IMPROVING ESTIMATES OF m sin i BY EXPANDING RV DATA SETS

ROBERT A. BROWN

1 Space Telescope Science Institute 3700 San Martin Drive Baltimore, MD 21218, USA; hbrown@stsci.edu

Received 2016 January 5; revised 2016 May 18; accepted 2016 May 19; published 2016 July 11

ABSTRACT

We develop new techniques for estimating the fractional uncertainty (F) in the projected planetary mass (m sin i) resulting from Keplerian fits to radial-velocity (RV) data sets of known Jupiter-class exoplanets. The techniques include (1) estimating the distribution of m sin i using projection, (2) detecting and mitigating chimeras, a source of systematic error, and (3) estimating the reduction in the uncertainty in m sin i if hypothetical observations were made in the future. We demonstrate the techniques on a representative set of RV exoplanets, known as the Sample of 27, which are candidates for detection and characterization by a future astrometric direct imaging mission. We estimate the improvements (reductions) in F due to additional, hypothetical RV measurements obtained in the future. We encounter and address a source of systematic error, “chimeras,” which can appear when multiple types of Keplerian solutions are compatible with a single data set.

Key words: methods: statistical – techniques: radial velocities

1. INTRODUCTION

In the next decade, we can anticipate the development of space telescopes capable of obtaining astrometric direct images (ADIs) and low-resolution spectra of dozens of known radial-velocity (RV) exoplanets. A prime objective is learning their true masses, which are interesting both in themselves, as fundamental astrophysical quantities, and as factors in estimating the atmospheric scale height, which is needed to interpret spectroscopy of the exoplanetary atmosphere. Traub et al. (2016) review the science program of WFIRST, which may be the first ADI mission.

Brown (2015) shows that a single ADI measurement of the apparent separation (s) between planetary companion and host star can yield an estimate of sin i, the sine of the inclination angle (i) of the planetary orbit. Meanwhile, RV measurements (RVMs) yield estimates of the projected mass (m sin i). Therefore, ADIs and RVMs can be combined to estimate m, the true planetary mass.

Our primary goal in this paper is to formulate and address the following question: what are the potential benefits of possible future RVMs, obtained between now and the start of the ADI mission? This question calls for (1) a technique to estimate the fractional uncertainty (F_{RV}^{m\sin i}) in m sin i from fitting an RVM data set (D), and (2) an analysis of how F_{RV}^{m\sin i} improves when hypothetical RVMs—possibly with increased measurement precision—are combined with the real data set (D_0).

The RVM data sets currently available to the public are largely limited to those listed in the discovery papers of RV planets, which in many cases were published many years ago. Therefore, the reader should appreciate that while this paper develops and demonstrates new statistical methods, we can currently demonstrate those methods only with the incomplete RV data currently available to the public.

2. RV DATA SETS AND ORBITAL SOLUTIONS

In our treatment, the elements of an RV orbital-solution vector (p) are:

\[ p = (a, e, \omega_p, t_o, m \sin i, V, \dot{V}) \equiv A[D], \]

\( a \) is the semimajor axis, \( e \) is the eccentricity, \( \omega_p \) is the planetary argument of periapsis, \( t_0 \) is an epoch of periapsis passage, \( V \) is the constant RV of the center of mass, \( V \) is its radial acceleration, and \( A \) is the process of least-squares fitting regarded as a function of \( D \), which is a data set of RVMs in the form:

\[ D = ((t_1, u_1, \delta u_1), (t_2, u_2, \delta u_2), \ldots (t_n, u_n, \delta u_n)), \]

where \( n \) is the number of RVMs in the data set, \( t \) is the epoch of an RVM, and \( u \) and \( \delta u \) are the value and uncertainty of an RVM.

When \( A \) is the usual least-squares fit with weights \( 1/\delta u^2 \), the RV orbital fits are often poor according to the \( \chi^2 \) metric, presumably due to systematic RV “stellar noise,” such as caused by star spots on a rotating star. Therefore, we proceed in this paper using an unweighted least-squares fitting process, \( A' \) (all weights assumed to be equal).

We note that the projected mass, \( m \sin i \), is the only fitted parameter in \( p \) that relates to the planetary mass. Here, \( i \) is the unknown inclination angle of the planetary orbit.

