of the masses reported here, as they are unreasonably small for our sample of planets. Furthermore, Figure 3.18 shows 10 draws of the eccentricity PDF from the joint posterior, from a mass-radius-eccentricity HB model that uses a two-component Gaussian mixture model to parameterize the eccentricity distribution. The top panel shows a linear scale and the bottom panel shows a log scale in order to see the larger dispersion component more clearly.

Table 3.2: Posterior modes and 68% credible intervals of hyperparameters for mass-radius-eccentricity HB model, using a one-component Gaussian mixture model for eccentricity.

				68% credible interval	
Hyperparameter	Mean	Median	Mode	+	−
$M_0\,[M_\oplus]$	0.016	0.01	2.96×10^{-06}	0.029	0.003
γ_{MR}	0.85	0.80	0.12	1.37	0.12
$\sigma_{MR}\,[M_\oplus]$	0.26	0.25	0.24	0.087	0.045
σ_{hk}	0.32	0.31	0.31	0.06	0.04

Table 3.3: Posterior modes and 68% credible intervals of hyperparameters for mass-radius-eccentricity HB model, using a two-component Gaussian mixture model for eccentricity.

				68% credible interval	
Hyperparameter	Mean	Median	Mode	+	−
$M_0\,[M_\oplus]$	0.043	0.029	1.79×10^{-06}	0.08	0.007
γ_{MR}	0.71	0.61	0.04	1.25	0.138
$\sigma_{MR}\,[M_\oplus]$	0.49	0.48	0.49	0.11	0.10
$\sigma_{low\,hk}$	0.028	0.027	0.025	0.014	0.006
$\sigma_{high\,hk}$	0.74	0.76	0.93	0.001	0.370
$f_{low\,hk}$	0.74	0.75	0.75	0.067	0.083
$f_{high\,hk}$	0.26	0.25	0.25	0.056	0.094

3.5 Summary of Results

In theory, we are able to infer the distribution parameters for a mass-radius power-law relation by marginalizing over eccentricity, while simultaneously computing probabilistic estimates for the masses and eccentricities of these systems. This would allow us to extract masses and eccentricities without doing computationally expensive numerical simulations often not feasible for this subclass of observations. Furthermore, we harness the statistical power of using small fractional uncertainties by using planet-to-star radius ratios, stellar masses and stellar radii measurements, in addition to well-measured TTV amplitudes, as our observables. For simulated data, our model is robust and performs well in all our test cases. When applying the model to the real *Kepler* data, we find that the model returns unreasonably small mass estimates, and thus a poor estimate of the mass radius relationship.

3.6 Discussion

We are currently investigating the application of the HB model developed in this study to real *Kepler* data. Next, we will try parameterizing the uncertainty in TTV amplitudes as a Student-t distribution instead of a Normal distribution. This may

help the model be more robust to outliers. It is possible that in cases where the systems are far from resonance, the Δ term will be larger, and in the analytical relations shown in §3.2.4, there may be more weight given to the eccentricity terms for a given TTV amplitude, which would push the masses estimates closer to zero. This effect should have been present in the simulated data scenarios but it is possible that the large fraction of non-detections and small TTV amplitudes in the real data set changes the impact of this feature.

The plethora of short-period tightly-packed inner planetary systems (STIPS) has provided a unique opportunity to constrain the masses of the sub-Neptune class planets, where mass constraints for these planets are out of reach for current RV surveys. This has allowed scientists to begin to probe the bulk properties of these sub-Neptunes, and to empirically constrain mass-radius relationships for small planets. N-body simulations have been used to obtain probabilistic parameter estimates of planet masses for tightly packed, information rich, multi's and are useful in cases where non-sinusoidal TTV behavior exists (e.g., Carter et al., 2012; Nesvorný & Morbidelli, 2008). However, these simulations are costly, and unnecessary for many of the *Kepler* planet candidates pairs near first-order MMR, where the analytical approximations relate the amplitude of the sinusoidal fit to the TTV curve to the mass and eccentricity of the planet pairs (Lithwick et al., 2012). Previous studies have assumed that when the phase of the longitude of conjunction is zero, the free eccentricity is zero, and used this to estimate the mass of the planet (Lithwick et al., 2012). This assumption is no longer needed in our probabilistic assessment of the masses and eccentricities of planet near first order MMR.

One important feature of our study is that we include cases where TTV signals are very small or not observed by predicting they may be there by the planets orbital periods which are well-known. We can predict the sinusoidal fit to these TTV curves by using the expression for the TTV period (super period), within the regime of these analytical approximations. In these cases, the TTV amplitude might be consistent with a zero mass planet. By including these non-detections, we can approach a more random sample, and minimize sampling bias.

By applying a HB model to this problem, we not only work to constrain the behavior of these planets as a population, but we are also able to get parameter estimates for the mass and eccentricity of each unique planet pair in in our population study irrespective of knowing the TTV phase or if there is a free eccentricity for a

planet that is inconsistent with zero. A caveat to this interpretation is that the results are only as accurate as the model specification. By evaluating the robustness of our model, we have improved the confidence in our interpretations in spite of the fact that the results are conditional on the model.

Parameter estimates in the HB framework experience "shrinkage" when compared to parameter estimates calculated for each system independently, improving the accuracy of these measurements. We develop the HB model in this study as a tool to extract population information, and as mentioned above, the model specification is specific to the quantity and quality of the data set. In cases where models are not constrained by the data, (e.g., we get back our priors for population level parameters), those models are considered poor parameterizations. This can impact the marginal posteriors for the latent variables, or parameter estimates that make up the population, further emphasizing the importance of developing robust models in the HB framework. As we approach a more robust parameterization of the data in the HB framework, our results are approaching more accuracy as dictated by the data set. Furthermore, in cases where the quality and quantity of the data set is improved, a more complex model may be able to be constrained, probabilistic parameter estimates improved upon, and more information extracted from the data in general (e.g., joint distributions for multiple populations such as mass-radius-eccentricity and orbital period).

Resonant repulsion has been a suggested mechanism to cause planets to move outside of MMR, where a dissipative force transfers orbital energy into heat while angular momentum is conserved Lithwick & Wu (2012), acting to re-excite small eccentricities in circularized planet pairs near first order mean motion resonance. Tides have been ruled out as a dissipative mechanism for sub-Neptune mass planets, as the tidal circularization timescale is on the order of, or greater than, the age of the system, or it requires initial eccentricities of the planets to be larger than is physically plausible (Silburt & Rein, 2015). Planet-planetesimal disk interactions are a plausible explanation for this dissipative force (Chatterjee & Ford, 2015). Assuming that the planets migrate into first-order MMR via smooth disk migration initially, and after the gaseous disk has cleared, many micro scattering evens with the remaining planetesimal disk can dissipate orbital energy in such a way where these planet pairs move just wide of first-order MMR (Chatterjee & Ford, 2015). During this phase of resonant repulsion, material is still potentially accreting, and

dissipation of orbital energy could potentially cause flash heating events that may impact the composition and structure of these planets. Furthermore, these planets may experience more than one resonance acting on them, which may change the resonant repulsion timescale, or the magnitude of heating.

3.6.1 Biases in this Study

Intrinsic to the *Kepler* mission design are selection effects creating biases that can skew the interpretation of our population studies if extended to planets "in general". *Kepler* detects planets that transit from the vantage point from Earth, which is more likely for planets that are on short-period orbits. In this study we limit our interpretation of the mass-radius-eccentricity distribution to sub-Neptune like planets that are near a first-order MMR. These planets may be a unique class of planets in terms of composition and structure.

Completeness is another issue impacting our sample. Planets in our sample come from less tightly packed low information content near resonant pairs in comparison to systems such as Kepler-36 (Carter et al., 2012). Tightly packed systems may not have TTV curves well approximated by sinusoids and are poor candidates for applying the Lithwick et al. (2012) analytical relations. There could be more planet-pairs near first-order MMR where one planet transit and the other does not.

Another concern affecting studies of the mass-radius relation is the large uncertainty in stellar properties that affects both the transit data and radial velocity data used in these studies (e.g., Weiss & Marcy, 2014; Wolfgang et al., 2015). The HB model is designed for large uncertainties, however, in these studies we have so far treated the uncertainties as random or recreated the posteriors for parameter estimates using summary statistics and parameterizing these as Gaussian. Doing this is a good first step in constructing comprehensive HB models, but in order to account for biases or noise unique to each independent measurement in these population studies, we need to use the full posteriors for parameter estimates for our population constituents.

An interesting question would be if the dynamical mechanisms responsible for creating tightly-packed systems and multi's with larger spacing, are independent of planet composition or structure. Perhaps there is an interplay between the composition and structure with dynamical evolution (e.g., tightly packed systems

experience more mass loss since they have smaller Hill spheres). Future population exploration will aim to determine the relationship between stochastic processes that produce the architectures we are seeing for these sub-Neptune multi's from *Kepler*.

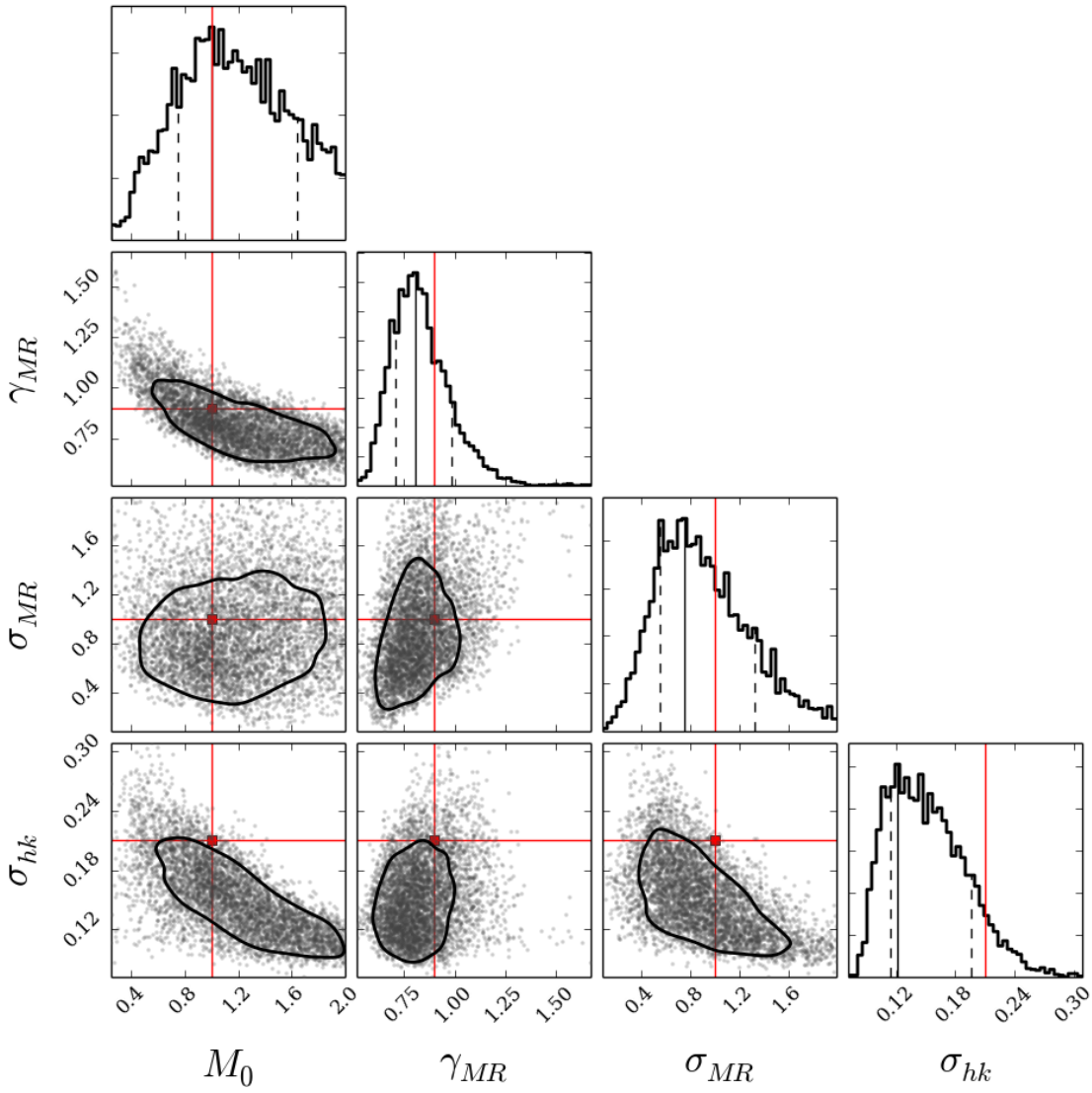

Figure 3.7: Posterior distributions for hyperparameters of our mass-radius-eccentricity HB model using 120 simulated *Kepler* planet candidate pairs. In this simulated data case, we assign a Uniform prior distribution to the dispersion for the projected eccentricity distribution, and test the ability of the data to constrain a model on increased complexity. Shown in red are the true values for the mass constant $M_o = 1.0\,M_\oplus$, the power-law index $\gamma_{MR} = 0.9$ and the scatter in mass for a given radii $\sigma_{MR} = 1.0\,M_\oplus$, and the dispersion for the projected eccentricity $\sigma_{hk} = 0.21$. All true values fall with the posterior distribution for the marginals shown above. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves.

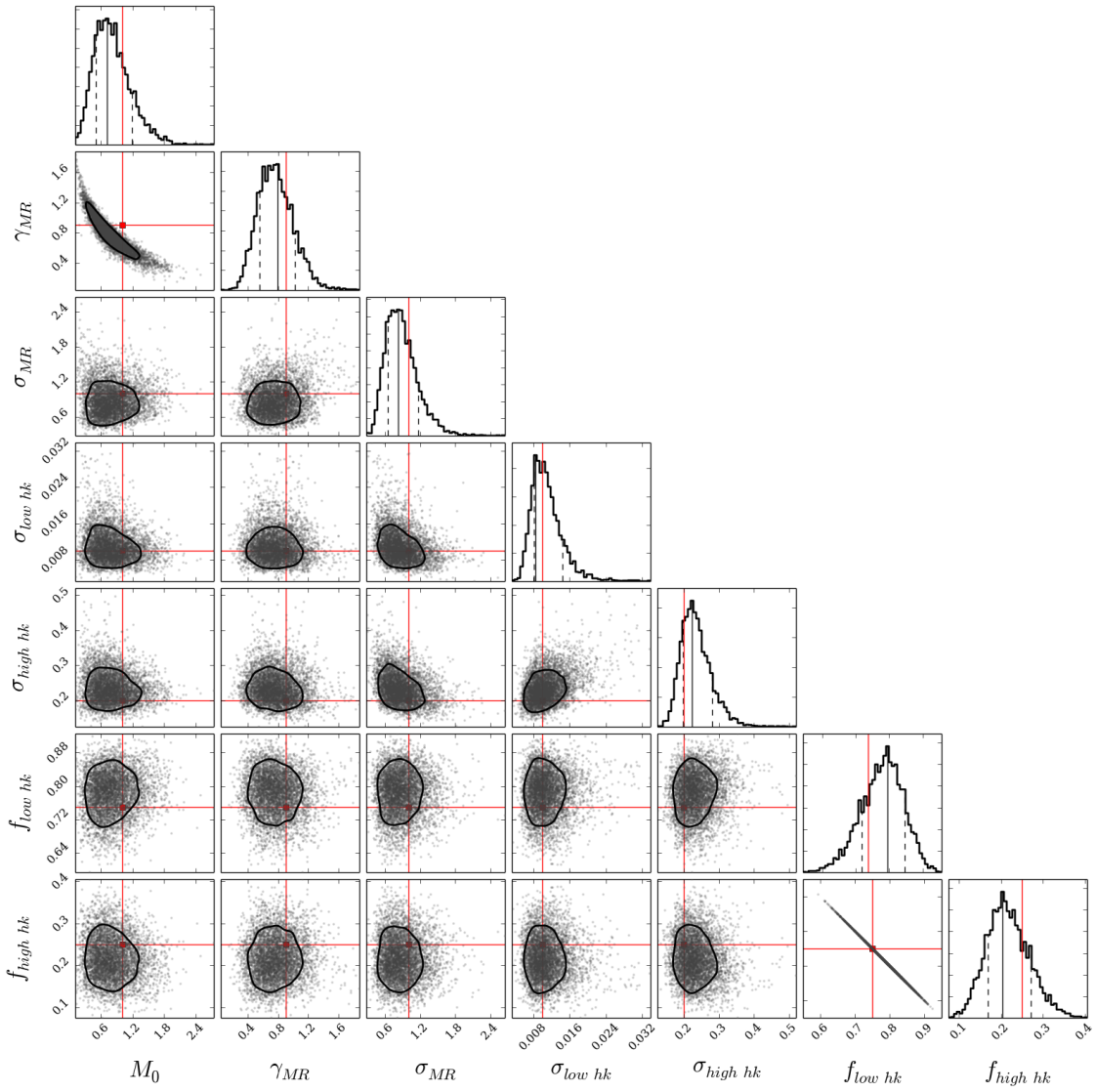

Figure 3.8: Hyperparameters for simulated *Kepler* data. Here, the hyperparameters for this HB model are $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, $\sigma_{MR} = 1.0\,M_\oplus$, and the two-component Gaussian mixture model hyperparameters are now $\sigma_{low} = 0.03$, $\sigma_{high} = 0.65$, $f_{low} = 0.75$, and $f_{high} = 0.25$. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves.

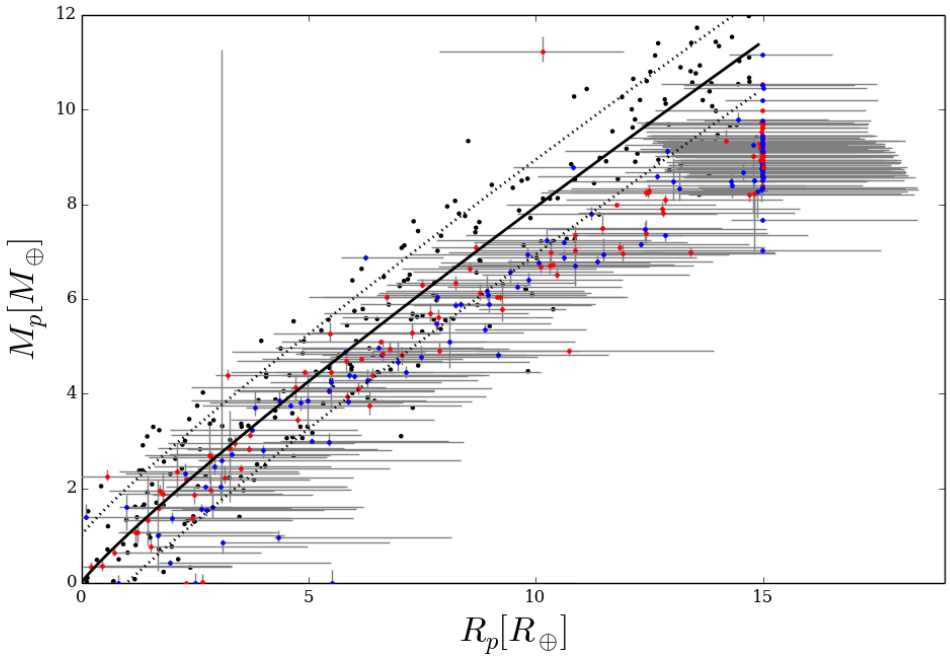

Figure 3.9: Posterior modes and credible intervals for planet radius vs. planet mass from simulated mass-radius-eccentricity HB model. The gray bars represent the 68% credible interval of the marginal posteriors for planet mass and radii. The red points represent the posterior modes for the interior planets and the blue points represent the posterior modes for the exterior planets. The black points represent the true planet masses and planet radii generated for this particular model. The thick solid black curve is the true mass-radius relation that the simulated data are drawn from: $M_o = 1.0\,M_\oplus$, $\gamma_{MR} = 0.9$, for $M_p = M_o R_p^{\gamma_{MR}}$. The region of scatter in mass for a given radii, $\sigma_{MR} = 1.0\,M_\oplus$, is shown as the region between the thick dotted black lines.

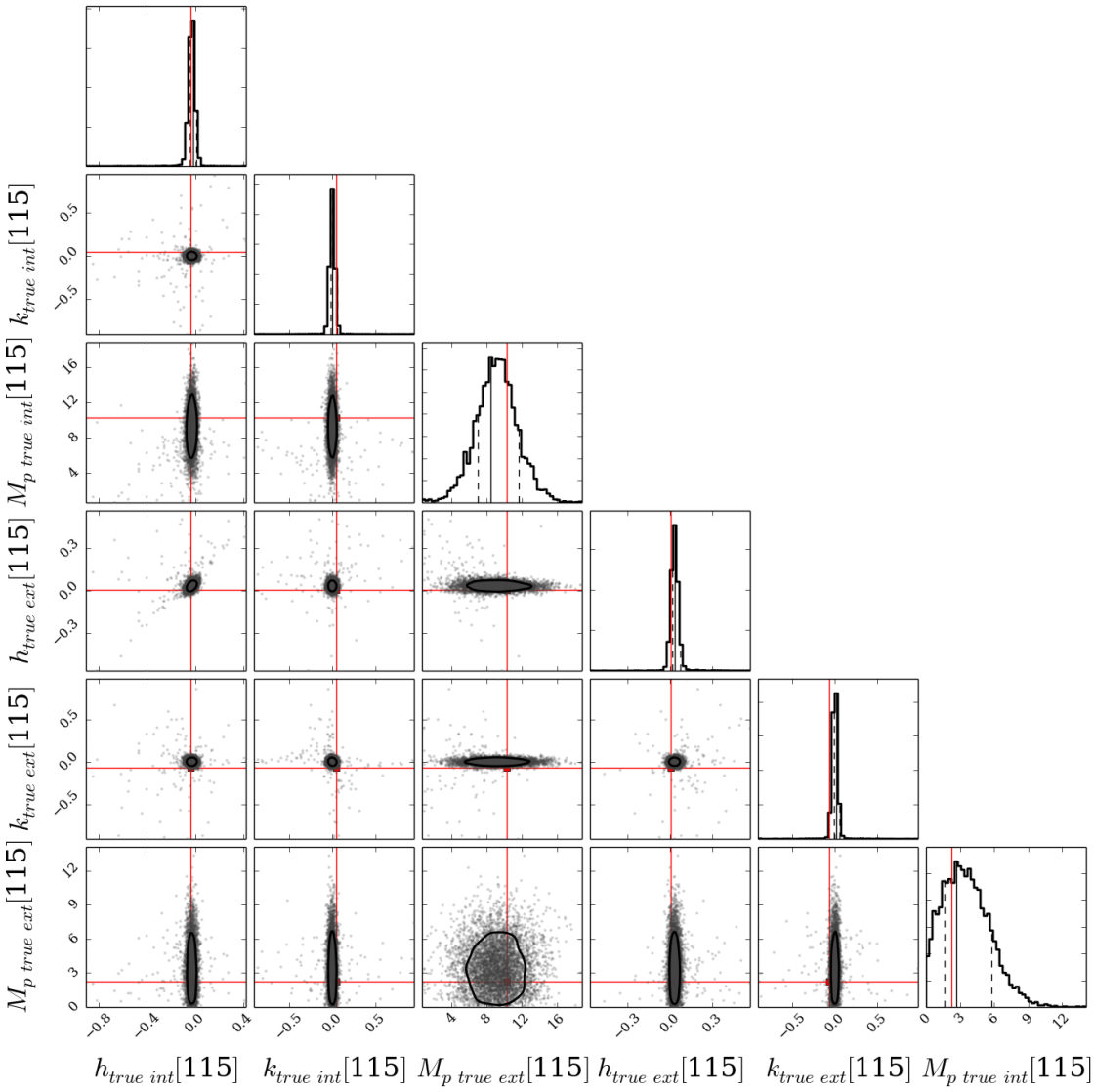

Figure 3.10: Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 115^{th} simulated system in the population. In this case, the simulated h and k values are both very close to zero. We can see that the histograms of the marginal posteriors for the interior and exterior planet mass constrain the true generated value to within the 68% credible intervals shown as the dotted black lines. Posterior modes for the marginals are shown as the solid black lines. The 68% credible region for the 2D posteriors is shown as the solid black curve.

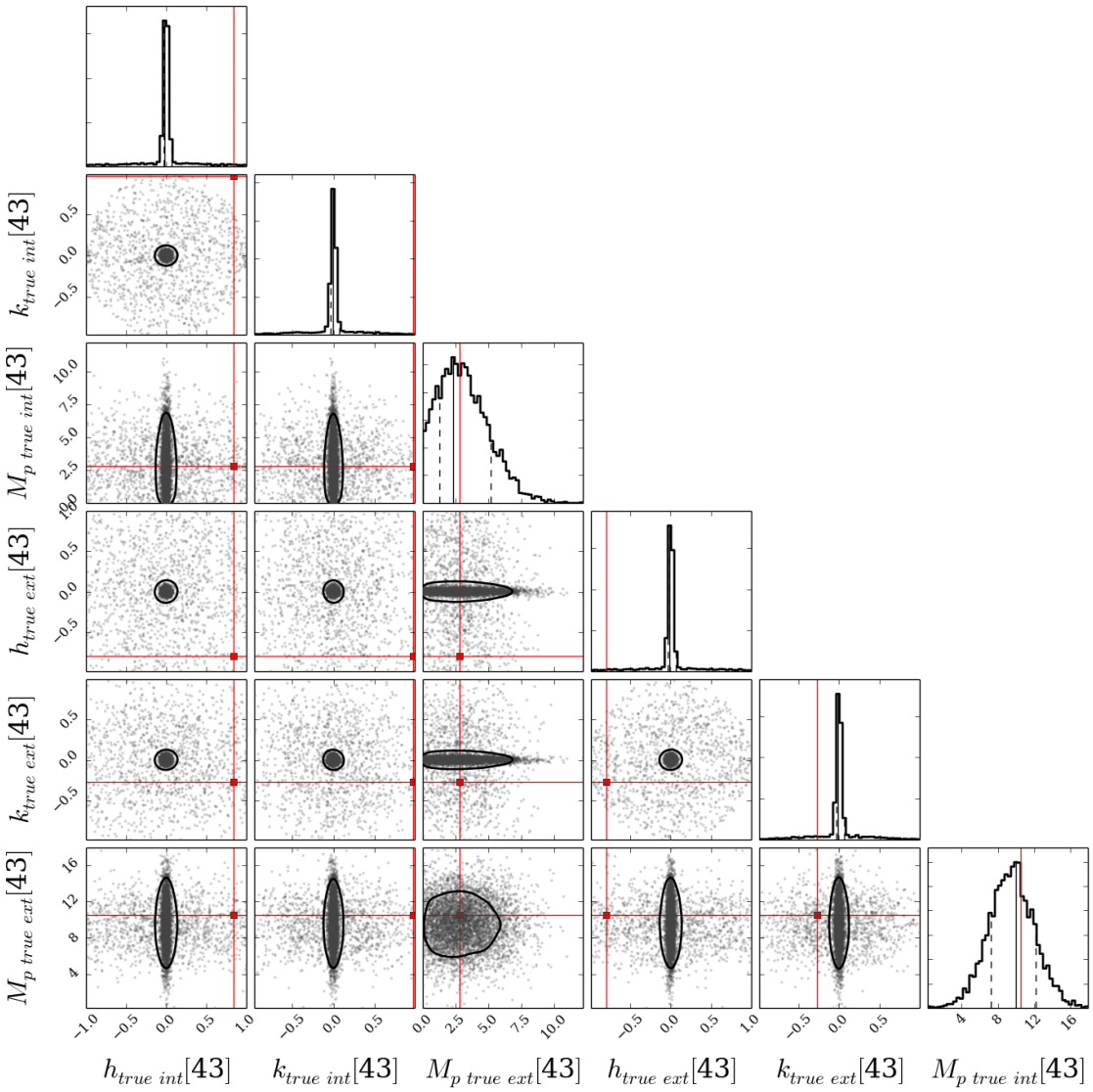

Figure 3.11: Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 43^{th} simulated system in the population. The 68% credible intervals are shown as the dotted black lines on the histograms of the marginal posteriors. Posterior modes for the marginals are shown as the solid black lines. The 68% credible regions for the 2D posteriors are shown as the solid black curve. In this particular case, the measurements of the TTV amplitudes were unavailable, so we included injected TTV measurement uncertainties set to 2 into the data set for this planet. In this case the posteriors for the projected eccentricity do not encompass the truths, but the simulation does will for recovering the true planet masses.

