Draft version December 16, 2017
Typeset using LaTeX **modern** style in AASTeX61

THE EFFECTS OF BINARITY ON PLANET OCCURRENCE RATES MEASURED BY TRANSIT SURVEYS

L. G. Bouma,[1] J. N. Winn,[1] and K. Masuda[1]

[1] *Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540, USA*

Submitted to AAS journals.

ABSTRACT

This study develops simplified models of signal-to-noise limited transit surveys, in order to clarify the biases that stellar binarity introduces to occurrence rate measurements. Particular attention is paid to dilution of observed planet radii, errors in star counts, and detection efficiencies. Free parameters in our models include the true radius distribution of planets, and the relative numbers of planets orbiting secondaries, primaries, and single stars. In the simplest possible case, all stellar systems are either single or twin binaries, and all planets are identical. If so, then we find that ignoring binarity leads to an underestimate of the occurrence at the true radius by a multiplicative factor of 0.76 (assuming a 10% twin binary fraction). Using more realistic models for the stellar population and planetary radii, we find that the occurrence of Earth-sized planets can be overestimated by at most $\approx 50\%$ when ignoring binarity – at present far smaller than systematic uncertainties on real $\eta_\oplus$ measurements. We quantify how gaps in the intrinsic planet radius distribution are filled in when binarity is neglected. Our models suggest that for most of the observed radius distribution ($r > 2r_\oplus$), ignoring binarity only leads to errors in inferred occurrence rates at the $\lesssim 10\%$ level. One corollary is that binarity is unlikely to strongly influence the HJ rate discrepancy.

Keywords: methods: data analysis — planets and satellites: detection — surveys

Corresponding author: L. Bouma
luke@astro.princeton.edu

1. INTRODUCTION

A group of binarity-ignoring astronomers wants to measure the mean number of planets of a certain type per star of a certain type. They perform a wide-field photometric search for planets that transit, and discover all for which

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right) = (\mathrm{const}) \cdot \delta_{\mathrm{obs}} L_{\mathrm{sys}}^{1/2} d^{-1} > \left(\frac{\mathrm{S}}{\mathrm{N}}\right)_{\mathrm{floor}}. \quad (1)$$

The observed transit depth δ_{obs} constitutes the signal, S. The survey is shot-noise dominated, and so the noise N scales as the inverse root of the photon flux received from any stellar system, $\mathrm{N} \propto (L_{\mathrm{sys}}/d^2)^{-1/2}$, for L_{sys} the system luminosity and d its distance. The solid angle coverage, telescope area, and survey duration are part of the "const" term of Eq. 1. The transit duration also affects detectability, but we omit it in this work for brevity.

To compute an occurrence, at every planetary and stellar property the astronomers choose the stars (among those initially surveyed) around which the planets of interest appeared to be searchable. Correcting for the geometric transit probability p_{tra}, they report an apparent rate density,

$$\Gamma_{\mathrm{a}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}}) = \frac{n_{\mathrm{det}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}})}{N_{\mathrm{s,a}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}})} \times \frac{1}{p_{\mathrm{tra}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}})}. \quad (2)$$

where $\mathcal{P}_{\mathrm{a}}$, $\mathcal{S}_{\mathrm{a}}$ are the apparent planetary and stellar parameters. The quantity $N_{\mathrm{s,a}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}})$ is the number of unresolved point-sources on the sky that appear to be searchable, and $n_{\mathrm{det}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}})$ is the number of detected planets, per unit $\mathcal{P}_{\mathrm{a}}$ and $\mathcal{S}_{\mathrm{a}}$.

There are many potential pitfalls. Some genuine transit signals can be missed by the detection pipeline. Some apparent transit signals are spurious, from noise fluctuations, failures of 'detrending', or instrumental effects. Stars and planets can be misclassified due to statistical and systematic errors in the measurements of their properties. Poor angular resolution causes false positives due to blends with background eclipsing binaries. *Et cetera.*

Here we focus on problems that arise from the fact that many stars exist in multiple star systems. For simplicity, we consider only binaries, and we assume that they are all bound and spatially unresolved.

An immediate complication is that, due to dynamical stability or some aspect of planet formation, the occurrence rate of planets might differ between binary and single-star systems. If "occurrence rate" is defined as the mean number of planets within set radius and period bounds per star in a given mass interval, it must implicitly marginalize over stellar multiplicity. This means marginalizing over "occurrence rates in single star systems", "occurrence rates about primaries", and "occurrence rates about secondaries" (see Wang et al. 2015a).

Outside of astrophysical differences, every term in Eq. 2 is observationally-biased. There are errors in $n_{\mathrm{det}}(\mathcal{P}_{\mathrm{a}}, \mathcal{S}_{\mathrm{a}})$ due to planet radius misclassification. There are errors

in $N_{\rm s,a}(\mathcal{P}_{\rm a}, \mathcal{S}_{\rm a})$ because an apparently searchable point on the sky might include two searchable stars. There are errors in $p_{\rm tra}(\mathcal{P}_{\rm a}, \mathcal{S}_{\rm a})$ because stars in binaries may have different masses and radii than assumed in the single-star case. Binarity might also affect the observers' ability to correctly select searchable stars (but see Sec. 3.1 for discussion of this issue).

Correcting for binarity's observational biases is non-trivial. For instance, in counting the number of selected stars, even after realizing that binaries count as two stars, one must note that the multiplicity fraction of *selected* stars is greater than that of a volume limited sample. This is the familiar Malmquist bias: binaries are selected out to larger distances than single stars because they are more luminous. As a separate challenge, finding the correct number of detected planets in a radius bin requires knowledge of the true planetary radii. Observers deduce apparent radii. In binaries, the apparent and true radii differ because of diluting flux, and possibly because the planet is assumed to orbit the wrong star (*e.g.*, Furlan et al. 2017).

To gain intuition for the many observational biases at play, we consider a set of idealized transit surveys:

- Model #1: fixed stars, fixed planets, twin binaries;
- Model #2: fixed primaries, power-law planets, varying secondaries;
- Model #3: fixed primaries, broken power-law planets and varying secondaries.

In Sec. 2, we clarify our terminology and state assumptions that apply throughout this study. We then present our transit survey models in Secs. 3.1-3.3, where each subsection corresponds to a model listed above. We interpret these calculations throughout, and in Sec. 4 connect them to topical questions in the interpretation of transit survey occurrence rates. In particular, we mention the "hot Jupiter rate discrepancy", the relevance towards measurements of $\eta_\oplus$, and the implications for detailed study of Fulton et al. (2017)'s recently discovered "valley". We conclude in Sec. 5.

2. DEFINITIONS AND PRELIMINARIES

2.1. *Occurrence rate, occurrence rate density*

Define the occurrence rate density, Γ, as the expected number of planets per star per unit of planetary and stellar phase space. Since we will mainly be concerned with the rate density's dependence on planetary radii r, we write

$$\Gamma(r) = \frac{{\rm d}\Lambda}{{\rm d}r}, \tag{3}$$

where Λ is the occurrence rate. In this notation, "the occurrence rate of planets of a particular size" translates to an integral of Eq. 3, evaluated over a radius interval.

The above definition implicitly marginalizes the rate density over stellar multiplicity. Considering only single and binary systems, the rate density for any selected sample

of stars is a weighted sum of rate densities for each system type:

$$N_{\rm tot}\Gamma_{\rm sel}(r) = N_0\Gamma_0(r) + N_1\Gamma_1(r) + N_2\Gamma_2(r), \tag{4}$$

where $i = 0$ corresponds to single stars, $i = 1$ to primaries of binaries, and $i = 2$ to secondaries of binaries. $N_{\rm tot} = \sum_i N_i$ is the total number of selected stars, and N_0, N_1, N_2 are the number of selected single stars, primaries, and secondaries. For our purposes binary systems are not identifiable, and so $N_1 = N_2$.

The rate density for each type of star, $\Gamma_i(r)$, can be written as the product of a shape function and a constant:

$$\Gamma_i(r) = \frac{{\rm d}\Lambda_i}{{\rm d}r} = Z_i f_i(r), \quad \text{for } i \in \{0, 1, 2\}. \tag{5}$$

The shape function is normalized to unity. The Z_i's can be interpreted as each system type's occurrence rate Λ_i, integrated over all planetary radii. They are equivalent to the number of planets per single star, primary, or secondary.

