MEMO FOR INTERNAL USE
Preprint typeset using LaTeX style emulateapj v. 12/16/11

A TOY ANALYTIC TRANSIT SURVEY: VARYING THE LIGHT RATIO OF BINARIES

LGB, KM, JNW
Memo for internal use

ABSTRACT

In this memo, we assume the stars in the population of binary systems have varying light ratios.
Subject headings: planets and satellites: detection

STATEMENT OF PROBLEM

Ah, transiting planets! We learn so much by studying them. But how much can we actually learn, and how much is messed up by binarity?

Imagine the following survey, similar to that outlined by [Pepper et al, 2003]:

- You are going to observe the entire sky for a duration $T_{\rm obs}$, with a detector of area A, and known bandpass. Your detector is photon-noise limited.
- You are interested only in detecting planets of radius R_p, and orbital period P. For instance, $R_p = R_\oplus$, $P = 1$ year.
- You only want to detect them around stars of radius R_1, and luminosity L_1. For instance, G2V dwarfs.
- Since your detections will be S/N limited, you want only to observe stars for which your target can be detected with $\mathrm{S/N} > (\mathrm{S/N})_{\rm min}$.
- For a photon-noise limited survey, the signal to noise limit is equivalent to a magnitude (flux) limit. So you select all the points on the sky above a flux limit, *i.e.* with flux $F_{\rm pt} > F_{\rm min}$, for $F_{\rm min}$ defined by your target planet and stellar types, and your survey specifications.
- You carry out a transit survey, and detect many transiting planets.

You now wish to derive an occurrence rate for planets of radius R_p and orbital period P. Assume your universe is a universe in which the true population of "points" (stellar systems, all unresolved) comprises single and double star systems. Single star systems have luminosity in the observed bandpass L_1, radii R_1, and effective temperature $T_{\rm eff,1}$. Double star systems have luminosity in the observed bandpass $L_d = (1+\gamma_R)L_1$, for $\gamma_R = L_2/L_1$ the ratio of the luminosity of the secondary to the primary.

In this memo γ_R varies across the population of star systems. It does so because the underlying mass ratio varies, and $\gamma_R = q^\alpha$. Since we are interested in solar type binaries, we take the distribution of the mass ratio $q = M_2/M_1$ by coarsely approximating [Rhagavan et al 2010, Fig 16]:

$$\mathrm{prob}(q) = \begin{cases} c_q & 0.1 < q \le 1 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

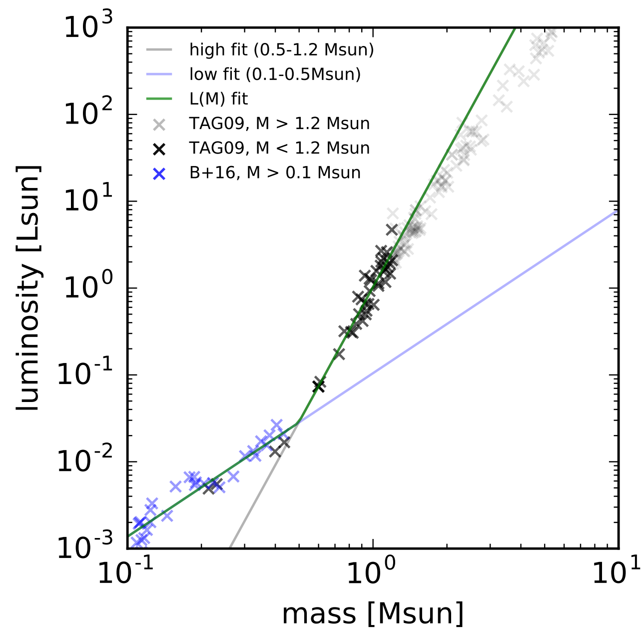

FIG. 1.— Empirical fit to main sequence dwarf mass luminosity data from Torres et al. [2009] and Benedict et al. [2016]. The "low fit" is a least squares fit to data from $0.1 - 0.5 M_\odot$, and the "high fit" is to data above that, and below the Kraft break ($1.2 M_\odot$). The $L(M)$ relation taken for subsequent numerics is the maximum of the two fits.

for $c_q = 1/9$ to normalize the distribution to 1. The mass-luminosity relation can, for analytic convenience, be approximated as $L = M^\alpha$, with the lore-value of α being 3.5. While we use this in subsequent analytic development, for numerics we fit a lines to mass-luminosity data collected by Torres et al. [2009] for dwarfs above M, and Benedict et al. [2016] for dwarfs below M. To convert Benedict et al. [2016]'s reported M_V values to absolute luminosities, we interpolated over E. Mamajek's table[1]. In log-log space, we let the intersection point of the lines float, and make various cuts on the data as indicated in Fig 1. The fit parameters are available in a footnote[2].

