

A Unified Bayesian Study of Exoplanet Populations: From Survey Biases to Circumbinary Systems in Kepler and TESS

Ms. Sournamalya Bhavani , Deepika Sree. A , Adiya Premjit
Star Labs Researches

Abstract - Exoplanets, planets that dwell outside our solar system and orbit around a sun-like star , are also free-flowing planets. The huge trend of discovering exoplanets since the 90's has expanded and today we are discovering and processing a huge volume of new information as we speak. In this work we present a statistical analysis of exoplanet occurrence and the host star parameters. By utilizing the data from the NASA Exoplanet archive and also combining the telescopic data from Kepler and TESS survey , we have constructed a unified dataset and model planet occurrence and their count process using a Bayesian regression framework. We employ two hierarchical models- Bayesian and poisson models to quantify the dependency of the planet occurrence on the key to stellar properties , which are metallicity, effective temperature, T_{eff} , mass, and radius.

The model's performance is evaluated through posterior predictive checks and information criteria , so that it allows a comparison. The comparison is between alternative statistical formulations. In order to account for survey- and method-specific observational biases, including differences between transit and radial velocity detections, we incorporate detection efficiency corrections into the modeling framework, enabling a consistent comparison between Kepler and TESS. For the further characterisation of the planetary populations, we incorporate mass–radius relationships to distinguish between different classes of planets, such as super-Earths and sub-Neptunes, and to separate small and giant planets. We extend the analysis to include planets in binary star systems (circumbinary or “Tatooine” planets) to investigate how stellar multiplicity influences planet occurrence. Furthermore, we examine correlations between stellar metallicity and planetary properties, with implications for atmospheric characterisation and potential habitability.

We demonstrate a novel empirical scaling relations between planetary radius and stellar characteristics that leads to the discovery of a survey-dependent scaling with stellar radius and a minor reliance on stellar effective temperature ($R_p \propto T_{\text{eff}}^{0.80 \pm 0.06}$). Specifically, TESS shows a much stronger association ($\alpha = 1.04$ after bias correction) than Kepler ($\alpha \approx 0.40$), this indicates that survey design and observational biases are also important in determining estimated population trends. We demonstrate that these discrepancies are only partially due to observational biases by using detection-efficiency weighting, suggesting fundamental differences in the underlying stellar populations. By expanding the research to circumbinary systems, we also discover that although these planets make up a minor portion of the sample (~0.8%), they have systematically bigger radii and different scaling behaviour from single-star systems, indicating distinct creation or evolutionary processes. Furthermore, we verify that planetary mass and star metallicity are positively correlated, which is in line with the core accretion hypothesis.

Our results show a statistically significant relationship between the properties of the host star and planet occurrence, with the metallicity and stellar mass as important predictors. We also find an interesting factor that the survey-dependent selection effects introduce measurable biases in inferred distributions, highlighting the importance of accounting for detection mechanisms in population studies. This approach actually improves the predictive performances and robust uncertainty estimates compared to simpler models. Our work provides a reproducible and extensible framework for the exoplanet analysis while emphasizing over the bayesian interference and with the awareness modelling to improve the astrophysical conclusion. One of the important factors about our work is that our work also has an interactive dashboard for all the astronomers out there to analyze the exoplanet and their host star properties, a special side for our blooming astronomers to learn about exoplanets enjoyably.

Keywords: exoplanets – Bayesian inference – stellar properties – detection bias – planetary multiplicity – mass–radius relation

INTRODUCTION:

The first exoplanet orbiting a sun-like star was discovered in 1995 . Since then we have found a lot of exoplanets which revolutionized the exoplanet area. By observational and large-scale surveys ,we can understand the exoplanets' characteristics , nature, their occurrence and so on . It's actually very fascinating that these discoveries help us in understanding the universe better everyday. The earlier discoveries majoritarily used Radial velocity method , which revealed the presence of massive planets through Doppler shifts in stellar spectra (**Marcy & Butler 2005**). The transit surveys, especially by NASA's famous Kepler mission , helped us in identifying and studying the small planets like Kepler-37B which has a radius about 0.31 times of earth's radius. They mainly helped us by providing statistically significant samples for population-level studies. The very recent expansion of the TESS mission

by targeting bright, nearby stars, and enabling detailed follow-up observations and comparative analyses across different stellar environments. But the foremost important question about the Exoplanet is by targeting bright, nearby stars, enabling detailed follow-up observations and comparative analyses across different stellar environments. The most well established study conducted by **Fischer & Valenti (2005)**, shows that the correlation between stellar metallicity and the occurrence of giant planets, They also helped us in understanding the planet occurrence increases strongly with stellar metal abundance. Then the later studies on this category have extended the relationship of the planet sizes and their orbital configuration showing that metal-rich stars exhibits to host a greater diversity of planetary systems (**Petigura et al. 2017**) mean while the small planets exhibits more complex and sometimes weaker dependencies on metallicity (**Mulders et al. 2016**). These observations have provided a lot of new theories on planet formations, and help us in studying the core accretion paradigm, in which the higher metallicity environments contribute to the formation of massive planetary cores.

Despite the many advanced detection techniques, the population studies are substantially affected by the observing biases that are arising from multiple different techniques. The transit survey such as Kepler, TESS are essentially biased towards the short period planets that has favourable orbital inclinations, while radial velocity surveys advantageously detect massive planets that induce measurable stellar motion. These selection effects lead to systematic differences in inferred planet populations if not properly accounted for. For instance the super giants that are close around the metal rich stars that may be amplified by detection biases inherent to these methods (**Osborn et al. 2020**). As a result combining datasets from multiple detection techniques requires careful statistical treatment to ensure consistent population inference.

The recent efforts in this demographics has seen a trend of increased reliance on Bayesian statistical methods to address these challenges. This particular framework allows the incorporation of observational uncertainties, hierarchical structure, and prior knowledge, making them particularly well-suited for diverse datasets. The hierarchical Bayesian model gives the separation of population-level trends from survey-specific effects, providing a more robust approach to modeling exoplanet occurrence. These methods have helped us in studying the occurrence rates, planetary mass distributions, and correlations with stellar properties, demonstrating their effectiveness in extracting astrophysical signals from biased observations.

In addition to the host star correlations, the structure of planetary systems—including multiplicity and mass distribution—provides further insight into planet formation and evolution. Many observations have shown us that a fraction of stars host multiple planets, often arranged in compact and dynamically ordered systems (**Lissauer et al. 2011; Fabrycky et al. 2014; Winn & Fabrycky 2015**). The mass and the radius relationship of the exoplanet has emerged as a vital diagnostic for discriminating between different types of planets, such as super-Earths and sub-Neptunes, and inferring their internal compositions (**Weiss & Marcy 2014; Chen & Kipping 2017; Fulton et al. 2017**). The extensive study of planetary structures and compositions has advanced as a result of Bayesian and interior modelling (**Bloot et al., 2023**). Furthermore, researching planets in binary star systems (circumbinary or "Tatooine" planets) provides an exceptional opportunity to investigate the effect of stellar multiplicity in planet formation and orbital stability (**Doyle et al. 2011; Welsh et al. 2012**). Recent research has improved our understanding of planetary system topologies and occurrence rates by combining survey immense datasets with new statistical techniques (**Rosenthal et al. 2021 Pinamonti et al., 2022; Kunitomo & Matthews 2020**). Recent findings from current surveys have increased the diversity of such systems and highlighted the importance of combining detection techniques in validating planetary designs (**Stefánsson et al., 2025**).

Now, in our work we provide a unified statistical framework for exoplanet population analysis that assesses data from multiple detection methods, including transit and radial velocity observations. By using the data from NASA exoplanet archive, we are able to construct a dataset of exoplanets along with their host star properties. Planet occurrence is modelled as a count process within a hierarchical Bayesian framework that includes essential star properties such as metallicity, effective temperature, mass, and radius. Importantly, we explicitly account for detection method-dependent selection effects using bias correction terms, allowing for consistent comparisons across diverse datasets (**Foreman-Mackey et al. 2014; Hogg et al. 2010**).

Our approach goes beyond occurrence rates to incorporate statistical modelling of planetary system multiplicity and planet classification using mass-radius connections. We also analyse how the star metallicity influences planetary populations, with outgrowth for the atmosphere characterisation and potential habitability (**Fischer & Valenti 2005; Petigura et al. 2017**). By integrating multiple detection methods, stellar properties, and statistical techniques.

