The effect of batting during the evening in cricket

Eamon McGinn

Written while at:
The University of New South Wales
Sydney, NSW, Australia
2052

Email: eamon.m@gmail.com

Abstract

The team batting second in a day-night cricket match faces different playing conditions to the team batting first. This paper quantifies the effect on the runscoring ability of the team batting second by using the difference in differences estimator. This approach indicates that batting during the evening reduces the expected number of runs scored per over by 0.2. The effect explains around 12% of the total margin of loss for teams in the treatment group who lose the match. It is also found that batting during the evening increases the expected number of wickets lost at any given point in the innings.

Keywords: batting, cricket, day-night, treatment effect

1. Introduction

Cricket is a bat and ball game based around batsmen, bowlers, safe zones, catches, outs and runs. Like other bat and ball games, cricket generates a large amount of quantitative data and is amenable to quantitative analysis. As a result there are a number of papers which have looked at cricket from a quantitative perspective. In particular, there have been attempts to estimate the causal effect of factors like weather conditions or who wins the toss on the outcome of cricket matches. It has, for example, been difficult to identify a statistically significant effect of winning the toss on match outcomes (Morley & Thomas, 2005).

This paper extends this literature by identifying the effect of batting during the evening on a team's run scoring ability. The effect of batting during the evening is of interest in cricket as one form of the game, the one day international (ODI), can commence in the afternoon and run into the evening. This means that the team batting first bats during the day while the team that bats second bats mostly during the evening. The two teams face different batting conditions and this may affect their performance. Gaining a correct understanding of the effect of batting during the evening is therefore of interest for both strategic and administrative reasons.

From a strategic perspective, a correct quantification of the effect of batting during the evening may allow for improved decision making by team captains and coaches. Batting during the evening has been a strategic issue since the introduction of day-night matches to the ODI style of the game in 1977. Teams have generally perceived that it is a disadvantage to bat during the evening and, in the data used for this paper, around 78% of the teams that won the toss in a day-night match elected to bat first. This compares to 43% of those in day matches. This effect has also been commented on by players: Ian Bell, an English cricketer, has noted that "it was commonly agreed that you had to bat first if you won the toss and avoid batting under lights" (Bell, 2007). Dawson *et al.* (2009, p. 1790) also record how captains and coaches have noted the effect. The basis for this perception is supported by match results as teams batting first in day-night matches win 52% of the time, while only 43% of teams batting first in day matches win.

Administratively, if the team batting in the evening is systematically disadvantaged, some corrections to the rules of the game could be considered in order to make it fairer. A correction to the rules requires a quantification of any disadvantage. Any disadvantage from batting during the evening is of particular interest at the present time as there are currently plans to introduce evening batting in test matches (cricinfo, 2011). This innovation has already been introduced at the county cricket level in England, which is one level below international (Turbervill, 2011). While a test match would likely see both teams being exposed to batting during the evening, which would reduce the systematic nature of the disadvantage present in ODI matches, a significant effect on run scoring potential from batting during the evening should be taken into account when considering the costs and benefits of introducing evening test matches.

Although important for the game from both a strategic and administrative point of view, identification of the effect of batting during the evening is made difficult by the presence of unobserved factors such as team quality and preferences for evening play. These difficulties have not been an issue in past papers as their focus has been on factors (such as the coin toss and weather) which are essentially exogenous and are far less likely to suffer from endogeneity issues. In this paper, the coin toss is used as a source of exogenous variation while the difference in differences estimator is also used to control for unobserved heterogeneity in team quality. The difference in differences estimator has not previously been applied to the analysis of cricket matches and is another contribution of this paper.

It is found that batting during the evening reduces the expected number of runs scored each over by around 0.21. To put this effect into context, over the course of a 50 over match this could amount to around 10 runs. Of the 252 matches in the dataset where the team batting second in a day-night match lost the toss and the match, 24 (9.5%) of these matches have a margin of victory which is ten runs or less. Overall, the estimate explains an average of 12.2% of the margins of victory in these 252 matches. When considering the 95% confidence interval of the estimate, this ranges from 5.0% to 19.4%.

2. Background on cricket

Cricket is, for the most part, similar to other bat and ball games but there are some specific details which are important to be aware of. Teams are made up of eleven players with teams taking turns at batting and fielding, each turn being called an innings. The aim of the batting team is to score runs while the aim of the bowling team is to get ten of the batsmen out (this is also called 'taking a wicket' or 'dismissing' the batsman).

There are currently three varieties of cricket being played at an international level: test cricket, ODI and twenty20 (T20). A test match can run for up to five days and is the oldest form of the game; the first official test match was played in 1877. ODI matches have been played since 1971 and take place during a single day. They were primarily developed to make cricket more appealing for broadcast on television. ODIs can take place entirely during the day or can commence in the afternoon and run into the evening, this is called a day-night match. Day-night matches contribute to the ongoing popularity of ODI cricket as they allow for broadcasting of live matches during primetime while also allowing spectators to attend the match at convenient times. In a day-night match, the team batting first bats during the day while the team that bats second bats mostly during the evening. This means that the team batting second may face different batting conditions due to both climate and lights than the team batting first. Twenty20 matches are the most recently introduced style - the first international match was played in 2007 – and they are completed in around 3 hours.