2.2. *Searchable stars*

For a given planet radius, semimajor axis, and stellar luminosity, a fixed signal-to-noise floor yields a maximum searchable distance (Pepper et al. 2003; Pepper & Gaudi 2005). From Eq. 1, this maximum searchable distance scales as

$$d(\delta_{\rm obs}, L_{\rm sys}) \propto \delta_{\rm obs} \cdot L_{\rm sys}^{1/2}. \tag{6}$$

Assuming that stars are uniformly distributed in space, the number of searchable stars $N_{\rm s}$ is then proportional to

$$N_{\rm s}(\delta_{\rm obs}, L_{\rm sys}) \propto n\delta_{\rm obs}^3 L_{\rm sys}^{3/2}, \tag{7}$$

where n is the volume density of *e.g.*, single star, or binary systems. We neglect the dependence of n on stellar type.

2.3. *Relation between Apparent and Actual Stellar Properties*

We assume that the apparent properties of an unresolved binary are identical to the true properties of the primary:

$$M_{\rm a} = M_1, \quad R_{\rm a} = R_1, ... \tag{8}$$

We also take the stellar radius and luminosity to be uniquely related to the stellar mass. Given these assumptions, the total luminosity of a system is

$$L_{\rm sys} = L_1 + L_2 = L(M_{\rm a}) + L(qM_{\rm a}), \tag{9}$$

where $q = M_2/M_1$. Note that $L_{\rm sys}$ is the true value (because $M_{\rm a}$ is the true primary mass), while an observer with no a priori knowledge of the distance to a system would measure a flux, assume a stellar mass M_a, and estimate a system luminosity of $L(M_{\rm a})$.

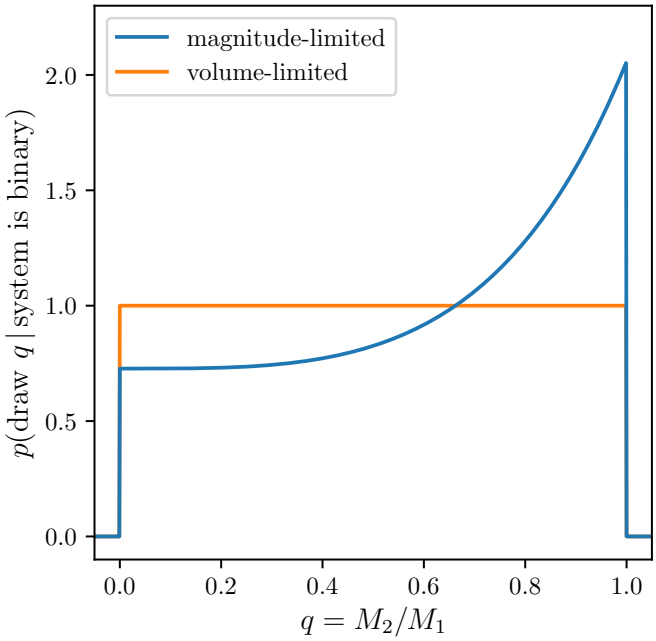

Figure 1. The mass ratio distribution for a magnitude-limited sample of binary stars, in which the underlying volume-limited distribution is uniform, qualitatively similar to *e.g.*, Raghavan et al. (2010)'s Fig. 16. At a given observed depth, the searchable binaries in a transit survey are magnitude-limited (see Sec. 2.4).

2.4. *Apparent number of searchable stars*

Given the observed signal $\delta_{\rm obs}$ and apparent stellar mass $M_{\rm a}$, the maximum searchable distance for singles and binaries are proportional to $\delta_{\rm obs} \cdot L(M_{\rm a})^{1/2}$ and $\delta_{\rm obs} \cdot [L(M_{\rm a}) + L(qM_{\rm a})]^{1/2}$. Thus, the apparent number of searchable stars (points in the sky), which will be selected by an ignorant observer, can be written as

$$N_{\rm s,a}(\delta_{\rm obs}, M_{\rm a}) = N_{\rm s}^0(\delta_{\rm obs}, L(M_{\rm a}))\,[1 + \mu({\rm BF}, M_{\rm a})]\,, \tag{10}$$

where $N_{\rm s}^0$ is the number of searchable singles (this agrees with the actual value), ${\rm BF} = n_{\rm b}/(n_{\rm s} + n_{\rm b})$ is the binary fraction in a volume-limited sample, n_s (n_b) is the number density of single (binary) systems in a volume-limited sample, and μ is the ratio of the number of double to single star systems that are searchable for an observed signal $\delta_{\rm obs}$. Applying Eq. 7,

$$\mu({\rm BF}, M_{\rm a}) = \frac{N_{\rm d}}{N_s^0} = \int_0^1 \frac{n_{\rm b}}{n_{\rm s}} \left[1 + \frac{L(qM_{\rm a})}{L(M_{\rm a})}\right]^{3/2} f(q){\rm d}q, \tag{11}$$

where $f(q)$ is the distribution of binary mass ratios in a volume-limited sample. If BF is independent of q, $L \propto M^\alpha$, and $f(q) \propto q^\beta$, this reduces to

$$\mu = \frac{\rm BF}{1 - {\rm BF}} \cdot \frac{1}{1+\beta} \int_0^1 (1 + q^\alpha)^{3/2}\, q^\beta {\rm d}q, \tag{12}$$

which may be written in a closed form using the hypergeometric function.

This can be understood as a Malmquist bias: at fixed observed transit depth, for singles and primaries with the same luminosity, binaries are searchable to a greater distance. In particular, the sample of searchable binaries is magnitude-limited. Given a searchable binary, the probability of drawing a mass ratio q will scale as

$$p(\text{draw } q \,|\, \text{system is binary}) \propto (1+q^{\alpha})^{3/2} q^{\beta} \tag{13}$$

where q^{β} is the volume-limited probability of drawing a binary of mass ratio q, and the first term is the Malmquist bias. We show this magnitude-limited mass ratio distribution for the $\beta = 0$ case in Fig. 1. In Monte Carlo simulations of transit surveys, it is important to draw binaries from a correctly biased mass-ratio distribution (*e.g.*, Bakos et al. 2013; Sullivan et al. 2015; Günther et al. 2017).

3. IDEALIZED MODELS OF TRANSIT SURVEYS

3.1. *Model #1: fixed stars, fixed planets, twin binaries*

Since the effects of binarity are most pronounced when the two components are similar, we first consider a universe in which all planets are identical, and all stars are identical except that some fraction of them exist in binaries. Expressed mathematically, from Eqs. 4 and 5 the rate density for a selected sample of stars at planet radius r is

$$\Gamma_{\rm sel}(r) = \delta(r - r_p) \cdot \frac{Z_0 + \mu(Z_1 + Z_2)}{1 + 2\mu}, \tag{14}$$

where μ is given by Eq. 11 with $f(q) = \delta(q-1)$, δ is the Dirac delta function, and the true planet radius is r_p. We will generally write rate densities in terms of μ's, rather than N_i's, because μ's are independent of planet radius.

We return to our group of binarity-ignoring astronomers. They select all the points on-sky that they think are searchable: those that are above some flux limit (at each $\delta_{\rm obs}$). The relevant stellar samples are illustrated in Fig. 2. At an apparent radius equal to the true radius ($r_{\rm a} = r_p$), the observer thinks all searchable stars are singles, with $d < d_0$. In fact, binaries with $r_{\rm a} = r_p$ are also searchable, out to $\sqrt{2}d_0$. However, in this model no such systems exist; all planets in twin binaries have true radii $r = r_p$, and so are observed with apparent radii $r_{\rm a} = r_p/\sqrt{2}$. This is because they are "diluted": for an unresolved point-source,

$$\delta_{\rm obs} = \left(\frac{r_{\rm a}}{R_{\rm a}}\right)^2 = \left[\frac{r}{R_{\rm host}}\right]^2 \times \frac{L_{\rm host}}{L_{\rm sys}(M_a, q)}. \tag{15}$$

At an apparent radius of $r_{\rm a} = r_p/\sqrt{2}$, there might also be planets with true radii $r = r_p/\sqrt{2}$ orbiting a single — however none of these exist in this model either. Following Eq. 2, the apparent rate density computed without considering binarity is

$$\Gamma_a(r_a) = \frac{1}{1+\mu} Z_0 \cdot \delta(r_{\rm a} - r_p) + \frac{\mu}{1+\mu}(Z_1 + Z_2) \cdot \delta\left(r_{\rm a} - \frac{r_p}{\sqrt{2}}\right), \tag{16}$$

