The Astronomical Journal, 129:1706–1717, 2005 March

QUANTIFYING THE UNCERTAINTY IN THE ORBITS OF EXTRASOLAR PLANETS

Eric B. Ford[1]
Department of Astrophysical Sciences, Peyton Hall, Princeton University, Ivy Lane, Princeton, NJ 08544-1001
Received 2004 May 27; accepted 2004 December 8

ABSTRACT

Precise radial velocity measurements have led to the discovery of ~100 extrasolar planetary systems. We investigate the uncertainty in the orbital solutions that have been fitted to these observations. Understanding these uncertainties will become more and more important as the discovery space for extrasolar planets shifts to longer and longer periods. While detections of short-period planets can be rapidly refined, planets with long orbital periods will require observations spanning decades to constrain the orbital parameters precisely. Already in some cases, multiple distinct orbital solutions provide similarly good fits, particularly in multiple-planet systems. We present a method for quantifying the uncertainties in orbital fits and addressing specific questions directly from the observational data rather than relying on best-fit orbital solutions. This Markov chain Monte Carlo (MCMC) technique has the advantage that it is well suited to the high-dimensional parameter spaces necessary for the multiple-planet systems. We apply the MCMC technique to several extrasolar planetary systems, assessing the uncertainties in orbital elements for several systems. Our MCMC simulations demonstrate that for some systems there are strong correlations between orbital parameters and/or significant non-Gaussianities in parameter distributions, even though the measurement errors are nearly Gaussian. Once these effects are considered, the actual uncertainties in orbital elements can be significantly larger or smaller than the published uncertainties. We also present simple applications of our methods, such as predicting the times of possible transits for GJ 876.

Key words: methods: statistical — planetary systems: general — techniques: radial velocities

1. INTRODUCTION

Recent detections of planets around other stars have spurred a wide range of research on planet formation and planetary system evolution. The first planet discovered around a solar-type star, 51 Pegasi b, was in a surprisingly short period orbit (Mayor & Queloz 1995). Other early planets such as 70 Virginis b revealed surprisingly large orbital eccentricities (Marcy & Butler 1996). Multiple-planet systems have revealed intricate dynamical interactions, such as the resonances in υ Andromedae (Butler et al. 1999; Chiang et al. 2001) and GJ 876 (Marcy et al. 2001).

The future of radial velocity planet searches promises to be exciting. Ongoing large surveys including a broad array of nearby main-sequence stars will continue to increase the number of known extrasolar planets. Refinements in the radial velocity technique should continue to improve measurement precision, permitting the detection of planets with smaller masses. The increasing time span of precision observations will permit the discovery of planets with larger orbital periods. Several stars presently known to harbor one planet are expected to reveal additional planets in longer period orbits (Fischer et al. 2001).

These advances will also bring new challenges. The 51 Peg–like planets could be rapidly confirmed by independent observers (Mayor & Queloz 1995; Marcy et al. 1997) because of their short orbital periods and large velocity amplitudes. However, for a given signal-to-noise ratio, the time required to obtain observations to confirm or refute a possible planetary candidate will typically scale with the orbital period of the planet. Already, one known planet (55 Cancri d) has an orbital period of over 14 years (Marcy et al. 2002).

It is also more difficult to obtain precise orbital elements for planets with large orbital periods. While early planet candidates were routinely observed for multiple periods before publication, recently planet candidates have been published when observations span only a single orbital period. Thus, it will become increasingly easy to overinterpret early orbital determinations. A further challenge is that the growing precision and time span of radial velocity measurements are expected to significantly increase the number of stars known to harbor multiple planets. Fitting multiple-planet systems requires many more free parameters, so that observational data may not tightly constrain all the orbital parameters or even distinguish between multiple possible orbital solutions, especially in the presence of significant noise and sparsely sampled data.

These trends imply that it will become increasingly important to understand the uncertainties in orbital elements and other parameters derived from such observations, and thus we must use the best possible statistical tools to analyze radial velocity data.

In this paper, we introduce a Bayesian analysis for constraining the orbital parameters of extrasolar planets with radial velocity observations. The Bayesian framework considers a joint probability distribution function for both the observed data ($\boldsymbol{d}$) and model parameters that cannot be directly observed ($\boldsymbol{x}$). This joint probability, $p(\boldsymbol{d}, \boldsymbol{x})$, can be expressed as the product of the probability of the observables given the model parameters, $p(\boldsymbol{d}|\boldsymbol{x})$, and a prior probability distribution function, $p(\boldsymbol{x})$, which is based on previous knowledge of the model parameters. Bayes's theorem allows one to compute a posterior probability density function, $p(\boldsymbol{x}|\boldsymbol{d})$, which incorporates the knowledge gained by the observations $\boldsymbol{d}$. That is

$$p(\boldsymbol{x}\,|\,\boldsymbol{d}) = \frac{p(\boldsymbol{d},\boldsymbol{x})}{\int p(\boldsymbol{d},\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}} = \frac{p(\boldsymbol{x})p(\boldsymbol{d}\,|\,\boldsymbol{x})}{\int p(\boldsymbol{d},\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}}. \quad (1)$$

Unfortunately, the lower integral can be extremely difficult to compute, particularly when $\boldsymbol{x}$ has a large number of dimensions. Even after this integral is performed, most questions require

[1] Current address: Department of Astronomy, 601 Campbell Hall, University of California at Berkeley, Berkeley, CA 94720-3411; eford@astron.berkeley.edu.

additional integration of $p(\boldsymbol{x}|\boldsymbol{d})$ over many of the model parameters. This paper describes the application of Markov chain Monte Carlo simulation using the Metropolis-Hastings algorithm and the Gibbs sampler to perform the necessary integrations. This technique allows us to accurately characterize the posterior probability distribution function for orbital parameters based on radial velocity observations.

In this paper we first describe our model for calculating radial velocities from planetary orbital parameters. In § 3, we summarize methods for identifying the maximum likelihood orbital solution for a set of observed radial velocities. Then, we discuss methods of characterizing the uncertainties of model parameters in § 4. We focus our attention on the application of Markov chain Monte Carlo simulation to estimate the uncertainties for orbital parameters in a Bayesian framework. In § 5, we apply this technique to analyze several published extrasolar-planet data sets. Finally, we summarize the potential of this technique as additional long-period planets and multiple-planet systems are discovered.

2. RADIAL VELOCITY MODEL

In radial velocity surveys, the velocity of the central star is precisely monitored for periodic variations that could be caused by orbiting companions. Each individual observation can be reduced to a measurement of the star's radial velocity and an estimate of the observational uncertainty based on photon statistics. Because the observations are averaged over hundreds of sections of the spectrum, the observational uncertainties of most current echelle-based radial velocity surveys are nearly Gaussian (Butler et al. 1998). Stellar activity can also contribute to the observed radial velocities. However, most of the stars investigated in this paper are chromospherically quiet, and so the contribution of stellar "jitter" to the observed radial velocities is believed to be negligible.

2.1. *Single-Planet Systems*

Such radial velocity observations are able to detect the acceleration of a star due to the gravitational perturbations from an orbiting planet. We model the motion of the planet as a Keplerian orbit using the following parameters: orbital period (P), velocity semiamplitude (K), eccentricity (e), argument of periastron (ω), and mean anomaly at the specified epoch (M_0). The perturbation to the radial velocity (Δv_*) of a star due to a planet on a Keplerian orbit is given by

$$\Delta v_*(t) = K\{\cos[\omega + T(t)] + e\cos\omega\}. \tag{2}$$

A planet's true anomaly [$T(t)$] is a function of time (t) and is related to the planet's eccentric anomaly [$E(t)$] by the relation

$$\arctan\left[\frac{T(t)}{2}\right] = \sqrt{\frac{1+e}{1-e}}\arctan\left[\frac{E(t)}{2}\right]. \tag{3}$$

The eccentric anomaly is related to the mean anomaly [$M(t)$] through Kepler's equation,

$$E(t) - e\sin E(t) = M(t) - M_0 = \frac{2\pi}{P}(t - t_0), \tag{4}$$

where M_0 is a constant, the orbital phase at $t = 0$, which is related to the time of pericenter (t_0).

2.2. *Multiple Planets*

For a star being perturbed by multiple planets, there is no analytic expression for the exact radial velocity perturbation. However, in many cases the radial velocity perturbation can be well modeled as the sum of multiple independent Keplerian orbits. The deviations from the simple sum-of-Keplerians model can be divided into two types: short-period interactions and secular interactions. The magnitude of the short-period interactions (deviations from independent Keplerian orbits on an orbital timescale) is often small compared with the magnitude of the Keplerian perturbation and the observational uncertainties. The secular interactions are typically modeled as changes in the Keplerian orbital parameters. While the secular interactions can have large effects on the observed radial velocities, the timescales are typically much longer than the time span of observations. Thus, we model the observed radial velocity of a star as

$$v_{*,\,\mathrm{model}}(t,j) = \sum_i \Delta v_{*,\,i}(t) + C_j, \tag{5}$$

where $\Delta v_{*,i}(t)$ is given by equation (2) using the orbital parameters of the ith planet and there is an unknown constant velocity offset (C_j), which provides no information about planetary companions. In practice, high-precision radial velocity surveys are typically calibrated such that different observatories have different zero-point offsets, and hence it is important to allow observations from different observatories to have independent constant velocity offsets.

