The behaviour of betting and currency markets on the night of the EU referendum

Tom Auld
Oliver Linton

The behaviour of betting and currency markets on the night of the EU referendum[*]

Tom Auld[†], Oliver Linton

31 October, 2017

Faculty of Economics, University Of Cambridge, The Austin Robinson Building, Sidgwick Ave, Cambridge, CB3 9DD, UK

Abstract

We study the behaviour of the Betfair betting market and the sterling/dollar exchange rate (futures price) during 24 June 2016, the night of the EU referendum. We investigate how the two markets responded to the announcement of the voting results. We employ a Bayesian updating methodology to update prior opinion about the likelihood of the final outcome of the vote. We then relate the voting model to the real time evolution of the market determined prices. We find that although both markets appear to be inefficient in absorbing the new information contained in vote outcomes, the betting market is apparently less inefficient than the FX market. The different rates of convergence to fundamental value between the two markets leads to highly profitable arbitrage opportunities.

Keywords: EU Referendum, Prediction Markets, Machine Learning, Efficient Markets Hypothesis, Pairs Trading, Cointegration, Bayesian Methods, Exchange Rates

1 Introduction

Were currency and prediction markets efficient overnight on 24 June 2016 as the results of the United Kingdom European Union membership referendum were announced?

This question is important as the EU referendum was one of the great political shocks of 2016. The results of the vote itself provide for a unique period in market history for which both financial and prediction market efficiency can be studied. The night is a special event for a number of reasons. Firstly, referendums are rare events with no similar votes in history for market participants to base expectations[1]. There was also a strong prior belief that the UK would vote to remain in the European Union. This provided fertile ground for inefficiencies and behavioural biases to arise. Secondly, there were 382 different voting areas that announced results at different times. The results were immediately widely distributed during an overnight period in which no other market relevant announcements were made. As such, the sole determinant of prices (the area results) was well known and quantitative[2] for a period of a few hours. Thirdly, there are 2 markets to study, a prediction market in the Betfair betting market and the pound dollar currency market.

The Efficient Market Hypothesis (EMH) holds that financial markets immediately reflect all available information in prices. If this is true, investors cannot receive above market returns except by chance. The weakest form of the hypothesis relates only to historical price information. Opinion is split on whether this form holds. Stronger forms relating to both fundamental (semi-strong) and private information (strong) also exist. Most studies conclude that the stronger forms of the EMH do not hold. For the night of the EU referendum, one existing working paper concludes that the pound market was slow to reflect the information contained in the vote results and hence the EMH in the semi-strong form did not hold. Regarding prediction markets, there is a consensus in the literature that prediction markets provide better estimates of future events than experts, and that the predictions of such markets are useful in a variety of situations.

This paper makes a number of contributions: This is the first high fre-

---

[1]There have only been two other UK wide referendums. The first, the European Communities membership referendum held in 1975, would be of little use for inferring voting patterns today. The other, on an unrelated subject, was the Alternative Vote Referendum in 2011 and had a turnout of only 42.2%, opposed to a typical figure of $60 - 70\%$ for general elections

[2]The percentage of those voting to leave the EU, as well as the percentage of those from the eligible electorate who voted were released for each of the voting areas

quency study in the literature comparing a prediction market with a financial market. We agree with other work that concludes that the EMH in semi-strong form did not hold in the currency market during the night of the referendum, but we demonstrate this was also the case for a prediction market. Further, we show that the Betfair market was more efficient than the sterling market, which provides some support for the view that prediction markets yield useful predictions. Small sample inference is required to predict Brexit early on in the night of the vote and we improve upon earlier prediction methods by using a rigorous Bayesian approach that is valid for small samples. Finally, we demonstrate that the different rates at which the vote was reflected in the two markets led to arbitrage opportunities, implying a failure of the EMH in the weak form.

The remaining sections of this paper are organized as follows. The next section reviews the literature. Section 3 discusses the data we use. In Section 4 we present the electoral model updating methodology, which employs Bayesian Machine Learning. Section 5 presents a theoretical model linking the behaviour of the two asset prices (Betfair contracts and Sterling futures) under standard economic assumptions. In Section 6 we present our empirical results. Section 7 concludes.

2 Literature Review

This is a study of market efficiency, and its manifestation in a prediction and a financial market on the night of the EU referendum. We summarize the debate concerning the Efficient Market Hypothesis (EMH) and refer to two studies of referendums recently held in the UK. However a comprehensive review of each field is beyond the scope of this paper. The reader is instead referred to Horn et al. (2014) for a review of prediction markets and Graefe (2016) for a specific review of political markets. There is a growing consensus in the literature that political prediction markets are more accurate at forecasting elections than are polls or experts.

The EMH (Fama (1965)) states that the prices of financial assets immediately discount all available information and therefore investors cannot make above average returns, except by chance. There are various forms of the hypothesis. In the Weak form, financial prices instantaneously discount all *market* information; in the Semi-Strong form, prices instantaneously discount all *publicly available* information; in the Strong form, prices instantaneously discount all information *both public and private*, including privileged information available to insiders. Many authors (eg Malkiel (2003)) argue that the EMH does not imply that pricing is perfect or that mispricings never oc-

cur, just that mispricings are random and it is not possible to systematically profit from them in advance.

There are various behavioural explanations that attempt to explain why the EMH may not hold. For a recent comprehensive review see Huang et al. (2016). One such theory is that of investor inattention, which is potentially relevant to us as results were announced outside of major market times. See Hirshleifer et al. (2013), Hirshleifer et al. (2009) and Hou et al. (2009) for examples, as well as DellaVigna & Pollet (2009) where the authors claim to show that earnings announcements on a Friday take longer for the market to react to. Other behavioural explanations include anchoring and systematic overconfidence. Another idea presented in Caballero & Simsek (2016) postulates that a study of any anomalies of the EMH require an analysis of the presence or absence of any arbitrage process that may exist to bring prices rapidly back to the "correct" value. There have been many opinions and studies published on the EMH and no consensus exists as to its validity. For a recent review, see GabrielaTitan (2015).

There are many studies of betting markets and their ability to predict future events including elections (Wolfers & Zitzewitz (2004)). A number of studies have evaluated the efficiency of sports betting markets, mostly finding some inefficiency (Goddard & Asimakopoulos (2004); Vlastakis et al. (2009); Badarinathi & Kochman (1996)). There are relatively few studies of referendums because they are unique events and require event-specific approaches. During the preparation of this paper a related study was published (Wu et al. (2017)) which investigated the real time response of the exchange rate to the announced vote outcomes. Their conclusion was that the "Brexit result could have been predicted with high confidence under realistic conditions". Examining social and psychological factors as well as Betfair data prior to the vote, the authors conclude that the mispricing "indicates both generic inefficiency and a specific inertia / durable bias in the market similar to herding during bubbles". The paper also examines trading behaviour in the pound around the announcement of specific results. We agree with the conclusion of a mispricing in the pound market. However, there are some shortcomings to the methodology. This concerns the use of OLS estimators with small number of samples. These considerations are explored formally in B. We improve upon this work by using a Bayesian electoral model valid for small sample inference, by using real-time Betfair price data from the night of the vote and introducing a theoretical model to relate the prices of the two markets. One other study that uses political Betfair data is that Wall et al. (2017). This study does not examine the efficiency of the betting market but instead relies on the largely accepted premise that prediction markets can provide meaningful forecasts of outcomes. Betfair data is used to control

4

for polling shocks and isolate campaign event effects in the 2014 Scottish Independence campaign in the months leading up to the referendum.

3 Data

The model uses the following data:

- EU Referendum Results: sourced from Electoral Commission (EC) website (Electoral Commission (2016)).

- Timing Results: we use the earliest confirmed time for each voting area from the three sources below:

  - Press Association: the time, to the nearest second, that each result was received and processed at the Association was used. This was a small number of seconds after it was publicly announced at the count.

  - Electoral Commission: returning officers for each area both inform the EC both just prior to announcement and immediately afterwards. The EC has made both times for each area available to us to the nearest second. This dataset provides a window within which each area's result must have been made public.

  - Bloomberg: 67 of the 382 results were published in real time on the Bloomberg terminal and the timestamps to the nearest second have been retrieved. Although only a small subset of the total, these results can be assumed to contain the most market-sensitive, and hence most informative, information.

- Priors for each voting area: prior to the vote the psephologist, C. Hanretty of Norwich University, published a blog titled "The EU referendum: what to expect on the night" (Hanretty (2016)). Expectations and 90% confidence intervals (CIs) for vote share based upon publicly available data could be downloaded for all but 4 of the 382 voting areas[3]. The work was reported in the media, including on the popular financial platform Bloomberg. It is reasonable to believe that market participants were aware of this information and had ready access to it. We do not reproduce this work, but we do to make use of it. The priors were based on a panel data analysis of the British Electoral Study

---

[3]Northern Ireland, the Isles of Scilly, Anglesey and Gibraltar were excluded due to the lack of availability of local authority demographic data.

(BES) from 2015 and demographic results at the local authority level. Firstly the priors are calculated directly from the BES, and secondly, a uniform swing is applied to each area to bring the results in line with polling information available on the date of Hanretty's publication (7 June). Various assumptions were used to generate the dataset. Hanretty characterises them as follows:

- the geographic patterns of the Leave and Remain votes have remained constant since May of [2015] (though the levels may have changed)
- that undecideds will break in roughly even numbers between the two camps
- Leave- and Remain-voting areas will vote at roughly equal rates.

- We make particular use of YouGov's poll on the day (YouGov (2016)). We also use other polls where there are gaps in Hanretty's priors.

- Historical General Election Data: we make extensive use of constituency level data for preceding general elections from the EC (Electoral Commission (2017)) and the website Electoral Calculus (Baxter (2017)). We also use historical polling information and measures of their accuracy for general elections. This was obtained from Wikipedia (Wikipedia (2017b)).

- GBPUSD Futures Price: we decided to use the GBPUSD future price traded on the Chicago Mercantile Exchange rather than the spot price. There are multiple exchanges where the spot trades and aggregation could be prohibitively difficult. It is well known that spot and futures prices for foreign exchange are extremely well correlated and are effectively contemporaneous on time scales of under one second. The data was downloaded from Bloomberg and timestamps of trades were reported to an accuracy of one second. Note that the futures contract was closed between 10 pm and 11 pm on 23 June, but before the announcement of results.

- Betfair Data: The betting website Betfair listed two contracts. These were traded on Betfair's exchange platform which acts as a limit order book. The first paid out £1 in the event of Brexit, the other paid £1 in the event of Remain. The sum of the prices of the contracts did not deviate sufficiently from £1 to enable a profitable arbitrage. Betfair supplied all trades with timestamps of one second granularity in both contracts between 10 pm on 23 June to 5 am on 24 June. We convert
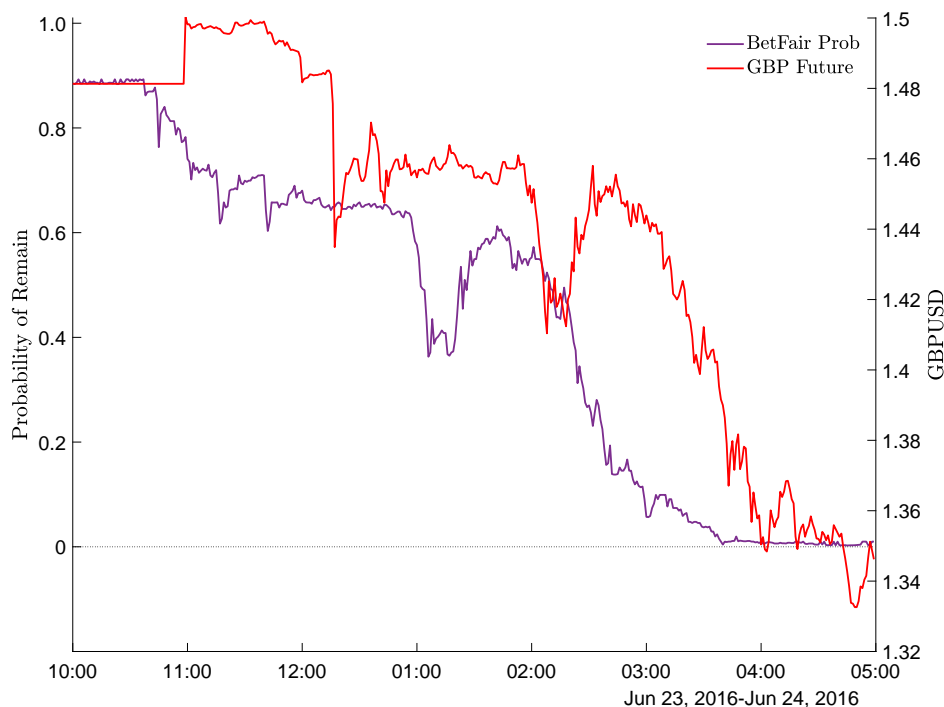
Figure 1: The Pound and Betfair Markets Overnight for the Period of the Announcement of Results.

all prices in the Remain contract to a synthetic price in the Brexit one by subtracting from £1. The resulting Betfair data set, along with the GBPUSD future price, is shown in Figure 3. We conduct analysis on the combined set of trades, which number 182,534. £51,016,907 in total was matched during this 7 hour window. This compares with 88,246 trades in the GBP future during this time with a total notional traded of around $5.5Bn.