3. NEW TECHNIQUES FOR ESTIMATING F_{RV}^{m\sin i}

To estimate the distribution of the statistic \( F_{RV}^{m\sin i} \) from a data set \( D \), we employ two Monte Carlo (MC) techniques in series: “projection” and “bootstrap.” Brown (2004) describes MC projection for RVM data sets, following the recipe in Section 15.6 of Press et al. (2007).

In the current application, projection calls for drawing a large random sample from the population of \( m \sin i \). This draw is achieved by fitting RV orbits to each of \( N = 10,000 \) “jiggled” data sets \( D_j \), each of which is statistically equivalent to the real data set \( D_0 \):

\[ D_j = A[D_0] = ((t_1, u_{1u}, \delta u_1), (t_2, u_{2u}, \delta u_2), \ldots (t_n, u_{nu}, \delta u_n)), \]

where \( uu \) is a normal random variate of mean \( u \) and standard deviation \( \delta u \), and \( F \) is the jiggling process regarded as a function. We obtain \( p_j \) by an unweighted least-squares fit, \( A' \), to the data set \( D_j \). We then obtain the sampled values \( (m \sin i)_j \)
by extracting the fifth element of each of the $N$ parameter vectors $p_i$, guided by Equation (1). The result is the projection sample \{(m \sin i)_j\}, for $j = 1$ to $N$.

Efron (1979) first described our second MC technique, the bootstrap, which estimates the distribution of a statistic without requiring the support of either a parameterized model or the assumption of a normal distribution. The bootstrap operates directly on a sample, which in this case is \{(m \sin i)_j\}, with cardinality 10,000.

In the current application, the desired statistic is $F_{RV}^m \sin i$, where $\mu_{m \sin i}$ and $\sigma_{m \sin i}$ are the mean and standard deviation of the population $m \sin i$. Thus, we can compute an estimate of $F_{RV}^m \sin i$ from the mean and standard deviation of the sample \{(m \sin i)_j\}. Further, we can create $N$ “resamples” or “bootstrap samples” by randomly selecting $N$ elements of \{(m \sin i)_j\} with replacement. We compute the statistic $F_{RV}^m \sin i$ for each resample, creating a grand sample \{$F_{RV,k}^m \sin i$\}, for $k = 1$ to $N$.

The mean of the $N$ values in \{$F_{RV,k}^m \sin i$\} is our grand estimate of $F_{RV}^m \sin i$ for the population, which we call $F_{RV}^m \sin i$.

4. PREPARING HYPOTHETICAL DATA SETS ($D_{hypo}$)

Here, we prepare a hypothetical data set ($D_{hypo}$), which is the union of the real data set $D_0$ and a synthetic data set $D_{syn}$:

$$D_{hypo} = D_0 \cup D_{syn}. \tag{5}$$

In the current treatment, we assume that $D_{syn}$ comprises $n'$ new RVMs spread evenly over the timespan of one exoplanetary year immediately prior to the assumed start of the ADI mission on 2025 January 1 (i.e., $t_{ADI} = $ Julian day 24660676.5). Given the planetary period $P$, the epochs of the synthetic RVMs are therefore

$$t'_i = t_{ADI} - P \left(1 - \frac{i-1}{n'}\right), \tag{6}$$

for $i = 1$ to $n'$. The synthetic data set is

$$D_{syn} = ((t'_1, u'_1, \delta'_1), (t'_2, u'_2, \delta'_2), \ldots (t'_{n'}, u'_{n'}, \delta'_{n'})),$$  

which $\delta'_{n'} = q \delta_{rms}$, where $\delta_{rms}$ is the root mean square of the uncertainties in the RVMs contained in $D_0$, and where

$$u'_{n'} = V + V(t'_1 - t_0) + \sqrt{G \left[\cos(\nu(t'_1) + \omega) + e \cos(\omega)\right]/\sqrt{1-e^2(m+m_\star)}} \sin i, \tag{8}$$

where $a$ is the semimajor axis, $e$ is the eccentricity, $\omega_0$ is the argument of periapsis of the star, $m_\star$ is the stellar mass, $G$ is the gravitational constant, and the true anomaly $\nu(t)$ is the root of the equation

$$\tan \frac{\nu}{2} = \sqrt{1+e} \tan \frac{E}{2}, \tag{9}$$

in which the eccentric anomaly $E$ is the root of Kepler’s Equation,

$$E - e \sin E = M, \tag{10}$$

and the mean anomaly $M$ is

$$M = 2\pi \left(\frac{t'_0 - t_0}{P}\right). \tag{11}$$

Table 1

| Case | $n'$ | $q$ |
|------|------|-----|
| A    | 0    | ... |
| B    | 10   | 1.0 |
| C    | 10   | 0.1 |
| D    | 100  | 1.0 |
| E    | 100  | 0.1 |

Note. For each planet, the standard deviations $\delta$ of the synthetic RVMs are equal to $q$ times the root-mean-square errors in the real $D_0$.