DATA

The dataset used in this study is obtained from the NASA Exoplanet Archive (**Akeson et al. 2013**), accessed programmatically using the Astroquery interface and direct archive queries. We use data products and parameter choices that are in line with recent large-scale analyses of planetary populations to provide accurate relevance to current exoplanet demographics research (**Kunimoto et al. 2022; Bryson et al. 2021; Zink et al. 2021**). By combining verified exoplanets and host star characteristics from several archive tables, we create a single dataset. In order to understand the survey-dependent biases, we choose planets found by the Kepler (**Borucki et al. 2010**) and TESS (**Ricker et al. 2015**) missions for comparative population analysis. These missions differ greatly in observational strategy, completeness, and target selection (**Kunimoto & Matthews 2022**).

In order to construct a sample for the population level analysis, we have utilized the dataset from famous missions the Kepler (**Borucki et al. 2010**) and TESS (**Ricker et al. 2015**) missions by using the `disk_facility` field. These surveys have a compelling difference in their observations, limit on completeness, target selection and their functions. The Kepler mission provides us a deep long baseline observation of a fixed field, which makes them highly sharp to small and long-period planets, meanwhile the TESS mission focuses on bright, nearby stars across the sky and is optimized for detecting short-period planets. Therefore the combination of these surveys enables a broader sampling of planetary systems while introducing survey-dependent biases that must be accounted for in statistical analysis (**Kunimoto & Matthews 2022; Bryson et al. 2021**).

The planetary name (`pl_name`), host star identifier (`hostname`), discovery facility (`disc_facility`), discovery method (`discovery_method`), discovery year (`disc_year`), stellar effective temperature (`st_teff`), stellar metallicity (`st_met`), stellar mass (`st_mass`), stellar radius (`st_rad`), system distance (`sy_dist`), and system-level characteristics such as the number of planets (`sy_pnum`) and stars (`sy_snum`) are extracted from the `pscomppars` table. Both stellar and system structures are fully described by these factors.

In addition to the survey selection, the dataset merges the exoplanets detected through both transit and radial velocity (RV) methods. These techniques explore complementary regions of parameter space and are characterised by distinct observational biases. The transit detections are slanted more toward planets with short orbital periods and favourable geometric alignments, whereas the radial velocity detections are sharper for massive planets capable of inducing measurable stellar reflex motion. Recent research has thoroughly examined how these selection processes affect inferred exoplanet populations (**Hsu et al. 2019; He et al. 2019; Yang et al. 2023**). We create a single dataset by merging transit and RV detections, which allows for comparison analysis between detection techniques and incorporates bias corrections in later modelling.

Our analysis is completely focused on the key host star parameters that are known to have an impact on system design and planet occurrence. These consist of star mass, stellar radius, metallicity ($[Fe/H]$), and stellar effective temperature (T_{eff}). To enable bias-aware modelling, other system-level factors like distance and discovery techniques are also kept. To provide consistency and dependability, a standardised cleaning workflow is used to preprocess the dataset. Also, to provide a distinct depiction of every stellar system, duplicate host stars are eliminated. Median imputation is used to address missing values in important parameters, particularly system distance (`sy_dist`) and stellar metallicity (`st_met`). Due to inadequate coverage, the highly sparse variables like stellar rotation period (`st_rotp`) are not included in this analysis. To preserve physical consistency, non-physical values—such as negative or zero parameters—are eliminated.

The Z-score normalisation, which is defined by subtracting the mean and dividing by the standard deviation, is used to standardise continuous variables in order to increase numerical stability in statistical modelling. The stellar effective temperature (`st_teff`), metallicity (`st_met`), stellar mass (`st_mass`), and stellar radius (`st_rad`) are all subject to this normalisation. Then, in order to facilitate population-level analysis, a number of derived features are created. The number of confirmed planets connected to each host star is known as planetary multiplicity. The following bins are used to classify stellar metallicity into discrete classes: low ($[Fe/H] < -0.5$), sub-solar ($-0.5 \leq [Fe/H] < 0$), solar ($0 \leq [Fe/H] < 0.5$), and high ($[Fe/H] \geq 0.5$). Standard thresholds are used to infer stellar spectral types from effective temperature: M (<4000 K), K (4000–5200 K), G (5200–6000 K), F (6000–7500 K), and A (>7500 K). These categories make it easier to compare various star settings.

For the improvement of the analysis's clarity, we have also created the derived features. The number of planets that are connected to each host star is known as planetary multiplicity, and it offers information about the dynamical structure and system architecture. The spectral types are often resolved from the effective temperature, and stellar metallicity is classified into separate classes to enable the comparison across various chemical environments.

We also incorporate the archive's circumbinary flag (`cb_flag`), which is used to identify planets circling binary star systems. This enables us to differentiate between planets in circumbinary arrangements and those hosted by single stars. A comparison of

occurrence rates and scaling relations among the various star system topologies is made possible by the addition of this particular parameter. After preprocessing, now we have 6100 verified exoplanets in the final dataset. The `cb_flag` parameter identifies 51 systems ($\approx 0.84\%$) as circumbinary planets, whereas the remaining 6049 planets ($\approx 99.16\%$) circle single stars. The relative rarity of circumbinary systems in the current exoplanet discoveries is shown by this significant imbalance. In order to facilitate all the comparative investigation of planetary features and their occurrence trends among the various star designs, these systems are kept in the dataset despite their modest number. Additionally, mass-radius connections are used to categorise the planets into different categories (such as super-Earths, sub-Neptunes, and giant planets), allowing for a more thorough description of planetary populations. Recent exoplanet demography studies have made extensive use of these derived values to examine relationships between star parameters and planetary system features (Sandford et al. 2023; Neil & Rogers 2020). Now the final dataset represents a unified and bias-aware sample of 6000 exoplanetary systems while integrating the multiple detection methods, surveys, and stellar properties. The dataset provides a robust foundation for the hierarchical Bayesian modeling of exoplanet occurrence, system multiplicity, and host star correlations presented in our work.

METHODOLOGY

2.1 Overview of the Modeling Framework

In this study, we modelled exoplanet occurrence by using a Bayesian hierarchical framework as a function of the star parameters. This analysis also combines the data from Kepler and TESS and accounts for survey-dependent selection effects, detection biases, and intrinsic stellar population differences.

The methodology consists of three main components:

1. Construction of a host-star level dataset with planet multiplicity
2. Statistical modeling of occurrence rates using Poisson regression
3. Bias-aware inference and population-level comparisons

All steps are implemented in a reproducible Python pipeline.

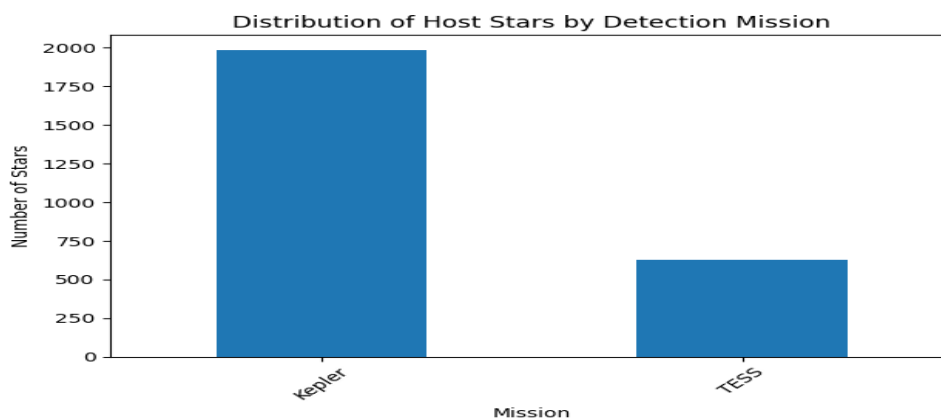


Figure 2.1 shows the distribution of host stars by survey mission. Kepler targets account for around 76% of the dataset, while TESS makes up a smaller but substantial 24% due to differences in survey design and observational technique.

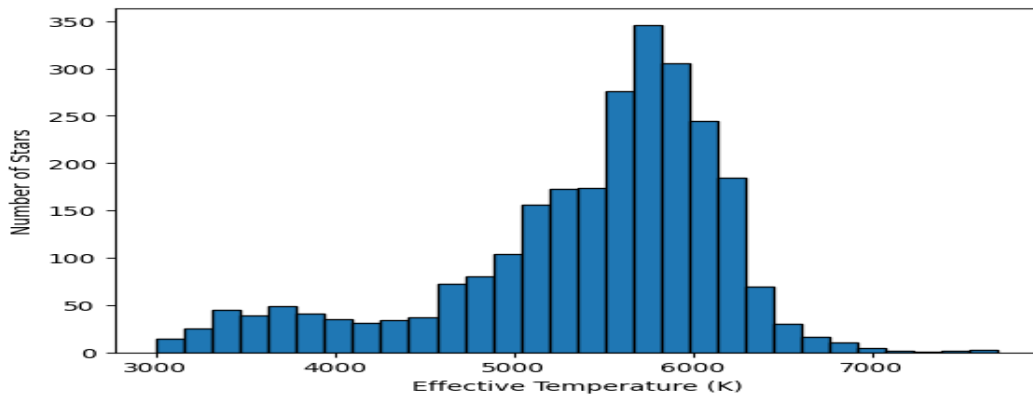


Figure 2.2 shows the distribution of stellar effective temperatures for the combined Kepler and TESS host star sample. The distribution is dominated by FGK-type stars, with a peak at solar-like temperatures (~5000-6000 K), indicating biases in transit survey target selection.