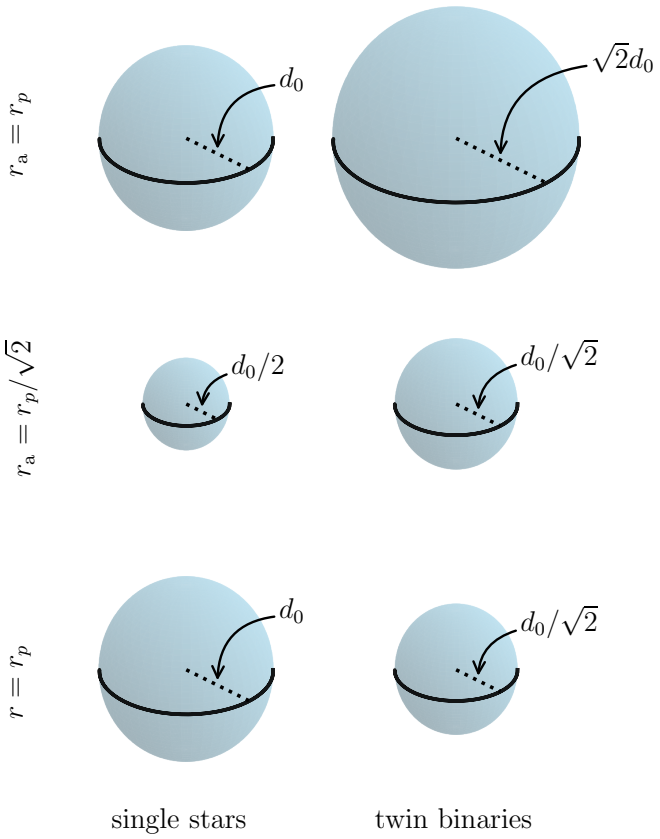

Figure 2. Cartoon of the searchable volumes for single stars (*left*), and twin binaries (*right*). This model (#1) assumes all stars have equal mass and luminosity, and all planets have the same true radius $r = r_p$. *Top:* At an apparent radius $r_a = r_p$, the observer searched twin binaries out to a larger distance, $\sqrt{2}\times$ that of single stars. Because of dilution, there are no planets in twin binaries with $r_a = r_p$. *Middle:* At an apparent radius $r_a = r_p/\sqrt{2}$, the searched volumes are half those at $r_a = r_p$. The only detected planets with $r_a = r_p/\sqrt{2}$ orbit twin binaries. *Bottom:* Planets with a true radius $r = r_p$ are searchable to a maximum distance d_0 around singles, and $d_0/\sqrt{2}$ around twin binaries. These distances are the same as what an observer selects based on apparent radii.

The rate density for selected stars (Eq. 14) and the apparent rate density (Eq. 16) differ in that (A) the total number of selected stars, $N_{\rm tot} = N_0 + N_1 + N_2$, was miscounted; and (B) the inferred radii in binary systems are all $\sqrt{2}$ too small.

To assess the severity of these errors, we need to assume a binary fraction, and to know the true number of planets per single, primary, and secondary star (Z_0, Z_1, and Z_2). For the former, Raghavan et al. (2010) found a multiplicity fraction[1] of 0.44 for primaries with masses from $0.7M_\odot$ to $1.3M_\odot$. "Twin binaries" are perhaps 10-20% of the binaries in a volume-limited sample, depending on how "twin" is defined. Thus

[1] The binary fraction is the fraction of systems in a volume-limited sample that are binary. It is equivalent to the multiplicity fraction if there are no triple, quadruple, or higher order multiples.

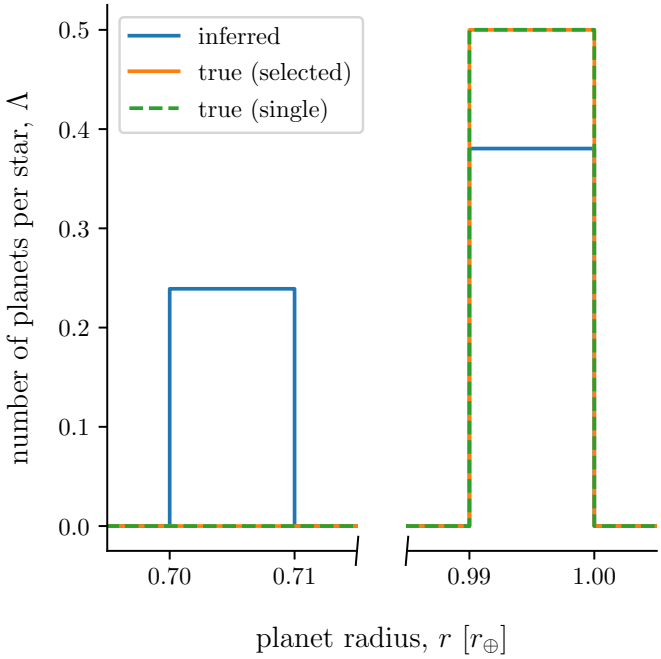

Figure 3. Inferred occurrence rates over $0.01 r_\oplus$ bins in planet radius, for Model #1. This model has fixed stars, fixed planets, and twin binaries. We assume a twin binary fraction of BF = 0.1, and that all stars (single, primary, and secondary) host planets at equal rates. If the true planet radius is r_p, all planets detected in binaries will have apparent radii $r_\mathrm{a} = r_p/\sqrt{2}$.

a representative twin binary fraction for this model is BF $\approx$ 0.05-0.1. For the latter, taking $Z_0 = Z_1 = Z_2 = 1/2$, we plot the resulting occurrence rates as a function of planet radius in Fig. 3. When the Z_i's are equal, the rate density for singles, $\Gamma_0(r)$ (dashed green line), is equal to the rate density for selected stars, $\Gamma_\mathrm{sel}(r)$ (solid orange line).

Correction to inferred rate density and inferred rate—We can define a rate density "correction factor", X_Γ, as the ratio of the apparent to single star rate density,

$$X_\Gamma(r) \equiv \left.\frac{\Gamma_a(r_\mathrm{a})}{\Gamma_0(r)}\right|_{r_\mathrm{a}\to r}. \tag{17}$$

Continuing in the assumption that the number of planets per single, primary, and secondary star are equal, we find a rate density correction factor at the true planet radius of $1/(1+\mu)$. This yields a correction of 0.76 if BF = 0.1, and 0.87 if BF = 0.05. Rephrasing the result, if the twin binary fraction were $(2^{3/2}+1)^{-1} \approx 0.26$, then the apparent rate would be half the true rate. Fortunately, in most contexts the twin binary fraction is not that high.

An alternative assumption is that secondaries do not host planets. In that case, $Z_0 = Z_1$, and $Z_2 = 0$. The correction to the rate density at the true planet radius becomes $(1+2\mu)/(1+\mu)^2$. This evaluates to 0.94 if BF = 0.1, and 0.98 if BF = 0.05.

While it seems unlikely that secondaries are planet-less, it is worth noting that Γ_a/Γ is sensitive to the relative number of planets per single, primary, and secondary.

On completeness—Our expressions for the apparent rate density (Eqs. 2 and 16) do not explicitly mention detection efficiency. If we were to take a smooth detection probability in S/N, we would need to include an explicit detection efficiency term in Eq. 2 is to inversely weight for planets that had low probabilities of being detected. However, our "detection probability" is a step function in S/N (Eq. 1). In our model, the maximum searchable distance is only a function of the apparent depth and true system luminosity. Even though the observers are selecting stars based on what they think are "true radii" (bottom row of Fig. 2), the actual searchability of a point-source is determined by apparent radius and system luminosity (Eq. 1; top rows of Fig. 2). Thus at a given apparent radius $r_{\rm a}$, the stars (single or binary) that are searchable are exactly the stars that are selected, since dilution affects the apparent radius and the maximum searchable distance in equal and opposite order.

3.2. *Model #2: Power law planets*

3.2.1. *Twin binaries*

Our binary-twin model provides a simple estimate of binarity's systematic effects, but perhaps it is too simple. We begin introducing realism by keeping all binaries as twins, but letting the radius distribution of planets vary. Specifically, we take

$$\Gamma_i(r) = Z_i f(r) = Z_i r^\delta / \mathcal{N}_r, \tag{18}$$

for $f(r)$ the radius shape function, $\mathcal{N}_r$ the shape function's normalization, and Z_i the number of planets per star. It turns out that in this model, the apparent rate density is quite similar to that written in Eq. 16, except that the radius dependence and normalization are slightly different:

$$\Gamma_a(r_a) = \frac{r_a^\delta}{\mathcal{N}_r}\left[\frac{Z_0}{1+\mu} + 2^{\frac{\delta+1}{2}}\frac{\mu}{1+\mu}\left(Z_1+Z_2\right)\right]. \tag{19}$$

The details of this equation come from a more general form derived in the Appendix (Eq. A22).