4 Electoral Model

In this section we present a Bayesian model for calculating an implied probability of Brexit. The model updates as new area results are announced throughout the night of the referendum. A summary of the model is presented in A.

4.1 Setup

We have n constituencies with sizes $s_1, s_2, \ldots, s_n \in \mathbb{N}$, ordered by time result. Suppose $p_i, q_i \in [0, 1]$ are the proportion of voters in favour of leaving the EU and the turnout percentage in constituency i. Then the proportion of the national vote is:

$$p_N = \sum_{i=1}^{n} p_i q_i s_i$$

and the event of leaving the EU occurs when:

$$p_N > \frac{1}{2}$$

4.2 Gaussian Copula Prior

In order to determine a real-time probability of Brexit we use a prior. This is updated via Bayes' rule as results arrive to give a distribution of unannounced results conditional on announced results. The variables in the model are the $2n$ values of vote share and turnout of the vector $r = (p_1, \ldots, p_n, q_1, \ldots, q_n)^\mathsf{T}$. Prior to the vote, Hanretty had effectively published marginal distributions for the individual variables. The available data naturally leads us to take a copula approach and this is how we will proceed.

Given CDFs for the unknown variables $\{F_{r_1}(r_1), \ldots, F_{r_{2n}}(r_{2n})\}$, the quantiles of the variables are the values of the CDFs and are uniformly distributed. Writing the quantiles as $X_i = F_{p_i}(p_i)$ and $Y_i = F_{q_i}(q_i)$, the Copula defines the dependence structure between the variables and is defined as:

$$C(x_1, \ldots, x_n, y_1, \ldots, y_n) =$$
$$P(X_1 < x_1, \ldots, X_n < x_n, Y_1 < y_1, \ldots, Y_n < y_n)$$

Sklar's Theorem states that any multivariate distribution which has continuous marginals can be expressed in terms of a unique copula.

We model the dependence using the multivariate Gaussian copula over $[0, 1]^{2n}$. Writing the $2n$ quantiles of r as a vector $z = (F_{r_1}(r_1), \ldots, F_{r_{2n}}(r_{2n}))^\mathsf{T}$ and the correlation matrix Σ_0, this is defined as:

$$C_{\Sigma_0}^{Gauss}(z) = \Phi_{\Sigma_0}(\Phi^{-1}(z_1), \ldots, \Phi^{-1}(z_{2n})),$$

where Φ_Σ is the CDF of a multivariate normal distribution with zero mean and covariance matrix Σ. $\Phi(\bullet)$ is the usual CDF of the $N(0, 1)$ distribution. We model the dependence between the variables as a two-factor model,

the factors being vote share and turnout, with correlations between factors allowed, specifically

$$\Sigma_0 = \begin{pmatrix} \Sigma_p & \Sigma_{pq} \\ \Sigma_{pq}^T & \Sigma_q \end{pmatrix},$$

where Σ_p is an $n \times n$ matrix describing the dependence of $(p_1, \ldots, p_n)^{\mathsf{T}}$ given a correlation ρ_p and is

$$\Sigma_p = (1 - \rho_p) I_n + (\rho_p) i_n i_n^{\mathsf{T}}.$$

Similarly, Σ_p is the correlation matrix of $(q_1, \ldots, q_n)^{\mathsf{T}}$ given correlation ρ_q and is

$$\Sigma_q = (1 - \rho_q) I_n + (\rho_q) i_n i_n^{\mathsf{T}}.$$

To model the correlation ρ_{pq} between turnout and vote share, the matrix Σ_{pq} is defined as

$$\Sigma_{pq} = \rho_{pq} \times \left[(1 - (\rho_q \rho_p)) I_n + (\rho_q \rho_p) i_n i_n^{\mathsf{T}} \right].$$

4.3   Prior Probability of Brexit

Given specifications for the dependency Σ_0 and marginal distributions

$$\{F_{r_1}(r_1), \ldots, F_{r_{2n}}(r_{2n})\}$$

the prior probability of Brexit is

$$\Pr(BREXIT)_0 = \Pr\left(\frac{\sum_{i=1}^n p_i q_i s_i}{\sum_{i=1}^n q_i s_i} > \frac{1}{2} \right), \tag{4.1}$$

where:

$$(p_1, \ldots, p_n, q_1, \ldots, q_n)^{\mathsf{T}} = r$$
$$\Phi^{-1}(F_1(r_1)), \ldots, \Phi^{-1}(F_{2n}(r_{2n})) \sim N(0, \Sigma_0).$$

As there is no analytical form for the integral in equation (4.1), a sampling method is required for evaluation.

4.4 Marginal Distributions

Hanretty provided expectations and 90% CIs for the marginals. These CIs are implied by responses of the BES coupled with Local Authority demographic data. They do not take into account the uncertainty of the national vote. There may be an argument that the distributions that Hanretty supplies as part of his econometric analysis of panel and local authority data are asymptotically normal. No such argument can be made once the uncertainty of the national vote share is taken into account. We interpret the prior as an expression of a degree of belief about the possible values vote shares could take. As such, we are not constrained by the normal distribution.

4.4.1 Marginal Calibration

To first calculate the expectations of the marginals, suppose μ_H is the vector of expectations provided by Hanretty. Given an expected level for the national average vote share μ_N, then μ_p can be formed by applying a uniform shift, $\alpha_N \times i$, to μ_H where μ_H are Henretty's expectations[4]. Note that for $\rho_{pq} \neq 0$, α_N is *not* exactly equal to μ_P less the share implied by μ_H weighted by $\nu_i s_i$, as the national vote share is a sum of the p_i weighted by q_i which are generally correlated[5].

We present two ways of adjusting for the uncertainty of the marginal distributions. Firstly, we can assume that investors have rational expectations and that the EMH holds. This leads us to set the overall level of variance so as to produce a prior model probability equal to that in the Betfair market prior to results being announced, which we label P_0. Another approach is to give up on the EMH and attempt a forecast by setting a level of variance of p_N, σ_p^2 equal to a generous estimate of what we think it could be, $\hat{\sigma}_N^2$. Either way, we add a constant variance σ_N^2 to each marginal variance and adjust σ_N^2 to achieve either result.

Firstly, we convert the CIs assuming a normal distribution as follows:

$$(\sigma_H^2)_{p_i} = ((90\% \text{ Confidence Interval})_i/(2 \times 1.645))^2,$$

where $(\sigma_H^2)_i$ is now the unadjusted variance implied for the i'th voting area. The national vote share is uncertain and treating this variation as independent of the idiosyncratic variances implied by Hanretty's values leads us to

---

[4]Note, Henretty himself forms μ by applying a uniform shift to the priors he calculates from the BIS and census data to agree with polling data at the time of his publication

[5]For instance if $\rho_{pq} > 0$, $E(\frac{\sum_i p_i q_i s_i}{\sum_i q_i s_i}) > \frac{\sum_i \mu_i \nu_i s_i}{\sum_i \nu_i s_i}$ since when p_i is above its mean it will tend to have a higher turnout and hence weight in the sum.

add to each area variance a constant variance, say σ_N^2,[6] whereby

$$\sigma_{p_i}^2 = (\sigma_H^2)_i + \sigma_N^2$$

The prior mean for p_N is below 50 so the part of the national vote share distribution that lies above 50%, which is the model probability of Brexit, will be greater given higher variance σ_N^2. Note that this calibration of σ_N^2 will have to be conducted only once the other prior parameters have been chosen. For example a higher value of correlation between areas ρ_p will yield a wider national vote share distribution and will thus require a smaller value of σ_p^2 for a given value of P_0.

For turnout we base area level expectations, say ν, on historical general elections. However we choose the variances for each area to be equal and labelled by σ_ν^2.

We evaluate the model for different marginal distributions. We consider the following:

4.4.2 Normal

$$r_i \sim N(\mu_i, \sigma_i^2)$$

$$F_i(r_i) = \Phi\left(\frac{r_i - \mu_i}{\sigma_i}\right)$$

This distribution is computationally cheap as it does not require the evaluation of the CDF or inverse. The prior is actually simply a multivariate Gaussian. However, a disadvantage is that it does not restrict the random variables p_i, q_i to $[0, 1]$.

4.4.3 Logit Normal

$$logit(r_i) \sim N(\mu_i, \sigma_i^2)$$

$$F_i(r_i) = \Phi\left(\frac{logit\,(r_i) - \mu_i}{\sigma_i}\right)$$

---

[6]We are not implying that area vote shares are independent of the national vote share, just that the idiosyncratic variation implied by Hanretty's study of survey respondents and local authority data is independent of the variation of the national vote share number. Our assumptions are not even that strong; as we apply a uniform shift to variances we are simply implying that the *difference* in the variations of the marginal distributions of individual areas is the same as the difference implied by Hanretty's study.

We can apply the logit transform to the variables to convert them to the real line, and then assume a normal distribution under that transformation. Again, a multivariate normal distribution is implied for the transformed variables. As there is no analytic solution to the moments a sampling method could be used. However, we will use a simple transformation which was found in practice to make no difference to the results as follows:

$$\mu_i = logit^{-1}(\mu_i)$$

$$\sigma_i = \frac{[logit\,(\mu_i + \sigma_i) - logit\,(\mu_i - \sigma_i)]}{2}$$

Note that as the logit function is only symmetric around the value 0.5, the above transformations will not (quite) preserve the differences in expected values of vote share. However, in practice the differences were not found to be meaningful.

4.4.4 Beta

$$r_i \sim Beta(\alpha_i, \beta_i)$$

$$F_i(r_i) = I_{r_i}(\alpha_i, \beta_i) = \int_0^{r_i} t^{\alpha_i - 1} (1 - t)^{\beta_i - 1} \, dt$$

The Beta distribution is restricted to $[0, 1]$ and is parameterized by two shape parameters α and β. For a mean of μ_i and variance of σ_i^2, the shape parameters will be equal to:

$$\alpha_i = \left(\frac{1 - \mu_i}{\sigma_i^2} - \frac{1}{\mu_i} \right) \mu_i^2$$

$$\beta_i = \alpha_i \left(\frac{1}{\mu_i} - 1 \right)$$

4.4.5 Logit Student with Location and Scale

$$logit(r_i) \sim t_\nu(\mu_i, \sigma_i^2)$$

$$F_i(r_i) = \tfrac{1}{2} + \left(\frac{logit(r_i) - \mu_i}{\sigma_i} \right) \Gamma \left(\frac{\nu+1}{2} \right) \times \frac{{}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}, \frac{3}{2}, -(logit(r_i) - \mu_i)^2 / (\sigma_i^2 \times \nu) \right)}{\sqrt{\pi \times \nu} \Gamma\left(\frac{\nu}{2} \right)}$$

The Student's t-distribution has a higher kurtosis than the normal distribution and including it enables us to study priors with greater fourth

moments for a given variance. The mean is simply μ whereas the variance, for $\nu > 2$, is $\frac{\nu}{\nu-2} \times \sigma$. We set the parameters in a similar way to the logit normal marginal, but with scale parameter σ_i

$$\sigma_i = \frac{\nu - 2}{\nu} \times \frac{[logit\,(\mu_i + \sigma_i) - logit\,(\mu_i - \sigma_i)]}{2}$$

The excess kurtosis is ∞ for $\nu \in (2, 4]$ and $\frac{6}{\nu-4}$ for $\nu > 4$. To explore the implications of infinite fourth moments (for $logit(r_i)$ not r_i) we set $\nu = 3$.