The true population of planets around these stars is then as follows: a fraction $\Gamma_{t,s}$ of stars in single star systems have a planet of radius R_p, with orbital period P. A fraction $\Gamma_{t,d}$ of each star in a double star system

[1] `pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt`, downloaded 2017.08.02

[2] m lo: 1.8818719873988132 – c lo: -0.9799647314108376 – m hi: 5.1540712426599882 – c hi: 0.0127626185389781 – M at merge: 0.4972991257826812 – L at merge: 0.0281260412126928.

has a planet of radius R_p, with orbital period P. For instance, if $\Gamma_{t,s} = \Gamma_{t,d} = 0.1$, on average each double system contributes 0.2 planets, and each single system 0.1 planets.

There are alternative choices for the planet population to be assumed in binary star systems. We have taken the limit that says "any planet around the secondary is as likely as around the primary". The opposite extreme says "there are only planets around the primaries, not the secondaries". An intermediate case might be allowing a third occurrence rate, for instance splitting $\Gamma_{t,d} \to (\Gamma_{t,d1}, \Gamma_{t,d2})$ where the latter terms represent the occurrence rate around the primary of a double star system, and the occurrence rate around the secondary of a double star system. Another intermediate would be to prescribe $\Gamma_t(M_\star)$ – in other words to allow the true occurrence rate to vary as a function of stellar mass.

Consider the same set of questions as from the first memo.

1. How many single and double star systems, respectively, are in the sample? Correspondingly, how many stars are in the sample?
2. How many planets are in the sample? (Orbiting single stars, and orbiting double stars respectively).
3. What is the true occurrence rate?
4. How many planets are detected?
5. What occurrence rate does astronomer A, who has never heard of binary star systems, derive for planets of radius R_p and period P?
6. What occurrence rate does astronomer B, who accounts for the "2 for 1" effect of binarity (*i.e.* that the sample actually has more stars than astronomer A thought) derive?
7. What about astronomer C, who accounts for "2 for 1" *and* misclassification due to diluted radii? In other words, astronomer C did a combination of high resolution imaging and RV followup on every candidate, and correctly classifies the planetary radii in every case.

1. HOW MANY STARS ARE IN THE SAMPLE?

Let N_s be the number of single star systems, and N_d the number of double star systems. N_s is the same as in the first memo.

Now

$$N_d(\gamma_R) = n_d \frac{4\pi}{3} d^3_{\rm max,d}(L_1^N, \gamma_R, F^N_{\rm lim}) \tag{2}$$

for

$$d_{\rm max,d}(L_1^N, \gamma_R, F^N_{\rm lim}) = d_{\rm max,s}(L_1^N, F^N_{\rm lim}) \times (1+\gamma_R)^{1/2}. \tag{3}$$

Since q is a random variable, γ_R is a random variable, and $d_{\rm max,d}$ is a random variable. Since N_d is a function of $d_{\rm max,d}$, this means the number of double star systems is a random variable. The mean number of double star systems in the sample is

$$\langle N_d \rangle = \int_0^\infty N_d \, {\rm prob}(N_d)\, {\rm d}N_d. \tag{4}$$

By applying the chain rule for probability density functions, the distribution ${\rm prob}(N_d)$ can be written

$${\rm prob}(N_d) = {\rm prob}(q(\gamma_R)) \left|\frac{{\rm d}q}{{\rm d}\gamma_R}\right| \left|\frac{{\rm d}\gamma_R}{{\rm d}d_{\rm max,d}}\right| \left|\frac{{\rm d}d_{\rm max,d}}{{\rm d}N_d}\right|. \tag{5}$$

Doing some algebra, and assuming $\gamma_R = q^3$, this can be shown to be

$$\begin{aligned}{\rm prob}(N_d) = \frac{2}{81} N_d^{-1/3} \left(\frac{3}{4\pi n_d}\right)^{2/3} d^{-2/3}_{\rm max,s} \times \\ \left[\left(\frac{3N_d}{4\pi n_d}\right)^{2/3} - d^2_{\rm max,s}\right]^{-2/3}\end{aligned} \tag{6}$$

over the interval

$$\text{lower bound} = \frac{4\pi n_d}{3}(\sqrt{0.1^3+1}d_{\rm max,s})^3 \tag{7}$$

$$\text{upper bound} = \frac{4\pi n_d}{3}(\sqrt{2}d_{\rm max,s})^3, \tag{8}$$

and outside the stated interval ${\rm prob}(N_d) = 0$.