It will be useful to have an understanding of some of the key rules and nomenclature of ODIs for reading the rest of the paper. The key words from cricket that will be used most frequently in this paper are overs, innings, wickets and runs. An over is a set of six legal balls and each innings in an ODI is generally made up of a maximum of 50 overs. A bowler can also bowl an illegal ball which does not count towards the six legal balls of the over. The most common illegal deliveries are a wide, where the ball is pitched too far to the left or right of the batsman, and a no ball, normally where the ball is pitched too high or the bowler proceeds too far down the pitch before releasing the ball. A team loses a wicket when one of their batsmen gets out. As each team has 11 batsmen, who bat in pairs, when a team loses 10 wickets their innings is over. Runs are scored when a batsman hits a ball and the batting partners exchange positions by crossing the pitch. In an ODI, the batting team's innings ends if they have batted for fifty overs, lost ten wickets or (if batting second) they have scored more runs than the first team; the match is over when both teams have batted; and the winner is the team that has scored more runs at the end of the match (a tie is also possible). As an indication, in the data set used for this paper, the average number of runs scored in an innings is 237.5.

3. Literature review and likely presence of unobservables

There is a body of quantitative analysis in cricket which is spread across a range of disciplines. The set of literature that is most applicable to this paper is where causal effects of factors such winning the coin toss and weather conditions on the outcome of the match are estimated.

Papers which have looked into the effect of winning the toss have found mixed results. Allsopp and Clarke (2004) considered ODIs and test match cricket and found no advantage from winning the toss. Similar results were found in Clarke and Allsopp (2001) and in de Silva and Swartz (1997) for ODIs. Saikia and Bhattacharjee (2010) also find similar results when investigating T20s. In contrast, Morley and Thomas (2005), when looking at one day matches in domestic English cricket, found a significant positive effect of winning the toss in one of the three logistic regressions undertaken. Forrest and Dorsey (2008) also found a statistically significant positive effect of the toss on individual matches and team relegation and promotion in English domestic four-day matches. Dawson *et al.* (2009) focused specifically on day-night ODIs and found that winning the toss and batting first increased the odds of a team winning the match. They also found that this effect increased during 1992 and 1993 when a new style of white ball was introduced. This result is supported by the findings in Bhaskar (2009) that winning the toss and batting during a day-night match significantly increases the chances of winning the match (Bhaskar, 2009, p. 18).

While the effect of the toss is the main focus of many of the above papers, Forrest and Dorsey (2008) also investigated the role of weather in affecting a team's end of season ranking and found significant effects, with weather disruptions reducing the probability of a match producing a winner. Dawson *et al.* (2009) and Saikia and Bhattacharjee (2010) also found some evidence that playing at home increased the chances of a team winning.

A broader literature review on sports other than cricket (particularly baseball) indicated that the effect of playing under lights or during the evening has not been the focus of any other papers. This is potentially due to the fact that, for most sports, the entire match will either be played during the day or during the evening leading to no systematic disadvantage for either team.

The papers outlined above mainly measure the effect of factors which could be broadly considered as luck (coin toss, weather and home field advantage). Analysis of factors related to luck is aided by the exogeneity of the factor being analysed: a coin toss is the essence of random allocation, weather is exogenous to a cricket match and home ground advantage normally rotates evenly between teams over time. This means that identification of the causal relationship between the factor and the outcome of the game is relatively simple.

In contrast, the effect of batting during the evening is more related to skill than luck and there are likely to be serious endogeneity concerns in identifying the effect. The first endogeneity concern is that teams in day-night matches systematically differ from teams in day only matches in terms of their quality. Second, some teams may prefer to bat during the evening while others may prefer to bat during the day. In the presence of these two factors, a regression model which simply contains an evening batting dummy variable would be confounding the effects of team quality, team batting preferences and batting during evening.

Considering team quality first, scheduling of matches in ODI cricket is coordinated by the International Cricket Council (ICC) and so matches are not randomly assigned to day only or day-night status. Non-random assignment may have to do with observable factors such as the availability of grounds with lighting (the first day-night match held in the Caribbean was in 2006, for example) (ESPN Cricinfo, 2012). However, the reason for non-random assignment is also likely influenced by unobservable variables such as the negotiating power of the teams that are playing, specific elements of broadcasting contracts and perceptions of how exciting the match will be. These factors could together be thought of as ‘quality’ with day-night matches possibly more likely to contain higher ‘quality’ pairings.

Figure 1 lends some support to the notion that teams of higher quality appear more frequently in day-night matches. The figure shows the 13 ODI teams that are currently ranked by the ICC, their average ranking points in the period May 1999 to December 2011 and the proportion of day-night matches that the team has been involved in. Overall, teams with more points (those that are better ranked) tend to play in more day-night matches. This relationship is statistically significant at the 1% level of significance.