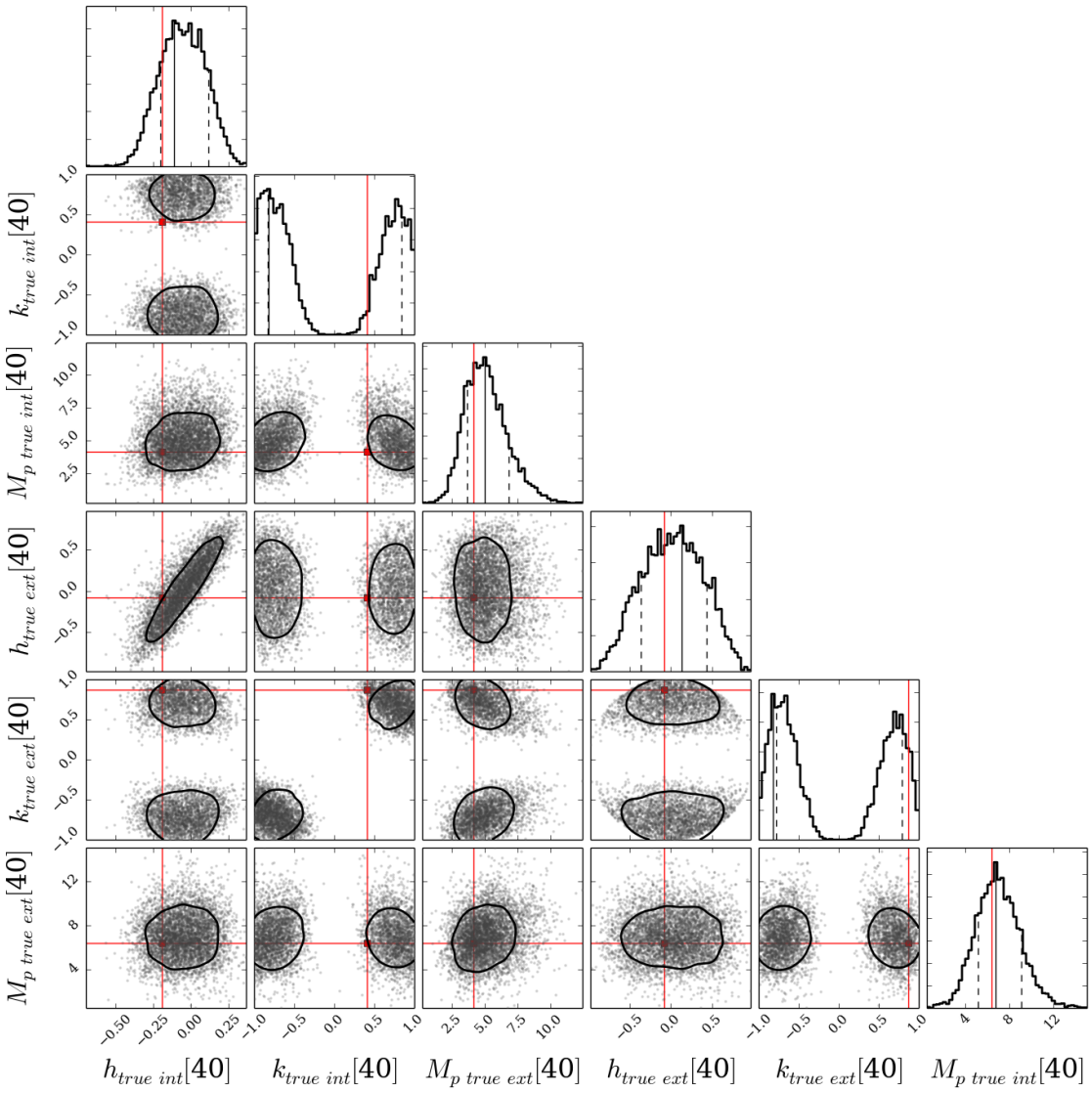

Figure 3.12: Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 40^{th} simulated system in the population. The 68% credible intervals are shown as the dotted black lines on the marginal posteriors. Posterior modes for the marginals are shown as the solid black lines. The 68% credible regions for the 2D posteriors are shown as the solid black curve. In this case we see a bimodal solution for k, for both the interior and exterior planet. Solutions where k is closer to zero produce larger mass planets, and cases where k deviates from zero produces smaller mass planets. In this case, the HB model is able to recover the true generated values well.

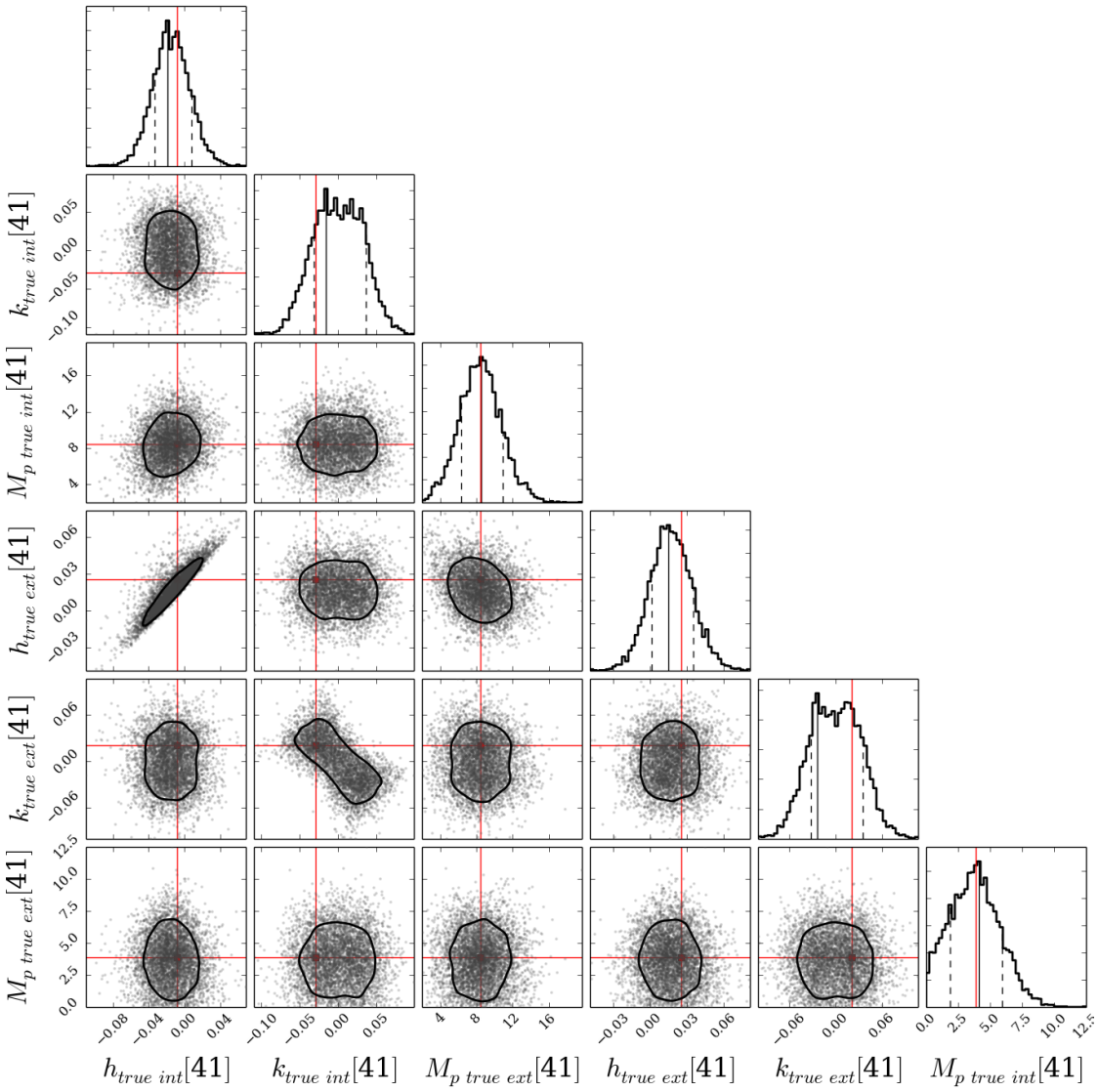

Figure 3.13: Planet mass vs. projected eccentricity (h and k) of interior and exterior planet pairs, using a mass-radius-eccentricity HB model with a two-component Gaussian mixture model for h and k. Here we show the 41^{st} simulated system in the population. The 68% credible intervals are shown as the dotted black lines on the marginal posteriors. Posterior modes for the marginals are shown as the solid black lines. The 68% credible regions for the 2D posteriors are shown as the solid black curve. In this case, k deviates a small amount from zero and you can see two modes start to break out. The masses and projected eccentricities are still well constrained in this case.

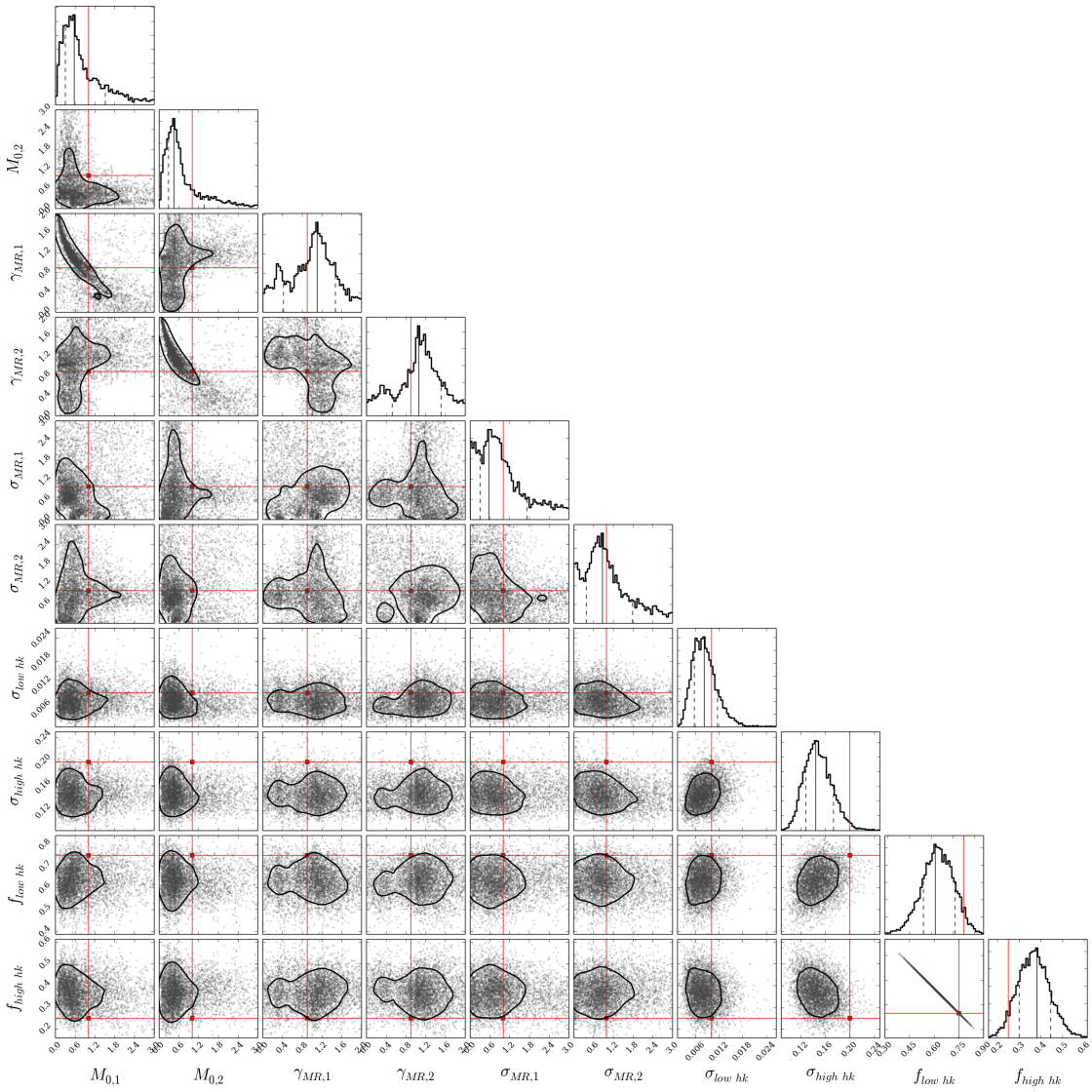

Figure 3.14: Hyperparameters for simulated *Kepler* data. Here, the hyperparameters for this HB model are $M_1 = M_2 = 1.0\,M_\oplus$, $\gamma_{MR_1} = \gamma_{MR_2} = 0.9$, $\sigma_{MR_1} = \sigma_{MR_2} = 1.0\,M_\oplus$, and the two-component Gaussian mixture model hyperparameters are now $\sigma_{low} = 0.03$, $\sigma_{high} = 0.65$, $f_{low} = 0.75$, and $f_{high} = 0.25$. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves.

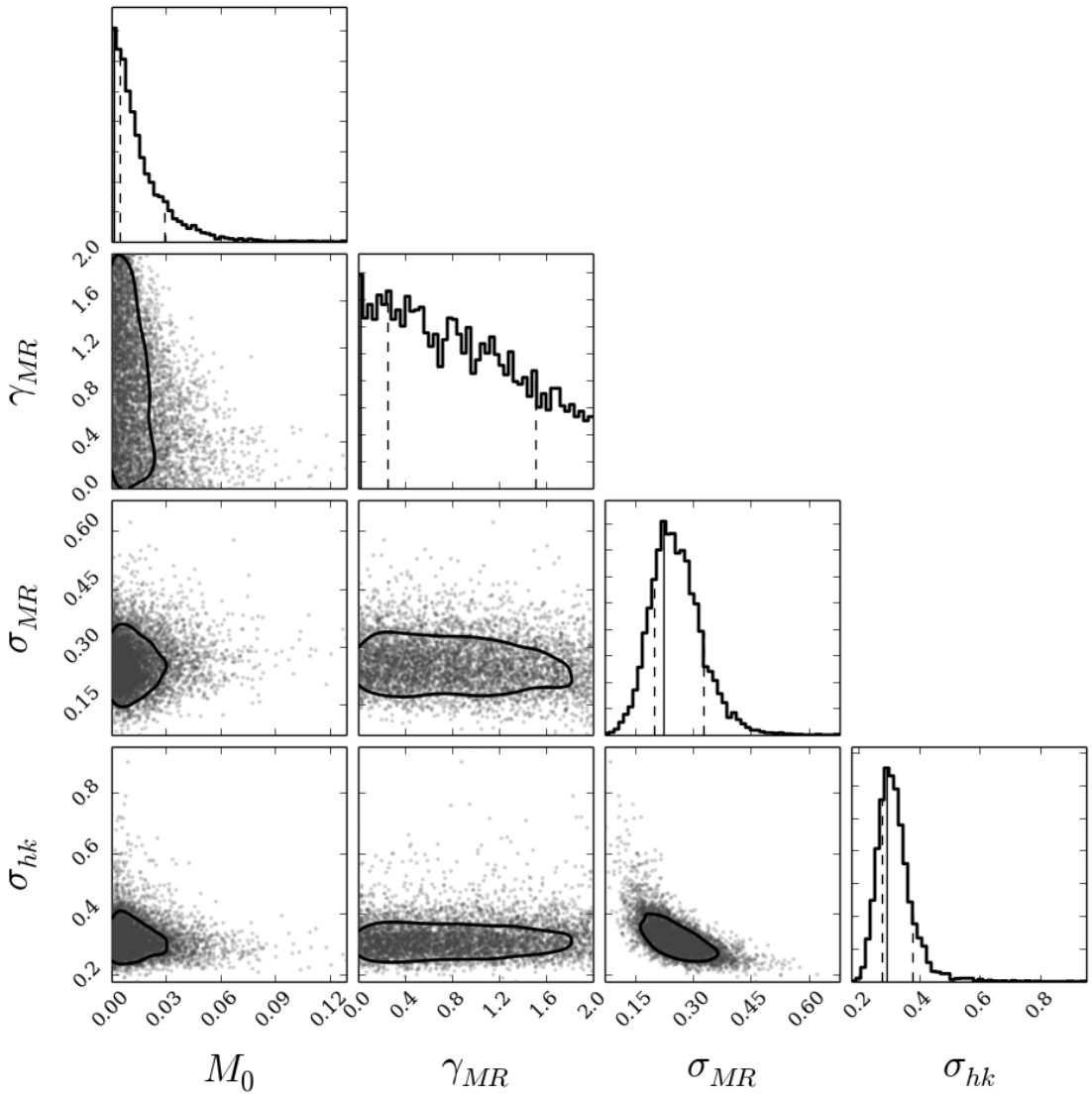

Figure 3.15: Hyperparameters for real *Kepler* data. Here we use a one-component Gaussian mixture model for the eccentricity distribution in our joint mass-radius-eccentricity HB model. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. The posterior modes and credible intervals for the hyperparameters are reported in Table 3.2.

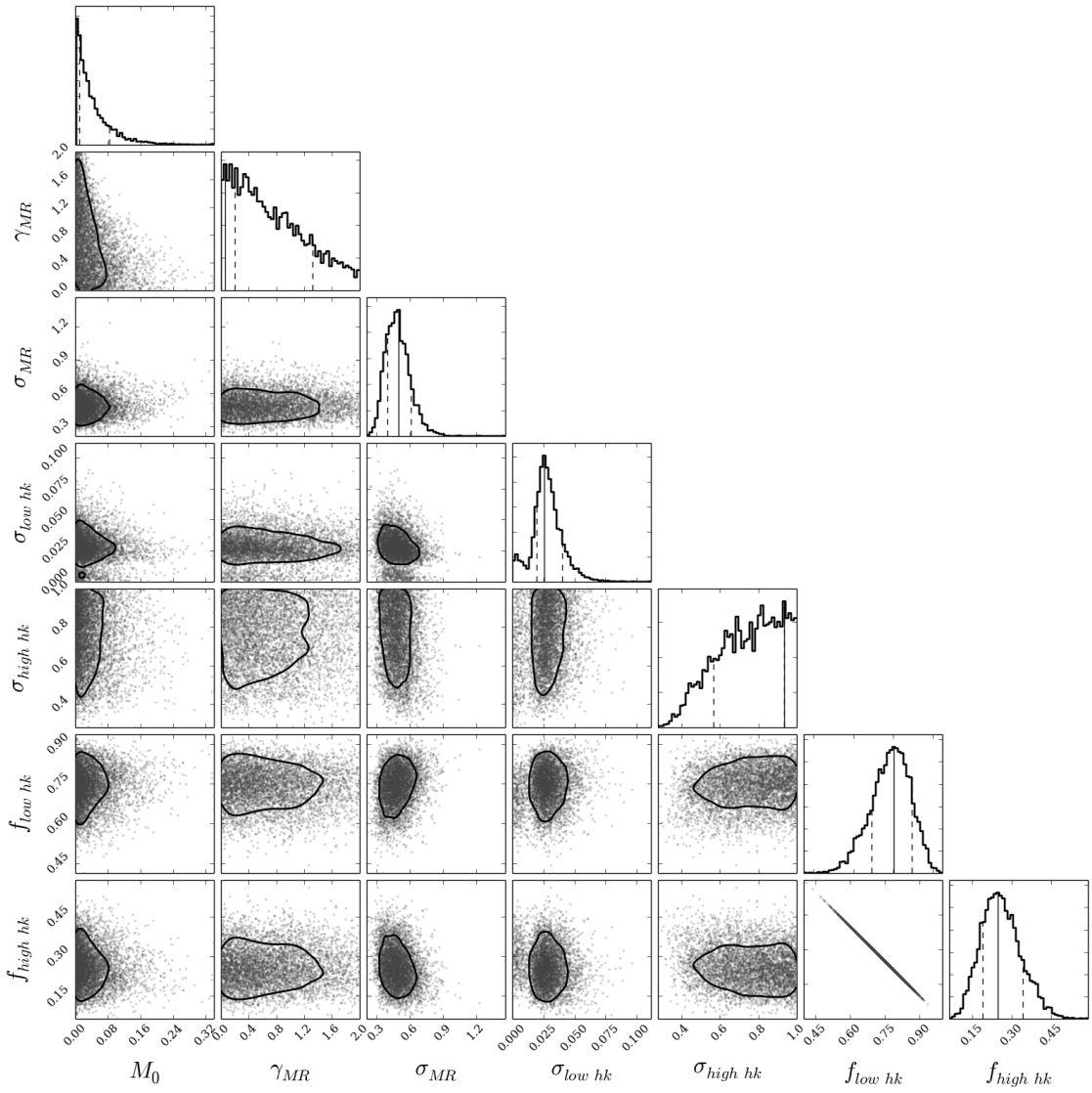

Figure 3.16: Hyperparameters for real *Kepler* data. Here we use a two-component Gaussian mixture model for the eccentricity distribution in our joint mass-radius-eccentricity HB model. The 68% credible interval is shown as black dashed lines on the marginal posteriors for these hyperparameters. Posterior modes are shown as black solid lines, and the 68% credible interval on the 2-D posteriors are show as black solid curves. The posterior modes and credible intervals for the hyperparameters are reported in Table 3.2.

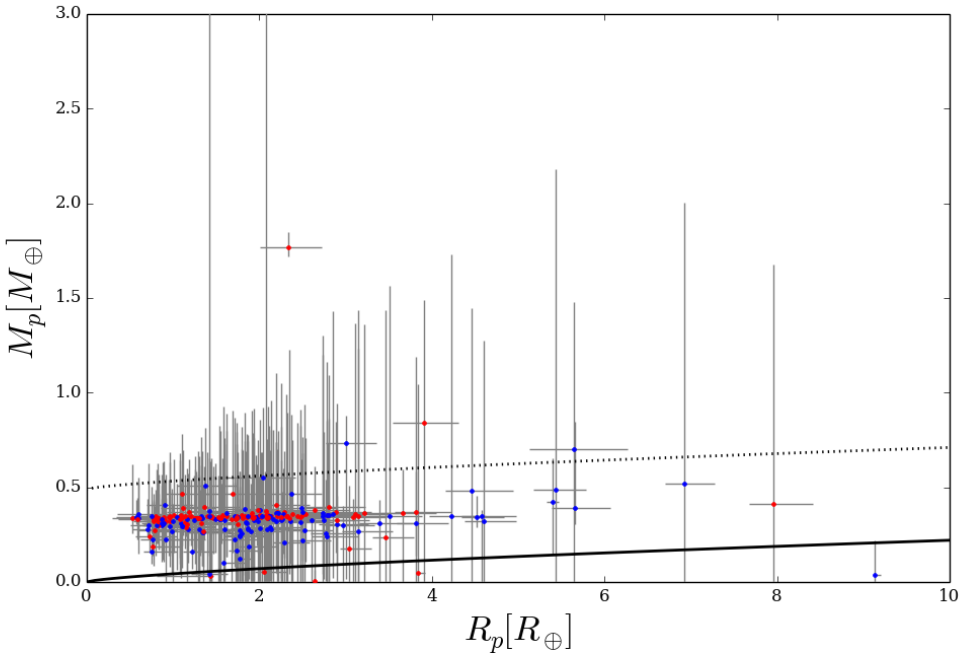

Figure 3.17: Planet mass vs. planet radius. Shown are the means from the latent variable posteriors for planet mass and planet radius when using a mass-radius-eccentricity HB model that parameterizes the projected eccentricity as a two-component Gaussian mixture model. The 68% credible intervals of the marginal posteriors for mass and radius are show as gray bars. The mass-radius relation using the posterior means for the mass constant M_o and the exponent γ_{MR} is shown as the solid back curve. The dotted black curve represents the posterior mean for the scatter in planet mass for a given radius σ_{MR}.

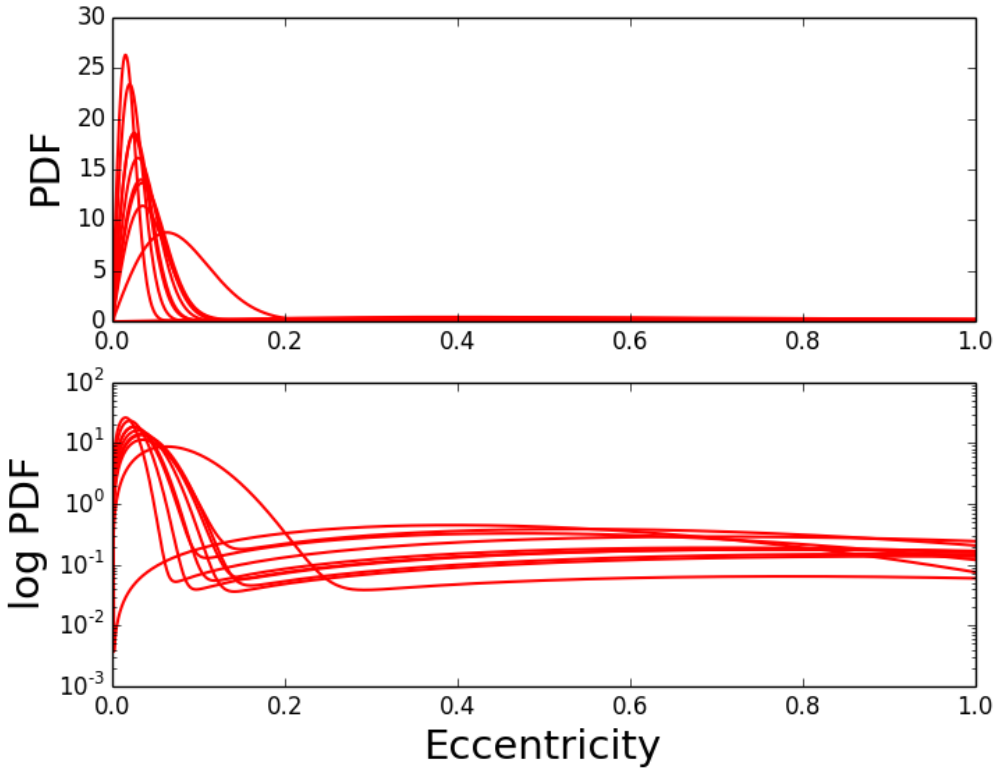

Figure 3.18: 10 draws of the eccentricity PDF from the joint posterior, from a mass-radius-eccentricity HB model that uses a two-component Gaussian mixture model to parameterize the eccentricity distribution. The top panel shows a linear scale and the bottom panel shows a log scale in order to see the larger dispersion component more clearly.

Table 3.4: Transit Timing Variation (TTV) amplitudes for 120 near-first-order mean motion resonance transiting exoplanet pairs.

