Edinburgh School of Economics

Discussion Paper Series

Number 72

Comparing the Performance of Baseball Players: A Multiple Output Approach

Gary Koop (University of Glasgow)

Date
April 2001

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

Comparing the Performance of Baseball Players: A Multiple Output Approach

GARY KOOP
Department of Economics, University of Glasgow,
Glasgow, G12 8RT, U.K.
E-mail: G.Koop@socsci.gla.ac.uk

April 2001

In this paper, we extend ideas from the economics literature on multiple output production and efficiency to develop methods for comparing baseball players which take into account the fact that there are many dimensions to batting performance. A key part of our approach is the estimation of an output aggregator based on the performance of the best players in baseball. An individual player can then be measured relative to the best and a number between zero and one characterizes his performance as a fraction of the best. The methods are applied using data from 1995-1999 on all regular players in baseball's major leagues.

1 Introduction

Academics and sports fans alike are often interested in a statistical comparison of baseball players. This is hard to do since baseball is fundamentally a multiple output sport. For instance, some players are power hitters and excel in hitting home runs, others hit more singles. To directly compare a power hitter like Mark McGwire with someone who hits for average like Tony Gwynn is fairly meaningless. In the sporting press, there are many different ways of aggregating performance in different offensive categories into a single number. Examples of such *output aggregators* include batting average, on base percentage and slugging percentage. There are two related problems with such standard ways of combining different offensive categories into one number. First, they are all imperfect measures of offensive performance. For instance, batting average underweights the contribution of home runs and all of the output aggregators do not correct for different player circumstances (e.g. Coors Park is notorious for being a hitter's park and, accordingly, it is easier for players on the Colorado Rockies to hit well). Second, the weights in all the output aggregators are somewhat ad hoc. For instance, slugging average weights home runs precisely four times as heavily as singles; batting average weights singles and home runs equally; and on base percentage weights walks and hits equally. All such weighting choices can be criticized. The purpose of this paper is to use statistical methods to estimate an output aggregator for offensive performance of baseball players and then compare individual players to this benchmark.

The statistical methods used in this paper are adapted from the economics literature and, hence, we adopt some economic terminology. That is, we view baseball players (like firms in an economic context) as producing outputs given firm characteristics (e.g. a batter produces the "output" hits which depends on his situation including what team he plays for, etc.). In an economics context, you typically see firms operating with different output mixes just as in baseball you see batters with different mixes of hits (e.g. singles, doubles, home

runs). By looking at the best firms in different regions of output space, the economist can trace out a *production possibility curve* which measures the maximum feasible combinations of outputs which can be produced. In baseball, we can look at the best batters in different regions of output space and trace out a comparable curve. For instance, Mark McGwire might heavily influence the production possibility curve in the power dimensions while Tony Gwynn might have influence in batting average dimensions. Once we have estimated this production possibility curve, we can compare individuals to it. Since the curve reflects best practice, an average player will lie inside the curve. As discussed below, a number between zero and one can be calculated which reflects how far the player is from the nearest point on the curve. Following the economics literature, we refer to this number as *efficiency*. So, for instance, we might find a certain player has an efficiency of 0.8. This number has a simple, intuitive interpretation: the player under consideration is only 80% as productive as the best players with a comparable output mix. In other words, our methodology allows for a comparison of any player with similar players and provides an easily interpretable, single number summary of a player's performance. Furthermore, our methodology incorporates all of the outputs that a batter produces and corrects for the player's situation (e.g. playing in a particularly good or bad hitter's park).

The literature which uses statistical methods to analyze baseball data is voluminous. However, to the best of our knowledge, there has been no previous academic study of the multiple-output character of baseball and the use of efficiency analysis to measure a player's performance. Many of the more influential papers use data on the performance of the team as a whole (e.g. Barry and Hartigan, 1993, Kaigh, 1995 or Ferrall and Smith, 1999) or focus on a particular measure or two of player performance (e.g. Albright, 1993, Berry, Reese and Larkey, 1999, Lackritz, 1996 and Schall and Smith, 2000) or discuss issues relating to salary determination (e.g. Chapman and Southwick, 1991, Depken, 2000, Hoaglin and Velleman, 1995, Kahn, 1993, Scully, 1974 and Watnik, 1998).

We derive a Markov Chain Monte Carlo (MCMC) algorithm for carrying out Bayesian

estimation of player efficiencies. We use this algorithm to carry out statistical inference relating to the performance of all batters who played regularly during the years 1995-1999. Our empirical findings are sensible. For instance, our list of the top 15 ranked players contains the names of many of the top players in the game.

The remainder of the paper is organized as follows. The next section discusses the data. The third section develops the statistical model, while the fourth section presents empirical results. The fifth section concludes.

2 The Data

The data used in this study was obtained from Sean Lahman's Baseball Archive (www.baseball1.com) which contains a myriad of statistics on everyone who has ever played major league baseball. In order to focus the study, we consider only performance indicators relating to batting. Similar models could be developed to investigate pitching, fielding or base running performance. Accordingly, we omit all pitchers from our sample and also omit players with fewer than 200 at bats in a given year. Having fewer than 200 at bats usually reflects injury or very irregular play. Furthermore, Berry, Reese and Larkey (1999) provides convincing evidence that a player's ability can change substantially over the course of his career. To avoid explicitly modelling such aging affects, we use data from 1995-1999 and assume that a player's ability is roughly constant over such a short time span. Our preferred specification (see below) is a logged one and, hence, does not accommodate zero values for any of the outputs. Accordingly, we add doubles and triples together into one output (since triples are quite rare) and delete all player years with zero values for any of the outputs. This removes only 19 observations (i.e. player-years).

The resulting data set contains 1,492 observations on 535 different players. As outputs we use:

- y_1= number of singles per at bat,

- y_2= number of doubles plus triples per at bat,
- y_3= number of home runs per at bat,
- y_4= number of walks (including intentional walks and hit-by-pitch's) per at bat.

The omission of runs scored and runs batted in (RBIs) from our list of outputs warrants explanation. These are not direct measures of batting performance and contain minimal information that is not already contained in $y_1, ..., y_4$. Runs and RBIs reflect the hitting ability of a player either to get on base or drive in other players who are already on base. This is already measured by the other outputs. Runs and RBIs also depend on the performance of other players (which is already measured in their batting statistics) and base running ability (which is not the focus of the present study). Furthermore, RBIs depend crucially on a player's location in the batting order and we do not have data on this. Accordingly, we do not include runs scored or RBIs as separate output measures.

Two players with the same level of ability can have different outputs for several reasons. We attempt to control for this through including the following explanatory variables:

- x_1= an intercept,
- $x_2 - x_5$= dummies for the years 1996, 1997, 1998 and 1999 (1995 is the omitted year),
- x_6= dummy for league (=1 for American League),
- $x_7 - x_{35}$= dummies for 29 of the 30 teams which existed for all or part of the 1995-1999 period (Baltimore Orioles are the omitted team).

The league dummy is included to reflect differences across the American and National Leagues. A primary difference is the use of the designated hitter (i.e. a batter designated to hit instead of the pitcher) in the American League. However, differences also might occur due to umpires in the different leagues having different styles (e.g. different strike

zones). The team dummies are included to reflect both physical effects (i.e. baseball parks have different characteristics making some easier to hit in than others) and proxy for the effect the other players on a team have on an individual's hitting performance (e.g. being a good hitter on a weak team will mean that opposing pitchers can "pitch around" you and not offer you any good pitches to hit, knowing that weaker hitters will follow you in the batting order).

Another possibly relevant issue is the treatment of outs. There are several ways a player can get out (e.g. strikeouts, ground outs, fly outs). Strikeouts are typically considered the worst way of getting out in that runners cannot advance on a strikeout exception in very unusual situations (although with strikeouts, the chances of a double play are small, in contrast to ground outs or even fly outs). The data set used in this paper has data on strikeouts, but not on other ways of getting out. In an early version of this paper, we included strikeouts per at bat as an undesirable output in our model (see Fernandez, Koop and Steel, 2001a,b, for discussions of multiple output modelling with undesirable outputs). Results for this model excessively penalized players with many strikeouts and did not seem reasonable (e.g. Mark McGwire was ranked as a very average player due to his high number of strikeouts). Perhaps, if data were available, including strikeouts, ground outs and fly outs as three separate undesirable outputs would be sensible. But in the absence of complete data on these, we decided to omit strikeouts altogether.