If the Z_i's are equal, the "correction factor" relative to the rate density for singles becomes quite simple:

$$\left.\frac{\Gamma_a(r_a)}{\Gamma_0(r)}\right|_{r_a\to r} = \frac{1+2^{\frac{\delta+3}{2}}\mu}{1+\mu}. \tag{20}$$

For the case of BF $= 0.1$, $\mu \approx 0.153$. Taking $\delta = -2.92$ from Howard et al. (2012), we get a correction factor $\Gamma_{\rm a}/\Gamma_0 = 1.004$. In other words, the apparent rate density is an *over*estimate of the rate density of selected stars, with a relative difference $\delta\Gamma_0 = |\Gamma_0 - \Gamma_{\rm a}|/\Gamma_0$ of 0.4%. A power law radius distribution $f(r) \propto r^\delta$ diverges for $\delta < 0$. From the work of Howard et al. (2012), this approximation is applicable for $r \gtrsim 2r_\oplus$.

3.2.2. *Varying binaries*

To confirm the intuition that non-twin binaries do not affect the above result, we now assume $f(q) = q^\beta/\mathcal{N}_q$, for $\mathcal{N}_q$ the normalization. This changes μ (Eq. 11). It may also affect the rate density, which could be a function of the stellar mass, which is now varying for secondaries (we assume that singles and primaries have the same mass). Absorbing this dependence into a power law as well,

$$\Gamma_i(r, M) = Z_i \times \frac{r^\delta}{\mathcal{N}_r} \times \frac{M^\gamma}{\mathcal{N}_M}, \tag{21}$$

where Z_i is dimensionless, and the normalization constants carry the units (each side has units $[r^{-1}M^{-1}]$). We assume that stars are a one-parameter family, given by $R \propto M \propto L^{\frac{1}{\alpha}}$, so that a given value of q determines everything about a secondary.

Using Eq. A22, we can show that when $Z_0 = Z_1 = Z_2$,

$$X_\Gamma \equiv \left.\frac{\Gamma_{\rm a}(r_{\rm a}, M_{\rm a})}{\Gamma_0(r, M)}\right|_{r_{\rm a}\to r,\ M_{\rm a}\to M,\ Z_i\text{'s equal}} \tag{22}$$

$$= \frac{1}{1+\mu}\left[1 + \frac{1}{\mathcal{N}_q}\frac{\rm BF}{1-{\rm BF}}\left(\int {\rm d}q\, q^\beta (1+q^\alpha)^{\frac{\delta+4}{2}} + \int {\rm d}q\, q^{\beta+\gamma+\delta+\frac{8}{3}}(1+q^\alpha)^{\frac{3}{2}}(1+q^{-\alpha})^{\frac{\delta+1}{2}}\right)\right]. \tag{23}$$

For $\alpha = 3.5$, $\beta = 0$, $\gamma = 0$, $\delta = -2.92$, the summed integrals in Eq. 23 give $(\ldots) \approx 1.41425$. For BF $= 0.44$, this yields $\Gamma_{\rm a}/\Gamma_0 = 1.014$; the apparent rate density is higher than the rate density around singles by 1.4%. To summarize, this model suggests that for a power-law radius distribution (applicable for $r \gtrsim 2r_\oplus$), binarity influences apparent planet occurrence rates around Sun-like stars at the ∼few percent level.

3.3. *Model #3: Fixed primaries, broken power-law planets and secondaries*

Though the details of the radius distribution $\Gamma_i(r)$ at $r < 2r_\oplus$ are still an active topic of research, we can consider how binarity effects this regime under various reasonable assumptions. For instance, we can assume that the true radius shape function is

$$f(r) \propto \begin{cases} r^\delta & \text{for } r \geq 2r_\oplus \\ \text{constant} & \text{for } r \leq 2r_\oplus. \end{cases} \tag{24}$$

As before, all single and primary stars in this model have identical properties. The secondaries have masses, radii, and luminosities that vary between systems: $R \propto M \propto L^{\frac{1}{\alpha}}$. Our "nominal model" remains the same: $\alpha = 3.5$, $\beta = \gamma = 0$, $\delta = -2.92$.

At apparent radii $r_{\rm a} > 2r_\oplus$, the results of this model are the same as those from Sec. 3.2.2. For $r_{\rm a} < 2r_\oplus$, the equations are quite tedious, though still tractable. To compute $\Gamma_{\rm a}(r_{\rm a})$, we simply insert Eq. 24 into Eq. A22, and integrate using a computer

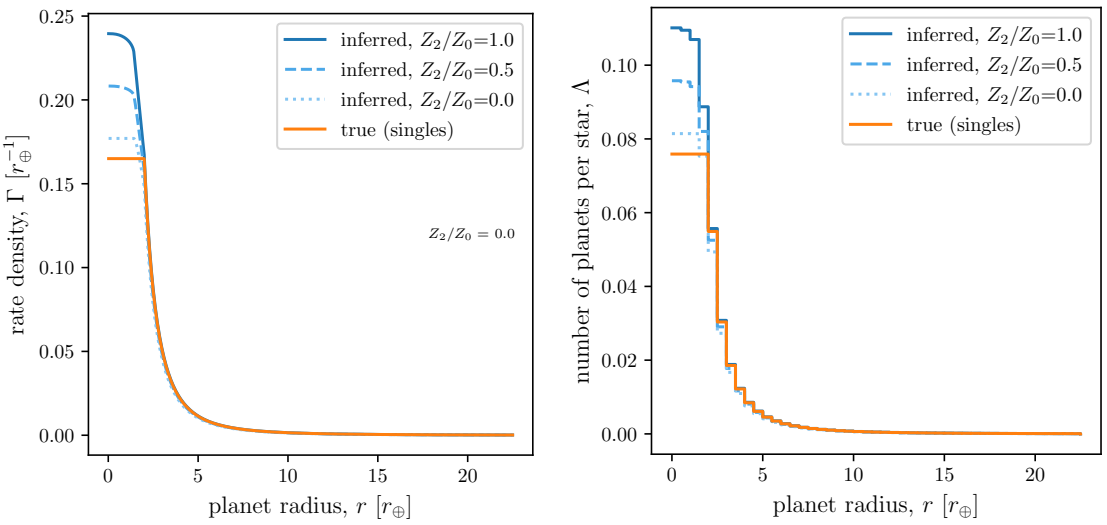

Figure 4. *Left:* rate density and *right:* rate (over $0.5r_\oplus$ bins) as a function of planet radius in Model #3. This model has fixed primaries and single stars, but varying secondaries. The true planet radius distribution is a power law with exponent -2.92 above $2r_\oplus$, below which it is uniform (similar to Howard et al. 2012). We use an arbitrary normalization of $Z_0 = 0.5$ throughout.

program. We refer the interested reader to our online implementation[2]. The output is validated against analytic predictions in the $r_a > 2r_\oplus$ and the $r_a < 2r_\oplus/\sqrt{2}$ regimes, and is plotted in Fig. 4.

The rate of Earth analogs—The most immediately arresting result there is a "bump" in the apparent rate density at $r_a < 2r_\oplus$: the true rate for singles is less than the inferred rate. For the case in which secondaries host as many (half as many) planets as single stars, this means an overestimate of the absolute occurrence rate by $\approx 50\%$ ($\approx 25\%$). The "bump" exists even for the $Z_2/Z_0 = 0$ case as a $\approx 10\%$ effect.

Hot Jupiter occurrence rates—Taking Fig. 4 and integrating the rate density, we can compare the apparent hot Jupiter occurrence rate with the true rate for singles. In the most extreme case of $Z_2/Z_0 = 0$, we find that $\Lambda_{HJ,0}/\Lambda_{HJ,a} = 1.13$, where

$$\Lambda_{HJ,a} = \int_{8r_\oplus}^{\infty} \Gamma_a(r)\, dr, \tag{25}$$

and similarly for $\Lambda_{HJ,0}$. If $Z_2/Z_0 > 0$, binarity affects this value less: when $Z_2/Z_0 = 0.5$, we find $\Lambda_{HJ,0}/\Lambda_{HJ,a} = 1.06$

3.4. *Further models: radius gaps, Gaussian HJ distributions*

The radius distribution specified by Eq. 24 misses some important features recently reported by state of the art occurrence rate studies.