While a closed analytic expression for $\langle N_d \rangle$ does exist, it is messy, and it does not yield much intuition. Instead, summarizing the important points and supporting them numerically:

- In a volume-limited sample of binary star systems in which the primary mass is fixed, and the mass ratio is drawn from a bounded uniform distribution, the distribution of γ_R will be biased towards low values ($\gamma_R \approx 0.1$). This is shown in Fig. 2.
- In a magnitude-limited sample of binary star systems in which the primary mass is fixed, and the mass ratio of the *population* is drawn from a bounded uniform distribution, the observed distribution of mass ratios will be biased towards high values. You will see more twins, because they are detectable out to a greater distance. This is shown in Fig. 3.
- In a magnitude-limited sample of binary star systems in which the primary mass is fixed, the distribution of γ_R will be biased towards low values (≈ 0.1), but less so than in a volume-limited sample. This is shown in Fig. 4.

In passing, the bias in ${\rm prob}(\gamma_R)$ towards low luminosity ratios can be seen analytically. If we assume $\gamma_R = q^\alpha$, then

$${\rm prob}(\gamma_R) = \frac{1}{9\alpha}\, \gamma_R^{\frac{1-\alpha}{\alpha}} \quad \text{for } (0.1)^\alpha < \gamma_R < 1, \tag{9}$$

and otherwise zero. For instance if $\alpha = 4$, ${\rm prob}(\gamma_R) \propto \gamma_R^{-3/4}$ and the domain extended from 1 to 10^{-4}, where the probability distribution peaks.

How many stars are in the sample?— We return to the original question: how many stars?

$$\langle N_{\rm stars} \rangle = N_s + 2\langle N_d \rangle, \tag{10}$$

as before. Since we are not giving analytic expressions for either of the $\langle \ldots \rangle$ terms, we just note that in the

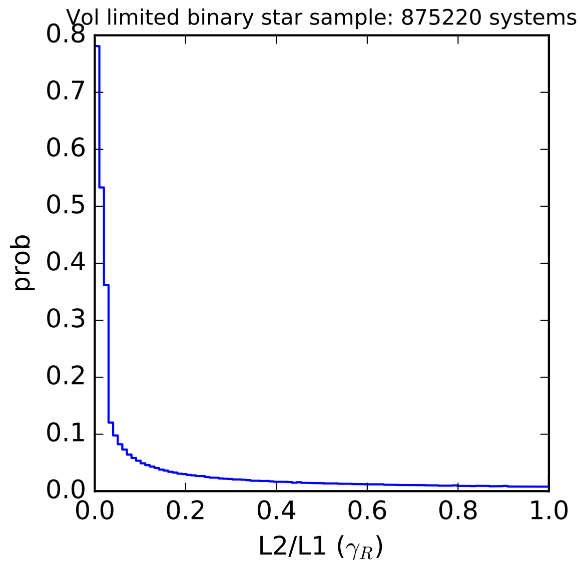

FIG. 2.— The distribution of the luminosity ratio for a volume limited sample of binary stars. BF = 0.45 [Duchene and Kraus 2013]; total number density from Bovy 2017; $M(L)$ relation from Fig. 1; mass ratios are drawn from Eq. 1.

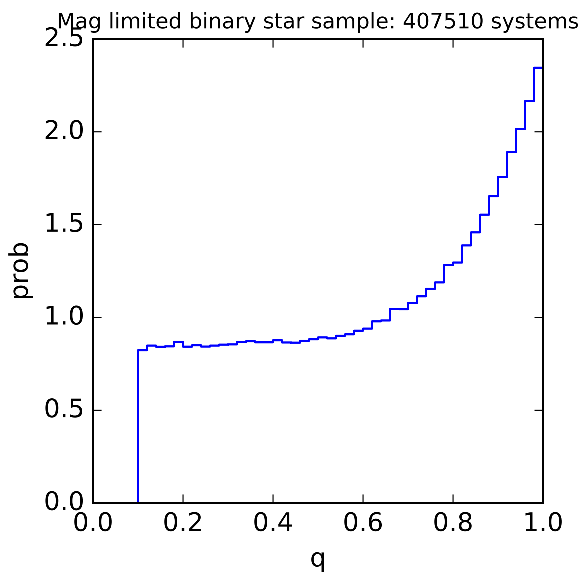

FIG. 3.— The distribution of the mass ratio for a magnitude limited sample of binary stars. BF = 0.45 [Duchene and Kraus 2013]; total number density from Bovy 2017; $M(L)$ relation from Fig. 1; mass ratios are drawn from Eq. 1 – a *uniform* distribution in a volume-limited sample!