4.5 Update

Calculation of the conditional distribution is most easily done by a re-ordering of the variables:

$$\tilde{r} = (p_1, q_1, \ldots, p_n, q_n)^\top .$$

Then

$$\Phi^{-1}(\tilde{F}_1(\tilde{r}_1)), \ldots, \Phi^{-1}(\tilde{F}_{2n}(\tilde{r}_{2n})) \sim N(0, \tilde{\Sigma}_0),$$

where $\tilde{x}$ indicates a similar re-ordering of rows and columns of x. To calculate the conditional distribution of the remaining variables after m results have been announced, partition $\tilde{\Sigma}_0$ into four block matrices as follows:

$$\tilde{\Sigma}_0 = \left(\begin{array}{cc} \tilde{\Sigma}_{m,m} & \tilde{\Sigma}_{m,\backslash m} \\ \tilde{\Sigma}_{\backslash m,m} & \tilde{\Sigma}_{\backslash m,\backslash m} \end{array} \right),$$

where: $\tilde{\Sigma}_{m,m}$, $\tilde{\Sigma}_{m,\backslash m}$, $\tilde{\Sigma}_{\backslash m,m}$ and $\tilde{\Sigma}_{\backslash m,\backslash m}$, are $2m \times 2m$, $2m \times 2(n-m)$, $2(n-m) \times 2m$ and $2(n-m) \times 2(n-m)$ matrices respectively. The multivariate Gaussian copula provides a simple update of the conditional distribution to yield another Gaussian copula as follows. Given the observations $p_1, q_1, \ldots, p_m, q_m$, write

$$\tilde{x}_m = \left(\tilde{F}_1(\tilde{r}_1)), \ldots, \Phi^{-1}(\tilde{F}_{2m}(\tilde{r}_{2m})) \right).$$

Then

$$\Phi^{-1}(\tilde{F}_{2m+1}(\tilde{r}_{2m+1})), \ldots, \Phi^{-1}(\tilde{F}_{2n}(\tilde{r}_{2n})) \,|\, \tilde{r_m} \sim N(\tilde{\Pi}_{\backslash m}, \tilde{\Sigma}_{\backslash m}),$$

which is a Gaussian copula with non-zero mean $\tilde{\Pi}_{/m}$ and covariance matrix $\tilde{\Sigma}_{/m}$ given by:

$$\tilde{\Pi}_{\backslash m} = \tilde{\Sigma}_{\backslash m,m} \tilde{\Sigma}_{m,m}^{-1} \tilde{x}_m$$
$$\tilde{\Sigma}_{\backslash m} = \tilde{\Sigma}_{\backslash m, \backslash m} - \tilde{\Sigma}_{\backslash m,m} \tilde{\Sigma}_{m,m}^{-1} \tilde{\Sigma}_{m,\backslash m}$$

As $p_1, \ldots, p_r, q_1, \ldots, q_r$ is now known, the model probability of Brexit can be computed by simulation via:

$$\mathrm{Pr}(\mathrm{BREXIT})_m = P\left(\frac{\sum_{i>m} p_i q_i s_i}{\sum_i q_i s_i} > \frac{1}{2} - \frac{\sum_{i \le m} p_i q_i s_i}{\sum_i q_i s_i} \right).$$

An advantage of the model is that it provides for closed form updates to the posterior distributions of the parameters, as the Gaussian copula has a conditional distribution. This avoids the need for a Markov Chain Monte Carlo sampling technique to calculate the integral in equation (4.1). This would be particularly arduous given the large number (2×382) of variables involved. An alternative copula, with well understand and closed form conditional distributions, is the Student's t copula Ding (2016). The conditional distribution is also a Student's t copula and will have fatter joint tails from the Gaussian. It would be interesting to see if there were any effects of the speed of convergence by changing to a t copula, but such a study is beyond the scope of this work. We note, however, that our framework does at least allow for higher kurtosis distributions in the national vote p_N through the use of the logit Student's t distribution with degrees of freedom above but close to 2. Another potential avenue for future exploration would be to use a marginal distribution with a *low* kurtosis. A Generalized Normal Distribution with a β parameter of between 3 and 8, for instance, would be an excellent candidate.

We now comment on the expected qualitative impact of the model as parameters change. For purely independent vote share results ($\rho_p, \rho_q = 0$), convergence will solely be due to the results as they come in and the distributions of the yet to be announced results will not be affected. For higher values of ρ_p, convergence will be faster. It is expected that the value of ρ_p and the variance σ_N^2 will have the greatest effect on the speed at which predictions change. The effect of ρ_q and ρ_{pq} on the model probability are effectively second order. Given that we will be setting ρ_{pq} as negative and turnout was above expectations, there will be some small second order effects from changes in the other parameters. Lowering ν or σ_v^2 will slow the speed of convergence whereas lower ρ_q and lower $|\rho_{pq}|$ will speed convergence, but these effects should be very small.

4.6 Parameter Choices

In an ideal hypothetical world we would have a sequence of Brexit referendums on which we could fit the covariance structure of the results Σ_0 using a method such as Maximum Likelihood or Feasible Generalized Least Squares. Unfortunately, we have no such data. We rely instead on general elections and polling data to chose plausible parameter values. We attempt to rely only on information available in the public domain prior to the announcement of results and will do our best to avoid any hindsight bias in the analysis.

4.6.1 Turnout

National Turnout There were reports of high turnout on the day of the vote itself (Gutteridge (2016)). We will use national turnout for general elections as a guide, but note that the Scottish independence referendum had an unprecedentedly high turnout of 85%. The general election turnout figures since 1945 are shown in Table 4.6.1. The average is 66.9% (6.7%) and for the last three elections the average is 64.2%. We use 67.6% which is the three-election average weighted upwards by half the six-election standard deviation. A reasonable range of expectations would be $65 - 70\%$.

Area Turnout Voting regions for the EU referendum were not the same as the constituencies used for general elections. However, the EC categorizes both the 381 voting areas in the referendum (excluding Gibraltar) and the (most recently 650) general election constituencies by 12 region codes. This enables us to make a more granular estimate of turnout per area v_i than simply assuming a uniform expectation. We use average turnout for each region for the 2010 and 2015 general elections as outlined in Table 4.6.1. Similar to the means of the expected vote share per region, $(\nu_1, \ldots, \nu_n)$ can be uniformly shifted to achieve the required expected national turnout.

Turnout Variance Instead of setting turnout variance by region we will simply use the same level for every area and use the standard deviation figure for the last six general election 6.7%.

4.6.2 Turnout Correlation by Area

Given elections in time periods $t = 1, \ldots, T$ and turnouts q_{it}, if we have predictions in advance for q_{it}, $\bar{q}_{it}$ then we can model the prediction errors $\triangle q_{it} = q_{it} - \bar{q}_{it}$ as being due to a national error ϵ_t and individual error terms, η_{it} where

Election Year	England	Wales	Scotland	N. Ireland	UK
2015	65.8%	65.7%	71.1%	58.1%	66.1%
2010	65.5%	64.7%	63.8%	57.6%	65.1%
2005	61.3%	62.6%	60.8%	62.9%	61.4%
2001	59.2%	61.6%	58.2%	68%	59.4%
1997	71.4%	73.5%	71.3%	67.1%	71.4%
1992	78%	79.7%	75.5%	69.8%	77.7%
Standard Deviation UK, 1992 - 2015					6.7%
1987	75.4%	78.9%	75.1%	67%	75.3%
1983	72.5%	76.1%	72.7%	72.9%	72.7%
1979	75.9%	79.4%	76.8%	67.7%	76%
1974 Feb	79%	80%	79%	69.9%	78.8%
1974 Oct	72.6%	76.6%	74.8%	67.7%	72.8%
1970	71.4%	77.4%	74.1%	76.6%	72%
1966	75.9%	79%	76%	66.1%	75.8%
1964	77%	80.1%	77.6%	71.7%	77.1%
1959	78.9%	82.6%	78.1%	65.9%	78.7%
1955	76.9%	79.6%	75.1%	74.1%	76.8%
1951	82.7%	84.4%	81.2%	79.9%	82.6%
1950	84.4%	84.8%	80.9%	77.4%	83.9%
1945	73.4%	75.7%	69%	67.4%	72.8%

Table 1: Historical UK General Election Turnout.

Region	2015 Turnout	2010 Turnout	Average Turnout
East	67.5%	67.6%	67.6%
East Midlands	66.5%	66.8%	66.6%
London	65.4%	64.5%	64.9%
North East	61.8%	61.1%	61.4%
North West	64.3%	62.3%	63.3%
Northern Ireland	58.1%	57.6%	57.8%
Scotland	71.0%	63.8%	67.4%
South East	68.6%	68.2%	68.4%
South West	69.5%	69.0%	69.2%
Wales	65.7%	64.8%	65.2%
West Midlands	64.1%	64.7%	64.4%
Yorkshire and The Humber	63.3%	62.9%	63.1%

Table 2: Turnout per EC Region in 2010 and 2015.

$$\begin{aligned}
\triangle q_{it} &= \epsilon_{it} + \eta_t \\
\epsilon_{it} &\sim N(0, \sigma_\epsilon^2) \\
\eta_t &\sim N(0, \sigma_\eta^2) \\
\text{COV}(\epsilon_t, \eta_{it}) &= 0
\end{aligned}$$

Then for $i \neq j$, ρ_q is given by:

$$\rho_q = Corr(\triangle q_{it}, \triangle q_{jt}) = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2} \qquad i \neq j$$

A regression of 2015 constituency turnout on 2010 turnout yields a coefficient of determination of 0.734 which provides evidence for simply using the turnout of the last election as the prediction. We do so. σ_η^2 is simply the variance of the national turnout (6.7%). σ_ϵ^2 requires looking at errors at the constituency level for each separate election. As there was the fifth constituency boundary review in 2008 we can form no easy prediction for area turnout for the 2010 election because constituencies changed. We simply use the 2015 election to estimate σ_η^2 with predictions provided by the 2010 election. This results in a estimate[7] of 3.0% for σ_ϵ and one for ρ_q of 0.835. In the absence of any other estimate or information pertinent to likely voting habits, this is what we use.

4.7   Vote Share

4.7.1   Area Vote Share

An expected national vote share μ_N is required to calculate the means μ of the marginal distributions in the prior. In the days and weeks before the referendum, various opinion polls were published. We use polls to calculate this expectation. Polls with samples in the week preceding the referendum are shown in Table 4.7.1, along with seven polls of polls. For general elections, exit polls measure how people declare they have voted on the day itself at a selection of particular, secret, polling stations. They are much more accurate than any pre-election polling (Curtice et al. (2011)), due to the fact that there is no measurement error of respondents and a diff-in-diff estimate is used based on the sample from the preceding election. There was no exit poll for the referendum as it was a one-off election. There was, however, a poll on the day conducted by YouGov which was published shortly after voting closed

---

[7]Estimates based on sample moments are consistent due to the Law of Large Numbers.

Date(s)	Remain	Leave	Undecided	Remain Lead	Organisation
22-Jun	55%	45%	--	10%	Populus
20–22 Jun	51%	49%	--	2%	YouGov
20–22 Jun	49%	46%	1%	3%	Ipsos MORI
20–22 Jun	44%	45%	9%	1%	Opinium
17–22 Jun	54%	46%	--	8%	ComRes
17–22 Jun	48%	42%	11%	6%	ComRes
16–22 Jun	41%	43%	16%	2%	TNS
20-Jun	45%	44%	11%	1%	Survation/IG Group
18–19 Jun	42%	44%	13%	2%	YouGov
16–19 Jun	53%	46%	2%	7%	ORB/Telegraph
17–18 Jun	45%	42%	13%	3%	Survation
			Polls of Polls		
23-Jun	52%	48%	--	4%	What UK Thinks: EU
23-Jun	50.6%	49.4%	--	1.2%	Elections Etc.
23-Jun	45.8%	45.3%	9%	0.5%	HuffPost Pollster
22-Jun	46%	44%	10%	2%	Number Cruncher Politics
23-Jun	48%	46%	6%	2%	Financial Times
22-Jun	51%	49%	--	2%	The Telegraph
23-Jun	44%	44%	9%	0%	The Economist
				2.0%	**Average Poll of Polls**

Table 3: Opinion Polling Prior to the EU Referendum. Source Wikipedia (2017a).

at 10 pm YouGov (2016). This poll measured how people voted versus how those same individuals reported their voting intention the preceding day. The result was a demographically weighted result of 48.38%. This was broadly in line with recent polls. As we consider this the most accurate poll we set 48.38%.