3 The Model

To illustrate the basic ideas underlying our model, consider the case where baseball players produce two outputs, power and batting average. The outputs of six hypothetical baseball players are plotted in Figure 1 and marked as (1),..,(6). If we look first at these points, it can be seen that there appears to be a trade-off between power and average. Players (1), (2) and (3) have relatively low values for power but having relatively high batting averages.

Players (4), (5) and (6) are power hitters. What our statistical methods do is to estimate a curve of the best power/average combinations that are observed in the data. This curve is plotted on Figure 1 and can be seen to be heavily influenced by players (1), (2), (5) and (6).

Direct comparisons of, say, player (1) to player (6) would be considered by many to be relatively meaningless. After all, there is no one better than either of these players at doing what they specialize in. However, player (4) is a power hitter and can be meaningfully compared to power hitters (5) or (6). Similarly player (3) can be meaningfully compared to players (1) or (2). Our statistical methods use the ratio of distances $O(4)/OA$ as a measure of the performance of player (4) and $O(3)/OB$ as a measure of the performance of player (3). Note that players are always being measured relative to a curve largely defined by the best comparable players (e.g. power hitters are compared to power hitters). Adopting terminology from the economics literature we call these measures of player performance *efficiencies*, although more accurately (but awkwardly) they could be called *performance as a proportion of the comparable best players'*.

Figure 1 is a hypothetical illustration where hitters are either power or average hitters. Of course, in reality there will be a continuum of hitters and many more outputs. Nevertheless, the same intuition holds. The performance of the best players for a given combination of outputs defines the production possibility curve at this combination of outputs. Other players are compared to curves largely defined by players with a similar mix of outputs.

Insert Figure 1 here

An alternative way of intuitively motivating our approach is in terms of functions relating a $k-$vector of explanatory variables, x, to a $p-$vector of outputs, y:

$$\theta(y) = h(x). \tag{1}$$

The positive function $\theta(y)$ should be interpreted in the same manner as the curve in Figure 1. That is, it maps out the output combinations achieved by the best players (for a given value for $h(x)$). The function $h(x)$ corrects for different player characteristics. It can be interpreted as the maximum amount of output which can be produced by a player with characteristics given by x where output is measured by the output aggregator $\theta(y)$.

Since $\theta(y)$ is, by definition, the set of output combinations achieved by the best players with characteristics x, a typical player will produce less than the value implied by $\theta(y)$. Accordingly, if we add a subscript i to indicate a particular player, and let τ_i be a measure of this shortfall with $0 < \tau_i < 1$ we can write:

$$\theta(y_i) = h(x_i)\tau_i \tag{2}$$

or

$$\tau_i = \frac{\theta(y_i)}{h(x_i)}. \tag{3}$$

τ_i is the measure of player i's efficiency used in this paper. As emphasized previously, it can be interpreted in terms of distances on graphs like Figure 1 or as a measure of performance relative to the best players with a similar output mix (i.e. player i is producing $\tau_i\%$ of the output of the best players with similar characteristics and a similar output mix).

The previous paragraphs have sketched out the basic ideas of our approach in an intuitive manner. In the remainder of this section, we formalize these ideas by assuming specific forms for $\theta(.)$ and $h(.)$, explicitly incorporating the longitudinal nature of the data, adding measurement error and making additional assumptions necessary to achieve a valid sampling model.

The model used extends that developed in Fernandez, Koop and Steel (2000). Let $y_{(i,t,j)}$ be the amount of the j^{th} output produced by, and $x_{(i,t,l)}$ the l^{th} explanatory variable of, player i at time t where $i = 1, .., N$ and $t = 1, .., T_i$. Note that we have an unbalanced

panel where, for some players, data is missing for some values of t. In order to model the production possibility curve, we define an output aggregator given by:

$$\theta_{(i,t)} = \left(\sum_{j=1}^{p} \alpha_j^q \, y_{(i,t,j)}^q \right)^{1/q}, \tag{4}$$

which depends on parameters $\alpha_j \in (0,1)$ for all $j = 1, \ldots, p$, with $\sum_{j=1}^{p} \alpha_j = 1$ and $\alpha = (\alpha_1, .., \alpha_p)'$. This is closely related to the commonly-used constant elasticity of scale function. As described in many places (e.g. Fernandez, Koop and Steel, 2000) this is a flexible form for the output aggregator which allows for a wide variety of shapes. The parameter, q, relates to the trade-off between the outputs and values of $q > 1$ imply a negative trade-off (i.e. increasing one output implies reducing at least one of the other outputs). We expect to find $q > 1$, but do not impose this in the prior. The α_j parameters ensure invariance to the scaling of the j^{th} output (e.g. measuring y_1 in terms of singles per 100 at bats instead of singles per at bat will only alter α_1 and not affect the measurement of efficiency). We can interpret $\theta_{(i,t)}$ as an aggregate output, which collapses the p-dimensional output vector into one dimension.

Defining the $NT-$vector containing the logs of the aggregate outputs as:

$$\log \theta = (\log \theta_{(1,1)}, \log \theta_{(1,2)}, \ldots, \log \theta_{(1,T_1)}, \ldots, \log \theta_{(N,T_N)})', \tag{5}$$

where $NT = \sum_{i=1}^{N} T_i$, we model $\log \theta$ as depending on an $NT \times k$ matrix of explanatory variables,

$$X = \begin{bmatrix} x_{(1,1,1)} & x_{(1,1,2)} & . & x_{(1,1,k)} \\ x_{(1,2,1)} & . & . & \\ . & . & . & \\ x_{(N,T_N,1)} & x_{(N,T_N,2)} & . & x_{(N,T_N,k)} \end{bmatrix}$$

and an N-vector of individual effects, $z = (z_1, z_2, ..., z_N)'$, and add on a term to reflect measurement error:

$$\log \theta = X\beta - Dz + \varepsilon. \tag{6}$$

The $(NT \times N)$-matrix D is selected to match the individual effect with the appropriate individual. That is,

$$D = \begin{bmatrix} \iota_{T_1} & 0 & . & 0 \\ 0 & \iota_{T_2} & . & . \\ . & 0 & . & 0 \\ 0 & 0 & . & \iota_{T_N} \end{bmatrix}$$

where ι_S is an S-dimensional vector of ones.

Some motivation for the specification in (6) is called for. One way of interpreting it is as a standard panel data stochastic frontier model (see, e.g., Schmidt and Sickles, 1984) in terms of the aggregate output, θ. Adding the term z_i is crucial to our analysis as it can be interpreted as a measure of the ability of player i in producing the aggregate output. The assumption that a player's ability is roughly constant over the short time span of the data implies that z_i is constant over time. Given that we are interested in measuring player performance relative to a benchmark established by the best players, z_i must have a one-sided distribution to ensure $z_i > 0$. That is, $\theta(y) = h(x)$ defines the frontier established by the best players and, hence, all players lie on or within this frontier and thus $z_i > 0$ (if we ignore measurement error). Given the log specification in (6), it follows that player efficiency (see equation 3 or Figure 1) is given by:

$$\tau_i = \exp(-z_i) \tag{7}$$

where $0 < \tau_i < 1$. Fernandez, Koop and Steel (2000) provides additional graphical motivation for the output aggregator given in equation (4) and the way in which an equation analogous to (6) provides a sensible efficiency measure.

In order to keep the number of parameters reasonably small relative to the number of observations, it is standard to make a distributional assumption for z. Such an assumption can either be treated as a hierarchical prior or as part of the likelihood function. We

assume:

$$p(z|\lambda_0, \lambda_1) \propto \prod_{i=1}^{N} f_G\left(z_i|, \lambda_0, \lambda_1\right) \tag{8}$$

where $f_G(z_i|a, b)$ denotes the p.d.f. of a Gamma distribution with mean a/b and variance a/b^2. We note that previous work with stochastic frontier models typically uses more restrictive distributions such as the exponential or the half Normal. However, we justify the use of the Gamma on the grounds that, with the large amount of data we have, it is appropriate to have a more flexible form to let the data determine the form of the efficiency distribution. Furthermore, histograms of the data indicate that the Gamma is probably more suitable than the exponential or half-Normal. For instance, histograms of $-\ln(y_{(i,t,j)})$ should give us a rough idea of the shape of the distribution of z_i. These typically have an interior mode and are skewed to the right. Such properties can easily be accommodated using the Gamma distribution, but not the exponential, half Normal or even the more general truncated Normal.