Since the uncertainties from individual observations are expected to closely follow a normal distribution, to evaluate the goodness of fit for a given model we calculate the usual $\chi^2(\boldsymbol{x})$ statistic,

$$\chi^2 = \sum_k [v_{*,\,\mathrm{model}}(t_k, j_k) - v_{*,\,\mathrm{obs}}(t_k, j_k)]^2/\sigma_k^2, \tag{6}$$

where t_k is the time of the kth observation, j_k is the observatory used for the kth observation, and σ_k is the observational uncertainty of the kth radial velocity observation.

3. MAXIMUM LIKELIHOOD ESTIMATORS

Given a set of model parameters ($\boldsymbol{x}$), which are typically the masses and orbital parameters of the planets, we wish to compare the model predictions with the observations using the model and $\chi^2(\boldsymbol{x})$ statistic described in § 2. The maximum likelihood estimate (MLE) of the model parameters, $\boldsymbol{x}_{\mathrm{MLE}}$, is obtained by finding the minimum $\chi^2(\boldsymbol{x}_{\mathrm{MLE}}) = \min_{\boldsymbol{x}}\chi^2$. These are the orbital parameters typically reported as the "best-fit" model. Finding the set of parameters that minimizes $\chi^2(\boldsymbol{x})$ can be challenging, particularly for multiple-planet systems. Often a combination of several methods is used to identify the "best-fit" orbital parameters.

3.1. *Periodograms*

Promising orbital periods can be recognized as sharp dips in a plot of the period versus the minimum χ^2 for that period with the phase and amplitude allowed to vary (a periodogram). There are well-developed methods for determining the significance of periodicities identified in this way and estimating the false-alarm probability (Horne & Baliunas 1986). While a single sinusoid cannot reproduce the signal of an eccentric Keplerian orbit, a periodogram allows the rapid identification of any periodicities in observational data, without requiring a simultaneous fit for amplitude, eccentricity, argument of periastron, or other parameters. While the use of periodograms for an initial exploration of parameter space may not be optimal for identifying all possible Keplerian variations in the radial velocities (e.g., large eccentricities spread power across multiple frequencies), they do

provide an efficient means for identifying potential orbital periods. Accurate estimates of the orbital period are valuable input parameters for subsequent algorithms, particularly the local minimization procedures.

3.2. *Local Minimization*

Once an initial estimate of the orbital parameters is available ($\boldsymbol{x}_0$), an iterative minimization algorithm such as Levenberg-Marquardt (L-M; Press et al. 1992) can be used to refine the model parameters by minimizing χ^2. The L-M algorithm will identify only a single local minimum for a given initial guess of model parameters, $\boldsymbol{x}_0$. Unfortunately, the $\chi^2(\boldsymbol{x})$ surface can have many local minima. In particular, for multiple-planet systems, the many degrees of freedom and the rugged $\chi^2(\boldsymbol{x})$ surface can render L-M minimization particularly sensitive to the initial-guess model parameters and vulnerable to finding a local minimum of $\chi^2(\boldsymbol{x})$ far away from and much less probable than the global minimum. In practice, local minimization algorithms such as L-M are most useful for fine-tuning the parameters of a local minimum identified by global search algorithms.

3.3. *Simulated Annealing*

Simulated annealing generalizes iterative algorithms such as L-M to reduce the risk of becoming trapped in a local minimum (Press et al. 1992). Random perturbations are applied at each iteration. The perturbations are initially large (high temperature) and are gradually reduced (lower temperature). Provided the temperature is reduced sufficiently slowly, simulated annealing can convert a local minimization algorithm into a global minimization algorithm. Unfortunately, the choice of the cooling curve is important. In addition, the number of iterations required may be prohibitively large. Nevertheless, simulated annealing can be a valuable tool when performing global nonlinear minimization.

3.4. *Genetic Algorithms*

A genetic algorithm (GA) is an optimization algorithm loosely based on biological evolution (Charbonneau 1995). GAs are much less likely to become trapped in local minima than are local minimization algorithms such as L-M. The disadvantage is that they require orders of magnitude more evaluations of the goodness-of-fit statistic. When fitting planetary orbits to a single set of radial velocity data, this can be merely an inconvenient delay for analytic models, but GAs can be prohibitively time-consuming if it is necessary to search a large number of data sets. Once a GA has identified a minimum, an L-M–type minimization algorithm can provide an efficient means of fine-tuning the solution. GAs have been applied to the υ And and GJ 876 systems (Stępiński et al. 2000; Laughlin & Chambers 2001). In these cases the GAs have verified the minima found by previous authors. Additional minima have been identified in the case of υ And (Stępiński et al. 2000).

4. ALGORITHMS FOR ESTIMATING UNCERTAINTIES

The previous discussion has focused on identifying the maximum likelihood or best-fit parameter values. It is also important to characterize the uncertainty in the estimation of the parameter values. Here we discuss three methods of estimating these uncertainties: constant-$\Delta\chi^2$ boundaries, resampling, and Markov chain Monte Carlo methods.

4.1. *Constant-$\Delta\chi^2$ Boundaries*

One method of estimating the uncertainty in orbital parameters fitted to the observed data is to evaluate $\chi^2(\boldsymbol{x})$ at points on a grid in the parameter space near $\boldsymbol{x}_{\rm MLE}$. Then confidence intervals can be calculated by finding the boundary (B) along which χ^2 is constant. The probability that the parameters lie inside the boundary can be calculated as

$$\frac{\int_B dx\, e^{-\Delta\chi^2(\boldsymbol{x})}}{\int dx\, e^{-\Delta\chi^2(\boldsymbol{x})}}, \tag{7}$$

where $\Delta\chi^2(\boldsymbol{x}) = \chi^2(\boldsymbol{x}) - \chi^2(\boldsymbol{x}_{\rm MLE})$ and the bottom integral extends over the entire parameter space. This approach is equivalent to assuming a uniform prior in $\boldsymbol{x}$ in a Bayesian framework. Unfortunately, evaluating these integrals often requires a prohibitive number of evaluations of $\chi^2(\boldsymbol{x})$, particularly when the number of parameters is large. For example, to fit the radial velocities caused by one, two, or three planets requires at least 6, 11, or 16 model parameters, respectively. If only 10 grid points are used in each dimension, then this technique would require 10^6, 10^{11}, or 10^{16} evaluations of $\chi^2(\boldsymbol{x})$. In practice, the integrands are can be poorly behaved, requiring many more evaluations in each dimension. Thus, this technique is usually not practical for estimating the uncertainties in orbital parameters, especially for multiple-planet systems.

Brown (2004) has applied a similar method to the case of HD 72659. Brown evaluated χ^2 by choosing random parameter values within a region of parameter space using Monte Carlo rather than choosing parameters along a grid. Still, the large range of parameter space made it impractical to explore the entire range of parameter space with a constant density. Thus, Brown initially sampled a wide range of parameter space and then manually identified several boxes of parameter space to sample at higher density. In total, Brown evaluated χ^2 at hundreds of millions of points to sample the allowed parameter space for HD 72659.

4.2. *Refitting to Synthetic Data Sets*

One method of estimating the uncertainty in orbital parameters fitted to the observed data is to apply the fitting technique repeatedly to many sets of simulated data. Since the observational errors are believed to be very nearly Gaussian and each radial velocity measurement has a corresponding uncertainty estimate, it is straightforward to construct simulated data sets by adding Gaussian random values to the actual data points. Each set of simulated data is meant to represent a possible set of measurement values. If the same fitting procedure is applied to the actual data and each simulated data set, then one can obtain the distribution of best-fit parameter values.

One disadvantage of the refitting technique is that one must identify the best-fit orbital parameters for each synthetic data set. In principle, one should apply a global minimization algorithm to each data set, but in practice the computational requirements often dictate that only a local minimization routine will be run for the synthetic data sets. Even when using a local minimization algorithm, the computational requirements can be a burden. For example, for a one-planet system with six free parameters, χ^2 must be calculated 13 times for each iteration of L-M for each synthetic data set. For obtaining confidence intervals roughly equivalent to 3 σ, a sample of at least $\sim 10^5$ synthetic data sets is necessary. Assuming that the local minimization routine requires an average of eight iterations to converge, this amounts to $\sim 10^7$ evaluations of χ^2. Still, the use of only a local minimization algorithm on the synthetic data sets would render this type of analysis vulnerable to underestimating the range of allowed parameters. If a global minimization

algorithm were used, even over a small fraction of the possible parameter space, the number of χ^2 evaluations required would increase by orders of magnitude.