case of a population with a single γ_R value, $N_d/N_s = \mathrm{BF}\times(1+\gamma_R)^{3/2}$. A lower bound for this ratio will always be the binary fraction. For a single run of a Monte Carlo code, where mean$(\gamma_R) = 0.30$, median$(\gamma_R) = 0.16$, and the distribution was that given in Fig. 4, the observed ratio was $N_d/N_s = 0.59$. This corresponds to a single-valued population with $\gamma_R \sim 0.23$, a number between the mean and median of the true distribution. It means ~ 1.2 stars in binary star systems for every star in a single star system from this sample.

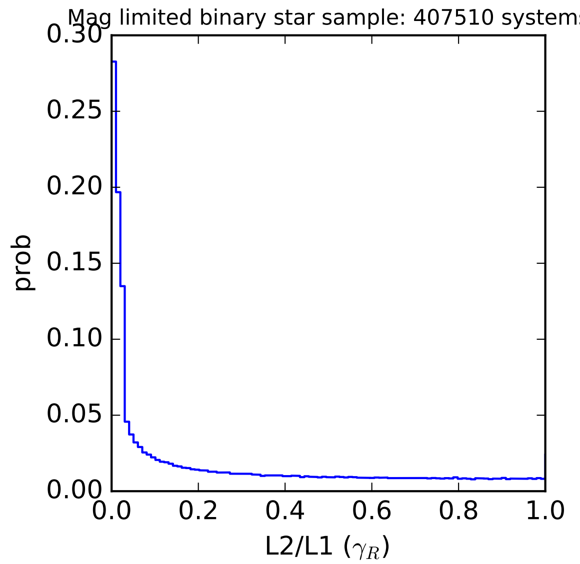

FIG. 4.— The distribution of the luminosity ratio for a magnitude limited sample of binary stars. BF = 0.45 [Duchene and Kraus 2013]; total number density from Bovy 2017; $M(L)$ relation from Fig. 1; mass ratios are drawn from Eq. 1.

We also note in passing that $N_d/N_s = 0.59$ is a smaller ratio than for the case of a fixed γ_R population with $\gamma_R = 1$, which gave $N_d/N_s = 1.27$. The difference is that in the latter population, the volume of searchable stars is greater. Based on the fixed-γ_R population scaling law $N_d/N_s \propto (1+\gamma_R)^{3/2}$ (derived in the first memo), we would expect the ratio of single stars to be $\approx (2/1.1)^{3/2} = 2.45$ times smaller, which is roughly (though not exactly) what we observe.

2. HOW MANY PLANETS ARE IN THE SAMPLE?

The mean number of planets in the sample is

$$\langle N_{\mathrm{planets}}\rangle = N_{\text{planets in single star systems}} + N_{\text{planets in double star systems}} \tag{11}$$

$$= \Gamma_{t,s}N_s + 2\Gamma_{t,d}\langle N_d\rangle. \tag{12}$$

The factor of 2 accounts for the fact that there are twice as many stars in double star systems. We are assuming planets can be around either component.

3. WHAT IS THE TRUE OCCURRENCE RATE?

The "true occurrence rate" is the average number of planets per star. Thus

$$\Gamma_t = \frac{\langle N_{\mathrm{planets}}\rangle}{\langle N_{\mathrm{stars}}\rangle} \tag{13}$$

$$\Gamma_t = \frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}\langle N_d\rangle}{N_s + 2\langle N_d\rangle}. \tag{14}$$

4. HOW MANY PLANETS ARE DETECTED?

The total number of planet detections is the sum of the number of planets detected in single star systems $N_{\mathrm{det},s}$ and the number of planets detected in double star systems $N_{\mathrm{det},d}$. The latter of these is a random variable. Since we selected stars for which there was enough light to make detections, there is no need to compute the S/N distribution of "threshold-crossing events".