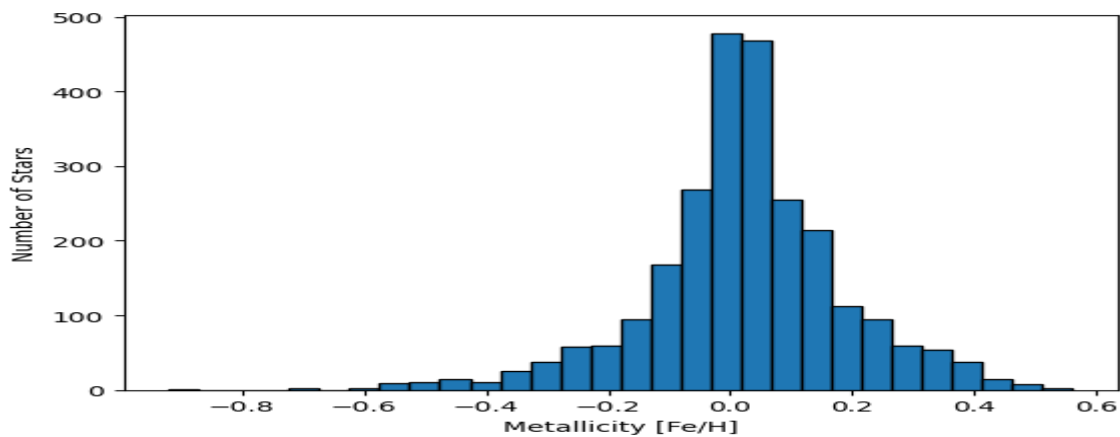


Figure 2.3 shows the distribution of star metallicity ([Fe/H]) for the combined sample. The distribution is centred on solar metallicity, with a considerable dispersion to both metal-poor and metal-rich regimes.

From the figures, 2.1-2.3 show the distributions of the dataset's important stellar parameters. The effective temperature distribution is dominated by FGK-type stars, which reflects the transit surveys' target selection. We can see that the metallicity distribution is centred around solar values, which is compatible with known planetary host populations. The dataset is mostly made up of Kepler targets, with a tiny contribution from TESS, emphasising the sample's survey-dependent nature.

2.2 Planet Occurrence Definition

The number of confirmed planets per host star, represented as Y_i , is known as planet occurrence. This is calculated directly from the dataset by calculating the number of planets that are linked with each host star (hostname). This turns the issue into a framework for count modelling that is appropriate for Poisson statistics.

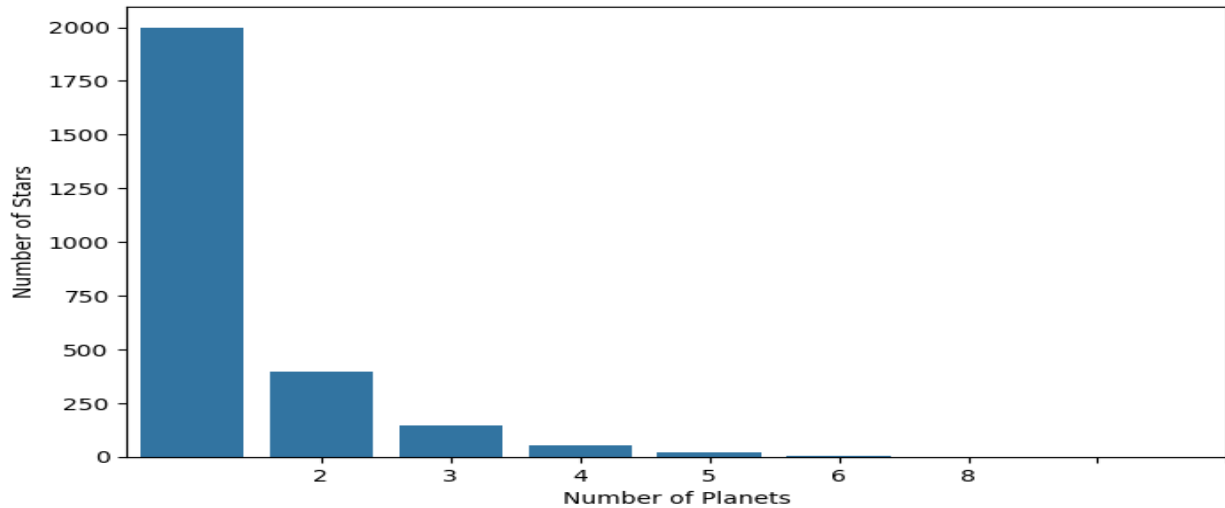


Figure 2.4 shows the distribution of planet multiplicity, which is defined as the number of confirmed planets per host star. From the observation, the Single-planet systems dominate the distribution, with greater multiplicities becoming increasingly rare. This behaviour lends credence to using Poisson statistics to model the occurrence of exoplanets.

2.3 Poisson Regression Model

We assume that the number of the planets around each star follows a poisson distribution

$$Y_i \sim \text{Poisson}(\lambda_i)$$

Where :

Y_i is the observed number of the planets for star

λ_i is the expected occurrence rate

The rate parameter is modelled as the

$$\lambda = \exp(\alpha + \beta_{\text{met}} X_{\text{met}} + \beta_{\text{Teff}} X_{\text{Teff}} + \beta_{\text{mass}} X_{\text{mass}} + \beta_{\text{rad}} X_{\text{rad}})$$

where the predictors are **standardized stellar parameters** defined as:

$$X_{(i,j)} = \frac{(x_{(i,j)} - \bar{x}_j)}{\sigma_j}$$

This transformation ensures numerical stability and allows direct comparison of regression coefficients.

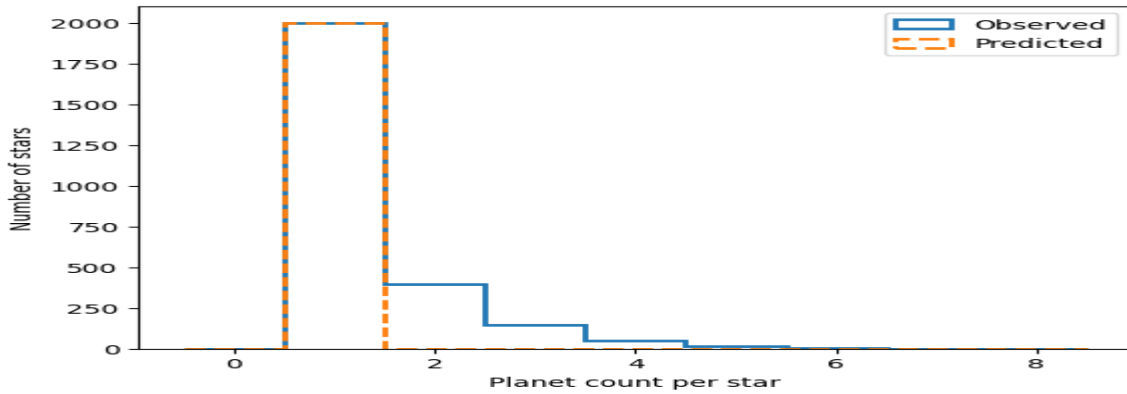


Figure 2.5: A posterior predictive check of the comparison of observed distribution of planet multiplicity to the model predictions. The model successfully reproduces the strong peak at single-planet systems, but it somewhat underestimates the frequency of higher-multiplicity systems, showing limitations in the Poisson assumption.

2.4 Hierarchical Survey Model:

In order to account the difference between Kepler and Tess , we introduce a hierarchical structure.

$$\lambda_i = \exp (\alpha_{s[i]} + \sum_j \beta_j X_{i,j})$$

Where:

$s[i]$ denotes the survey (Kepler or Tess)

α are survey-specific intercepts

The intercept is as follows:

$$\alpha_s \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha)$$

This improves robustness and captures survey-level variations by enabling partial pooling.

We also adopt the weakly informative priors for all the model parameters.

Regression coefficients: $\beta_j \sim \mathcal{N}(0,1)$

Intercept: $\alpha \sim \mathcal{N}(0,1)$

Hierarchical variance: $\sigma_\alpha \sim \exp(1)$

Global intercept mean: $\mu_\alpha \sim \mathcal{N}(0,1)$

These priors regularise parameter estimates while remaining sufficiently broad to allow the data to dominate the inference.