Unfortunately, even then the distribution estimated by means of resampling may not reflect the full range of possible parameter values, particularly in cases where the χ^2 surface is significantly asymmetric around the minimum. To illustrate this possibility, we use the actual observations for the extrasolar planet around HD 72659. In Figure 1, we show the best-fit value of χ^2 when all parameters except orbital period are allowed to vary, as a function of the orbital period (*solid line*). We also show with dotted lines the same curve, but using synthetic data sets generated from the actual observations of HD 72659. For all the data sets, orbital periods shorter than ~1700 days are strongly ruled out and χ^2 increases very slowly for orbital periods greater than the best-fit period. However, the variation in the location of the best-fit period across the synthetic data sets does not reflect the fact that the very slow increase in χ^2 for longer period orbits allows a very large range of orbital periods.

4.3. *Markov Chain Monte Carlo, the Metropolis-Hastings Algorithm, and the Gibbs Sampler*

Bayesian inference using Markov chain Monte Carlo (MCMC) simulations provides an alternative method for estimating the uncertainty of fitted parameters. The MCMC method has been applied to several other astronomical data sets and problems, including analysis of UV and X-ray spectra (Kashyap & Drake 1998; van Dyk et al. 2001), star formation history (Cid Fernandes et al. 2001; Panter et al. 2003), object detection (Hobson & McLachlan 2003), the Cepheid distance scale (Barnes et al. 2003), cluster weak lensing and the Sunyaev-Zeldovich effect (Marshall et al. 2003), Type Ia supernovae (Wang & Mukherjee 2004), and especially the cosmic microwave background (e.g., Knox et al. 2001; Verde et al. 2003).

The goal of the MCMC method is to generate a chain (i.e., sequence) of states (i.e., sets of parameter values $\boldsymbol{x}_i$) that are sampled from a desired probability distribution [$f(\boldsymbol{x})$]. Such a chain can be calculated by specifying an initial set of parameter values, $\boldsymbol{x}_0$, and a transition probability, $p(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n)$. The Monte Carlo aspect of MCMC simulation refers to randomness in the generation of each subsequent state. The Markov property specifies that the probability distribution for determination of $\boldsymbol{x}_{n+1}$ can depend on $\boldsymbol{x}_n$ but not previous states. If the Markov chain is reversible, that is, if

$$f(\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{x}') = f(\boldsymbol{x}')p(\boldsymbol{x}'|\boldsymbol{x}), \tag{8}$$

and it is aperiodic and irreducible, then it can be proved that the Markov chain will eventually converge to the stationary distribution $f(\boldsymbol{x})$ (Gilks et al. 1995). The requirement that the chain be irreducible guarantees that it will be possible for the chain to reach every state with nonzero probability from any initial state.

MCMC offers a relatively efficient method of performing the integrations necessary for a Bayesian analysis (eq. [1]), if we can calculate a Markov chain whose equilibrium distribution is equal to the joint posterior probability density function for the model parameters given the observed data.

For application to radial velocity measurements and orbit determination, the observational errors are believed to be very nearly Gaussian with accurately estimated variances. Thus, if the data are generated by the model specified by $\boldsymbol{x}$, then the probability of drawing the observed values, $p(\boldsymbol{d}|\boldsymbol{x})$, is roughly proportional to $\exp[-\chi^2(\boldsymbol{x})/2]$. If we choose a uniform prior in $\boldsymbol{x}$ [$p(\boldsymbol{x}) \sim 1$], then the posterior distribution, $p(\boldsymbol{x}|\boldsymbol{d})$, is also

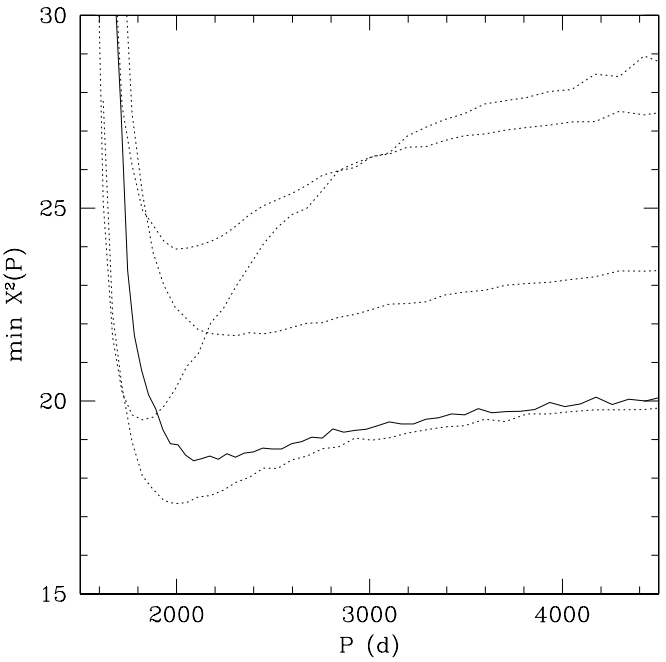

FIG. 1.—Minimum value of χ^2, varying all parameters except the orbital period, vs. orbital period for HD 72659. The solid line is for the actual observations and the dotted lines are based on resampled data. Note that the very slow rise in $\chi^2(P)$ for larger orbital periods allows a much greater range in the orbital period than suggested by the variation in the minimum of the $\chi^2(P)$ curves for resampled data.

roughly proportional to $\exp[-\chi^2(\boldsymbol{x})/2]$. This will allow us to construct Markov chains with an equilibrium distribution equal to the posterior distribution,

$$f(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{d}) = p(\boldsymbol{x})p(\boldsymbol{d}|\boldsymbol{x}) \sim \exp[-\chi^2(\boldsymbol{x})/2]. \tag{9}$$

The Metropolis-Hastings (M-H) algorithm involves the generation of a trial state ($\boldsymbol{x}'$) according to a candidate transition probability function, $q(\boldsymbol{x}'|\boldsymbol{x}_n)$, and randomly accepting the trial as the next state or rejecting the trial state in favor of the current state. The M-H algorithm specifies an acceptance probability, $\alpha(\boldsymbol{x}'|\boldsymbol{x})$, such that the transition probability

$$p(\boldsymbol{x}'|\boldsymbol{x}) = q(\boldsymbol{x}'|\boldsymbol{x})\,\alpha\,(\boldsymbol{x}'|\boldsymbol{x}) \tag{10}$$

is guaranteed to be reversible and irreducible, provided only that $q(\boldsymbol{x}'|\boldsymbol{x})$ allows transitions to all $\boldsymbol{x}$ for which $f(\boldsymbol{x})$ is nonzero. The M-H algorithm acceptance probability is

$$\alpha(\boldsymbol{x}'|\boldsymbol{x}) = \min\left[\frac{f(\boldsymbol{x}')q(\boldsymbol{x}|\boldsymbol{x}')}{f(\boldsymbol{x})q(\boldsymbol{x}'|\boldsymbol{x})}, 1\right] \tag{11}$$

if $f(\boldsymbol{x})q(\boldsymbol{x}'|\boldsymbol{x}) > 0$, and $\alpha(\boldsymbol{x}'|\boldsymbol{x}_n) = 1$ otherwise. Note that the M-H algorithm does not require that the normalization of $f(\boldsymbol{x})$ be known a priori. While the M-H algorithm guarantees that the chain will converge to $f(\boldsymbol{x})$, it does not specify when the chain will achieve convergence.

In principle, the basic M-H algorithm can be implemented as follows:

1. Initialize the chain with some $\boldsymbol{x}_0$ and $n = 0$.
2. Generate a trial state, $\boldsymbol{x}'$, according to $q(\boldsymbol{x}'|\boldsymbol{x}_n)$.
3. Calculate $\chi^2(\boldsymbol{x}')$ for the trial state and $\chi^2(\boldsymbol{x}_n)$ for the current state.
4. Determine the ratio $f(\boldsymbol{x}')/f(\boldsymbol{x}_n) \sim \exp\{-\frac{1}{2}[\chi^2(\boldsymbol{x}') - \chi^2(\boldsymbol{x}_n)]\}$.

5. Draw a random number, u, from a uniform distribution between 0 and 1.
6. If $u \leq \alpha(\boldsymbol{x}'|\boldsymbol{x}_n)$, as defined by equation (11), then set $\boldsymbol{x}_{n+1} = \boldsymbol{x}'$. If $u > \alpha(\boldsymbol{x}'|\boldsymbol{x}_n)$, then set $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n$.
7. Set $n = n + 1$.
8. Go to step 2.