4.7.2 Variance of Area Vote Share

Variances are chosen by shifting those implied by Hanretty by a constant amount σ_N^2 so that either the model probability agrees with the Betfair implied probability P_0 or σ_p^2 is set so as to imply a generous estimate $\hat{\sigma}_p^2$:

- For P_0 calibration: Between the release of the on the day poll and the first result there were known trades in the spread betting markets with implied probabilities between 10.7% and 37.5%. The average was

Election	Average Poll (Prior Week)	Result	Error
2015	34%	37.8%	-3.80%
2010	35%	36.9%	-1.58%
2005	31%	33.2%	-2.20%
2001	31%	31.7%	-0.92%
1997	30%	30.7%	-0.27%
1992	37%	41.9%	-4.46%
	$\hat{\sigma}_\epsilon$		**2.66%**

Table 4: Opinion Polls and Vote Share for the Conservatives for recent general elections.

    25.1% (7.1%). We use 25% but note that values in $(18\%, 32\%)$ would be reasonable (the mean $\pm$ the standard deviation). Note that these probabilities are broadly in line with the predictions of the Almanis prediction market (19) and Dr Tetlock's Good Judgment project (25%).

- For σ_p^2 calibration: Figure 4.7.2 shows that the average error in opinion polls from the prior week in the last six general elections was 2.66%. General election polling is a well-researched field with plenty of historical precedent and would probably be too confident a figure. As our aim is to produce a prediction based on a conservative prior, we will set $\sigma_p = 5\%$.

4.7.3 Correlation between Voting Areas

We analyse general election data in a similar manner to section 4.6.2 to estimate ρ_p. The generally pro-Brexit Conservative party vote share is used as a proxy for the Brexit vote. Predictions for that vote at the constituency level are based on applying the implied swing from opinion polls from the week prior to each election, to the level of the last election. See Table 4.7.2 for these polling results and for the results of the last six elections. Complications arise due to Westminster constituency boundary reviews in 1995, 2005 and 2008. These reviews change the number of constituencies and their composition of voters. They occur periodically in order to remove variations in the number of electors in each area, and have tended to favour the Conservatives (Rallings et al. (2008)). This is a well understood problem and the website Electoral Calculus (Baxter (2017)) publishes implied election results for elections preceding a review to enable ready comparison; we use these implied figures.

The implied standard error of σ_η using the data in Table 4.7.2 is 2.66%[8]. Relying on the last six elections, the constituency level error calculation yields an estimate of 4.18% for σ_ϵ, implying a correlation $\rho_p = 0.288$. However, our constituency level errors are probably estimated at too high a level as better predictions for constituency level results exist although we do not have ready access to them. For this reason, the value of σ_η^2 is probably estimated at too high a level and ρ_p too low. Consequently, this value of ρ_p will be treated as a lower bound.

The correlation ρ_p is likely to be the parameter with the largest effect on how quickly the model prediction will converge to the true result. It is therefore worth commenting that the implied correlation coefficient, as estimated, appears to be stable. Using only the last three elections results in an estimate of 0.324. The largest estimated value of σ_ϵ (5.69%) in any single election for the last six was in 1997, which was (unsurprisingly) also the largest error in the national vote share. If we combine this $\sigma_\eta = 2.66\%$, the $\rho_p = 0.18$. This is an artificially low estimate and parameter values below this level are highly unlikely.

4.8   Correlation between Vote and Turnout

There were conflicting reports concerning the probable impact of turnout, even within the same newspaper on the day of the result (Gutteridge (2016), Foster (2016)).

Predicting how unexpected turnout affects results requires a successful prediction of whether the difference in turnout is attributed to Leave or Remain supporters. This is a difficult problem. It was well understood in advance that younger voters, who are less likely to vote, would favour Remain, and that Brexit supporters were reported in surveys being more than twice as likely to vote as Remain ones Twyman (2016). YouGov were widely quoted as suggesting the relationship between turnout and Brexit vote share would be negative (see Foster (2016) for an example).

We again take a quantitative approach by examining general elections. Due to boundary changes we are restricted to studying only the 2015 general election as no implied turnouts for elections preceding a boundary review is available. We proxy support for Brexit at the 2015 general election by using combined vote shares of UKIP and the Conservatives (the parties with supporters most sympathetic to Brexit). We regress the swing of the combined UKIP and Conservatives vote share against the change in turnout at the

---

[8]As the model implicitly assumes a mean of zero this is the square root of the average of the squares of the error, not the sample variance.
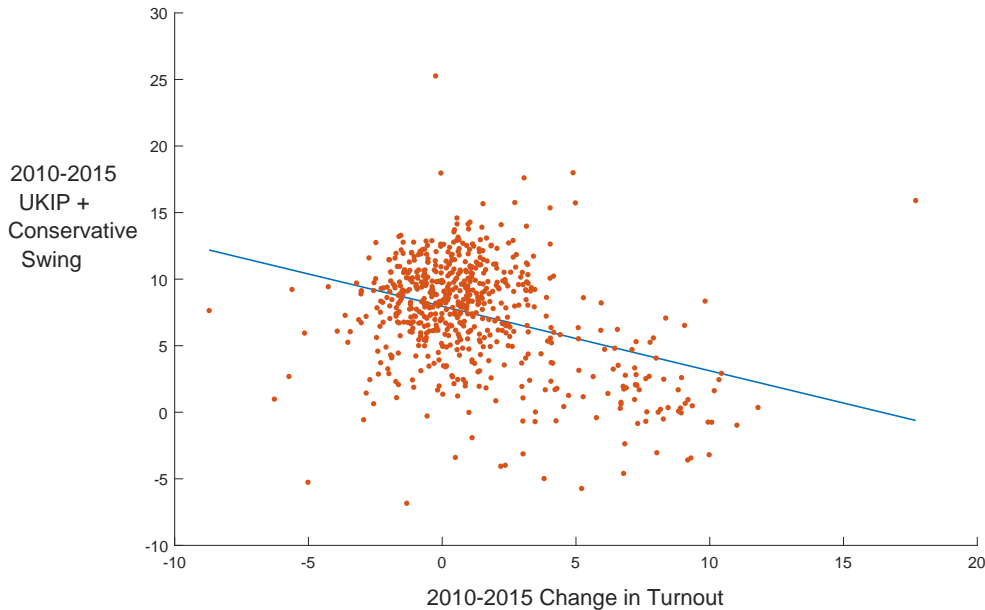
Figure 2: Change in UKIP + Conservative Vote Share versus Change in Constituency Turnout at the 2015 General Election.

constituency level for the 2010-2015 elections. Figure 2 shows the regression, which results in a statistically significant correlation coefficient of -0.361. This is indeed negative, agreeing with YouGov, and we use this value.

4.9 A Note on Model Correlation and Measured Correlation

The model correlation parameters ρ_p, ρ_q and ρ_{pq} are not strictly the correlations of the random variables $r = (r_1, \ldots, r_{2n})$. They are the correlations of Φ^{-1} of the quantiles or:

$$\left(\Phi^{-1}(F_1(r_1)), \ldots, \Phi^{-1}(F_{2n}(r_{2n}))\right).$$

When the marginal is normal, $\Phi^{-1}(F_i(r_i))$ will be a linear function of r_i correlations will be identical. When $\Phi^{-1}(F_i(\bullet))$ is non-linear, a simulation or other method would be strictly required to convert them. However, we omit this step as we do not expect it to be significant. As we shall see, moving away from the Gaussian marginals does not change the behaviour of the model greatly.

Parameter	Description	Value	Range
ν	National Turnout	67.6	$65-70\%$
σ_ν	Turnout Error	$(6.7\%)^2$	–
ρ_q	Area Turnout Correlation	0.835	–
μ_N	Expected National Vote Share	48.38	–
P_0	Prior Expectation of Brexit	25%	$18-32\%$
σ_p	National Vote Error	5%	–
ρ_p	Area Vote Correlation	≥ 0.288	≥ 0.18
ρ_{pq}	Area Vote and Turnout Correlation	-0.361	–

Table 5: Plausible parameter values.

4.10   Missing Priors

Of the 382 voting areas of the Referendum, Hanretty failed to publish priors for four areas. These are listed in Table 4.10. The 4 areas are:

1. Gibraltar: This makes up a tiny 0.05% of the electorate, was the first area to announce, and had overwhelming support for Remain (Reyes (2016b,a)). As the population is so distinct from that of the rest of the UK, the result is not informative. We therefore take it as given and do not include it in the model.

2. The Isles of Scilly and Isle of Anglesey make up only 0.11% of the electorate and are simply ignored.

3. Northern Ireland consists of about 1.26m voters in a total electorate of roughly 46.5m. We use opinion polls for the mean and a standard deviation equal to that of the average of the other areas. We use a poll published on June 20 (Shapiro (2016)) that showed Remain 11% ahead, or 9% higher than the rest of the UK at that time. We therefore set the mean equal to $\mu_N - 9\%$.

Area	Declaration (Actual / Expected)	Electorate (%)
Gibraltar	23:36:33 / 00:01	24,119 (0.05%)
Isles of Scilly	00:49:42 / 00:01	1,799 (0.004%)
Isle of Anglesey	02:18:00 / 02:30	51,445 (0.11%)
Northern Ireland	04:37:00 / 04:00	1,260,955 (2.71%)

Table 6: The 4 Voting Areas with Missing Priors. Source Hanretty (2016).

5 Theoretical Model

In this section we present a simple model of how the markets could behave overnight according to basic economic theory.

5.1 Assumptions

The assumptions of the model are labeled as follows:

NS   No shocks to the GBPUSD price beyond those that affect the probability of Brexit

ID   Identical beliefs in the GBPUSD and Betfair markets

EMHW   The Efficient Market Hypothesis holds in the weak form

EMHSS   The Efficient Market Hypothesis holds in the semi-strong form

CMI   Conditional Mean Independence of GBPUSD price given Brexit

RN   Risk Neutrality

Not all these assumptions are regarded as holding exactly but they are presented as an approximation. We discuss each one in detail below.

NS   The assumption of no external shocks beyond those affecting the likelihood of Brexit would not be expected to hold over longer time scales, but for the period under study its validity is reasonable. Indeed, the authors of Wu et al. (2017) describe the circumstances as "a natural experiment" with "near perfect conditions" to study such a situation. There were no other major economic releases, or significant news events (beyond Brexit). In advance, the Econoday Economic Calendar (Econoday (2016)) listed the final market moving news releases on the 23 June as the US New Home Sales Report at 10:00 am Eastern Time (ET) and the first one for 24 June (beyond the referendum) as Durable Goods Orders at 10 am ET. They predicted that the following would be the market focus for the 24th: "In a rare and potentially powerful wildcard, the markets will react to the Brexit outcome". This demonstrates that in advance there were beliefs that the main determinant of prices would be the outcome of the Referendum.

ID   When market participants are allowed to have heterogeneous beliefs, the idea of a "market implied probability" has no definition.
          In fact, even under RN, the price of a Betfair contract paying out
          £1 in the event of Brexit does not even equal the budget weighted
          mean belief of participants[9]. This assumption states that under
          heterogeneous beliefs both the Betfair and the financial market
          have the same *distributions*.

EMHW   When the EMH holds there can be an interpretation of a market
          probability for a prediction market, as all information has been
          aggregated into the price. The market price of the Betfair contract can be interpreted as $u^{-1}(P_{betfair} \times u(£1))$, where $u(\bullet)$ is
          the Bernoulli utility function and $P_{betfair}$ is the market implied
          probability. Correspondingly, $P_{betfair} = u(BETP)/u(£1)$ where
          $BETP$ is the price. Note that EMHW $\Rightarrow$ ID.