The Normally distributed error term, ε, reflects the usual measurement error and model imperfections. If we let $f_N^R(\varepsilon|a, A)$ denote the R-variate Normal p.d.f. with mean a and covariance matrix A, we assume:

$$p(\varepsilon|\Sigma) = f_N^{NT}\left(\varepsilon|0, h^{-1}I_{NT}\right), \tag{9}$$

where h is the error precision and I_{NT} is the identity matrix.

It is important to stress that equation (6), plus the assumptions in (8) and (9), are not enough to fully describe the likelihood function. For each player in each year, we have data on p outputs. A single equation such as (6) is not adequate to define a likelihood function for $p > 1$ endogenous variables. If α and q were known, then (6) would provide a valid likelihood function for $\log(\theta)$ but not the individual outputs which go into this aggregate output. In order to complete specification of a likelihood function, we need to say something about the individual outputs. As in Fernandez, Koop and Steel (2000), we

do so by defining the weighted output shares:

$$\eta_{(i,t,j)} = \frac{\alpha_j^q y_{(i,t,j)}^q}{\sum_{l=1}^p \alpha_l^q y_{(i,t,l)}^q}, \ j = 1, \ldots, p, \tag{10}$$

which can be interpreted as the contribution of an individual output to the aggregate. Since $\sum_{j=1}^p \eta_{(i,t,j)} = 1$, (10) provides the extra $(p-1)$ dimensions necessary to provide a valid probability density for the p outputs. Since $0 \leq \eta_{(i,t,j)} \leq 1$ it is logical to group these shares as $\eta_{(i,t)} = (\eta_{(i,t,1)}, \ldots, \eta_{(i,t,p)})'$, and assume independent sampling from

$$p(\eta_{(i,t)}|s) = f_D^{p-1}(\eta_{(i,t)}|s), \tag{11}$$

where $s = (s_1, \ldots, s_p)'$ and $f_D^{p-1}(\cdot|s)$ is the p.d.f. of a Dirichlet distribution with parameter vector s. Equations (10) and (11) are not only added simply to complete the likelihood function. As we shall see below, η has a simple intuitive interpretation relating to there being different types of hitters (e.g. those who hit for power, for average, etc.).

Define Y to be the $NT \times p$ matrix containing data on all outputs for all players in all years, with elements ordered in the same manner as $\log \theta$ in (5). Equations (6), (9), (10) and (11) imply the following likelihood function for Y:

$$\begin{aligned} p(Y|\beta, z, h, \alpha, q, s) = & f_N^{NT}\left(\log\theta|X\beta - Dz, h^{-1}I_{NT}\right) \\ & \prod_{i,t}\left[f_D^{p-1}(\eta_{(i,t)}|s)\left(\prod_{j=1}^p q^{1-\frac{1}{p}}\frac{\eta_{(i,t,j)}}{y_{(i,t,j)}}\right)\right]. \end{aligned} \tag{12}$$

In the present paper, we adopt a Bayesian approach. Accordingly, we interpret (8) as a hierarchical prior. In addition, we require priors for the parameters $(\beta, \lambda_0, \lambda_1, h, \alpha, q, s)$. Details on the priors are given in Appendix A. Suffice it to note here that we choose relatively noninformative priors which allow the data information to be predominant.

In order to carry out posterior inference, we develop a Markov Chain Monte Carlo (MCMC) algorithm which involves sequentially drawing from $p(\beta|Y, z, \lambda_0, \lambda_1, h, \alpha, q, s)$,

$p(h|Y,z,\beta,\lambda_0,\lambda_1,\alpha,q,s)$, $p(\alpha|Y,z,\beta,\lambda_0,\lambda_1,h,q,s)$, $p(q|Y,z,\beta,\lambda_0,\lambda_1,h,\alpha,s)$, $p(s|Y,z,\beta,\lambda_0,\lambda_1,h,\alpha,q)$, $p(\lambda_0|Y,z,\beta,\lambda_1,h,\alpha,q,s)$, $p(\lambda_1|Y,z,\beta,\lambda_0,h,\alpha,q,s)$ and $p(z|Y,\beta,\lambda_0,\lambda_1,h,\alpha,q,s)$. Details on the precise forms of these conditional posteriors, methods for drawing from them and a discussion of MCMC convergence are given in Appendix B. Suffice it to note here that we combine various techniques from Fernandez, Koop and Steel (2000), Koop, Steel and Osiewalski (1995) and Tsionas (2000).

4 Empirical Results

4.1 Posterior Inference on the Parameters of the Model

Table 1 contains posterior medians and the upper and lower bounds of a 95% highest posterior density interval (HPDI) for most of the parameters. Since the parameters themselves are of little interest, we do not discuss this table in any detail. The posterior for q indicates only a slight departure from the linear trade-off case where $q = 1$ (i.e. the case where the production possibility curve in a figure analogous to Figure 1 is a straight line). The α_is indicate that more weight is put on singles, doubles and triples than home runs or walks on a per at bat basis. Note that the relatively low weight on, say, home runs, does not necessarily penalize home run hitters in terms of their efficiency. For instance, in Figure 1, players (5) and (6) are very efficient despite have relatively low batting averages.

The posterior for λ_0 provides strong evidence for the importance of our use of the Gamma distribution (see equation 8) for the distribution of the z_is. That is, clear departures from the commonly-used exponential distribution are found. We will discuss issues relating to the parameters in s in a sub-section involving the $\eta_{(i,t,j)}$s below.

Table 1: Posterior Medians and 95% HPDIs for Parameters

	Median	Lower 95%	Upper 95%
q	1.266	1.063	1.534
α_1	0.276	0.234	0.316
α_2	0.433	0.412	0.458
α_3	0.181	0.138	0.228
α_4	0.110	0.096	0.120
λ_0	1.653	1.440	2.815
λ_1	13.589	11.423	19.457
h	138.52	125.44	152.33
s_1	20.142	13.646	28.009
s_2	8.551	6.038	12.399
s_3	1.403	1.021	1.893
s_4	3.830	2.668	5.286

For reasons of brevity, we do not present posterior results for each of the 35 elements of β. Only 6 of these are significantly different from zero (in the sense that the 95% HPDI does not include zero). However, a further 14 elements are possibly important in that they have 70% HPDIs which exclude zero. Furthermore, extensive experimentation suggests that posterior inference on the efficiencies is not greatly affected by the inclusion or exclusion of the explanatory variables which appear insignificant. Accordingly, the results presented in this section simply include all of the explanatory variables. The only strongly negatively significant variables is the intercept, while the 1997 and 1998 year dummies and team dummy for Detroit have 70% HPDIs which are entirely negative. Hence, we have some slight evidence that 1997 and 1998 were pitchers' years. The strongly positively significant variables all reflect teams. Since Baltimore is the omitted team dummy, these explanatory variables indicate Boston, Cleveland, New York Yankees, Colorado and Houston either play in hitters' parks or have hitters' teams relative to Baltimore (either of which would be an advantage for players on these teams). This list includes some notorious hitters' parks and, unsurprisingly, the Colorado dummy has the highest point estimate. Many of the other team dummies have 70% HPDIs which are entirely positive. The league dummy is not significant, but if all the team dummies are omitted it becomes significant.

In the preceding sections, we have argued that our model is a flexible and reasonable one. However, some readers may be interested in formal measures of model fit. Even for the single output case, there is no commonly recognized measure of fit used with this class of models. That is, the total error or deviation from the frontier in this model is given by:

$$u_{(i,t)} = \varepsilon_{(i,t)} - z_i$$

which does not have mean zero and, hence, a measure such as R^2 would be hard to interpret. Instead we measure the relative contributions of the two components to the total squared error. In particular, note that

$$E\left(u_{(i,t)}^2 | h, \lambda_0, \lambda_1\right) = h^{-1} + \frac{(\lambda_0 + \lambda_0^2)}{\lambda_1^2}$$

can be broken into a term reflecting measurement error (h^{-1}) and one reflecting efficiency. The expected value of the share due to measurement error is 0.22. This implies that 78% of the squared deviation from the frontier is being picked up by the (time-invariant) efficiency component. We interpret this as saying that the model is fitting well in that the role of measurement error is quite small.