In Equations (6)–(11) we adopt the orbital elements in $p_0$, which is the RV solution determined by fitting the real data set $D_0$, using an unweighted least-squares fit, $A'$. As a result, the addition of hypothetical RVMs does not substantively change the orbital solution, but it does increase in accuracy by adding supportive information.

In this paper, as a demonstration, we treat five particular cases of hypothetical data sets, as listed in Table 1. Note that Case A assumes no additional, synthetic RVMs, and uses only the RVMs in the real data set $D_0$.

5. RESULTS FOR $F_{RV}^m \sin i$

Here, we present our results for $F_{RV}^m \sin i$ for Cases A–E.

Our input catalog is the “Sample of 27,” which includes all known RV exoplanets that are (1) single-planet systems, (2) hosted by single stars, (3) possibly observable by the ADI mission with an assumed inner working angle $IWA = 0''1$, and (4) offer a publicly available RV data set. (See Table 5 and Appendices A and B.)

We prepare columns 6–10 in Table 2 using the recipes for MC projection and bootstrap described in Section 3. Each entry $F_{RV}^m \sin i$ is the mean of 10,000 realizations of $F_{RV}^m \sin i$ computed from jiggled data sets, each one fitted for a value of $m \sin i$, followed by 10,000 bootstrap resamples of $m \sin i$, from each of which we compute one estimate of the mean and one estimate of the standard deviation of $m \sin i$. From each such pair of values we calculate one value of $F_{RV}^m \sin i$ by Equation (4). Those steps produce a sample of 10,000 values of $F_{RV}^m \sin i$. Our grand estimate for Table 2 is $F_{RV}^m \sin i$, which is the mean of this sample of $F_{RV}^m \sin i$.

As explained in Section 6, the results in red typeface in Table 2 are affected by chimeras. The 16 affected exoplanets are excluded from the final analysis.

The three columns headed by a reference to “$m \sin i$” are a reality check on our Keplerian fits of jiggled data sets. The third column is the value of $m \sin i$ listed in the www.exoplanets.org catalog on 2014 March 1. The fourth column is the mean of $m \sin i$ from the 10,000 fits to jiggled data sets. These are purely Case A results, involving no synthetic RVMs. The fifth column is the fractional deviation between the third and fourth columns. In other words, it is a comparison between our Keplerian fits and those published by the RV practitioners.

In the fifth column, consider at first only the 11 values of fractional deviation in black typeface. The average absolute fractional deviation is 1.5%. For these 11 exoplanets, we are therefore confident that we are performing the Keplerian fits correctly as stated. The small differences with the catalog
results for $m \sin i$ could be explained by any of several factors, including the questions of weighted versus unweighted least-squares fits, the inclusion or not of a customary “stellar jitter” term to lower the chi-square metric (not used here), and the particular suite of Keplerian parameters chosen to fit.

By contrast, most of the 16 exoplanets with red typeface in the fifth column of Table 2 show elevated values of fractional deviation. Those are the exoplanets with chimeras in Case A, as we discuss in Section 6.

Many of the values of $\frac{\Delta (m \sin i)}{m \sin i}$ in columns 6–10 in Table 2 are anomalous—such as those that are greater than one. Other values seem reasonable, in the range $1\%–10\%$. Furthermore, for those planets showing a sequence of reasonable values, the sequence shows monotonic improvement (decline) with the progression of Cases A–E. This is as expected; from the discussion of Table 1 in Section 7, we expect that the information content of the synthetic data sets increases monotonically in ascending alphabetic order, Case A to Case E.