The choice of $q(\boldsymbol{x}'|\boldsymbol{x})$ and deciding when a chain has converged are the primary practical complications. The M-H algorithm can be optimized for a particular problem by a judicious choice of $q(\boldsymbol{x}'|\boldsymbol{x})$. Poor choices can lead to extremely inefficient sampling and hence slow convergence. The most efficient choice for $q(\boldsymbol{x}'|\boldsymbol{x})$ would be $p(\boldsymbol{x}'|\boldsymbol{d})$, the posterior probability distribution itself. However, this is not possible for this application, since the whole purpose of the Markov chain is to calculate the posterior distribution. In such cases, a common choice for $q(\boldsymbol{x}'|\boldsymbol{x})$ is a Gaussian distribution centered around $\boldsymbol{x}$. Still, there remain important choices about the correlations and scale of the candidate transition probability distribution. We will address correlations later in this section, but we first discuss the choice of scale. If the trial states are chosen with too large a dispersion, then a large fraction of the trial states will be rejected, causing the chain to remain at each state for several trials and to converge very slowly. If the trial states are chosen with too small a dispersion (σ_q), then the small step size will cause the chain will behave like a random walk, that is, the number of steps required for the chain to traverse a distance L in parameter space would scale roughly as L^2/σ_q^2.

Monitoring the fraction of trial states that are accepted is one way to verify that the scale chosen for $q(\boldsymbol{x}'|\boldsymbol{x})$ is not too inefficient. Optimal values for the acceptance rate have been determined for some simple cases. For example, when $\boldsymbol{x}$ has only one dimension and the posterior distribution function is known to be Gaussian, then an acceptance rate of $\sim$0.44 is optimal, among the class of Gaussian candidate transition probability distribution functions centered on the current value of $\boldsymbol{x}$. For parameter spaces with many dimensions, a similar analysis yields an optimal acceptance rate of $\sim$0.25 (Gelman et al. 2003).

Thus, it is common to choose a Gaussian $q(\boldsymbol{x}'|\boldsymbol{x})$ centered on $\boldsymbol{x}$ with some guess for the scale parameters. If, after running the Markov chain for some time, it becomes clear that the acceptance rate is significantly different from the desired acceptance rate, then the scale parameters are adjusted, the previous Markov chain is discarded, and a new Markov chain is begun. It may be necessary to repeat this several times to determine an acceptable set of scale parameters. Note that Markov chains that serve as the basis for altering the step size are not combined with the final Markov chain for inference.

For multidimensional parameter spaces, it is not always obvious how to change $q(\boldsymbol{x}'|\boldsymbol{x})$ so as to obtain the desired acceptance rate. Even if all the parameters are uncorrelated, there are multiple possible scale parameters. For this reason, we chose to use a special case of the M-H algorithm, widely known as the Metropolis-Hasting algorithm within the Gibbs sampler. The Gibbs sampler generates a trial $\boldsymbol{x}'$ by altering only a subset of the parameters in $\boldsymbol{x}$ for each step. We combine the Gibbs sampler with a Gaussian candidate transition function; that is, for the parameters in $(\boldsymbol{x})$ to be updated, the candidate transition probability function is

$$q(x'_\mu|x_\mu) = \frac{1}{\sqrt{2\pi\beta_\mu^2}} \exp\left[-\frac{(x'_\mu - x_\mu)^2}{2\beta_\mu^2}\right] \qquad (12)$$

for valid x'_μ (i.e., if the model dictates that x'_μ be positive definite, then trial states with negative x'_μ are rejected). We use the index μ to distinguish elements of the vector of parameters and the index n to indicate the nth step of the Markov chain. Each β_μ is a parameter that controls the size of the steps for the parameter indicated by μ. Thus, our acceptance probability reduces to

$$\alpha(\boldsymbol{x}'|\boldsymbol{x}) = \min\left\{\exp\left[\frac{\chi^2(\boldsymbol{x}) - \chi^2(\boldsymbol{x}')}{2}\right], 1\right\} \qquad (13)$$

for valid $\boldsymbol{x}'$, and $\alpha(\boldsymbol{x}'|\boldsymbol{x}_n) = 0$ for invalid parameter values.

While there are several variations, we choose the parameters to be altered in the next trial state according to randomly generated permutations of the model parameters. If the parameters are chosen one at a time, then it is easy to monitor the acceptance rates for steps involving each parameter separately. If any of these acceptance rates differ from the desired acceptance rate, then the guilty scale parameters, β_μ, can be identified and adjusted for calculating the next Markov chain more efficiently.

The other major difficulty in applying MCMC is determining how long the Markov chain should be before using it for inference. In practice, one performs multiple tests that can positively identify chains that have not converged. The failure of such tests to demonstrate nonconvergence suggests, but does not prove, that the chain has converged (Chen et al. 2002). In our experience, we have found it advisable to construct multiple Markov chains for comparison. If all the chains have converged, then the distributions of all quantities of interest should be similar to within statistical uncertainties (Gelman & Rubin 1992).

4.4. *Comparison*

Both refitting to simulated data sets and MCMC provide reasonable methods of assessing the uncertainty in model parameters. Refitting to simulated data estimates the distribution of the MLE for the simulated data, which is not necessarily the most desirable estimator. Transcribed into the language of Bayesian statistics, the MLE is not necessarily the mean or median of the parameter's marginal distribution. MCMC samples from the joint posterior distribution given a prior and the observational data.

While changing the prior would require completely redoing an MLE analysis, importance sampling can be applied to a realization of a Markov chain to compute confidence intervals for several different priors simultaneously (Press et al. 1992; Gilks et al. 1995). This could become valuable when considering alternative formation scenarios (e.g., alternative distributions for initial eccentricities and inclinations) or when additional constraints become available (e.g., new observations or long-term orbital stability studies). Note that for importance sampling to be accurate, the posterior distributions under the different priors must not be too dissimilar.

Further, calculating the best-fit model for each set of simulated data requires repeatedly solving a complex and time-consuming nonlinear minimization problem. Similarly, most techniques that are less sensitive to local minima (e.g., GAs) are less efficient than MCMC for multiple-planet systems, which have a large number of free parameters. Calculating a longer Markov chain is a relatively simple task and computationally more efficient. Thus, we expect the MCMC technique to generalize to multiple-planet systems much better than refitting.

While both methods may be practical when fitting to a small number of model parameters, MCMC is particularly useful for high-dimensional parameter spaces. This advantage is particularly

important for long-period planets and multiple-planet systems, where there are multiple free parameters that can be traded off against each other to obtain similarly good fits. The radial velocity observations of such systems can result in large valleys in the $\chi^2(\boldsymbol{x})$ surface that permit a broad range of parameter values. Within these valleys, there is often a rugged $\chi^2(\boldsymbol{x})$ surface, meaning that there are many nearby local minima for $\chi^2(\boldsymbol{x})$. While the rugged $\chi^2(\boldsymbol{x})$ surfaces present difficulties for local minimization routines, the MCMC method is able to jump between these local minima and accurately calculate the posterior probability distribution for model parameters. It is important to realize that if there were multiple local minima separated by a barrier region of parameter space that resulted in significantly larger values of χ^2, then MCMC would have great difficulty sampling from the posterior distribution unless special care was taken to identify the well-separated local minima and a candidate transition probability function was carefully chosen to allow transitions between the local minima. Thus, it is still important to conduct a single global search over the parameter space to recognize if there are distinct orbital solutions separated by a large $\chi^2(\boldsymbol{x})$ barrier.

In summary, MCMC has the following advantages compared with the other methods we have discussed:

1. MCMC naturally allows for correlated and non-Gaussian uncertainties in fit parameters.
2. The results can be simply interpreted as the joint posterior distribution for a given prior and set of observational data.
3. In some cases the resulting distribution can be efficiently updated to account for additional observations or alternative priors, provided they do not greatly alter the posterior distribution.
4. Calculating the next step in a Markov chain is much faster than performing an additional minimization with resampled data.
5. Computationally, MCMC is more efficient than other techniques, particularly for high-dimensional parameter spaces.

5. EXAMPLE APPLICATIONS

To illustrate the application of the MCMC method to fitting radial velocity data, we will present several example applications to the presently known extrasolar planets. For calculating the Markov chain, we use a set of parameters

$$\boldsymbol{x} = (\log P, \log K, e\cos\omega, e\sin\omega, M_0, \boldsymbol{C}), \qquad (14)$$

where $\boldsymbol{C}$ is the set of mean velocity offsets (one offset for each observatory). It is common practice to choose a "noninformative" prior that is uniform in the logarithm of a positive definite magnitude (e.g., orbital period and velocity semiamplitude), as suggested by several scaling arguments (Gelman et al. 2003). In addition, we found that the use of $\log P$ and $\log K$ instead of P and K increased the rate of convergence for systems in which the orbital period was not tightly constrained. Similarly, we found that the use of $e\sin\omega$ and $e\cos\omega$ instead of e and ω significantly increased the rate of convergence for systems with small eccentricities, where ω is not tightly constrained.

The distribution of published orbital parameters is roughly consistent with being uniform in $\log P$, $\log K$, e, ω, and M_0 (aside from cutoffs at small orbital periods and large planetary masses and not including the very short period planets, for which tidal forces have presumably circularized the orbits). Therefore, for the histograms and contour plots presented below we have sampled from the Markov chains using importance sampling so that they correspond to priors that are flat on ($\log P$, $\log K$, e, w, M_0).