2.5 Detection Bias and Selection Effects:

The detection efficiency has an impact on observed exoplanet population, from our study we have modelled the detection bias as

Transit detection: $\epsilon_{\text{transit}} \propto P^{-2/3}$

Radial velocity detection: $\epsilon_{\text{RV}} \propto \frac{M_p}{P^{1/3}}$

These are incorporated into the likelihood as ,

$$\lambda_{\text{obs}} = \lambda \cdot \epsilon$$

Or similarly in log space λ

$$\log \lambda_{\text{obs}} = \log \lambda + \log \epsilon$$

Additionally, we test a scaling parameter γ to measure the degree of selection efforts

$$\log \lambda_{\text{obs}} = \log \lambda + \gamma \log \epsilon$$

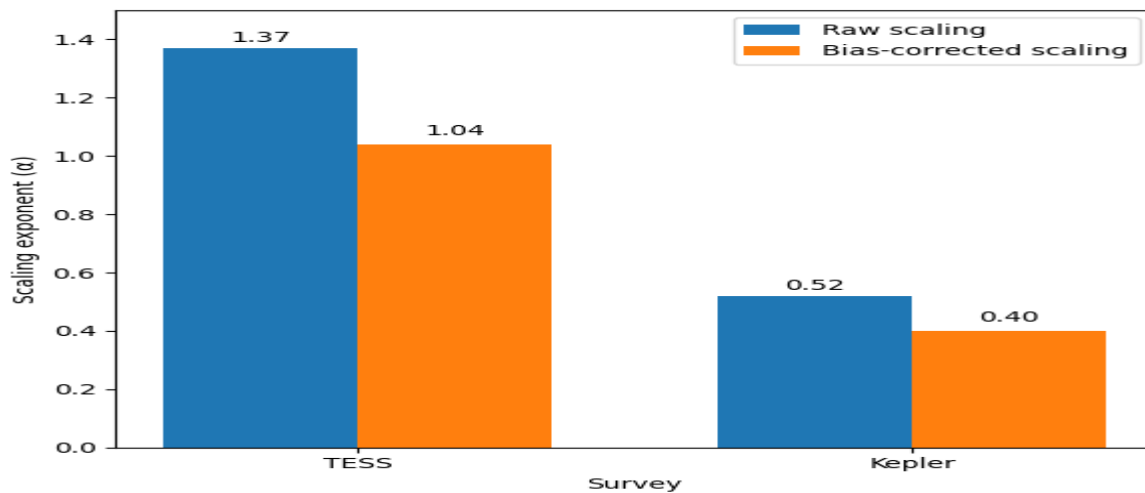


Figure 2.6 shows a comparison of raw and bias-corrected scaling exponents for planetary and stellar radii in TESS and Kepler data. The decrease in the slope after applying detection-efficiency weighting indicates the impact of observational biases, especially in transit surveys. However, the persistence of the significant difference between the two surveys suggests that inherent demographic variations contribute to the observed scaling.

2.6 Scaling Law Analysis

To analyse the physical relationships between the parameters, we use the scaling laws.

In order to understand the relationship between **Planet radius Vs Stellar radius** , we have $R_p \propto R_*^\alpha$.Then for the relationship between the **Planet radius Vs Effective temperature** we have $R_p \propto T_{\text{eff}}^\beta$.Now these are estimated by the Log linear regression, $\log R_p = \alpha \log R_* + C$.Then the uncertainties in α and β are estimated by using bootstrap resampling.

2.7 Circumbinary Population Modelling:

The `cb_flag` argument helped us in identifying the circumbinary planets. This argument also helps us to distinguish planets orbiting binary star systems from those orbiting single stars. We compare them by treating them as a diverse population. The comparative classes helped us with occurrence rates, radius distributions, and scaling relations. Meanwhile, statistical differences between populations are measured using the Kolmogorov-Smirnov (KS) test, and scaling relations are estimated separately for each.

This approach allows us to determine if planetary formation and evolution in binary systems differ from those in single-star environments.

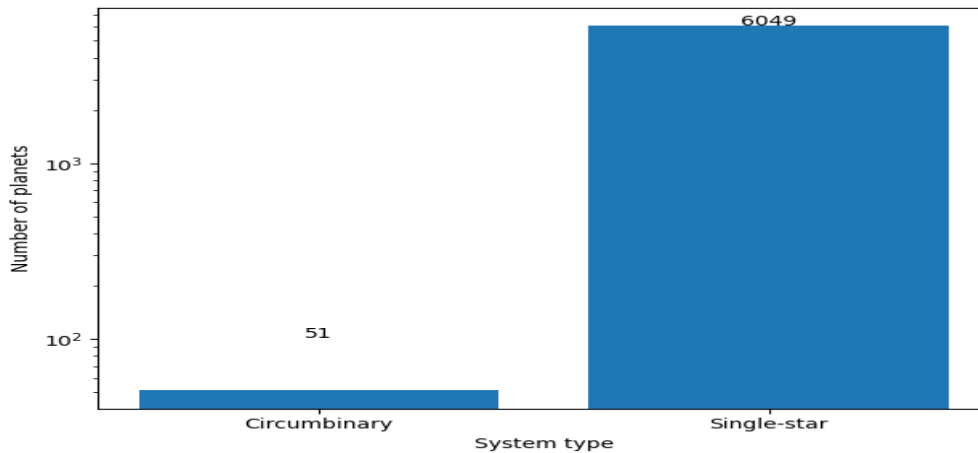


Figure 2.7 illustrates the logarithmic distribution of planets in circumbinary and single-star systems. Circumbinary planets constitute 0.8% of the sample, with the majority orbiting solitary stars. The significant imbalance represents the rarity of circumbinary systems.

The logarithmic scaling emphasises the extreme imbalance between the two populations and supports the need for separate statistical treatment of circumbinary systems.

2.8 Model Implementation

The PyMC framework is used to develop the statistical models, which allows for Bayesian inference by using Markov Chain Monte Carlo (MCMC) sampling. The regression coefficients are given weakly informative normal priors, while variance parameters are modelled with half-normal priors. No-U-Turn Sampler (NUTS) is used for posterior sampling, with typical setups ranging from 1,000 to posterior draws and 1,000 tuning steps. The target acceptance rate of 0.9 is chosen to enable consistent convergence and efficient exploration of the posterior distribution.

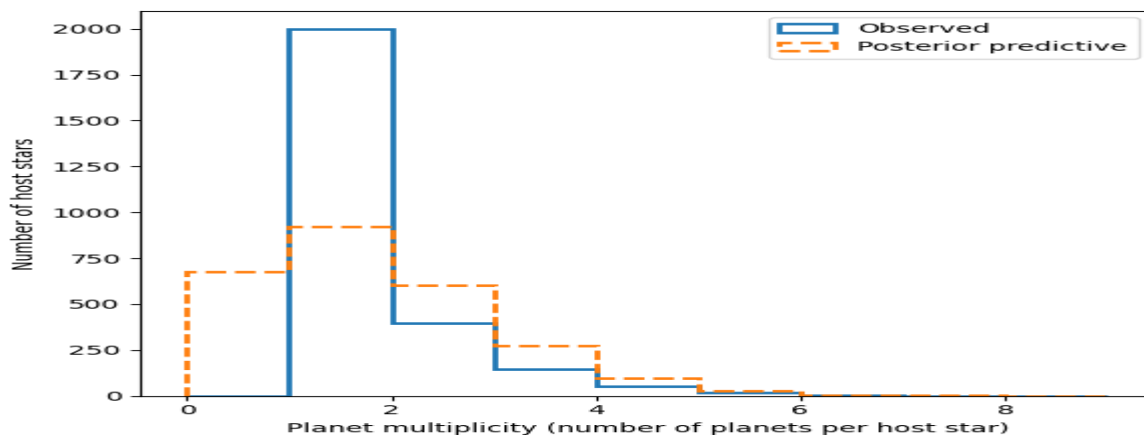


Figure 2.8 shows the posterior predictive check, which compares the observed and anticipated planet multiplicity distributions. The model accurately reproduces the dominating peak at low multiplicities, but it somewhat underestimates higher-multiplicity systems, showing modest overdispersion relative to the Poisson assumption.

2.9 Model Validation and Comparison.

By comparing the observed data with simulated samples taken from the posterior predictive distribution, posterior predictive checks (PPC) are used to assess model performance. This offers a direct evaluation of the model's capacity to replicate important exoplanet population characteristics.