EMHSS   EMH holding in the semi-strong form will imply that prices immediately discount all publicly available information including
          any results of voting areas already announced in the referendum. As a solution concept, the semi-strong form will imply
          that $P_{betfair}$ will equal the electoral model probability at all times.
          Note that EMHSS $\Rightarrow$ EMHW $\Rightarrow$ ID.

CMI   Prior to the vote there were several predictions that the pound
          would sell off significantly in the event of Brexit but rally a little
          otherwise (see Wu et al. (2017)). Let

$$D = \begin{cases} 1 & BREXIT \text{ at time } T \\ 0 & REMAIN \text{ at time } T \end{cases}$$

          The CMI assumption can be written mathematically as:

$$\mathbb{E}_t(GBP_T|D, p_N) = \mathbb{E}_t(GBP_T|D),$$

          where p_N is the national share of those voting for Brexit, while
          T is the time when the decision is announced and t is any time
          with $t < T$. Equivalently, the sterling rate is affected by p_N only
          through its affect on whether Brexit occurs. The GBPUSD price
          would be expected to be the same if the vote for Brexit were either
          50.01% or 99.99%. This is a strong assumption, particularly when

---

[9]C. Manski argued in 2004 (Manski (2004)) that as the market price is the price at
which net buyers must equal net sellers then the mean belief will lie in an interval whose
midpoint is the market price P but with width $2P(1 - P)$ (assuming RN).

there is significant probability mass around outcomes very close $p_N = 50\%$.

RN   Again, this is a strong assumption which is not believed to hold in practice, but it is a useful approximation that is likely to be roughly valid. However, it is noted that any deviations from RN may have larger effects than they otherwise might have due to the increased risk of holding the pound during the period under study[10].

5.2 Model

We rely on all the assumptions of the model except EMHSS. These imply that for the currency price GBP_t and for the Betfair Contract Price, $BETP_t$, which pays out £1 in the event of Brexit,

$$
\begin{aligned}
GBP_t &= \mathbb{E}_t\left(GBP_T\right) \\
&= \mathbb{E}_t\left(GBP_T | D, p_N\right) \\
&= \mathbb{E}_t\left(GBP_T | D\right) \\
&= \Pr\left(D = 1 | \mathcal{F}_t\right) \mathbb{E}_t\left(GBP_T | D = 1\right) \\
&\quad + \Pr\left(D = 0 | \mathcal{F}_t\right) \mathbb{E}_t\left(GBP_T | D = 0\right) \\
&= \Pr\left(D = 1 | \mathcal{F}_t\right)\left(\mathbb{E}_t\left(GBP_T | D = 1\right) - \mathbb{E}_t\left(GBP_T | D = 0\right)\right) \\
&\quad + \mathbb{E}_t\left(GBP_T | D = 0\right)
\end{aligned}
$$

$$
\begin{aligned}
BETP_t &= \mathbb{E}_t\left(BETP_T\right) \\
&= \Pr\left(D = 1 | \mathcal{F}_t\right) \times £1 \\
&= \Pr\left(D = 1 | \mathcal{F}_t\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
GBP_t &= BETP_t \times \left(\mathbb{E}_t\left(GBP_T | D = 1\right) - \mathbb{E}_t\left(GBP_T | D = 0\right)\right) \\
&\quad + \mathbb{E}_t(GBP_T | D = 1).
\end{aligned}
$$

5.3 Implications

The assumptions of the model imply that the two markets should move together tick by tick. We can relax the assumption of no external shocks to

---

[10]Higher perceptions of risk were evident from, for instance, higher implied volatility from options pricing as well as increased margin requirements from brokers for sterling related products.

prices beyond Brexit by assuming that there are shocks but that they do not persist (are stationary) and are not serially correlated.

Write $GBP_H = \mathbb{E}_t\left(GBP_T | D = 0\right)$, $GBP_L = \mathbb{E}_t\left(GBP_T | D = 1\right)$ and $\triangle GBP = \triangle GBP_H - \triangle GBP_L$ and assume that these are constants, i.e., independent of t. Prices at time t will thus satisfy:

$$
\begin{aligned}
GBP_t &= GBP_L + (1 - BETP_t) \times (GBP_H - GBP_L) + \epsilon_t \\
&= GBP_L + (1 - BETP_t) \times (\triangle GBP) + \epsilon_t,
\end{aligned}
$$

where ϵ_t is a martingale difference. Thus the two markets will be cointegrated with cointegrating vector $(\triangle GBP, -1)$. The presence of serial correlation in deviations from the cointegrating vector ϵ_t would imply that markets were informationally inefficient, violating the EMHW.

The stronger assumption of EMHSS implies that:

$$
\begin{aligned}
GBP_t &= GBP_L + \Pr(D = 0 | \mathcal{F}_t) \times (\triangle GBP) + \epsilon_t \qquad (5.1) \\
BETP_t &= \Pr(D = 1 | \mathcal{F}_t),
\end{aligned}
$$

where $\Pr(D = 0 | \mathcal{F}_t)$ is the model probability of Remain and is correctly evaluated given all available information. Taking $\triangle GBP$ as given, an arbitrage opportunity may exist if the error term deviates from zero. This would be facilitated by taking a position in the Betfair contract and taking an opposing position in the pound, in the ratio of the cointegrating vector.

5.4   Relaxation of assumptions

Relaxing RN will imply that

$$
GBP_t = u^{-1}\left(u\left(GBP_L\right) + u^{-1}\left(1 - BETP_t\right) \times u\left((GBP_H - GBP_L)\right)\right)
$$

where $u(\bullet)$ is the utility function. When we relax CMI we can still assume that the GBPUSD price will decrease with the vote share p_N. Thus GBP_t is will now be a non-linear but monotonic function of $(1 - BET_p)$. Therefore, only NS and ID are required to imply that the pound price should be moving contemporaneously with some non-linear but monotonic function of the Betfair price.

6 Results

6.1 Electoral Probability Model

We used Matlab to generate results, sampling from the relevant multivariate distributions to evaluate model probabilities. Where calibration was required, we found that a simple gradient descent was adequate.

6.1.1 Forecast Model

Figure 6.1 shows the results for calibrating the model prior to the generous standard deviation figure of 5% using the parameters from Table 4.8. The kurtosis of the distributions is lower for the normal model, due to the logit mapping squeezing the distribution. The logit t-distribution has higher kurtosis, as expected. Both logit distributions are highly significantly different to the normal distribution, as shown by the Jarque-Bera (JB) test pValue. All marginals had a very generous prior 90% confidence interval of around 16.5% width. Figure 6.1 shows the evolution of the model forecast as the night progressed. The results of all marginals are almost identical except for the higher kurtosis logit-t marginal, which had a surprisingly quicker rate of convergence. The logit-t model predicted Brexit at 1:32am on the 13th result with 95% accuracy, and on the 19th result at 1:45am with 99% accuracy. The other models took until 1:44:56am (16th result) and 2:05am (34th result) to get to those certainties respectively. This compares with the BBC projecting Brexit at 4:39:32am[11]. The Betfair market took until 04:21:00 am to imply 99% probability, and the pound took even longer to react. This is considered a successful prediction which violates EMHSS.

6.1.2 Robustness to Parameter Changes

Table 7 shows the figures when making parameters more conservative by roughly 5%. This involves lowering ν, lowering σ_ν, increasing the magnitude of ρ_{pq}, increasing ρ_q, decreasing ρ_p and increasing the initial variance σ_p. Changing only σ_p and ρ_p on their own changes the speed of convergence a little. Lowering ρ_p further to the very conservative lower bound of 0.18 slows the result significantly but still predicts Brexit with 99% probability about an hour and a half before the Betfair market. This suggests that our conclusions are robust to parameter changes. Only changes to σ_p and ρ_p meaningfully affect the model, as expected.

---

[11]According to our Bloomberg scrape.

Forecast Model Prior
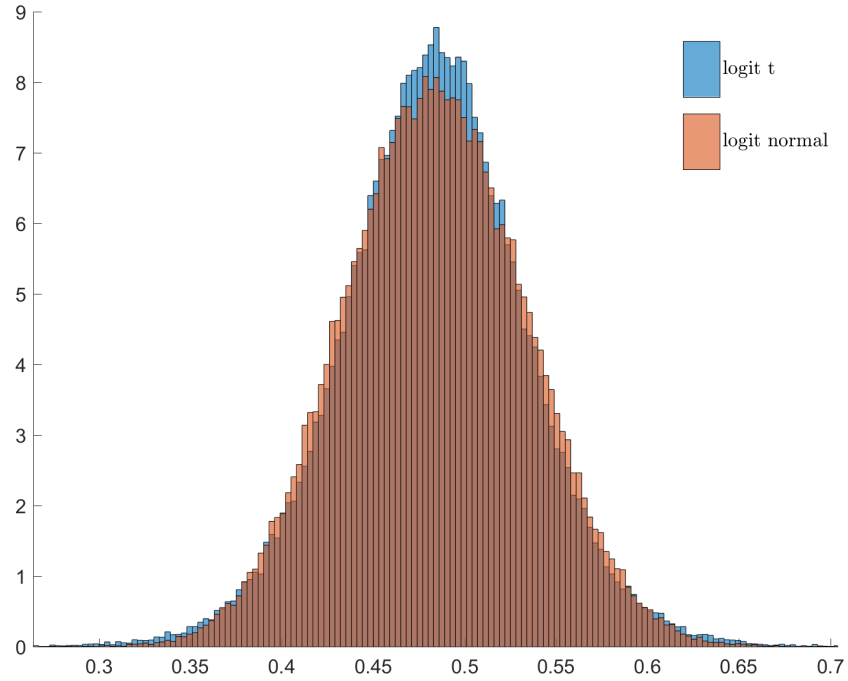
Marginal	Kurtosis	90% CI	99% CI	JB p
Normal	3.0	(40.1%, 56.7%)	(35.4%, 61.4%)	> 0.5
Logit Normal	2.9	(40.2%, 56.6%)	(35.9%, 60.1%)	< 0.001
Beta	3.0	(40%, 56.6%)	(35.5%, 61.3%)	0.29
Logit t	3.6	(40%, 56.6%)	(34.4%, 62.4%)	> 0.5

95% result	99% result
16 (1:44:57)	34 (2:05:00)
16 (1:44:57)	34 (2:05:00)
16 (1:44:57)	34 (2:05:00)
13 (1:32:00)	19 (1:45:00)

Figure 3: Prior Distributions for Forecast Model.