Our main goal is to estimate the improvements in $\mathcal{F}_{RV}^{m \sin i}$ when hypothetical RVMs are added to the real data set. We measure this improvement by the metric $f$, which is the fractional change in $\mathcal{F}_{RV}^{m \sin i}$ for Cases B–E as compared with Case A:

$$f(B-E) = \frac{\mathcal{F}_{RV}^{m \sin i}(B-E)}{\mathcal{F}_{RV}^{m \sin i}(A)}.$$  

Equation (12) applies separately to any exoplanet in the Sample of 27. As an example, for Case D and HD 89307 b, $f(D, HD \ 89307 \ b) = 0.028/0.066 = 0.42$. For that particular exoplanet, adding 100 synthetic RVM of the same root-mean-square accuracy as in the real data set reduces the uncertainty in $m \sin i$ by 42%.

6. CHIMERAS

RV chimeras are a phenomenon previously described—but not given a name—by Brown (2004), who studied the chimera in HD 72659 b. Chimeras arise in the projection sample $m \sin j \{ (m \sin i) \}$ when the data set supports two or more types of Keplerian solution. (We borrow the term “chimera” from genetics, where it refers to an organism composed of two or more genetically distinct tissues, such as an organism that is partly male and partly female.) We find that usually one solution type is dominant and centered near the primary solution. The other solution type—the chimera—may exploit a lack of constraint in the data set, sometimes producing a multi-modal distribution of RV parameters. As an example, Figures 1–2 show the chimera of epsilon Eri, discovered in the current study.

The skewness metric is useful for detecting outliers and asymmetric distributions, such as those caused by chimeras. Our application of the metric uses the absolute value of

| Sample of 27 | Catalog $m \sin i$ | Case A $m \sin i$ | $\Delta (m \sin i)$ | Case A $m \sin i$ | Case B | Case C | Case D | Case E |
|-------------|------------------|------------------|------------------|------------------|--------|--------|--------|--------|
| 1 | GJ 649 b | 0.325 | 0.360 | 0.106 | 0.064 | 0.042 | 0.034 | 0.026 | 0.012 |
| 2 | HD 147513 b | 1.180 | 1.112 | -0.057 | 0.100 | 0.044 | 0.038 | 0.022 | 0.011 |
| 3 | HD 72659 b | 3.174 | 3.684 | -0.161 | 0.543 | 0.960 | 0.671 | 0.019 | 0.010 |
| 4 | HD 70642 b | 1.909 | 1.961 | 0.027 | 0.036 | 0.027 | 0.023 | 0.014 | 0.007 |
| 5 | 7 Cma b | 2.432 | 4.539 | 0.866 | 0.579 | 1.168 | 1.380 | 0.666 | 0.134 |
| 6 | HD 50554 b | 4.399 | 5.078 | 0.154 | 0.091 | 0.034 | 0.018 | 0.014 | 0.004 |
| 7 | GJ 832 b | 0.644 | 4.258 | 5.068 | 1.111 | 0.312 | 0.739 | 1.096 | 1.149 |
| 8 | 14 Her b | 5.215 | 5.151 | -0.012 | 0.006 | 0.005 | 0.005 | 0.003 | 0.002 |
| 9 | GJ 849 b | 0.831 | 0.821 | -0.012 | 0.489 | 2.767 | 2.524 | 1.869 | 1.046 |
| 10 | HD 79498 b | 1.346 | 1.332 | -0.010 | 0.042 | 0.035 | 0.034 | 0.023 | 0.016 |
| 11 | 16 Cyg B b | 1.640 | 1.665 | 0.015 | 0.030 | 0.028 | 0.026 | 0.018 | 0.012 |
| 12 | HD 39091 b | 10.088 | 10.086 | -0.000 | 0.006 | 0.006 | 0.005 | 0.004 | 0.002 |
| 13 | HD 216437 b | 2.168 | 2.173 | 0.002 | 0.022 | 0.020 | 0.018 | 0.013 | 0.007 |
| 14 | HD 220773 b | 1.450 | 2.473 | 0.706 | 1.696 | 1.358 | 0.979 | 0.506 | 0.128 |
| 15 | HD 50499 b | 1.745 | 1.749 | 0.003 | 0.043 | 0.036 | 0.030 | 0.020 | 0.010 |
| 16 | GJ 676 A b | 4.897 | 4.810 | -0.018 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 |
| 17 | HD 89307 b | 1.791 | 1.770 | -0.012 | 0.066 | 0.046 | 0.042 | 0.028 | 0.018 |
| 18 | epsilon Eri b | 1.054 | 0.935 | -0.099 | 0.100 | 0.058 | 0.062 | 0.036 | 0.027 |
| 19 | HD 153435 b | 0.957 | 0.936 | -0.022 | 0.035 | 0.024 | 0.021 | 0.012 | 0.006 |
| 20 | HD 117207 b | 1.819 | 3.425 | 0.885 | 1.330 | 0.242 | 0.214 | 0.144 | 0.008 |
| 21 | GJ 179 b | 0.824 | 0.815 | -0.012 | 0.009 | 0.051 | 0.046 | 0.030 | 0.014 |
| 22 | HD 10647 b | 0.925 | 0.938 | 0.014 | 0.588 | 0.692 | 0.324 | 0.039 | 0.019 |
| 23 | HD 87883 b | 1.256 | 4.288 | 1.443 | 1.554 | 0.226 | 0.124 | 0.024 | 0.013 |
| 24 | HD 90562 b | 1.333 | 1.400 | 0.065 | 0.113 | 0.642 | 0.190 | 0.493 | 0.020 |
| 25 | GJ 317 b | 1.175 | 1.315 | 0.119 | 0.339 | 0.016 | 0.013 | 0.008 | 0.003 |
| 26 | HD 142022 b | 4.468 | 7.518 | 0.683 | 1.433 | 0.080 | 0.024 | 0.014 | 0.003 |
| 27 | HD 106252 b | 6.959 | 6.641 | -0.046 | 0.017 | 0.013 | 0.012 | 0.009 | 0.006 |
skewness, defined as
\[
\text{skewness} = \frac{\mu_3}{\mu_2^{3/2}}.
\]
where \(\mu_k\) is the \(k\)th central moment of the distribution \(\{m \sin i\}\):
\[
\mu_k = \langle (m \sin i - \langle m \sin i \rangle)^k \rangle.
\]