KOI	$V_{obs\,Re\,int}$ (days)	$\sigma_{V_{obs\,Re\,int}}$	$V_{obs\,Re\,ext}$ (days)	$\sigma_{V_{obs\,Re\,ext}}$	$V_{obs\,Im\,int}$ (days)	$\sigma_{V_{obs\,Im\,int}}$	$V_{obs\,Im\,ext}$ (days)	$\sigma_{V_{obs\,Im\,ext}}$
153.0	-0.00066	0.00042	0.0004	0.0004	-0.00102	0.00042	-0.00063	0.00042
171.0	-0.00147	0.00054	-3e-05	0.0006	-0.00094	0.0017	-0.00151	0.00183
222.0	0.00147	0.00042	0.00048	0.00042	-0.00139	0.00129	-0.00344	0.00113
238.0	-0.00117	0.00126	-0.00156	0.00115	-0.00605	0.00525	-0.00278	0.00499
239.0	0.00033	0.00038	-0.00013	0.00044	nan	2.0	nan	2.0
244.0	0.002	0.00022	0.00178	0.0002	0.00097	0.00011	0.00064	8e-05
251.0	0.00013	0.00024	0.00023	0.00026	0.0	0.00224	0.00385	0.0023
255.0	0.00087	0.00607	0.0042	0.00494	0.00144	0.00088	-0.00111	0.00072
266.0	0.00314	0.00135	0.00205	0.00142	nan	2.0	nan	2.0
274.0	0.03533	0.00433	-0.0242	0.00356	0.03103	0.00525	-0.02084	0.00542
283.0	-0.00032	0.00036	0.00027	0.00037	-0.0005	0.00418	0.00071	0.00355
295.0	-0.00099	0.00062	-1e-05	0.00058	-0.00108	0.00191	0.00117	0.00186
301.0	0.00096	0.00096	-0.00096	0.00096	0.00159	0.00309	0.00334	0.00338
304.0	-0.00177	0.00517	0.00294	0.00545	-8e-05	0.00035	0.00042	0.00035
312.0	0.00185	0.00122	-0.0006	0.00115	0.00108	0.00108	-0.00239	0.0011
321.0	-0.0007	0.00057	-0.00035	0.00052	-0.0022	0.00208	-0.00516	0.00176
332.0	0.00309	0.00091	-0.00085	0.00102	0.00219	0.00596	-0.00915	0.00572
341.0	0.0007	0.00183	-0.00027	0.00158	-0.00086	0.00095	-0.00122	0.00075
354.0	-0.00444	0.00268	-0.00487	0.00226	-0.00032	0.00093	-1e-05	0.0008
370.0	0.00074	0.0025	-0.00678	0.00288	0.00213	0.00119	-0.0034	0.00147

413.0	-0.00158	0.00083	0.00048	0.00072	0.00155	0.00194	0.00256	0.00193
430.0	-0.00246	0.00398	0.0053	0.00373	0.0004	0.0007	0.001	0.00062
457.0	0.00052	0.00046	-0.00066	0.00047	0.00052	0.00058	-0.0015	0.00073
508.0	-0.00147	0.00073	0.00078	0.00072	-0.00288	0.00115	0.00181	0.00127
523.0	-0.01056	0.01107	0.01718	0.00448	-0.00647	0.00415	0.00897	0.0014
543.0	0.00111	0.00133	0.0032	0.00141	-0.00139	0.00068	-0.00057	0.0006
546.0	-0.00188	0.00203	0.00032	0.00195	0.0003	0.00155	-0.00098	0.00132
551.0	7e-05	0.00115	-0.00279	0.00143	-0.00043	0.00147	-0.00151	0.00186
568.0	-0.00486	0.00165	0.00081	0.00185	0.00019	0.0008	-2e-05	0.00085
572.0	-0.00289	0.00329	-0.0001	0.00314	0.00078	0.00167	-0.00249	0.00135
574.0	0.00291	0.00254	-0.00053	0.00236	0.0016	0.00099	-0.00273	0.00108
579.0	0.00116	0.00067	-0.00076	0.00067	0.00038	0.00072	-0.00064	0.0007
605.0	-0.00014	0.00039	-0.00016	0.00038	0.00126	0.00317	0.0008	0.00336
626.0	-0.00158	0.00432	-0.00224	0.00456	-0.00118	0.0013	-0.00136	0.00143
627.0	0.00202	0.00157	0.0021	0.00162	0.0017	0.00072	0.00038	0.00074
647.0	-0.00098	0.00123	-0.00029	0.00118	0.01087	0.0082	-0.001	0.0066
678.0	-0.00032	0.00076	0.00071	0.0007	-0.00045	0.00092	0.00111	0.00103
679.0	0.00194	0.0067	0.00796	0.00684	0.00233	0.00196	0.002	0.00196
691.0	0.0039	0.00448	0.00496	0.00462	0.0002	0.00127	0.00036	0.00113
837.0	-0.00103	0.00185	0.00097	0.0018	0.00227	0.00129	-0.00038	0.00118
870.0	-0.00052	0.00064	0.00553	0.00066	-0.00449	0.00105	0.00862	0.00087
912.0	-0.0024	0.00142	0.00206	0.00144	-0.00096	0.00058	-0.00049	0.00053
992.0	nan	2.0	nan	2.0	0.00254	0.0033	0.006	0.00286

1106.0	-0.00252	0.00157	0.0004	0.00174	-0.00763	0.0053	0.00495	0.00524
1165.0	-0.00369	0.0035	-0.00763	0.00392	0.00084	0.00064	0.00136	0.00067
1175.0	0.00439	0.00709	-0.00963	0.00701	0.00084	0.00562	0.00265	0.00588
1215.0	0.00716	0.00205	-0.00923	0.00212	0.00651	0.00424	-0.00931	0.00295
1241.0	0.08674	0.00908	0.03683	0.008	0.03337	0.0037	-0.0014	0.00382
1270.0	-0.00347	0.00051	-0.00073	0.00057	-0.02526	0.00155	-0.00278	0.00165
1279.0	0.00158	0.0038	-0.00037	0.00339	0.00163	0.00139	-0.00044	0.00123
1316.0	0.00056	0.0036	0.00066	0.00355	nan	2.0	nan	2.0
1363.0	0.00061	0.00157	-0.00132	0.00144	-5e-05	0.0023	-0.00166	0.00243
1404.0	0.00125	0.00233	0.00408	0.00262	-0.0058	0.00473	-0.00707	0.00491
1529.0	-0.01723	0.00568	0.005	0.00495	-0.04657	0.00387	0.00718	0.00316
1593.0	-0.00026	0.00241	-0.0031	0.00239	0.00958	0.00398	-0.00422	0.00347
1760.0	0.00136	0.00203	0.00029	0.0021	-0.00477	0.00378	7e-05	0.00387
1783.0	0.02314	0.00788	0.00366	0.00172	nan	2.0	nan	2.0
1824.0	1e-05	0.00064	-0.00046	0.0006	-0.00083	0.00079	-0.00178	0.00074
1843.0	-0.00018	0.00064	0.00138	0.00067	-0.00263	0.00257	0.00378	0.00282
1873.0	-0.00519	0.01627	-0.03515	0.00899	0.01545	0.01369	-0.01025	0.00905
1889.0	-0.00044	0.00413	-0.00247	0.00439	-0.00377	0.00195	-0.00617	0.00219
1891.0	-0.00164	0.00263	-0.00161	0.00265	-0.00232	0.00185	0.00193	0.00174
1899.0	-0.00061	0.00685	0.00261	0.00724	0.00207	0.00197	-0.0036	0.00181
1908.0	-0.001	0.00176	-0.00014	0.00162	-0.01133	0.00315	-0.00267	0.0029
1932.0	-0.00133	0.00752	-0.00056	0.00754	-0.00404	0.00554	-0.01507	0.00583
1940.0	nan	2.0	nan	2.0	0.00145	0.00084	-0.00176	0.00089

1977.0	-0.00531	0.00246	-0.00309	0.00225	-0.00454	0.00085	-0.00157	0.001
1992.0	-0.00032	0.00272	0.00075	0.00252	0.02145	0.00743	0.00612	0.00596
1996.0	nan	2.0	nan	2.0	0.0016	0.00165	-0.00084	0.0015
2022.0	-0.0003	0.00107	-0.00075	0.00122	0.00132	0.00161	-4e-05	0.00169
2034.0	-0.00085	0.00268	-0.00307	0.00278	0.00086	0.00069	0.00083	0.00069
2036.0	0.00453	0.00309	0.00296	0.00304	-0.0002	0.00174	0.00115	0.00161
2111.0	0.00188	0.00158	-0.00145	0.00162	0.00249	0.00116	1e-05	0.00131
2113.0	-0.00152	0.00168	0.00106	0.00166	0.00289	0.00183	-0.00062	0.00158
2158.0	-4e-05	0.00215	-0.00198	0.00204	-1e-05	0.00445	-0.00166	0.00449
2162.0	0.00147	0.00407	0.00817	0.00466	-0.02154	0.01172	-0.00347	0.00897
2173.0	0.00111	0.00202	-0.00445	0.00253	-0.00441	0.00323	-0.00039	0.00345
2209.0	0.00375	0.00459	0.00044	0.00411	nan	2.0	nan	2.0
2236.0	0.00146	0.00376	-0.00132	0.00346	-0.00627	0.0038	0.00156	0.00421
2243.0	0.00226	0.00222	-0.00222	0.0023	-0.00537	0.00287	0.00415	0.00301
2257.0	-0.00136	0.00343	0.00988	0.00346	-0.00132	0.00612	0.00423	0.00593
2279.0	0.00509	0.00371	-0.00123	0.00426	-0.00244	0.0035	-0.00238	0.00286
2303.0	-0.00083	0.00216	-1e-05	0.00246	0.00493	0.00283	-0.00194	0.00269
2333.0	-0.00343	0.00244	0.00088	0.00248	-0.00455	0.00254	0.00331	0.00247
2414.0	0.6685	0.51724	-0.12853	0.05484	-0.05907	0.39819	-0.09264	0.05923
2420.0	-0.00444	0.00455	-0.00985	0.00478	0.00016	0.00316	-0.00301	0.00314
2422.0	0.00507	0.00378	-0.0066	0.00325	0.00259	0.00613	0.00318	0.00647
2442.0	-0.00061	0.00415	0.00562	0.00481	-0.0016	0.00358	-0.00024	0.00425
2458.0	0.00126	0.00158	-0.0024	0.00159	0.00403	0.00229	-0.00085	0.00226

2498.0	-0.0002	0.00312	0.00033	0.00246	0.00519	0.00456	-0.00518	0.00448
2595.0	0.00052	0.00284	-0.0053	0.00369	-0.00053	0.00456	0.00069	0.00517
2612.0	0.00471	0.00383	-0.00599	0.00352	-0.00099	0.00377	-0.00146	0.00422
2672.0	0.00855	0.00178	0.01571	0.00187	0.01296	0.00748	0.01716	0.00314
2696.0	-0.00337	0.00371	-0.01496	0.00319	-0.0033	0.00286	-0.00024	0.00357
2704.0	0.00176	0.0012	4e-05	0.00114	-0.00068	0.00073	-0.00103	0.0007
2711.0	0.00298	0.00156	-0.00111	0.00136	0.00446	0.003	0.00244	0.00297
2739.0	-0.00084	0.00591	-0.0086	0.00663	-0.00291	0.00251	-0.0032	0.00233
2768.0	-0.00034	0.00532	-0.00108	0.00428	-0.00615	0.00703	-0.00298	0.00539
2840.0	-0.0044	0.00272	0.00104	0.00344	0.00973	0.00563	0.00534	0.0052
2977.0	-0.00162	0.00293	0.00544	0.00284	0.0038	0.00308	0.00141	0.00363
3022.0	0.00207	0.00252	-0.0012	0.00241	-0.00287	0.00381	-0.00028	0.0036
3052.0	0.00012	0.00422	0.00182	0.00389	-0.00354	0.00592	-0.0083	0.00427
3077.0	-0.00051	0.00243	-0.00368	0.00247	0.00213	0.00501	0.00951	0.00513
3196.0	0.00043	0.00317	0.00606	0.0033	-0.00013	0.00287	0.00251	0.00294
3214.0	-0.00351	0.00466	0.00078	0.00427	nan	2.0	nan	2.0
3340.0	-0.00711	0.0057	0.0029	0.00551	0.00074	0.00464	-0.00531	0.0049
3384.0	-0.0027	0.00241	0.00094	0.00271	0.00071	0.00304	-0.00531	0.00298
3503.0	-0.23747	1.08911	-0.39543	0.1281	nan	2.0	nan	2.0
4021.0	-0.0011	0.0032	-0.00234	0.00341	-0.00088	0.00231	0.00221	0.00215
4097.0	0.00094	0.00069	0.00081	0.00073	-0.00208	0.00231	0.00082	0.00228
4149.0	0.00782	0.00582	-0.00764	0.0053	-0.00224	0.0058	-0.00066	0.00638
4246.0	0.00365	0.00279	-0.00301	0.00277	nan	2.0	nan	2.0

4287.0	nan	2.0	nan	2.0	-0.003	0.00308	0.00636	0.00298
4288.0	0.00568	0.00394	-0.00711	0.0036	-0.00496	0.00441	-0.00167	0.00488
4382.0	6e-05	0.00366	-0.0018	0.0034	-0.00084	0.00549	-0.00261	0.00563
4435.0	-0.00194	0.00846	-0.00729	0.0085	nan	2.0	nan	2.0
4504.0	0.00144	0.00314	0.00172	0.00305	-0.00397	0.00463	0.00414	0.00482
4541.0	-0.00908	0.00375	0.00409	0.0059	nan	2.0	nan	2.0
4657.0	0.00827	0.00439	0.00269	0.00451	nan	2.0	nan	2.0
4913.0	0.00153	0.00374	-0.00107	0.00352	nan	2.0	nan	2.0

Table 3.5: Planet-to-star radius ratio and orbital period for 120 near-first-order mean motion resonance transiting exoplanet pairs.

KOI	$R_p/R_{\star obs\,int}$	$\sigma_{+\,R_p/R_{\star obs\,int}}$	$\sigma_{-\,R_p/R_{\star obs\,int}}$	P_{int}(days)	$R_p/R_{\star obs\,ext}$	$\sigma_{+\,R_p/R_{\star obs\,ext}}$	$\sigma_{-\,R_p/R_{\star obs\,ext}}$	P_{ext}(days)
153.0	0.0246	0.00052	-0.00039	4.75	0.03275	1e-05	-0.00389	8.93
171.0	0.02172	0.00148	-0.00046	5.97	0.01443	0.0005	-0.00023	13.07
222.0	0.03316	0.00068	-0.00033	6.31	0.02639	0.00079	-0.00047	12.79
238.0	0.02101	0.00049	-0.00033	17.23	0.01095	0.00056	-0.00054	26.69
239.0	0.0134	0.00068	-0.00271	3.62	0.03808	0.00028	-0.00191	5.64
244.0	0.01844	0.00011	-4e-05	6.24	0.03156	0.00027	-0.0001	12.72
251.0	0.04766	0.00028	-0.00163	4.16	0.01496	0.00094	-0.00051	5.77
255.0	0.01219	0.00093	-0.00112	13.6	0.04501	0.00069	-0.00044	27.52
266.0	0.01221	1e-05	-0.00077	25.31	0.00941	0.0002	-0.00057	47.74
274.0	0.00789	0.00029	-0.0002	15.09	0.00785	0.00142	-0.00041	22.8
283.0	0.01891	0.00031	-0.0001	16.09	0.00738	0.00023	-0.00032	25.52
295.0	0.01542	0.00032	-0.00016	5.32	0.00975	0.00032	-0.00035	10.11
301.0	0.01304	0.00023	-0.0002	6.0	0.00807	0.00028	-0.00024	11.45
304.0	0.00527	2e-05	-0.00076	5.52	0.022	0.0003	-0.00011	8.51
312.0	0.01516	6e-05	-0.00089	11.58	0.01389	0.00034	-0.00023	16.4
321.0	0.01246	0.00027	-0.00017	2.43	0.00717	0.00034	-0.00017	4.62
332.0	0.01583	0.00087	-0.00108	5.46	0.00589	0.00042	-0.00037	6.87
341.0	0.0212	0.00207	-0.00159	4.7	0.02902	0.00205	-0.00146	7.17
354.0	0.0101	0.00033	-0.0004	7.38	0.02126	0.00037	-0.00022	15.96
370.0	0.01088	0.00021	-0.00036	22.95	0.01898	0.00084	-0.00109	42.88

413.0	0.03334	6e-05	-0.00214	15.23	0.02341	0.00087	-0.00048	24.67
430.0	0.01358	0.00098	-0.00054	9.34	0.04025	0.00331	-0.00188	12.38
457.0	0.02559	0.00024	-0.00066	4.92	0.02423	0.00102	-0.00027	7.06
508.0	0.02718	0.00253	-0.00102	7.93	0.02519	0.00076	-0.0003	16.67
523.0	0.0311	0.00043	-0.00436	36.85	0.06424	0.00039	-0.00101	49.41
543.0	0.01602	0.00054	-0.00055	3.14	0.02504	0.0006	-0.00038	4.3
546.0	0.0218	0.00164	-0.00183	9.83	0.02948	5e-05	-0.00145	20.68
551.0	0.02129	0.00069	-0.00045	5.69	0.02318	0.0008	-0.00042	11.64
568.0	0.00896	0.00062	-0.00024	2.36	0.0149	0.0006	-0.00028	3.38
572.0	0.0105	0.00035	-0.00041	4.94	0.02153	0.0	-0.00184	10.64
574.0	0.01616	0.00077	-0.00045	10.4	0.03406	0.00326	-0.00247	20.13
579.0	0.01663	0.0005	-0.00023	2.02	0.01705	0.00064	-0.00044	3.76
605.0	0.03107	0.00012	-0.00234	2.63	0.0115	0.00035	-0.00088	5.07
626.0	0.00777	0.00037	-0.00057	8.03	0.01848	0.00019	-0.00064	14.59
627.0	0.01012	0.00035	-0.00022	4.17	0.01852	0.00032	-0.0003	7.75
647.0	0.01354	0.00027	-0.00021	5.17	0.00625	0.00029	-0.00052	8.11
678.0	0.01463	0.00056	-0.00021	4.14	0.01484	0.00053	-0.00035	6.04
679.0	0.00702	0.00042	-0.00025	16.26	0.01652	0.00024	-0.00021	31.81
691.0	0.01001	0.00037	-0.00039	16.23	0.0235	0.00036	-0.00021	29.67
837.0	0.01841	0.00074	-0.00067	4.14	0.0275	0.00117	-0.00054	7.95
870.0	0.03077	4e-05	-0.00218	5.91	0.03111	6e-05	-0.00404	8.99
912.0	0.01857	0.0006	-0.00086	6.67	0.03694	0.00098	-0.00047	10.85
992.0	0.01112	0.00073	-0.00066	4.58	0.01922	0.00072	-0.0005	9.93

1106.0	0.02132	0.00243	-0.00112	7.43	0.01345	0.00074	-0.00075	15.98
1165.0	0.00764	0.0001	-0.00075	4.29	0.02097	0.00046	-0.00028	7.05
1175.0	0.01042	0.00197	-0.0009	17.16	0.01147	0.00032	-0.0007	31.59
1215.0	0.016	4e-05	-0.00164	17.32	0.0151	0.00049	-0.00024	33.01
1241.0	0.01074	0.00033	-0.00024	10.5	0.02456	0.00025	-0.00205	21.41
1270.0	0.03365	0.00014	-0.00496	5.73	0.52845	0.47114	-0.20455	11.61
1279.0	0.01043	0.00136	-0.00096	9.65	0.01689	0.0005	-0.0002	14.37
1316.0	0.00806	0.00032	-0.00141	7.65	nan	0.001	0.001	12.38
1363.0	0.02014	0.00078	-0.00053	3.55	0.01793	0.00113	-0.00066	5.7
1404.0	0.02757	0.00076	-0.00085	13.32	0.01912	0.0009	-0.00139	18.91
1529.0	0.00926	0.00055	-0.00045	11.87	0.01468	0.00031	-0.00094	17.98
1593.0	0.0223	0.00094	-0.00085	9.69	0.02198	0.00103	-0.00082	15.38
1760.0	0.02068	0.00095	-0.0006	5.52	0.01812	0.00074	-0.0011	8.78
1783.0	0.07175	0.00123	-0.00094	134.48	0.04208	0.00045	-0.00396	284.04
1824.0	0.0113	0.00023	-0.00014	1.68	0.01285	0.00026	-0.00022	3.55
1843.0	0.02939	0.00024	-0.00521	4.19	0.01285	0.0006	-0.00144	6.36
1873.0	0.02866	6e-05	-0.00654	34.92	0.04457	0.00107	-0.00063	71.31
1889.0	0.01668	0.0009	-0.00058	9.18	0.03346	0.0004	-0.00452	14.31
1891.0	0.01612	0.00099	-0.00041	8.26	0.02725	0.00375	-0.00104	15.96
1899.0	0.00739	0.00066	-0.00059	10.52	0.01914	0.00248	-0.00081	19.76
1908.0	0.02297	0.00028	-0.0023	12.55	0.01879	0.00493	-0.00136	24.09
1932.0	0.01329	0.00027	-0.00125	14.84	0.01641	0.00077	-0.00064	22.82
1940.0	nan	0.001	0.001	6.74	0.0316	0.0001	-0.00347	10.99

1977.0	0.01142	0.00045	-0.0007	7.42	0.01772	0.00109	-0.00029	9.39
1992.0	0.01825	1e-05	-0.00151	12.8	0.01507	0.00031	-0.0011	27.32
1996.0	0.01487	0.00098	-0.00082	7.07	0.02823	0.0008	-0.00071	10.13
2022.0	0.01897	0.00054	-0.00042	5.93	0.01978	0.00074	-0.00034	12.25
2034.0	0.01645	0.00075	-0.00101	2.37	0.03663	0.0006	-0.001	3.61
2036.0	0.01886	0.00538	-0.00237	5.8	0.03088	0.00019	-0.00499	8.41
2111.0	0.0144	0.0005	-0.00036	3.29	0.01877	0.00101	-0.00032	7.19
2113.0	0.02825	0.00108	-0.00059	12.33	0.03067	0.00124	-0.00046	15.94
2158.0	0.00837	0.00031	-0.00031	4.56	0.00636	0.00053	-0.00022	6.68
2162.0	0.01719	0.0007	-0.00038	108.59	0.01672	0.00012	-0.00124	199.67
2173.0	0.01471	0.00056	-0.00043	37.82	0.01736	0.00242	-0.00254	53.58
2209.0	0.01398	0.00201	-0.00114	18.3	0.00974	0.0004	-0.00096	35.5
2236.0	0.01709	0.0007	-0.00106	12.13	0.03189	0.00038	-0.0072	19.99
2243.0	0.01496	0.00051	-0.0005	5.19	0.01493	0.00074	-0.00044	8.46
2257.0	0.02433	0.005	-0.00183	32.56	0.01962	0.00108	-0.00111	59.28
2279.0	0.01164	0.00039	-0.00075	12.51	0.01581	0.00052	-0.00042	27.02
2303.0	0.01088	0.00041	-0.00037	4.89	0.01091	0.00035	-0.00077	8.93
2333.0	0.00897	0.00044	-0.00024	3.93	0.0104	0.00043	-0.00042	7.63
2414.0	0.01124	0.0005	-0.00028	22.6	0.01212	0.0006	-0.00042	45.35
2420.0	0.00871	0.00058	-0.00061	5.47	0.01284	0.00066	-0.00038	10.42
2422.0	0.01967	0.00044	-0.00111	26.78	0.01817	0.00128	-0.00103	57.39
2442.0	0.01553	0.00094	-0.00085	12.31	0.02186	0.00079	-0.00085	25.19
2458.0	0.01649	0.00092	-0.00036	2.94	0.01571	0.00069	-0.0006	4.35

2498.0	0.00928	0.00052	-0.0004	6.74	0.0096	0.00034	-0.00083	13.06
2595.0	0.00864	0.00023	-0.00036	9.18	0.008	0.00042	-0.00022	14.61
2612.0	0.00549	0.00143	-0.00063	4.61	0.00539	0.00187	-0.00051	7.57
2672.0	0.03017	0.00058	-0.00028	42.99	0.05242	0.00061	-0.00059	88.52
2696.0	0.01864	0.0	-0.00141	44.56	0.03058	0.00182	-0.00211	96.46
2704.0	0.06849	0.01824	-0.0041	2.98	0.10114	0.00213	-0.00294	4.87
2711.0	0.01347	0.00034	-0.00033	9.02	0.01199	0.00045	-0.00038	17.34
2739.0	0.00961	0.00046	-0.00081	7.1	0.01644	0.00052	-0.00051	10.11
2768.0	0.01261	0.00099	-0.00087	11.83	0.0174	0.00543	-0.00411	14.75
2840.0	0.00852	0.0005	-0.00022	3.68	0.00803	0.00036	-0.00054	7.45
2977.0	0.00719	0.00012	-0.00046	2.79	0.00787	0.00159	-0.00064	4.14
3022.0	0.0128	0.00084	-0.00055	2.76	0.01251	0.00073	-0.00083	5.05
3052.0	0.00895	0.00039	-0.00052	10.13	0.00883	0.00054	-0.00047	15.61
3077.0	0.02173	0.00749	-0.00151	4.1	0.01933	0.00119	-0.00142	7.56
3196.0	0.00424	0.00024	-0.00021	4.96	0.00535	0.00024	-0.00036	6.88
3214.0	0.00759	0.0	-0.00118	11.49	0.00592	0.0003	-0.00029	25.09
3340.0	0.0083	0.00034	-0.00056	8.95	0.00966	0.00037	-0.00054	13.73
3384.0	0.00965	0.00046	-0.00075	10.55	0.01196	0.00176	-0.00089	19.92
3503.0	0.00833	0.00054	-0.00094	21.19	0.00865	0.002	-0.00055	31.82
4021.0	0.00769	0.00028	-0.00045	4.93	0.00966	0.00034	-0.0004	7.24
4097.0	0.0136	0.00125	-0.00042	2.79	0.0088	0.00044	-0.00063	4.45
4149.0	0.01168	0.00215	-0.0007	9.55	0.01182	0.00049	-0.00059	14.71
4246.0	0.00852	0.00055	-0.00056	6.98	0.01174	0.00011	-0.00341	8.76

4287.0	0.00458	0.00027	-0.00027	9.62	0.0059	0.00026	-0.00038	13.12
4288.0	0.00567	0.00027	-0.00022	6.28	0.00561	0.00038	-0.00022	9.09
4382.0	0.0071	0.00036	-0.00038	2.95	0.00619	0.00049	-0.00035	3.98
4435.0	0.00763	0.00034	-0.00039	10.86	0.00735	0.00028	-0.00064	17.87
4504.0	0.01664	0.00139	-0.00036	4.13	0.01706	0.00097	-0.00097	8.12
4541.0	0.0179	0.00063	-0.0013	8.53	0.01659	0.00106	-0.00126	17.24
4657.0	0.00656	0.00028	-0.00049	7.58	0.00705	0.00042	-0.00067	10.43
4913.0	0.0085	0.00067	-0.0005	6.79	0.00969	0.00064	-0.00078	8.97

Table 3.6: Host star mass and radius for 120 near-first-order mean motion resonance transiting exoplanet pairs.