4.2 What do Point Estimates of Player Efficiencies Tell Us?

Space constraints prevent us from presenting efficiency results for each of the 535 players. Accordingly, Figure 2 contains a histogram of the posterior mean efficiency for all players. Table 2 presents posterior means and standard deviations for the 15 players with the highest posterior mean efficiencies. In addition, Table 3 presents results for players with posterior means of efficiencies which are at the minimum, first quartile, median and third quartile of the efficiency distribution, respectively. In order to provide a comparison with other common measures of batting performance the tables also include batting average and number of home runs per at bat.

Figure 2 indicates a skewed efficiency distribution with an interior mode, but one with relatively low dispersion (i.e. the efficiency estimates run from 0.648 through 0.987 with most of the players clustered in the interval [0.8,1.0]. This is unsurprising given that a batter who hits safely one time in four risks demotion to the minor leagues whereas a batter who hits safely one time in three is close to winning a batting title.

****Insert Figure 2 here***

The results in Table 2 are quite reasonable, with many of the top names in baseball being listed. It is also reassuring to see that a variety of types of hitters are ranked highly. The top 15 list contains hitters who are well known as having high on-base percentages (e.g. Tony Gwynn, Wade Boggs, Chuck Knoblauch) as well as power hitters (e.g. Barry Bonds, Frank Thomas, Gary Sheffield) as well as those who combine some power with a high batting average (e.g. Edgar Martinez, Mike Piazza, Bernie Williams). The presence of explanatory variables means that two identical players on different teams can have different efficiencies. The fact that Larry Walker (with batting average of 0.344 over the 1995-1999 period) played in Colorado accounts for his absence from the top 15 list (although with an efficiency of 0.967 his is not far off the list). The fact that Jose Offerman, Jason Kendall and Wade Boggs played at least some of their years on relatively poor hitting teams benefitted them in terms of efficiency performance.

A notable absence from this list is the name of Mark McGwire (a record setting home run hitter). With a posterior mean efficiency of 0.960, he is not far off of our list. However, his exceptionally fine home run performance in 1998 and 1999 was partly counterbalanced by a weaker performance in earlier years due to injury. Sammy Sosa's weaker (relative to 1998 and 1999) years in 1995-1997 and lower batting average implies he is ranked even lower. However, the relatively low ranking of these two famous home run hitters brings up an interesting point. Our methodology involves comparing individual players with the best where the "best" can be established in any year. With regards to home runs, the standard

set in 1998 and 1999 was so exceptional that, relative to it, the 1995-1997 McGwires and Sosas do not look that efficient. Our approach implicitly averages a player's performance over 1995-1999 and, hence, McGwire and Sosa average exceptional scores in 1998 and 1999 with lower efficiencies in 1995-1997 and, thus, rank lower than their exceptional 1998 and 1999's would suggest.

Table 2: Posterior Results for the Top 15 Players

Rank	Name	Efficiency (Mean)	Efficiency (St. Dev.)	Batting Average	Home Runs per at bat
1	E. Martinez	0.987	0.010	0.334	0.052
2	M. Grace	0.984	0.012	0.318	0.025
3	T. Gwynn	0.982	0.013	0.352	0.022
4	G. Sheffield	0.981	0.014	0.298	0.064
5	F. Thomas	0.979	0.015	0.314	0.061
6	J. Kendall	0.977	0.016	0.312	0.018
7	J. Offerman	0.975	0.017	0.300	0.010
8	M. Piazza	0.975	0.017	0.338	0.069
9	R. Alomar	0.974	0.017	0.312	0.033
10	J. Olerud	0.974	0.017	0.304	0.035
11	B. Bonds	0.972	0.019	0.294	0.076
12	B. Larkin	0.972	0.019	0.305	0.034
13	W. Boggs	0.971	0.019	0.303	0.010
14	C. Knoblauch	0.971	0.019	0.303	0.023
15	B. Williams	0.971	0.019	0.324	0.044

The players in Table 3 are less well-known than those in Table 2 (e.g. J. R. Phillips played only in 1995 and even then only in 92 games). Nevertheless, based on an examination of batting average it does seem as though our model is providing with a reasonable ranking of efficiency. The fact that Charlie O'Brien (who is known for his defensive skills) is ranked so poorly emphasizes the fact that the present model takes into account only batting performance.

Table 3: Posterior Results for Players Ranked at Quartiles

Rank	Name	Efficiency (Mean)	Efficiency (St. Dev.)	Batting Average	Home Runs per at bat
Minimum	J. R. Phillips	0.648	0.057	0.195	0.039
1st quartile	C. O'Brien	0.851	0.049	0.230	0.031
Median	J. Bates	0.891	0.056	0.267	0.025
Third quartile	D. Cedeno	0.928	0.039	0.281	0.009

4.3 Are Point Estimates Accurate Enough to Meaningfully Rank Players?

The results in the previous sub-section are all based on point estimates of efficiencies. One advantage of the Bayesian approach over non-Bayesian alternatives is that it is simple to derive posterior standard deviations for player-specific efficiencies to see if rankings based on point estimates really are statistically significant (Fernandez, Koop and Steel, 2000b, discusses this issue in detail). An examination of Tables 2 and 3 indicate that posterior standard deviations can be fairly large, indicating that caution must be taken in comparing individual players. For instance, the estimated efficiencies of all our top 15 players are within one standard deviation of one another. A naive interpretation of this finding is that our top 15 ranking is pretty meaningless. However, correlations between the efficiencies of different players implies that posterior standard deviations may not tell the full story. In order to shed additional light on this issue, various posterior probabilities which address questions like "what is the probability that player A is more efficient than player B?" can be calculated. Such probabilities can be obtained by simply calculating the proportion of MCMC replications for which player A is more efficient than player B. Of course, there are a myriad of possible comparisons which could be made. As a general rule, we find that it is not easy to distinguish between players which are ranked very close to each other in terms of point estimates, but it is possible to distinguish between players farther apart in the ranking. . The following selection of comparisons illustrates these points.

A first finding is that there is a 0.021 probability that Mark McGwire (ranked 34th

with an estimated efficiency of 0.960) should be on our top 15 list (i.e. in roughly 2% of the MCMC draws he falls in the top 15 in terms of efficiency). Hence, even though McGwire's efficiency estimate is within one standard deviation of the those of the top 15 players, we can reliably say he is less efficient that they are. A justification for his relatively poor performance is given in the previous subsection. A second finding is that there is a 0.086 probability that Edgard Martinez truly is the most efficient player (i.e. in over 90% of MCMC replications one of the other players is most efficient). Tables 4 and 5 offer more systematic comparisons. Table 4 takes the players ranked 1st, 5th, 10th and 15th and compares them. In particular, the element in the i^{th} column and j^{th} row in Table 4 is the probability that player i is more efficient than player j. Table 5 is of the same form, but takes the 5 quartile players (i.e. minimum, first quartile, median, third quartile and maximum efficient players).

Table 4: Probability that Player in Column is More Efficient than Player in Row

	E. Martinez	F. Thomas	J. Olerud	B. Williams
E. Martinez				
F. Thomas	0.671			
J. Olerud	0.705	0.546		
B. Williams	0.757	0.600	0.563	

Table 5: Probability that Player in Column is More Efficient than Player in Row

	E. Martinez	D. Cedeno	J. Bates	C. O'Brien	J. R. Phillips
E. Martinez					
D. Cedeno	0.962				
J. Bates	0.976	0.724			
C. O'Brien	0.997	0.833	0.808		
J. R. Phillips	1.000	0.999	0.997	0.993	

We find these results to be moderately encouraging. Of course, there is a fair bit of statistical uncertainty associated with estimating the large number of parameters in the model. However, Table 5 indicates that our approach can reliably say that players ranked in different quartiles from one another can be said to be different in their efficiencies. Note

that the lowest figures in this table are associated with J. Bates, a player with an unusually high posterior standard deviation. In Table 4, we are comparing players ranked roughly one percentile apart and it is unsurprising that results are not as strong. Nevertheless, there is a good chance that our top 15 rankings are at least roughly sensible.