We find strong correlations between (1) evidence of chimeras in histograms of \(\{m \sin i\}\), such as Figure 1, (2) elevated values of the skewness of \(\{m \sin i\}\), as shown by the results in red typeface in Table 3, and (3) bias in the calculation of \(m_{\text{RV}}\) sin \(i\), as shown by the results in red typeface in Table 2. On the basis of these correlations—particularly that between items (2) and (3)—we include in the final analysis only the subset of results \(m_{\text{RV}}\) sin \(i\) in Table 2 for which the Case-A skewness in Table 3 is less than a chosen threshold value, cutoff = 0.3.

Our estimator of \(f\) in Equation (12) is its mean value, \(\langle f \rangle\), computed on the valid subset of 11 exoplanets. Because the value of the cutoff defines the valid subset, the estimate \(\langle f \rangle\) is a function of the cutoff, too. If the cutoff is lowered, the cardinality of the valid subset is reduced and \(f\) gets noisy. If the cutoff is sufficiently increased, then chimera-corrupted values of \(m_{\text{RV}}\) sin \(i\) are included, and again \(f\) gets noisy. The choice for the cutoff involves a tradeoff, as shown in Figure 3. We choose cutoff = 0.3, which minimizes noise toward both limits. Importantly, this choice minimizes the widths of the colored bands, which are plus and minus the standard deviation of the mean, \(\langle f \rangle\), as a function of the cutoff.

### 7. RESULTS FOR \(f\)

Table 4 gives our final results for \(f\), the factor by which \(F\) is improved (reduced) by hypothetical new RVMs that might be obtained in the future. We compute \(f\) from Equation (12) using the values of \(m_{\text{RV}}\) sin \(i\) in Table 2. Only the 11 chimera-free RV exoplanets contributed to these results.

It should be kept in mind that the improvement factor \(f\) in Equation (12) is the average over that sample, and that \(f\) is not an estimate of the expected improvement for any particular RV planet. Conversely, our techniques are perhaps most relevant to studies of individual RV planets, to support the planning and scheduling RV observations, rather than to surveys designed for statistical results.

It is somewhat surprising that the values of \(\langle f \rangle\) in this study should have such small uncertainties. The 11 RV planets contributing to the final analysis are a heterogeneous sample, both in terms of the RV orbits and the number and accuracy of RVMs in the real data sets.