Forecast Model Path
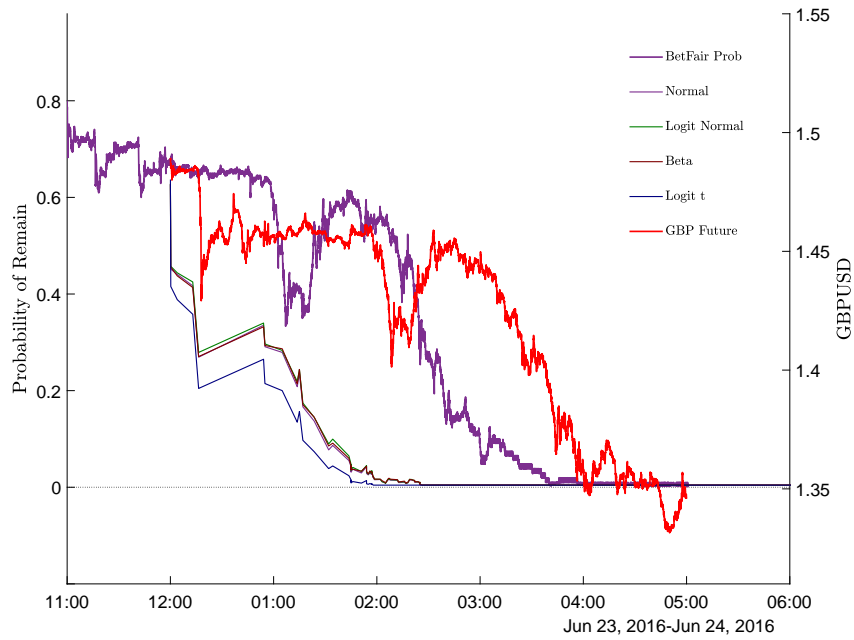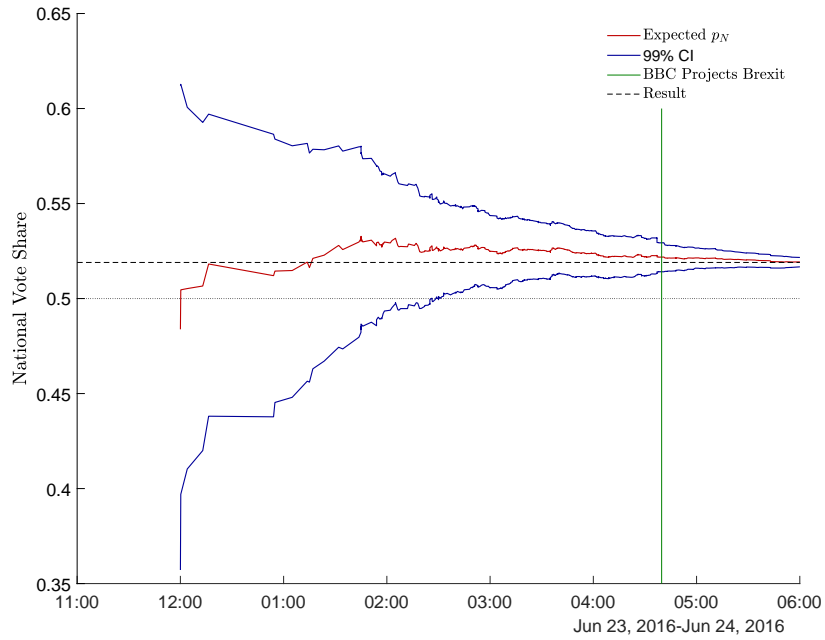
Figure 4: National Vote Distribution Evolution (Logit Normal) and Model Probability Paths for Forecast Model.

Result	σ_P	ρ_p	ρ_q	ρ_{pq}	ν	σ_ν	95% result[a]
1 [b]	5%	0.288	0.877	−0.379	64.2%	6.4%	16 (1:44:57)
2 [c]	5%	0.274	0.835	−0.361	67.6%	6.7%	16 (1:44:57)
3 [d]	5.3%	0.288	0.835	−0.361	67.6%	6.7%	17 (1:45:00)*
4 [e]	5%	0.18	0.835	−0.361	67.6%	6.7%	35 (2:05:00)*

[a]*Indicates a change from initial values

[b]Result 1 changes all parameters except ρ_p and σ_p.

[c]Result 2 changes ρ_p only.

[d]Result 3 changes σ_p only.

[e]Result 4 sets ρ_p at the lower limit of our plausible range, 0.18.

Table 7: How the Times to Predict Brexit Vary with More Conservative Parameter Values (Normal Marginal).

6.1.3 Other Parameter Values

Next we turn to the question of how the model evolves when calibrating to the initial Betfair probability. Results are shown in Figure 6.1.3. The models converge exceptionally quickly, with all of them predicting Brexit with 99% confidence by the 12th result at 1:23:34. Calibration to $P_0 = 25\%$ yields an initial standard deviation of 2.4% (2.7% for logit). This is lower than recent general election polling errors and may indicate that the market was irrationally overconfident. To answer the question of what prior the market expresses, we lower ρ_p while setting the variance so as to fix P_0 to market, and fixing the other parameter values. This is justified, as the preceding section demonstrated that ρ_p is the main determinant of model behaviour after σ_p has been fixed. Figure 6.1.3 shows the results and suggests that a model roughly consistent with the Betfair market would necessitate a ρ_p of around 0.02. This is simply implausible. There is no hope for the pound market where a value of $\rho_p \in (0, 0.01)$ appears to be required. We conclude that the market could not have behaved consistently for any reasonable prior belief.

Calibration to Betfair Market
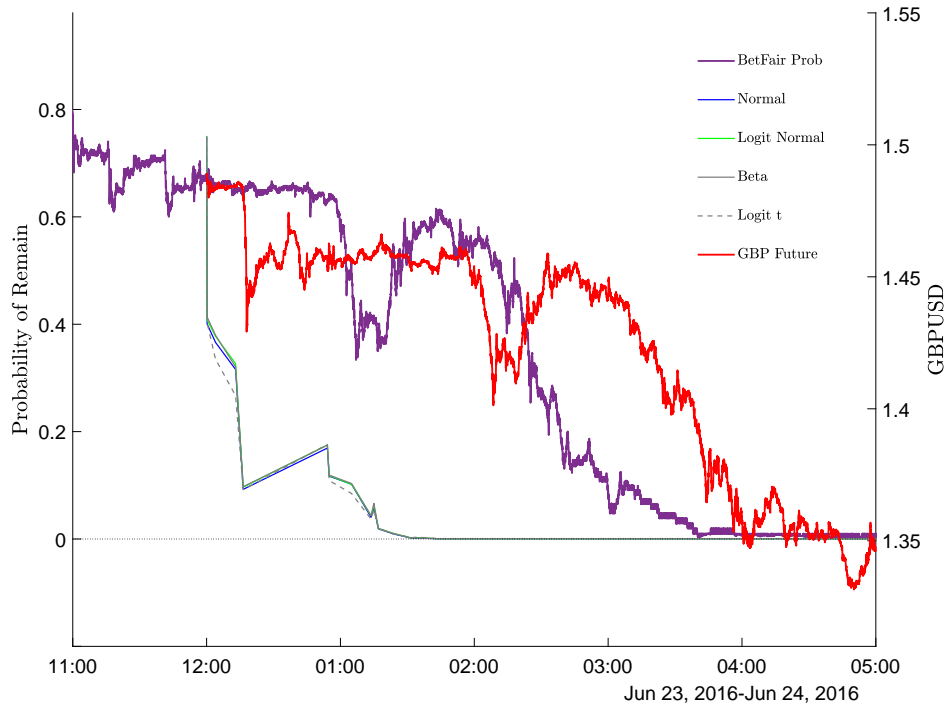
Marginal	Stdev	Kurtosis	90% Interval	99% Interval	JB p
Normal	2.40%	3.0	44.5%, 52.3%	42.3%, 54.5%	0.404
Logit Normal	2.40%	3.0	44.5%, 52.3%	42.3%, 54.5%	0.012
Beta	2.40%	3.0	44.4%, 52.3%	42.2%, 54.6%	0.264
Logit t	2.70%	4.2	44.1%, 52.7%	40.7%, 56.0%	<0.001

95% result	99% result
10 (1:15:00)	11 (1:17:00)
10 (1:15:00)	12 (1:23:34)
10 (1:15:00)	11 (1:17:00)
10 (1:15:00)	12 (1:23:34)

Figure 5: Results when Calibrating Initial Model Probability to Betfair Market with the Statistics of the Calibrated Model Priors and Prediction Times.

31
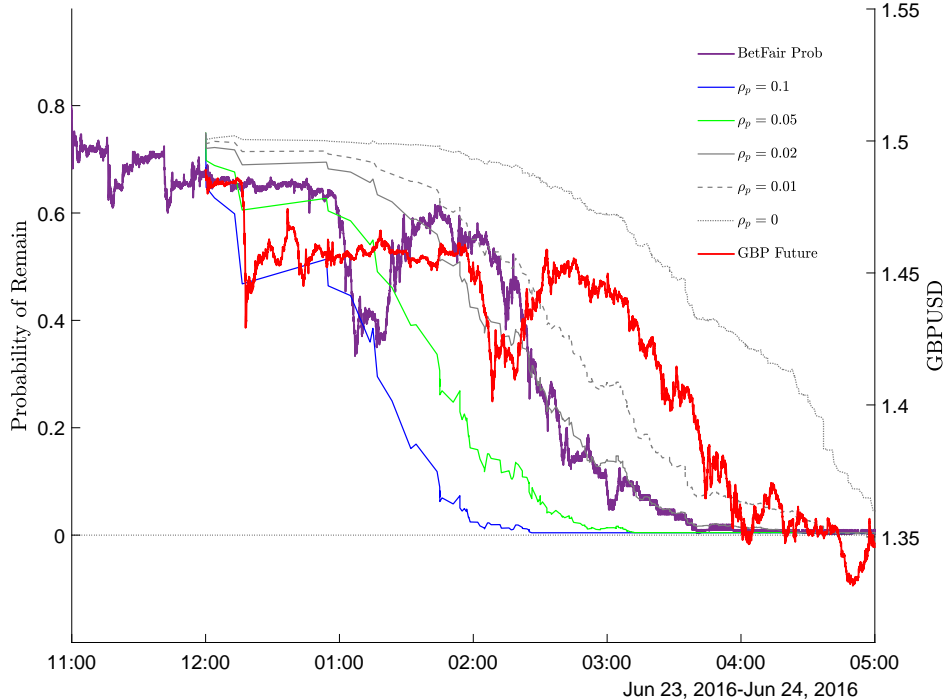
Figure 6: Model Probability Paths when Lowering ρ_p.

6.2 Behaviour of Markets and Arbitrage Opportunity

The evolution of the two markets on the night is shown in Figure 7. We calculate the cointegrating ratio by evaluating both markets at the beginning and end of the 7 hour period from 10 pm on 23 June to 5 am on 24 June. Following the notation of equation 5.1, this gives implied values of $GBP_L = 1.345$, $GBP_H = 1.499$ and $\triangle GBP = 0.154$. This is similar to the predictions made of the pound conditional on the outcome of the referendum made by various commentators in advance (see Wu et al. (2017)).

The cointegration error $\epsilon_t = GBP_t - (1 - BetP) \times \triangle GBP$ is plotted in Figure 8 along with the sample Auto-Correlation Function (ACF) in Figure 9. Note that the error is the return of the portfolio generated by buying £1 and selling $\triangle GBP$ Betfair Brexit contracts. Testing the Null of stationarity of the error with the KPSS test results in rejection with a pValue of < 0.01[12]. This is a formal rejection of the theoretical model presented in section

---

[12]Testing a null hypothesis of a unit root via an augmented Dickey Fuller test using the saturate and reject method of lag selection results in non-rejection of the null with a
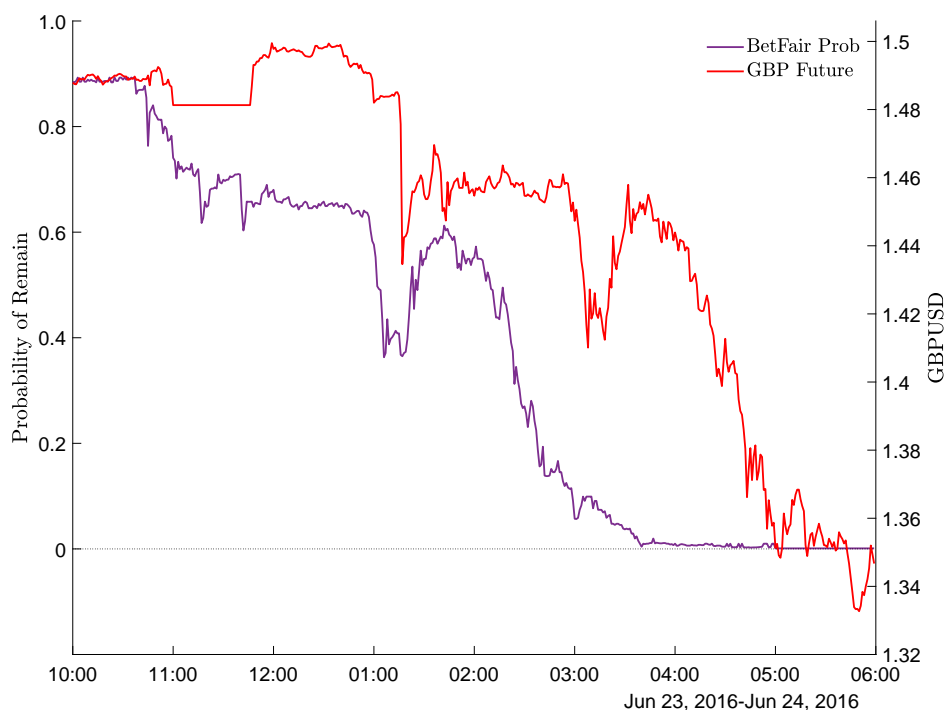
Figure 7: Last Prices by Minute.

