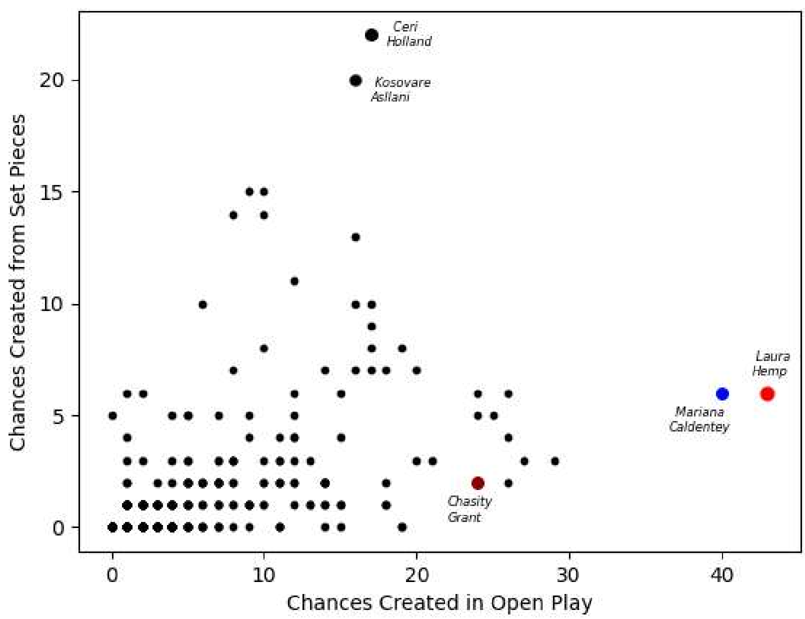
Evaluating Playmakers in the Women's Super League (WSL):

Opportunity-adjusted model of chance creation.

This report is a brief extension of the analyses carried out in *Evaluating playmakers in 9 leagues*. One of the conclusions drawn there was that across all the (men's) leagues examined, the same drivers explained the numbers of chances a player is expected to create. These drivers were stable across four different tiers of the English football system; across all Big Five leagues; and (in the case of League One) between successive seasons. Do these same drivers also apply in the women's game? If so, we would be one step closer to claiming this as a Law of Football. To test this possibility, we analyse playmaking in the WSL using the same methodology as we used in the men's leagues. Aside from this so-called Law, it is of interest to know who the most productive chance creators were in the WSL over the 2025-26 season.

In *Use Chances Created, not Assists or Expected Assists,* we were able to show that Chances Created (CC) is a more reliable statistic than Assists (A) or Expected Assists (xA). Rather more surprisingly, someone's propensity to create chances in season 2024-25 was better able to predict not only their CC numbers in the following season, but also their future Assists and xA. That is, CC in t_1 is the best predictor in t_2 of CC, A, and xA. So, the analyses we carry out here, and in the previous report, are more than of passing interest in identifying playmakers, but have solid predictive value for the coming season.

In a nutshell, we build a model of how many chances we expect from the typical player, given the opportunities accorded her (the number of minutes she is on the pitch, how many passes she gets to make in the final third, number of crosses, and number of passes she gets to make in the non-final third). Some players make an excess of chances above the typical; others fall below what might be expected. This excess (or deficiency), is our index of playmaking performance. The procedural detail is in the two previous reports.



