

APPENDIX – ESTIMATION OF BID/ASK SPREAD

In this Appendix, we provide details of the different methods considered to estimate bid/ask spreads: two leading methods from literature and the conservative method implemented. It is important to recognize both the difficulty and consequence of the bid/ask spread estimation. In order to quantify buy-all or sell-all arbitrage opportunities, it is necessary to have the ask and bid price data on-hand. However, PredictIt and the IEM only provide data on the daily level, including high, low, open, and close prices and daily trading volume for each contract. From this information, there is no way to tell whether the price given at a particular time (i.e. open or close) was made at the existing ask price or bid price.

Further, timing of trades within the day is also consequential, as prices are constantly fluctuating. If Joe Biden trades high in the morning and Elizabeth Warren trades high in the afternoon, an arbitrage trader will not be able to sell contracts at each of these high prices to execute an arbitrage trade without a great deal of luck. Put another way, traders that recognize and immediately execute trades are likely not trading each of those contracts at the daily high.

Due to the observed existence of sell-all arbitrage in PredictIt markets, the emphasis here will be the estimation of bid prices. As the bid prices are needed to estimate the sell-all arbitrage profits, it is important to calculate these conservatively, as not to overstate the arbitrage profit opportunity observed by a trader at that time. The following three algorithms represent methods of estimating the bid/ask spread so that the arbitrage profit opportunity can be estimated as accurately as possible. In an effort to estimate the bid/ask spread at a particular point in time, all three methods estimate the price spread at the end of a trading day. This is an arbitrary choice (compared to the beginning or any other point in the day), but remains consistent among the methods.

1 CORWIN AND SCHULTZ (2012)

The high-low price estimator proposed by Corwin and Schultz (2012) estimates the price spread at the end of a trading day (Day n) by considering the price to be a diffusion process with constant spread and variance over two consecutive trading days (Day n & $n+1$). The method of Corwin and Schultz (2012) differs from the prominent Roll (1984) metric by considering high & low prices, as well as close prices. This provides a richer dataset from which to estimate the bid-ask spread.

The bid-ask spread estimation proposed by Corwin and Schultz (2012) was originally presented in terms of log-prices, reflecting the pricing of stock prices in the positive semi-infinite interval: $x \in (0, \infty)$. Their method has been adapted here to use arithmetic (rather than logarithmic) operations, reflecting

the pricing of prediction market contracts on the unit interval: $x \in [0, 1]^* p$. The algorithm is presented here in terms of daily low (L), high (H), and close (C) prices, with the subscripts (n or $n + 1$) referring to the day, the accent-bar ($\bar{\cdot}$) indicating an adjusted quantity, and the tilde ($\tilde{\cdot}$) indicating an intermediate quantity. The daily low and high prices of the following day are shifted such that they bound the close price of the current day:

$$\text{If } L_{n+1} > C_n : \delta = L_{n+1} - C_n \rightarrow \begin{pmatrix} \bar{L}_{n+1} \\ \bar{H}_{n+1} \end{pmatrix} = \begin{pmatrix} L_{n+1} - \delta = C_n \\ H_{n+1} - \delta \end{pmatrix}$$

$$\text{If } H_{n+1} < C_n : \delta = C_n - H_{n+1} \rightarrow \begin{pmatrix} \bar{L}_{n+1} \\ \bar{H}_{n+1} \end{pmatrix} = \begin{pmatrix} L_{n+1} + \delta \\ H_{n+1} + \delta = C_n \end{pmatrix}$$

$$\tilde{\beta}_n = (H_n - L_n)^2 + (\bar{H}_{n+1} - \bar{L}_{n+1})^2$$

$$\tilde{\gamma}_n = (\max(H_n, \bar{H}_{n+1}) - \min(L_n, \bar{L}_{n+1}))^2$$

$$\tilde{\alpha}_n = \frac{\sqrt{\tilde{\beta}_n} (\sqrt{2} - 1)}{3 - 2\sqrt{2}} - \sqrt{\frac{\tilde{\gamma}_n}{3 - 2\sqrt{2}}}$$

$$S_n = \max\left(0, 2 \frac{\exp(\tilde{\alpha}_n) - 1}{\exp(\tilde{\alpha}_n) + 1}\right)$$

$$\alpha_n^{cs} = C_n + \frac{S_n}{2}, \beta_n^{cs} = C_n - \frac{S_n}{2}$$

where S is the estimated bid-ask spread that can be used to approximate the bid (β^{cs}) and ask (α^{cs}) prices.

2 ABDI AND RANALDO (2017)

The bid-ask spread estimate proposed by Abdi and Ranaldo (2017) is conceptually similar to the autocovariance spread estimate developed by Roll (1984). Their approach differs from that of Roll (1984) by considering the covariance of mid-to-close and close-to-mid price ranges of two consecutive trading days centering, rather than close-to-close price ranges. Like Corwin and Schultz (2012), this method takes advantage of daily high/low trade prices to improve the accuracy of the bid-ask spread estimation. The Abdi and Ranaldo (2017) method is presented below, again using arithmetic operations

with the prices rather than logarithmic ones reflecting the limited price range on the unit interval.

$$M_n = \frac{L_n + H_n}{2}, M_{n+1} = \frac{L_{n+1} + H_{n+1}}{2}$$

$$\tilde{C}_n = \frac{M_n + M_{n+1}}{2}$$

$$\tilde{\gamma}_n = \max\left(0, (C_n - M_n) \cdot (C_n - M_{n+1})\right)$$

$$S_n = 2\sqrt{\tilde{\gamma}_n}$$

$$\alpha_n^{ar} = \tilde{C}_n + \frac{S_n}{2}, \beta_n^{ar} = \tilde{C}_n - \frac{S_n}{2}$$

where S is the estimated bid-ask spread that can be used to approximate the bid (β^{ar}) and ask (α^{ar}) prices.