[2] https://github.com/lgbouma/binary_biases

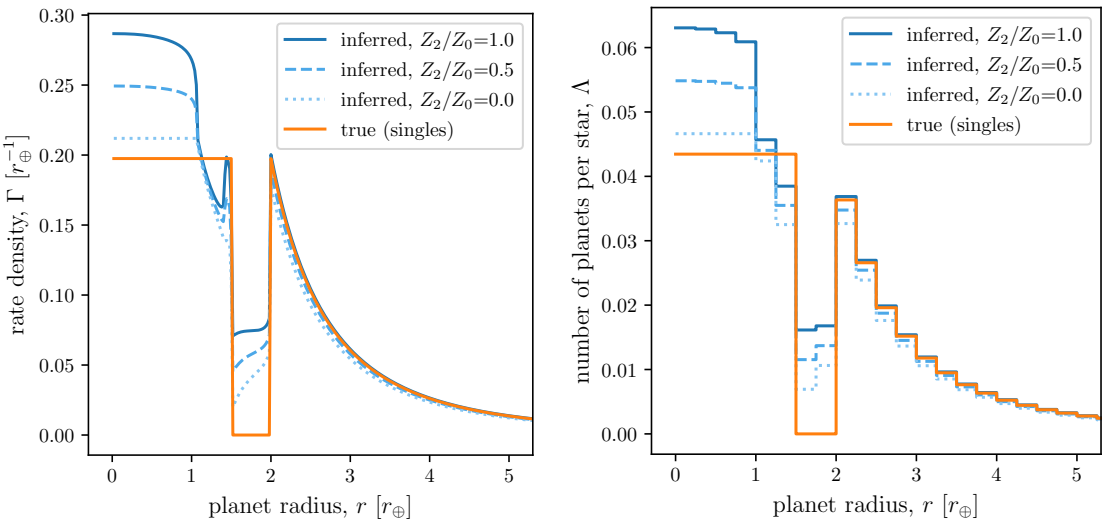

Figure 5. *Left:* rate density and *right:* rate (over $0.25r_\oplus$ bins) as a function of planet radius in a model with a radius gap (Eq. 26). Similar to Model #3, this model has fixed primaries and single stars, but varying secondaries.

Precise features of the radius valley—In particular, Fulton et al. (2017) recently reported a "gap" in the radius-period plane (Petigura et al. 2017a; Johnson et al. 2017). The existence of the gap has been independently corroborated from a sample of KOIs with asteroseismically-determined stellar parameters (Van Eylen et al. 2017). Precise measurement of the gap's features, in particular its width, depth, and shape, will require accurate occurrence rates. To illustrate binarity's role in this problem, we make identical assumptions as in Sec. 3.3, but instead assume an intrinsic radius distribution

$$f(r) \propto \begin{cases} r^{\delta} & \text{for } r \geq 2r_\oplus, \\ 0 & \text{for } 1.5r_\oplus < r < 2r_\oplus, \\ \text{constant} & \text{for } r \leq 1.5r_\oplus. \end{cases} \tag{26}$$

The resulting true and inferred rates are shown in Fig. 5. If left uncorrected, binarity makes the gap appear more shallow, and flattens the step-function edges. Of course, other effects could also "blur" the gap in the planet radius dimension. In particular, the valley's period-dependence is almost certainly not flat (Van Eylen et al. 2017; Owen & Wu 2017). This means that any study measuring the gap's width or depth in the face of binarity must either perform tests at fixed orbital period, or else marginalize over period and account for the associated blurring in the planet radius dimension.

Alternative models for the HJ distribution—In the recent work by Petigura et al. (2017b), hot Jupiters appear as an island in period-radius space, rather than as a continuous component of a power law distribution. Consider instead a Gaussian

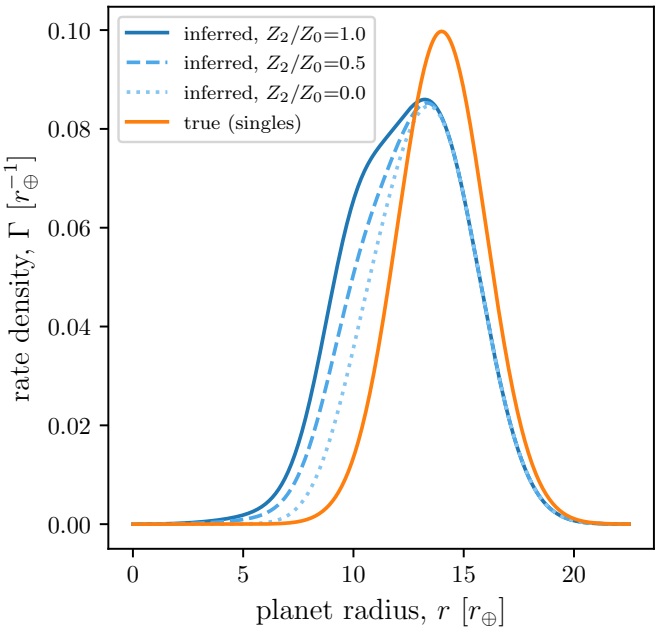

Figure 6. Rate density for a population of planets with true radii r drawn from $\mathcal{N}(14r_\oplus, 2r_\oplus)$. This is similar to the hot Jupiter distribution presented by Petigura et al. (2017b).

radius shape function,

$$f(r) \propto \exp\left(-\frac{(r-\bar{r})^2}{2\sigma^2}\right), \tag{27}$$

with $\bar{r} = 14r_\oplus$ and $\sigma = 2r_\oplus$. As always, $\Gamma_i(r) = Z_i f(r)$. We then compute $\Gamma_{\rm a}(r_{\rm a})$, and plot it in Fig. 6. Integrating to find hot Jupiter rates, we find the opposite effect as in the power law model. For $Z_2/Z_0 > 0$, the apparent HJ rate is *greater* than the true rate for singles. The effect is maximal when there are as many hot Jupiters orbiting secondaries as singles, in which case $\Lambda_{\rm HJ,0}/\Lambda_{\rm HJ,a} = 0.84$ For the case when no hot Jupiters orbit secondaries, $\Lambda_{\rm HJ,0} \approx \Lambda_{\rm HJ,a}$ to within one percent.

4. DISCUSSION

How has binarity been considered in occurrence rate measurements?—This study has shown that under a reasonable set of simplifying assumptions, binarity can introduce systematic errors to star and planet counts in transit surveys, which then bias derived occurrence rates. Thus far, real-world occurrence rate calculations[3] using transit survey data have mostly ignored stellar multiplicity (*e.g.*, Howard et al. 2012; Fressin et al. 2013; Foreman-Mackey et al. 2014; Dressing & Charbonneau 2015; Burke et al. 2015). For *Kepler* occurrence rates specifically, it seems that no one has yet carefully assessed binarity's importance, or lack thereof. While we do not resolve

[3] A list of occurrence rate papers is maintained at https://exoplanetarchive.ipac.caltech.edu/docs/occurrence_rate_papers.html

the problem, we do suggest the approximate scale of the necessary corrections in a survey-independent manner.

Of course, on a system-by-system level stellar multiplicity affects the interpretation of planet candidates. High resolution imaging campaigns have measured the multiplicity of almost all *Kepler* Objects of Interest (Howell et al. 2011; Adams et al. 2012, 2013; Horch et al. 2012, 2014; Lillo-Box et al. 2012, 2014; Dressing et al. 2014; Law et al. 2014; Cartier et al. 2015; Everett et al. 2015; Gilliland et al. 2015; Wang et al. 2015b,c; Baranec et al. 2016; Ziegler et al. 2017). The results of these programs have been collected by Furlan et al. (2017), and they represent an important advance in understanding the KOI sample's multiplicity statistics. In particular, they can be immediately applied to rectify binarity's effects on the mass-radius diagram (Furlan & Howell 2017).

The high resolution imaging campaign is also beginning to connect with occurrence rate calculations. The most recent rate studies have used Furlan et al. (2017)'s catalog to test the effects of removing KOI hosts with known companions, which helps reduce contamination in the "numerator" of the occurrence rate (Fulton et al. 2017; Petigura et al. 2017b). However, without an understanding of the multiplicity statistics of the non-KOI stars, the true number of searchable stars, and thus the true occurrence rates, will remain biased. The first-order correction that we suggest, given the impracticality of performing high-resolution imaging of every selected star in a transit survey, is to model the detection pipeline's efficiency while accounting for binarity. For *Kepler*, this would require high resolution imaging of a comparison sample of non-KOI host stars. If the associated multiplicity statistics are then included in a model of the pipeline's detection efficiency, and the number of selected stars is appropriately counted, it would correct many of binarity's biases.

The rate of Earth analogs—Per *Kepler*'s primary science objective, the rate of Earth-like planets orbiting Sun-like stars has been independently measured by Youdin (2011); Petigura et al. (2013); Dong & Zhu (2013); Foreman-Mackey et al. (2014), and Burke et al. (2015). These efforts have found that the one-year terrestrial planet occurrence rate varies between ≈ 0.03 and ≈ 1 per Sun-like star, depending on assumptions that are made when retrieving the rate (Burke et al. 2015's Fig. 17). In our Model #3, the inferred rate can be up to 50% higher than the true rate around single stars. Though this bias seems large, it is currently small compared to the other systematic factors that dominate the dispersion in $\eta_\oplus$ measurements. If a future analyses determine absolute values of $\eta_\oplus$ to better than a factor of two, binarity might merit closer attention.