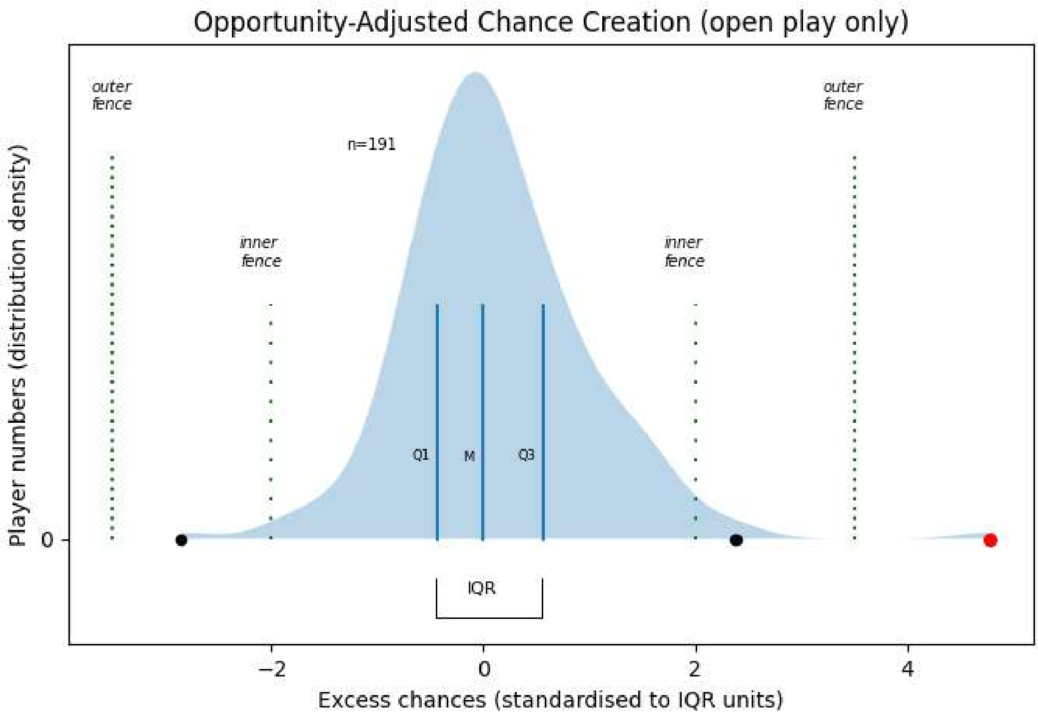
In the above scattergram, each dot represents a different player in the WSL, measured on two aspects: chances created from open play (along the x-axis), and chances created from set pieces (y-axis). To be consistent with the previous investigation we will analyse only chances created from open play, and furthermore we have filtered out all players whose game-time for the season was less than 450 minutes. This leaves 191 players *in* the analysis, and 57 rejected for lack of minutes.

There are two players who immediately catch the eye. They are: Laura Hemp (CC = 43 in open play – red dot); and Mariana Caldentey (CC = 40 in open play – blue dot). Both are well ahead of the third position (CC = 29). Scattergrams like this are a standard graphical device and from this information alone, we might conclude that LH and MC were neck and neck in chance creation. However, once the opportunities to make chances are adjusted for, MC's performance seems less remarkable. On the other hand, Chasity Grant (marked in dark red) who appears to be unremarkable on the scatterplot, turns out to be the second best performer in the opportunity-adjusted assessment of CC, as shown in the table below.

WSL	Mins	CC	ExpCC	Excess	Excess in IQRs
Lauren Hemp	1384	43	21.188	21.812	4.778
Chasity Grant	1227	24	13.273	10.727	2.388
Riko Ueki	1727	24	13.355	10.645	2.371
Melvine Malard	1520	27	18.152	8.848	1.983
Alyssa Thompson	1420	29	20.368	8.632	1.937
Sam Kerr	685	14	6.293	7.707	1.737
Signe Gaupset	927	18	10.517	7.483	1.689
Amanda Nildén	1778	14	6.618	7.382	1.667
Vivianne Miedema	1451	24	16.875	7.125	1.612
Kiko Seike	1616	26	19.177	6.823	1.547

CC are the number of chances created in open play, and ExpCC are the expected numbers of CCs, estimated from on the regression model, which is based on the opportunities for chance creation that each player experiences. Excess = (CC minus ExpCC) is our index of playmaking creativity. Finally, Excess is re-expressed in terms of IQR units. (IQR = Interquartile Range = Q3 – Q1). Those with Excess > 2 IQR units are classified as outliers, and those with Excess > 3.5 IQR units are classified as far outliers. A downloadable file is available that covers all players, not just the top 10.

A graphical representation of the entire distribution of Excess (IQRs) is in the 'violin plot', below. The distribution has been smoothed by kernel smoothing. There is one outlier on the negative side (someone who must have wasted many opportunities), and three on the positive side. One of these (Lauren Hemp) is a far outlier, marked in red. And the other two are so close to each other (2.388 and 2.371, see Table) that they appear as one. The graph makes in clear that Lauren Hunt had an outstanding season in this respect, outperforming the next best by some distance. Furthermore, Mariana Caldentey slipped a long way out of the top 10, to 37th place. From the opportunities play gave her, it was estimated she should have made 36.846 chances, so that her actual tally of 40 was not much above expected.