4.4 What do Point Estimates of the $\eta_{(i,t,j)}$s Tell us?

Player-specific efficiencies are the focus of the present study. Nevertheless, it is worthwhile to briefly digress to discuss the $\eta_{(i,t,j)}$s as these are an innovative part of the present model. Remember that they can be interpreted as being the share of the aggregate output attributed to output j for player i in time t (see equation 10). Figure 3 calculates histograms of all the $\eta_{(i,t,j)}$s. Based on the modes of these histograms, we can infer that, for a typical player, the output singles receives roughly 54% of the weight in the aggregate output, doubles and triples receive roughly 28%, home runs roughly 4% and walks roughly 12%. The weights reflect both the shape of the production possibility curve (i.e. α and q, which are estimated from the data) as well as the actual value of the output in each category (i.e. $y_1, .., y_4$).

Figure 3 illustrates a relatively high variability across players. For instance, the contribution of singles to aggregate output varies from 0.45 to 0.70. The contribution of doubles and triples varies from 0.16 to over 0.36, etc. In other words, there are many different types of players with very differing performance in terms of each of the individual outputs. This indicates that any ranking of players based on a single output could be very misleading and emphasizes the value of finding a statistically justifiable way of aggregating outputs.

Insert Figure 3 here

The shape of the histograms in Figure 3 are, with one exception, roughly normal. The exception is home runs, which is skewed and has a few interesting bumps in the right tail.

The bins furthest to the right are for Mark McGwire and Sammy Sosa in the years 1998 and 1999.

4.5 The Relationship Between Efficiency and Salaries

The chief purpose of the present paper relates to the issue of player performance. The related issue of how closely player performance relates to salary is an important one. However, there are so many complications (e.g. the fact that players become free agents only after a certain number of years) which imply that modelling salaries in baseball is a difficult task which would lead us far from the main focus of the paper (see, e.g., Scully, 1974, Depken, 2000, Hoaglin and Velleman, 1995). In addition, the efficiency estimates only include batting performance and, hence, are an incomplete measure of a player's worth to his team. Nevertheless, for the reader interested in some basic information on the relationship between salary and performance, we have collected data on the salary (including any bonus) of 478 of our 535 players. The salary data is taken from Sean Lahman's Baseball Archive (www.baseball1.com) but is only available through 1998 and has some missing players which accounts for the fact that only 478 players were available. Furthermore, given free agency and the time it takes for a player to develop, it is plausible that the most recent salary is the best signal of a player's performance. Accordingly, the salary data is for the most recent year available (usually 1998, but occasionally for earlier years for players who retired).

The mean, median, standard deviation, minimum and maximum of the salaries are \$1,654,065, \$700,000, \$1,981,836, \$100,000 and \$10,000,000, respectively. The simple correlation between salary and efficiency is only 0.29. Figure 4 contains an XY plot of salary and efficiency, and the reason for the low correlation between salary and performance is readily apparent. There are a large number of very efficient players who are paid very poorly. Most of these are young players yet to test the free agent market. For instance, Derek Jeter has an efficiency estimate of 0.963 but in 1998 was only paid \$750,000. Jason

Giambi, with efficiency of 0.961, was only paid $315,000 in 1998.

Insert Figure 4 here

5 Conclusion

In this paper we have developed statistical methods for comparing the performance of individual baseball players which explicitly take into account the many outputs players produce. A key component in our method is an output aggregator which is estimated based on the performance of the best players in the various output dimensions. Individual players are then compared to the benchmark set by the best. The result is a number between zero and one which has a simple, intuitive, interpretation. If a player's efficiency estimate is 0.75, then we can say he is producing 75% of the output of the best comparable players, where the output is measured using the output aggregator. Such a number, we argue, has advantages over traditional measures of batting performance in that it is simple to interpret and does not require the choice of ad hoc weights to create an aggregate output. Bayesian methods for statistical inference were developed and used to estimate the efficiencies of 535 players in the years 1995-1999.

Given the great, and increasing, importance of sport in our economic and social life developing such techniques is of direct interest. However, there is also an indirect sense in which developing statistical techniques in a sports-related context is of interest. In many social and biological sciences, cases exist where the researcher is interested in evaluating the performance of an individual against a standard set by others. In such cases, models similar to the one used in this paper are potentially of interest. Many researchers find it easier to conceptualize statistical modelling issues in comfortable surroundings such as that provided by baseball. Accordingly, a second purpose of this paper is to discuss the issues involved in multiple output production modelling in what we hope is a familiar context, in order to popularize them in other fields.

6 References

Albright, S.C. (1993), "A Statistical Analysis of Hitting Streaks in Baseball (with discussion and rejoinder)," *Journal of the American Statistical Association*, 88, 1175-1196).

Barry, D. and Hartigan, J.A. (1993), "Choice Models for Predicting Divisional Winners in Baseball," *Journal of the American Statistical Association*, 88, 766-774.

Berry, S., Reese, C.S. and Larkey, P. (1999), "Bridging Different Eras in Sport (with discussion and rejoinder)," *Journal of the American Statistical Association*, 94, 661-686.

van den Broeck, J., Koop, G., Osiewalski, J. and Steel, M.F.J. (1994), "Stochastic Frontier Models: A Bayesian Perspective," *Journal of Econometrics*, 61, 273-303.

Chapman, K. (1991), "Testing the Matching Hypothesis: The Case of Major League Baseball," *American Economic Review*, 81, 1352-1360.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, 49, 327-335.

Depken II, C.A. (2000), "Wage Disparity and Team Productivity: Evidence from Major League Baseball," *Economics Letters*, 67, 87-92.

Ferrall, C. and Smith, A. (1999), "A Sequential Game Model of Sports Championship Series: Theory and Estimation," *Review of Economics and Statistics*, LXXXI, 704-719.

Fernandez, C., Koop, G. and Steel, M.F.J. (2000), "A Bayesian Analysis of Multiple-Output Production Functions," *Journal of Econometrics*, 98, 47-79.

Fernandez, C., Koop, G. and Steel, M.F.J. (2001a), "Multiple Output Production with Undesirable Outputs: An Application to Nitrogen Surplus in Agriculture," manuscript available at www.gla.ac.uk/departments/economics/.

Fernandez, C., Koop, G. and Steel, M.F.J. (2001b), "Alternative Efficiency Measures for Multiple Output Production," manuscript available at www.gla.ac.uk/departments/economics/.

Fernandez, C., Osiewalski, J., and Steel, M.F.J. (1997), "On the Use of Panel Data in Stochastic Frontier Models With Improper Priors," *Journal of Econometrics*, 79, 169-193.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith (eds.), *Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics*, 169-194. Oxford: Oxford University Press.

Hoaglin, D. and Velleman, P. (1995), "A Critical Look at Some Analyses of Major League Baseball Salaries," *The American Statistician*, 49, 277-284.

Kahn, L. (1993), "Free Agency, Long-Term Contracts and Compensation in Major League Baseball: Estimates from Panel Data," *Review of Economics and Statistics*, 75, 157-164.

Kaigh, W. (1995), "Forecasting Baseball Games," *Chance*, 8, 33-37.

Koop, G., Osiewalski, J., and Steel, M.F.J. (1997), "Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers," *Journal of Econometrics*, 76, 77-105.

Koop, G., Steel,. M.F.J. and Osiewalski, J. (1995), "Posterior Analysis of Stochastic Frontier Models using Gibbs Sampling," *Computational Statistics*, 10, 353-373.

Lackritz, J. (1996), "Two of Baseball's Great Marks: Can They Ever Be Broken?" *Chance*, 9, 12-18.

Reinhard, S., Lovell, C.A.K., and Thijssen, G. (1999), "Econometric Application of Technical and Environmental Efficiency: An Application to Dutch Dairy Farms," *American Journal of Agricultural Economics*, 81, 44-60.

Schall, T. and Smith, G. (2000), "Do Baseball Players Regress Towards the Mean?" *The American Statistician*, 54, 231-235.

Schmidt, P. and Sickles, R. (1984), "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, 2, 367-374.