One relevant caveat to our assessment of binarity's importance for $\eta_\oplus$ measurements is that none of our models included the rate density's period-dependence. However, close binaries usually provoke dynamical instabilities, leading to fewer long-period planets per star (*e.g.*, Holman & Wiegert 1999; Wang et al. 2014; Kraus et al. 2016).

This effect might further bias transit survey measurements of $\eta_\oplus$ beyond our rough estimate.

The hot Jupiter rate discrepancy—There is at least one context in which ignoring binarity has been suggested as a possible source of discrepant occurrence rate measurements. Hot Jupiter occurrence rates measured by transit surveys ($\approx 0.5\%$) are marginally lower than those found by radial velocity surveys ($\approx 1\%$; see Table 1). Though the discrepancy has weak statistical significance ($< 3\sigma$), one reason to expect a difference is that the corresponding stellar populations have distinct metallicities. As argued by Gould et al. (2006), the RV sample is biased towards metal-rich stars, which have been measured by RV surveys to preferentially host more giant planets (Santos et al. 2004; Fischer & Valenti 2005). Investigating the discrepancy from the metallicty angle, Guo et al. (2017) measured the *Kepler* field's mean metallicity to be $[\mathrm{M/H}]_{\mathrm{Kepler}} = -0.045 \pm 0.009$, which is lower than the California Planet Search's mean of $[\mathrm{M/H}]_{\mathrm{CPS}} = -0.005 \pm 0.006$. The former value agrees with that measured by Dong et al. (2014). Based on their measurements, Guo et al. then argued that the metallicity difference could account for a $\approx 20\%$ relative difference in the measured rates between the CKS and *Kepler* samples – not a factor of two. Guo et al. concluded that "other factors, such as binary contamination and imperfect stellar properties" must also be at play.

Aside from surveying stars of varying metallicities, radial velocity and transit surveys differ in how they treat binarity. Radial velocity surveys typically reject both visual and spectroscopic binaries (*e.g.*, Wright et al. 2012). Transit surveys typically observe binaries, but the question of whether they were searchable to begin with is left for later interpretation. In spectroscopic follow-up of candidate transiting planets, the prevalence of astrophysical false-positives may also lead to a bias against confirmation of transiting planets in binary systems.

Ignoring these complications, in this work we showed that binarity biases transit survey occurrence rates through its effects on star counts and the apparent radii of detected planets. Our results from Sec. 3.3 indicate that binarity could lead to underestimated HJ rates relative to singles by a multiplicative factor of at most ≈ 1.13 (assuming a power-law radius distribution, and that no secondaries host HJs). We later pointed out in Sec. 3.4 that this power-law radius distribution is probably unrealistic. If we instead assumed a Gaussian radius distribution (e.g., Petigura et al. 2017b), apparent HJ rates are actually *greater* than the true rate around singles.

If we believe the assumed power-law radius distribution, we can ask whether this "correction factor" might help resolve the discrepancy. Explicitly, we ask: what is the probability of Wright et al. (2012)'s result, given a rate drawn from the stated bounds of Petigura et al. (2017b)? In other words, we first take the true HJ rate per thousand stars as $\Lambda_{\mathrm{HJ}} = 5.7 \pm 1.3$, with Gaussian uncertainties. We then draw from a Poisson distribution and compute the probability of detecting at least 10 hot Jupiters in a sample of 836 stars. Without accounting for binarity or metallicity, only

4% of RV surveys would detect at least 10 hot Jupiters. If we multiply $\Lambda_{\rm HJ}$ by 1.2 to account for Guo et al. (2017)'s measured metallicity difference between the *Kepler* field and the local solar neighborhood, 9% of RV surveys would detect at least 10 hot Jupiters. If we multiply once more by 1.13 to account for binarity's supposed bias, we find that 14% of RV surveys would detect at least 10 hot Jupiters, still suggesting a weak discrepancy.

This suggestion should be taken with caution, because it (likely incorrectly) assumes a power-law radius distribution. Assuming a Gaussian similar to that reported by Petigura et al. (2017b), we find that binarity is unlikely to resolve the HJ rate discrepancy. Of course, such calculations can be at most suggestive – a true resolution of binarity's effects may require a detailed understanding of the *Kepler* field's multiplicity statistics, and the mission's completeness (both for candidate detection and follow-up). For instance, if hot Jupiters are less likely to be confirmed in binary systems, this might further bias the rates low.

Does a detected planet orbit the primary or secondary?—Ciardi et al. (2015) studied the effects of stellar multiplicity on the planet radii derived from transit surveys. They modeled the problem for *Kepler* objects of interest by matching a population of binary and tertiary companions to KOI stars, under the assumption that the KIC-listed stars were the primaries. They then computed planet radius correction factors assuming that *Kepler*-detected planets orbited the primary or companion stars with equal probability (their Sec. 5). Under these assumptions, they found that any given planet's radius is on average underestimated by a multiplicative factor of $\approx$1.5. Ziegler et al. (2017) recently reported a similar value for the real population of systems observed by the Robo-AO KOI survey, under the same assumption.

Our models indicate that assuming a detected planet has equal probability of orbiting the primary or secondary leads to an overestimate of binarity's population-level effects. A planet orbiting the secondary does lead to extreme corrections, but these cases are rare outliers, because the searchable volume for secondaries is so much smaller than that for primaries (see Fig. 2). This means that when high-resolution imaging discovers a binary companion in a system that hosts a detected transiting planet, the planet is much more likely to orbit the primary. This statement is independent of the fact that planets are often confirmed to orbit the primary by inferring the stellar density from the transit duration.

Utility in future occurrence rate measurements—*TESS* is expected to discover over 10^4 giant planets (Ricker et al. 2014; Sullivan et al. 2015). Though they will be difficult to distinguish from false positives, one possible use of this large sample will to measure an occurrence rate of short-period giant planets with minimal error from counting statistics. Our models suggest that binarity will only become an appreciable fraction of the error budget at the $\approx$ 10% precision level. This should be sufficient for a rate

measurement precise enough to indicate a preference between $\approx 0.5\%$ and $\approx 1\%$ of Sun-like stars hosting hot Jupiters.

5. CONCLUSION

This study presented simple models for the effects of binarity on occurrence rates measured by transit surveys. The simplest of these models (Model #1) provided an order-of-magnitude estimate of how much binarity might affect occurrence rates. The most realistic of these models (Model #3) suggests that binarity leads to overestimates in transit survey occurrence rates at planetary radii $< 2r_\oplus$, with at most $\approx 50\%$ relative error. Although this is smaller than current systematic uncertainties on the occurrence rates of Earth-sized planets, this means that binarity could eventually become an important component of the $\eta_\oplus$ error budget. Further, we showed that assuming a power-law radius distribution, then hot Jupiter rates measured by transit surveys could be biased to infer $\approx 1.13\times$ fewer hot Jupiters per star than surveys that only measure occurrence rates about single stars (*i.e..,* radial velocity surveys). However, assuming a more realistic Gaussian radius distribution indicates that the effects of binarity on measured hot Jupiter rates are either negligible, or else lead to *over*estimated rates. The overall result is that our models do not suggest that binarity can resolve the HJ rate discrepancy. Finally, we also pointed out that binarity can substantially "fill in" gaps in the apparent radius distribution (Fig. 5). This will likely become an important consideration for precise measurements of the depth, width, and slope of Fulton et al. (2017)'s radius gap.

It was a pleasure discussing this work with F. Dai and T. Barclay.