Figure 1: Percentage of day-night matches and average ICC ranking points

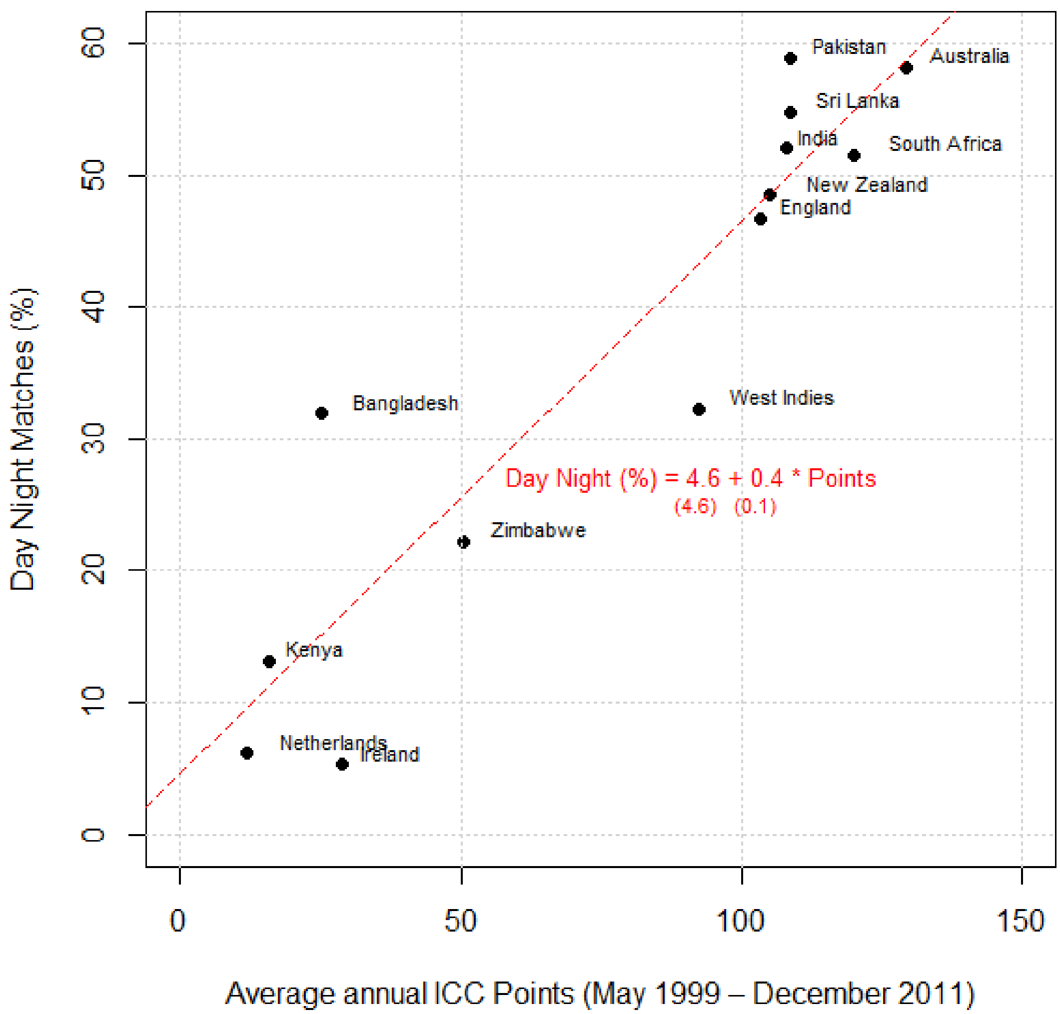


Figure 1 also suggests that there may be a way to measure team quality: through the ICC points data. The algorithm used to calculate ICC ODI points and rankings was introduced in 2003 and has subsequently been used to calculate points and rankings for earlier periods. Although providing an indication of team quality, the ICC points are designed to not take into account some important factors including the competition in which a match takes place, the location of a match and the margin of victory (Kendix, 2011). A higher quality team would be expected to perform better in each of these areas but, in the ICC points, this is not taken into account. Another factor not accounted for in the ICC points is the quality of a team's leadership and management (Mukherjee, 2012).

Given the imperfect measurement of team quality afforded by the ICC rankings, quality will be treated as an unobservable characteristic in the rest of this paper. ICC rankings are included as a proxy for team quality in a sensitivity test in section 7.

The second unobservable characteristic is a team's preference for batting during the evening. There are a number of reasons why teams may differ in their preference for batting during the evening. For example, some teams may prefer cooler batting conditions or may consider that fielding is more difficult under artificial light. Batting under lights may be seen as a particular advantage if dew is expected to form, as this is perceived to make spin bowling more difficult (Bhaskar, 2009). Table 1 shows a breakdown of batting decisions after winning the toss. It appears that, after winning the toss, there is variation in the strength of team's preferences for batting second. Although there is some evidence that teams differ in their preference for batting during the evening, this does not make it possible to control for this preference with team specific dummy variables. The preference for batting during the evening will likely be determined by many factors which are not recorded, such as pitch condition, as well as a number of unobservable factors such as the selected strategy, perceptions of the opposition and expectations about changes in the weather.

Table 1: Choice to field following coin toss win

	Number of matches		Coin toss wins		Elect to field following coin toss win		Percentage of matches electing to field following coin toss win		
	Day only	**Day-night**	**Day only**	**Day-night**	**Day only**	**Day-night**	**Day only**	**Day-night**	**Difference**
Canada	30	6	14	3	11	0	79	0	-79*
West Indies	151	72	76	23	54	5	71	22	-49***
Sri Lanka	124	148	57	79	35	11	61	14	-47***
South Africa	115	117	52	65	31	15	60	23	-37***
Netherlands	27	4	13	3	9	1	69	33	-36
England	127	112	66	64	41	17	62	27	-36***
Kenya	56	17	30	6	10	0	33	0	-33
Pakistan	107	156	59	76	28	12	47	16	-32***
India	142	171	80	81	47	22	59	27	-32***
Bangladesh	132	68	63	34	36	9	57	26	-31***
New Zealand	113	109	59	58	39	21	66	36	-30***
Australia	117	164	51	83	19	13	37	16	-22***
Zimbabwe	159	44	83	20	40	7	48	35	-13
Ireland	46	4	20	1	10	1	50	100	50

Note: One star indicates significance at the 10% level, two stars at the 5% level and three stars at the 1% level based on a test for difference in proportions. In this case, the difference is often less statistically significant when the sample size is small (such as Canada, Netherlands, Kenya and Ireland).