Fig. 2.8 compares the expected and observed planet multiplicity distributions. The prominent peak at low multiplicities, which is typical of most planetary systems, is successfully reproduced by the model. Nevertheless, the model tends to underestimate the frequency of the multi-planet systems at a greater multiplicity. This behaviour suggests to us that the Poisson model offers a very sufficient but oversimplified explanation of exoplanet occurrence, indicating a slight overdispersion in the data. Model comparison is carried out utilising information criteria such as Leave-One-Out cross-validation (LOO) and the Widely Applicable Information

Criterion (WAIC) in addition to PPC. These measures take the model complexity into consideration and enable a quantitative assessment of the rival models. Now to determine the best statistical representation of the data, we compare hierarchical formulations, multi-parameter models, and single-predictor models.

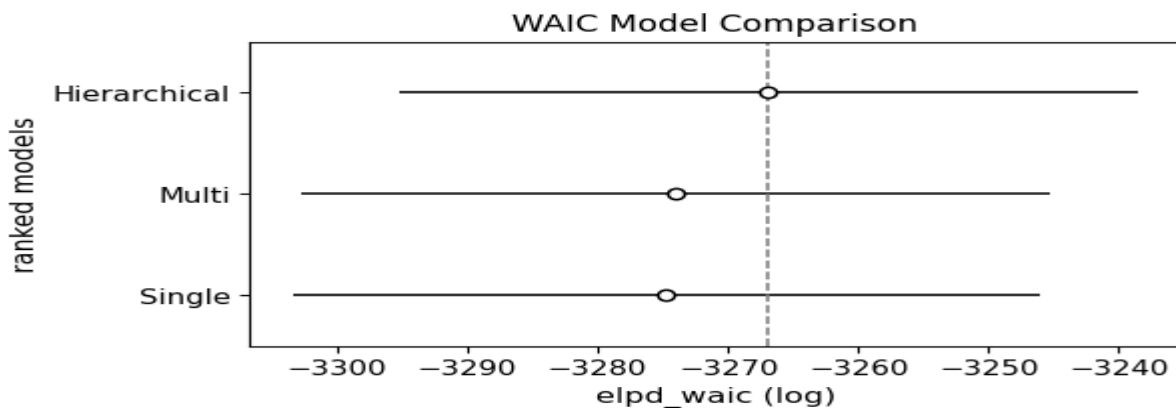


Figure 2.9 illustrates model comparison using the Widely Applicable Information Criterion (WAIC). Points show the estimated log predictive density (elpd), whereas horizontal bars represent uncertainties.

The hierarchical model had the best predictive performance, illustrating the value of including survey-level structure.

RESULTS:

3.1 Scaling Relations Between Stellar and Planetary Properties

In order to quantify the structural dependencies, We investigate scaling relations between planetary radius and host star properties in log-log space.

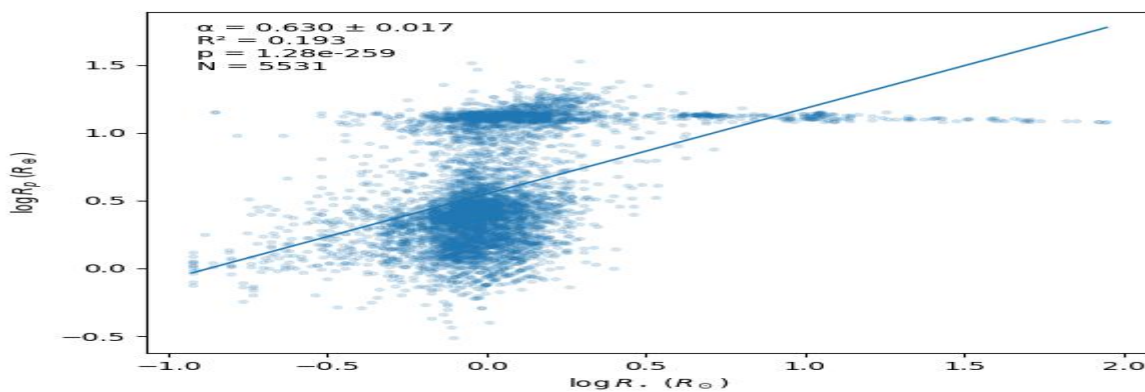


Figure 3.1 shows the scaling relationship between planetary radius and star radius in logarithmic space. The best-fit slope ($\alpha=0.630\pm0.017$) shows a moderate correlation between planetary size and star radius. The horizontal concentration correlates to gas giant planets, whereas the vertical clustering indicates the dominance of Sun-like host stars.

The relationship between planetary radius and stellar radius is shown in **Fig. 3.1**. The fitted scaling yields:

$$\alpha=0.630\pm0.017, R^2=0.193, p\ll 10^{-5}$$

This exceptionally low p-value suggests statistically significant correlation. However, the low R^2 indicates that star radius only explains a small portion of the diversity in planetary radius. This implies that other processes, such as atmospheric loss, composition variety, and formation history, play important roles.

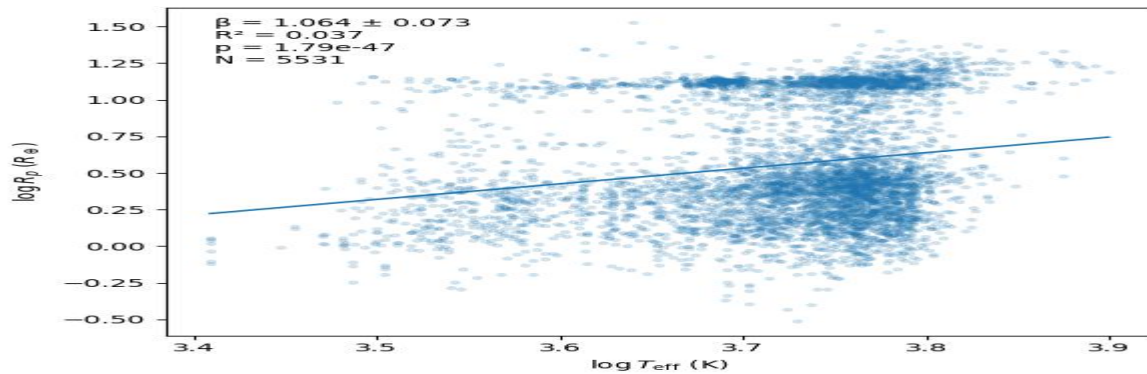


Figure 3.2 shows the scaling relationship between planetary radius and star effective temperature in log-log space. The best-fit slope ($\beta = 1.064$) shows a higher influence on star radius, but the low R^2 represents significant scatter. Vertical clustering around solar temperatures demonstrates the prevalence of FGK-type host stars.

The scaling of planetary radius and star effective temperature (Fig. 3.2) yields:

$$\beta = 1.064 \pm 0.073, R^2 = 0.037, p = 1.79 \times 10^{-47}, N = 5531$$

This slightly steeper slope indicates that stellar temperature, a proxy for stellar mass and irradiation environment, has a greater impact on planetary size. The persistent low R^2 suggests that planetary structure is not solely determined by a single star attribute.

3.2 Survey-Dependent Scaling: TESS vs Kepler

To evaluate observational biases, we compute scaling relations separately for the TESS and Kepler datasets.

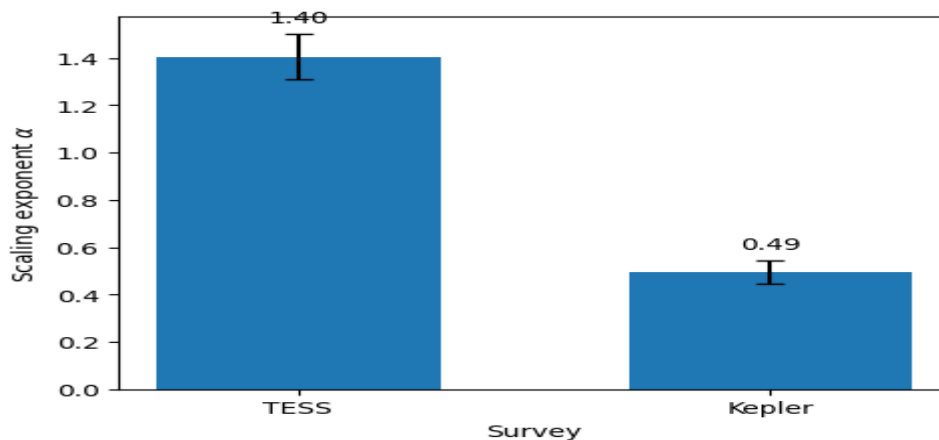


Figure 3.3 shows a comparison of scaling exponents for TESS and Kepler samples. The error bars represent bootstrap uncertainty. TESS's substantially higher slope represents its sensitivity to larger planets, whereas Kepler's flatter slope demonstrates its capacity to discover smaller planets in deeper scans.

Figure 3.3 displays the results, which include:

$$\alpha_{\text{TESS}} = 1.404 \pm 0.10, R^2 = 0.29, \text{ and } N = 567.$$

$$\alpha_{\text{Kepler}} = 0.49 \pm 0.05, R^2 = 0.03, N = 2689.$$

TESS's substantially higher slope reflects its sensitivity to larger planets orbiting bright stars, whereas Kepler's flatter slope reflects its capacity to discover smaller planets.