3 IMPLEMENTED METHOD

The method implemented in our paper estimates the bid price directly from the close price with some corrections such that the price is reasonably bounded. The first correction applied is that the price may not exceed the daily high price minus the price increment. As the daily high price is assumed to derive from buy order matched at the prevailing ask price (Corwin and Schultz, 2012), this correction acts to ensure that the close price could possibly be from a sell order matched to the prevailing bid price (i.e. is not certainly a buy order at ask price). The second correction ensures that the bid price be non-negative.

$$\hat{\beta}(bid) = \max\left(0, \min(close, high - increment)\right)$$

where the price increment is 1ϕ for PredictIt and 0.1ϕ for IEM. Similarly, the ask price can be estimated using the low price plus price increment and payoff price as lower and upper bounds:

$$\hat{\alpha}(ask) = \min\left(p, \max(close, low + increment)\right)$$

4 COMPARISON

The implemented method was selected specifically due to its simplicity and its conservatism. While there is a very high correspondence between the bid

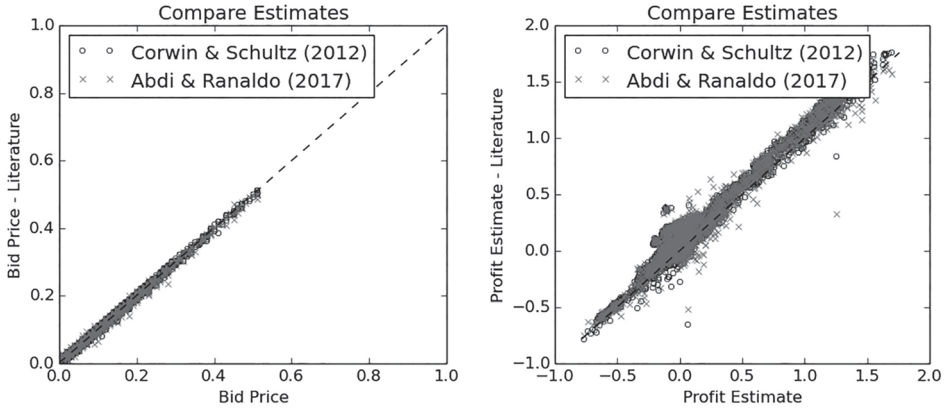


Figure A.1. Comparison of implemented bid price estimate to leading estimates from literature: Corwin and Schultz (2012), Abdi and Rinaldo (2017), as described in this Appendix. At left, a comparison of daily bid price estimates for all candidates from the 2020 Democratic nomination market hosted by PredictIt. At right, a comparison of the daily arbitrage profit estimate for all markets studied.

prices estimated from implemented method and those considered from literature (see Figure A.1, left), the differences between the prices compound when combined to form the estimate arbitrage profit. The arbitrage profit estimated using these three methods similarly shows a high correlation, but there is significantly more variation between them and the profit tends to be somewhat higher (see Figure A.1, right).

This higher predicted profit is illustrated in Figure A.2 for the 2020 Democratic and Republican party nomination markets hosted by PredictIt.

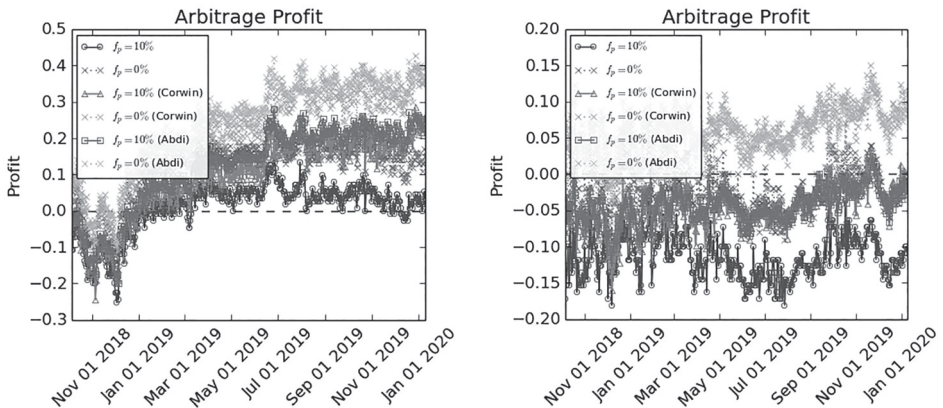


Figure A.2. Comparison of estimated arbitrage profit in the 2020 Democratic (left) and Republican (right) nomination markets hosted by PredictIt. Profit estimates are calculated with and without profit fees ($f_p = 10\%$, 0%) and using the implemented bid price estimate and two from literature, as described in this Appendix.

Using the implemented bid price estimate, the Democratic market shows both market mispricings ($\Pi(f_p = 0\%) > 0$) and a small arbitrage profit ($\Pi(f_p = 10\%) > 0$) since early 2019, and the Republican market shows only occasional mispricings and no arbitrage profit. When the bid price estimates for literature are used, mispricings are predicted for both markets and significant ($>10\text{¢}/\text{share}$) arbitrage profits in the Democratic nomination market for nearly the entire time period.

The authors have observed that the mispricing & arbitrage profit predictions using the implemented bid price estimation tend to agree with the mispricing/arbitrage profit state. In contrast, the literature-based estimates tend to overstate the mispricing and arbitrage profits. As this paper focuses on the existence, duration, and magnitude of mispricing and arbitrage profits, it is important to (1) be as accurate as possible and (2) be conservative in estimations when possible. Our experience leads us to conclude that the implemented bid price estimation method fits these criteria better than the two explored alternatives.