4. Controlling for the presence of unobservables

The likely presence of these endogeneity issues means that estimation of the effect of batting during the evening requires an identification strategy which can address the presence of the unobservable characteristics of team quality and preferences for batting during the evening.

Eliminating the effect of batting preferences can be done by only considering teams who have lost the toss. The team that wins the coin toss is able to select when they bat so they are able to express their preferences for batting. Losing the toss means a team does not get to express its preferences for batting. This will mean that, for the teams who lose the toss, there will be a random mix of batting preferences among teams batting first and teams batting second.

The difference in differences estimator can then be used to control for unobservable quality differences between day only and day-night matches. The structure of cricket matches, where there are two successive innings, allows the data to be broken up into four groups depending on whether the match is day only or day-night and whether it is the first or second innings. Just comparing first and second innings in a day-night match could be affected by strategic differences between innings (Stern, 2009). Comparing second innings between day only and day-night matches is likely to be affected by the endogeneity issues discussed above. The difference in differences estimator is based on a comparison of both of these differences and eliminates these measurement concerns.

Using the conventional terminology of Heckman *et al.* (1999) to express the idea more formally, the average treatment effect on the treated (ATT) of batting during the evening is the effect of interest. The ATT can be defined as:

$$E[\Delta_i | X_i, D_i = 1] = E\left[Y_{1,i} | X_i, D_i = 1\right] - E\left[Y_{0,i} | X_i, D_i = 1\right] \qquad (1)$$

Where E is the expected value, Δ_i is the treatment effect, X_i is a set of covariates, D_i is a dummy variable indicating team *i*'s treatment status (equal to one if the team is batting in the evening and zero otherwise) and Y_{ti} is the outcome for team *i* with treatment status equal to *t* (set to one when batting during the evening and zero otherwise).

To control for the likely presence of the unobservable factors discussed above, a difference in differences (DID) estimator is used. The DID estimator can be derived from equation 1 by adding and subtracting $E\left[Y_{0,i,1} | X_i, D_i = 1\right]$ which is the non-treatment outcome for the treated measured in the non-treatment period. The DID estimator is therefore defined as:

$$\begin{aligned} \mathrm{E}[\Delta_\mathrm{i}] = {} & \mathrm{E}\left[\mathrm{Y}_{1,\mathrm{i},2} | \mathrm{X}_\mathrm{i}, \mathrm{D}_\mathrm{i} = 1\right] - \mathrm{E}\left[\mathrm{Y}_{0,\mathrm{i},1} | \mathrm{X}_\mathrm{i}, \mathrm{D}_\mathrm{i} = 1\right] \\ & - \left(\mathrm{E}\left[\mathrm{Y}_{0,\mathrm{i},2} | \mathrm{X}_\mathrm{i}, \mathrm{D}_\mathrm{i} = 1\right] - \mathrm{E}\left[\mathrm{Y}_{0,\mathrm{i},1} | \mathrm{X}_\mathrm{i}, \mathrm{D}_\mathrm{i} = 1\right]\right) \end{aligned} \qquad (2)$$

Where D_i is a dummy variable indicating team *i*'s treatment status and Y_{tij} is the outcome for team *i* with treatment status equal to *t* and batting either first or second, according to *j*.

In order to estimate this from the data the following independence assumption is used:

$$D_i \perp\!\!\!\perp (Y_{1,i,2} - Y_{1,i,1}, Y_{0,i,2} - Y_{0,i,1}) \tag{3}$$

If this holds then the following relationship is true:

$$E[Y_{0,i,2}|X_i, D_i = 1] - E[Y_{0,i,1}|X_i, D_i = 1] = E[Y_{0,i,2}|X_i, D_i = 0] - E[Y_{0,i,1}|X_i, D_i = 0] \tag{4}$$

In this case, the independence assumption (equation 3) effectively means that if the day-night matches were rescheduled as day matches then there would have been the same differential in runs scored between teams who batted first and second. If this assumption holds then this means that the treatment effect can be estimated from observable data by substitution of equation 4 into equation 2 (Moffitt, 1991, p. 313). The difference in differences estimator can then be implemented directly using an ordinary least squares regression.

5. Data

To implement the identification strategy, a set of over-by-over data on 1,423 ODI matches played between May 1999 and December 2011 has been gathered. This gives 126,768 overs of data to work with. The data has been collated from commentary provided on the cricinfo website (ESPN Cricinfo, 2011). The data include descriptive match information such as date played, location, teams, whether it was a day-night game, who won the toss and which team batted first. It also contains, for each team, how many runs they have scored and how many wickets they have lost, both in total and at the end of each over.