KOI	$R_{\star\,obs}(R_\odot)$	$\sigma_{+\,R_{\star\,obs}}$)	$\sigma_{-\,R_{\star\,obs}}$	$M_{\star\,obs}(M_\odot)$	$\sigma_{+\,M_{\star\,obs}}$	$\sigma_{-\,M_{\star\,obs}}$
153.0	0.709	0.028	-0.034	0.696	0.04	-0.028
171.0	1.173	0.768	-0.136	1.195	0.281	-0.176
222.0	0.57	0.09	-0.09	0.59	0.09	-0.09
238.0	1.092	0.184	-0.099	1.012	0.063	-0.058
239.0	0.989	0.399	-0.095	1.056	0.183	-0.134
244.0	1.309	0.023	-0.023	1.187	0.06	-0.06
251.0	0.52	0.06	-0.06	0.54	0.06	-0.06
255.0	0.51	0.06	-0.06	0.53	0.06	-0.06
266.0	1.768	0.783	-0.607	1.269	0.269	-0.262
274.0	1.659	0.038	-0.038	1.184	0.074	-0.074
283.0	0.958	0.163	-0.043	1.071	0.049	-0.083
295.0	1.533	0.796	-0.416	1.263	0.235	-0.247
301.0	1.209	0.891	-0.159	1.187	0.297	-0.17
304.0	1.825	0.72	-0.671	1.173	0.299	-0.201
312.0	1.188	0.233	-0.113	1.168	0.092	-0.109
321.0	0.973	0.174	-0.047	1.058	0.064	-0.072
332.0	1.259	0.285	-0.186	1.126	0.101	-0.107
341.0	0.946	0.171	-0.078	0.892	0.063	-0.042
354.0	0.965	0.147	-0.045	1.075	0.041	-0.081
370.0	1.85	0.05	-0.05	1.319	0.101	-0.101
413.0	0.818	0.32	-0.066	0.874	0.093	-0.082
430.0	0.48	0.06	-0.06	0.51	0.06	-0.06
457.0	0.739	0.12	-0.054	0.817	0.078	-0.089
508.0	0.99	0.143	-0.049	1.081	0.048	-0.08
523.0	1.066	0.557	-0.108	1.07	0.245	-0.119
543.0	0.761	0.193	-0.055	0.84	0.089	-0.09
546.0	1.145	0.689	-0.132	1.195	0.253	-0.188
551.0	0.944	0.357	-0.081	1.037	0.131	-0.128
568.0	0.82	0.244	-0.058	0.92	0.066	-0.09
572.0	1.099	0.36	-0.199	0.896	0.138	-0.08

574.0	0.755	0.163	-0.056	0.836	0.084	-0.093
579.0	0.785	0.194	-0.056	0.868	0.075	-0.086
605.0	0.56	0.09	-0.09	0.58	0.09	-0.09
626.0	1.014	0.16	-0.055	1.05	0.044	-0.063
627.0	1.072	0.25	-0.051	1.165	0.095	-0.094
647.0	1.089	0.464	-0.13	1.028	0.172	-0.114
678.0	1.56	1.028	-0.695	0.929	0.144	-0.088
679.0	1.141	0.282	-0.097	1.137	0.096	-0.092
691.0	1.053	0.185	-0.117	0.93	0.071	-0.065
837.0	0.766	0.087	-0.057	0.823	0.061	-0.077
870.0	0.57	0.09	-0.09	0.59	0.1	-0.1
912.0	0.59	0.09	-0.09	0.61	0.1	-0.1
992.0	0.917	0.353	-0.078	1.01	0.112	-0.126
1106.0	1.009	0.41	-0.096	1.07	0.197	-0.134
1165.0	1.114	0.378	-0.239	0.839	0.142	-0.054
1175.0	1.791	0.463	-0.455	1.062	0.177	-0.131
1215.0	1.692	0.42	-0.363	1.125	0.187	-0.112
1241.0	4.23	0.15	-0.15	1.32	0.13	-0.13
1270.0	0.791	0.179	-0.06	0.884	0.066	-0.098
1279.0	1.031	0.185	-0.101	0.95	0.066	-0.061
1316.0	2.024	0.319	-0.47	nan	0.5	0.5
1363.0	0.986	0.393	-0.089	1.093	0.189	-0.148
1404.0	0.45	0.09	-0.09	0.48	0.1	-0.1
1529.0	1.011	0.433	-0.095	1.066	0.199	-0.129
1593.0	0.914	0.379	-0.079	1.0	0.12	-0.125
1760.0	0.936	0.339	-0.082	0.978	0.079	-0.083
1783.0	0.957	0.338	-0.084	1.027	0.131	-0.118
1824.0	1.378	0.831	-0.304	1.123	0.306	-0.145
1843.0	0.45	0.08	-0.05	0.464	0.08	-0.05
1873.0	1.064	0.442	-0.106	1.149	0.216	-0.165
1889.0	0.852	0.366	-0.068	0.934	0.097	-0.098
1891.0	0.743	0.119	-0.055	0.825	0.072	-0.092
1899.0	1.175	0.617	-0.21	1.064	0.277	-0.158
1908.0	0.583	0.057	-0.053	0.589	0.058	-0.053

1932.0	3.177	0.704	-1.108	2.046	0.36	-0.397
1940.0	nan	1.4	1.4	0.573	0.06	-0.045
1977.0	0.63	0.051	-0.065	0.616	0.065	-0.052
1992.0	0.998	0.385	-0.094	1.074	0.183	-0.139
1996.0	0.561	0.041	-0.039	0.561	0.042	-0.025
2022.0	0.911	0.317	-0.074	1.001	0.1	-0.119
2034.0	0.868	0.373	-0.068	0.963	0.097	-0.1
2036.0	0.523	0.1	-0.05	0.541	0.1	-0.05
2111.0	1.028	0.427	-0.113	1.022	0.18	-0.123
2113.0	0.806	0.299	-0.06	0.897	0.08	-0.089
2158.0	1.46	0.788	-0.569	0.896	0.103	-0.081
2162.0	0.844	0.363	-0.059	0.935	0.099	-0.11
2173.0	0.718	0.042	-0.025	0.759	0.032	-0.042
2209.0	0.8	0.364	-0.068	0.822	0.11	-0.069
2236.0	0.848	0.289	-0.067	0.937	0.091	-0.106
2243.0	1.022	0.378	-0.093	1.09	0.194	-0.132
2257.0	0.834	0.263	-0.067	0.917	0.077	-0.096
2279.0	0.913	0.357	-0.087	0.933	0.107	-0.105
2303.0	1.319	0.999	-0.205	1.334	0.286	-0.253
2333.0	1.097	0.18	-0.113	0.987	0.066	-0.062
2414.0	0.91	0.189	-0.108	0.773	0.076	-0.029
2420.0	0.918	0.367	-0.077	1.007	0.116	-0.123
2422.0	0.985	0.355	-0.121	0.912	0.096	-0.062
2442.0	0.746	0.148	-0.062	0.77	0.116	-0.07
2458.0	0.909	0.452	-0.127	0.795	0.146	-0.06
2498.0	0.769	0.059	-0.036	0.793	0.048	-0.045
2595.0	1.604	0.883	-0.384	1.315	0.273	-0.253
2612.0	0.924	0.375	-0.119	0.824	0.129	-0.064
2672.0	1.06	0.389	-0.197	0.85	0.13	-0.061
2696.0	1.666	1.018	-0.355	1.411	0.303	-0.282
2704.0	0.189	0.125	-0.054	0.169	0.148	-0.049
2711.0	1.146	0.633	-0.15	1.07	0.252	-0.119
2739.0	0.96	0.416	-0.081	1.057	0.183	-0.13
2768.0	0.857	0.327	-0.064	0.952	0.081	-0.095

2840.0	0.848	0.386	-0.101	0.781	0.126	-0.053
2977.0	0.981	0.417	-0.092	1.015	0.162	-0.116
3022.0	0.901	0.38	-0.071	0.999	0.117	-0.116
3052.0	0.946	0.395	-0.132	0.821	0.135	-0.062
3077.0	0.742	0.068	-0.057	0.759	0.072	-0.063
3196.0	1.392	0.866	-0.293	1.21	0.293	-0.192
3214.0	3.132	1.202	-1.714	1.502	0.213	-0.479
3340.0	0.947	0.355	-0.104	0.925	0.114	-0.1
3384.0	1.09	0.198	-0.081	1.074	0.066	-0.066
3503.0	0.847	0.365	-0.057	0.918	0.093	-0.103
4021.0	1.937	0.975	-0.483	1.509	0.306	-0.312
4097.0	0.62	0.057	-0.057	0.643	0.053	-0.066
4149.0	1.259	0.907	-0.192	1.266	0.255	-0.221
4246.0	1.235	0.573	-0.229	1.046	0.236	-0.121
4287.0	1.517	0.807	-0.412	1.194	0.283	-0.202
4288.0	1.219	0.211	-0.161	1.07	0.079	-0.075
4382.0	1.173	0.56	-0.143	1.195	0.238	-0.182
4435.0	1.013	0.405	-0.092	1.069	0.189	-0.124
4504.0	0.803	0.337	-0.065	0.868	0.099	-0.085
4541.0	0.931	0.374	-0.091	0.958	0.123	-0.107
4657.0	0.773	0.245	-0.058	0.856	0.087	-0.094
4913.0	1.48	0.82	-0.344	1.277	0.269	-0.234

Chapter 4
The Period-Eccentricity Distribution for Eclipsing Binaries from *Kepler*

4.1 Introduction

In the era of *Kepler*, there has been an increase in the quality and quantity of photometric observations of eclipsing binary (EB) stars (Koch et al., 2010; Borucki et al., 2011c,b; Batalha et al., 2013; Burke et al., 2014). More than 2400 eclipsing binary stars now have measured light curves from *Kepler* (Prša et al., 2011; Slawson et al., 2011). This has provided high quality constraints for stellar astrophysics since EBs provide an opportunity to probe stellar astrophysics using minimum theoretical assumptions (Torres et al., 2010). The large quantity of uniform, high duty-cycle observations presents an opportunity to study EBs in a vigorous statistical manner (Prša et al., 2012). From this, we hope to learn about the distribution of orbital properties such as the eccentricity and period that inform our understanding of the dynamical evolution of EBs.

Stellar evolution is likely coupled with planet formation and dynamical evolution. For example, tidal circularization plays an important role in sculpting of eclipsing binaries and planetary systems orbital properties (Wang & Ford, 2011; Plavchan et al., 2014a; Shabram et al., 2015). The behavior of such systems share common dynamical processes, but operate in different mass and composition regimes, offering an opportunity for comparison. Thus, investigating properties of EBs statistically

can shed light onto star and planet formation.

The observational uncertainties result in uncertainty in the parameter estimates of EB eccentricities. As a first step in an encompassing characterization of the population of EBs, we work with physical parameters that have been explored in detail, where uncertainties are believed to be well-estimated. In addition to characterizing the eccentricities of individual EBs, the distribution of orbital eccentricities is particularly important for constraining planet formation. We use hierarchical Bayesian (HB) modeling to estimate population level parameters for the eccentricity distributions of the *Kepler* EBs. This method has been applied to and tested on similar domains, such as the *Kepler* short-period planet candidates that both transit and occult, presented in Chapter 2. Investigating the EB sample is a natural extension of these previous analyses. Previous analyses of the eccentricity distribution of EBs using the *Kepler* EBs have been limited by the lack of the incorporation of measurement uncertainties.

With a large sub-sample of 795 well characterized EBs, we can perform inference using more complex models for the eccentricity distribution than previous EB studies (e.g., Latham et al., 2002; Mayor et al., 2001), including potential dependences on additional parameters such as orbital period. This was not practical for analogous exoplanet studies since the sample size was much smaller ($\sim$ 50, Shabram et al. (2015)) . In particular, we investigate the eccentricity distribution of EBs, including a joint distribution for the orbital eccentricity and period. The resulting posterior distributions can shed light on tidal circularization with implications for both EB and planet formation. The end state of EBs after billions of years of tidal evolution provide a way to constrain the internal structure of stars and planets, as well as the physics of tidal dissipation.

In §4.2, we describe the measurements used in this study. In §4.3, we describe the HB model, and the priors selected for this study. In §4.4, we present the results of our HB analysis for the eccentricity distribution and period-eccentricity distribution for EBs. In §4.5, we discuss our conclusions.

4.2 Observations

The observational data we use come from the 3rd *Kepler* EB catalog (Kirk et al. in prep http://keplerebs.villanova.edu/) that analyzed light curves to identify eclipsing binaries using the EBAI-code described in Prša et al. (2008, 2011, 2012). Dr. Prša sent us outputs for tier 1 binaries, (i.e., the EBs that had small measurement uncertainties for eccentricity). Specifically, we have obtained a selection of estimates for $h = e\cos\omega$ and $k = e\sin\omega$, as well as orbital period, for 795 EBs from this catalog. Binaries with orbital periods of 5 to 10 days are included to explore circularization.

The h and k measurement for our sample of EBs used in this study are shown in Figure 4.1. The bottom left panel shows h vs. k, where the blue points represent each EB measurements. Measurement uncertainties are shown as solid black lines, but are very small and only visible in some cases. Histograms of the measurements are show in the top left and bottom right panels. In Figure 4.2, we show histograms of the orbital period measurements for the EB sample used in this study. The orbital periods of the EBs in our sample range from 0.26 to 623 days.

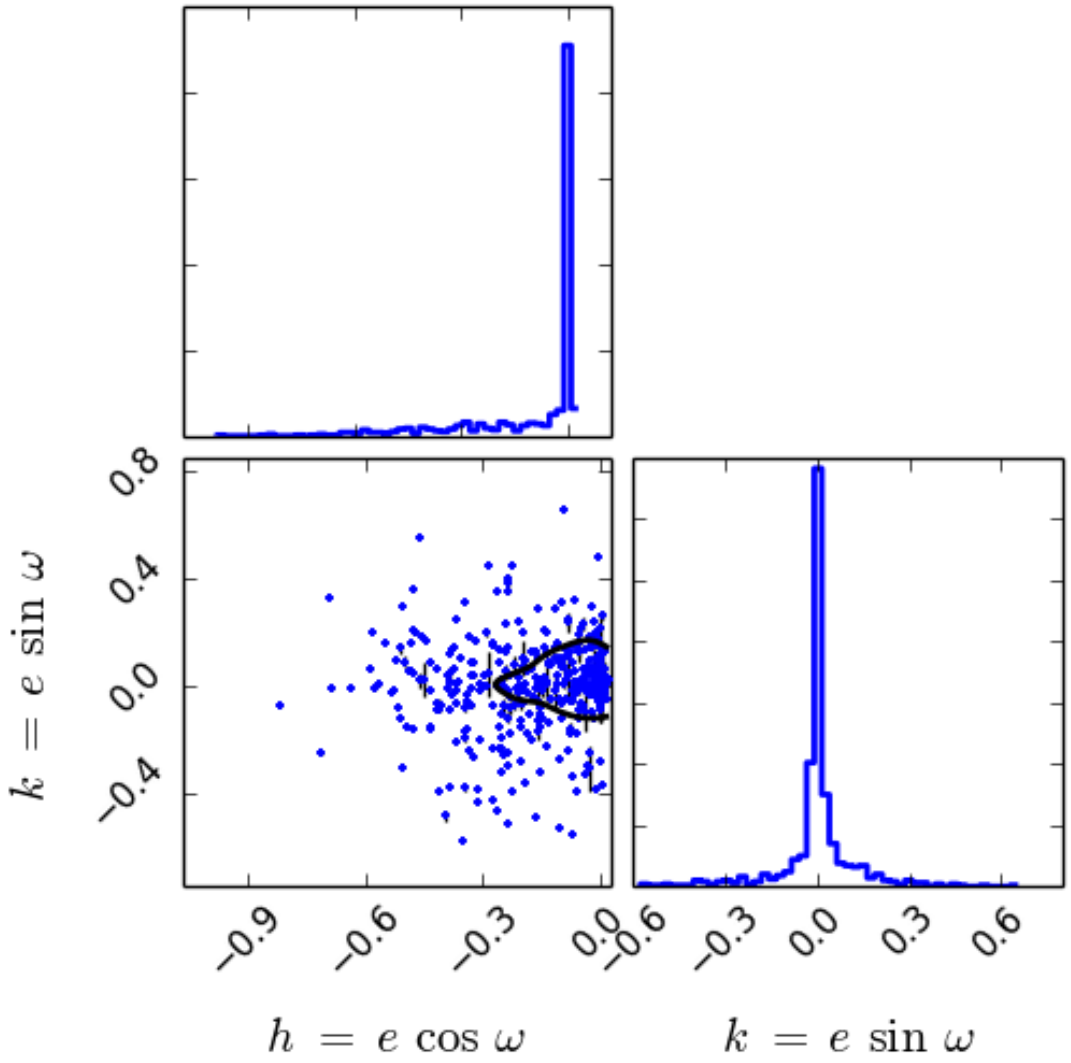

Figure 4.1: Projected eccentricity ($h = e\cos\omega$ and $k = e\sin\omega$) for *Kepler* EBs. The blue dots indicate the scatter in h and k for our sample of 795 EBs. The black bars show the measurement uncertainties, and are only visible in some cases, as the measurement uncertainties for this sample are small. The 68% credible region of the data is shown as the black curve in the bottom left panel. Histograms of the h and k measurements are shown in blue.

4.3 Method

We aim to characterize the eccentricity distribution and joint period-eccentricity distribution of a sample of EBs from *Kepler*. First, we parameterize the eccentricity

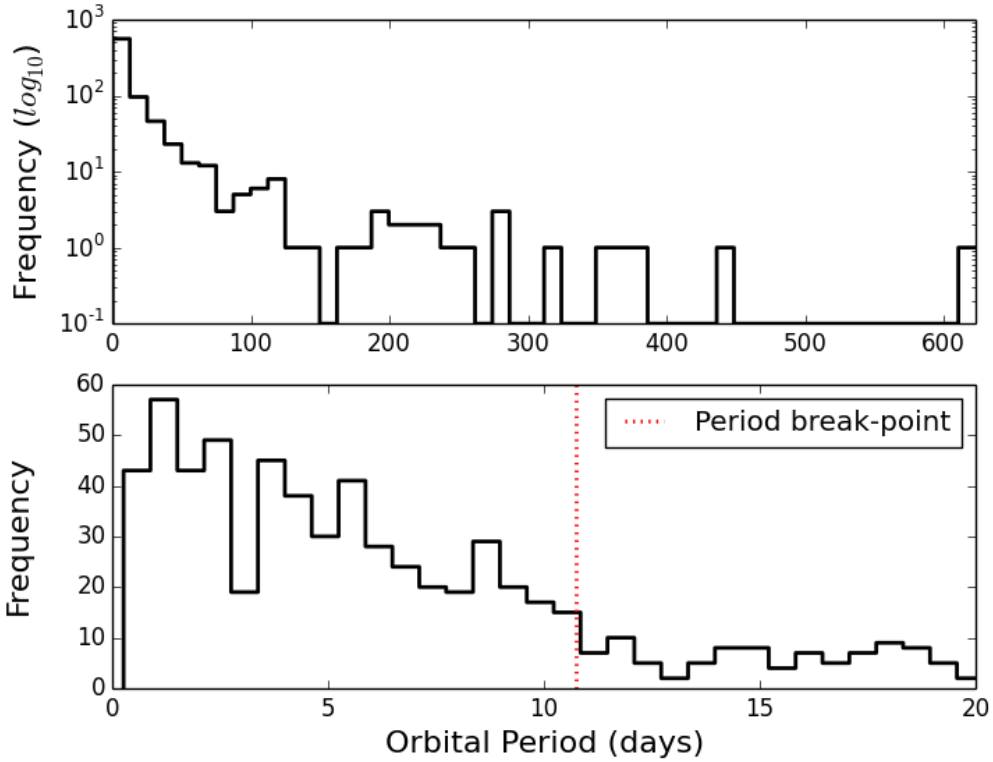

Figure 4.2: *Kepler* EB orbital periods. The top panel shows the histogram of the full range in period for our sample of EBs, ranging from 0.00325 to 623 days. The frequency of orbital period values for the top panel is shown on a log scale for clarity. The bottom panel zooms in on orbital period for clarity in the region of interest, showing the frequency of orbital period between 0 and 20 days. The red dotted line shows the posterior mode of the inferred period break point at which the EBs below this threshold have a different eccentricity distribution than those above this threshold. See §4.3 and 4.4 for details regarding this result. We use the full sample of EBs in our analysis.

distribution as a mixture of two Gaussian distributions for h and k. If h and k were each parameterized by a simple Gaussian, $e = \sqrt{h^2 + k^2}$ would take on a Rayleigh distribution. Because $e < 1$ for bound orbits, we always truncate the eccentricity distribution to $0 \leq e < 1$ and renormalize, so that the total probability integrates to unity. Later, we set up a joint period-eccentricity distribution where we infer a period transition between two, two-component Gaussian mixture model eccentricity distributions. The methods used in this study are described in detail in Chapter 2, specifically, §2.3. In §2.3 we validate the HB model used in this study. We show a

graphical model of our period-eccentricity distribution HB model in Figure 4.4.

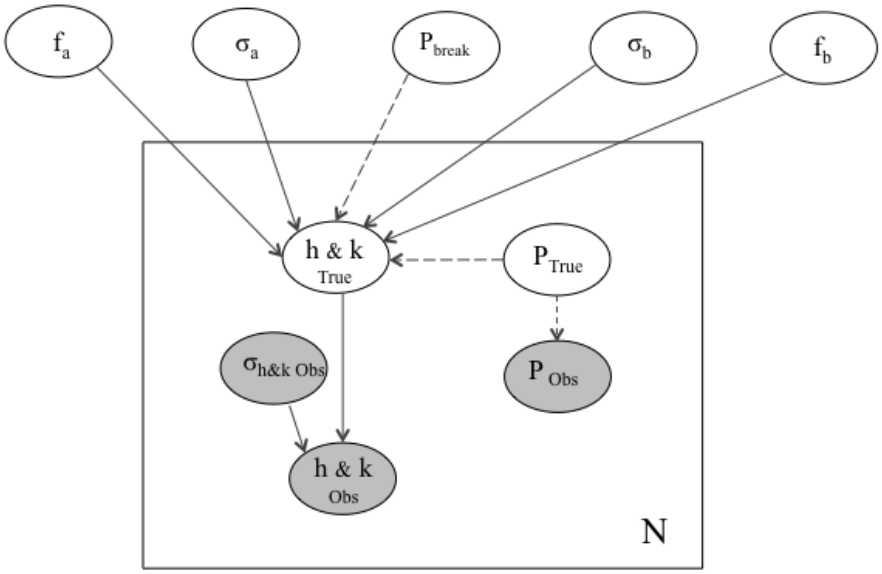

Figure 4.3: Graphical model of Bayesian network. Here we show the relationship between the hyperparameters, latent variables and observables in the joint period-eccentricity hierarchical Bayesian model. The observables are shown in gray. The latent variables are shown in white within the boxed region. The hyperparameters that describe the features of the population model are shown in white above the boxed region. The solid arrows represent probabilistic relationships and can be read as "distributed as", where the dashed arrows represent definitions or physical models and can be read "defined as".

4.3.1 The Hierarchical Bayesian Model

4.3.1.1 The Population Model

We consider eccentricity-period distributions that are a mixture of normals for h and k (truncated so $e < 1$) for any given period. We assume either: 1) the same eccentricity distribution for all orbital periods considered, or 2) a piece-wise

version, where the parameters for the eccentricity distribution depend on whether the orbital period are below or above a "period break" point, P_{break}. The HB model we use for our joint period-eccentricity distribution is as follows:

$$\begin{aligned} p(\boldsymbol{f_a}, \boldsymbol{\sigma_a}, \boldsymbol{f_b}, \boldsymbol{\sigma_b}, P_{break}, h_{true}, k_{true}, P_{true} | h_{obs}, k_{obs}, \sigma_{h_{obs}}, \sigma_{k_{obs}}, P_{obs}) \\ \propto p(\boldsymbol{f_a}) p(\boldsymbol{\sigma_a}) p(\boldsymbol{f_b}) p(\boldsymbol{\sigma_b}) p(P_{break}) \\ p(h_{true}, k_{true} | \boldsymbol{\sigma_{ref}}) p(\boldsymbol{\sigma_{ref}} | f_a, f_b, \sigma_a, \sigma_b, P_{break}, P_{true}) \\ p(h_{obs}, k_{obs} | h_{true}, k_{true}, \sigma_{h_{obs}}, \sigma_{k_{obs}}) p(P_{true} | P_{obs}) \end{aligned} \quad (4.1)$$

In the expression for the HB model in Equation 4.1 above, P_{break} is the period transition at which the EB eccentricity distribution changes. We characterize the eccentricities of the EBs for any given orbital period using a two-component Gaussian mixture model where the mixture fractions are $\boldsymbol{f_a}$ and $\boldsymbol{f_b}$, and the dispersions of the two components are $\boldsymbol{\sigma_a}$ and $\boldsymbol{\sigma_b}$, respectively (each with N_{comp} components). The true values of the projected eccentricities for each EB are given as h_{true} and k_{true}. The dispersion vector, $\boldsymbol{\sigma_{ref}}$, represents either $\boldsymbol{\sigma_a}$ or $\boldsymbol{\sigma_b}$, depending on if the EB has a period that is below (use $\boldsymbol{\sigma_a}$) or above (use $\boldsymbol{\sigma_b}$) the period break point. The true orbital period is represented as P_{true}, with the observed orbital period represented as P_{obs}. The measurements of the projected eccentricity are given as summary statistics h_{obs}, k_{obs}, $\sigma_{h_{obs}}$, and $\sigma_{k_{obs}}$. The "p()" represents each separable function in our HB model and are described in detail in §4.3.1.2 through 4.3.1.4.

4.3.1.2 Hyperpriors

Equations 4.2 through 4.7 below show the hyperprior distributions for our HB model. Here, $N_{comp,a}$ and $N_{comp,b}$ represents the number of mixture components used in the Gaussian mixture models to parameterize the eccentricity distribution. Except where otherwise specified, in this study we use two mixture components. We assign a Dirichlet distribution as the prior for the mixture fractions of the eccentricity distributions on either side of the period break point ($\boldsymbol{f_a}$ and $\boldsymbol{f_b}$) with $\alpha = 1$. For the associated dispersions for each of the h and k distributions, we assign a Uniform distribution between 0 and 1. We assign a Uniform distribution for

the range in period break point values P_{break} between the minimum and maximum periods in our sample of EBs, 0.00325 and 623 days, respectively.

$$\boldsymbol{f_a} \sim Dirichlet(\alpha = 1)\, for\, a = 1...N_{comp,a} \tag{4.2}$$

$$\boldsymbol{\sigma_a} \sim Uniform(0, 1)\, for\, a = 1...N_{comp,a} \tag{4.3}$$

$$\boldsymbol{f_b} \sim Dirichlet(\alpha = 1)\, for\, b = 1...N_{comp,b} \tag{4.4}$$

$$\boldsymbol{\sigma_b} \sim Uniform(0, 1)\, for\, b = 1...N_{comp,b} \tag{4.5}$$

$$P_{break} \sim Uniform(P_{min} = 0.00325, P_{max} = 623.0) \tag{4.6}$$

$$\tag{4.7}$$

4.3.1.3 Mid-Level Model Specification

Here we describe the mid-level of our HB model, that relates the latent variables in our study to the population level parameters. We use a Categorical distribution for the category to assign each EB to a mixture component. Whether f_a and σ_a or f_b and σ_b are used for a given EB, depends on whether the orbital period is greater or less than P_{break}. The true projected eccentricity values h_{true} and k_{true} are drawn from a two-component Gaussian mixture model centered at zero, with a dispersion $\boldsymbol{\sigma_a}$ if it is drawn from the h and k distribution appropriate for periods below the period break point, and $\boldsymbol{\sigma_b}$ if it is drawn from the h and k distribution appropriate for periods above the period break point. The mid-level model variables are vectors of N (number of EBs), and in this study $N = 795$.

$$a_{j,i} \sim Categorical(\boldsymbol{f_{a,i}})\, for\, j = 1...N_{comp,a},\, i = 1...N \tag{4.8}$$

$$b_{j,i} \sim Categorical(\boldsymbol{f_{b,i}})\, for\, j = 1...N_{comp,b},\, i = 1...N \tag{4.9}$$

$$\boldsymbol{\sigma_{ref,i}} \leftarrow \begin{cases} \sigma_{a_{j,i}}\ (P_{true,i} < P_{break})\, for\, j = 1...N_{comp,a},\, i = 1...N \\ \sigma_{b_{j,i}}\ (P_{true,i} > P_{break})\, for\, j = 1...N_{comp,b},\, i = 1...N \end{cases} \tag{4.10}$$

$$h_{true,i}, k_{true,i} | \boldsymbol{\sigma_{ref,i}}, P_{true,i} \sim Normal((\mathbf{0}), (\boldsymbol{\sigma_{ref,i}}\, \mathbf{1}))\, for\, \sqrt{(}h^2 + k^2) < 1,\, i = 1...N \tag{4.11}$$

4.3.1.4 Base-Level Model Specification

The base-level of our HB model describes the measurements used to constrain the model. This level links the variables for which we have measurements to their associated latent variables described in §4.3.1.3. The observables in our study are the orbital periods and the projected eccentricity $h = e\cos\omega$ and $k = e\sin\omega$. Here we assume that the observed values of h and k are Normally distributed about the true values of h and k, with a known dispersion. Because the measurement uncertainty is very small for orbital period, the orbital periods are taken to be known.

$$h_{obs,i}, k_{obs,i} | h_{true,i}, k_{true,i}, \sigma_{h_{obs,i}}, \sigma k_{obs,i} \sim Normal((\boldsymbol{h_{true,i}, k_{true,i}}), (\boldsymbol{\sigma_{h_{obs,i}}, \sigma k_{obs,i}}\, \mathbf{1})) \quad for\ \sqrt{(h^2 + k^2)} < 1,\ i = 1...N \tag{4.12}$$

$$P_{obs,i} \leftarrow P_{true,i}\ for\ i = 1...N \tag{4.13}$$

4.3.2 Evaluating the Hierarchical Model

We sample from the posterior of our HB model using MCMC. We use the publicly-available code Just Another Gibbs Sampler (JAGS; Plummer, 2003), to calculate Markov chains. JAGS uses Gibbs sampling when possible, and otherwise reverts to standard Metropolis–Hastings. We simultaneously sample from both the posterior distributions for the population parameters and the posterior predictive distributions for each observable. We evaluate the Gelman-Rubin (R) ratio to test for non–convergence, and accept chains with an R < 1.01. We also look at the autocorrelation function for the Markov chains and accept cases that have a zero crossing at a lag of ≤ 5, after the chains have been thinned by a factor of 500 iterations. Our results and figures are based on the last 1000 iterations (after thinning). The exact JAGS input model used in our study can be found in Figure 4.4.

```
model
{
  p.break ~ dunif(min(p), max(p))
  for (j in 1:Nm) {
    e.sigma.b[j] ~ dunif(0.0E+00, 1)
    e.phi.b[j] <- 1/(e.sigma.b[j] * e.sigma.b[j])
    a[j] <- 1
  }
  f.b ~ ddirch(a[])
  e.sigma.a ~ dunif(0.0E+00, 1)
  e.phi.a <- 1/(e.sigma.a * e.sigma.a)
  for (n in 1:Ndata) {
    c.b[n] ~ dcat(f.b[])
    e.phi.ref[n] <- ifelse(p[n] < p.break, e.phi.a, e.phi.b[c.b[n]])
    h[n] ~ dnorm(0.0E+00, e.phi.ref[n]) T(-1, 1)
    k[n] ~ dnorm(0.0E+00, e.phi.ref[n]) T(-sqrt(1 - h[n] * h[n]), sqrt(1 - h[n] * h[n]))
    hhat[n] ~ dnorm(h[n], 1/(hhat.sigma[n] * hhat.sigma[n])) T(-1, 1)
    khat[n] ~ dnorm(k[n], 1/(khat.sigma[n] * khat.sigma[n])) T(-sqrt(1 - hhat[n] * hhat[n]), sqrt(1 - hhat[n] * hhat[n]))
  }
}
```

Figure 4.4: JAGS model for EB period-eccentricity distribution, with R syntax highlighting.