For each planetary system, we initialize multiple chains with parameter values near the published values of the orbital elements or the best-fit models identified with a GA. We also start chains with initial parameters drawn from a Gaussian distribution centered on the published values with standard deviations 3 times the published uncertainties. In cases where our Markov chains indicated uncertainties greater than the published values, we constructed additional chains with initial parameters chosen over an even wider range.

For these simulations, we have used a Gaussian candidate transition probability function (eq. [12]) and a Gibbs sampler that randomly chooses to update one or two parameters at each step with equal probability. (When two parameters have significant correlations, perturbing two parameters in a single step can allow the Markov chain to explore parameter space more quickly.) All parameters other than the one or two chosen to be updated are left unchanged. The order of the parameters to be altered is determined by a randomly generated permutation. We monitor the acceptance rates for steps involving each parameter separately. If a chain has an acceptance rate below 0.25 or above 0.55, then we adjust the relevant β_μ and start computing a new chain. To reduce the dependence on the initial parameter values, we discard the initial 10,000 states or the first 10% of the chain, whichever is longer. Since consecutive states can are often highly correlated, we lose little information by considering only every mth set of parameter values, with m greater than or equal to 10 times the number of free parameters.

We test for nonconvergence of each chain by comparing the marginalized distribution of each parameter in the first half and full chain. We also compare the results of multiple chains run with multiple values of $\boldsymbol{\beta}$. This helps us recognize chains where an inappropriate step size may have resulted in the chain's becoming trapped in a small region of parameter space for the duration of the chain. In addition, we run multiple chains with widely dispersed initial conditions to verify that $\hat{R}^{1/2}$, the Gelman-Rubin test statistic, is consistent with convergence. Since we initialize the chains with a wide range of initial parameter values, it is easy to recognize that the distributions of parameters based on only the early portion of the chains are dependent on the values chosen for the initial state of the Markov chain. The Gelman-Rubin test statistics can determine if the Markov chains have yet to converge by comparing the variance of each parameter within each chain with the variance of the parameter across multiple chains. For Markov chains that converge, the Gelman-Rubin test statistic approaches 1.0 from above. For our results we have checked that it is always less than 1.2 and usually less than 1.1 (Gilks et al. 1995). The plots presented are based on only the Markov chain with the most desirable acceptance rates.

These very conservative choices are computationally inefficient, but the inefficiency can be tolerated for the purposes of this paper. For each system, we have computed tens of Markov chains, including a wide range of $|\beta|$. Each of our Markov chains typically contains 10^6–10^{10} steps, depending on the number of fit parameters and the choice of $|\beta|$. (When $|\beta|$ results in too high or low an acceptance rate, convergence requires many more steps.) Ideally, one would attempt to improve the computational efficiency while maintaining confidence that the chains have converged. Based on our experience from the systems studied in this paper, we believe that one could construct only five Markov chains with a single value of $|\beta|$ (chosen to give a desirable acceptance rate) for each system and still be reasonably confident that the Markov chains had converged. We believe that further refinements to the candidate transition probability function could make MCMC even more efficient. Using additional changes of

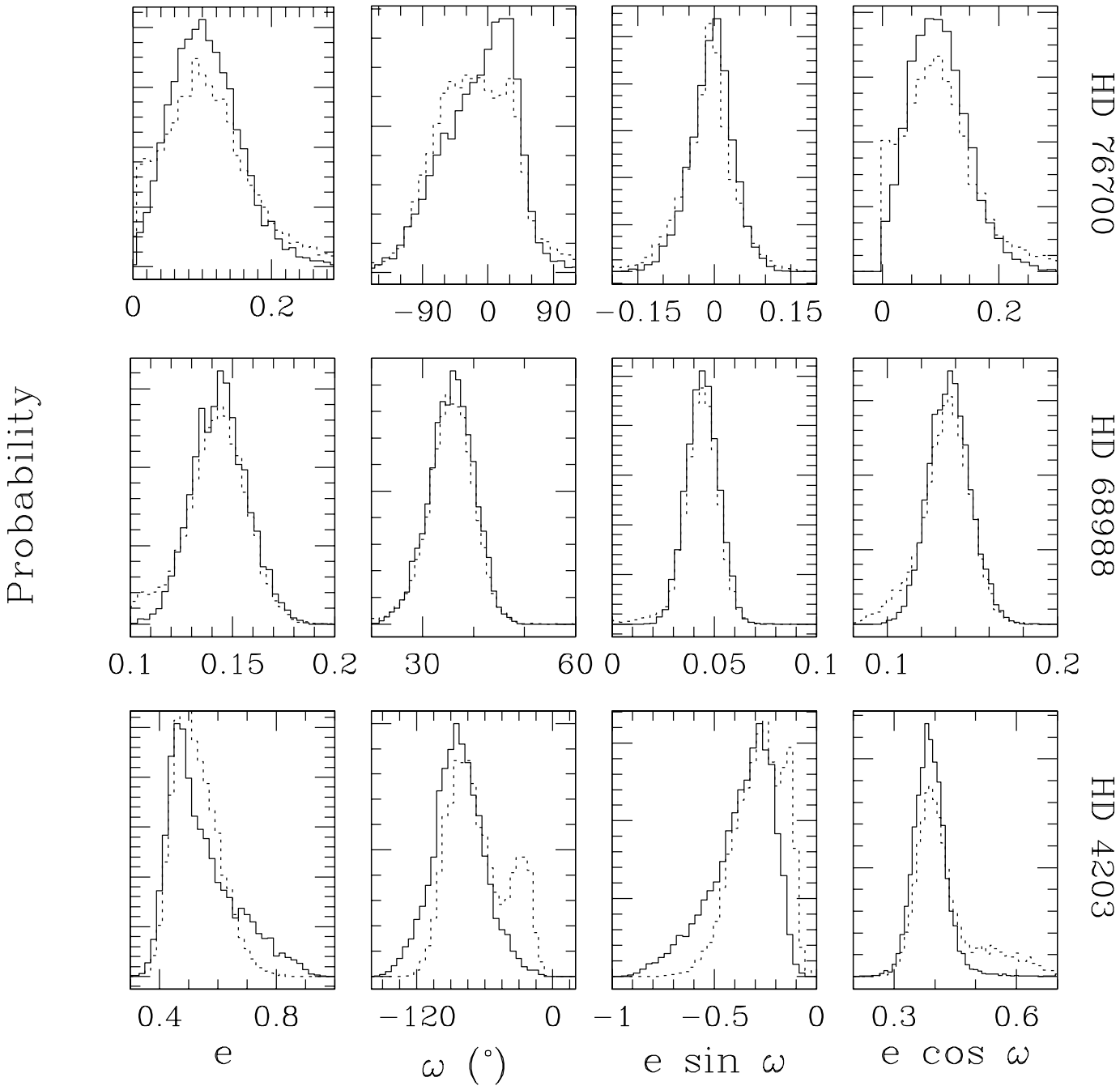

FIG. 2.—Distribution of eccentricities (*far left*), longitudes of periastron (*center left*), and two related quantities, $e \sin \omega$ (*center right*) and $e \cos \omega$ (*far right*), for the companions to HD 76700 (*top*), HD 68988 (*middle*), and HD 4203 (*bottom*). The solid lines represent the results of our MCMC simulations, while the dotted lines show the results of our fits to resampled data based on the published observations and uncertainties (Vogt et al. 2002; Tinney et al. 2003).

variables could reduce covariances between model parameters, speed convergence, and reduce the autocorrelation of the states in each chain. We hope to present such optimizations in a future paper.

5.1. *Eccentricities*

Despite the nearly Gaussian uncertainties for the radial velocity data, the uncertainties for orbital parameters can be significantly non-Gaussian. In particular, since the orbital eccentricity is confined between zero and one, we might expect significant deviations from normality when the eccentricity is small or nearly unity. In Figure 2, we show the distribution of eccentricities and arguments of periastron for HD 76700, 68988, and 4203 based on the published observations and uncertainties (Vogt et al. 2002; Tinney et al. 2003). There are several interesting features in these plots:

1. For HD 76700, the distributions of orbital elements calculated by resampling and MCMC are similar. We find the distribution of eccentricities barely includes $e = 0$ (assuming a uniform prior in $\boldsymbol{x}$ as defined in eq. [14]) and excludes $e \leq 0.017$ at the 90% confidence level, in contrast to the published solution, $e = 0 \pm 0.04$. While the published orbital solution does not constrain ω, our simulations constrain the argument of pericenter to $-100^\circ \leq \omega \leq 60^\circ$ (90% confidence interval). In addition, the