Software: `numpy` (Walt et al. 2011), `scipy` (Jones et al. 2001), `matplotlib` (Hunter 2007), `pandas` (McKinney 2010)

APPENDIX

A. APPARENT OCCURRENCE RATE — GENERAL FORMULA

Recall our definition of apparent rate density (Eq. 2). Let us explicitly write $\mathcal{P} = r$ and $\mathcal{S} = M$; R and L are uniquely determined from the assumed stellar mass–radius–luminosity relation. We neglect the dependence on planetary orbital period. The planets with $(r_\mathrm{a}, M_\mathrm{a})$ are associated with systems of many different planetary and stellar properties, so $n_\mathrm{det}(r_\mathrm{a}, M_\mathrm{a})$ is given by the convolution of the true rate density, $\Gamma(r, M)$, and $\mathcal{N}(r_\mathrm{a}, M_\mathrm{a}; r, M)$, the number (per unit $(r_\mathrm{a}, M_\mathrm{a})$) of searchable stars that give $(r_\mathrm{a}, M_\mathrm{a})$ when the true system actually has (r, M). Mathematically,

$$n_\mathrm{det}(r_a, M_a) = \sum_i n^i_\mathrm{det}(r_a, M_a) \tag{A1}$$

$$= \sum_i \int \mathrm{d}r\mathrm{d}M\, \mathcal{N}^i_\mathrm{s}(r_a, M_a; r, M) \cdot \Gamma_i(r, M) \cdot p_\mathrm{tra}(r, M), \tag{A2}$$

where i specifies the type of true host stars (0: single, 1: primary, 2: secondary). The problem reduces to the evaluation of

$$\mathcal{N}_{\rm s}^{i}(\mathcal{P}_{\rm a},\mathcal{S}_{\rm a};\mathcal{P},\mathcal{S}) \tag{A3}$$

for planets around single stars, primaries in binaries, and secondaries in binaries.

Single stars—For $i = 0$,

$$\mathcal{N}_{\rm s}^{0}(r_{\rm a},M_{\rm a};r,M) = \delta(r_{\rm a}-r)\delta(M_{\rm a}-M)N_{\rm s}^{0}(r,M), \tag{A4}$$

so

$$n_{\rm det}^{0}(r_{\rm a},M_{\rm a}) = N_{\rm s}^{0}(r_{\rm a},M_{\rm a})\cdot\Gamma_{0}(r_{\rm a},M_{\rm a})\cdot p_{\rm tra}(r_{\rm a},M_{\rm a}). \tag{A5}$$

Primaries in binaries—The number of primaries with apparent parameters (r_a, M_a) given the true parameters (r, M) is

$$\mathcal{N}_{\rm s}^{1}(r_a,M_a;r,M) = \int {\rm d}q\, f(q)\mathcal{N}_{{\rm s},q}^{1}(r_a,M_a,q;r,M), \tag{A6}$$

where $f(q)$ is the volume-limited binary mass ratio distribution.

Since we assume $\mathcal{S}_{\rm a} = \mathcal{S}_1$,

$$\mathcal{N}_{{\rm s},q}^{1}(r_a,M_a,q;r,M) \propto \delta(M_a-M). \tag{A7}$$

In this case, $\mathcal{N}_{{\rm s},q}^{1}$ is non-zero only at $r_a = R_a\sqrt{\delta_{\rm obs}}$, and the observed depth is

$$\delta_{\rm obs} = \left[\frac{r}{R(M_a)}\right]^2 \times \frac{L(M_a)}{L_{\rm sys}(M_a,q)} \equiv \left[\frac{r}{R(M_a)}\right]^2 \times \mathcal{A}(q,M_{\rm a})^2 \tag{A8}$$

The normalization of $\mathcal{N}_{{\rm s},q}^{1}$ is given by the number of binaries that are searchable for a signal $\delta_{\rm obs}$ and that have the mass ratio q:

$$N_{\rm s}^{0}(\delta_{\rm obs},L(M_a))\cdot\frac{n_b}{n_s}\left[\frac{L_{\rm sys}(M_a,q)}{L(M_{\rm a})}\right]^{3/2} = N_{\rm s}^{0}(\delta_{\rm obs},L(M_a))\cdot\frac{\rm BF}{1-\rm BF}\cdot\frac{1}{\mathcal{A}(q,M_{\rm a})^3}. \tag{A9}$$

Thus,

$$\begin{aligned}\mathcal{N}_{{\rm s},q}^{1}(r_a,M_a,q;r,M) = {} & N_{\rm s}^{0}(\delta_{\rm obs},L(M_a))\cdot\frac{\rm BF}{1-\rm BF}\cdot\frac{1}{\mathcal{A}(q,M_{\rm a})^3} \\ & \times\,\delta\left[r_a - r\mathcal{A}(q,M_{\rm a})\right]\delta(M_a-M).\end{aligned} \tag{A10}$$

Secondaries in binaries—In this case, $M = qM_1 = qM_a$, so

$$\mathcal{N}_{{\rm s},q}^{2}(r_a,M_a,q;r,M) \propto \delta\left(M_a - \frac{M}{q}\right). \tag{A11}$$

Again $\mathcal{N}_{\rm s}^{2}$ is non-zero only at $r_a = R_a\sqrt{\delta_{\rm obs}}$, but this time

$$\delta_{\rm obs} = \left[\frac{r}{R(qM_a)}\right]^2 \times \frac{L(qM_a)}{L_{\rm sys}(M_a,q)}. \tag{A12}$$

The normalization remains the same as the previous case (we are counting the searchable stars at a given observed depth $\delta_{\rm obs}$, total luminosity of the binary is the same). Thus,

$$\mathcal{N}_{\rm s}^2(r_a, M_a; r, M) = \int {\rm d}q\, f(q) \mathcal{N}_{{\rm s},q}^2(r_a, M_a, q; r, M), \tag{A13}$$

where

$$\mathcal{N}_{\rm s}^2(r_a, M_a; r, M; q) = N_{\rm s}^0(\delta_{\rm obs}, L(M_a)) \cdot \frac{\rm BF}{1-{\rm BF}} \cdot \frac{1}{\mathcal{A}(q, M_{\rm a})^3} \times \delta\left(r_a - r\sqrt{\left[\frac{R(M_a)}{R(qM_a)}\right]^2 \frac{L(qM_a)}{L_{\rm sys}(M_a, q)}}\right) \delta\left(M_a - \frac{M}{q}\right). \tag{A14}$$

One might worry in Eq. A14 that we opt to write $\mathcal{N}_{\rm s}^2 \propto \delta(M_a - M/q)$, rather than $\propto \delta(M_a q - M)$ or some other delta function with the same functional dependence, but a different normalization once integrated. We do this because the delta function in Eq. A14 is defined with respect to the measure ${\rm d}M_{\rm a}$, not ${\rm d}M$. This is because $\mathcal{N}_{\rm s}^2$ is defined as a number per $r_{\rm a}$, per $M_{\rm a}$.

Number of detected planets—Marginalizing per Eq. A2, we find

$$\begin{aligned} n_{\rm det}^0(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}_{\rm s}^0(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma^0(r, M) \cdot p_{\rm tra}(M) \\ &= N_{\rm s}^0(\delta_{\rm obs}, L(M_{\rm a})) \cdot \Gamma^0(r_{\rm a}, M_{\rm a}) \cdot p_{\rm tra}(M_{\rm a}), \end{aligned} \tag{A15}$$

and

$$\begin{aligned} n_{\rm det}^1(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}_{\rm s}^1(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma^1(r, M) \cdot p_{\rm tra}(M) \\ &= N_{\rm s}^0(\delta_{\rm obs}, L(M_{\rm a})) \cdot p_{\rm tra}(M_{\rm a}) \cdot \frac{\rm BF}{1-{\rm BF}} \int \frac{{\rm d}q}{\mathcal{A}^4} f(q)\, \Gamma^1\left(\frac{r_a}{\mathcal{A}}, M_{\rm a}\right), \end{aligned} \tag{A16}$$

where

$$\mathcal{A}(q, M_{\rm a}) = \sqrt{\frac{L(M_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)}}. \tag{A17}$$

Finally,

$$n_{\rm det}^2(r_{\rm a}, M_{\rm a}) = \int {\rm d}r{\rm d}M\, \mathcal{N}_{\rm s}^2(r_{\rm a}, M_{\rm a}; r, M) \cdot \Gamma^2(r, M) \cdot p_{\rm tra}(M) \tag{A18}$$

$$= N_{\rm s}^0(\delta_{\rm obs}, L(M_{\rm a})) \cdot \frac{\rm BF}{1-{\rm BF}} \int \frac{q{\rm d}q}{\mathcal{A}^3 \mathcal{B}} f(q)\, \Gamma^2(r_a/\mathcal{B}, qM_{\rm a})\, p_{\rm tra}(qM_{\rm a}), \tag{A19}$$

where

$$\mathcal{B}(q, M_{\rm a}) = \frac{R(M_{\rm a})}{R(qM_{\rm a})}\sqrt{\frac{L(qM_{\rm a})}{L_{\rm sys}(M_{\rm a}, q)}}. \tag{A20}$$

General formula for apparent occurrence rate—Using the results above, the apparent rate density,

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{N_{\rm s,a}(r_{\rm a}, M_{\rm a}) p_{\rm tra}(r_{\rm a}, M_{\rm a})} \times \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}), \tag{A21}$$

evaluates to

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{1+\mu({\rm BF}, M_{\rm a})} \times \left\{ \Gamma^0(r_{\rm a}, M_{\rm a}) + \frac{\rm BF}{1-{\rm BF}} \left[\int \frac{{\rm d}q}{\mathcal{A}^4} f(q)\, \Gamma^1\left(\frac{r_{\rm a}}{\mathcal{A}}, M_{\rm a}\right) \right.\right.$$
$$\left.\left. + \int \frac{q{\rm d}q}{\mathcal{A}^3\mathcal{B}} f(q)\, \Gamma^2\left(\frac{r_{\rm a}}{\mathcal{B}}, qM_{\rm a}\right) \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3} \right] \right\}. \tag{A22}$$

This equation is used to derive Eqs. 19 and 23, and is numerically integrated to create Figs. 4, 5, and 6. It is validated in limits in which it is possible to write down the answer (e.g., Eq. 16), and also against a Monte Carlo realization of the twin binary models (Secs. 3.1 and 3.2).