4.4 Results for *Kepler* Eclipsing Binaries

4.4.1 The Eccentricity Distribution

First, we evaluate the eccentricity distribution of our sample of eclipsing binaries independent of orbital period. We use a two-component Gaussian mixture model to characterize the eccentricity distribution. For the hyperparameters, we relabel the two indices from 1 and 2 to "low" and "high", where these are assigned based on the dispersion (i.e., $\sigma_{low} = min(\sigma_1, \sigma_2)$, and f_{low} is the mixture fraction associated with this corresponding σ. Roughly half (53%) EBs come from a distribution with a very small dispersion ($\sigma_{low} = 0.0067 \pm^{0.0003}_{0.0003}$). Most of these EBs have eccentricities that are very close to zero. We are able to constrain the the dispersion of the low eccentricity component to better than 1 part in 100. The other 47% of the population comes from an eccentricity distribution with a broad dispersion ($\sigma_{high} = 0.21 \pm^{0.007}_{0.005}$). The marginal posterior distribution for mixture fractions results in $f_{low} = 53 \pm^{2.2}_{1.6}$ % and $f_{high} = 47 \pm^{1.8}_{1.9}$ %. We report the posterior modes and the credible intervals for the hyperparameters in our two-component HB model in Table 4.1. Figure 4.5 shows the 2D marginal posteriors and marginals for the hyperparameter posteriors. Figure 4.6 shows cumulative distributions for the EB projected eccentricity distribution, where the *Kepler* EB projected eccentricity data are shown as the dotted black curve. The gray shaded regions shows draws from the posterior of the eccentricity distribution.

Table 4.1: Posterior modes and 68% credible intervals of hyperparameters for the two-component Gaussian mixture model for h and k.

		68% credible interval	
Hyperparameter	Mode	+	−
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

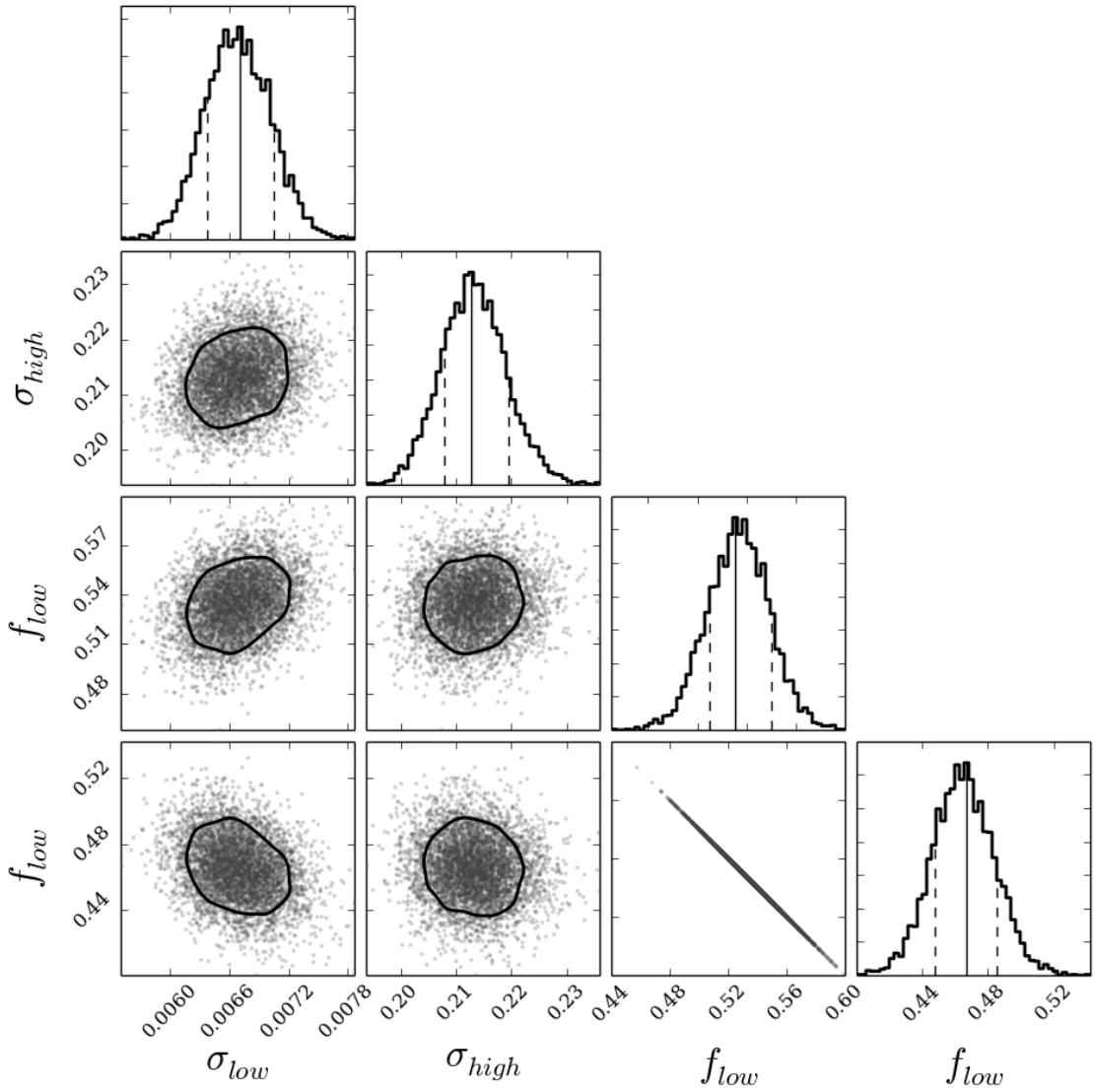

Figure 4.5: Posteriors for the hyperparameters in the two-component Gaussian mixture model described in Chapter 2, applied to *Kepler* EB projected eccentricity measurements. Approximately half the population comes from an eccentricity distribution with a very small dispersion ($\sigma_{low} = 0.0067$). The other half comes from a distribution with a larger dispersion ($\sigma_{high} = 0.21$), meaning some EBs will have significant eccentricities in this half of the population. The black contour for the 2D joint posteriors represent the 68% credible region. For the marginal distributions shown as histograms, the solid black line represents the posterior mode, and the dotted black lines represents the 68% credible interval. The posterior modes and the 68% credible intervals of the marginal posteriors can be found in Table 4.1.

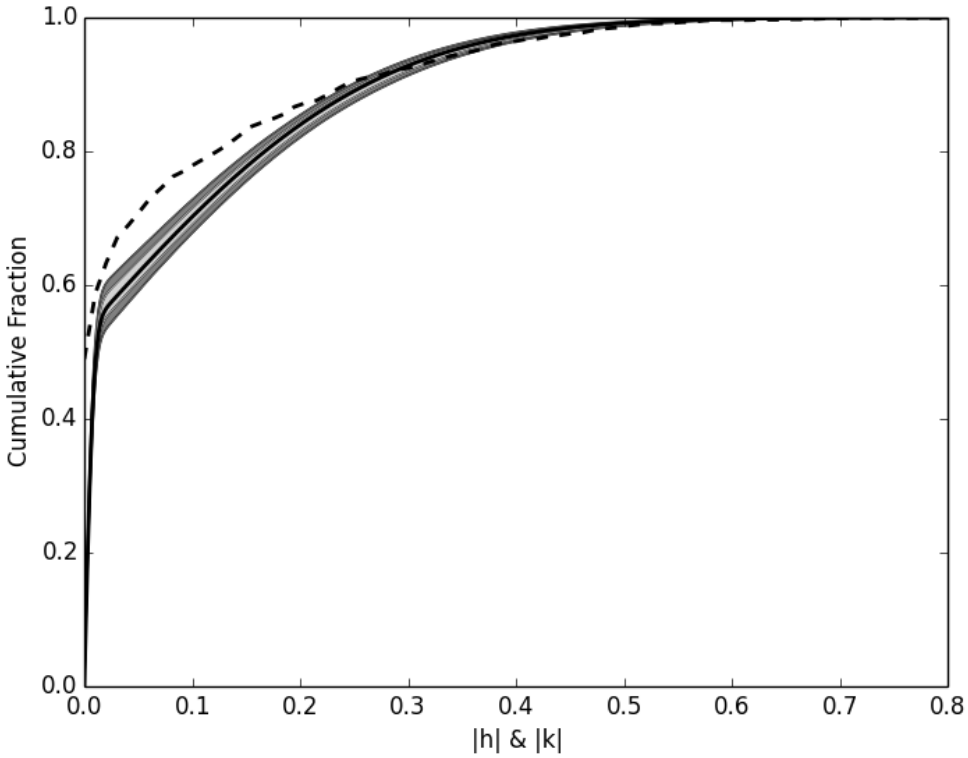

Figure 4.6: Cumulative distributions for the EB projected eccentricity distribution. The *Kepler* EB projected eccentricity data are shown as the dotted black curve. The light gray shaded region shows the 68% credible region for the projected eccentricity distribution using posterior draws of the hyperparameters. The dark grey region represents the 95% credible region. The solid black curve is the cumulative distribution for the projected eccentricity using the posterior mode for the hyperparameters.

4.4.2 The Period-Eccentricity Distribution

We explore the period-eccentricity distribution for our sample of *Kepler* eclipsing binaries. We develop a joint HB model where the eccentricity distribution differs for EBs with orbital periods greater or less than a break in the period distribution, where this break point is also a parameter to be inferred.

In our first scenario, we characterize the eccentricity distribution for EBs that have orbital periods that fall below the period break point to be a one-component Gaussian distribution, and EBs that have orbital periods that fall above the period break point to be a two-component Gaussian mixture model. In this case $N_{comp,a} = 1$

and $N_{comp,b} = 2$ in Equations 4.2 through 4.10. In this case the inferred period break point is at $1.5\pm^{0.04}_{0.03}$ days. This means that for EBs with orbital periods above 1.5 days, the eccentricity distribution is better characterized by a two-component Gaussian mixture model, where the population that has EBs with orbital periods below 1.5 days has an eccentricity distribution that is better characterized by a one-component Gaussian distribution. In this regime, the population of EBs with eccentricities above the period break point have a very similar eccentricity distribution as when fitting the eccentricity distribution with a two-component Gaussian mixture model independent of orbital period (e.g., the populations have roughly equal mixture fractions, and the low eccentricity dispersion components are both on the order of 0.006 and the high dispersion components are both around 0.21). In the simplest case we explore in §4.4.1, this is the dominant population that emerges.

Table 4.2: Posterior modes and 68% credible intervals of hyperparameters for a one-component (low orbital periods) and a two-component (high orbital periods) Gaussian mixture model for h and k with period break point.

		68% credible interval	
Hyperparameter	Mode	+	−
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
$P_{break}[days]$	1.5	0.04	0.03

In our second scenario when characterizing the joint period-eccentricity distribution, we assign two independent two-component Gaussian mixture models for the eccentricity distribution on either side of the period break point. We find that the period break point in this regime is at $10.74\pm^{0.26}_{0.027}$ days. There is a second shorter mode at approximately 11 days, and we suspect the possibility of an EB with an eccentricity value that has a probability of being above and below the break point that falls in between these two modes causing the split. The eccentricity distribution for the population with orbital periods below the period break point has

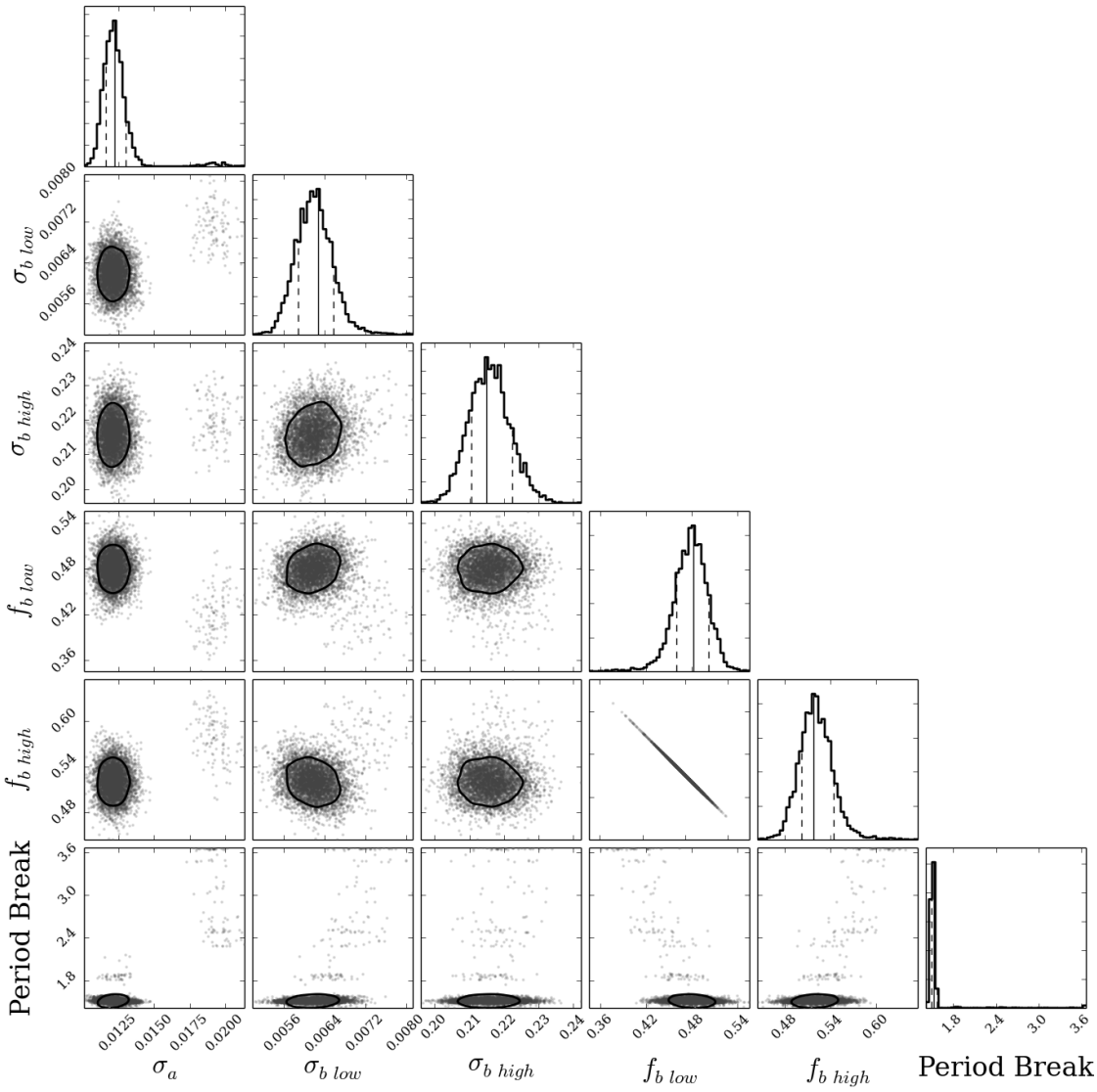

Figure 4.7: The black contour for the 2D joint posteriors represent the 68% credible region. For the marginal distributions shown as histograms, the solid black line represents the posterior mode, and the dotted black lines represents the 68% credible interval. The posterior modes and the 68% credible intervals of the marginal posteriors can be found in Table 4.2.

$\sim 75\%$ of EBs that come from an eccentricity distribution with a small dispersion of $0.0055\pm^{0.0004}_{0.0002}$. The other $\sim 25\%$ of EBs with orbital periods below the period break point come from a population with a eccentricity distribution characterized by a larger dispersion of $0.1\pm^{0.0047}_{0.0048}$. The population of EBs with orbital periods about the ~ 11 day period break point is characterized by an eccentricity distribution

with only $\sim 10\%$ of the EBs coming from an eccentricity distribution characterized by a small dispersion of $0.018\pm^{0.0035}_{0.0018}$, and the other $\sim 90\%$ of the EBs with orbital periods about this period break point coming from an eccentricity distribution with a large dispersion of $0.26\pm^{0.0074}_{0.011}$. This means that many of the EB population above the period break point have cases where eccentricities are significantly different from zero, and take on a wide range of eccentricity values.

Table 4.3: Posterior modes and 68% credible intervals of hyperparameters for a two-component (low orbital periods) and a two-component (high orbital periods) Gaussian mixture model for h and k with period break point.

		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

4.5 Discussion

Interestingly, we find that the period break of ~ 10.74 days when analyzing the EB population with a two-component Gaussian mixture model on either side of the period break point. Figure 4.2 shows a histogram of the periods of the binaries used in our study, with the inferred period break point for this HB model regime plotted as the vertical dotted red line. This is consistent with previous findings from Latham et al. (2002); Mayor et al. (2001), where they observe a period break point at around 10 days for EBs from disk and halo parent populations, indicative of tidal circularization operating on this population of short-period EBs. Latham et al. (2002) find that EBs with periods shorter than 10 days have EBs on nearly

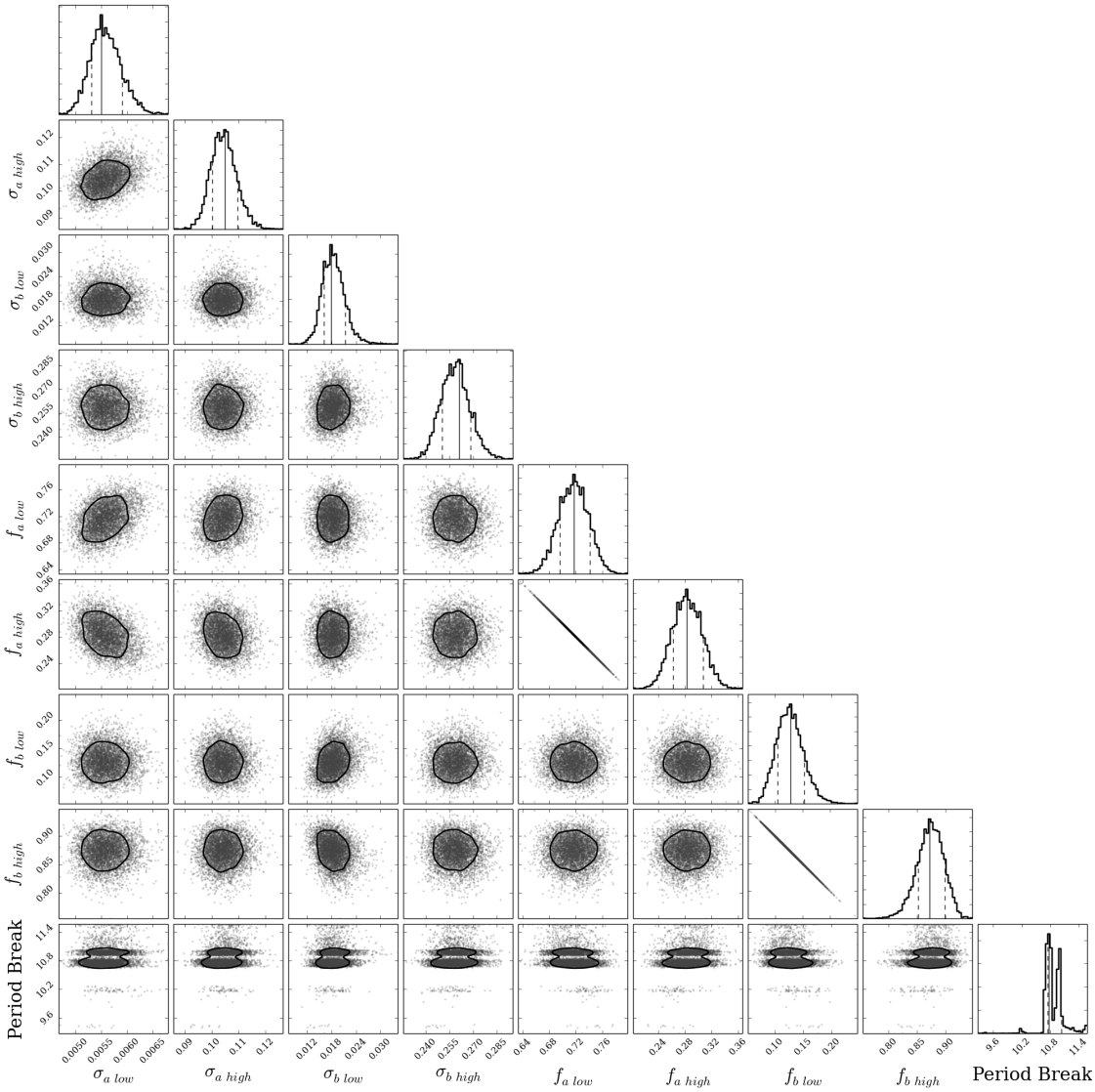

Figure 4.8: The black contour for the 2D joint posteriors represent the 68% credible region. For the marginal distributions shown as histograms, the solid black line represents the posterior mode, and the dotted black lines represents the 68% credible interval. The posterior modes and the 68% credible intervals of the marginal posteriors can be found in Table 4.3.

circular orbits, and EBs above 20 days take on a wide range of eccentricities. In the marginal posterior for our inferred period break point in this HB model regime, there is a second shorter mode at approximately 11 days. We suspect that there is an EB with an eccentricity value that has a probability of being above and below the break point that falls in between these two modes and is causing this split.

Many of the observational biases in this study are similar to those described in §2.7.1 where we explore the eccentricity distribution using projected eccentricity measurements h and h of short period planet candidates detected in both transit and occultation by *Kepler*. In particular, our sample of EBs is incomplete for long period binaries. The geometrical transit probability limits the sample towards larger periods. Furthermore, our sample of EBs are a field population not a cluster population, for which we currently do not have good age estimates. However, results from Latham et al. (2002), where they evaluate populations of EBs from the disk of the galaxy versus the halo, indicate that characteristics that are distinct between these two populations, such as metallicity, do not impact the eccentricity distribution. Thus, Latham et al. (2002) suggest that these differences have little impact over the process that leads to short-period binaries on circularized orbits.

In the future, it would be interesting to explore how the period-eccentricity distribution correlates with the EB mass ratio, tidal circularization times, and ages. Furthermore, it would be interesting to look for more than one split in the orbital period and eccentricity distribution explored in this study, to see if we can find another period-break point around 20 days similar to that found in Latham et al. (2002).

Table 4.4: Orbital period measurements and projected eccentricity measurements with uncertainties for *Kepler* EBs.

ID	Period (days)	$e\cos\omega$	$\sigma_{e\cos\omega}$	$e\sin\omega$	$\sigma_{e\sin\omega}$
01026032	8.460440	-0.041580	0.000030	0.002120	0.004620
01433980	1.592630	-0.000390	0.000100	0.014740	0.001300
01571511	14.022450	-0.042220	0.000570	0.231540	0.023800
01995732	77.362720	-0.131260	0.000040	0.079460	0.004080
02019076	7.129230	0.000060	0.000020	-0.003280	0.003720
02161623	2.283470	0.001420	0.000050	-0.040360	0.002230
02162994	4.101600	0.000200	0.000010	-0.001040	0.000440
02305372	1.404690	0.004280	0.000020	-0.028200	0.000680
02306740	10.306990	-0.023320	0.000010	0.295810	0.000560
02307206	204.031380	-0.010910	0.000030	0.199010	0.020770
02442084	49.788600	-0.502620	0.000490	-0.304310	0.002910
02445134	8.412010	0.000960	0.000220	-0.006190	0.008270

02447893	0.661620	-0.000360	0.000020	-0.017730	0.000680
02576692	87.878530	-0.142050	0.000020	0.144120	0.001180
02580872	15.926630	-0.052980	0.000010	0.193950	0.000880
02693092	39.841520	-0.247330	0.000550	-0.075970	0.031460
02708156	1.891270	0.000080	0.000040	-0.075030	0.001480
02719873	17.279290	-0.001700	0.000000	0.026350	0.000530
02852560	11.961310	-0.332960	0.000090	-0.244140	0.001010
02860594	5.499940	0.001200	0.000870	-0.023980	0.002110
02860788	5.259740	-0.000060	0.000030	0.001230	0.001090
02973509	6.627610	-0.102520	0.000020	0.105210	0.000810
02998124	28.597880	-0.043640	0.000020	0.029830	0.001950
03003991	7.244780	-0.000270	0.001070	0.010250	0.020680
03098194	30.476530	-0.200500	0.000050	-0.284650	0.000740
03102024	13.782520	-0.283510	0.000240	0.448870	0.001530
03115480	3.693700	0.000200	0.000080	0.002190	0.007740
03120320	10.265610	-0.036190	0.000060	0.011830	0.006760
03127817	4.327140	0.000190	0.000040	0.004330	0.001740
03218683	0.771670	-0.000930	0.000020	0.005320	0.000480
03241344	3.912680	-0.004910	0.000070	0.043890	0.003760
03241619	1.703340	-0.000000	0.000020	-0.012940	0.000540
03247294	67.418830	-0.385770	0.000470	-0.373580	0.002800
03248332	7.363610	-0.040780	0.000020	0.140250	0.002010
03323289	33.693200	-0.250160	0.000010	0.023940	0.001040
03327980	4.231020	-0.000650	0.000010	-0.000360	0.000360
03335816	7.422010	0.000040	0.000080	0.004280	0.011520
03341934	18.839500	-0.033560	0.000060	-0.294160	0.003840
03347485	7.544840	-0.000070	0.000870	-0.016390	0.037790
03348093	7.964420	0.006940	0.000030	0.266570	0.015140
03352751	3.495460	0.000920	0.000110	0.009870	0.007150
03428468	114.908680	-0.711120	0.001240	-0.251560	0.006470
03439031	5.952030	-0.000990	0.000000	-0.005040	0.000380
03440230	2.881100	0.001380	0.000030	0.010250	0.000990
03442054	117.681380	-0.377020	0.000240	0.138120	0.004540
03449540	3.212010	0.000000	0.000000	-0.007470	0.000330

03530668	1.946200	-0.000760	0.019900	0.013980	0.003400
03542573	6.942800	-0.000020	0.000020	0.005230	0.001720
03544694	3.845700	-0.001220	0.000030	0.002250	0.003320
03644542	119.679800	-0.504380	0.000180	0.297150	0.001080
03654950	8.134720	0.000240	0.000040	-0.023910	0.002730
03730067	0.294080	-0.000390	0.000020	-0.011240	0.000510
03735629	9.328350	0.000020	0.000010	-0.002840	0.001380
03757778	36.514400	-0.238520	0.000060	-0.245640	0.000990
03765771	5.567720	0.000160	0.000920	-0.014160	0.122190
03834364	2.908460	0.000080	0.000180	0.000710	0.007820
03836439	1.540400	0.000360	0.000150	-0.001080	0.007180
03838496	5.980400	0.000980	0.000020	-0.004750	0.000920
03848919	1.047260	-0.000310	0.000010	0.001940	0.000280
03851193	1.341080	0.004280	0.000790	0.010890	0.003650
03858884	25.951930	-0.433640	0.000040	-0.162240	0.000580
03865298	9.530640	-0.229990	0.000030	0.042890	0.002680
03938073	31.024310	-0.432230	0.000040	-0.012340	0.005760
03964545	3.012480	0.000010	0.000050	-0.007210	0.004040
03970117	110.295120	-0.481790	0.000040	0.159790	0.000540
03970233	8.254910	-0.010440	0.000270	0.101810	0.019260
03971315	9.892280	-0.000140	0.000390	0.021420	0.055350
03973002	3.984200	0.000090	0.000020	-0.002180	0.002150
03973504	4.318730	0.000140	0.000050	0.008480	0.008300
04037428	4.264730	0.000320	0.000080	0.049200	0.028230
04042026	34.820050	-0.330140	0.000100	0.089800	0.003340
04054905	274.709020	-0.305380	0.000010	-0.200890	0.000090
04069213	5.194260	-0.020780	0.001010	-0.015270	0.015260
04075064	61.422810	-0.237250	0.000360	-0.513330	0.002210
04076952	9.761170	0.018210	0.000350	0.033500	0.000420
04079530	17.727100	-0.210990	0.000120	0.240230	0.002060
04149684	4.320850	0.003810	0.000150	-0.003670	0.007820
04150611	94.198200	-0.365640	0.000100	-0.067710	0.004000
04157488	5.197420	0.008480	0.000020	0.001510	0.001270
04174507	3.891820	0.001340	0.000160	0.009350	0.018480