### Table 3

The Skewness Metric Computed from the Projection Samples \(\{m \sin i\}\)

| Sample of 27 | A       | B       | C       | D       | E       |
|-------------|---------|---------|---------|---------|---------|
| GJ 649 b    | 0.441   | 0.053   | 0.015   | 0.019   | 0.026   |
| HD 147513 b | 0.006   | 0.008   | 0.014   | 0.007   | 0.051   |
| HD 72659 b  | 16.738  | 35.718  | 43.665  | 17.603  | 16.333  |
| HD 70642 b  | 0.013   | 0.011   | 0.015   | 0.032   | 0.008   |
| 7 CMa b     | 2.970   | 2.113   | 2.164   | 17.603  | 25.676  |
| HD 79498 b  | 0.510   | 0.199   | 0.211   | 0.026   | 0.015   |
| 16 Cyg b    | 0.065   | 0.003   | 0.030   | 0.017   | 0.010   |
| HD 39691 b  | 0.013   | 0.003   | 0.058   | 0.039   | 0.007   |
| HD 216437 b | 0.017   | 0.003   | 0.031   | 0.033   | 0.045   |
| HD 220773 b | 5.692   | 7.658   | 12.598  | 9.567   | 25.676  |
| HD 50499 b  | 0.005   | 0.027   | 0.019   | 0.021   | 0.033   |
| GJ 636 A b  | 0.026   | 0.011   | 0.014   | 0.010   | 0.019   |
| HD 89307 b  | 0.112   | 0.051   | 0.002   | 0.040   | 0.014   |
| epsilon Eri b | 66.296 | 49.492  | 0.723   | 0.004   | 0.019   |
| HD 154345 b | 0.111   | 0.038   | 0.023   | 0.049   | 0.012   |
| HD 117207 b | 2.970   | 0.025   | 0.004   | 0.014   | 0.003   |
| GJ 179 b    | 0.139   | 0.084   | 0.025   | 0.009   | 0.036   |
| HD 10647 b  | 41.009  | 32.221  | 70.945  | 0.036   | 0.026   |
| HD 87883 b  | 2.531   | 22.782  | 22.936  | 0.306   | 0.347   |
| HD 30562 b  | 51.674  | 4.285   | 20.115  | 28.715  | 4.201   |
| GJ 317 b    | 50.547  | 0.009   | 0.019   | 0.020   | 0.013   |
| HD 142022 b | 3.004   | 1.512   | 0.642   | 0.001   | 0.003   |
| HD 106252 b | 0.030   | 0.005   | 0.000   | 0.022   | 0.029   |

Note. As in Table 2, red typeface in Case A indicates corruption by chimeras, and the exoplanet is excluded from the final analysis.
Consider a three-dimensional space spanned by \( \log n' \), \( \log q \), and \( f_{\text{calc}} \). We perform a least-squares fit of a plane to the four points in Table 4. Those points are the values in rows 2–4 in each of the columns headed by B–E. The planar solution is

\[
\begin{align*}
\text{f}_{\text{calc}} &= a \log n' + b \log q + c \\
&= 0.381 \log n' + 0.148 \log q + 1.235. \tag{15}
\end{align*}
\]

In Table 4, the fifth row gives the value of \( f_{\text{calc}} \) produced from Equation (15) for Cases B–E, and the last row gives the individual contributions to chi-square. The contributions sum to \( \chi^2 = 6.26 \), with one degree of freedom. The probability of random noise producing this value of \( \chi^2 \) or higher is \( \sim 1\% \). Therefore, the fit of a plane to this data is not first-rate, but it is pretty good considering the unexpectedly small uncertainties in \( f \). In any case, the fit is good enough to estimate the local power-law dependences on \( n' \) and \( q \).

In Equation (15), the coefficients of \( \log n' \) and \( \log q \) are the effective power laws for the dependence of \( f \) on \( n' \) and \( q \)—in this region of parameter space, for these particular RV planets. If the number \( n' \) of new RVMs were overwhelmingly greater than the number of existing, real data points—which \( n' \) is not, in this study—we would expect an \( n' \) power law approaching the value 0.50, instead of 0.38. For \( q \), we would expect a power law of 1.0 for \( n' \) very large, instead of the low value 0.148 from this study.

8. CONCLUDING REMARKS

We develop three new techniques in this paper: (1) estimating the distribution of \( m \sin i \) using Monte Carlo projection, (2) detecting and mitigating chimeras using the metric skewness, and (3) estimating the reduction in the uncertainty in \( m \sin i \) if hypothetical observations are made in the future. We have demonstrated the effectiveness of these techniques on a uniform subsample of 11 exoplanets drawn from the Sample of 27.