The matches in the data set involve 24 different teams with a total of 616 day-night matches being played. However, not every team has batted in both the first and second innings of both a day match and a day-night match. It is desirable to exclude teams that do not satisfy this criterion in order to establish a strong common support in the data set. The teams in the common support are Australia, Bangladesh, Canada, England, India, Ireland, Kenya, Netherlands, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies and Zimbabwe. The sensitivity of the results to excluding other teams is tested in section 7. Data for matches where no result was recorded has also been eliminated. Similarly, data has been eliminated for eleven matches where there were inconsistencies in the data recorded by cricinfo, particularly where the official scorecards and the commentary disagreed on the runs scored.

After eliminating teams outside the common support, matches with no result and matches with errors, 14 teams, 1,319 matches and 117,915 observations are left. Of these remaining matches, 596 are day-night matches. Further restricting the data to teams who have lost the toss gives a final dataset containing 58,280 overs.

6. Model and results

Based on the identification strategy described in section 4, the reduced form model to be estimated is:

$$\mathrm{runs_i} = \beta_0 + \beta_1 \mathrm{day.night_i} + \beta_2 \mathrm{second.innings_i} + \beta_3(\mathrm{day.night_i} \times \mathrm{second.innings_i}) + \beta \mathrm{X_i} + \mathrm{u_i}$$

$$u_i \sim N(0, \sigma^2) \quad (5)$$

Where $runs_i$ is the number of runs scored in over i, $day.night_i$ is a dummy variable indicating whether the match observation was from a day-night match, $second.innings_i$ is a dummy variable indicating whether the observation was for a team batting second, X_i is a set of covariates and u_i is the random error which is assumed to be distributed normally. In this equation the treatment effect will be given by β_3.

A range of covariates are available for inclusion and so a number of different specifications of the model are estimated which include different covariates. These models range from a basic model, with no covariates, to models which incorporate non-linear effects and team specific dummies. This allows for an assessment of how sensitive the estimated treatment effect is to the inclusion (or exclusion) of particular covariates. The results from these models are shown in Table 2.

Table 2: Model results

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	4.85***	3.85***	3.75***	4.19***	3.67***	3.88***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Day.night	0.31***	0.26***	0.22***	0.24***	0.22***	0.14***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Second.innings	0.06	0.11***	0.10**	0.09**	0.17***	0.17***
	(0.17)	(0.00)	(0.01)	(0.02)	(0.00)	(0.00)
Day.night *second.innings	**-0.36*** **	**-0.24*** **	**-0.21*** **	**-0.22*** **	**-0.21*** **	**-0.21*** **
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
Over		0.10***	0.10***	0.10***	0.33***	0.32***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total.out		-0.42***	-0.41***			
		(0.00)	(0.00)			
At.home			0.25***	0.24***	0.24***	0.20***
			(0.00)	(0.00)	(0.00)	(0.00)
Region of play dummies			✓	✓	✓	✓
Wicket dummies				✓	✓	✓
Non-linear effects					✓	✓
Team dummies						✓
Reset Test p-value	NA	0.00	0.00	0.00	0.16	0.36
Breusch Pagan Test p-value	0.00	0.00	0.00	0.00	0.00	0.00

Note: P-values are shown in brackets below parameter estimates. P-values are based on robust standard errors. One star indicates significance at the 10% level, two stars at the 5% level and three stars at the 1% level.

Many of the covariates are included for clear reasons based on the structure and rules of the game. Overs and wickets provide a good example of this. To account for the expected non-linear effect of wickets lost, a model with wickets lost represented by dummy variables is included. Also, as overs and wickets are substitutes, it is potentially important to consider the interaction between these two factors in producing runs, therefore, an interaction term, with wickets lost defined as a discrete variable, is included.

Home-ground advantage is considered using a dummy as the home team is expected to be more familiar with the weather and ground conditions, and may benefit from crowd support (Dawson *et al.*, 2009). A set of dummy variables is also used to control for the region of play to take into account the potential for batting conditions to differ between countries, the regions included are Europe, the Indian subcontinent, Southern Africa, the Caribbean, Oceania and other. Finally, batting and fielding team dummies are included to adjust for the relative strength of particular teams. A version of the model which includes the square and cube of the over number, and interaction terms between this and the number of wickets lost, is also estimated in order to capture potential non-linear effects. This is particularly relevant given the non-linear functional form used by Duckworth and Lewis (1998).

One variable which would, ideally, be included is whether a "powerplay" is active in the current over. Although rules have varied over time, a powerplay activates fielding restrictions which are beneficial for the batsmen and it can be activated at various points through the innings (with certain restrictions). Comprehensive and consistent data on whether a powerplay was active or not was not available in the data gathered from Cricinfo. This is a potential area for future research.

Model 5 and 6 build from model 4, which is based on dummy variables for the number of wickets lost, rather than from model 3, which is based on a single variable for number of wickets lost. In model 4, the marginal effect of losing a wicket varies from around -0.15 runs per over (when the 8^{th} wicket is lost) to -0.90 runs per over (when the 1^{st} wicket is lost). This is not captured in model 3, where the effect is restricted to average -0.41 runs per over for each wicket lost. The dummy variable approach of model 4 is therefore preferred as it allows for variation in the marginal effect of losing a wicket.