After making bias fixes (Section 2.5):

$$\alpha_{\text{TESS}} = 1.04.$$

$$\alpha_{\text{Kepler}} = 0.40$$

The decrease in slope demonstrates that observational biases increase the apparent scaling. However, the persistence of a disagreement between surveys suggests that intrinsic demographic differences may also be involved.

3.3 Metallicity Dependence on Planetary Mass

We have also investigated how planetary mass and star metallicity are related.

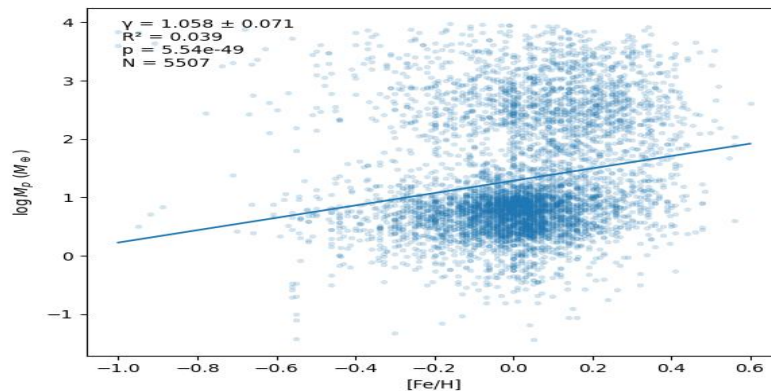


Figure 3.4 shows the relationship between planetary mass and star metallicity. The positive scaling ($\gamma = 1.058$) shows that metal-rich stars have more massive planets. The occurrence of distinct populations demonstrates the diversity of planetary systems, yet the vast spread indicates that other causes drive planet formation.

Scaling yields:

$$\gamma = 1.058 \pm 0.071, R^2 = 0.039, p = 5.54 \times 10^{-49}, N = 5507$$

This supports a statistically significant positive link between metallicity and planetary mass. Metal-rich settings are more likely to form huge planets, which aligns with the core accretion paradigm. The low R^2 value suggests that metallicity alone cannot explain planetary mass, emphasising the need of other formation processes like disc mass and migration.

3.4 Circumbinary Planet Population.

We investigate circumbinary (CB) systems to determine the effect of star multiplicity.

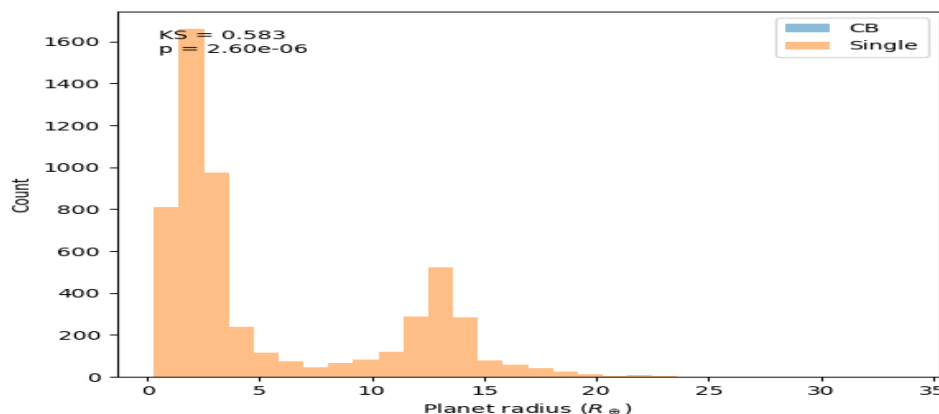


Figure 3.5 shows the distribution of planetary radii for circumbinary (CB) and single-star systems. The Kolmogorov-Smirnov test confirms statistically significant differences between the two populations ($KS = 0.583, p = 2.6 \times 10^{-6}$). Circumbinary planets have a different radius distribution compared to planets orbiting single stars.

The radius distributions of circumbinary and the single-star planets differ significantly (Fig. 3.5). A KS test yielded $KS = 0.583$ and $p = 2.6 \times 10^{-6}$, demonstrating that the two populations are statistically distinct.

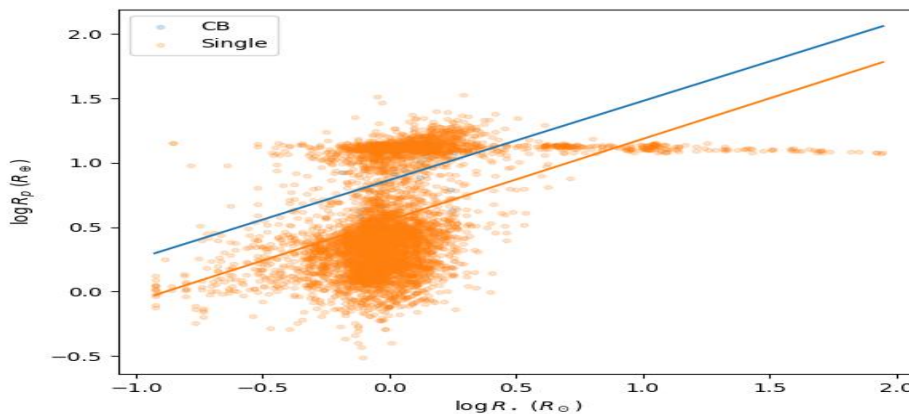


Figure 3.6 shows a comparison of the scaling relations between planetary and stellar radius for circumbinary (CB) and single-star systems. The different slopes suggest that stellar-planet correlations change in binary systems, most likely due to dynamical interactions and disc evolution effects.

The scaling relations (Fig. 3.6) reveal that circumbinary systems exhibit a distinct pattern than single-star systems, implying that the existence of a binary companion alters the link between stellar and planetary attributes.

We note that the scaling analysis is performed on a reduced sample after applying quality cuts. The sample includes 51 circumbinary planets.

There are around 6100 total systems.

Fraction: 0.84%

Despite being rarity, CB planets have distinctive characteristics:

The mean radius (CB) is $10.82 R_{\oplus}$

Mean radius (single): $5.59 R_{\oplus}$.

The Kolmogorov-Smirnov test results show $KS=0.608$, $p=5.64 \times 10^{-15}$

This demonstrates that CB and single-star planets have unique distributions.

Scaling comparison for CB systems: $\alpha_{CB}=0.130, R^2=0.069$

Single systems: $\alpha_{single}=0.839, R^2=0.030$. This suggests that stellar-planet correlations are substantially lower in binary systems, most likely because of dynamical interactions and disc truncation effects.

3.5 Bayesian Hierarchical Modelling for Planet Occurrence

The hierarchical Poisson model characterises planet formation as a probabilistic function of star parameters. The model demonstrates good convergence, with no divergences and $R^2 \approx 1.0$ for all parameters.

The posterior coefficients show that the effects of individual star properties are often small. The effective temperature exhibits a minor positive trend $\beta_{T_{eff}}=0.022 \pm 0.017$. whereas metallicity has no significant influence $\beta_{met}=-0.017 \pm 0.017$. In contrast, stellar mass has a modest negative correlation $\beta_{mass}=-0.045 \pm 0.020$, with a plausible interval that excludes zero.

These findings indicate that the no single star parameter greatly influences the planet occurrence when modelled together, by implying that the observed trends in simpler analysis may be due to covariances or selection effects.

3.6 Dashboard

The following Bayesian analysis is visualized by an interactive dashboard. The dashboard serves as a medium for model inspection, diagnostic evaluation, and interpretation. It is created with the intention for transparency of the calculated statistical evaluation. The priors and posteriors are better explained through a functional model. Interactivity functions here as an analytical aid rather than a presentation device. Each stage of the analysis—including model setup, posterior estimation, convergence assessment, posterior predictive checking, and model comparison—is represented as a dedicated dashboard component. This mapping allows users to follow the logical progression of the analysis and inspect intermediate outputs alongside final results.