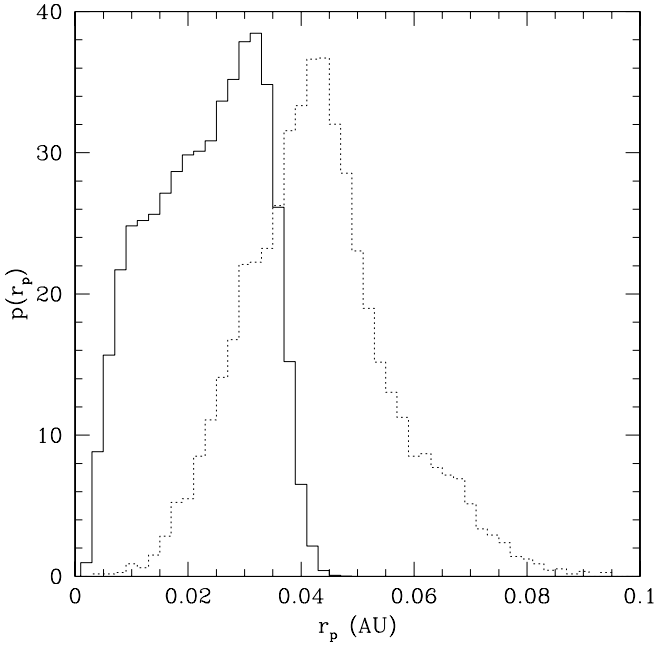

FIG. 3.—Probability distribution for the periastron distance of the companion to HD 80606. Here we assume a stellar mass of 1.1 $M_{\odot}$ and inclinations corresponding to $\sin i = 1$. The solid line shows the results of MCMC simulation, and the dotted line shows the results of fitting to resampled data. Considering the strong dependence of the tidal circularization rate on the periastron distance, the uncertainty in the pericenter distance makes it impossible to constrain theories of tidal circularization based on the present observations.

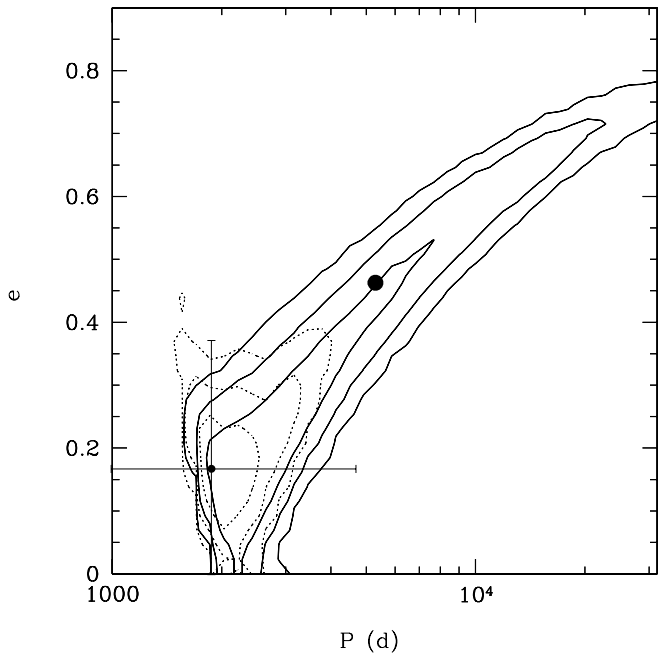

FIG. 4.—Probability distribution for HD 72650 marginalized over all variables except the period and eccentricity. The solid contours show the 1, 2, and 3 σ confidence intervals, defined to contain 68.3%, 95.4%, and 99.73% of the probability distribution, based on MCMC simulations. The dotted contours also enclose the 1, 2, and 3 σ confidence intervals, but based on resampling methods. The point with error bars indicates the published orbital solution and uncertainties. Both sets of contours assume a uniform prior in log P and e. The large filled circle indicates the best-fit orbital solution based on more recent observations made after the rest of this plot had been prepared (G. W. Marcy 2003, private communication). Unfortunately, there are often large uncertainties in the orbital elements derived for long-period planets.

distribution of ω has a noticeable asymmetry. For planets with small eccentricities such as HD 76700, the variables $e \sin \omega$ and $e \cos \omega$ generally have smaller uncertainties and are more Gaussian than the uncertainties in e and ω.

2. For HD 68988, the eccentricity distribution is clearly separated from zero, and the hypothesis that $e < 0.1$ is strongly rejected within the single-planet Keplerian model. Again, resampling and MCMC methods result in similar uncertainty estimates.

3. HD 4203 has a significant eccentricity; however, the published observations still permit strongly non-Gaussian uncertainties in e. In this case, MCMC results in an uncertainty distribution with a more significant high-eccentricity tail and a narrower distribution in ω than is predicted by resampling.

HD 80606 has a very large eccentricity ($e \simeq 0.93$) and a very small radial separation at periastron ($r_p \simeq 0.035$ AU) (Naef et al. 2001). While the orbital period is very well constrained, e and thus r_p have significant uncertainties (Fig. 3), especially in the context of tidal dissipation, since the rate of tidal circularization is a steep function of r_p (Rasio et al. 1996). Our MCMC simulations show that the current observations still permit orbital solutions with pericenter distances as small as a stellar radius. It is interesting to note that MCMC and resampling methods result in significantly different distributions for r_p. The high correlation between variables does slow the convergence of the Markov chains, but we have run several chains for an extended duration and none resulted in a significant probability of $r_p \geq 0.045$ AU. If additional observations were to constrain the pericenter distance, this system might eventually provide an interesting test for theories of tidal circularization and orbital decay.

5.2. *Long-Period Systems*

Constraining the orbital parameters of long-period systems is a significant challenge, since good solutions typically require observations spanning at least one orbital period. For eccentric systems, it is much more difficult to identify optimal times to observe the system while still during the first or second orbit. In Figure 4, we show the joint distribution of the period and eccentricity for the system HD 72659. Note that the MCMC simulations reveal a much larger uncertainty in orbital period and eccentricity than suggested by resampling or by the published error bars. After this plot was prepared, G. W. Marcy (2003, private communication) informed us that more recent observations of HD 72659 yield a best fit of $P = 5828$ days and $e = 0.47$ (*large filled circle*), demonstrating the potential for MCMC to provide more accurate uncertainty estimates than resampling.

In Figure 5 we show the marginalized probability distributions for P, K, e, and ω for several of the known extrasolar planets with the longest orbital periods (Fischer et al. 2002b; Jones et al. 2002a, 2002b, 2003; Butler et al. 2003). While the parameters are well constrained for some systems, significant uncertainties in the periods and eccentricities appear to be common. Both the mean value and uncertainty of orbital elements determined by MCMC simulations can be significantly different from the values and uncertainties suggested by resampling, particularly when at least one parameter has a significantly skewed distribution.

5.3. *Multiple Orbital Solutions*

For some systems the observational data do not determine a unimodal orbital solution; in particular, there can be two similarly good fits that are well separated from each other. This is particularly common in multiple-planet systems. In Figure 6, we show