The treatment effect is reasonably stable across all of the models that include covariates, varying from -0.24 to -0.21 runs per over (Table 2). The estimated treatment effect does increase as covariates are added to the model. This can be attributed to the additional covariates accounting for some of the variation in runs scored per over. For example, the move from model 5 to model 6 adds dummy variables for each team, this may help explain some variation in runs scored per over as there is variance in average scoring ability between teams.

To put the estimated treatment effect into context, over the course of a 50 over match a reduction in runs scored per over by 0.21 could result in a total reduction of around 10 runs. Of the 252 matches in the dataset where the team batting second in a day-night match lost the toss and the match, 24 (9.5%) of these matches have a margin of victory which is less than the disadvantage that is estimated for the team batting second. Further, in these 252 matches, the winning teams scored a total of 18,697 more runs than the losing teams and in these matches losing teams faced a total of 11,066 overs. The results of analysis indicate that the batting during the evening reduced the number of runs scored in these overs by 2,283 (*11,066×-0.21*). This means that, overall, the estimate explains an average of 12.2% of the margins of victory in these 252 matches. When considering the 95% confidence interval of the estimate, this ranges from 5.0% to 19.4%.

Most of the covariates behave as expected, are statistically significant and maintain their sign across the various specifications. In particular, the number of runs per over increases as the game proceeds (i.e. as over number increases), there is a home ground advantage of around 0.20-0.25 runs per over and losing wickets slows down the run rate (particularly lower down the batting order). Dummy variables representing the region of play were generally not statistically significant at conventional levels.

Considering some tests of the various models, to check for omitted interactions and higher powers we conducted a Ramsey test on each of the models. This test suggests that the models that do not contain non-linear effects may suffer from missing variables. The inclusion of non-linear effects for Over and Total Out corrects this. The statistical significance of these non-linear effects aligns with the non-linear functional form for the relationship between runs, overs and wickets identified by Duckworth and Lewis (1998) but does suggest the possible presence of multicollinearity. The table below shows the variance inflation factor (VIF) for selected parameters from models 4 and 6. The results indicate that multicollinearity is likely to be an issue for some variables (if a VIF of five is taken to raise concerns about multicollinearity). The presence of multicollinearity works to increase the estimated variance of the parameters and so, given that the key parameter estimates in the above regressions are highly statistically significant, it is unlikely that the possible presence of multicollinearity has had a practical effect on conclusions formed from the models' results.

Table 3: Variance inflation factors for selected models and variables

Parameter	Model 4	Model 6
Day.night	3.41	3.57
Innings.2	1.87	1.98
Day.night*second.innings	4.48	4.57
Over	2.43	140.08
At.home	1.03	1.07
Region of play dummies	1.43	5.62
Wicket dummies	2.45	1023.21

The Breusch-Pagan test indicated the presence of heteroskedasticity in all of the models and so the results in Table 2 report robust standard errors based on a White-corrected covariance matrix. Ultimately, the large sample size means that there is very little difference when using robust standard errors.

7. Sensitivity Analysis

Three additional models were also estimated to show the sensitivity of the treatment effect to restrictions placed on the data. These models were based on alterations to Model 6. The first model includes a variable for team quality based on the official ICC rankings. The results from this model indicate that an additional ICC ranking point increases expected runs scored per over by around 0.01, the effect is highly statistically significant. To put this into context, the best average ranked team, Australia, would be expected to score around 0.94 more runs per over on average and other variables constant than the lowest ranked team, the Netherlands. The key result from this model is that the estimated effect of batting during the evening only changes from -0.21 to -0.20 runs per over.

The second model run to test the sensitivity of the results expands the set of teams used from the 14 teams that form the common support to all teams that have played an ODI. This change increases the number of observations from 58,280 to 62,209. The estimated treatment effect decreases from -0.21 to -0.19 runs per over. Moving away from the common support should be expected to result in a smaller treatment effect as teams outside the common support are less likely to play in day-night matches and are also likely to score fewer runs per over. This will work to decrease the average runs per over scored in day matches and will dampen the measured treatment effect.

Table 4: Sensitivity test results

Parameter	Including ICC ranking	Including all teams	Including only test teams
Constant	2.84***	3.71***	3.72***
	(0.00)	(0.00)	(0.00)
Day.night	0.14***	0.13**	0.10***
	(0.01)	(0.01)	(0.00)
Innings.2	0.17***	0.16***	0.22***
	(0.00)	(0.00)	(0.00)
Day.night*second.innings	**-0.20*** **	**-0.19*** **	**-0.11*** **
	(0.00)	**(0.00)**	**(0.01)**
Over	0.32***	0.32***	0.33***
	(0.00)	(0.00)	(0.00)
At.home	0.19***	0.19***	0.17***
	(0.00)	(0.00)	(0.00)
Ranking	0.01***		
	(0.00)		
Region of play dummies	✓	✓	✓
Wicket Dummies	✓	✓	✓
Non-linear effects	✓	✓	✓
Team dummies	✓	✓	✓
Reset Test p-value	0.33	0.02	0.00
Breusch Pagan Test p-value	0.00	0.00	0.00

Note: P-values are shown in brackets below parameter estimates. P-values are based on robust standard errors. One star indicates significance at the 10% level, two stars at the 5% level and three stars at the 1% level.

