

Measurement error and the hot hand

Daniel F. Stone*

Oregon State University

February 2012

Abstract

This paper shows the first autocorrelation of basketball shot results is a highly biased and inconsistent estimator of the first autocorrelation of the *ex ante* probabilities the shots are made. Shot result autocorrelation is close to zero even when shot probability autocorrelation is close to one. The bias is caused by what is equivalent to a severe measurement error problem. The results imply that the widespread belief among players and fans in the hot hand is not necessarily a cognitive fallacy.

*Email: dan.stone@oregonstate.edu. I first thank Zheng Cao, whose work inspired this paper. I give special thanks to Bruce McGough and Jon Faust, and also thank Jeremy Arkes, Todd Pugatch, Joe Price, Frank McIntyre, Liz Schroeder and Kevin Thom for helpful comments.

1 Introduction

One of the most well known and counter-intuitive results from research on sports statistics is (the claim) that—in contrast to the almost universally held popular belief—there is no “hot hand” in basketball. Gilovich et al. (1985) is the seminal paper on this topic. The authors wrote “[the term the hot hand implies] the probability of a hit should be greater following a hit than following a miss (i.e., positive association),” and presented strong evidence that both fans and players believe this is generally true. However, the authors found almost no evidence of this positive association, or autocorrelation, of shot results occurring in NBA field goal data. While this could be explained by players taking more difficult shots or being defended more closely when “hot” (as discussed by, e.g., Dorsey-Palmateer and Smith (2004)), Gilovich et al. also found free throws and shots taken in a controlled experiment at a constant level of difficulty, exhibit the same lack of autocorrelation. Since then these results have been corroborated (see, e.g., Koehler and Conley (2003) and Cao (2011)) and withstood numerous criticisms (e.g., Wardrop (1995)). The recent best-selling book *Nudge* summarizes what appears to be the new conventional wisdom on the topic, at least for those who know this literature: “It turns out that the ‘hot hand’ [in basketball] is just a myth... To date, no one has found it,” (Thaler and Sunstein (2008), p.30). Given this lack of statistical evidence for the hot hand, belief in the hot hand is consequently often cited in both the academic literature and popular press as a cognitive error.¹ Very recently Arkes (2010) found evidence that NBA free throws do exhibit positive serial correlation, but on average only around 3%.²

¹In addition to the psychology literature, false belief in the hot hand is also regularly cited in the behavioral economics literature (e.g., Rabin (1998)). See Lehrer (2009) for another example from the popular press, in addition to Thaler and Sunstein (2008).

²Arkes’ results differed from those of previous literature primarily because, first, he used a relatively large dataset (data on all free throws from the 2005-06 NBA season), and second, he pooled the analysis across players, using a fixed effects approach to account for heterogeneity in average ability. Dorsey-Palmateer and Smith (2004) is one of numerous papers that finds evidence of a hot hand existing in a context other than basketball. See Reifman (2011) for a general discussion of the hot hand in sports; other literature is discussed further

In this paper I present an alternative view of the hot hand. I argue this view implies the widespread belief in the hot hand is not necessarily a cognitive fallacy—that the results of both Gilovich et al and Arkes are consistent with the hot hand not only existing, but being of a large magnitude. The alternative definition of the hot hand is that it exists if the *probability* of a shot being made is positively correlated with the probability for the next shot. Similarly, the magnitude of the hot hand phenomenon can be defined based on the magnitude of this correlation of shot probabilities, as opposed to shot results. Although these probabilities are unobserved, they describe shooting ability more accurately than shot results, since probabilities characterize the data-generating process, while shot results are just realizations of data. Thus, this alternative definition is arguably preferable to the definition of Gilovich et al. And whether or not it is preferable, it seems clear the alternative definition is at least reasonable and of interest—that if the alternative definition is satisfied then the hot hand, in a very real sense, does exist. For example, suppose a player who shoots 50% on average has a 70% probability of making a particular shot (because, e.g., he feels extra confident). And suppose this implied that the player’s expected probabilities of making his next three shots were, say, 66%, 63%, 60%—all well above his overall average. It seems clear then that this player would be, in expectation, experiencing the hot hand.^{3,4}

below.

³If the probabilities were not serially correlated, but still stochastic, a player would have higher probabilities for some shots than others and thus, in a sense, be hotter at some times than others. For example, the probability could be 70% on one shot, and so the player would be hot for that shot, but still be expected to be the player’s overall mean of 50% on the next shot. But this is not what people mean when they refer to the hot hand: they are referring to above average ability that is at least somewhat persistent, i.e., that likely lasts for at least two consecutive shots. Otherwise, it would be completely irrelevant for behavior: there would be no point in passing to someone hot if the hot state immediately fully dissipates. The original definition of Gilovich et al was based on persistence as well. I also note a hot player will not actually stay hot with certainty if the probabilities are stochastic, but think it is clear that fans, players etc understand this.

⁴It is also worth noting that, if this alternative definition were to be satisfied, players would sometimes experience the cold hand as well: if their shot probability was, e.g., 30% for one shot, it would be expected to be less than 50% on subsequent shots. This is not at all problematic; most fans and players seem to believe in the cold hand’s existence as well, and this phenomenon would also occur if the definition of Gilovich et al. was satisfied.

The alternative hot hand definition implies analysis of shot result data is subject to what is equivalent to a measurement error problem, as shot results are essentially noisy measures of the *ex ante* probabilities with which each shot is made. To illustrate using the example above, the player could easily miss two or more of the four shots—there is actually a 44% chance of this occurring.⁵ Shot result data would then indicate he was not hot, cold or even very cold, when in fact he was hot. That is, shot result data appear to often mismeasure a player’s hot/cold state.

Section 2 develops a simple AR(1) model of shot probabilities to formally develop these ideas. The model has just two parameters: the unconditional mean and the autocorrelation, ρ . The shock is white noise with a variance equal to a function of these parameters. The alternative hot hand definition proposed above implies $\rho > 0$ would mean the hot hand exists, and is of a larger magnitude (which I call “stronger”) when ρ is larger (and of a smaller magnitude, or “weaker,” when ρ is closer to zero). The especially interesting question is how strong the hot hand is, since if it exists, but is very weak (say, $\rho < 0.05$), this would arguably be practically very similar, with respect to the implications for both psychology and optimal basketball behavior, to the case of the hot hand not existing at all.⁶ I then use the model to formally show that the measurement error problem not only exists, but is inherently severe in the basketball context, causing analysis of shot data to vastly under-estimate the first autocorrelation of shot probabilities in unboundedly large samples. If the mean shot probability is 0.5 and $\rho = 0.4$, the probability limit of estimated shot autocorrelation is 0.057; when ρ is larger, estimated shot autocorrelation is actually even lower. Shot autocorrelation is

⁵This assumes the probabilities are equal to their expected values (70%, 66%, 63% and 60%).

⁶Another necessary condition for the hot hand phenomenon to be of a large magnitude—for a player to go through periods in which his ability is substantially better than normal—is that the variance of the shock must be sufficiently large. In the model of Section 2 this variance shrinks as ρ increases. But the variance is still reasonably large so long as ρ is not too close to one. For example, the shock is distributed $U[-0.1, 0.1]$ if $\rho = 0.8$. Moreover, this relation between the shock’s variance and ρ is relaxed in Section 3.

also even lower for other values of the mean shot probability. I also conduct a simple simulation analysis, presented in Section 3, to analyze a version of the model with heteroscedastic shocks. The results are not as extreme as those found in Section 2, but still extreme. For example, average estimated shot autocorrelation from the simulations is less than 0.1 for all ρ up to 0.8, for all parameters considered. For some values of the parameters, average simulated shot autocorrelation is less than 0.04, and is statistically insignificant over 80% of the time, even when $\rho = 0.9$.

To be clear, I am not claiming this AR(1) model is literally the “true model” of shot probabilities. I use this model, first, because it is simple and easily analyzed. Second, it is at least a reasonable approximation to a plausible true model. Loosely speaking, the AR(1) model might be true. Moreover, the paper’s results hold for similar models. Thus, although the fit of the AR(1) model is admittedly not empirically tested, the model is sufficient for making the paper’s main point—that standard analysis of shot result autocorrelation may not tell us much about shot probability autocorrelation. Consequently, if one agrees that probability autocorrelation is an important parameter for understanding the hot hand, then a lack of shot result autocorrelation does not tell us the hot hand is merely a “myth,” or even of a small magnitude. The paper’s results do not imply shot probability autocorrelation cannot possibly be estimated with shot result data, just that the measurement error issue should be addressed to do so and may make this estimation very difficult. This may open up an interesting line of future research; implications are discussed further in Section 4.

There is a substantial body of work showing tests of serial correlation have low power when shot probabilities follow a Markov process (see, e.g., Miyoshi (2000); see Bar-Eli et al. (2006) and Oskarsson et al. (2009) for good reviews of the entire literature). However, there does not appear to be any work showing tests of autocorrelation not only have low power,

but are highly biased and inconsistent with respect to shot probability autocorrelation, when shot probabilities are determined by an autoregressive process, due to the measurement error problem. I would like to stress that I am not claiming that previous literature confuses shot result and probability autocorrelations, only that the important difference between them is neglected. A secondary contribution of this paper is thus methodological: the results highlight that measurement error bias can arise even when all data are recorded with complete accuracy. Although the focus of this paper is the hot hand in basketball, the results imply serial correlation is likely underestimated in any context in which the true variable of interest is essentially measured with error due to randomness. This could be relevant to analysis of the hot hand in other sports and non-sports contexts, such as hitting streaks in baseball, or the persistence of mutual fund manager performance (see, e.g., Hendricks et al. (1993)), both of which are of course affected by luck in addition to skill.

2 A Model of Shot Probabilities

Suppose basketball shooting data are obtained from a controlled setting: a single player takes T shots from the same location with no defense, distractions, etc. Let x_t denote the result of shot t , with $x_t = 1$ if the shot is made and $x_t = 0$ otherwise. A standard approach used to analyze the hot hand for the player is to examine the sample first autocorrelation of shot results:

$$\widehat{\rho}_{x,1} = \frac{\widehat{Cov}(x_t, x_{t-1})}{(\widehat{Var}(x_t)\widehat{Var}(x_{t-1}))^{0.5}}. \quad (1)$$

\widehat{Cov} and \widehat{Var} denote sample covariance and variance, respectively. Another, almost equivalent approach is to estimate a regression of x_t on x_{t-1} using ordinary least squares (OLS). Both

approaches yield consistent estimates of the first autocorrelation of shots made. Another common approach is to use “runs” tests; these yield very similar results as shown by Wardrop (1999).

Now suppose the data generating process (DGP) is as follows:

$$p_t = \rho p_{t-1} + (1 - \rho)\mu_p + \epsilon_t, \quad (2)$$

$$x_t = \begin{cases} 1 & \text{with probability } p_t \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

with $\rho \in [0, 1)$ and $\mu_p \in (0, 1)$. That is, p_t is the probability of making shot t , and is determined by an AR(1) process. Assume for now that $\mu_p \geq 0.5$. Suppose that $\epsilon_t \sim U[-(1 - \rho)(1 - \mu_p), (1 - \rho)(1 - \mu_p)]$ for all t , so $Var(\epsilon_t) = (1/3)(1 - \rho)^2(1 - \mu_p)^2$. This distribution guarantees $p_t \in [0, 1]$ for all values of ρ, μ_p . In fact, of all the continuous, symmetric, weakly-single peaked (with mean zero) and time-invariant distributions that satisfy this property, this distribution has the highest variance, for any particular values of ρ, μ_p .⁷ This fact will be used further below. Suppose also that ϵ_t is independent of $\epsilon_{t'}, p_{t'}$ for all $t' < t$ (ϵ_t is white noise). This assumption is admittedly questionable and will be relaxed in Section 3. It is then straightforward to show that $E(p_t) = \mu_p$, $Var(p_t) = Var(\epsilon_t)/(1 - \rho^2) = \frac{(1 - \mu_p)^2(1 - \rho)}{3(1 + \rho)}$, and p_t 's first autocorrelation is ρ , for all t . Note that this setup can be interpreted as that of state-space modeling, in which (2) is the state equation and (3) is the observation.

Given this DGP, ρ is naturally the key parameter for understanding the hot hand; if $\rho > 0$

⁷This can be seen by noting that, using recursive substitution, p_t can be written as $\mu_p + \sum_{k=0}^{\infty} \rho^k \epsilon_{t-k}$. If ϵ_{t-k} equalled its upper bound, $(1 - \rho)(1 - \mu_p)$, for all k , then p_t would equal one. Thus, if ϵ_t had an upper bound greater than $(1 - \rho)(1 - \mu_p)$, then p_t could be greater than one. Since it is assumed that the distribution of ϵ_t is symmetric around zero, the lower bound for ϵ_t must be $-(1 - \rho)(1 - \mu_p)$, and it can similarly be shown that $p_t \geq 0$. Thus, the support of ϵ_t is as large as possible such that $p_t \in [0, 1]$ with certainty. And the uniform distribution has a greater variance than any other continuous, symmetric, weakly single-peaked distribution with the same support.

the hot hand exists (the player likely shoots better than average in shot t if he/she shot better than average in $t - 1$), and the greater ρ is, the “streakier” outcomes will be.⁸ However, $\widehat{\rho}_{x,1}$ is an inconsistent estimator of ρ . This will be shown formally below, but the basic intuition can be seen immediately: although $E(x_t) = p_t$, it is (obviously) not the case that x_t is in general equal to p_t . That is, x_t can be thought of as a measure of p_t that, while unbiased, contains error. And it is well known that when a variable is measured with error, this can cause estimation results to be biased and inconsistent.⁹

To formally demonstrate the inconsistency, let w_t denote the error for observation t , i.e. $x_t = p_t + w_t$. Thus $Pr(w_t = 1 - p_t | p_t) = p_t$, $Pr(w_t = -p_t | p_t) = 1 - p_t$, $E(w_t | p_t) = 0$ and $Var(w_t | p_t) = p_t(1 - p_t)$. Then

$$\widehat{\rho}_{x,1} = \frac{\widehat{Cov}(p_t + w_t, p_{t-1} + w_{t-1})}{(\widehat{Var}(x_t)\widehat{Var}(x_{t-1}))^{0.5}} = \frac{\widehat{Cov}(\rho p_{t-1} + (1 - \rho)\mu_p + \epsilon_t + w_t, p_{t-1} + w_{t-1})}{(\widehat{Var}(x_t)\widehat{Var}(x_{t-1}))^{0.5}}, \quad (4)$$

and thus, since $Cov(\epsilon_t, w_{t-1}) = 0$, $Cov(w_t, p_{t'}) = 0$ for all t, t' and $Cov(w_t, w_{t'}) = 0$ for $t \neq t'$,¹⁰ then $\widehat{\rho}_{x,1}$ converges in probability to

$$\text{plim } \widehat{\rho}_{x,1} = \frac{\rho Var(p_t)}{Var(x_t)} = \frac{Var(p_t)}{Var(p_t + w_t)} \rho = \frac{Var(p_t)}{Var(p_t) + Var(w_t)} \rho < \rho. \quad (5)$$

This result highlights how it is measurement error ($Var(w_t) > 0$) that causes the estimator

⁸It is an unfortunate feature of this DGP that the distribution of ϵ_t is also a function of ρ . The DGP was constructed this way, as opposed to one with an error distribution independent of ρ (or a more complicated model, e.g. one with a latent variable), so that the main analytical results could be derived easily for all parameter values. The importance of this assumption is discussed later in this section, and the assumption is dropped in Section 3.

⁹See, e.g., Stock and Watson (2007), p.319-321, for a discussion of how errors-in-variables (measurement error affecting the independent variables) causes regression coefficients to be inconsistent.

¹⁰To show $Cov(w_t, p_{t'}) = 0$, note $E(w_t p_{t'}) = E(E(w_t p_{t'} | p_{t'})) = E(p_{t'} E(w_t | p_{t'})) = E(p_{t'} E(E(w_t | p_t, p_{t'})))$ by the law of iterated expectations. $E(w_t | p_t, p_{t'}) = 0$, since the distribution of w_t conditional on p_t is independent of $p_{t'}$, and as noted above $E(w_t | p_t) = 0$. Thus $E(w_t p_{t'}) = 0$ implying $Cov(w_t, p_{t'}) = 0$. By similar arguments it can be shown that $Cov(\epsilon_t, w_{t-1}) = 0$ and $Cov(w_t, w_{t'}) = 0$ for $t \neq t'$.

to be inconsistent. The precise expression for $\text{plim } \widehat{\rho}_{x,1}$ can then be obtained using the fact that $\text{Var}(x_t) = \mu_p(1 - \mu_p)$ and substituting for $\text{Var}(p_t)$,

$$\text{plim } \widehat{\rho}_{x,1} = \frac{(1 - \mu_p)(1 - \rho)}{3\mu_p(1 + \rho)}\rho. \quad (6)$$

If μ_p were less than 0.5, it could analogously be shown that $\text{plim } \widehat{\rho}_{x,1} = \frac{\mu_p(1 - \rho)}{3(1 - \mu_p)(1 + \rho)}\rho$.¹¹ These results imply the upper bound over μ_p , occurring when $\mu_p = 0.5$, for $\text{plim } \widehat{\rho}_{x,1}$ is $\frac{(1 - \rho)}{3(1 + \rho)}\rho$. This expression has a maximum of approximately 0.057 that occurs when $\rho = 0.4$; when $\rho = 0.8$, $\text{plim } \widehat{\rho}_{x,1}$ is less than 0.03 and when $\rho = 0.9$, $\text{plim } \widehat{\rho}_{x,1}$ is less than 0.02! As μ_p moves away from 0.5 the bias only worsens as well.

These results indicate that $\widehat{\rho}_{x,1}$ is not just a biased estimator of ρ , but that the bias is extreme. The simplicity of the model allows it to parsimoniously convey the force driving the bias: $\frac{\text{Var}(p_t)}{\text{Var}(p_t) + \text{Var}(w_t)}$ (inversely) determines the bias caused by measurement error, and due to the nature of basketball shot data, it appears likely that $\text{Var}(p_t) \ll \text{Var}(w_t)$. When p_t has a mean of 0.5, measurement error is generally large compared to variation in p_t ; when p_t 's mean is higher, measurement error is smaller, but variation in p_t is likely smaller as well.

Given the AR(1) model of shot probabilities, $\text{Var}(p_t)$ is increasing in $\text{Var}(\epsilon_t)$ and ρ . Given the particular distribution of ϵ_t used above, $\text{Var}(\epsilon_t)$ is decreasing in ρ . This seems reasonable, since when ρ is larger, for fixed values of the other parameters, $\text{Var}(\epsilon_t)$ must become smaller for the p_t 's to stay in $[0, 1]$. Thus, in this model an increase in ρ has competing effects on $\text{Var}(p_t)$, and the negative effect dominates at least for $\rho > 0.4$. However, it would also be reasonable to assume $\text{Var}(\epsilon_t)$ is independent of ρ . If that were the case, then $\text{Var}(p_t)$ would be unambiguously increasing in ρ , and hence $\text{plim } \widehat{\rho}_{x,1}$ would also unambiguously increase in

¹¹This result would follow from use of an analogous assumption for the distribution of ϵ_t , $U[-(1 - \rho)\mu_p, (1 - \rho)\mu_p]$.

ρ . Hence, it is not generally true that $\text{plim } \widehat{\rho}_{x,1}$ decreases in ρ for large ρ . One might wonder then if the bias is still extreme for large ρ for distributions of ϵ_t that do not depend on ρ . But as noted above, $\text{Var}(\epsilon_t)$ is greater using the assumed distribution, as compared to other distributions of a similar type (continuous, symmetric, single-peaked, and time-invariant), for given values of ρ, μ_p . This means that if $\text{Var}(\epsilon_t)$ did not depend on ρ , and ϵ_t had the same type of distribution, then $\text{Var}(p_t)$ could only be lower, and hence bias would only be greater.

The assumptions that ϵ_t 's distribution is continuous, symmetric and single-peaked seem very reasonable. But the assumption of time-invariance (conditional homoskedasticity) was noted above to be highly questionable. This assumption is relaxed in the simulation analysis presented in the next section.

3 Simulation

Suppose shot probabilities are still determined by (2), but now $\epsilon_t|p_{t-1} \sim U[-\alpha\delta_t, \alpha\delta_t]$, with $\delta_t \equiv \min\{\rho p_{t-1} + (1-\rho)\mu_p, 1 - (\rho p_{t-1} + (1-\rho)\mu_p)\}$ and $\alpha < 1$. This distribution also ensures $p_t \in [0, 1]$, while allowing the distribution of ϵ_t to depend on p_{t-1} in a simple but natural way: ϵ_t varies more when it has “room” to, i.e., when p_{t-1} is closer to 0.5. Note $\text{Var}(\epsilon_t)$ will generally be larger when α is larger. Note also that it is still the case that $\text{Cov}(\epsilon_t, p_{t-1}) = 0$ and it can also be assumed that conditional on p_{t-1} , ϵ_t and $\epsilon_{t'}$ are independent for all $t' < t$.

To see how estimation bias is affected by changing the distribution of ϵ_t this way I simulate 1,000 samples of shot probabilities determined by this process and corresponding shot results, each with 1,000 observations, for various parameter values. For each sample,

$$x_t = \beta_0 + \beta_1 x_{t-1} + v_t \tag{7}$$

is estimated using OLS for each sample. The sample sizes are conservatively large, as compared to those used in experimental work (e.g., Gilovich et al. (1985)). Two values of μ_p are used, 0.5 and 0.75, two values of α , 0.25 and 0.5, and eight values of ρ , 0.1, 0.2, ..., 0.9, so there are $2 \times 2 \times 8 = 32$ simulations in total. To get a sense of the effects of the parameters, note p_t ranges (on average across simulations) from 0.37–0.63 and 0.20–0.80 when $\mu_p = 0.5$, $\alpha = 0.25$ for $\rho = 0.1$ and $\rho = 0.9$, respectively (0.24–0.76 and 0.08–0.92 when $\alpha = 0.5$). When $\mu_p = 0.75$, p_t ranges from 0.68–0.82 and 0.38–0.90 if $\alpha = 0.25$ (0.61–0.88 and 0.15–0.96 if $\alpha = 0.5$). Note again that the OLS slope from a regression of p_t on p_{t-1} is a consistent estimate of ρ .

Results are presented in Figure 1. The average estimate of β_1 ranges from 0.015–0.11 when $\rho \leq 0.8$ and from 0.032–0.17 when $\rho = 0.9$ for all values of α and μ . These results imply the results Section 2, while somewhat distorted, are in the ballpark of what occurs even with heteroscedasticity. While it is probably not the case that shot autocorrelation actually declines as ρ increases, it does seem that shot autocorrelation remains very small. The figure also shows the likelihood of an estimate being negative can be greater than 40% for $\rho > 0.6$ and is greater than 10% for all values of μ and α for all $\rho \leq 0.5$. Thus, researchers may find negative autocorrelation even when the hot hand is relatively strong.¹² Power is also very low when $\rho \leq 0.2$ (less than 10%) and is generally less than 50% for ρ up to 0.7.¹³

¹²Runs tests yield extremely similar results, consistent with the finding of Wardrop discussed above; they are not reported but available upon request.

¹³This is determined using homoscedasticity-only standard errors to be consistent with some of the work done in previous literature; power would likely be even *lower* if appropriate heteroscedasticity-robust standard errors were used.

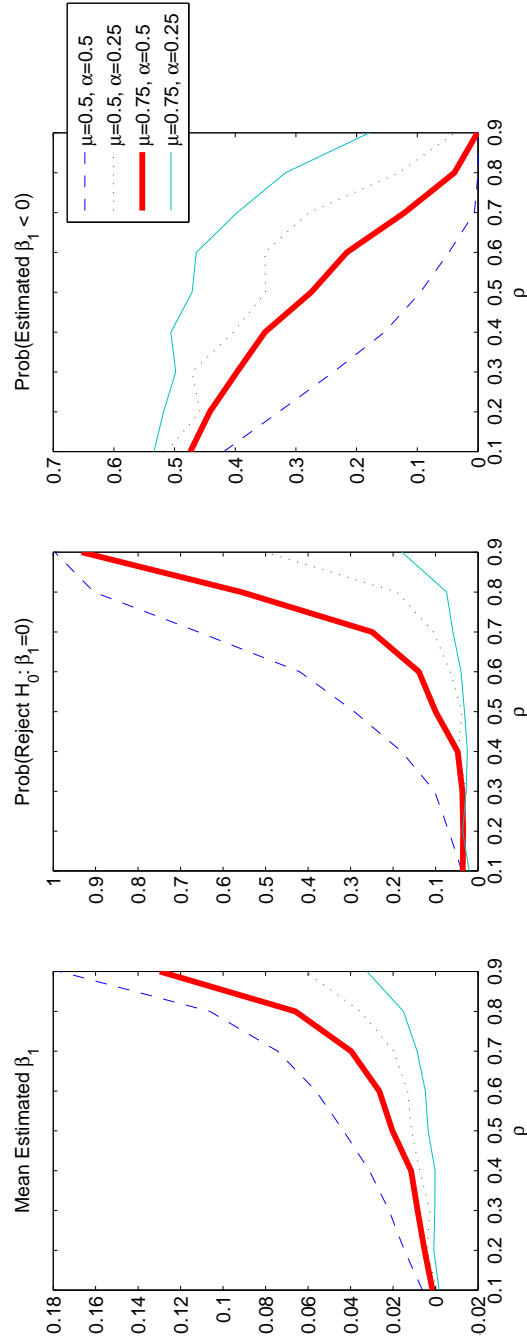


Figure 1: Simulation results

4 Discussion

This paper has shown that standard analysis of the hot hand underestimates the first auto-correlation of shot probabilities substantially. This raises the following important question: if it seems we cannot learn about shot probabilities from shot results, why should we even care about these probabilities? That is, is ρ at all relevant to observable phenomena, and if not, is this paper anything more than an abstract probabilistic exercise?

First and foremost, even if ρ could never be detected in the data, this paper would help to clarify the interpretation of shot result correlation, and the distinction from probability correlation. This distinction is important unto itself, but especially since, as discussed above, the lack of shot correlation is often cited as evidence of flaws in human intuition. People may indeed be very poor pattern detectors, but referring to the myth of the hot hand in basketball seems to not be a good way to support this claim, since myth implies non-existence, and non-existence seems to not have been proven. Even statements about the lack of evidence in the data for the hot hand could be misleading, if it is not clarified that measurement error bias prevents potential evidence from being observed.

One might argue next that practitioners—fans, players etc—think mainly about shot results, and not probabilities, when thinking about hot hands, and so this analysis of probabilities does not actually have implications for psychology. For example, according to the probability hot hand definition a player could be hot—shoot at a probability above his average—for 10 straight shots, but miss them all, and if practitioners observed this they would surely say the player was cold. But ability (shot probability) is still what is relevant to practitioners’ decision problem of who should take the next shot, so ability must be considered by practitioners at some level. In other words, when a player says, “Give me the ball, I’m hot,” the player is implicitly

saying he thinks he is shooting at a higher probability than normal, not that he will definitely make his next shot. A related argument would be that if practitioners think they can draw inferences about the hot hand from shot results, they must still be making mistakes. However, it is not clear that statistical analysis of shot result data indicates practitioners' inferences are necessarily incorrect. Practitioners have much more information than the average statistician, at least for particular situations. For example, they can see whether a made shot was lucky (perhaps due to an inadvertent bank off the backboard) or was of high or low difficulty. They can also observe changes in offensive and defensive sets, and individual player match-ups. Players know how they felt when they took their last shot. This information is, to some extent, unquantifiable and therefore cannot be used for statistical analysis. But it still could be sufficient to justify relatively precise beliefs about whether players are in hot/cold states—the information would effectively reduce measurement error (the degree to which an observed shot result was affected by unobserved factors). Thus, although statisticians may not be able to see the hot hand, practitioners may still be able to. Furthermore, even if practitioners do not have sufficient information to justify their beliefs about hotness in most situations, it is still very possible they have enough information (from their extensive personal experience and observation) to know that players' abilities at least sometimes get hot. Much of the previous literature seems to have concluded from shot result analysis that this basic, broad belief is wrong; this paper shows this is not at all necessarily the case.

I should also clarify that I am not claiming ρ cannot possibly be estimated precisely with shot result data, just that the measurement error issue should be addressed when this estimation is attempted. There are almost certainly better estimators out there than those discussed above (the sample autocorrelation and OLS coefficient). In fact, for the very simple model of shot probabilities analyzed in this paper, an instrumental variable regression estimation

strategy, using the second lagged shot made as an instrument, yields a consistent estimate of ρ . However, the practical value of this particular approach is likely limited, as simulations suggest it requires an extremely large amount of data to be reasonably precise. This is due to the instrument being very weak; moreover, this may not be a robust approach as the instrument may not even be valid for other DGPs. Developing more useful estimators that account for measurement error would be an interesting line of future work (see Albert and Williamson (2001) for detailed discussion of related issues). The most appropriate estimator very well may depend on the context; e.g., the baseball and mutual fund examples mentioned above may warrant different methods. I do suspect that developing a useful estimator for basketball may be very difficult to the severity of the measurement error bias in the particular context. Another approach would be to try to “back out” probability correlations from shot correlations by making stronger assumptions on the probability model and its parameters. For example, if it was assumed the model of this paper was correct, $\mu = 0.75$ and $\alpha = 0.5$, then the results of Arkes (who found free throw shot autocorrelation is approximately 0.03) would correspond to a ρ of around 0.65. A range of parameter values/models could be examined to determine a range of ρ 's implied by Arkes' results. Regardless, the measurement error issue only strengthens the conclusion from Arkes' results on the hot hand's existence.

Finally, I again note that the results of this paper likely have implications for hot hand analysis in other contexts (sports and other). As discussed in the review papers cited above, the evidence for/against the hot hand varies substantially across sports. This variation may be partly explained by the finding (see equation (4)) that measurement error is likely to be a bigger problem in contexts where the variance of p_t (true ability) is smaller and the variance of w_t (measurement error) is larger.¹⁴

¹⁴Oskarsson et al. (2009) discusses how evidence of the hot hand has been found more often in sports that are more “controllable,” such as bowling and archery, as opposed to more “chaotic” sports like basketball.

References

- Albert, J. and Williamson, P. (2001). "Using model/data simulations to detect streakiness." *The American Statistician*, vol. 55(1), 41–50.
- Arkes, J. (2010). "Revisiting the Hot Hand Theory with Free Throw Data in a Multivariate Framework." *Journal of Quantitative Analysis in Sports*, vol. 6(1).
- Bar-Eli, M. and Avugos, S. and Raab, M. (2006). "Twenty years of "hot hand" research: Review and critique." *Psychology of Sport and Exercise*.
- Cao, Z. (2011). "Revisiting the hot hand: reality and perceptions." Working paper.
- Dorsey-Palmateer, R. and Smith, G. (2004). "Bowlers' hot hands." *The American Statistician*, vol. 58(1), 38–45.
- Gilovich, T., Vallone, R. and Tversky, A. (1985). "The hot hand in basketball: On the misperception of random sequences." *Cognitive Psychology*, vol. 17(3), 295–314.
- Hendricks, D. and Patel, J. and Zeckhauser, R. (1993). "Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988." *Journal of Finance*, vol. 48(1), 93–130.
- Koehler, J.J. and Conley, C. (2003). "The 'Hot Hand' Myth in Professional Basketball." *Journal of Sport & Exercise Psychology*, vol. 25(2). p.253–259.
- Lehrer, J. (2009). *How We Decide*. Houghton Mifflin Harcourt.
- Miyoshi, H. (2000). "Is the "hot-hands" phenomenon a misperception of random events?" *Japanese Psychological Research*, vol. 42(2). p.128-133.
- Oskarsson, A.T., Van Boven, L., McClelland, G.H. and Hastie, R. (2009). "What's next? Judging sequences of binary events." *Psychological bulletin*, vol. 135(2). p.262–285.

- Rabin, M. (1998). "Psychology and Economics." *Journal of Economic Literature* , 36(1). P.11–46.
- Reifman, A. (2011). *Hot Hand: The Statistics Behind Sports' Greatest Streaks*. Potomac Books Inc.
- Stock, J.H. and Watson, M.W. (2007). *Introduction to econometrics: 2/E*. Yale University Press.
- Thaler, R.H. and Sunstein, C.R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Wardrop, R.L. (1995). "Simpson's paradox and the Hot Hand in basketball." *The American Statistician*, vol. 49(1). p.24–28.
- Wardrop, R.L. (1999). "Statistical tests for the hot-hand in basketball in a controlled setting." Unpublished manuscript.