We now turn from individual performance to the drivers behind expected performance, which is the sliding scale against which over- and under-performance are measured. The regression model for WSL is in the following table.

	Mins x1000	Pass 1/3rd x100	Cross x100	Pass 2/3rd x100	Const x1
parameters	4.739	5.857	0.649	-1.332	-0.746
t-values	4.61	10.93	0.24	-8.25	-1.04

Suppose someone had M minutes of play, attempted P1 passes in the final third, C crosses, and P2 passes not in the final third, then her expected number of chances created (if she was a bang on typical player) would be:

$$CC_{expected} = 0.004739\ M \ + \ 0.05857\ P1 + \ 0.00649\ C \ - \ 0.01332\ P2 \ - \ 0.746.$$

This looks very much like the kind of regression equations we found in the men's game. In order to get an accurate fix on just how similar it is, we followed the same procedure as used for the men's leagues.

1. Estimate each person's expected CC from the regression equation of their league (say E1, for English 1st tier, the Premier League).
2. Now re-estimate those same players' expected CCs using the regression equation from a different league (say, WSL).

3. Correlate the estimates from (1) and (2).

If the correlation is high, then there is agreement between leagues about how chance creation occurs. To follow our example through, Pearson's correlation R = 0.9930, indicating very good agreement. See darker green in the top right. If on the other hand we swap the roles of E1 and WSL so we are making two assessments of players in WSL, not E1, we get R = 0.9933 (darker green, bottom left). Note that, unlike in a standard correlation matrix, the coefficients are not symmetrical.

Model coefficients (from each of the 10 leagues)

Players (in the 10 leagues)

	E1	E2	E3	E3a	E4	F1	G1	I1	S1	WSL
E1	1	0.9881	0.9977	0.9984	0.9881	0.9910	0.9982	0.9968	0.9943	0.9930
E2	0.9776	1	0.9712	0.9842	0.9246	0.9903	0.9918	0.9932	0.9793	0.9889
E3	0.9976	0.9858	1	0.9986	0.9939	0.9834	0.9943	0.9956	0.9919	0.9830
E3a	0.9986	0.9908	0.9985	1	0.9874	0.9880	0.9980	0.9980	0.9924	0.9888
E4	0.9920	0.9736	0.9964	0.9930	1	0.9693	0.9857	0.9871	0.9830	0.9691
F1	0.9906	0.9948	0.9824	0.9870	0.9574	1	0.9946	0.9947	0.9966	0.9983
G1	0.9981	0.9945	0.9946	0.9979	0.9762	0.9940	1	0.9987	0.9932	0.9966
I1	0.9958	0.9954	0.9937	0.9971	0.9694	0.9933	0.9987	1	0.9951	0.9924
S1	0.9941	0.9899	0.9882	0.9892	0.9705	0.9961	0.9937	0.9953	1	0.9925
WSL	0.9933	0.9955	0.9845	0.9902	0.9703	0.9984	0.9968	0.9944	0.9925	1

Note, the high level of correlation throughout the matrix does not mean that approximate same numbers are computed in either case. A little extra work is needed for that. The simple (y = a + bx) regression equation that estimates E1 from WSL is:

$$E1 = 0.058903 + 0.90324\ WSL \qquad (R = 0.9930)$$

and the equation that estimates WSL from E1 is:

$$WSL = 0.232525 + 1.083854\ E1 \qquad (R = 0.9933)$$

In particular, $a \neq 0$, and $b \neq 1$. Think of it more like translating between the two temperature scales. They also have a linear relationship, but in translating ºC to ºF, we use the equation F = 32 + 1.8 C, where it is also the case that $a \neq 0$, and $b \neq 1$.

It is clear from the correlations shaded in green that the same kinds of factors drive CC in the women's game as they do in the men's game. But if slight nuances can be teased out, they would be that WSL is most like Ligue 1 and the Bundesliga, and least like the English League Two. But then again, League Two is the odd one out for almost all leagues.