The final specification used to test the sensitivity of the results reduces the set of teams used from the common support to only Test playing nations.[1] This change reduces the number of observations to 51,895. The estimated treatment effect changes from -0.21 to -0.11 runs per over. This is a much larger change than from the second sensitivity test where the sample was expanded. Restricting the analysis to test teams could be expected to result in a smaller treatment effect as test teams are likely to be more experienced playing in a range of conditions and to also be of higher quality, resulting in a reduced difference in performance between day only and day-night matches. This also suggests that non-test playing nations may be more disadvantaged by batting during the evening than test playing teams.

A further area which could be of interest is the effect of batting during the evening on wickets lost. This has not been the focus of this paper but wickets are more important in test cricket and, with the potential introduction of day-night test matches, there may be a need to consider the effect of batting during the evening on both run scoring ability and wickets lost. The results below are based on applying the same DID approach that was used in all other models in this paper but with wickets lost as the dependent variable. Given that wickets lost is a truncated count variable, a quasipoisson model was used to estimate the parameters. The results from the quasipoisson model shown below are essentially identical to those from a poisson model that was also run (results of which are not shown).

[1] The test playing nations in the dataset are: Australia, Bangladesh, England, India, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies and Zimbabwe.

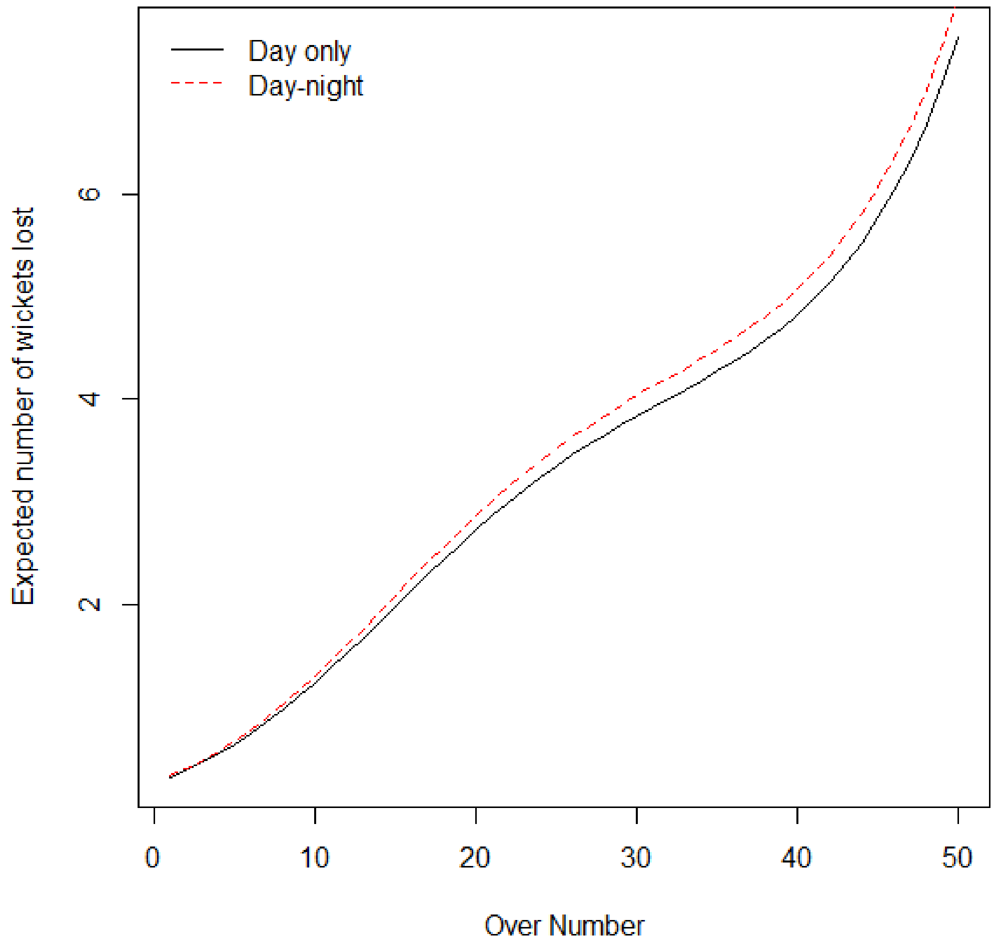
Table 5: Result for model with wickets lost as the dependent variable

Parameter	Parameter estimates
Constant	-1.36***
	(0.00)
Day.night	-0.04***
	(0.00)
Innings.2	-0.04***
	(0.00)
Day.night*second.innings	**0.09*** **
	(0.00)
Over	0.21***
	(0.00)
At.home	-0.04***
	(0.00)
Region of play dummies	✓
Non-linear effects	✓
Team dummies	✓

Note: P-values are shown in brackets below parameter estimates. One star indicates significance at the 10% level, two stars at the 5% level and three stars at the 1% level.

The results indicate that batting during the evening increases the expected number of wickets lost, this result is statistically significant at the 1% level. The use of the quasipoisson model makes direct interpretation of the marginal effects shown above difficult. To provide an indication of the scale of this effect, two sets of predicted values were generated from the match data. The first set had all matches set to day only status while the second set had all matches set to day-night status. The mean predicted value of wickets lost at each over was then calculated. The results are shown in the figure below.

Figure 2: Mean predicted values of wickets lost



The difference between day only and day-night matches increases throughout the match, reaching a peak at over 50 when teams batting during the evening are expected to have lost around 0.38 more wickets, on average, than teams batting during a day only match. This indicates that, while the estimated effect is statistically significant, it is unlikely to be strategically significant in the match.