The number of planets detected in single star systems is

$$N_{\mathrm{det},s} = N_s \Gamma_{t,s} f_{s,\mathrm{geom}}, \tag{15}$$

where the product $N_s\Gamma_{t,s}$ is the number of planets in the single star systems of the sample, and $f_{s,\mathrm{geom}}$ is the geometric transit probability. The mean number of planets detected in double star systems is

$$\langle N_{\mathrm{det},d}\rangle = 2\langle N_d\rangle \Gamma_{t,d} f_{d,\mathrm{geom}}, \tag{16}$$

where now $2\langle N_d\rangle\Gamma_{t,d}$ is the mean number of planets in the double star systems of the sample. The mean number of detected planets $\langle N_{\mathrm{det}}\rangle$ is the sum of the two equations above.

5. ASTRONOMER A IGNORES BINARITY

Astronomer A has never heard of binary star systems. What occurrence rate does he derive for planets of radius R_p and period P?

The *total* occurrence rate (number of planets divided number of "stars") for Astronomer A would be $(N_{\mathrm{det}}/f_{s,\mathrm{geom}})/(N_s + N_d)$. However, even though Astronomer A does not know about binaries, the radii he derives for any planets in binary systems are too small, by a factor $\sqrt{\mathcal{D}}$. The question asks what occurrence rate is derived for planets of radius R_p and period P. The answer is

$$\Gamma_{\mathrm{A,\ planets\ of\ R_p}} = \frac{N_{\mathrm{det,s}}/f_{s,\mathrm{geom}}}{N_s + N_d} \tag{17}$$

$$= \frac{\Gamma_{t,s} N_s}{N_s + N_d}. \tag{18}$$

This astronomer will also think there is a second population of planets, with radius $R_p\sqrt{\mathcal{D}}$ (a constant number for $\gamma_R =$ constant). He will then also claim have derived a second occurrence rate,

$$\Gamma_{\mathrm{A,\ planets\ of\ R_p}\sqrt{\mathcal{D}}} = \frac{N_{\mathrm{det,d}}/f_{d,\mathrm{geom}}}{N_s + N_d}, \tag{19}$$

where at least for the twin binary case the geometric completeness term is the same as for Eq. 17.

6. ASTRONOMER B COUNTS HOST STARS CORRECTLY

Astronomer B can somehow account correctly for the "2 for 1" effect of binarity, *i.e.* that the sample actually has more stars than astronomer A thought.

By the same token as above,

$$\Gamma_{\mathrm{B,\ planets\ of\ R_p}} = \frac{N_{\mathrm{det,s}}/f_{s,\mathrm{geom}}}{N_s + 2N_d}, \tag{20}$$

and

$$\Gamma_{\mathrm{B,\ planets\ of\ R_p}\sqrt{\mathcal{D}}} = \frac{N_{\mathrm{det,d}}/f_{d,\mathrm{geom}}}{N_s + 2N_d}. \tag{21}$$

7. ASTRONOMER C COUNTS HOST STARS CORRECTLY AND FIGURES OUT DILUTED RADII

Astronomer C knows which planets are in binary systems, and which are in single star systems. She knows that the purported population of planets with radii $R_p\sqrt{\mathcal{D}}$ does not exist. All detected planets from this survey have radii R_p. She computes an occurrence rate

$$\Gamma_{\mathrm{C,\ planets\ of\ R_p}} = \frac{N_{\mathrm{det,s}}/f_{s,\mathrm{geom}} + N_{\mathrm{det,d}}/f_{d,\mathrm{geom}}}{N_s + 2N_d} \tag{22}$$

$$= \frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}N_d}{N_s + 2N_d}. \tag{23}$$

With $\approx$ a full semester at Keck, a system-by-system analysis, and perfect understanding of the completeness of the detection efficiency for single and double star systems, Astronomer C has found the true occurrence rate (cf. Eq. 14).

8. REPRESENTATIVE NUMBERS FOR A FEW CASES

We have not given analytic expressions for the number of double star systems N_d because even with simplified $M(L)$ relations, they are unwieldy. However, we can ask similar numerical questions as from the first memo, and compare the answers.