5. There is a limit to what can be achieved from a statistical test of a time series spanning eight hours where there appears to be a very strong lead/lag relationship of the order of 2 hours between the two components. The ACF appears to exhibit non-stationary persistence. The most convincing explanation for the rejection of stationarity of the error and the shape of the ACF is that the referendum shocks were first felt in the betting market and later in the currency market. This is shown in Figure 10 which graphs implied probabilities of the forecast model and the markets. (The RN and the CMI assumptions are required to imply these market probabilities). The forecast model leads the Betfair price, which in turn leads the pound probability. The average horizontal distance on this plot between the relevant lines is 113 minutes for forecast-Betfair and 185 for forecast-pound. It appears that the fundamental information (forecast) led the betting market by nearly 2 hours which in turn led the pound by over an additional hour.

Relaxing the CMI, RN and EMHSS assumptions made in section 5 but still relying on NS, implies that the markets move contemporaneously, ac-

pValue of 0.168. This provides further evidence of non-stationarity.

33
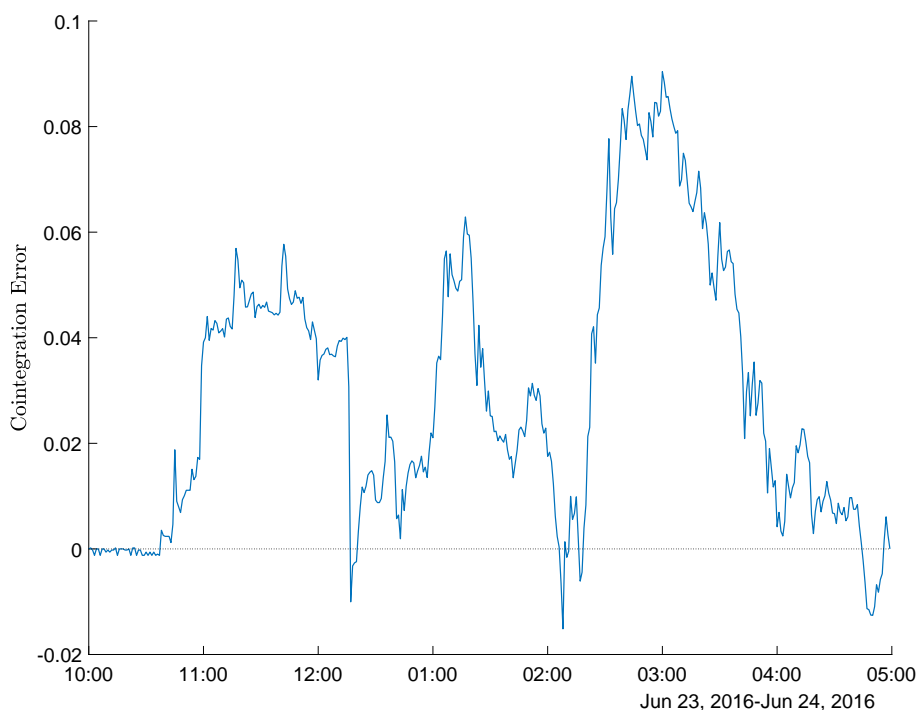
Note: The error is the return, before transaction costs, of buying GB-PUSD and selling the Betfair contract for Remain, in the cointegrating ratio. There appear to be profitable arbitrage opportunities that require no successful forecast of GBP

Figure 8: The Theoretical Cointegration Error.

cording to some possibly non-linear but monotonic function of price. What would need to be assumed for the EMHW to actually hold? One would have to believe that there was a shock to the pound that caused a change in price of 9 cents, or around 7%, which was independent of the referendum results and which was subsequently reversed around 2 hours later, at around the time that Brexit became apparent. This is not plausible.

If we do not reject the null of stationarity, then the ACF suggests that the error term, that is, the return of the portfolio, is highly serially correlated. This could be formally tested. However, given the shape of the ACF and the evolution of the prices of the 2 assets over time, there is little point confirming what we already know. This is that the Betfair market led the pound market and that a portfolio of selling the pound and buying the Betfair contract (i.e. selling the error term) would be profitable. There are known trades in the
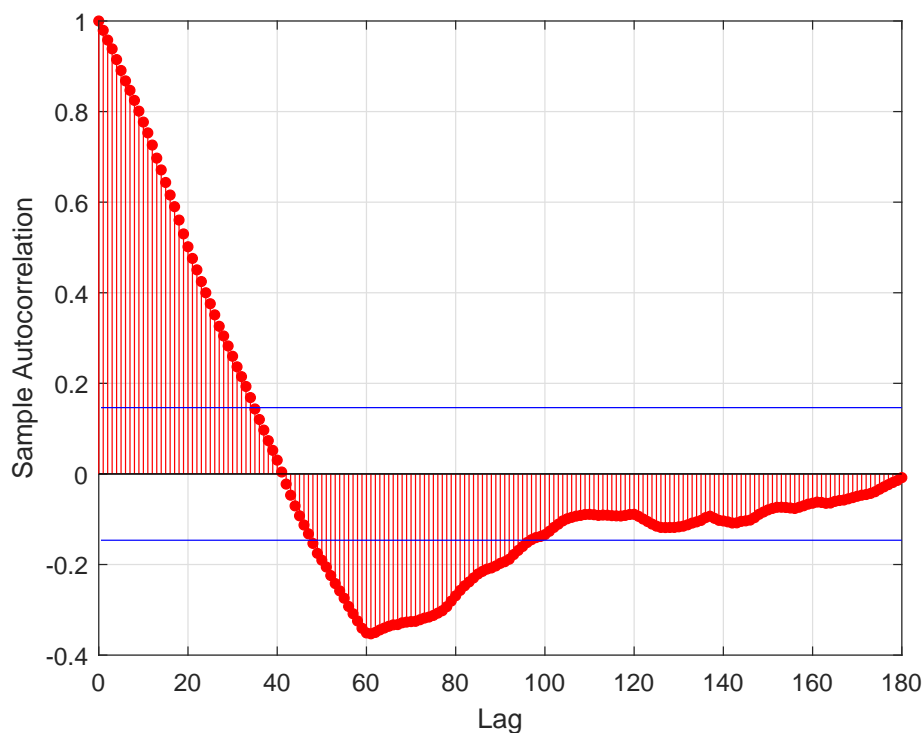
Figure 9: Sample Auto Correlation Function of the Error.

market at around 4 am that would have made roughly 9 cents of profit per £1 of the pound sold, or an unleveraged return of up to 7% in about two hours[13]. Supporters of the EMH could object to the conclusion that the EMH fails by simply selling the pound uncovered due to the increased risk an investor would be being exposed to (the pound, after all, was exceptionally volatile during the period). However, the arbitrage strategy of covering the short with a position in the spread bet largely eliminates risk. The arbitrage suggests that failure of the EMH in its weak form is not reliant on the forecast model predictions.

---

[13]In terms of transaction costs, selling the pound would cost about 2-3 hundredths of a cent at that time, whereas the Betfair cost is $3-5\%$ levied on any bets that pay out. This would slightly change the ratio of the portfolio but not significantly affect profits or these conclusions.
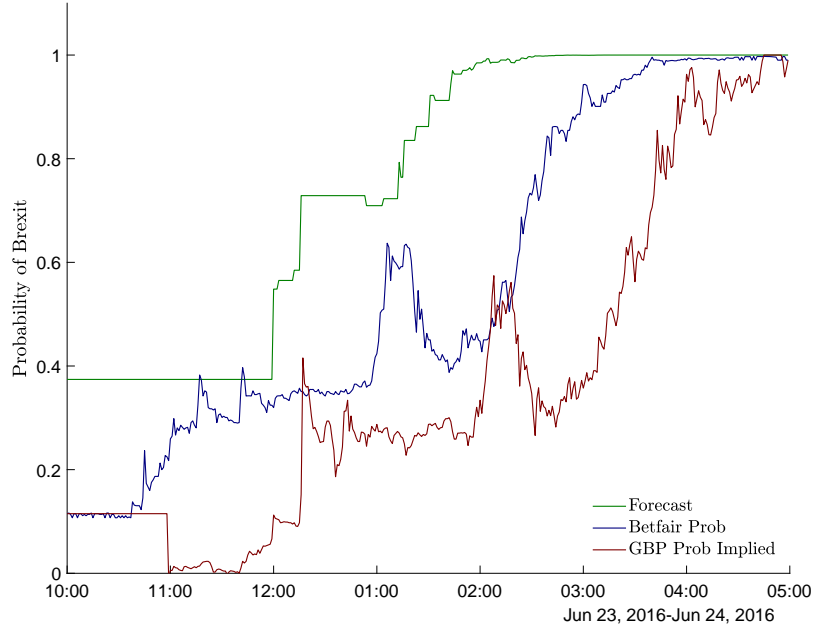
Figure 10: The Forecast Model Probability, the Betfair Probability and the GBP Implied Probability under RN and CMI Assumptions.

7   Conclusion

This paper examined the efficiency of the Betfair and pound/dollar markets as the results of the United Kingdom European Union membership referendum were announced. This event provided a unique opportunity to study the interaction between a flow of information, a prediction market and a financial market, where there was a sole, public determinant of prices. Other work has identified the pound market as being inefficient during the period under investigation but we were able to answer questions about the efficiency of the prediction market as well as the relative speed at which the two markets digested the information flow.

We have presented a rigorous Bayesian real-time model of the probability of Brexit for the period under consideration. This is based on a copula that is not constrained to normal distributions. The Bayesian method improves upon earlier estimation methods as it does not rely on any asymptotic properties of estimators for small samples. The conclusions of the model are as follows:

1. Not only was the currency market informationally inefficient so too

36

was the betting market, and both markets violated semi-strong EMH on the night of the vote.

2. The betting market, although inefficient, was more efficient than the currency market. The betting market took less than 2 hours to reflect the information contained in the vote whereas the currency market took over 3 hours.

3. The mispricing is inconsistent with any plausible prior and flow of information. Specifically, an assumption of consistent behaviour implies a simply unbelievable prior belief (very close to perfect independence of the results of different voting areas).

We also present a close to risk-free arbitrage opportunity in the two markets. The arbitrage result suggests that a violation of EMH in the weak form has occurred. The conclusion that there is a failure of the weak form of the hypothesis is not reliant on any flow of fundamental information or the electoral probability model.

Our results suggest that market participants suffered a behavioural bias as the results unfolded. It appears that traders and gamblers simply could not believe that the UK was voting to leave the EU. Further, it appears that on this occasion the betting market, although slow, adjusted much more quickly than the financial markets. Any future possible UK referendum on this subject will present an opportunity to study whether this inefficiency persists or whether efficient behaviour is exhibited, possibly due to the publication of this and other studies. If an inefficiency were to persist it would be interesting to observe whether the betting markets again lead the financial markets.