Scully, G. (1974), "Pay and Performance in Major League Baseball," *American Economic Review*, 64, 915-930.

Tsionas, E. (2000), "Full Likelihood Inference in Normal-Gamma Stochastic Frontier Models," *Journal of Productivity Analysis*, 13, 183-205.

Watnik, M. (1998), "Pay for Play: Are Baseball Salaries Based on Performance?" *Journal of Statistics Education*, 6, 1-5.

7 Appendix A: The Prior for $(\beta, \lambda_0, \lambda_1, h, \alpha, q, s)$

We use a prior of the form:

$$p(\beta, \lambda_0, \lambda_1, h, \alpha, q, s) = p(\beta)p(\lambda_0)p(\lambda_1)p(h)p(\alpha)p(q)p(s).$$

For β, we assume:

$$p(\beta) \propto f_N^k(\beta|b_0, H_0^{-1}),$$

and make the noninformative choices of $b_0 = 0_k$ and $H_0 = 10^{-4} \times I_k$. Next we assume

$$p(\lambda_0) = f_G(\lambda_0|\underline{\lambda}_{00}, \underline{\lambda}_{01}),$$

with positive prior hyperparameters $\underline{\lambda}_{00}, \underline{\lambda}_{01}$. We choose these so that E(λ_0)=3 and var(λ_0)=1 and hence $\underline{\lambda}_{00} = 9, \underline{\lambda}_{01} = 3$. In other words, we are centering the prior over a Gamma distribution with an interior mode. Furthermore,

$$p(\lambda_1) = f_G(\lambda_1|\underline{\lambda}_{10}, \underline{\lambda}_{11}),$$

with positive prior hyperparameters $\underline{\lambda}_{10}, \underline{\lambda}_{11}$ which we choose as $\underline{\lambda}_{10} = 3$ and $\underline{\lambda}_{11} = -\ln(\tau^*)$. If we had $\lambda_0 = 3$ (i.e. its prior mean) then these values would imply that the prior median efficiency is τ^*, which is a natural quantity to elicit. We set $\tau^* = 0.70$. The reader is referred to van den Broeck, Koop, Osiewalski and Steel (1994) for a further discussion of prior elicitation in this class of models and justification of such choices as being reasonable, but noninformative relative to the data.

For h we take:

$$p(h) = f_G(h|n_0/2, a_0/2),$$

and set $n_0/2 = 1$ (which leads to an Exponential prior for h) and $a_0/2 = 10^{-6}$. These values imply large prior uncertainty.

Since the components of α are all in the interval $(0, 1)$ and sum up to one, an obvious choice of a prior distribution is a Dirichlet with p.d.f.

$$p(\alpha) = f_D^{p-1}(\alpha|a),$$

with hyperparameter vector $a = (a_1, \ldots, a_p)'$. We use the diffuse choice of $a = \iota_p$, which makes the prior Uniform over the $(p-1)$-dimensional unit simplex.

For q we take an Exponential prior,

$$p(q) \propto f_G(q|1, d).$$

We adopt the diffuse choice of $d = 10^{-6}$.

We assume p independent Gamma distributions for the components of s:

$$p(s) = \prod_{j=1}^{p} p(s_j) = \prod_{j=1}^{p} f_G(s_j|b_j, c_j).$$

We make a diffuse choice of $b_j = 1$ and $c_j = 10^{-6}$ for all j.

8 Appendix B: MCMC Algorithm

Our MCMC algorithm involves sequentially drawing from the posterior conditionals described in the body of the text. The first of these is:

$$p(z|Y,\beta, \lambda_0, \lambda_1, h, \alpha, q, s) \propto \prod_{i=1}^{N} z_i^{\lambda_0 - 1} f_N^1(z_i|m_i, R_i) I(z_i \geq 0),$$

where m_i is the i^{th} element of the $N-$vector

$$m = (D'D)^{-1}\{D'(X\beta - \log(\theta)) - h^{-1}\lambda_1\},$$

and R_i is the i^{th} diagonal element of the $N \times N$ diagonal matrix

$$R = h^{-1}(D'D)^{-1}.$$

Since the elements of z are conditionally independent they can be drawn separately. In practice, we use a modified version of the acceptance algorithm of Koop, Steel and Osiewalski (1995) to draw from this non-standard density. We use $f_G(z_i|\lambda_0, \lambda_1)$ to take candidate draws, z_i^{can}, which are then are accepted with probability:

$$\frac{\exp\left\{-\frac{h}{2}\sum_{t=1}^{T_i}\left(\log\left(\theta_{(it)}\right) - x_{it}\beta + z_i\right)^2\right\}}{\exp\left\{-\frac{h}{2}\sum_{t=1}^{T_i}\left(\log\left(\theta_{(it)}\right) - x_{it}\beta + z^{opt}\right)^2\right\}}$$

where $z^{opt} = \max\left\{-\frac{\sum\left(\log(\theta)_{it} - x_{it}\beta\right)}{T_i}, 0\right\}$ and $x_{it} = \left(x_{(i,t,1)}, x_{(i,t,2)}, .., x_{(i,t,k)}\right)'$.

The p.d.f. of the posterior conditional distribution of λ_0 is given by:

$$p(\lambda_0|Y, z, \beta, h, \alpha, q, s, \lambda_1) \propto \lambda_1^{N\lambda_0}\lambda_0^{\underline{\lambda}_{00}-1}\Gamma(\lambda_0)^{-N}\exp\left[\left(\sum_{i=1}^{N}\ln(z_i) - \underline{\lambda}_{01}\right)\lambda_0\right].$$

This distribution is non-standard, but a simple acceptance algorithm can be derived following Tsionas (2000). Candidate draws, λ_0^{can}, are taken from $f_G(\lambda_0|1, A)$ where A is chosen to maximize the acceptance probability. To obtain the upper bound which appears in the acceptance algorithm we must find λ_0^* which solves the following equation:

$$B + \underline{\lambda}_{00}\lambda_0^{*-1} - N\Psi(\lambda_0^*) = 0$$

where $\Psi(.)$ is the digamma function (i.e. $\Psi(x) \equiv \frac{d\ln\Gamma(x)}{dx}$ where $\Gamma(.)$ is the Gamma function) and $B = N\ln(\lambda_1) + \sum_{i=1}^{N}\ln(z_i) - \underline{\lambda}_{01}$. We solve this equation for λ_0^* using a simple one-dimensional grid search. The optimal value of A is then $A^* = \lambda_0^{*-1}$.Candidate draws from $f_G(\lambda_0|1, A^*)$ are then accepted with probability

$$\frac{R(\lambda_0^{can})}{R(\lambda_0^*)}$$

where

$$\ln(R(\lambda_0)) = (B + A^*)\,\lambda_0 - \ln(A^*) - N\ln(\Gamma(\lambda_0)) + (\underline{\lambda}_{00} - 1)\ln(\lambda_0).$$

The conditional posterior distribution of λ_1 is:

$$p(\lambda_1|Y, z, \beta, h, \alpha, q, s, \lambda_0) = f_G\left(\lambda_1|\underline{\lambda}_{10} + N\lambda_0, \underline{\lambda}_{11} + \sum_{i=1}^{N} z_i\right).$$

The conditional posterior distribution of β is:

$$p(\beta|Y, z, \lambda_0, \lambda_1, h, \alpha, q, s) = f_N^k(\beta|b_*, H_*^{-1})$$

where

$$H_* = H_0 + hX'X,$$

and

$$b_* = H_*^{-1}\{H_0 b_0 + hX'(\log(\theta) + Dz)\}.$$

For h we obtain:

$$p(h|Y, z, \lambda_0, \lambda_1, \beta, \alpha, q, s) = f_G\left(h|\frac{n_0 + NT}{2}, \frac{a_0 + (\log(\theta) - X\beta + Dz)'(\log(\theta) - X\beta + Dz)}{2}\right).$$

For α we obtain the following non-standard p.d.f.:

$$\begin{aligned} p(\alpha|Y, z, \lambda_0, \lambda_1, \beta, h, q, s) \quad \propto \quad & \\ & \prod_j \alpha_j^{a_j + s_j qNT - 1} \prod_{i,t}\left(\sum_j \alpha_j^q y_{(i,t,j)}^q\right)^{-\sum_j s_j} \\ & \exp\left\{-\frac{1}{2\sigma^2}(\log(\theta) - X\beta + Dz)'(\log(\theta) - X\beta + Dz.)\right\} \end{aligned}$$

Note that $\log(\theta)$ depends on α. Based on the experience of Fernandez, Koop and Steel (2000, 2001a,b), we use a random walk Metropolis-Hastings algorithm which seems to work very well. In particular, we use a $(p-1)$-variate Normal candidate-generating density (proposals outside the support of α are never accepted). The mean of the Normal is given by the previous draw and the variance is calibrated in preliminary MCMC runs so that the acceptance probability is in the region of $5-50\%$. Chib and Greenberg (1995) provide a simple discussion of various Metropolis-Hastings algorithms and discuss the issue of what reasonable acceptance rates are. The acceptance probability is merely the ratio of $p(\alpha|Y,z,\lambda_0,\lambda_1,\beta,h,q,s)$ evaluated at the candidate draw to this density evaluated at the last accepted draw.