Table 1. Occurrence rates of hot Jupiters (HJs) about FGK dwarfs, as measured by radial velocity and transit surveys.

Reference	HJs per thousand stars	HJ Definition
Marcy et al. (2005)	12±2	$a < 0.1$ AU; $P \lesssim 10$ day
Cumming et al. (2008)	15±6	–
Mayor et al. (2011)	8.9±3.6	–
Wright et al. (2012)	12.0±3.8	–
Gould et al. (2006)	$3.1^{+4.3}_{-1.8}$	$P < 5$ day
Bayliss & Sackett (2011)	10^{+27}_{-8}	$P < 10$ day
Howard et al. (2012)	4±1	$P < 10$ day; $r_p = 8 - 32 r_\oplus$; solar subset[a]
–	5±1	solar subset extended to $Kp < 16$
–	7.6±1.3	solar subset extended to $r_p > 5.6 r_\oplus$.
Moutou et al. (2013)	10±3	*CoRoT* average; $P \lesssim 10$ day, $r_p > 4 r_\oplus$
Petigura et al. (2018, in prep)	$5.7^{+1.4}_{-1.2}$	$r_p = 8 - 24 r_\oplus$; $P = 1 - 10$ day; CKS stars[b]
Santerne et al. (2018, in prep)	9.5±2.6	*CoRoT* galactic center
–	11.2±3.1	*CoRoT* anti-center

NOTE— The first four studies use data from radial velocity surveys; the rest are based on transit surveys. Many of these surveys selected different stellar samples. "–" denotes "same as above".

[a] Howard et al. (2012)'s "solar subset" was defined as *Kepler*-observed stars with $4100\,\mathrm{K} < T_{\rm eff} < 6100\,\mathrm{K}$, $Kp < 15$, $4.0 < \log g < 4.9$. They required signal to noise > 10 for planet detection.

[b] Petigura et al. (2018, in prep)'s planet sample includes all KOIs with $Kp < 14.2$, with a statistically insignificant number of fainter stars with HZ planets and multiple transiting planets. Their stellar sample begins with Mathur et al. (2017)'s catalog of 199991 *Kepler*-observed stars. Successive cuts are: $Kp < 14.2$ mag, $T_{\rm eff} = 4700 - 6500$ K, and $\log g = 3.9 - 5.0$ dex, leaving 33020 stars.

REFERENCES

Adams, E. R., Ciardi, D. R., Dupree, A. K., et al. 2012, The Astronomical Journal, 144, 42

Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2013, The Astronomical Journal, 146, 9

Bakos, G. ., Csubry, Z., Penev, K., et al. 2013, PASP, 125, 154

Baranec, C., Ziegler, C., Law, N. M., et al. 2016, The Astronomical Journal, 152, 18

Bayliss, D. D. R., & Sackett, P. D. 2011, The Astrophysical Journal, 743, 103

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, The Astrophysical Journal, 809, 8

Cartier, K. M. S., Gilliland, R. L., Wright, J. T., & Ciardi, D. R. 2015, The Astrophysical Journal, 804, 97

Ciardi, D. R., Beichman, C. A., Horch, E. P., & Howell, S. B. 2015, The Astrophysical Journal, 805, 16

Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, Publications of the Astronomical Society of the Pacific, 120, 531

Dong, S., & Zhu, Z. 2013, The Astrophysical Journal, 778, 53

Dong, S., Zheng, Z., Zhu, Z., et al. 2014, The Astrophysical Journal Letters, 789, L3

Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2014, The Astronomical Journal, 148, 78

Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45

Everett, M. E., Barclay, T., Ciardi, D. R., et al. 2015, The Astronomical Journal, 149, 55

Fischer, D. A., & Valenti, J. 2005, The Astrophysical Journal, 622, 1102

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, The Astrophysical Journal, 795, 64

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, The Astrophysical Journal, 766, 81

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, The Astronomical Journal, 154, 109

Furlan, E., & Howell, S. B. 2017, arXiv:1707.01942 [astro-ph]

Furlan, E., Ciardi, D. R., Everett, M. E., et al. 2017, The Astronomical Journal, 153, 71

Gilliland, R. L., Cartier, K. M. S., Adams, E. R., et al. 2015, The Astronomical Journal, 149, 24

Gould, A., Dorsher, S., Gaudi, B. S., & Udalski, A. 2006, Acta Astronomica, 56, 1

Günther, M. N., Queloz, D., Demory, B.-O., & Bouchy, F. 2017, Monthly Notices of the Royal Astronomical Society, 465, 3379

Guo, X., Johnson, J. A., Mann, A. W., et al. 2017, The Astrophysical Journal, 838, 25

Holman, M. J., & Wiegert, P. A. 1999, The Astronomical Journal, 117, 621

Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2012, The Astronomical Journal, 144, 165

—. 2014, The Astrophysical Journal, 795, 60

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, The Astrophysical Journal Supplement Series, 201, 15

Howell, S. B., Everett, M. E., Sherry, W., Horch, E., & Ciardi, D. R. 2011, The Astronomical Journal, 142, 19

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90

Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, arXiv:1703.10402 [astro-ph]

Jones, E., Oliphant, T., Peterson, P., et al. 2001, Open source scientific tools for Python

Kraus, A. L., Ireland, M. J., Huber, D., Mann, A. W., & Dupuy, T. J. 2016, The Astronomical Journal, 152, 8

Law, N. M., Morton, T., Baranec, C., et al. 2014, The Astrophysical Journal, 791, 35

Lillo-Box, J., Barrado, D., & Bouy, H. 2012, Astronomy and Astrophysics, 546, A10

—. 2014, Astronomy and Astrophysics, 566, A103

Marcy, G., Butler, R. P., Fischer, D., et al. 2005, Progress of Theoretical Physics Supplement, 158, 24

Mathur, S., Huber, D., Batalha, N. M., et al. 2017, The Astrophysical Journal Supplement Series, 229, 30

Mayor, M., Marmier, M., Lovis, C., et al. 2011, ArXiv e-prints, 1109, arXiv:1109.2497

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51

Moutou, C., Deleuil, M., Guillot, T., et al. 2013, Icarus, 226, 1625

Owen, J. E., & Wu, Y. 2017, arXiv:1705.10810 [astro-ph]

Pepper, J., & Gaudi, B. S. 2005, The Astrophysical Journal, 631, 581

Pepper, J., Gould, A., & Depoy, D. L. 2003, Acta Astronomica, 53, 213

Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, Proceedings of the National Academy of Science, 110, 19273

Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017a, arXiv:1703.10400 [astro-ph]

Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2017b, arXiv:1712.04042 [astro-ph]

Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, The Astrophysical Journal Supplement Series, 190, 1

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Santos, N. C., Israelian, G., & Mayor, M. 2004, Astronomy and Astrophysics, 415, 1153

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, The Astrophysical Journal, 809, 77

Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2017, arXiv:1710.05398 [astro-ph]

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science & Engineering, 13, 22

Wang, J., Fischer, D. A., Horch, E. P., & Huang, X. 2015a, The Astrophysical Journal, 799, 229

Wang, J., Fischer, D. A., Horch, E. P., & Xie, J.-W. 2015b, The Astrophysical Journal, 806, 248

Wang, J., Fischer, D. A., Xie, J.-W., & Ciardi, D. R. 2015c, The Astrophysical Journal, 813, 130

Wang, J., Xie, J.-W., Barclay, T., & Fischer, D. A. 2014, The Astrophysical Journal, 783, 4

Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, The Astrophysical Journal, 753, 160

Youdin, A. N. 2011, ApJ, 742, 38

Ziegler, C., Law, N. M., Baranec, C., et al. 2017, arXiv:1712.04454 [astro-ph]