04247023	47.190620	-0.332150	0.000200	0.182870	0.003150
04247791	4.100870	0.000100	0.000020	-0.001110	0.000850
04249218	0.316200	0.005100	0.000020	-0.059760	0.001140
04276114	3.722410	-0.000180	0.000320	-0.022650	0.019030
04281895	9.543590	-0.303860	0.000030	-0.001910	0.003860
04285087	4.486030	-0.000090	0.000010	0.003740	0.000720
04346875	4.694220	0.000100	0.000090	0.007940	0.004200
04352168	10.643750	-0.155000	0.000030	-0.118660	0.001460
04365461	3.428830	0.000010	0.000020	-0.000300	0.001030
04375101	61.259600	-0.210620	0.000000	-0.000170	0.001560
04380283	1.744720	0.000140	0.000010	-0.000270	0.000470
04445630	5.627910	0.022570	0.000000	0.021240	0.000360
04473933	103.592620	-0.245610	0.000930	-0.001520	0.003880
04540632	31.005400	-0.091450	0.000080	0.656630	0.000800
04544587	2.189100	-0.237210	0.000010	0.146480	0.000320
04574310	1.306220	0.000350	0.000020	0.001800	0.000690
04578594	252.155000	-0.176690	0.000050	0.064240	0.004630
04586482	623.000000	-0.511120	0.000330	-0.118140	0.005340
04633434	22.271210	-0.025690	0.000000	0.096860	0.000840
04659255	198.013280	-0.063060	0.000020	-0.009820	0.019550
04660997	0.562560	0.000360	0.000020	0.003650	0.000320
04661634	73.894180	-0.434450	0.000040	-0.150490	0.000640
04663623	358.100000	-0.001700	0.000020	-0.003200	0.004350
04665989	2.248070	0.000180	0.000010	0.001360	0.000560
04670267	2.006100	-0.000070	0.000020	0.000290	0.000420
04671584	5.593320	-0.000140	0.000710	-0.051400	0.077880
04677321	1.572180	0.000510	0.000070	0.004210	0.001730
04678171	15.288710	-0.003480	0.000000	-0.006730	0.001020
04681152	1.835920	0.008860	0.000570	-0.007440	0.019010
04729553	0.961290	0.000480	0.000020	0.007450	0.000430
04736208	63.681890	-0.251000	0.000050	-0.194650	0.000990
04737267	9.524080	-0.006980	0.000020	-0.013520	0.000340
04737302	9.455590	-0.010460	0.000020	0.003950	0.001380
04739194	1.246590	0.002760	0.000040	-0.003080	0.000920

04739791	0.898920	-0.001040	0.000120	-0.010510	0.003620
04740676	3.454240	0.000080	0.000030	0.043100	0.011560
04751083	4.532370	0.012500	0.000110	0.029150	0.010900
04753988	7.304430	-0.005180	0.000020	0.009420	0.002770
04758368	3.749940	-0.003650	0.000100	0.002390	0.002610
04773155	25.706000	-0.264000	0.000080	0.354220	0.000770
04815612	3.856800	0.000020	0.000030	0.025770	0.002430
04826439	2.474300	-0.000060	0.000000	0.004590	0.000230
04839180	8.852690	-0.002460	0.000010	-0.005680	0.000810
04840263	1.915650	0.000110	0.000030	-0.001200	0.001270
04840818	10.049540	0.000010	0.000050	0.005460	0.006710
04847832	30.960240	-0.058000	0.000020	-0.396000	0.000850
04851217	2.470280	-0.031490	0.000020	0.003240	0.000630
04851464	5.548260	0.001760	0.000090	0.120250	0.003380
04902030	1.757610	0.000000	0.000010	-0.033620	0.000440
04912991	8.885040	0.000000	0.000040	0.009990	0.010820
04931073	26.951240	-0.407760	0.000140	-0.070700	0.004890
04937143	9.848880	-0.212780	0.000140	-0.006640	0.017050
04940201	8.816630	0.001370	0.000050	-0.040900	0.010440
04946680	8.665260	-0.115980	0.000710	-0.018970	0.059030
04947726	4.726090	-0.078800	0.000120	0.124590	0.005580
04948863	8.643590	0.004710	0.000050	0.024800	0.013320
05016163	5.245660	-0.000070	0.000000	-0.001140	0.000410
05017058	2.323890	0.000140	0.000020	-0.000710	0.000910
05022440	3.693950	-0.000040	0.000120	0.008900	0.016490
05023948	3.649290	-0.000190	0.000010	0.018670	0.001350
05024292	4.301030	0.010710	0.000030	0.020630	0.002010
05025294	5.462690	-0.197240	0.001310	0.104880	0.067400
05036538	2.122020	0.000090	0.000010	0.015910	0.000440
05036966	62.735000	-0.201630	0.000280	-0.060470	0.024360
05039441	2.151380	-0.002910	0.000220	0.004110	0.002930
05080652	4.144360	0.000040	0.000010	0.003440	0.000810
05090690	24.023320	0.000020	0.000060	-0.013010	0.021240
05091614	21.142460	-0.144620	0.000040	0.030250	0.007360

05097446	1.288060	0.000890	0.000080	-0.000140	0.000530
05110423	5.370430	0.000230	0.000190	-0.042020	0.033170
05113053	3.185100	0.000070	0.000020	-0.001730	0.000970
05115178	12.851720	-0.210820	0.000050	-0.010030	0.005950
05181455	5.579630	-0.000060	0.000060	-0.000340	0.008890
05193386	21.378290	-0.007030	0.000000	0.013860	0.000060
05215700	1.312400	0.000420	0.000020	0.000310	0.000560
05216727	1.513020	0.000000	0.000050	-0.002630	0.002360
05218014	10.845290	-0.221590	0.000500	0.083100	0.027820
05263802	6.700510	-0.006890	0.000050	-0.001860	0.005250
05264304	120.580000	-0.367230	0.000030	-0.016460	0.005990
05266937	5.917090	-0.000070	0.000020	0.010030	0.000670
05269407	0.958870	0.000600	0.000060	-0.015220	0.001290
05284133	8.784580	-0.182860	0.000050	-0.103300	0.002420
05285607	3.899400	0.000010	0.000060	-0.007680	0.003240
05286786	9.948750	-0.035910	0.000510	-0.093910	0.071500
05288543	3.457080	-0.002480	0.000030	0.009430	0.000480
05294739	3.736000	0.008260	0.000020	0.008490	0.000490
05300878	1.279440	-0.000460	0.000010	-0.007250	0.000300
05310435	4.931360	-0.000820	0.000060	0.004840	0.002530
05347784	9.584060	-0.010920	0.000020	0.000980	0.001500
05356593	45.977870	-0.092240	0.000290	0.102850	0.009540
05359678	6.230610	0.000040	0.000020	-0.000780	0.001190
05370302	3.904330	-0.002200	0.000260	-0.283040	0.003600
05372966	9.286360	-0.025350	0.000050	0.054310	0.005240
05384802	6.083090	-0.000010	0.000060	-0.009330	0.008790
05392897	42.399300	-0.413230	0.000440	-0.392350	0.000970
05393558	10.217100	-0.023070	0.000010	0.141630	0.001260
05394544	0.431970	0.001120	0.000800	0.021190	0.028380
05444392	1.519530	-0.000230	0.000010	0.000130	0.000240
05450814	6.487530	-0.026620	0.000720	-0.306680	0.084880
05456023	131.978310	-0.028070	0.000070	0.180910	0.013840
05460835	21.539270	-0.162090	0.000030	-0.006060	0.009260
05462901	5.270730	0.000180	0.000030	0.025670	0.004990

05471619	0.962840	0.001520	0.000030	-0.023950	0.000700
05473556	11.258820	-0.081780	0.000010	0.115310	0.000870
05479973	1.795270	0.001840	0.000110	0.005920	0.002550
05513861	1.510210	-0.000040	0.000020	0.000560	0.000420
05530881	5.082520	0.000170	0.000060	0.017700	0.001340
05534702	1.025470	0.000390	0.000030	-0.006010	0.000760
05534814	9.718780	-0.000090	0.000140	-0.022680	0.042990
05535280	74.985300	-0.686410	0.000030	-0.008530	0.005570
05553624	25.762070	-0.106440	0.000060	-0.532130	0.000720
05597970	6.717380	-0.000280	0.000010	0.017060	0.000890
05598639	1.297550	0.000270	0.000020	-0.008150	0.000800
05621294	0.938900	0.002100	0.000060	-0.003460	0.000960
05622250	16.294210	-0.320980	0.000110	0.037560	0.007950
05632781	11.025200	-0.253480	0.000010	-0.106130	0.000440
05652260	8.929260	0.001310	0.000030	-0.008470	0.006180
05700330	53.220410	-0.476100	0.000200	0.357560	0.001020
05728139	10.668160	0.000130	0.001230	0.001890	0.004980
05728283	6.198280	-0.056050	0.000050	-0.118840	0.001250
05730394	5.519180	-0.000620	0.000040	0.003180	0.000600
05731312	7.946380	-0.418060	0.000280	-0.088270	0.007390
05735878	16.832080	-0.415870	0.000090	-0.063590	0.002000
05736461	7.346160	-0.109870	0.000020	0.041560	0.001260
05738506	9.501480	-0.153080	0.000050	-0.151790	0.002060
05738698	4.808770	0.000370	0.000010	-0.003000	0.000440
05768927	4.390510	0.000040	0.000040	0.047740	0.002640
05781192	9.459990	-0.031450	0.000010	0.047890	0.001410
05786154	197.920780	-0.342070	0.000020	-0.161410	0.000380
05796185	42.559240	-0.233920	0.003190	0.388890	0.030310
05802470	3.791870	-0.000020	0.000030	-0.008320	0.002020
05807579	17.844330	-0.256300	0.000070	0.089620	0.002720
05809827	1.222110	0.000080	0.000030	0.004100	0.000780
05817566	8.412200	0.000120	0.000020	-0.006360	0.000830
05866104	36.976000	-0.345400	0.001740	-0.378660	0.011400
05871918	12.643300	-0.158310	0.001050	-0.166340	0.036300

05872150	20.355100	-0.032300	0.000400	0.060290	0.008470
05876805	18.173040	-0.177850	0.000030	-0.196750	0.000620
05955321	11.637880	-0.005950	0.000000	0.482230	0.000410
05961350	5.262670	-0.000560	0.000880	0.060980	0.076520
05962716	1.804610	0.000380	0.000110	-0.005960	0.003620
05979863	16.621850	-0.033660	0.000060	-0.020630	0.004770
05982353	19.563100	-0.056650	0.000360	-0.018460	0.016100
05983348	25.150780	-0.007300	0.000010	0.178590	0.002060
05986209	23.737970	-0.192080	0.000150	-0.364810	0.001890
06029130	12.591630	0.011080	0.000010	0.011130	0.000960
06029214	0.819070	0.000310	0.000070	-0.008520	0.002120
06042116	5.407160	-0.065640	0.000030	0.015110	0.002150
06042663	30.971770	-0.074860	0.000000	-0.174940	0.000250
06044064	5.063210	-0.007370	0.000040	0.037840	0.000320
06045264	0.909310	0.000190	0.000010	-0.000460	0.000260
06058875	1.129870	0.000140	0.000010	0.002910	0.000530
06060580	2.313370	0.000060	0.000200	0.016310	0.018410
06063448	76.017700	-0.445500	0.000050	0.166300	0.000630
06103049	0.643170	0.001680	0.000100	-0.001850	0.002900
06109688	14.086790	-0.066970	0.000030	0.322720	0.000930
06128027	5.455770	-0.004990	0.000150	0.060940	0.007450
06131659	17.527830	0.000060	0.000010	0.000030	0.002130
06145939	17.745410	0.000000	0.000010	-0.017860	0.008540
06147573	25.836830	-0.363750	0.000020	0.016160	0.002070
06182019	3.664960	-0.000640	0.000590	0.062550	0.026040
06182849	6.399110	0.000230	0.000020	0.018210	0.000860
06185717	11.702200	-0.054270	0.000140	0.131470	0.019120
06187893	0.789170	0.000340	0.000160	0.014350	0.002550
06197038	9.751710	0.007470	0.000030	0.018470	0.001740
06205460	3.722830	0.004380	0.000000	-0.036870	0.000040
06206751	1.245330	0.000310	0.000040	-0.001280	0.001050
06220470	8.144140	-0.154280	0.000020	-0.076370	0.000320
06227560	4.803510	-0.001240	0.000000	0.040940	0.000550
06228703	8.400410	0.000030	0.000030	0.003430	0.002420

06231401	6.091960	-0.023130	0.000060	-0.000710	0.006000
06233466	15.873310	-0.028910	0.000050	0.004440	0.001870
06233903	5.990850	-0.002290	0.000020	0.001030	0.001430
06301030	11.991160	-0.307750	0.000010	0.010340	0.001140
06307062	75.379010	-0.577340	0.000020	-0.009970	0.002310
06307537	29.744520	0.005860	0.000000	0.020520	0.000050
06312521	3.015550	0.000050	0.000030	-0.026540	0.003280
06359798	14.154040	-0.409660	0.000040	-0.091590	0.001040
06362386	4.592400	-0.001010	0.000060	-0.005580	0.001380
06363494	1.815630	-0.000320	0.000100	-0.002150	0.003610
06367628	3.779690	-0.000210	0.000020	-0.000640	0.000530
06387450	3.661330	-0.001080	0.003340	0.019150	0.027790
06421188	16.434350	-0.193320	0.000080	0.094620	0.003150
06431670	29.911480	-0.331160	0.000020	0.039330	0.001750
06443392	0.776980	-0.000250	0.000140	-0.000910	0.006850
06448768	16.486800	-0.159310	0.000070	0.291870	0.000840
06449358	5.776790	-0.000250	0.000020	0.042880	0.000260
06449552	20.148810	-0.103960	0.000070	-0.276940	0.002290
06452742	1.375540	0.000040	0.000050	-0.007050	0.003170
06464285	0.843650	0.000550	0.000060	-0.031250	0.000930
06466939	2.285890	-0.000030	0.000000	0.001270	0.000440
06468938	7.217010	-0.034720	0.000020	-0.004290	0.002300
06504534	28.162600	-0.090700	0.000110	0.068880	0.018980
06521542	4.425750	0.000020	0.000000	0.006950	0.000430
06522750	17.445640	-0.070160	0.000010	-0.058980	0.001220
06523216	14.313160	-0.042250	0.000260	-0.069530	0.002030
06525196	3.420600	0.000050	0.000010	-0.000430	0.000950
06531485	0.676990	0.000050	0.000130	0.021410	0.005440
06531496	14.323910	-0.001140	0.000070	0.030090	0.009670
06543674	2.391030	0.000050	0.000010	0.003000	0.000280
06545018	3.991460	0.002700	0.000020	0.003990	0.000490
06546508	6.107060	-0.000840	0.000040	-0.006680	0.007110
06548447	10.768380	-0.106410	0.000010	-0.047910	0.000030
06579806	9.880480	-0.053240	0.000320	0.126330	0.037890

06591789	5.088410	-0.013170	0.000010	0.013750	0.001000
06594972	10.819170	-0.011090	0.000010	-0.017030	0.001170
06607610	63.518030	-0.108460	0.000030	0.158480	0.000890
06610219	11.301000	-0.012110	0.000000	0.196420	0.000420
06614926	3.026760	0.000080	0.000080	0.104550	0.001090
06620003	3.428550	-0.000370	0.000040	-0.007570	0.006800
06629332	4.310560	0.000010	0.000050	0.026410	0.008470
06629588	2.264470	-0.003170	0.000100	0.001330	0.002410
06665223	119.412020	0.000060	0.000000	0.037250	0.001780
06672229	23.832370	-0.058430	0.000020	-0.175540	0.001430
06694186	5.554220	0.010070	0.000200	0.001420	0.013240
06695889	1.106560	-0.000140	0.000020	-0.018020	0.001120
06697716	1.443250	0.000050	0.000020	0.012740	0.001030
06706287	2.535390	0.000100	0.000010	-0.004820	0.000470
06707942	9.014200	-0.081800	0.000990	0.042070	0.080630
06756669	5.851530	0.002470	0.001170	0.220090	0.057030
06762829	18.795270	-0.003950	0.000540	-0.035760	0.003180
06766748	7.002860	-0.004260	0.000020	-0.007230	0.000740
06778050	0.945830	0.000050	0.000010	0.001670	0.000390
06778289	30.130150	-0.051230	0.000020	0.220310	0.001070
06781535	9.122090	-0.251240	0.000040	0.017940	0.004540
06806632	9.469160	-0.323950	0.000250	0.089050	0.004890
06841577	15.537540	-0.117580	0.000120	0.107730	0.006830
06842345	17.381610	0.000010	0.000040	-0.071940	0.016300
06859813	10.882420	-0.235380	0.000480	-0.116440	0.016660
06863229	1.994920	0.000540	0.000110	0.002750	0.000310
06863840	3.852730	0.000230	0.000000	0.001950	0.000460
06864859	40.877840	-0.634480	0.000010	-0.007770	0.001110
06866228	7.702370	-0.015170	0.000650	0.071010	0.067670
06877673	36.758890	-0.098150	0.000000	-0.120120	0.000070
06891512	3.505130	-0.000060	0.000010	-0.004420	0.000350
06927629	7.735230	0.000020	0.000010	-0.000380	0.000780
06939670	4.238860	-0.000140	0.000050	0.013730	0.002750
06949550	7.841070	-0.264770	0.000000	-0.003700	0.000540

06956216	4.818830	-0.006970	0.000060	0.008180	0.000840
06960446	10.149000	0.006160	0.000040	0.008010	0.004880
06962018	1.269890	0.000360	0.000010	0.021530	0.000490
06965293	5.077820	0.010100	0.000060	0.015670	0.011740
07009828	10.642450	-0.000030	0.000020	0.001280	0.001810
07021177	18.645330	-0.070720	0.000030	-0.554310	0.000600
07025540	2.148210	-0.000050	0.000010	0.010740	0.000480
07025851	4.681000	-0.000160	0.000020	0.013920	0.001830
07033748	4.392710	-0.000020	0.000010	0.000730	0.000370
07037405	207.150520	-0.158720	0.000020	0.107060	0.000190
07037540	14.405910	-0.005260	0.000090	0.103850	0.007500
07049486	26.718420	-0.274880	0.000100	-0.074430	0.004830
07097571	2.213960	0.000680	0.000140	0.014600	0.001050
07105574	20.726070	-0.526440	0.000030	-0.016470	0.002740
07118545	14.797220	-0.276310	0.000390	0.157590	0.003410
07121885	33.937540	-0.242520	0.000290	0.232440	0.004910
07122259	20.882170	-0.159100	0.002460	-0.056000	0.039200
07125636	6.490660	-0.015850	0.000010	-0.002530	0.001550
07128918	7.118800	0.004480	0.000020	0.006330	0.002850
07129465	5.491820	0.000040	0.000010	0.009880	0.000750
07132542	66.360990	-0.381030	0.000010	-0.033210	0.000810
07136958	29.015850	-0.195490	0.000040	-0.065380	0.002770
07137798	2.253540	-0.002250	0.000060	0.051170	0.004270
07177553	17.996470	-0.391190	0.000030	-0.008730	0.005340
07199774	4.708130	0.000120	0.000040	0.025930	0.003060
07200102	14.665760	-0.237740	0.000130	0.349520	0.001390
07203179	2.245740	0.000160	0.000010	0.003060	0.000390
07212066	3.840490	0.002630	0.000120	-0.001070	0.012990
07212722	2.316170	0.000270	0.000020	0.001720	0.000880
07257373	10.466900	-0.000020	0.000000	0.003740	0.000430
07258889	0.903900	0.000090	0.000020	-0.017080	0.000590
07270230	6.998090	-0.000300	0.000890	-0.056010	0.081930
07336754	12.155240	0.012500	0.000030	0.011660	0.003220
07353970	198.443500	-0.020710	0.000000	-0.099960	0.002080

07362852	7.036710	-0.001530	0.000010	0.003290	0.000560
07368103	2.182520	0.000940	0.000140	0.000900	0.003390
07369523	5.318840	0.000180	0.000010	-0.002770	0.000450
07374746	2.733890	-0.000070	0.000010	0.000870	0.001140
07376500	5.877130	-0.385710	0.000110	-0.085920	0.003340
07385478	1.655480	0.000090	0.000080	0.005160	0.002970
07455553	6.516500	0.000460	0.000710	0.020080	0.076520
07506446	0.352600	-0.000370	0.000010	-0.005770	0.000250
07541502	44.827320	-0.474670	0.001140	0.205150	0.009280
07584826	0.622310	0.005060	0.000040	-0.008460	0.000980
07593110	3.549380	-0.002970	0.000040	-0.056110	0.003900
07597703	9.217800	0.000000	0.000000	0.001370	0.000480
07605600	3.326190	0.000640	0.000140	0.012230	0.001940
07620844	58.621570	-0.064610	0.000080	0.149200	0.007710
07624297	18.019650	-0.011140	0.000070	-0.010130	0.013900
07630658	2.151140	0.000180	0.000020	0.005490	0.000770
07670485	8.467710	-0.009760	0.000010	0.014440	0.003040
07671594	1.410350	0.001110	0.000020	0.044580	0.002740
07677005	38.058140	-0.475480	0.000100	-0.157890	0.001260
07690843	0.786260	0.000770	0.000130	0.005470	0.003300
07691527	4.800230	-0.067550	0.000010	0.056440	0.000870
07695093	4.032770	0.000680	0.000260	0.021310	0.009480
07708193	7.942450	-0.000280	0.000070	0.011010	0.005320
07750740	5.874200	0.000160	0.000060	0.012530	0.004740
07751562	17.042160	-0.198690	0.000070	0.071150	0.002390
07770471	1.157800	0.000480	0.000020	-0.022500	0.000690
07802136	10.980250	-0.087470	0.000080	0.023650	0.012640
07816680	8.585200	-0.000030	0.000010	0.027800	0.001660
07818448	0.618950	-0.000230	0.000030	0.000360	0.000840
07821010	24.238240	-0.349540	0.000160	-0.575960	0.000530
07830637	121.599180	-0.172630	0.000030	0.130120	0.001250
07838906	5.614950	0.001800	0.001180	0.024340	0.083900
07846730	11.028230	-0.019900	0.000040	-0.035290	0.000550
07847927	70.376270	-0.135030	0.000110	0.118950	0.006590

07866921	91.499890	-0.187640	0.000010	-0.026920	0.001580
07877824	9.449430	-0.192430	0.000010	0.010570	0.000920
07898576	8.646900	-0.013050	0.000010	-0.037660	0.000210
07919653	4.210480	-0.000030	0.000020	-0.000680	0.001310
07938468	7.226950	0.000780	0.000020	-0.000440	0.000770
07938870	0.580730	0.000560	0.000120	0.011010	0.000780
07943535	4.719340	-0.150300	0.000160	-0.019620	0.018370
07947631	2.516560	-0.000190	0.000100	0.007220	0.011440
07950775	8.966610	-0.208770	0.000060	0.035900	0.005980
07960547	6.767340	-0.016100	0.000010	-0.030910	0.000860
07970629	2.051070	0.000010	0.000020	-0.003210	0.000960
07970760	7.649410	-0.000990	0.000170	0.005280	0.000700
07971389	26.349610	0.000040	0.000020	-0.023430	0.006280
07985167	1.384480	0.000180	0.000050	-0.014340	0.001950
07987749	17.030850	-0.142110	0.000030	0.021950	0.005620
08008128	14.854420	-0.411540	0.000890	-0.074070	0.023970
08009500	38.476640	-0.346290	0.001280	-0.192310	0.018580
08016214	3.174970	-0.000690	0.000030	-0.001830	0.000360
08043714	6.814990	0.004710	0.000400	0.120050	0.036480
08043961	1.559210	-0.000130	0.000020	-0.000650	0.000500
08044608	106.175940	-0.585820	0.000380	0.064510	0.009950
08052474	245.440000	-0.143270	0.000290	0.206690	0.009350
08074045	0.536390	0.001680	0.000060	0.000910	0.000610
08075618	17.561210	-0.000010	0.000010	-0.001380	0.004910
08081389	1.489440	0.000080	0.000020	-0.001810	0.000720
08088354	2.897830	0.000590	0.000080	-0.008660	0.006630
08094140	0.706430	0.000370	0.000020	0.000390	0.000810
08097825	2.936850	-0.000380	0.000020	-0.001740	0.000660
08098300	4.305920	0.005310	0.000010	-0.018440	0.000470
08106973	439.108380	-0.042110	0.000000	0.061820	0.001160
08111622	15.446060	-0.519460	0.000260	0.098200	0.005080
08112013	1.790520	0.001620	0.000110	0.034560	0.024860
08128965	7.140380	-0.000080	0.000080	0.033370	0.000650
08129189	53.647080	-0.204880	0.000040	-0.015800	0.005920

08145789	1.670630	-0.000170	0.000100	0.009320	0.007100
08177798	2.900410	0.000040	0.000130	0.001150	0.009030
08180020	5.803130	0.009610	0.000880	0.003380	0.023660
08183389	32.440340	-0.344710	0.000260	-0.083780	0.009130
08193315	2.623870	-0.000060	0.000050	0.001460	0.001320
08196180	3.671660	-0.153040	0.000040	0.020480	0.002090
08229048	5.478950	0.000140	0.000070	-0.001830	0.007710
08231877	2.615530	-0.000070	0.000130	-0.009900	0.011460
08234477	4.028670	0.001110	0.000920	0.000220	0.000480
08243263	3.518550	0.000010	0.000050	0.007820	0.003940
08244173	2.184130	-0.000040	0.000010	0.013480	0.000420
08248812	0.685370	0.000500	0.000030	0.008310	0.000450
08262223	1.613020	0.000210	0.000030	0.002120	0.001230
08264097	2.897470	-0.001560	0.000210	-0.033610	0.005700
08285254	7.535900	0.000670	0.000030	0.004550	0.001920
08288719	1.510100	-0.000040	0.000070	-0.018680	0.004660
08296467	10.303360	-0.194600	0.000030	0.198670	0.000710
08299947	5.365600	0.000160	0.000010	0.001440	0.001080
08301013	4.428060	0.001490	0.000010	-0.000210	0.000430
08302455	4.883980	0.000130	0.000020	-0.013420	0.001810
08312222	12.823280	-0.054700	0.000020	-0.090910	0.001860
08314801	37.183280	-0.386340	0.000000	0.019000	0.000070
08316503	5.065330	-0.109540	0.000040	0.006910	0.002950
08345358	9.380630	0.000030	0.000050	-0.005160	0.007870
08356054	17.081230	-0.028590	0.000030	-0.250100	0.003180
08364119	7.735910	0.018210	0.000010	0.020140	0.000640
08374499	5.251920	0.013020	0.000030	-0.009030	0.000960
08378656	3.629440	0.014820	0.000020	0.016480	0.000080
08378922	43.263310	-0.113220	0.000030	0.289290	0.000850
08381592	5.784180	-0.004450	0.000030	0.008380	0.004070
08411947	1.797680	-0.002080	0.000010	-0.002780	0.000580
08414159	11.931130	-0.000300	0.000060	-0.037860	0.003630
08414914	100.302500	-0.477710	0.000140	0.051050	0.005950
08415863	27.533800	-0.220380	0.000290	-0.075210	0.016960

08429450	2.705140	0.000000	0.000010	0.001000	0.000350
08430105	63.327110	-0.249880	0.000090	0.009630	0.000200
08444552	1.178100	0.000190	0.000030	-0.033090	0.001700
08445775	26.508270	-0.000030	0.000020	0.038700	0.007820
08453191	11.553530	0.000080	0.000010	-0.012620	0.001190
08453324	2.524540	0.000020	0.000240	-0.006730	0.015080
08460600	6.352090	-0.000380	0.000460	0.031350	0.014410
08479386	11.720760	0.000020	0.000000	0.001340	0.000480
08488876	5.801880	-0.137200	0.001200	0.012180	0.067920
08491745	111.319930	-0.153490	0.000030	-0.001280	0.000340
08504570	4.007700	0.000250	0.000020	-0.000280	0.001380
08509346	6.099030	0.008000	0.000160	0.013580	0.016540
08543278	7.549300	-0.013040	0.000050	0.029850	0.011170
08552540	1.061930	0.000510	0.000010	-0.001460	0.000330
08553788	1.606160	0.000180	0.000110	0.001580	0.002840
08553907	42.032210	-0.469430	0.000090	0.204160	0.000890
08559589	223.113050	-0.162980	0.000140	0.127970	0.006750
08559863	22.470500	-0.034880	0.000030	0.003360	0.005250
08560861	31.973230	-0.036630	0.000030	-0.008830	0.004060
08569819	20.849920	0.004150	0.000060	-0.370270	0.001030
08572936	27.795810	-0.166150	0.000070	-0.490440	0.000650
08574270	15.120190	-0.228410	0.000390	-0.084240	0.020340
08580438	6.496030	-0.000300	0.005620	-0.002810	0.009910
08581658	3.481660	-0.000710	0.000030	0.006450	0.000720
08589754	8.939380	-0.000460	0.000010	-0.024360	0.000720
08590527	0.739770	-0.000270	0.000360	-0.018800	0.011620
08590780	8.653890	0.000100	0.000000	-0.008600	0.000560
08604993	6.570420	0.011140	0.000060	-0.005400	0.005190
08605074	37.601620	-0.132930	0.000060	0.088070	0.002770
08608310	0.950380	0.000290	0.000020	-0.000670	0.000450
08608490	1.082810	0.000500	0.000010	0.016780	0.000380
08610483	48.799370	-0.326220	0.000130	-0.268840	0.001350
08618226	5.882040	0.000180	0.000020	0.064450	0.002490
08621026	4.059430	-0.001100	0.000070	0.013310	0.001260