The conditional posterior for q is also non-standard and is given by:

$$p(q|Y,z,\beta,\lambda_0,\lambda_1,h,\alpha,s) \propto q^{NT(p-1)}\exp(-dq)\exp\{-A(q)\}$$

where

$$A(q)=\frac{h}{2}(\log(\theta)-X\beta+Dz)'(\log(\theta)-X\beta+Dz)+\sum_{i,t,j}s_j\log\left(\frac{\sum_{l=1}^{p}\alpha_l^q y_{(i,t,l)}^q}{\alpha_j^q y_{(i,t,j)}^q}\right).$$

In this case, too, we use a random walk Metropolis-Hastings algorithm with a Normal candidate-generating density calibrated as above.

We draw separately each of the p components of s. The conditional posterior distribution of s_j, $j=1,\ldots,p$ (also given the remaining components of s) has p.d.f. on $(0,\infty)$:

$$p(s_j|Y,z,\lambda_0,\lambda_1,\beta,h,\alpha,q,\{s_h \;:\; h\neq j\}) \propto \frac{\Gamma(\sum_l s_l)^{NT}}{\Gamma(s_j)^{NT}}s_j^{b_j-1}\exp\left[-s_j\left\{c_j+\sum_{i,t}\log\left(\frac{\sum_l \alpha_l^q y_{(i,t,l)}^q}{\alpha_j^q y_{(i,t,j)}^q}\right)\right\}\right].$$

In order to draw from this distribution, a random walk Metropolis-Hasting algorithm is used in the same manner as for q or α.

The results of Fernandez, Osiewalski and Steel (1997) and Fernandez, Koop and Steel (2000) imply that the posterior is proper and the MCMC algorithm will converge. However, we have yet to find a proof that posterior means and higher moments exist for parameters other than the efficiencies (although numerical evidence suggests that higher moments of all parameters do exist). Hence, except for the efficiencies, we present posterior medians and HPDIs (which do exist) instead of posterior means and standard deviations.

Our final results are based on 260,000 MCMC replications. The first 10,000 are discarded to mitigate startup effects. Of the remaining 250,000, we retain every 10th replication to break serial correlation in our draws (and minimize storage costs). The posterior properties reported are, thus, based on 25,000 replications.

MCMC convergence is monitored both formally and informally. Informally, several experimental runs with dispersed starting values provided posterior results which were essentially identical. Formally, we used a diagnostic suggested by Geweke (1992) which compares the means based on first 10% and final 50% of the draws relative to their numerical standard errors. If the sampler has converged, this statistic is standard Normal. We calculated this diagnostic for the efficiencies of the first five players and obtained statistics which were all less than 1.0, indicating convergence.

Figure 1

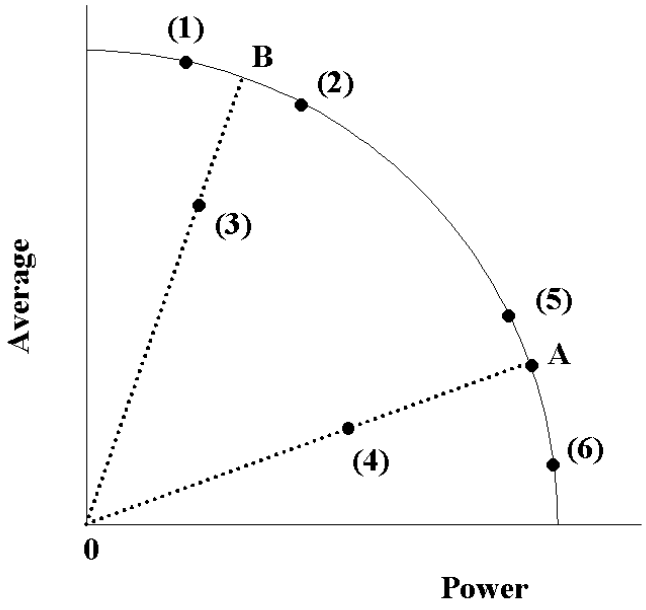

Figure 1:

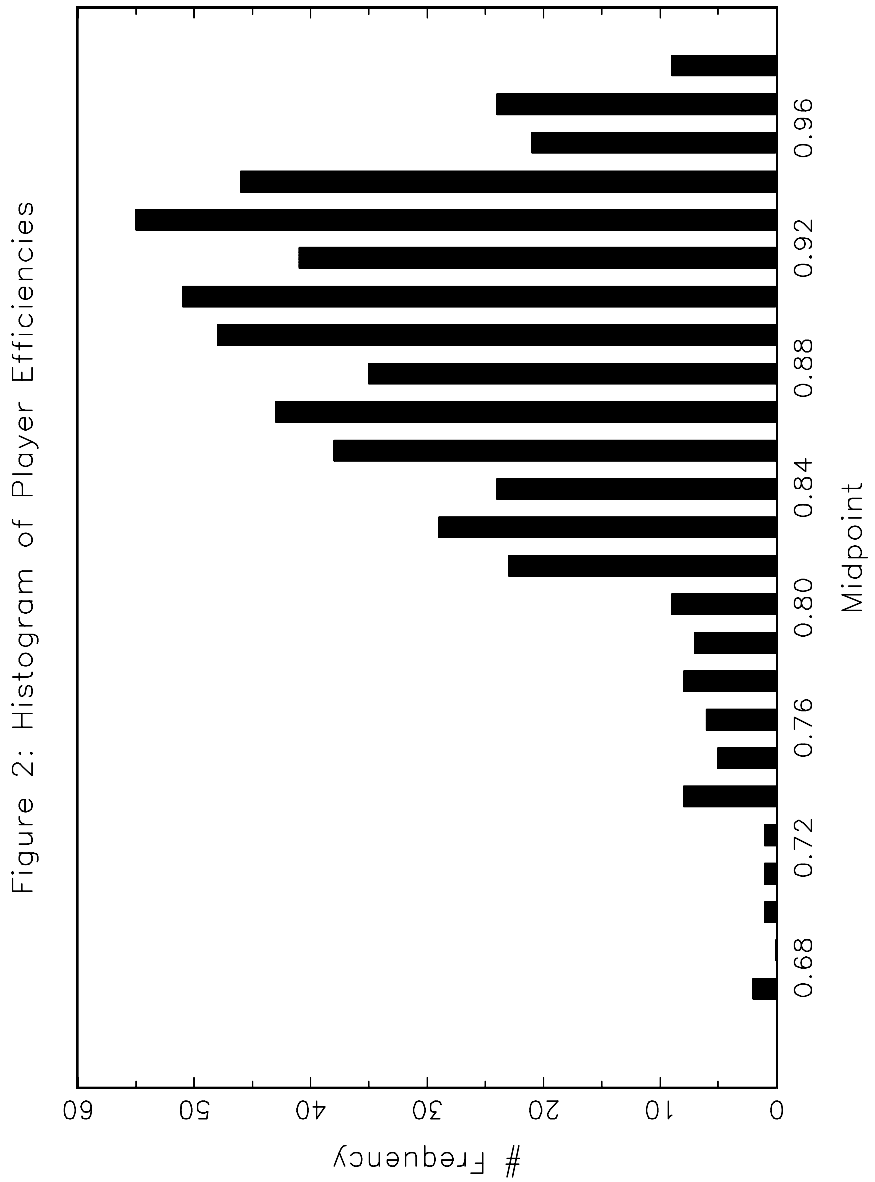
Figure 2: Histogram of Player Efficiencies
Frequency
0
10
20
30
40
50
60
0.68
0.72
0.76
0.80
0.84
0.88
0.92
0.96
Midpoint