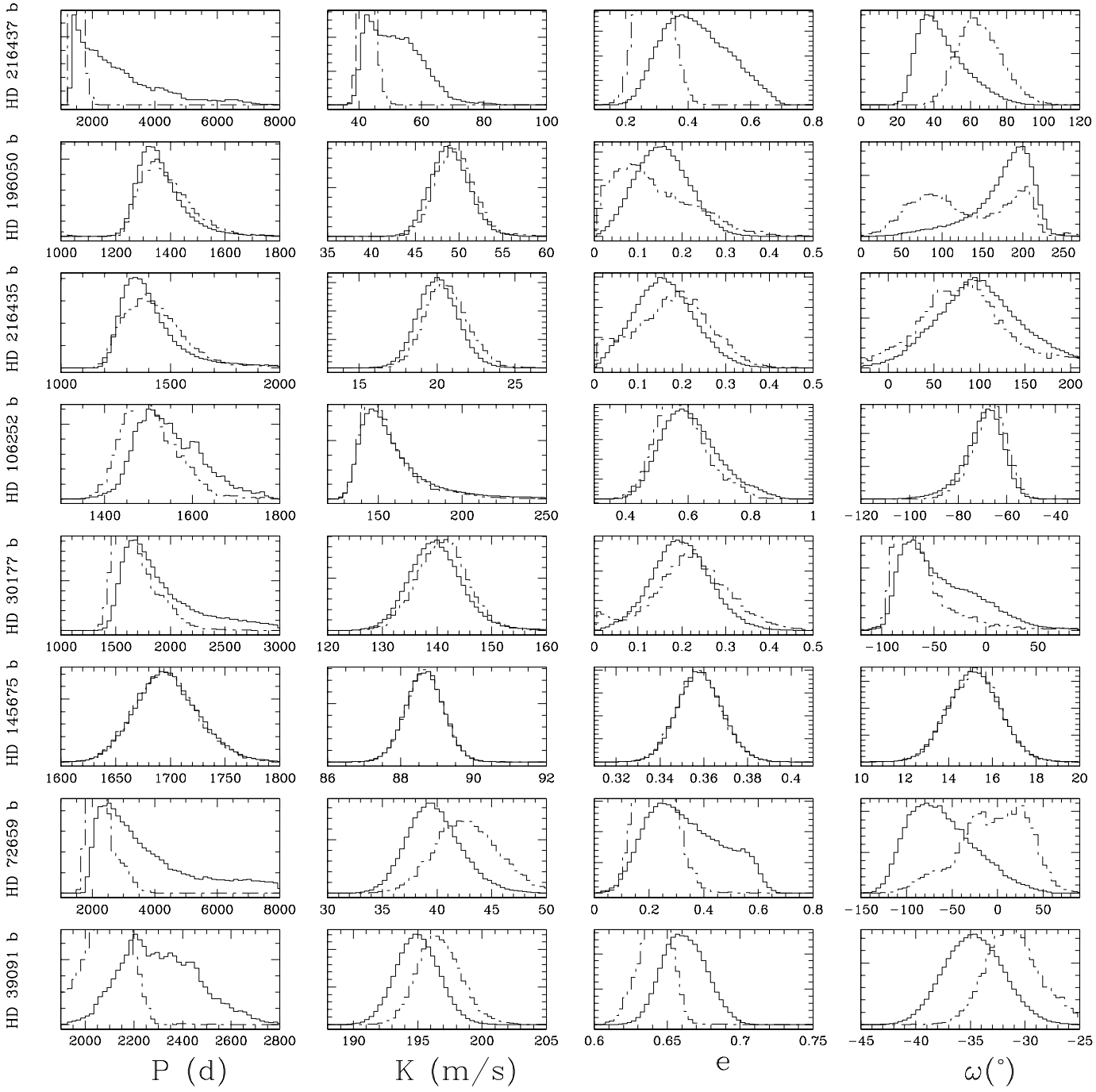

FIG. 5.—Probability distribution marginalized over all but one variable for the eight extrasolar planets with the longest orbital periods (we restrict ourselves to systems with only one known planet). The solid lines show the results of our MCMC simulations. The dashed lines show the results of our fits to resampled data. The area under each curve is normalized to unity; however, the calculated distributions sometimes have tails that extend beyond the domain of these figures.

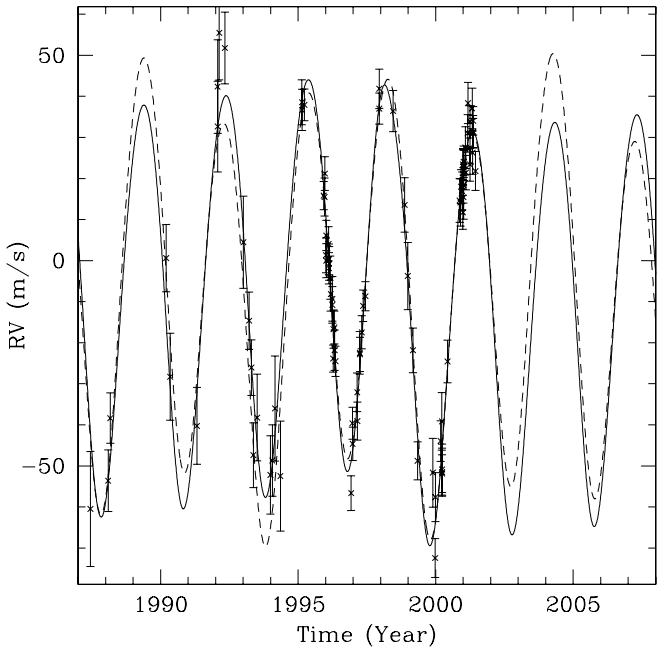

FIG. 6.—The published radial velocity data for 47 UMa along with two fits to the data, each employing two noninteracting planets on Keplerian orbits (Butler et al. 1996; Fischer et al. 2002a; Laughlin et al. 2002). The solid line is the published solution, while the dashed line is the solution for a much larger period (3 times the published best-fit period) and eccentricity ($e \simeq 0.8$). The χ^2 of the alternative fit is actually less than that of the published fit, but the difference is not significant. While the dashed solution is not favored by orbital stability requirements, 47 UMa illustrates that the observations on multiple-planet systems can sometimes be well modeled by very different orbital solutions.

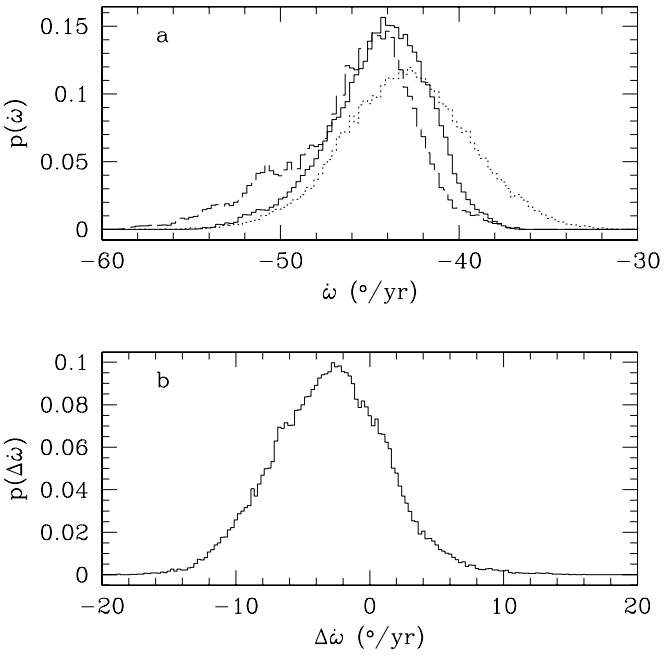

FIG. 7.—*Top*: Probability distributions for the precession rates of the periapses of GJ 876b (*dotted line*) and 876c (*dashed line*). The solid line is the instantaneous average of the two precession rates. *Bottom*: Distribution for the difference in the precession rates.

two possible orbital solutions for 47 UMa. One corresponds to the published orbital solution and the other corresponds to a solution in which the outer planet has a much longer orbital period and large eccentricity.

We have begun conducting MCMC simulations to estimate the parameter distributions for 47 UMa and other multiple-planet systems. For multiple-planet systems, several orbital parameters can be highly correlated, greatly slowing convergence. We are exploring changes of variables that could accelerate convergence and permit more detailed investigations of the uncertainties in the elements of such multiple-planet systems.

5.4. *Interacting Multiple Planet Systems*

The most easily observable deviation from Keplerian motion is generally expected to be precession of the longitude of periastron. This suggests we adopt a precessing Keplerian model,

$$\Delta v_{*,i}(t) = K_i\{\cos\,[T_i(t) + \omega_{0,i} + \dot{\omega}_i t] + e_i \cos\,(\omega_{0,i} + \dot{\omega}_i t)\}, \quad (15)$$

for each planet, where $\dot{\omega}_i$ is the precession rate of the longitude of periapse of the ith planet and t is the time of the observation. Such a model incorporates the dominant observable perturbation to the basic Keplerian model while introducing only one additional parameter per planet. This model is particularly good when the perturbations excite a single dominant secular eigenmode, as in GJ 876, in which case the values of $\dot{\omega}_i$ should be nearly equal and time-independent. Further, in this model $\dot{\omega}$ can be fitted to radial velocity measurement without time-consuming N-body integrations. Although radial velocity measurements constrain only $m \sin i$, the precession rates can in principle determine the planets' inclinations and masses. Of course, this approach assumes that there are no additional perturbations, for example, due to general relativistic effects, the quadrupole moment of the star, or additional companions.

The GJ 876 system is an example of an interacting system in which mutual perturbations are detectable (Marcy et al 1998). Modeling GJ 876 with two precessing Keplerian orbits and applying MCMC, we investigate the precession rates of the two orbits. In the top panel of Figure 7, we show the probability distributions for $\dot{\omega}_1$ (*dotted line*), $\dot{\omega}_2$ (*dashed line*), and the instantaneous average of the two (*solid line*). In the bottom panel we plot the probability distribution for the difference of the precession rates. In this case, the two precession rates are comparable, allowing for a secular resonance in which $\Delta\omega$ remains small (Lee & Peale 2002). Based on the published observations, we cannot yet detect a significant difference between the precession rates of the two planets. Rates differing by more than $\sim 10^\circ$ yr^{-1} can be excluded according to this model.

5.5. *Possible Transit Times*

Radial velocity measurements can be used to identify when a transit could occur, if the system were to have a favorable inclination. This can greatly aid observers, who can concentrate their efforts at the appropriate times. Correlations between best-fit orbital parameters can have a significant impact on the search for transits. Thus, it is important that observers understand the uncertainty in the predicted transit times. In particular, for strongly interacting multiple-planet systems, there can be significant uncertainties in the possible transit times.