08668810	313.100000	-0.391570	0.000190	0.040100	0.013640
08686150	51.425890	-0.236390	0.000250	0.405000	0.002140
08700506	43.797800	-0.264510	0.000120	-0.462710	0.000780
08701327	8.508840	-0.002010	0.000010	-0.003380	0.000660
08711548	18.778160	-0.156260	0.000020	-0.127540	0.000860
08719897	3.151420	-0.000030	0.000010	0.005300	0.000570
08736245	5.069480	-0.000590	0.000020	0.003010	0.000680
08738115	37.594270	-0.397760	0.005230	-0.484020	0.020460
08746310	6.855620	0.000100	0.000080	0.013730	0.010000
08773948	33.469520	-0.017430	0.000170	0.241950	0.000750
08780968	5.234490	-0.000260	0.000020	-0.005500	0.000930
08784288	3.655120	0.000040	0.000010	-0.012100	0.000400
08801343	2.739870	0.004190	0.000170	0.021140	0.038250
08823397	1.506500	0.000290	0.000020	-0.009030	0.000470
08841616	1.679580	-0.000690	0.000140	0.007660	0.003060
08845221	5.403460	0.001350	0.000070	0.003250	0.000480
08848271	9.991660	-0.282880	0.001160	0.044700	0.082110
08868650	4.447130	-0.001010	0.000210	-0.014430	0.000460
08879427	16.313280	-0.450820	0.000940	0.024590	0.067030
08879915	3.442630	0.000330	0.000000	-0.000060	0.000540
08892720	61.491140	-0.489350	0.000110	0.059610	0.003560
08906676	8.209520	0.000380	0.000060	0.006200	0.008680
08935352	10.908800	0.000050	0.000030	-0.002670	0.001100
08937019	5.663650	-0.257290	0.000040	-0.149920	0.001050
08938628	6.862220	-0.002740	0.000030	0.016580	0.007010
08939650	3.655320	-0.000100	0.000040	-0.002490	0.002630
08953059	5.499940	-0.100600	0.000160	0.072790	0.004750
08953296	0.784300	0.002140	0.000070	-0.012380	0.001350
08953426	5.331750	0.000010	0.000020	0.005460	0.001010
08957954	4.359850	0.000090	0.000010	-0.008670	0.000580
08971432	0.624390	0.002540	0.000400	-0.091080	0.011660
08973000	28.027870	-0.224260	0.000100	0.451300	0.000720
08984706	10.135400	0.014720	0.000010	0.005590	0.000750
09001468	17.329970	-0.273740	0.000510	-0.425490	0.003600

09007918	1.387210	0.001320	0.000160	0.010800	0.002970
09020426	0.913390	-0.000060	0.000010	0.001670	0.000240
09029486	6.277160	0.000170	0.000010	-0.002730	0.000980
09048145	8.667830	0.000410	0.001080	-0.003720	0.065480
09083523	0.918420	0.000520	0.000050	0.054230	0.001140
09091810	0.479720	-0.000820	0.000010	-0.000490	0.000240
09098810	8.258180	0.008820	0.000010	0.136400	0.000940
09119405	18.646320	-0.432640	0.000030	0.087240	0.000910
09143254	2.275050	-0.000090	0.000080	-0.017830	0.007580
09159301	3.044770	0.000520	0.000000	-0.037670	0.000180
09163674	1.546290	0.000090	0.000010	0.000610	0.000510
09164836	48.118820	-0.386290	0.000360	0.045250	0.026270
09172506	50.440310	-0.690180	0.000330	0.330170	0.001300
09179531	6.510780	0.007310	0.000010	-0.014550	0.001040
09207508	2.015150	0.000210	0.000020	-0.022460	0.000700
09210828	1.656410	-0.000250	0.000020	-0.002410	0.001170
09236858	2.537080	0.001170	0.000040	0.006570	0.001050
09246715	171.277700	-0.339710	0.000030	-0.057940	0.001490
09266285	5.613870	-0.003120	0.000070	-0.010000	0.004410
09284741	20.729260	-0.275560	0.000070	-0.231130	0.001110
09291748	15.926990	-0.017650	0.000070	0.067540	0.004240
09328852	0.645850	0.002370	0.000060	0.006560	0.001930
09344623	14.759480	-0.146170	0.000010	-0.119440	0.000620
09346253	34.064580	-0.327090	0.000400	-0.006940	0.003110
09353182	10.476260	-0.035270	0.000020	0.050100	0.000520
09357275	1.588300	0.000340	0.000070	0.000680	0.002150
09392702	3.909310	-0.000350	0.000060	-0.000410	0.002470
09402652	1.073100	-0.000030	0.000010	-0.001340	0.000260
09412462	10.186590	-0.021280	0.000010	-0.024860	0.000500
09418994	32.003800	-0.134460	0.000260	-0.221050	0.008130
09446824	4.202340	-0.002120	0.000120	0.145810	0.005140
09451096	1.250390	0.000620	0.000020	-0.008590	0.000680
09451127	5.117400	0.000280	0.002270	-0.006320	0.028990
09468296	5.748830	0.006690	0.000220	0.034320	0.041410

09474222	13.691790	-0.337870	0.000050	-0.032410	0.001860
09474485	1.025160	-0.000090	0.000010	0.002140	0.000340
09474969	21.570510	-0.414700	0.000000	0.064590	0.000060
09489411	6.688930	-0.001650	0.000010	-0.016500	0.000250
09509207	14.199840	-0.369220	0.000090	-0.208080	0.001110
09520700	4.273670	0.000100	0.000010	0.006040	0.000690
09528430	90.082520	-0.354760	0.000060	0.102170	0.001340
09529856	37.482220	-0.461310	0.000150	0.020850	0.029920
09532123	8.215000	-0.207080	0.000070	-0.021220	0.009180
09540450	2.154710	-0.000080	0.000010	0.006380	0.000620
09569866	1.480320	0.000070	0.000020	-0.024840	0.000630
09576197	7.964430	-0.007470	0.000370	0.009820	0.003640
09579192	12.083540	-0.290160	0.001160	0.199910	0.013190
09579499	7.946720	-0.081370	0.000680	0.237030	0.033520
09592575	2.637580	-0.000000	0.000020	-0.004700	0.000920
09592855	1.219320	0.000040	0.000020	0.000310	0.000690
09597095	2.745640	0.000370	0.000110	-0.006660	0.006410
09602595	3.556510	0.001600	0.000000	-0.018910	0.000110
09613070	2.331850	0.000480	0.000040	0.018680	0.002120
09635529	4.703910	0.000110	0.000010	-0.007490	0.000600
09637299	1.882440	-0.000090	0.000020	0.002550	0.000840
09641031	2.178150	-0.000010	0.000010	0.000840	0.000560
09649222	5.918620	0.000840	0.000360	0.020440	0.011680
09652632	4.977620	-0.002010	0.000020	-0.009130	0.000870
09654476	0.396320	0.004420	0.000020	-0.012810	0.000510
09658118	24.060000	-0.152470	0.000070	-0.324290	0.001210
09665503	11.567980	-0.241500	0.000060	0.125640	0.001860
09673173	21.294740	-0.566140	0.000020	0.005800	0.001890
09692336	0.687900	0.000830	0.000050	0.013280	0.000610
09714358	6.474180	0.014080	0.000060	-0.020270	0.008290
09715925	6.308200	-0.198570	0.000140	-0.002500	0.022980
09762519	7.515150	-0.037920	0.000040	-0.000280	0.005840
09775253	8.788500	-0.000140	0.000000	0.145360	0.000330
09777062	19.230040	-0.156110	0.000070	-0.104630	0.004040

09777090	5.508140	-0.006560	0.000090	0.008950	0.013150
09785454	184.520000	-0.499630	0.000740	0.092100	0.018490
09786017	4.498000	0.001110	0.000080	0.005680	0.001010
09813678	0.505080	0.000770	0.000230	0.006310	0.010300
09818732	45.035590	-0.319390	0.000050	0.165320	0.000850
09821078	8.429440	0.000730	0.000010	-0.025980	0.001270
09833618	1.408500	0.000950	0.000050	-0.024450	0.001050
09833806	0.393180	0.002890	0.000260	-0.010740	0.006860
09837083	0.623390	-0.000660	0.000070	0.026320	0.003380
09837578	20.733750	0.008650	0.000000	0.127810	0.000880
09838060	23.815830	-0.072000	0.000140	-0.014350	0.048430
09838975	18.692870	-0.052580	0.000070	0.164070	0.003140
09850387	2.748500	0.000310	0.000030	0.001670	0.001000
09851142	8.480300	-0.202930	0.000720	-0.128200	0.003250
09851944	2.163900	0.000020	0.000020	0.002100	0.000570
09873869	4.994780	-0.000100	0.000070	0.003370	0.002830
09881258	4.057030	0.000100	0.000020	-0.000010	0.001020
09892471	8.268080	-0.000420	0.000000	0.086990	0.000530
09896435	18.076700	-0.011460	0.000110	0.039320	0.028730
09899416	1.332560	-0.000140	0.000020	-0.035500	0.000510
09906590	1.686340	-0.000090	0.000090	0.012650	0.004800
09907505	39.206810	-0.230560	0.000150	-0.036990	0.019840
09911112	2.333650	-0.000250	0.000240	-0.067970	0.028640
09912977	1.887870	0.000010	0.000010	0.000520	0.000330
09913481	7.009480	0.000060	0.000030	0.004000	0.001370
09934208	9.058490	-0.089320	0.000020	0.137840	0.001220
09935245	0.784860	0.005510	0.000060	-0.007500	0.000750
09936698	5.712150	0.000310	0.000030	0.012800	0.002560
09944421	7.095210	-0.096800	0.000040	-0.065980	0.003750
09957351	0.495550	0.001850	0.000070	-0.006520	0.001950
09965206	16.227550	-0.024390	0.000010	-0.035190	0.001920
09966115	8.947430	-0.000170	0.000040	0.066990	0.002310
09970396	235.299400	-0.136590	0.000020	0.004290	0.004610
09971475	5.357410	0.000060	0.000020	0.034330	0.001540

10001167	120.390970	-0.126740	0.002640	-0.048030	0.002170
10015516	67.692050	0.017090	0.026970	-0.048010	0.001660
10020423	7.448380	-0.018620	0.004840	-0.026660	0.010320
10026136	9.080190	-0.232380	0.000030	-0.180420	0.000620
10026457	9.934420	-0.102790	0.000030	0.091460	0.003160
10030778	29.473160	-0.011700	0.000080	0.024410	0.032690
10031409	4.143880	0.000060	0.000010	-0.008980	0.000800
10031808	8.589640	-0.020010	0.000020	0.152690	0.000990
10063044	1.009670	0.000270	0.000070	-0.002300	0.001570
10074700	365.650000	-0.019710	0.000010	0.146360	0.000840
10090246	2.286110	-0.001290	0.000080	0.004580	0.001880
10091110	383.860000	0.000030	0.001430	-0.011020	0.035150
10095469	0.677760	0.001230	0.000000	-0.012460	0.000080
10095512	6.017190	0.001100	0.000030	-0.002070	0.002240
10129482	0.846290	-0.000900	0.000040	-0.013240	0.001690
10152836	220.930000	-0.550330	0.000450	0.163030	0.004920
10155080	3.860900	-0.000050	0.000040	0.000680	0.002960
10156064	4.855940	0.000200	0.000020	0.002110	0.001570
10191056	2.427500	0.000900	0.000020	0.000390	0.000570
10198109	17.918740	-0.251390	0.000090	-0.232530	0.001400
10206340	4.564400	0.000100	0.000010	-0.008440	0.000320
10208759	4.575030	-0.000250	0.000010	0.001270	0.000310
10215422	24.847080	-0.291840	0.000020	-0.000540	0.003770
10215869	5.017270	0.000500	0.000050	0.003110	0.002200
10216186	0.605940	-0.000400	0.000100	0.030220	0.000670
10253421	0.428350	0.001090	0.000030	-0.029250	0.000960
10258558	20.867860	-0.020750	0.000020	0.039780	0.003320
10259029	7.061140	-0.134390	0.000100	0.019710	0.016290
10264202	1.035150	0.000350	0.000030	-0.027550	0.001310
10274244	13.683690	-0.069880	0.000010	0.089960	0.000740
10275074	4.362880	-0.000080	0.000090	0.012340	0.006350
10275887	9.726750	0.000190	0.000020	0.015160	0.000680
10292238	143.119260	-0.581690	0.000250	0.202270	0.002050
10294608	3.708980	0.000020	0.000000	-0.010200	0.000410

10296163	9.296760	-0.246860	0.000420	-0.260930	0.005110
10332789	18.740730	-0.084140	0.000030	-0.237960	0.001310
10345862	58.288220	-0.310710	0.000030	-0.436230	0.000160
10346522	3.989140	0.002020	0.000010	-0.051670	0.000540
10352603	32.778990	-0.344990	0.000080	0.310650	0.000660
10363300	0.934930	0.000400	0.000060	-0.006720	0.001680
10383620	0.734560	0.000570	0.000040	0.008580	0.000470
10383696	8.487700	-0.065290	0.000050	-0.000420	0.007820
10385682	6.207460	0.000010	0.000000	-0.001490	0.000470
10420279	45.433250	-0.313880	0.000120	-0.384660	0.000840
10454401	12.180720	-0.038300	0.000010	-0.328260	0.000620
10454725	0.831770	0.002080	0.000100	0.027660	0.000900
10462826	50.467030	-0.184720	0.000120	0.059840	0.010610
10480952	4.074910	0.000080	0.000010	-0.002990	0.000620
10483644	5.110770	-0.001070	0.000240	0.011420	0.032640
10486425	5.274820	0.000230	0.000040	0.008870	0.003840
10487119	6.749360	-0.000430	0.000000	-0.038680	0.000500
10489521	3.224500	0.000120	0.000020	-0.001540	0.000850
10490960	5.682410	-0.077700	0.000010	-0.052400	0.000870
10491031	4.398140	0.000110	0.000010	-0.007020	0.000590
10518735	19.515000	-0.516910	0.000110	-0.077700	0.002700
10549576	9.089470	-0.003260	0.000100	-0.004220	0.003660
10581918	1.801860	0.000820	0.000020	-0.017370	0.000670
10592163	14.762950	-0.271960	0.000230	0.149330	0.005580
10593759	6.265210	0.001170	0.000130	0.004270	0.015510
10601579	8.098750	-0.017870	0.000040	-0.015900	0.003360
10616829	1.072930	0.002550	0.000130	-0.015270	0.002670
10619109	2.045180	0.000680	0.000080	0.004170	0.001990
10651945	25.367000	-0.480250	0.000150	0.056650	0.005420
10657664	3.273700	-0.000940	0.000070	-0.014400	0.000370
10661783	1.231360	0.000870	0.000060	0.001480	0.000480
10677186	100.401300	-0.258570	0.000220	-0.234900	0.003460
10686876	2.618430	0.000600	0.000250	0.008420	0.001830
10711551	10.692220	-0.051150	0.000010	-0.061290	0.002710

10711913	19.408020	-0.078820	0.000330	-0.161080	0.024540
10727668	2.305900	0.000540	0.000100	-0.011240	0.000690
10735564	1.674980	0.000250	0.000040	0.007150	0.002150
10736223	1.105090	0.000660	0.000020	-0.028770	0.000640
10748621	286.180000	-0.296950	0.000030	0.039770	0.002510
10751515	56.406490	-0.113040	0.000040	0.146750	0.001740
10753734	19.406690	-0.495100	0.000070	-0.154850	0.000900
10794242	7.143710	-0.032560	0.000010	0.065890	0.000210
10809677	7.042180	0.000120	0.000260	-0.019010	0.044180
10849244	24.261680	-0.166090	0.000010	0.031180	0.001500
10857342	2.415920	0.003280	0.000010	-0.012120	0.000170
10857589	32.532530	-0.129480	0.000020	-0.134200	0.000100
10858720	0.952380	0.000050	0.000010	0.002080	0.000220
10909274	39.238140	-0.459360	0.000240	0.554460	0.000640
10920314	0.505170	0.001690	0.000070	0.012490	0.000520
10923260	7.436020	-0.374890	0.000110	0.146570	0.001970
10936427	14.361020	0.001440	0.000010	0.031540	0.000470
10937609	2.570430	0.000260	0.000060	-0.006700	0.002660
10965963	6.640240	-0.036520	0.000010	0.035310	0.000850
10973583	6.256340	-0.005500	0.000330	0.092530	0.058570
10987439	10.674600	-0.031560	0.000140	0.073960	0.020880
10991989	0.974480	-0.000360	0.000500	-0.013200	0.015180
10992733	18.525920	-0.343950	0.000060	-0.167760	0.000920
11071207	8.049640	-0.187490	0.000030	-0.173930	0.000910
11073223	11.578950	-0.234990	0.000090	-0.294410	0.001120
11076279	1.472580	0.002480	0.000040	-0.002380	0.001470
11124509	8.892910	-0.000010	0.000310	-0.071370	0.058900
11134079	1.260570	0.000040	0.000100	-0.019200	0.003870
11147276	3.133060	-0.000150	0.000020	-0.004260	0.001680
11198723	6.840310	-0.003690	0.000030	-0.003110	0.002930
11200773	2.489550	0.000340	0.000650	-0.000980	0.027580
11228612	2.980480	0.000110	0.000020	-0.000590	0.001000
11230837	0.446080	-0.000940	0.000040	-0.027100	0.001490
11232745	9.634240	-0.000090	0.000060	-0.008860	0.011780

11233911	4.959900	0.000340	0.000020	-0.001370	0.001110
11234677	1.587420	0.001650	0.000030	0.081900	0.000890
11235323	19.668400	0.000110	0.000000	-0.006860	0.000050
11249624	65.583720	-0.820240	0.000520	-0.074200	0.008720
11250867	2.252820	0.000120	0.000040	0.001950	0.002080
11252617	4.478120	0.000400	0.000420	0.002990	0.022090
11287726	4.737710	0.000400	0.000040	0.010320	0.001980
11359305	28.487120	-0.452900	0.000650	0.185800	0.007360
11391181	8.617340	-0.170300	0.000020	-0.038860	0.001970
11391667	1.083740	0.000380	0.000240	0.016840	0.019180
11401845	2.161400	0.000440	0.000040	-0.000970	0.001520
11403216	4.053250	-0.000020	0.000060	0.010320	0.001900
11404644	5.902600	0.001380	0.001420	0.032070	0.017490
11409698	10.267830	-0.234160	0.000370	-0.018340	0.003080
11455484	6.142750	0.017880	0.000360	0.022490	0.002460
11499757	12.314400	-0.202280	0.000020	-0.156470	0.000730
11502172	25.431960	-0.090750	0.000040	0.008200	0.013450
11506235	20.413060	-0.242170	0.000360	0.052320	0.027400
11615481	65.432010	-0.529600	0.000010	-0.026590	0.000430
11616200	1.718650	0.001470	0.000140	-0.002510	0.001830
11616594	14.587610	-0.000170	0.000020	-0.000430	0.000740
11619964	10.368560	-0.065010	0.000030	-0.030820	0.004200
11656302	22.264810	-0.158370	0.000070	0.011900	0.003960
11662440	13.639460	-0.326530	0.000180	-0.075030	0.007340
11671429	112.463710	-0.365270	0.000110	0.250850	0.000980
11671660	8.710340	0.002480	0.000090	-0.019150	0.003350
11704044	13.765340	0.000090	0.000000	-0.024100	0.000890
11724210	5.746020	-0.000560	0.000340	-0.016710	0.009170
11768970	15.541350	-0.225920	0.001070	-0.395040	0.010140
11769146	282.965830	-0.083210	0.000410	0.247940	0.019000
11811454	2.274630	0.000250	0.000060	0.001400	0.001920
11817750	9.762550	0.006410	0.000030	0.000000	0.003870
11826400	5.889370	0.001590	0.001530	0.033130	0.042200
11858541	5.674350	-0.055790	0.000090	0.014420	0.009700

11906217	37.910220	-0.512500	0.001360	0.140810	0.018740
11913071	3.747840	0.000580	0.000060	0.004520	0.001570
11922782	3.512930	0.000100	0.000010	0.003760	0.000710
11923819	33.159510	-0.259470	0.000060	0.135720	0.001740
11959569	46.148740	-0.128160	0.000000	0.079000	0.000060
11966557	60.297610	-0.138030	0.000130	0.201510	0.004330
11967004	1.941640	-0.000120	0.000310	0.024070	0.020470
11968490	1.078890	-0.000360	0.000090	-0.007770	0.004660
11975363	3.518420	0.000060	0.000010	0.002390	0.000500
12004679	5.042490	0.001770	0.000000	0.000600	0.000530
12004834	0.262320	0.000430	0.000030	0.003620	0.000890
12013615	8.202870	0.005460	0.000010	0.010300	0.000500
12017140	22.845240	-0.079470	0.000120	0.049480	0.006660
12022517	3.442710	0.002830	0.000060	-0.016470	0.004670
12022718	9.255330	0.000930	0.000120	0.044200	0.032330
12023089	0.623440	0.000280	0.000140	-0.006360	0.006260
12071006	6.096030	-0.000100	0.000020	0.006600	0.000800
12105785	31.953030	-0.070170	0.000210	-0.340200	0.007630
12108333	0.705450	0.000140	0.000010	-0.008580	0.000260
12164634	89.331220	-0.307280	0.000040	-0.019580	0.003720
12164751	2.630090	0.007100	0.000030	-0.022050	0.000730
12208887	53.501800	-0.013160	0.000020	-0.381520	0.002680
12217907	43.204590	-0.254300	0.000020	-0.317950	0.000190
12251779	14.844230	-0.385100	0.000070	0.121530	0.001260
12257908	2.615920	0.002580	0.000020	0.053440	0.000050
12302391	25.321730	-0.096110	0.000110	0.311800	0.003270
12306808	37.878480	-0.087350	0.000010	-0.050850	0.001300
12316447	17.907700	-0.368420	0.000110	0.009600	0.009990
12356746	1.004910	-0.000150	0.000060	-0.003460	0.002430
12367310	8.627490	0.011190	0.000200	0.067710	0.021510
12418816	1.521870	0.000170	0.000010	0.000160	0.000420
12470530	8.207260	-0.372780	0.000040	0.066510	0.001600
12557713	7.214720	-0.016360	0.000080	0.006720	0.014730
12644769	225.885000	-0.018200	0.000010	-0.123150	0.000870

12645761	5.419140	0.000140	0.000100	-0.008130	0.009110

Chapter 5
Conclusions and Future Work

5.1 Conclusions

First, we examined a sample of short-period planet candidates from *Kepler* for which both a transit and an occultation were observed. We characterized the eccentricity distribution of a sample of 50 short-period planet candidates using transit and occultation measurements from NASA's *Kepler* Mission. We evaluated the sensitivity of our hierarchical Bayesian modeling and tested its robustness to model misspecification using simulated data. We then analyzed actual data assuming a Rayleigh distribution for eccentricity, and found the posterior mode for the dispersion parameter to be $\sigma = 0.081 \pm^{0.014}_{0.003}$. We found that a two-component Gaussian mixture model for $e\cos\omega$ and $e\sin\omega$ provided a better model than either a Rayleigh or Beta distribution. Based on application of our favored model to real *Kepler* data, we found that $\sim 90\%$ of planet candidates in our sample come from a population with an eccentricity distribution characterized by a small dispersion (~ 0.01), and $\sim 10\%$ come from a population with a larger dispersion (~ 0.22).

Moreover, we investigated how the eccentricity distribution correlates with selected planet and host star parameters. We found evidence that suggests systems around higher metallicity stars and planet candidates with smaller radii may come from a more complex eccentricity distribution, but caution that these results may be limited by large uncertainty in photometrically derived stellar metallicity and a small sample size. Furthermore, we demonstrated that HB models are well-suited for characterizing the eccentricity distribution using transit and occultation data. Two broad classes of planet formation mechanisms have been proposed to explain

the formation of hot Jupiters, and studies suggest that multiple mechanisms likely contribute to the formation of hot-Jupiters. Results of these two mechanisms would likely have different distributions of orbital eccentricities, with disk migration leading to the smaller dispersion of eccentricities, and planet-planet scattering leading to a population with a larger dispersion of eccentricities. This motivates us to consider interpreting evidence of a two-component mixture model for the eccentricity distribution in terms of two formation models.

Next, we analyzed a sample of planet pairs from *Kepler* in near first-order mean motion resonances with significant transit timing variation signals. We characterized the mass-radius-eccentricity distribution of this sub-population of transiting planets using Transit Timing Variations (TTV) observations from NASA's *Kepler* mission. *Kepler's* precise measurements of transit times (Mazeh et al., 2013; Rowe & Thompson, 2015a) constrain the planet-star mass ratio, eccentricity and pericenter directions for hundreds of planets. In particularly favorable cases, strongly-interacting planetary systems allow TTVs to provide precise measurements of masses and orbital eccentricities separately (e.g., Kepler-36, Carter et al. (2012)). For each of these precisely characterized systems, there are several planetary systems harboring at least two planets near a mean motion resonance (MMR) for which TTVs provide a joint constraint on planet masses, eccentricities and pericenter directions (Hadden & Lithwick, 2014).

Unfortunately, a near-degeneracy between these parameters leads to a posterior probability density with highly correlated uncertainties. Nevertheless, the population encodes valuable information about the distribution of planet masses, orbital eccentricities and the planet mass-radius relationship. We characterized the distribution of masses and eccentricities for near-resonant transiting planets by combining a hierarchical Bayesian model with an analytic model for the TTV signatures of near-first-order-resonant planet pairs (Lithwick et al., 2012) to accelerate exploration of this complex parameter space. By developing a rigorous statistical framework for analyzing the TTV signatures of a population of planetary systems, we improved upon previous analyses. For example, our analysis includes transit timing measurements of near-resonant transiting planet pairs regardless of whether there is a significant detection of TTVs, thereby avoiding biases due to only including TTV detections.

Finally, we applied the methods from our eccentricity distribution analysis

presented in Chapter 2 on a sample of 795 eclipsing binary (EB) systems, for which we have transit and occultation projected eccentricity measurements. First, we applied a two-component Gaussian mixture HB model to the sample of EB projected eccentricities. In this HB model regime, we found that the mixture fractions and dispersions are $f_{low} = 53 \pm^{2.2}_{1.6}$ %, $f_{high} = 47 \pm^{1.8}_{1.9}$ %, and $\sigma_{low} = 0.0067 \pm^{0.0003}_{0.0003}$, $\sigma_{high} = 0.21 \pm^{0.007}_{0.005}$, respectively. Here, "low" refers to the smaller dispersion component and "high" refers to the component where EBs may have significant eccentricities. Next, we applied an HB model using a piece-wise function, where the parameters for the eccentricity distribution depend on whether the orbital period are below or above a threshold in orbital period. In this scenario, the threshold, which we call the period break point, is simultaneously inferred along with the eccentricity distribution for each sub-population below and above the break point. In the case where we allowed the eccentricity distribution on either side of the period break point to be modeled as a mixture of two Gaussian distributions, we found that the inferred period break point is at $10.74 \pm^{0.26}_{0.027}$ days. This is consistent with previous findings from Latham et al. (2002); Mayor et al. (2001), where they observed a period break point at around 10 days for EBs from disk and halo parent populations, indicative of tidal circularization operating on this population of short-period EBs.

Furthermore, the eccentricity distribution of EBs below $\sim$ 10.74 days, has population level parameters $f_{low} = 0.72 \pm^{0.024}_{0.022}$, $f_{high} = 0.28 \pm^{0.025}_{0.02}$ %, and $\sigma_{low} = 0.0055 \pm^{0.0004}_{0.0002}$, $\sigma_{high} = 0.1 \pm^{0.0047}_{0.0048}$ for the mixture fractions and dispersions, respectively. The eccentricity distribution for EBs with periods above $\sim$ 10.74 days has population level parameters $f_{low} = 0.13 \pm^{0.024}_{0.023}$, $f_{high} = 0.87 \pm^{0.027}_{0.02}$ %, and $\sigma_{low} = 0.018 \pm^{0.0035}_{0.0018}$, $\sigma_{high} = 0.26 \pm^{0.0074}_{0.011}$ for the mixture fractions and dispersions, respectively. This suggests that $\sim$ 72% of EBs below the inferred period break point are very circularized, where as $\sim$ 87% of EBs above the inferred period break point can take on a wide range in eccentricity values including some with significant eccentricities. It will be interesting to apply our analysis beyond orbital period and eccentricity in the future, perhaps investigating any correlations of eccentricity with the EB mass ratio, tidal circularization times and EB metallicity. Furthermore, an important part of the development of an HB model is testing for model robustness and refining model set up iteratively. Model comparison techniques are being developed in order to quantitatively evaluate whether the data are able to constrain

a given model better than another given model. We hope to develop practical model comparison techniques for these analyses in the future.