A   Model summary

A.1   Variables

$p_i, q_i, s_i :$ Voting area vote percentage, turnout and size, order by time of announcement

$p_n, \mu_N, \sigma_i^2 :$ National vote share, mean and variance for Brexit

$\mu_{p_i} :$ Expectation of p_i

$\mu_H :$ Vector of expectations provided by Hanretty study

$\sigma_{p_i}^2 :$ Marginal variance of p_i

$\upsilon_i :$ Expectation of q_i

$\sigma_\upsilon^2 :$ Marginal variance of q_i, independent of i

$P_0 :$ Betfair implied prior probability of Brexit

$r :$ Vector of variables $= (p_1, \ldots, p_n, q_1, \ldots, q_n)^\intercal$

$\widetilde{r} :$ Re-ordered vector by time of announcement$= (p_1, q_1, \ldots, p_n, q_n)^\intercal$

$F_i, \widetilde{F}_i :$ Marginal CDFs of i'th components of r and $\tilde{r}$

$\rho_\theta, \rho_\phi :$ Inter-area prior vote share and turnout correlation

$\rho_{\theta\phi} :$ Intra-area prior vote and turnout correlation

$\Sigma_0 :$ Covariance matrix of copula prior

$\widetilde{\Sigma}_0 :$ Covariance matrix of copula prior with sequentially reordered rows and columns

$\widetilde{\Pi} :$ Mean of prior after the announcement of m results

$\widetilde{\Sigma} :$ Covariance matrix of prior after the announcement of m results

A.2 Prior Probability of Brexit

$$P(\text{BREXIT})_0 = P\left(\frac{\sum_i p_i q_i s_i}{\sum_i q_i s_i} > \frac{1}{2}\right)$$

$$\Phi^{-1}(F_1(r_1)), \dots, \Phi^{-1}(F_{2n}(r_{2n})) \sim N(0, \Sigma_0)$$

$$\Sigma_0 = \begin{pmatrix} \Sigma_p & \Sigma_{pq} \\ \Sigma_{pq}^T & \Sigma_q \end{pmatrix}$$

$$\Sigma_p = (1 - \rho_p) I_n + (\rho_p) i_n i_n'$$

$$\Sigma_q = (1 - \rho_q) I_n + (\rho_q) i_n i_n'$$

$$\Sigma_{pq} = \rho_{pq} \times [(1 - (\rho_q \rho_p)) I_n + (\rho_q \rho_p) i_n i_n']$$

A.3 Prior Marginal Calibration

$$\mu_p = \mu_H + \alpha_N \times i$$

$$\sigma_{p_i}^2 = (\sigma_H^2)_i + \sigma_N^2$$

$$\sigma_N^2, \alpha_N : E(p_N) = \mu_N, \quad P_0 = P(\text{BREXIT})_0 \ (\text{EMH}) \ or \ \sigma = \hat{\sigma}_N^2 \ (\text{FORECAST})$$

A.4 Update

$$\tilde{\Sigma}_0 = \begin{pmatrix} \tilde{\Sigma}_{m,m} & \tilde{\Sigma}_{m,\not{m}} \\ \tilde{\Sigma}_{\not{m},m} & \tilde{\Sigma}_{\not{m},\not{m}} \end{pmatrix}$$

$$\tilde{x}_m = (F_{p_1}(p_1), F_{q_1}(q_1), \dots, F_{p_m}(p_m), F_{p_m}(q_m))'$$

$$\tilde{\Pi}_{\not{m}} = \tilde{\Sigma}_{\not{m},m} \tilde{\Sigma}_{m,m}^{-1} \tilde{x}_m$$

$$\tilde{\Sigma}_{\not{m}} = \tilde{\Sigma}_{\not{m},\not{m}} - \tilde{\Sigma}_{\not{m},m} \tilde{\Sigma}_{m,m}^{-1} \tilde{\Sigma}_{m,\not{m}}$$

$$P(\text{BREXIT})_m = P\left(\frac{\sum_{i \geq m} p_i q_i s_i}{\sum_i q_i s_i} > \frac{1}{2} - \frac{\sum_{i \leq m} p_i q_i s_i}{\sum_i q_i s_i}\right)$$

$$\Phi^{-1}(\tilde{F}_{2m+1}(\tilde{r}_{2m+1})), \dots, \Phi^{-1}(\tilde{F}_{2n}(\tilde{r}_{2n})) | \tilde{x}_m \sim N(\tilde{\Pi}_{\not{m}}, \tilde{\Sigma}_{\not{m}})$$

B Review of Probability Model in Wu et al. (2017)

The model under consideration in Wu et al. (2017) performs (in the one factor case) the following Weighted Least Squares regression following the announcement of k results

$$p_i = \alpha\mu_i + \beta + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2) \quad i = 1, \ldots, k.$$

The national vote share and thus the probability of Brexit is then simulated by generating M realisations and evaluating the relevant sum. A correct application of this method would involve sampling unknowns $(\alpha, \beta, \sigma_\epsilon^2)$ from the joint distribution

$$N\left(\left(\begin{array}{c} \hat{\alpha} \\ \hat{\beta} \end{array}\right), \left(\begin{array}{cc} \hat{\sigma}_\alpha^2 & \rho_{\alpha\beta}\hat{\sigma}_\alpha\hat{\sigma}_\beta \\ \rho_{\alpha\beta}\hat{\sigma}_\alpha\hat{\sigma}_\beta & \hat{\sigma}_\beta^2 \end{array}\right)\right), \chi^2(k-2),$$

where the slope and the intercept from linear regression are mutually correlated. Then the correct covariance and variance of unannounced results would be:

$$
\begin{aligned}
\mathrm{cov}(p_i, p_j) &= E_{(\alpha,\beta,\sigma_\epsilon^2)}\left[\mathrm{cov}(p_i, p_j|\alpha, \beta, \sigma_\epsilon^2)\right] \\
&\quad + \mathrm{cov}_{(\alpha,\beta,\sigma_\epsilon^2)}\left[E(p_i|\alpha, \beta, \sigma_\epsilon^2), E(p_j|\alpha, \beta, \sigma_\epsilon^2)\right] \\
&= \mathrm{cov}_{(\alpha,\beta,\sigma_\epsilon^2)}\left[\alpha\mu_i + \beta, \alpha\mu_j + \beta\right] \\
&= \mu_i\mu_j\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + [(\mu_i + \mu_j)\rho_{\alpha\beta}\hat{\sigma}_\alpha\hat{\sigma}_\beta] \\
\mathrm{var}(p_i) &= \hat{\sigma}_\epsilon^2 + \mu_i\mu_j\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + [(\mu_i + \mu_j)\rho_{\alpha\beta}\hat{\sigma}_\alpha\hat{\sigma}_\beta].
\end{aligned}
\tag{B.1}
$$

However, an assumption that $\hat{\alpha}$ and $\hat{\beta}$ are uncorrelated appears to be used when in fact they are close to being perfectly anti-correlated with[14]

$$\rho_{\alpha\beta} = \frac{-\sum_1^k \mu_i}{\sqrt{n\sum_1^k \mu_i^2}} = -\frac{\bar{\mu}}{\sqrt{\bar{\mu^2}}} \approx -1.$$

The calculation in Wu et al. (2017) would thus be calculating the variance structure of the unknown referendum results as:

$$
\begin{aligned}
\mathrm{cov}(p_i, p_j) &= \mu_i\mu_j\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 \\
\mathrm{var}(p_i) &= \sigma_\epsilon^2 + \mu_i\mu_j\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2,
\end{aligned}
$$

---

[14]A simulation of their results yielded $\rho_{\alpha\beta} \in (-0.97, -1)$.

which is different to the values in equation B.1.

Another issue with the method is that the assumption of a normal distribution for $(\hat{\alpha}, \hat{\beta})$ with correctly specified errors is highly questionable, particularly in small samples. $(\hat{\alpha}, \hat{\beta})$ only follows such a distribution in finite samples if the errors are normal, and if not, it would be biased but consistent. The normal distribution is an asymptotic result for non-normal errors, and even then a correct evaluation assumes no heteroskedasticity; otherwise error estimates are likely to be too low and implied probabilities of Brexit to be too confident. This could be overcome using robust errors, but only for large datasets. Using robust errors in small samples can produce severely biased estimators.

The model in Wu et al. (2017) and that presented in this paper use different approaches to estimate the covariance structure of the conditional distributions used to form predictions. That of Wu et al. (2017) requires no prior (beyond expectations) and attempts to infer the covariance structure from an OLS regression of results announced so far. Our model, by contrast, starts with a prior for the covariance structure and updates that prior as results come in. Both methods will produce the same covariance and results in larger samples but will be different for small samples. The different approach is illustrative of the differences between a Frequentist and Bayesian approach to inference. However, we suggest that the Frequentist approach presented in Wu et al. (2017) is not appropriate for the small numbers of results available at the times of predictions (<20 results). More sophisticated corrections for small sampling estimation would be desirable. We believe our Bayesian approach is a more suitable way to proceed in the case of this application.

Acknowledgments

References

Badarinathi, R., & Kochman, L. (1996). Football betting and the efficient market hypothesis. *The American Economist*, *40*, 52–55.

Baxter, M. (2017). Electoral Calculus; Historical Data and Plots. URL: `www.electoralcalculus.co.uk` Retrieved July 2017.

Caballero, R. J., & Simsek, A. (2016). *A Model of Fickle Capital Flows and Retrenchment*. Working Paper 22751 National Bureau of Economic Research.

Curtice, J., Fisher, S. D., & Kuha, J. (2011). Confounding the commentators: how the 2010 exit poll got it (more or less) right. *Journal of Elections, Public Opinion and Parties*, *21*, 211–235.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, *64*, 709–749.

Ding, P. (2016). On the conditional distribution of the multivariate t distribution. *The American Statistician*, *70*, 293–295.

Econoday (2016). 2016 Economic Calendar. URL: `http://www.econoday.com`.

Electoral Commission (2016). EU Referendum Results. URL: `www.electoralcommission.org.uk` retrieved July 2016.

Electoral Commission (2017). Previous General Elections. URL: `www.electoralcommission.org.uk` retrieved July 2017.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*, 34–105.

Foster, A. (2016). Brexit: How important will turnout be to the EU referendum result? The Express. Retrieved July 2017.

GabrielaTitan, A. (2015). The efficient market hypothesis: Review of specialized literature and empirical research. *Procedia Economics and Finance*, *32*, 442–499.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, *23*, 51–66.

Graefe, A. (2016). Political markets. *Forthcoming (subject to changes) in the SAGE Handbook of Electoral Behavior*, .

Gutteridge, N. (2016). Working class revolution? Reports of huge EU referendum turnout which "would favour leave". The Express. Retrieved August 2016.

Hanretty, C. (2016). The EU referendum: what to expect on the night. Blog. Retreived July 2016.

Hirshleifer, D., Hsu, P.-H., & Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics*, *107*, 632–654.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, *64*, 2289–2325.

Horn, C. F., Ivens, B. S., Ohneberg, M., & Brem, A. (2014). Prediction markets–a literature review 2014. *The Journal of Prediction Markets*, *8*, 89–126.

Hou, K., Xiong, W., & Peng, L. (2009). A tale of two anomalies: The implications of investor attention for price and earnings momentum, .

Huang, J. Y., Shieh, J. C., & Kao, Y.-C. (2016). Starting points for a new researcher in behavioral finance. *International Journal of Managerial Finance*, *12*, 92–103.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, *17*, 59–82.

Manski, C. F. (2004). *Interpreting the Predictions of Prediction Markets*. Working Paper 10359 National Bureau of Economic Research.

Rallings, C., Johnston, R., & Thrasher, M. (2008). Changing the boundaries but keeping the disproportionality: The electoral impact of the fifth periodical reviews by the Parliamentary Boundary Commissions for England and Wales. *The Political Quarterly*, *79*, 80–90.

Reyes, B. (2016a). Gibraltar will vote to remain in EU - poll. Gibraltar Chronicle.

Reyes, B. (2016b). Support for Remain strengthens in Gibraltar - poll. Gibraltor Chronicle.

Shapiro, Y. (2016). Eu Referendum: Growing support in Northern Ireland for Brexit but balance of power lies with the undecided. Belfast Telegraph.

Twyman, J. (2016). Analysis: the impact of turnout on the EU referendum. YouGov. Retrieved August 2017.

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, *28*, 426–444.

Wall, M., Costello, R., & Lindsay, S. (2017). The miracle of the markets: Identifying key campaign events in the Scottish independence referendum using betting odds. *Electoral Studies*, *46*, 39–47.

Wikipedia (2017a). Opinion Polling for the United Kingdom European Union membership referendum. URL: `www.wikipedia.org` retrieved July 2017.

Wikipedia (2017b). Opinion polling for the United Kingdom General Election, 2015, 2010, 2005, 2001, 1997, 1992. URL: `www.wikipedia.org` retrieved July 2017.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*, 107–126.

Wu, K., Wheatley, S., & Sornette, D. (2017). The British Pound on Brexit Night: A Natural Experiment of Market Efficiency and Real-Time Predictability. *Swiss Finance Institute Paper no. 17-12*, .

YouGov (2016). YouGov EU Referendum On the Day Poll.