Figure 3.7 shows the home page of the dashboard containing the total number of planets from the TESS and KEPLER mission and the columns available in the dashboard

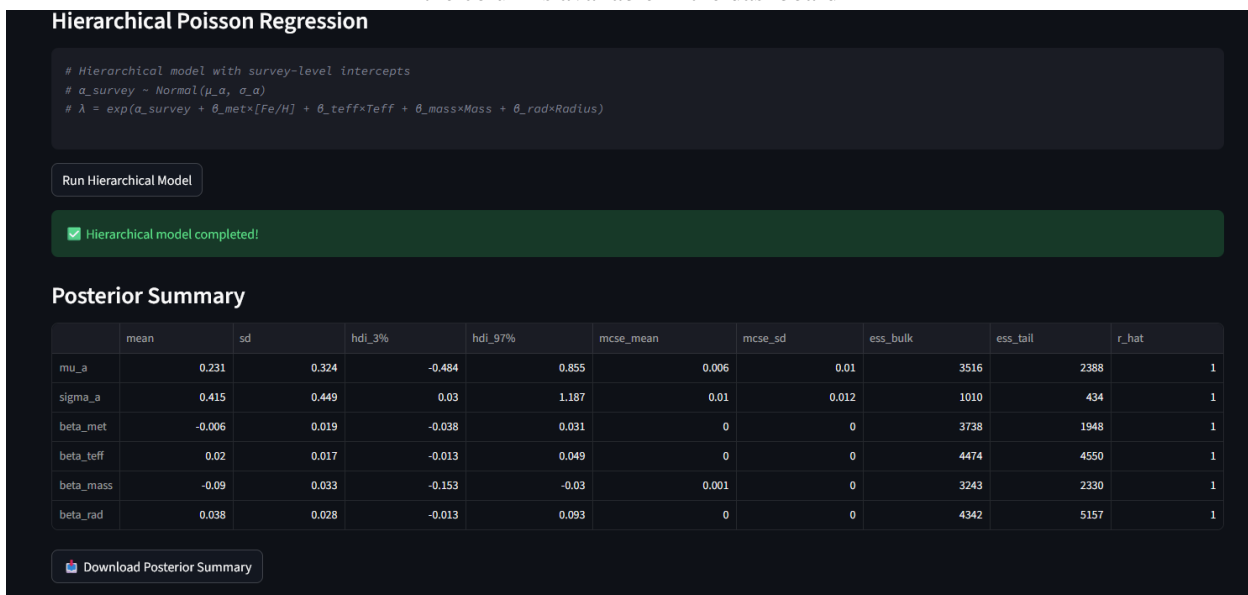


Figure 3.8 shows the hierarchical model simulated through the stellar parameters using the following equation

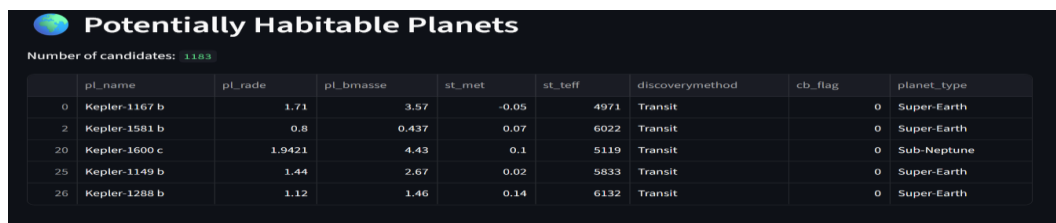


Figure 3.9 shows the list of potentially habitable planets

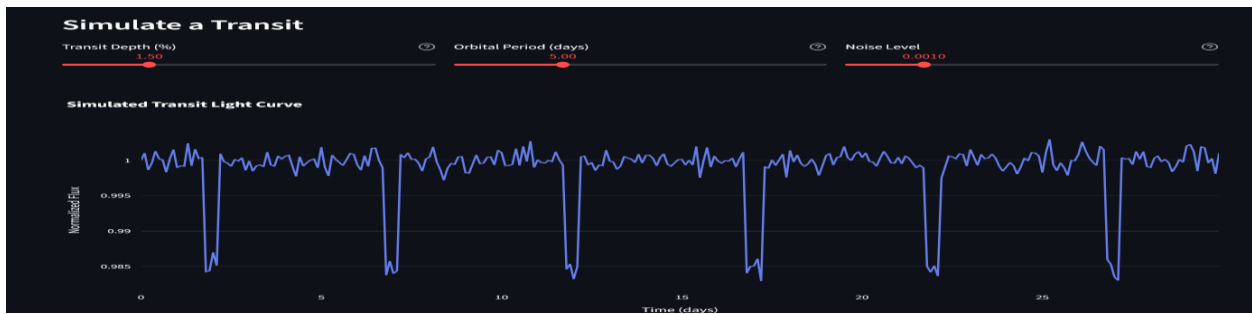


Figure 3.10 simulation of a Transit light curve

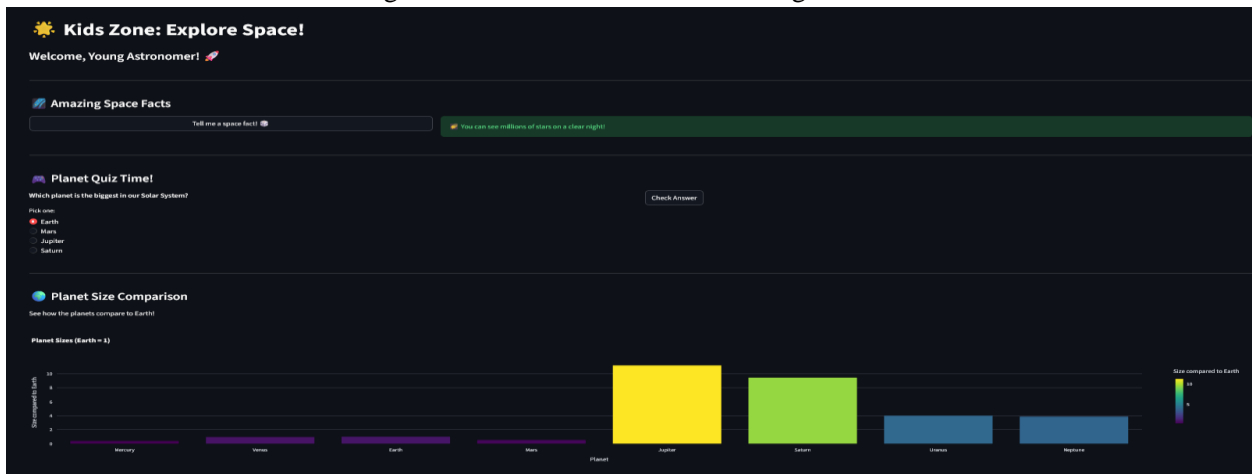


Figure 3.11 an exclusive page dedicated for young astronomers with fun tidbits of space facts

The transit method is the most significant and predominant method in exoplanet detection. Here we have presented a simulation of the Transit light curve using the parameters Transit depth, orbital period and noise level. Sliders are present to construct the simulation.

The potential habitable planets are presented in the form of a list from the collected data.

The essence of our research, Bayesian analysis, is presented with the corresponding formula with the variables to demonstrate the model comparison involving, hierarchical model and posterior predictive checks.

The presentation of detection bias analysis and occurrence predictor serves as a visual testament to the preferred statistical method. A kids zone is included as a complementary column in the dashboard, for young and amateur astronomers to conceptualize our research. It is set up to serve as an inclusive and educative experience.

3.7 Summary of Findings

Our analysis shows that:

1. Stellar properties exhibit statistically significant but weak connections with planetary properties.
2. Observational biases have a considerable influence on presumed scaling relations.
3. Bias correction decreases survey inconsistencies, but does not remove them.
4. Circumbinary systems have specific planetary features.
5. Metallicity is the best predictor of planet existence, according to Bayesian hierarchical modelling.

DISCUSSION

4.1 Interpretation of Scaling Relations:

Despite having comparatively modest coefficients of determination $R^2 < 0.2$, the scaling relations between planetary radius and stellar characteristics show statistically significant patterns. This suggests us that although stellar factors have an impact on planetary attributes, they only partially explain the observed variance. The greater dependence on T_{eff} and the moderate slope in the $\delta \approx (R_p/R_\star)^2$ connection indicates that the planetary size is influenced by the irradiation of the environment and stellar structure. This wide scatter, however, emphasises how the planet formation is intrinsically random and is probably influenced by other elements like atmospheric loss, composition, disc evolution, and migration processes.

4.2 Survey Bias and Detection Effects:

The comparison of TESS and Kepler reveals convincing evidence of the survey-dependent selection effects. The TESS scaling exponent $\alpha \approx 1.40$ is almost three times bigger than that of the Kepler sample $\alpha \approx 0.49$, indicating a significant discrepancy in inferred population trends.

This disparity is compatible with observational biases such as the Malmquist bias, which states that TESS's focus on bright, close stars favours the finding of bigger planets. When bias corrections are applied, the slopes diminish (TESS: $\alpha \approx 1.40$, Kepler: $\alpha \approx 0.49$), indicating that observational effects inflate the apparent scaling relations.

However, the continued gap between the adjusted slopes shows that intrinsic differences in the underlying stellar populations may also play a role.