5.2 Future Work

This dissertation research has largely focused on sub-samples of *Kepler* planets, tackling each specific problem using hierarchical Bayesian (HB) inference. In §1.2, we explain that without properly accounting for the impact of selection effects, detection-efficiency/completeness and the false-positive-rates/reliability of the *Kepler* sample, biases will affect estimates for the occurrence rate of Earth-like planets on year long orbits around solar-like stars, $\eta_{\oplus}$, and other population level science will remain significantly less accurate. To address the problem of properly accounting for selection effects, completeness, and reliability, future research should develop a comprehensive analysis of the *Kepler* candidate sample, building a framework of hierarchical Bayesian models for probabilistic population-level analysis. This research should incorporate the forthcoming advanced *Kepler* data products and account for selection effects and biases better than previous studies. The results of these inferences can help develop tools for high-impact exoplanet population synthesis with the ability to incorporate theoretical models with ease, which are useful for testing hypotheses, planning of future science objectives, and motivating new research projects and mission design.

The goals of future research will be to achieve the following:

1. A fully marginalized frequency of small, potentially habitable planets orbiting F, G, K, and M stars, enabling a comprehensive result for the *Kepler* mission objective to discover the occurrence rate of Earth analogs.

2. An encompassing distribution linking planet and host star properties observed by *Kepler*, uncovering correlations between planet and host star properties that will support or refute planet formation theories.

3. Flexibility to incorporate data products from continued ground based observations of exoplanet systems, the upcoming Transiting Exoplanet Survey Satellite (TESS), and the European Space Agency's upcoming mission designed for astrometry (GAIA) into the developed population models, informing future mission objectives and design.

4. Exoplanet population synthesis tools based on *Kepler* data allowing for interchangeable theoretical models and a platform to compare existing population models to the *Kepler* sample, providing a framework for theorists to compare theoretical models to the *Kepler* data.

Finalizing the *Kepler* mission objectives in a comprehensive Bayesian framework is not only paramount for the full realizations of the *Kepler* mission but sets a new standard for the quality of analysis required to draw conclusions from data in future exoplanet surveys. The computing resources and theory of today make it possible to implement these techniques as a gold standard in future population analysis efforts.

The *Kepler* pipeline detects transiting planets that **1)** have enough SNR to be detected and **2)** transit their host star as seen by *Kepler*. This introduces selection effects into the sample of planets. To reach towards studying the exoplanet population as a whole, we must work to account for these selection effects in our population studies. Future research should go beyond previous studies by accounting for these selection effects in two ways: first by testing robust population inference models; and secondly, by incorporating the forthcoming advanced data products from the *Kepler* mission planned for release on or before September 2017. These data products will include a catalog of exoplanet candidates **1)** obtained from a uniform search process using auto-vetting (McCauliff et al., 2015; Catanzarite et al., 2015), **2)** the associated measurements of pipeline *completeness* (detection efficiency) (Christiansen et al., 2013) and *reliability* (false positive rate) for the *Kepler* pipeline (Désert et al., 2015), and **3)** *posterior distributions* of uniform planet and star properties that serve to properly account for the *non-negligible measurement uncertainties* (Rowe & Thompson, 2015a) and are a record of the uncertainty introduced by unaccounted-for systematics and stochastic processes in nature. These data products will be optimized for statistical analyses built on a Bayesian framework - the state-of-the art in probabilistic data analysis.

Constructing a sophisticated probabilistic population analysis using the upcoming data products described above is paramount for the full and accurate realization of the *Kepler* mission's science goals. Developing comprehensive *Kepler* data products that are optimized for a Bayesian framework has been a large collaborative effort years in the making, and these products are only now nearing completion. With this advancement on the horizon, timely development of tested

state-of-the-art population-level analysis tools are needed. Using the upcoming *Kepler* data products is necessary to properly calculate the occurrence rates of the *Kepler* sample, among other population distributions of interest. The completed baseline data products being developed by the *Kepler* science team will include the *posterior distributions* of the properties describing planet candidate and host stars obtained through transit photometry and follow-up observations. These *posterior distributions* are probabilistic estimates for the planet candidate and host star parameters, encapsulating the information regarding systematics and noise.

By using the upcoming catalog of planet candidate and host star posterior distributions in a hierarchical Bayesian Framework and employing well-motivated noise models, the planet candidate population can be characterized in a robust probabilistic manner. Furthermore, characterizing the *Kepler* planets and their occurrence rates with more accuracy is possible with hierarchical Bayes because incorporating the cataloged posteriors of the parameter measurements being studied factors in their large fractional measurement uncertainty (e.g., the ratio of the %68.3 credible interval to the median value), and provides the framework to properly characterize the noisy incomplete *Kepler* data.

Jessie Christiansen of the *Kepler* science team has done work to account for many different effects that can impact the empirical detection efficiency of the current *Kepler* pipeline (Christiansen et al., 2013). This can now be incorporated into an advanced auto-vetting process. The current catalog of *Kepler* planet candidates was constructed using a pipeline that selected many threshold crossing events (or incidents of potential transits) (Tenenbaum et al., 2013, 2012; Jenkins et al., 2010b). The Threshold Crossing Event Review Team then went through and selected cases that appeared to be real transits (Batalha et al., 2010). Although this method is best for detecting planets, it created a catalog of planet candidates that were classified by different people with different biases. To solve this problem we will use the upcoming catalog of planet candidates, which will be selected via an auto-vetter. The new auto-vetter can be used to better classify the threshold-crossing events and create a more *uniformly selected* list of planet candidates (Catanzarite et al., 2015; Jenkins et al., 2014). This auto-vetter, although worse at finding planet candidates than humans, is reproducible, removing many of the human selection effects in the current catalog. All current population studies of the *Kepler* sample to date ignore this issue.

In Chapter 2, we show that the fractional uncertainties and population size impacts the ability of data to constrain particular hierarchical Bayesian models in eccentricity parameter space. More complex models may require more observations, more precise measurements, or both. For example, in Chapter 4, we successfully constrain a period-eccentricity distribution model for eclipsing binary stars from *Kepler*, using a sample of 795 precisely measured projected eccentricities. We attempted to apply this model the the projected eccentricities of a sample of 50 short-period planet candidates from *Kepler* in Chapter 2, however, the small sample size limited our ability to infer the period-eccentricity distribution. With this in mind, the procedure for developing a comprehensive population model, incorporating the full catalog of *Kepler* planet candidates, can be structured to harness the power of the large sample size, which could support more complex models.

Furthermore, by initially characterizing the eccentricity distribution of a subset of *Kepler* planet candidates in a probabilistic manner, future research can include the impact of eccentricity in future models. Eccentricity has largely been ignored for *Kepler* population studies, but the impact can be non-negligible and may influence the empirically measured detection efficiency of the *Kepler* planet candidates Kipping (2014b). Planets on eccentric orbits can have shortened or lengthened transit durations, which transit injection methods may need to account for in the future. The research in this dissertation can inform advancements in the characterization of the detection efficiency via the incorporation of the eccentricity distribution.

Many practical choices for population models exist. Future research should aim to test different population models and the impact of the model on the results. Robust models will be less sensitive to outliers in the data, and posterior solutions will be insensitive to small departures in the choice of the model. To accomplish this, models such as a single Normal distribution, single beta, or power law (Pareto) distribution can be used when characterizing planet and host star parameters. The complexity of the models can then be increased gradually, trying mixture models to look for evidence of multiple populations as implemented in this dissertation research. In some cases, Gaussian process models, a nonparametric Bayesian approach, can be used to model the detailed structure in the data. After incorporating these findings, transitioning back to simpler models can improve computational tractability. With

this procedure, the population models can be refined to increase the number of variables being modeled in a robust manner, and capturing correlations between these variables. Using the results of the full joint-posteriors, occurrence rates can be calculated by marginalizing over all other parameters to obtain results with improved accuracy. The discovery of correlations between planet and host star properties within the joint posterior can help to support or refute planet formation theories.

There are multiple approaches for developing population synthesis models:

(1) Use simple analytical relations as physical models to infer information about planet parameters that are not observable with *Kepler*. These simple relations will relate model parameters to transit observables from the *Kepler* data. Examples of this approach from the literature include: the envelope fraction in Wolfgang & Lopez (2015), transit timing variation amplitude for mass-radius-eccentricity relation in Shabram et al. (2015 *accepted to ApJ*), and the rocky gaseous transition in Rogers (2015a).

(2) Use semi-analytical approximations for planet formation models from theorists (e.g., Schlichting, 2014; Ida & Lin, 2004) as physical models in a hierarchical bayesian model. By coupling these analytic planet formation models with the inference models, we can infer the values of the tunable parameters in these analytical models that give a good match to *Kepler* data.

(3) Forward model: Obtain simulated catalogs of planets from detailed theoretical population synthesis models (e.g., Kenyon & Bromley, 2015; Raymond et al., 2014). The theoretical population synthesis models used in this forward process will be computationally expensive and poorly approximated by analytical relations. Future research can convert these catalogs into simulated *Kepler* data and run these simulated observations through inference models developed for the *Kepler* data. We can then compare the population level parameters of the simulated data to the population level parameters of the updated *Kepler* data. These comparisons can help tune and constrain theoretical planet formation simulations. Furthermore, these inference models can serve as analytical approximations for these computationally expensive planet formation models.

In this dissertation, we use summary statistics of the planet and host star properties. Using summary statistics and posterior distributions of probabilistic parameter estimates in lieu of pixel-level data renders the hierarchical model more

computationally tractable. This allows us to treat each observation of each system as independent variables. In reality, the pixel-level observations of each system are not independent (e.g., data from one pixel can be affected by photons hitting other pixels on the same detector from electronic artifacts), however, using these summary statistics and in turn posteriors is an excellent approximation. In future research, we encourage the use of full posterior distributions of probabilistic parameter estimates, as the reporting of full posteriors in the literature and cataloged data becomes the norm.

Accounting for selection effects from geometrical transit probability, pipeline sensitivity, reliability, and the large measurement uncertainties is critical when studying the *Kepler* planet candidates as a representative sample of all planets. In future research, an innovative aspect will be including the large measurement uncertainties in the data. An example case is illustrated in Figure 1 below. Figure 1 shows a toy model of the detection efficiency function versus the SNR. In previous studies, point estimates of the SNR were used in the detection efficiency function. This introduces many problems because in reality the SNR is imprecisely known. The figure shows how using a point estimate instead of the posterior distribution for the SNR leads to systematic biases. When including the detection efficiency in hierarchical Bayesian models, we need to marginalize over the SNR to properly include the large measurement uncertainty. This effect is particularly pronounced for planets near the detection threshold, and thus is critically important when pushing the limits of *Kepler* data to measure the frequency of Earth-size planets.

The *Kepler* mission objective to accurately characterize the occurrence rate of Earth analogs has yet to be accomplished. Many studies have chiseled away at this goal but have not yet yielded a complete analysis. Future research will be able to complete this analysis at a state-of-the-art level, exactly when the advanced data products become available. Furthermore, many outstanding questions about planet formation remain. By searching for and examining correlations between planet and host star properties, we can start to gain insight about the mechanisms that impact how a planetary system plays out. For instance, it would be interesting to learn if there exists fundamental differences between planetary systems that form around cool dim stars versus more massive hotter stars (Burke et al., 2015; Moorhead et al., 2011). The advanced *Kepler* data will be able to place empirical constraints on these mechanisms.

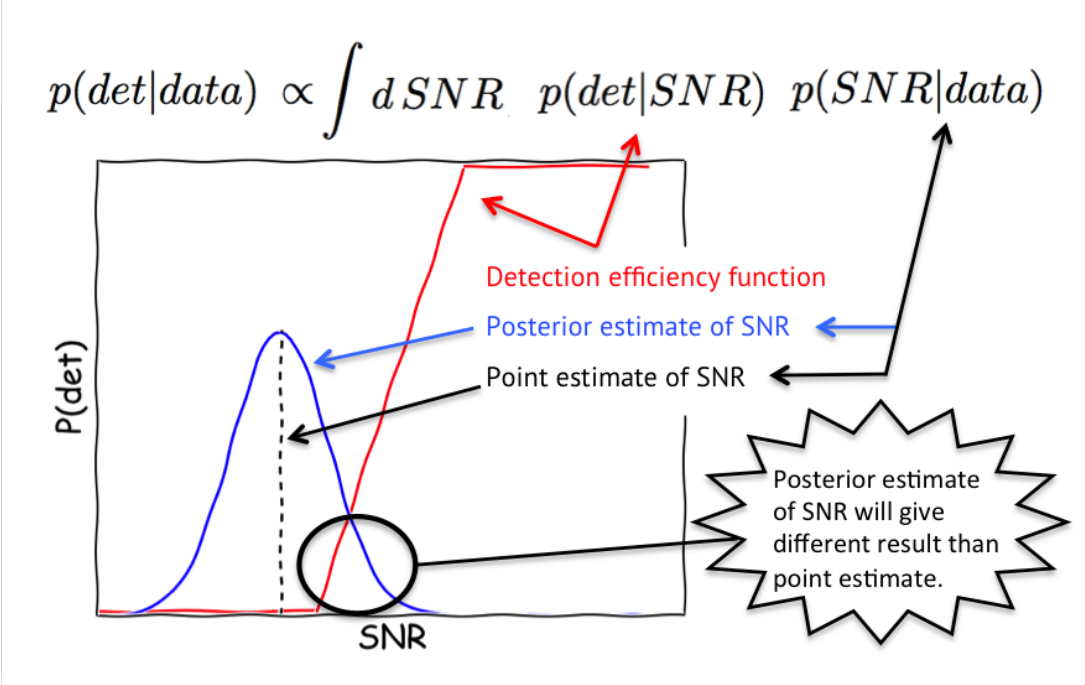

Figure 5.1: Probabilistic detection efficiency toy model. The red curve shows a detection efficiency function, the black dashed line shows a point estimate of the SNR, and the blue curve shows the full posterior measurement of the SNR for this example observation. Using the posterior for the measured SNR instead of a point estimate avoids biases in the inference models.

The inference models developed in future research could provide the means for theorists to synthesize exoplanet populations for their studies. An example would be understanding planet density. The *Kepler* sample provides us with information about planet radii, but only some systems have measured masses (calculated from transit-timing-variations for planets in near-mean-motion resonances, or from radial velocity observations) (Weiss & Marcy, 2014). Theoretical models incorporated into the population inference models will help to inform us about the distributions of planet bulk density and composition regimes (Wolfgang & Lopez, 2015). These studies provide clues as to whether the planet has formed in situ or migrated.

By understanding the types of planetary systems that exist around different types of stars, we can inform future mission objectives and design. A straight forward example for future transit observations is the SNR for the observed planet-to-star radius ratio. If planetary systems around most stars are comparable in size and orbital architecture, then stars with smaller radii, such as M stars, would

typically be better targets for detecting planetary systems. On the other hand, if planetary systems scale with host star surface density (Youdin & Kenyon, 2013; Johnson et al., 2010), the frequency of small planets around M stars that are amenable to Doppler/transit spectroscopy characterization might be low. M stars peak in the infrared where low energy photons and increased read noise make observations more challenging than in the optical. Therefore, characterizing even the nearest such planets with existing/near future observatories may be impractical.

Furthermore, transit spectroscopy is practical for planets with puffy atmospheres and low surface gravities. Although many targets in the *Kepler* field are too faint for current transit spectroscopy follow-up, understanding the distribution in planet density and composition could provide insight for future atmospheric studies with missions such as the James Webb Space Telescope (Seager et al., 2009). Moreover, the *Kepler* sample will be much larger and more uniform than future exoplanet observation missions. For instance, TESS will find planets around brighter targets but will not characterize the population well (Sullivan et al., 2015). By using the population information gained through the robust analysis of the *Kepler* sample, planets found by TESS can be better selected for follow-up observations and atmospheric characterization.

Bibliography

Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, PASP, 125, 989

Albrecht, S., Winn, J. N., Johnson, J. A., et al. 2012, ApJ, 757, 18

Barnes, J. W. 2007, PASP, 119, 986

Batalha, N. M., Rowe, J. F., Gilliland, R. L., et al. 2010, ApJ, 713, L103

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24

Bodenheimer, P., Hubickyj, O., & Lissauer, J. J. 2000, Icarus, 143, 2

Boley, A. C. 2009, ApJ, 695, L53

Boley, A. C., Morris, M. A., & Ford, E. B. 2014, ApJ, 792, L27

Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011a, ApJ, 728, 117

—. 2011b, ApJ, 736, 19

Borucki, W. J., Koch, D. G., Batalha, N., et al. 2011c, in American Astronomical Society Meeting Abstracts #218, #218.01

Boss, A. P. 1997, Science, 276, 1836

Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ, 142, 112

Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al. 2013, PASP, 125, 889

Buchhave, L. A., Latham, D. W., Johansen, A., et al. 2012, Nat, 486, 375

Burke, C. J. 2008, ApJ, 679, 1566

Burke, C. J., Bryson, S. T., Mullally, F., et al. 2014, ApJS, 210, 19

Burke, C. J., Mullally, F., Christiansen, J., et al. 2015, in AAS Meeting Abstracts, Vol. 225, #122.04

Cameron, A. G. W. 1978, Moon and Planets, 18, 5

Carter, J. A., Agol, E., Chaplin, W. J., et al. 2012, Science, 337, 556

Catanzarite, J., Jenkins, J. M., Burke, C. J., McCauliff, S. D., & Kepler Science Operations Center. 2015, in AAS Meeting Abstracts, Vol. 225, #257.33

Chatterjee, S., & Ford, E. B. 2015, ApJ, 803, 33

Christiansen, J. 2014, in AAS Meeting Abstracts, Vol. 224, #113.05

Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279

Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2013, ApJS, 207, 35

Claret, A., & Bloemen, S. 2011, A&A, 529, 75

Dawson, R. I. 2014, ApJ, 790, L31

Dawson, R. I., & Johnson, J. A. 2012, ApJ, 756, 122

Dawson, R. I., Johnson, J. A., Morton, T. D., et al. 2012, ApJ, 761, 163

Dawson, R. I., & Murray-Clay, R. A. 2013, ApJ, 767, L24

de Kort, J. J. M. A. 1954, Ricerche Astronomiche, 3, 109

Demory, B.-O. 2014, ApJ, 789, L20

Demory, B.-O., & Seager, S. 2011, ApJS, 197, 12

Désert, J.-M., Charbonneau, D., Torres, G., et al. 2015, ApJ, 804, 59

Dong, S., & Zhu, Z. 2013, ApJ, 778, 53

Fabrycky, D., & Tremaine, S. 2007, ApJ, 669, 1298

Fabrycky, D. C., & Winn, J. N. 2009, ApJ, 696, 1230

Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., et al. 2014, ApJ, 790, 146

Fang, J., & Margot, J.-L. 2012, ApJ, 761, 92

Ford, E. B. 2014, Proceedings of the National Academy of Science, 111, 12616

Ford, E. B., Quinn, S. N., & Veras, D. 2008, ApJ, 678, 1407

Ford, E. B., & Rasio, F. A. 2006, ApJ, 638, L45

—. 2008, ApJ, 686, 621

Ford, E. B., Rowe, J. F., Fabrycky, D. C., et al. 2011, ApJS, 197, 2

Ford, E. B., Fabrycky, D. C., Steffen, J. H., et al. 2012a, ApJ, 750, 113

Ford, E. B., Ragozzine, D., Rowe, J. F., et al. 2012b, ApJ, 756, 185

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81

Gelman, & Rubin. 1992, Statistical Science, 7, 457

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004, From Disks to Planets, ed. Chapman & Hall

Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, ApJS, 197, 6

Gillon, M., Demory, B.-O., Benneke, B., et al. 2012, A&A, 539, A28

Hadden, S., & Lithwick, Y. 2014, ApJ, 787, 80

Hansen, B. M. S., & Murray, N. 2015, MNRAS, 448, 1044

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166

Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, Science, 330, 653

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15

Huber, D., Chaplin, W. J., Christensen-Dalsgaard, J., et al. 2013, ApJ, 767, 127

Huber, D., Silva Aguirre, V., Matthews, J. M., et al. 2014, ApJS, 211, 2

Ida, S., & Lin, D. N. C. 2004, ApJ, 604, 388

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010a, ApJ, 713, L120

—. 2010b, ApJ, 713, L87

Jenkins, J. M., McCauliff, S., Burke, C., et al. 2014, in IAU Symposium, ed. N. Haghighipour, Vol. 293, 94–99

Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, PASP, 122, 905

Kane, S. R., Ciardi, D. R., Gelino, D. M., & von Braun, K. 2012, MNRAS, 425, 757

Kenyon, S. J., & Bromley, B. C. 2015, ApJ, 806, 42

Kipping, D. M. 2011, MNRAS, 416, 689

—. 2014a, MNRAS, 444, 2263

—. 2014b, MNRAS, 444, 2263

Kipping, D. M., Bastien, F. A., Stassun, K. G., et al. 2014, ApJ, 785, L32

Kley, W., & Nelson, R. P. 2012, ARA&A, 50, 211

Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJ, 713, L79

Latham, D. W., Stefanik, R. P., Torres, G., et al. 2002, AJ, 124, 1144

Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, Nat, 470, 53

Lithwick, Y., & Wu, Y. 2012, ApJ, 756, L11

Lithwick, Y., Xie, J., & Wu, Y. 2012, ApJ, 761, 122

Lopez, E. D., & Fortney, J. J. 2014, ApJ, 792, 1

Loredo, T. J. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 371, Statistical Challenges in Modern Astronomy IV, ed. G. J. Babu & E. D. Feigelson, 121

Lucy, L. B. 2013, A&A, 551, A47

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20

Matsumura, S., Takeda, G., & Rasio, F. A. 2008, ApJ, 686, L29

Mayor, M., Udry, S., Halbwachs, J.-L., & Arenou, F. 2001, in IAU Symposium, Vol. 200, The Formation of Binary Stars, ed. H. Zinnecker & R. Mathieu, 45

Mazeh, T., Nachmani, G., Holczer, T., et al. 2013, ArXiv e-prints, arXiv:1301.5499

McCauliff, S. D., Jenkins, J. M., Catanzarite, J., et al. 2015, ApJ, 806, 6

Moorhead, A. V., Ford, E. B., Morehead, R. C., et al. 2011, ApJS, 197, 1

Morton, T. D., & Johnson, J. A. 2011, ApJ, 729, 138

Morton, T. D., & Winn, J. N. 2014, ApJ, 796, 47

Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A., & Teyssandier, J. 2011, Nat, 473, 187

Naoz, S., Farr, W. M., & Rasio, F. A. 2012, ApJ, 754, L36

Nesvorný, D., & Morbidelli, A. 2008, ApJ, 688, 636

Parviainen, H., Deeg, H. J., & Belmonte, J. A. 2013, A&A, 550, A67

Payne, M. J., Deck, K. M., Holman, M. J., & Perets, H. B. 2013, ApJ, 775, L44

Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013a, Proceedings of the NAS, 110, 19273

Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013b, ApJ, 770, 69

Plavchan, P., Bilinski, C., & Currie, T. 2014a, PASP, 126, 34

—. 2014b, PASP, 126, 34

Plummer, M. 2003, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), ISSN 1609-395X.

Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, Icarus, 124, 62

Prša, A., Guinan, E. F., Devinney, E. J., et al. 2008, ApJ, 687, 542

Prša, A., Guinan, E. F., Devinney, E. J., et al. 2012, in IAU Symposium, Vol. 282, IAU Symposium, ed. M. T. Richards & I. Hubeny, 271–278

Prša, A., Batalha, N., Slawson, R. W., et al. 2011, AJ, 141, 83

Ragozzine, D., & Wolf, A. S. 2009, ApJ, 698, 1778

Rasio, F. A., & Ford, E. B. 1996, Science, 274, 954

Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, arXiv:1310.0696

Raymond, S. N., Kokubo, E., Morbidelli, A., Morishima, R., & Walsh, K. J. 2014, Protostars and Planets VI, 595

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 20

Rogers, L. A. 2015a, ApJ, 801, 41

—. 2015b, ApJ, 801, 41

Rogers, L. A., & Seager, S. 2010, ApJ, 712, 974

Rowe, J. F., & Thompson, S. E. 2015a, ArXiv e-prints, arXiv:1504.00707

—. 2015b, ArXiv e-prints, arXiv:1504.00707

Sanchis-Ojeda, R., Winn, J. N., Marcy, G. W., et al. 2013, ApJ, 775, 54

Schlichting, H. E. 2014, ApJ, 795, L15

Seader, S., Jenkins, J. M., Tenenbaum, P., et al. 2015, ApJS, 217, 18

Seager, S., Deming, D., & Valenti, J. A. 2009, Transiting Exoplanets with JWST, ed. H. A. Thronson, M. Stiavelli, & A. Tielens, 123

Shabram, M., Demory, B.-O., Cisewski, J., Ford, E. B., & Rogers, L. 2015, ArXiv e-prints, arXiv:1511.02861

Silburt, A., & Rein, H. 2015, ArXiv e-prints, arXiv:1508.04646

Slawson, R. W., Prša, A., Welsh, W. F., et al. 2011, AJ, 142, 160

Steffen, J. H., & Hwang, J. A. 2015, MNRAS, 448, 1956

Steffen, J. H., Ragozzine, D., Fabrycky, D. C., et al. 2012a, Proceedings of the National Academy of Science, 109, 7982

Steffen, J. H., Fabrycky, D. C., Ford, E. B., et al. 2012b, MNRAS, 421, 2342

Steffen, J. H., Fabrycky, D. C., Agol, E., et al. 2013, MNRAS, 428, 1077

Sterne, T. E. 1940, Proceedings of the National Academy of Science, 26, 36

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, ArXiv e-prints, arXiv:1506.03845

Tenenbaum, P., Christiansen, J. L., Jenkins, J. M., et al. 2012, ApJS, 199, 24

Tenenbaum, P., Jenkins, J. M., Seader, S., et al. 2013, ApJS, 206, 5

—. 2014, ApJS, 211, 6

Torres, G., Andersen, J., & Giménez, A. 2010, , 18, 67

Torres, G., Winn, J. N., & Holman, M. J. 2009, in IAU Symposium, Vol. 253, IAU Symposium, ed. F. Pont, D. Sasselov, & M. J. Holman, 482–485

Valencia, D., Guillot, T., Parmentier, V., & Freedman, R. S. 2013, ApJ, 775, 10

Valsecchi, F., & Rasio, F. A. 2014, ApJ, 786, 102

Van Eylen, V., & Albrecht, S. 2015, ApJ, 808, 126

Wang, J., & Ford, E. B. 2011, MNRAS, 418, 1822

Weiss, L. M., & Marcy, G. W. 2014, ApJ, 783, L6

Weiss, L. M., Marcy, G. W., Rowe, J. F., et al. 2013, ApJ, 768, 14

Winn, J. N. 2010, ArXiv e-prints, 1001.2010, arXiv:1001.2010

Wolfgang, A., & Lopez, E. 2015, ApJ, 806, 183

Wolfgang, A., Rogers, L. A., & Ford, E. B. 2015, ArXiv e-prints, arXiv:1504.07557

Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, PASP, 123, 412

Wu, Y., & Lithwick, Y. 2013, ApJ, 772, 74

Youdin, A. N., & Kenyon, S. J. 2013, From Disks to Planets, ed. T. D. Oswalt, L. M. French, & P. Kalas, 1

Vita

Megan Iris Shabram

440 Davey Lab
University Park, PA 16802, USA
Cell (760) 554-8561 ** Office (814) 863-5565
mus336@psu.edu
https://www.linkedin.com/in/meganshabram

Education

Doctor of Philosophy, *Astronomy and Astrophysics*
Pennsylvania State University, University Park, PA, **expected 2015**
Concentration: Exoplanet Formation Theory

Master of Science, *Astronomy and Astrophysics*
University of Florida, Gainesville, FL, **May 2012**
Concentration: Exoplanet Formation Theory

Bachelor of Science, *Physics*
University of California, Santa Cruz, CA, **March 2009**
Concentration: Astrophysics

Selected Research Publications

- Shabram, M., Demory, B.-O., Ford, E., et al. 2015, The Astrophysical Journal, accepted
- Shabram, M., Boley, A. 2013, The Astrophysical Journal, 767, 63
- Shabram, M., Fortney, J., et al. 2011, The Astrophysical Journal, 727, 65

Selected Presentations

- Shabram, M., Ford, E., et al. 2014, Extreme Solar Systems III 2015
- Shabram, M., Ford, E., et al. 2014, ExoStats2014
- Shabram, M., Ford, E., et al. 2013, Kepler Science Conference
- Shabram, M., Fortney J. J., et al. 2011, AAS meeting 217 ♯402.04