For the other 16 exoplanets in the Sample of 27, corruption by chimeras introduces systematic error into estimates of \( m \sin i \)—and no doubt other Keplerian parameters. These errors can be
It is possible that RVMs already obtained have eliminated chimeras in many RV exoplanets. Many more data sets and RVMs exist but are off limits, in private hands, and unavailable for research by the scientific public.

By applying these new techniques to new cases of hypothetical data sets, it is now possible to optimize future RV programs in support of future ADI missions, such as WFIRST. In fact, an RV practitioner can now estimate the improvement in the accuracy of \( m \sin i \) due to a newly obtained RVM or a planned suite of RVMs, in real time, at the telescope.

We thank Stuart Shaklan, Marc Postman, Wes Traub, and Dave Latham for their support and encouragement of this work over many years. We are grateful to Bob Vanderbei for his advice on statistical issues from time to time. Jason Wright has been helpful with the acquisition of publically available RV data sets for the Sample of 27.

APPENDIX A

THE SAMPLE OF 27

Our input catalog, which we call the “Sample of 27,” includes all 27 known RV exoplanets that are (1) single-planet systems, (2) hosted by single stars, (3) possibly observable by the ADI mission with an assumed inner working angle, \( IWA = 0.1 \), and (4) provided with a publically available RV data set.

To develop this catalog, we start with all 436 RV planets in the catalog at the website www.exoplanets.org on 2014 March 1. In the first cut, we eliminate all but 55 exoplanets on the basis of the criterion \( s_{\text{raw}} = (1 + e) a / d > 0.1 \). The term on the left side of this inequality, which we call the “raw separation,” is the maximum possible value of \( s \) for an exoplanet at stellar distance \( d \), with semimajor axis \( a \), eccentricity \( e \), but unknown inclination \( i \). The Sample of 55 is listed in Table 1.

In the second cut, we eliminate all multiple-planet systems, for simplicity in the orbital solutions. This cut creates the Sample of 28.

In the third cut, we eliminate the one planet that never achieves \( s > 0.1 \) despite the fact \( s_{\text{raw}} > 0.1 \). As explained in Appendix B, we use a new completeness metric—global observational completeness (GOC)—to test the entire Sample of 28 for GOC > 0, and we find that only one of them—kappa CrB b—does not satisfy the GOC criterion.

It is important to recognize that the Sample of 27 is not a sample developed for statistical purposes. It is simply a list of RV exoplanets useful for demonstrating our new techniques for estimating \( \mathcal{F}_{\text{RV}} \).
We compute GOC from the orbital solution $p_0$ that results from fitting the “catalog” RV data set ($D_0$). If the inclination $i$ were known and combined with $p_0$, we could compute the apparent separation $s$ at any time. We do not know $i$, of course—in fact, we are trying to estimate it. Nevertheless, we know how $i$ is distributed, namely uniformly in $\cos i$ and with the random variate $\arccos(1 - \mathcal{R})$, where $\mathcal{R}$ is a uniform random variate on the range 0–1. That information is enough to compute GOC by Monte Carlo experiment, as follows. First, we draw a random value of time $t$ from a uniform random variate on the range 0–10$^9$ days—in other words, anytime. Second, we draw a value of $i$ from its random variate. Third, we combine each value of $i$ with $p_0$ to compute $s$ and $\Delta\text{mag}$ from Equations (8), (20), (21) in Brown (2015). Fourth, we apply the detectability criteria $s > IWA$ and $\Delta\text{mag} < \Delta\text{mag}_0$, and get a 1 if both criteria are satisfied and get a 0 otherwise. Fifth, we sum the values for the one million trials, and divide by one million.

Table 1 shows the values of GOC for all the planets in the Sample of 28.

REFERENCES

Brown, R. A. 2004, ApJ, 610, 1079
Brown, R. A. 2015, ApJ, 805, 188
Efron, B. 1979, AnSta, 7, 1
Plavchan, P., Latham, D., Gaudi, B. S., et al. 2015, Radial Velocity Prospects Current and Future: A White Paper Report Prepared by the Study Analysis Group 8 for the Exoplanet Program Analysis Group (ExoPAG), http://adsabs.harvard.edu/abs/2015arXiv150301770P
Press, W. B., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 2007, Numerical Recipes (3rd ed.; New York: Cambridge Univ. Press)
Traub, W. A., Breckinridge, J., Greene, T. P., et al. 2016, JATIS, 2, 1