We have fitted a model based on equation (15) to the published radial velocity data for GJ 876. Using a genetic algorithm, we found a model with precession rates of $-44.^\circ0$ yr^{-1} and $-46.^\circ8$ yr^{-1} for the inner and outer planets, respectively. These are similar to the precession rates found for the best-fit

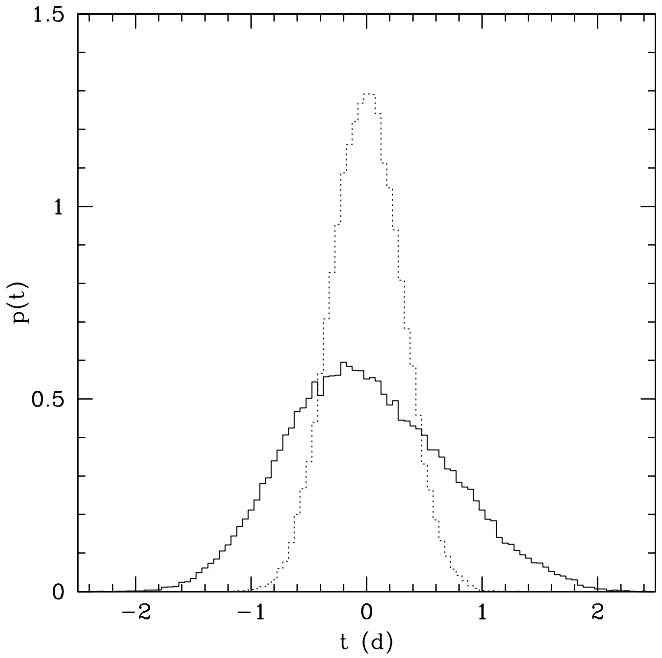

FIG. 8.—Times of the center of potential transits for GJ 876b approximately 2 months after the last published radial velocity, assuming $\sin i = 1$. The dotted line shows the prediction using a superposition of two noninteracting Keplerian orbits to model the data, while the solid line allows each orbit to precess at an arbitrary rate. While our model does not include the short-period perturbations important for actually forecasting potential transits, it does demonstrate the uncertainty of transit times in an interacting system.

orbital solution obtained by means of self-consistent N-body integrations and local minimization (Laughlin & Chambers 2001). Clearly, this comparison does not include important short-period perturbations to the orbits, which should be included when forecasting transits and planning observations. In addition, a transit search would want to target the time of ingress or egress, which introduces additional dependences on the stellar radii and orbital parameters. Nevertheless, we use this model to demonstrate the uncertainty in potential transit times in an interacting system and the applicability of MCMC to predicting potential transit times. The actual system is most likely precessing at a similar rate, and the unknown precession rate contributes to the uncertainty in predicting potential transit times. In Figure 8, we show the probability distributions for the center of the transit of GJ 876b approximately 2 months after the last published observation, assuming $\sin i = 1$. The model that includes the precession rates (*solid line*) allows for a significantly wider range of transit times than the model that assumes fixed periapses. This significantly increases the duration over which GJ 876 must be monitored to ensure that the ingress or egress falls within the window of observations.

6. CONCLUSIONS

Markov chain Monte Carlo methods provide a valuable tool for matching planetary orbits to radial velocity data. These techniques can be particularly useful for characterizing the uncertainties and correlations in estimated orbital parameters for extrasolar planets. Presently, several systems have large uncertainties in orbital parameters, particularly the long-period planets and multiple-planet systems. Our Markov chain Monte Carlo simulations reveal that the published orbital solutions can significantly over- or underestimate the uncertainty in orbital parameters. In addition to quantifying the uncertainties of various orbital parameters, Markov chain Monte Carlo allows specific questions to be investigated directly from the radial velocity data themselves, bypassing fits to orbital parameters. As more multiple-planet systems are discovered, such methods could become increasingly valuable.

We thank Gilbert Holder and Hiranya Peiris for discussions of Markov chain Monte Carlo techniques. We thank Geoff Marcy and Debra Fischer for discussions of observational issues and sharing radial velocity observations before publication. We thank Gregory Laughlin for bringing to our attention the uncertainties in possible transit times, and John Chambers for valuable discussions about GJ 876. This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration.

REFERENCES

Barnes, T. G., III, Jefferys, W. H., Berger, J. O., Mueller, P. J., Orr, K., & Rodriguez, R. 2003, ApJ, 592, 539 (erratum 611, 621 [2004])
Brown, R. A. 2004, ApJ, 610, 1079
Butler, R. P., Marcy, G. W., Fischer, D. A., Brown, T. M., Contos, A. R., Korzennik, S. G., Nisenson, P., & Noyes, R. W. 1999, ApJ, 526, 916
Butler, R. P., Marcy, G. W., Vogt, S. S., & Apps, K. 1998, PASP, 110, 1389
Butler, R. P., Marcy, G. W., Vogt, S. S., Fischer, D. A., Henry, G. W., Laughlin, G., & Wright, J. T. 2003, ApJ, 582, 455
Butler, R. P., Marcy, G. W., Williams, E., McCarthy, C., Dosanjh, P., & Vogt, S. S. 1996, PASP, 108, 500
Charbonneau, P. 1995, ApJS, 101, 309
Chen, R., Liu, J. S., & Wang, X. 2002, IEEE Trans. Signal Processing, 50, 255
Chiang, E. I., Tabachnik, S., & Tremaine, S. 2001, AJ, 122, 1607
Cid Fernandes, R., Sodré, L., Jr., Schmitt, H. R., & Leão, J. R. S. 2001, MNRAS, 325, 60
Fischer, D. A., Marcy, G. W., Butler, R. P., Laughlin, G., & Vogt, S. S. 2002a, ApJ, 564, 1028
Fischer, D. A., Marcy, G. W., Butler, R. P., Vogt, S. S., Frink, S., & Apps, K. 2001, ApJ, 551, 1107
Fischer, D. A., Marcy, G. W., Butler, R. P., Vogt, S. S., Walp, B., & Apps, K. 2002b, PASP, 114, 529
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2003, Bayesian Data Analysis (2nd ed.; Boca Raton: Chapman & Hall/CRC)
Gelman, A., & Rubin, D. B. 1992, Stat. Sci., 7, 457
Gilks, W. R., Richardson, S., & Spiegelhalter, D. J., eds. 1995, Markov Chain Monte Carlo in Practice (Boca Raton: Chapman & Hall/CRC)
Hobson, M. P., & McLachlan, C. 2003, MNRAS, 338, 765
Horne, J. H., & Baliunas, S. L. 1986, ApJ, 302, 757
Jones, H. R. A., Butler, R. P., Marcy, G. W., Tinney, C. G., Penny, A. J., McCarthy, C., & Carter, B. D. 2002a, MNRAS, 337, 1170
Jones, H. R. A., Butler, R. P., Tinney, C. G., Marcy, G. W., Penny, A. J., McCarthy, C., & Carter, B. D. 2003, MNRAS, 341, 948
Jones, H. R. A., Butler, R. P., Tinney, C. G., Marcy, G. W., Penny, A. J., McCarthy, C., Carter, B. D., & Pourbaix, D. 2002b, MNRAS, 333, 871
Kashyap, V., & Drake, J. J. 1998, ApJ, 503, 450
Knox, L., Christensen, N., & Skordis, C. 2001, ApJ, 563, L95
Laughlin, G., & Chambers, J. E. 2001, ApJ, 551, L109
Laughlin, G., Chambers, J., & Fischer, D. 2002, ApJ, 579, 455
Lee, M. H., & Peale, S. J. 2002, ApJ, 567, 596
Marcy, G. W., & Butler, R. P. 1996, ApJ, 464, L147
Marcy, G. W., Butler, R. P., Fischer, D. A., Laughlin, G., Vogt, S. S., Henry, G. W., & Pourbaix, D. 2002, ApJ, 581, 1375
Marcy, G. W., Butler, R. P., Fischer, D. A., Vogt, S. S., Lissauer, J. J., & Rivera, E. J. 2001, ApJ, 556, 296
Marcy, G. W., Butler, R. P., Vogt, S. S., Fischer, D. A., & Lissauer, J. J. 1998, ApJ, 505, L147
Marcy, G. W., Butler, R. P., Williams, E., Bildsten, L., Graham, J. R., Ghez, A. M., & Jernigan, J. G. 1997, ApJ, 481, 926
Marshall, P. J., Hobson, M. P., & Slosar, A. 2003, MNRAS, 346, 489
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Naef, D., et al. 2001, A&A, 375, L27
Panter, B., Heavens, A. F., & Jimenez, R. 2003, MNRAS, 343, 1145

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C (2nd ed.; New York: Cambridge Univ. Press)
Rasio, F. A., Tout, C. A., Lubow, S. H., & Livio, M. 1996, ApJ, 470, 1187
Stępiński, T. F., Malhotra, R., & Black, D. C. 2000, ApJ, 545, 1044
Tinney, C. G., Butler, R. P., Marcy, G. W., Jones, H. R. A., Penny, A. J., McCarthy, C., Carter, B. D., & Bond, J. 2003, ApJ, 587, 423
van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. 2001, ApJ, 548, 224
Verde, L., et al. 2003, ApJS, 148, 195
Vogt, S. S., Butler, R. P., Marcy, G. W., Fischer, D. A., Pourbaix, D., Apps, K., & Laughlin, G. 2002, ApJ, 568, 352
Wang, Y., & Mukherjee, P. 2004, ApJ, 606, 654