8. Conclusion

After controlling for unobserved variables, it has been identified that a team who is forced to bat during the evening in cricket is disadvantaged by around 0.21 runs per over. To put this effect into context, over the course of a 50 over match this could amount to around 10 runs. Of the 265 matches in the dataset where the team batting second in a day-night match lost the toss and the match, 24 (9.5%) of these matches have a margin of victory which is less than the disadvantage that has been estimated for the team batting second. Overall, the estimate explains an average of 12.2% of the margins of victory in these 265 matches. When considering the 95% confidence interval of the estimate, this ranges from 5.0% to 19.4%.

This result suggests the potential need for a strategic and policy response. For captains the clear conclusion is to choose to bat if they win the toss in a day-night match – this already appears to be the preference of most teams (Bhaskar, 2009). For the ICC there may be a need to consider technical solutions, such as changes to lighting or the ball, as well as changes to rules, such as an additional powerplay for teams batting during the evening, to remove the disadvantage from batting during the evening. The most pressing issue is likely to be in the move towards day-night test matches. A test match would likely see both teams being exposed to batting during the evening, which would reduce the systematic nature of the disadvantage present in ODI matches. However, the results in this paper suggest that the introduction of batting during the evening into test cricket would introduce a new strategic aspect to the game. This dimension should be taken into account when considering the costs and benefits of introducing evening test matches.

References

Allsopp, P. E., & Clarke, S. R. (2004, November). Rating teams and analysing outcomes in one-day and test cricket. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 167*(4), 657-667.

Bell, I. (2007, August 21). At first we thought batting under lights was like wearing a blindfold. *The Guardian*. Retrieved from http://www.guardian.co.uk/sport/2007/aug/21/cricket.englandcricketseries3

Bhaskar, V. (2009, January). Rational Adversaries? Evidence from Randomised Trials in One Day Cricket. *The Economic Journal, 119*(534), 1-23.

Clarke, S. R., & Allsopp, P. (2001, April). Fair Measures of Performance: The World Cup of Cricket. *The Journal of the Operational Research Society, 52*(4), 471-479.

cricinfo. (2011, December 14). *Cricket must respect the fan - Dravid.* Retrieved February 20, 2012, from ESPN cricinfo: http://www.espncricinfo.com/ci/content/story/545358.html

Dawson, P., Morley, B., Paton, D., & Thomas, D. (2009). To bat or not to bat: An examination of match outcomes in day-night limited overs cricket. *Journal of the Operational Research Society, 60*(12), 1786-1793.

de Silva, B. M., & Swartz, T. B. (1997). Winning the Coin Toss and the Home Team Advantage in One-Day International Cricket Matches. *New Zealand Statistician, 32*(2), 16-22.

Duckworth, F. C., & Lewis, A. J. (1998, Mar). A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society, 49*(3), 220-227.

ESPN Cricinfo. (2011). *Records - One Day Internationals - Match Results*. Retrieved December 16, 2012, from ESPN Cricinfo: http://stats.espncricinfo.com/ci/engine/records/team/match_results.html?class=2;id=2011;type=year

ESPN Cricinfo. (2012). *Beausejour Cricket Ground*. Retrieved December 16, 2012, from ESPN Cricinfo: http://www.espncricinfo.com/westindies/content/ground/59518.html

Forrest, D., & Dorsey, R. (2008). Effect of toss and weather on County Cricket Championship outcomes. *Journal of Sports Sciences, 26*(1), 3-13.

Heckman, J. J., Lalonde, R. J., & Smith, J. A. (1999). he economics and econometrics of active labor market programs. In D. Card, *Handbook of Labor Economics* (Vol. 3A, pp. 1865-2097). Amsterdam: North-Holland.

Kendix, D. (2011, June). ODI and Test Team Rankings: How they work and their suitability for determining tournament qualification. Retrieved June 17, 2012, from http://static.icc-cricket.yahoo.net/ugc/documents/DOC_BA4FD677CDC62391608CA0F6662C7DEB_1338990291442_280.pps

Moffitt, R. (1991). Program evaluation with non-experimental data. *Evaluation Review, 15*(3), 291-314.

Morley, B., & Thomas, D. (2005). An investigation of home advantage and other factors affecting outcomes in English one-day cricket matches. *Journal of Sports Sciences, 23*(3), 261-268.

Mukherjee, S. (2012, January 9). Identifying the greatest team and captain − A complex network approach to cricket matches. *Physica A – Statistical Mechanics and its Applications, 391*(23), 6066-6076.

Saikia, H., & Bhattacharjee, D. (2010). On the Effect of Home Team Advantage and Winning the Toss in the Outcome of T20 International Cricket Matches. *Assam University Journal of Science and Technology, 6*(2), 88-93.

Stern, S. E. (2009). An adjusted Duckworth-Lewis target in shortened limited overs cricket matches. *Journal of the Operational Research Society, 60*(2), 236-251.

Turbervill, H. (2011, September 12). First day-night County championship game between Kent and Glamorgan fails to wow crowds. *The Telegraph*. Chatham, Kent, England. Retrieved February 20, 2012, from http://www.telegraph.co.uk/sport/cricket/counties/8758981/First-day-night-County-championship-game-between-Kent-and-Glamorgan-fails-to-wow-crowds.html