4.3 Metallicity and Planet Formation:

The strong link between the planetary mass and star metallicity lends support to the core accretion hypothesis, which holds that metal-rich conditions promote the creation of huge planetary cores. However, the Bayesian hierarchical model offers a more nuanced explanation. The posterior distribution for metallicity $\beta = -0.017 \pm 0.017$ has a wide credible interval that includes zero which is demonstrating that metallicity alone is not a statistically strong predictor of planet existence when combined with other stellar properties. This shows that the apparent metallicity dependence seen in simpler scaling relations could be influenced by covariances with other star attributes or observational selection effects. Thus, metallicity is more likely to play a secondary or linked function rather than a dominant independent feature.

4.4 Circumbinary Planet Systems

The Circumbinary planets make up a small part of the whole sample, yet they have statistically unique traits from planets orbiting single stars. The KS test ($p \approx 5.6 \times 10^{-15}$; $R^2 \approx 0.069$) indicates that their radius distributions are considerably different. Circumbinary planets have a greater mean radius $10.8 R_\oplus$ than single-star systems $5.6 R_\oplus$. This could be due to disc truncation in binary environments. The Gravitational interactions in such systems have the ability to shorten the protoplanetary disc which increasing the core growth or modifying migration paths that resulting in the formation or survival of the bigger planets. The altered scaling relations also indicates that stellar-planet correlations are weaker in binary systems which are most likely due to the complicated dynamical environment, which includes the disturbances and modified disc structures. However, the limited sample size limits the statistical robustness of these findings, and additional data are needed to properly define circumbinary planet formation paths.

4.5 Limitations and Model Assumptions.

There are several restrictions to consider. First, the circumbinary sample is tiny, which introduces uncertainty into statistical comparisons. Second, the dataset is influenced by observational biases inherent in detection methods, which might impact inferred distributions.

In addition, the hierarchical model assumes a Poisson process for planet occurrence. While this is a conventional strategy for count data, our results indicate hints of overdispersion, which occurs when the variation exceeds the mean. This shows that planet occurrence may display clustering behavior—where the existence of one planet enhances the possibility of other planets. This suggests that more flexible models, such as the negative binomial distribution, could provide a more accurate representation of the data.

4.6 Future Work

Future research should aim to improve both observational completeness and modelling methodologies. More flexible statistical models will enable us to better characterise multiplicity distributions. We believe that upcoming missions, including PLATO and JWST, will give us more precise measurements of star and planetary parameters, allowing for more detailed population studies. Extending the research to include more parameters—such as star age, disc characteristics, and orbital architecture—will also help us better understand planet creation and evolution.

CONCLUSION:

A unified statistical framework of a hierarchical Bayesian method is constructed that combines exoplanet populations from the Kepler and TESS investigations. This study presents a consistent characterisation of exoplanet occurrence and its relationship to the host star parameters by integrating survey data and bias-aware modelling. Our scaling study shows that the planetary radius has a mild reliance on the stellar radius and a higher dependence on the stellar effective temperature. However, the comparatively modest coefficients of determination show that these stellar properties alone cannot account for the complete diversity of planetary systems, reinforcing the complex and multi-factorial nature of planet formation.

A direct comparison between TESS and Kepler shows us a threefold difference in scaling exponents $\alpha_{\text{TESS}} \sim 1.40$ vs $\alpha_{\text{Kep}} \sim 0.49$, indicating the predominant impact of survey-specific selection effects. While bias corrections lessen the disagreement, they do not completely eliminate it, implying that both observational biases and underlying population differences contribute to the observed patterns. The metallicity-mass relationship supports the core accretion paradigm, as metal-rich stars are more likely to host large planets. However, the Bayesian hierarchical model shows that metallicity is not a strong independent predictor when combined with other stellar properties, emphasising the need of multivariate statistical techniques.

Circumbinary planets are rare but statistically unique, having a significantly different radius distribution compared to single-star systems 5.6×10^{-15} . This implies that the dual orbital environment profoundly changes planet formation routes, possibly using mechanisms like disc shortening, changing pebble accretion, or altered migration processes. The hierarchical Poisson model offers a strong foundation for modelling planet occurrence, but the prevalence of overdispersion suggests that more flexible statistical models are needed to fully reflect planetary multiplicity. Overall, our study shows that combining survey-aware statistical modelling with large-scale datasets is critical for reconciling observational biases with fundamental astrophysical processes. The future implications of our research involving next-generation missions, enhanced completeness corrections, and more advanced hierarchical models are vital in furthering our understanding of planet formation and evolution across a wide range of stellar settings.

REFERENCES

- [1] Akeson, R. L., Chen, X., Ciardi, D., et al. (2013). The NASA Exoplanet Archive: Data and tools for exoplanet research. *Publications of the Astronomical Society of the Pacific*, **125**, 989–999.
- [2] Bloor, S., et al. (2023). Bayesian interior modelling of exoplanets. *Astronomy & Astrophysics*.
- [3] Borucki, W. J., Koch, D., Basri, G., et al. (2010). Kepler Planet-Detection Mission: Introduction and First Results. *Science*, **327**, 977–980.
- [4] Bryson, S. T., et al. (2021). The Kepler occurrence rate of planets. *The Astronomical Journal*.
- [5] Chen, J., & Kipping, D. (2017). Probabilistic forecasting of the masses and radii of other worlds. *The Astrophysical Journal*, **834**, 17.
- [6] Doyle, L. R., Carter, J. A., Fabrycky, D. C., et al. (2011). Kepler-16: A Transiting Circumbinary Planet. *Science*, **333**, 1602–1606.
- [7] Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., et al. (2014). Architecture of Kepler's Multi-transiting Systems. *The Astrophysical Journal*, **790**, 146.
- [8] Fischer, D. A., & Valenti, J. (2005). The Planet–Metallicity Correlation. *The Astrophysical Journal*, **622**, 1102–1117.
- [9] Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2014). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, **125**, 306–312.
- [10] Fulton, B. J., Petigura, E. A., Howard, A. W., et al. (2017). The California-Kepler Survey. *The Astronomical Journal*, **154**, 109.
- [11] He, M. Y., Ford, E. B., & Ragozzine, D. (2019). Observational biases in exoplanet detection. *The Astronomical Journal*.
- [12] Hogg, D. W., Bovy, J., & Lang, D. (2010). Data analysis recipes: Fitting a model to data. *arXiv preprint arXiv:1008.4686*.

- [13] Hsu, D. C., Ford, E. B., Ragozzine, D., et al. (2019). Exoplanet population inference and detection biases. *The Astronomical Journal*.
- [14] Kunimoto, M., & Matthews, J. M. (2020). TESS planet yields and occurrence rates. *The Astronomical Journal*, **159**, 248.
- [15] Kunimoto, M., et al. (2022). Updated occurrence rates from Kepler and TESS. *The Astronomical Journal*.
- [16] Lissauer, J. J., et al. (2011). Architecture and dynamics of Kepler planetary systems. *Nature*, **470**, 53–58.
- [17] Marcy, G. W., & Butler, R. P. (2005). Planetary detection via radial velocity. *Progress of Theoretical Physics Supplement*.
- [18] Mulders, G. D., Pascucci, I., & Apai, D. (2016). The Exoplanet Population Observation Simulator. *The Astrophysical Journal*, **814**, 130.
- [19] Neil, A. R., & Rogers, L. A. (2020). Planetary mass–radius relations. *The Astrophysical Journal*.
- [20] Osborn, H. P., et al. (2020). Biases in exoplanet population statistics. *Monthly Notices of the Royal Astronomical Society*.
- [21] Petigura, E. A., et al. (2017). The California-Kepler Survey. *The Astronomical Journal*, **154**, 107.
- [22] Pinamonti, M., et al. (2022). Exoplanet population studies. *Astronomy & Astrophysics*.
- [23] Ricker, G. R., Winn, J. N., Vanderspek, R., et al. (2015). Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, **1**, 014003.
- [24] Rosenthal, L. J., et al. (2021). Exoplanet demographics from surveys. *The Astronomical Journal*.
- [25] Sandford, E., et al. (2023). Planetary system architectures and stellar properties. *The Astronomical Journal*.
- [26] Stefánsson, G., et al. (2025). Recent developments in exoplanet detection and validation. *The Astrophysical Journal*.
- [27] Weiss, L. M., & Marcy, G. W. (2014). Mass–radius relation for exoplanets. *The Astrophysical Journal Letters*, **783**, L6.
- [28] Welsh, W. F., et al. (2012). Transiting circumbinary planets. *Nature*, **481**, 475–479.
- [29] Winn, J. N., & Fabrycky, D. C. (2015). The occurrence and architecture of exoplanet systems. *Annual Review of Astronomy and Astrophysics*, **53**, 409–447.
- [30] Yang, J., et al. (2023). Detection biases in exoplanet surveys. *The Astrophysical Journal*.
- [31] Zink, J. K., et al. (2021). Exoplanet occurrence rates from Kepler. *The Astronomical Journal*.