8.1. *If we ignore binarity, for what fraction of detections do we misclassify the radii? How much?*

Ignoring binarity, we will detect $N_{\mathrm{det,s}}$ planets around single stars, and $N_{\mathrm{det,d}}$ planets around double stars. The latter set will be assumed to have radii $R_p\sqrt{\mathcal{D}}$, where $\mathcal{D}$ is the vector of dilutions appropriate for each binary star system. The fraction of detections with misclassified radii can then be written

$$\frac{N_{\mathrm{det,d}}}{N_{\mathrm{det,s}} + N_{\mathrm{det,d}}} = \frac{1}{1+\alpha}, \tag{24}$$

for

$$\alpha \equiv \frac{2N_d\Gamma_{t,d}f_{d,\mathrm{geom}}}{N_s\Gamma_{t,s}f_{s,\mathrm{geom}}} \tag{25}$$

For the nominal G2V dwarf case of BF = 0.45, and the same distributions used to generate Figs. 2 – 4, we found $N_d/N_s = 0.59$. We are assuming $\Gamma_{t,d}/\Gamma_{t,s} = 1$. If we maintained the (incorrect) assumption that $f_{d,\mathrm{geom}}/f_{s,\mathrm{geom}} = 1$, this would give a misclassification rate of 45%.

This latter assumption is wrong because although we are letting the occurrence rate be the same across binary and single systems, our binary systems have secondaries with varying masses. Thus they have varying radii. Taking [Demircan & Kahraman 1991]'s empirical fit to eclipsing binary data,

$$R = 1.06M^{0.945} \tag{26}$$

for all the stars in our desired mass range ($M < 1.66M_\odot$, D&K 1991's stated bounds). For a zero eccentricity system, $f_{\mathrm{geom}} = R_\star/a$. Since $f_{d,\mathrm{geom}} = 0.5 \times (f_{d1,\mathrm{geom}} + f_{d2,\mathrm{geom}})$, *i.e.* it is the average of the transit probabilities for both the primary and secondary (d1 and d2), we can write (dropping the "geom" subscript)

$$\frac{f_d}{f_s} = \frac{1}{2}\left(1 + \left(\frac{M_1}{M_{d2}}\right)^{1/3}\frac{R_{d2}}{R_1}\right) \tag{27}$$

$$= \frac{1}{2}\left(1 + q^{-1/3+0.945}\right). \tag{28}$$

Computing it all out in a numerical simulation, the only thing that changes analytically is that

$$N_d f_d \to \sum_{i=1}^{N_d} f_{d,i}, \tag{29}$$

so when considering the ratio f_d/f_s, we are in fact interested in

$$\frac{\langle f_d \rangle}{f_s} = \frac{\frac{1}{N_d}\sum_{i=1}^{N_d} f_{d,i}}{f_s}, \tag{30}$$

where $f_{d,i}$ is the vector of geometric transit probabilities across all systems. For the same numerical realization with $N_d/N_s = 0.59$, and using the mass radius relation given by Eq. 26, I get $\langle f_d \rangle / f_s = 0.84$ (smaller secondaries lead to a smaller average transit probability in double star systems), for a radius misclassification rate of 0.51.

8.2. *Twin binaries: if we ignore binarity, how wrong is our occurrence rate for planets of radius R_p?*

For brevity, write $\Gamma_{\mathrm{A,\ planets\ of\ R_p}} = \Gamma_{\mathrm{A,R_p}}$, and similarly for C. Then the relative difference between the two occurrence rates is

$$\left|\frac{\Gamma_{\mathrm{A,R_p}} - \Gamma_{\mathrm{C,R_p}}}{\Gamma_{\mathrm{A,R_p}}}\right| = \left|1 - \left(\frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}N_d}{\Gamma_{t,s}N_s} \cdot \frac{N_s + N_d}{N_s + 2N_d}\right)\right| \tag{31}$$

$$= \left|1 - \left(\left[1 + \frac{2\beta}{\chi}\right] \cdot \frac{1+\beta}{1+2\beta}\right)\right| \tag{32}$$

for

$$\beta \equiv N_d/N_s, \quad \chi \equiv \Gamma_{t,s}/\Gamma_{t,d}. \tag{33}$$

Again, our nominal case has $\chi = 1$, so this simplifies to

$$\left|\frac{\Gamma_{\mathrm{A,R_p}} - \Gamma_{\mathrm{C,R_p}}}{\Gamma_{\mathrm{A,R_p}}}\right| = \beta. \tag{34}$$

For our nominal case, this means a relative error of 59%. As discussed in Sec. 1, this is less than our earlier (twin binary model) case because of the smaller volume of searchable binary stars.