Figure 2:

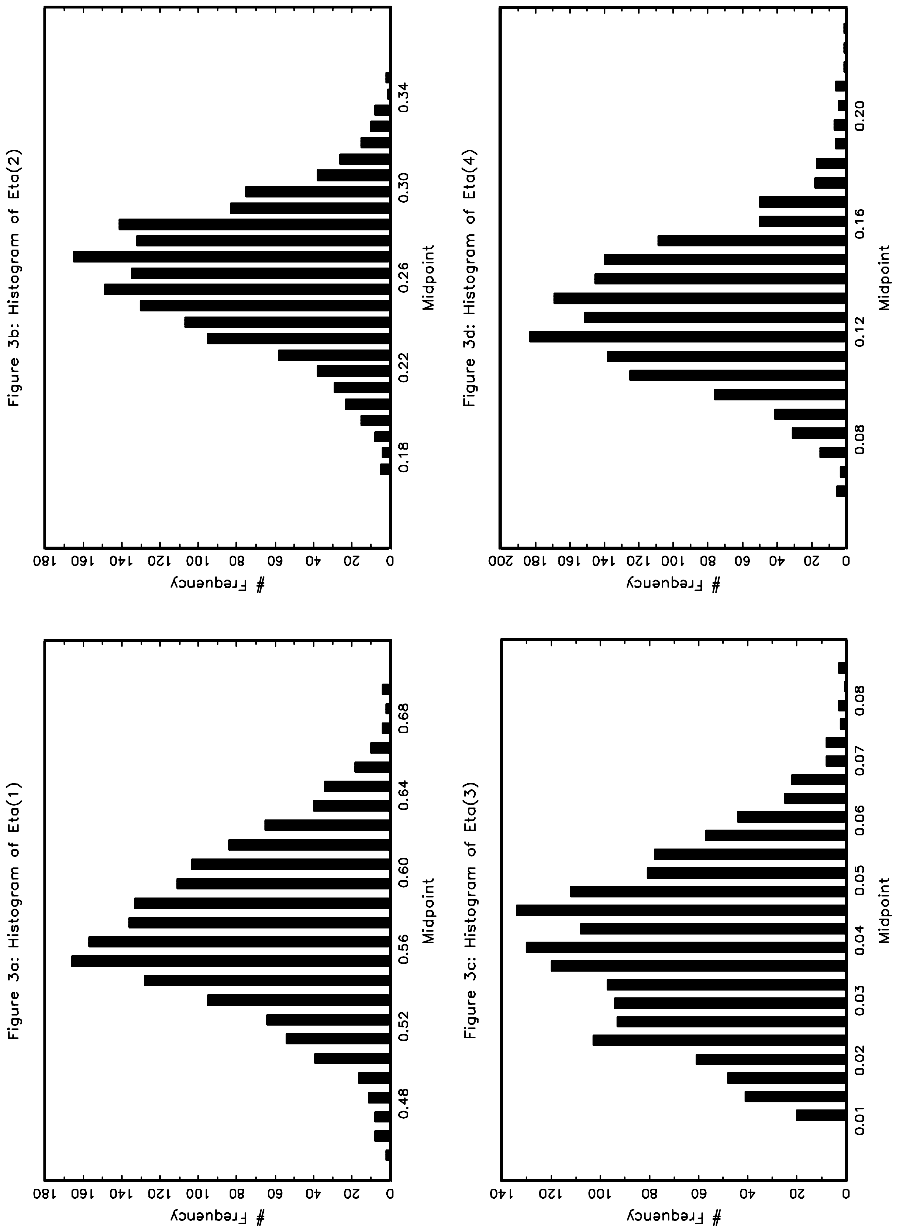
Figure 3a: Histogram of Eta(1)
Frequency
Midpoint
0.48
0.52
0.56
0.60
0.64
0.68
Figure 3b: Histogram of Eta(2)
Frequency
Midpoint
0.18
0.22
0.26
0.30
0.34
Figure 3c: Histogram of Eta(3)
Frequency
Midpoint
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Figure 3d: Histogram of Eta(4)
Frequency
Midpoint
0.08
0.12
0.16
0.20

Figure 3:

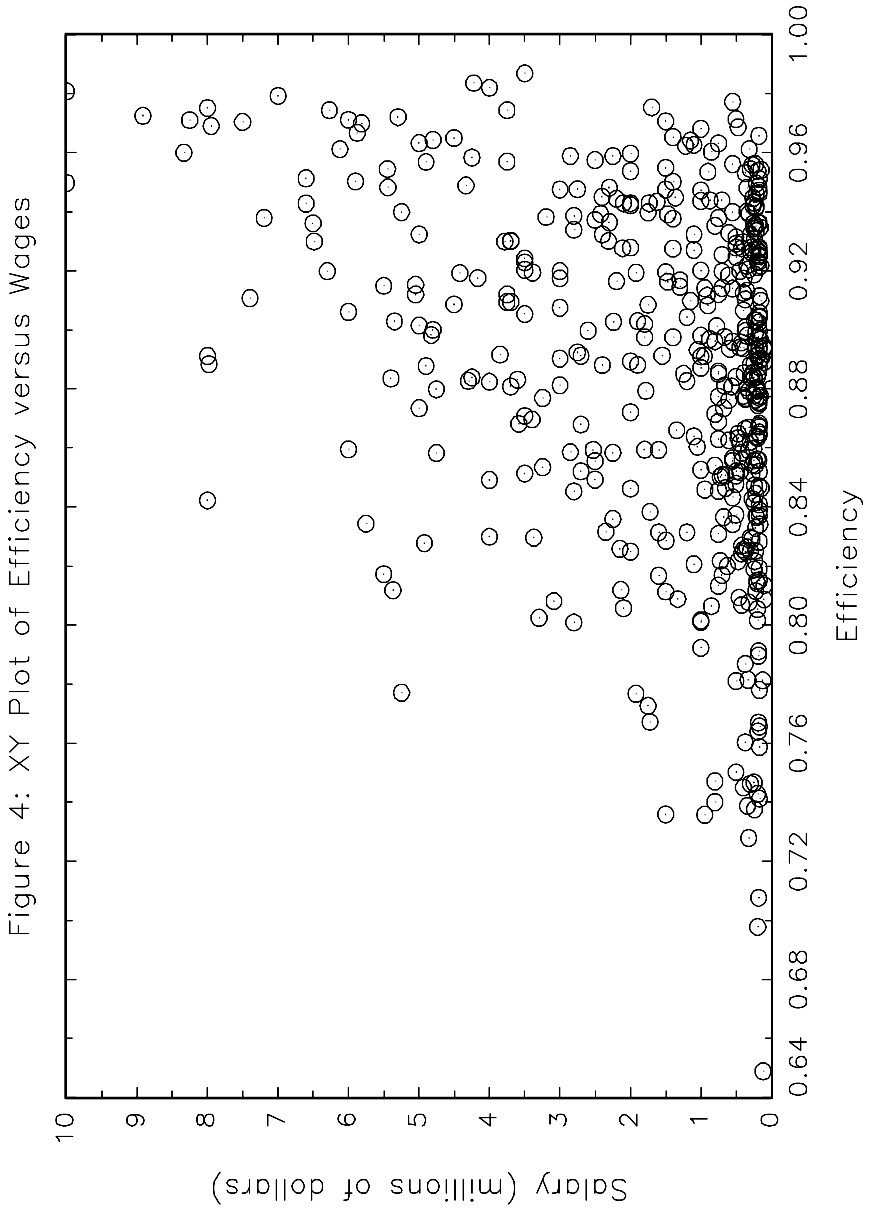
Figure 4: XY Plot of Efficiency versus Wages
Salary (millions of dollars)
Efficiency
0
1
2
3
4
5
6
7
8
9
10
0.64
0.68
0.72
0.76
0.80
0.84
0.88
0.92
0.96
1.00

Figure 4: