Reverse Engineering:
Obtaining Average Spin Components from Statcast Data

Alan M. Nathan*

Department of Physics, University of Illinois, Urbana, IL 61801

(Dated: July 8, 2024)

The publicly available spin-related Hawkeye data include the total spin ω and the angle $\Phi = \arctan(\omega_z/\omega_x)$, both on a pitch-by-pitch basis. Also included is the *average* spin efficiency at release for a given pitcher and pitch type. In this brief document, I will show how to use that information to obtain mean values, $\langle\omega_x\rangle$, $\langle\omega_y\rangle$, and $\langle\omega_z\rangle$, for the three spin components.

In approaching this problem, it is helpful to express both the spin vector $\vec{\omega}$ and the velocity vector at release $\vec{v}$ in spherical coordinates as follows:

$$\omega_x = \omega\sin\Theta\cos\Phi \quad \omega_y = \omega\cos\Theta \quad \omega_z = \omega\sin\Theta\sin\Phi\,, \tag{1}$$

and

$$\hat{v}_x \equiv v_x/v = \cos\theta\sin\phi \quad \hat{v}_y \equiv v_y/v = \cos\theta\cos\phi \quad \hat{v}_z \equiv v_z/v = \sin\theta\,. \tag{2}$$

Here the polar angle Θ is the angle of the spin with respect to the y axis and the azimuthal angle Φ is the angle of the projection of the spin in the xz plane with respect to the x axis. Also θ and ϕ are the vertical and horizontal release angles, respectively, both of which are obtainable from the public Statcast data as follows:

$$\phi = \arctan\left(\frac{v_x}{v_y}\right) \quad \theta = \arcsin\left(\frac{v_x}{v}\right). \tag{3}$$

As already noted, both the total spin rate ω and the azimuthal angle Φ are publicly available, but the polar angle Θ is not. If Θ were known, then everything would be known about the spin vector and one could proceed with the lift-side separation exactly.

What *is* known publicly is the average spin efficiency at release for a given pitcher and pitch type, which I denote by $\langle\sin\theta_S\rangle$, where θ_S is the angle of the spin vector with respect

* a-nathan@illinois.edu

to $\vec{v}$. It is actually more compact to give an expression for $\langle\cos\theta_S\rangle$, which can be written in polar coordinates as follows:

$$\langle\cos\theta_S\rangle \;=\; \hat{v}_x \sin\Theta\cos\Phi + \hat{v}_y\cos\Theta + \hat{v}_z\sin\Theta\sin\Phi\,, \tag{4}$$

where the $\hat{v}$ terms are given in Eq. 2. If the mean value of the spin efficiency $\langle\sin\theta_S\rangle$ is known for a given pitcher and pitch type, then $\langle\cos\theta_S\rangle$ is also known and the left-hand-side of Eq. 4 is determined, at least up to a sign. The equation can then be solved to find the only unknown parameter Θ.

As mentioned above, the sign of $\langle\cos\theta_S\rangle$ is ambiguous, as illustrated by the following example. Suppose the spin efficiency is 0.50, so that $\sin\theta$=0.50. That means θ is either 30° or 150°, depending on whether the spin vector points forward (i.e., toward the catcher) or backward (i.e, away from the catcher), respectively. In the forward case, $\langle\cos\theta_S\rangle$=0.833 (resulting in $\omega_y < 0$) whereas in the backward case $\langle\cos\theta_S\rangle$=-0.833 (resulting in ω_y >0). That is the origin of the sign ambiguity. However, the sign can be reasonably guessed, since ω_y is expected to be positive for a LHP and negative for a RHP, a prescription that is followed in the discussion below.

We next discuss two methods for solving Eq. 4.

METHOD 1

Solving Eq. 4 is actually quite easy, once the tricks are known. The equation can be cast in the form

$$A\sin\Theta + B\cos\Theta \;=\; C\,, \tag{5}$$

where

$$A \;=\; \hat{v}_x\cos\Phi + \hat{v}_z\sin\Phi \quad B \;=\; \hat{v}_y \quad C \;=\; \pm|\langle\cos\theta_S\rangle|\,, \tag{6}$$

and where the + or - sign are for RHP or LHP, respectively. In turn this can be written

$$R\sin(\Theta + X) \;=\; C\,, \tag{7}$$

where

$$R \;=\; \sqrt{A^2+B^2} \quad X \;=\; \arctan\left[\frac{B}{A}\right]. \tag{8}$$

Eq. 7 can then be solved to find

$$\Theta = \arcsin\left[\frac{C}{R}\right] - X\,, \tag{9}$$

where Θ should be interpreted to be the value averaged over all pitches for a given pitcher and pitch type. Once Θ is found, one can then use Eq. 1 to find the average values of the three spin components, where Φ is taken as the average value.

METHOD 2

There is a second approximate method to find Θ, which comes from the observation that $\vec{v}$ is primarily in the $-y$ direction, so that $\hat{v}_y \approx -1$. As a result, Eq. 4 can be rearranged and approximated by

$$\cos\Theta \approx \mp|\langle\cos\theta_S\rangle|\,, \tag{10}$$

where the - and + signs are for RHP and LHP, respectively. Eq. 10 uniquely determines Θ, from which the average values of the three spin components can be found via Eq. 1.

TESTING THE METHODS

With full access to the 3D spin data, it is possible to test the reverse engineering schemes discussed in the preceding section. The analysis utilized 675k pitches from the 2021 season. We examine averages over pitches and pitch types, with a threshold of 150 pitches, of which there are 794 different combinations. Fig. 1 compares the exact average spin components with those determined from the two reverse engineering methods. The two quantities compare extremely well, especially for ω_x and ω_z. There are a handful of obvious outliers for ω_y, all of which occur due to the "wrong sign". That is, the prescription for resolving the sign ambiguity described above, whereby $\omega_y < 0$ for RHP and $\omega_y > 0$ for LHP, seems to be violated. Curiously of the 30 such averages, 24 of them are for changeups. A good example is the **"airbender changeup"** of RHP Devin Williams, who threw 575 of these in 2021, with ω_y values narrowly distributed about a mean value of +1711 rpm and a mean spin efficiency of 79%. Indeed, it seems to be well known that this pitch has **"reverse gyro"**.

One interesting, albeit not surprising, feature is that the approximate method of Model 2 seems to do about as well as the exact method of Model 1 in describing the data.

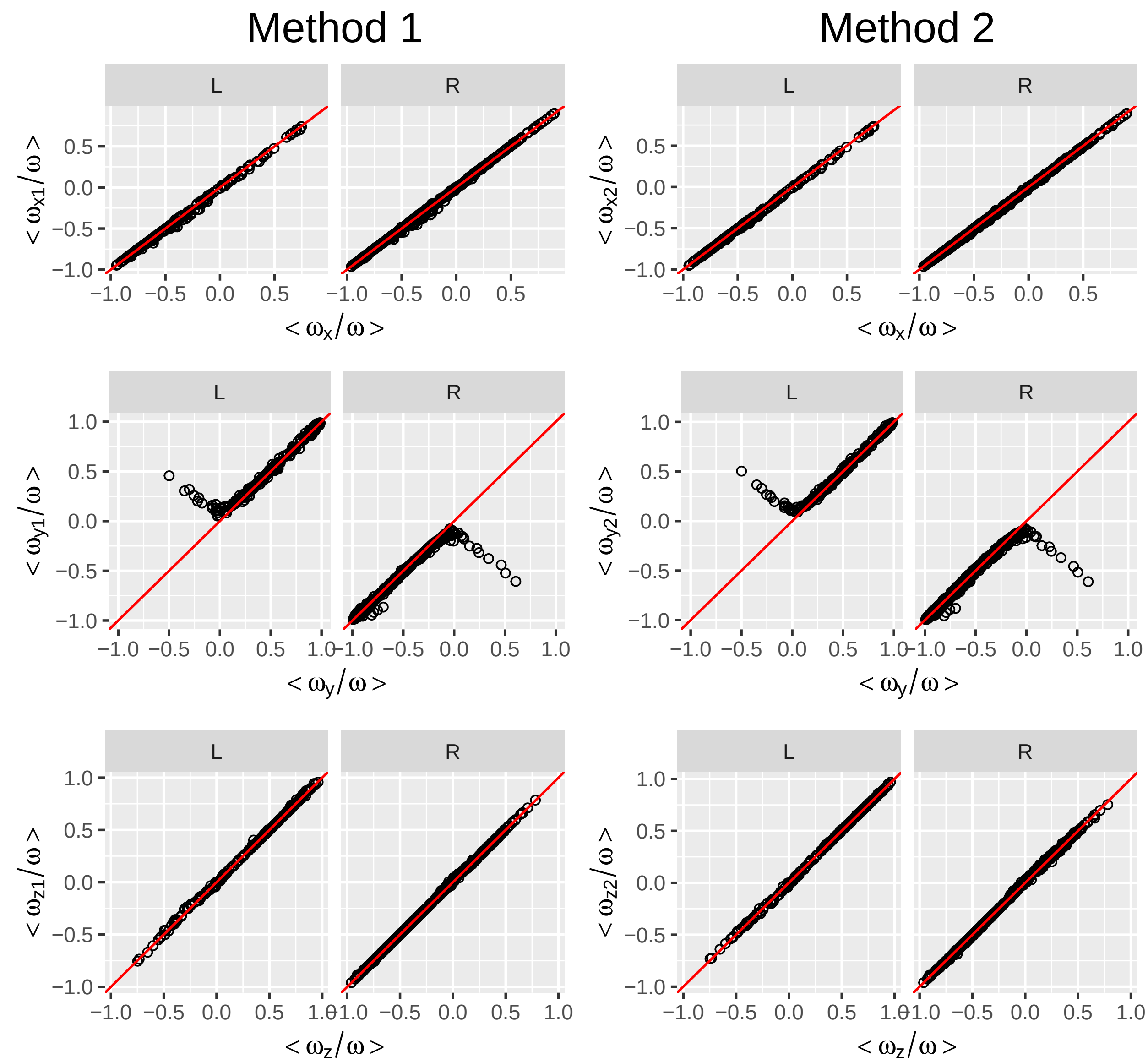

FIG. 1: Comparing exact spin components with those derived from both methods of reverse engineering, with the red line indicating equality. Each of the 794 points is an average for a given pitcher and pitch type, with a threshold of 150 pitches and with RHP and LHP shown separately.