Bitcoin Price Discovery

Eric Ghysels[1] Giang Nguyen[2]

[1]University of North Carolina – Chapel Hill

[2]Penn State University

NBER Big Data and High-Performance Computing for Financial Economics

Cambridge, MA

July 14, 2018

Paper Overview

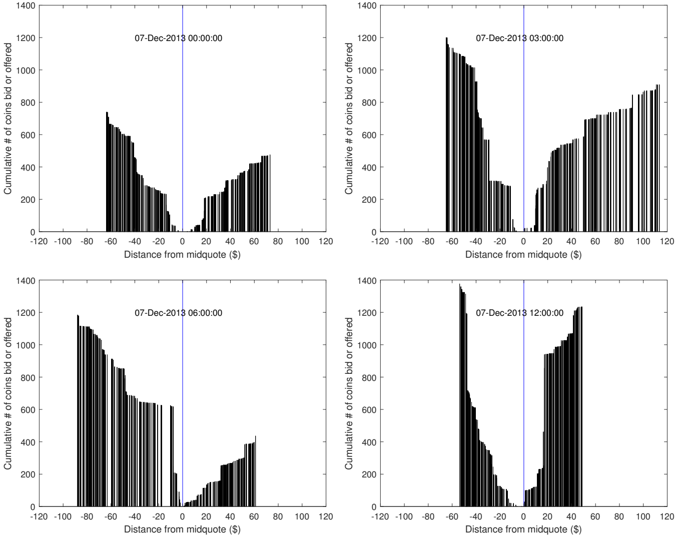
- Comprehensive study of microstructure of a bitcoin trading platform
  - Bitcoin is traded around the clock on many exchanges globally
  - Exchange design: limit order market (traders provide and take liquidity)
- Full limit order book snapshot data at high frequency allow for investigating:
  1. Does information content of orders increase with order aggressiveness?
  2. Does information asymmetry worsen liquidity?
  3. Is learning in market non-Markovian?

Contribution to literature on dynamic limit order markets

- Modeling limit order market is highly complex
- Current theoretical studies have to impose restrictive assumptions to focus on a certain dynamic and make model tractable
- Most empirical studies of limit order markets focus on trades and limit orders at the top tier(s) only
- Our full order book data (150 price levels on each side) showing complete supply and demand schedules:
  - shed lights on dynamics of limit orders behind the best quote
  - reflect realistic action space available to traders
  - provide a complete view of market liquidity
  - can be insightful for theory development and/or interpretation
- Bitcoin LOB market: excellent laboratory to test LOM theories (free market place, no exchange rules on minimum order size, tick size, level playing field in terms of pre-trade transparency, etc...)

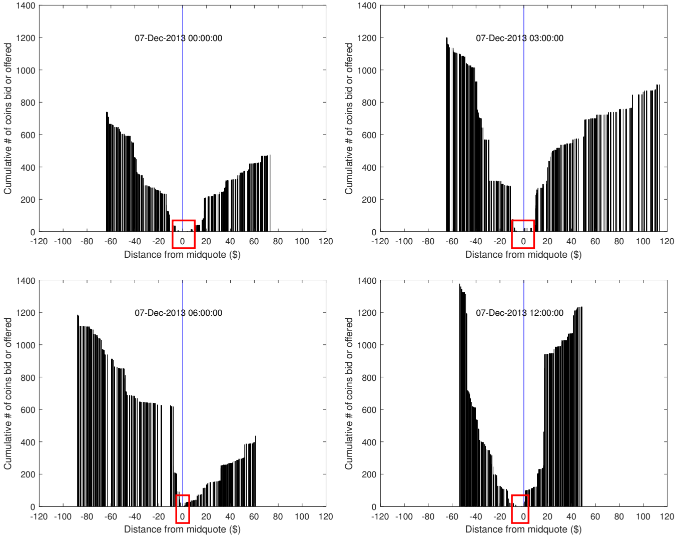
Limit Order Book Snapshots

The multiple facets of liquidity



Limit Order Book Snapshots

Top few layers of book do not tell the whole story



Price Discovery in Dynamic Limit Order Market

Theories:

- Goettler, Parlour, and Rajan (JFE2009): information content of limit order book depends on informed traders' order strategies
- Rosu (WP2016): more informed traders improve market learning and narrow bid-ask spread
- Ricco, Rindi, and Seppi (WP2018): price discovery and liquidity depends on nature of adverse selection (large value shock or greater fraction of informed traders), price discovery is history-dependent

Experimental:

- Bloomfield, O'Hara, and Saar (JFE2005): informed traders use more market orders (taking liquidity) when value shock is large, but shift to use limit orders (providing liquidity) when value shock is small

Data

- BTC-e cryptocurrency trading platform
- Sample period: 12/7/2013 - 9/24/2014
- Currency pair: BTC/USD
- BTC-e was a major bitcoin exchange at the time (approx. 20% share of global bitcoin trading volume)
- BTC-e (together with Mt. Gox) leads other exchanges in price discovery (Brandvold, Molnar, Vagstad, & Valstad, 2015)

Data Collection

- Data collected by Jacob Sagi by directly accessing BTC-e's servers
- Algorithm pings servers every 0.1 second:
  - takes snapshot of limit order book up to **150** price levels on each side
  - downloads transaction history (last 150 transactions)
- Two computers independently download data $\rightarrow$ two parallel (similar but not exactly the same) datasets spanning 292 days
- Transaction history datasets: merged and duplicates removed
- Snapshot datasets: merging is complex (need to maintain correct sequencing of snapshots given varying latency of each computer)
- Final dataset: complete view of limit order book at ultra-high frequency (sub-second) to allow most comprehensive study of dynamics of liquidity provision

Descriptive Statistics of Limit Order Book

	Tier 1	Tier 5	Tier 10	Tier 20	Tier 50	Tier 100	Tier 150
Panel A: Distribution of Depth Across Price Tiers							
Ask: Cum. Depth	4.1	17.5	30.0	51.3	108.3	203.5	322.3
Ask: % Cum. Depth	1.3	5.4	9.2	15.6	33.0	61.9	100.0
Bid: Cum. Depth	2.6	11.1	19.8	36.3	89.3	190.6	321.3
Bid: % Cum. Depth	0.8	3.5	6.2	11.3	27.2	58.1	100.0
Panel B: Spreads as Fraction of Bid-Ask Midpoint (bps)							
Ask: Distance from Best Bid	19.5	36.7	47.3	61.8	94.2	139.2	181.4
Ask: Volume-weighted Spread	19.5	27.6	33.9	42.8	62.2	88.6	116.6
Bid: Distance from Best Ask	19.5	36.1	46.5	61.1	95.0	144.9	194.3
Bid: Volume-weighted Spread	19.5	27.7	33.9	43.2	64.2	94.4	126.2

Descriptive Statistics of Trading Activity

	Buyer-initiated Trades				Seller-initiated Trades			
	Mean	P5th	Median	P95th	Mean	P5th	Median	P95th
Trade Frequency	6,710	1,712	4,413	15,220	6,197	1,179	3,648	15,427
Volume (# BTC)	5,626	873	3,321	18,584	5,752	795	3,065	18,808
Dollar Volume ($ m)	3.44	0.49	1.939	11.21	3.51	0.44	1.76	11.27
Trade Size (# BTC)	0.84	0.01	0.10	3.25	0.93	0.01	0.10	3.78
Dollar Trade Size ($)	512.62	5.13	56.11	2,007.58	566.87	5.50	59.93	2,291.98

Hypothesis 1

Information content of more aggressive orders **increase** in high-volatility environment

- Market order (immediate execution but costly) vs. limit order (earn the spread but incur waiting cost)
- Large value shock: informed traders use market and most aggressive limit orders to realize trading profits
  - market orders and aggressive limit orders have high information content
- Low value shock: informed traders choose less aggressive limit orders
  - less aggressive orders have higher information content

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

Empirical strategy:

1. Identification of large value shock environment
   - Theories: value shock size important for informed traders' strategies $\rightarrow$ important for information content of different order types
2. Measurement of information content of different order types
3. Test for changes in information content of different orders type in high value shock environment and low value shock environment, benchmarked by "normal" environment

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

1. Identification of large value shock environment:

- High-low range: proxy for return earned by informed traders with perfect information who buys at lowest and sells at highest

- Realized volatility (sqrt of sum of squared 5-minute returns)

- Partition sample into 3 sub-samples:

   1. High value shock days (62): Hi-lo range $\geq$ Q3 AND RV $\geq$ Q3
   2. Low value shock days (58): Hi-lo range $\leq$ Q1 AND RV $\leq$ Q1
   3. Average days (172): rest of sample

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

1. Identification of large value shock environment (cont'd): verify with news analysis

News Type	High N=62	Average N=172	Low N=58
Market Acceptance	2	4	1
Regulatory	14	14	0
Security/Hack	15	16	1
Total days with news	31	34	2

Table: Comprehensive search of news articles on cryptocurrency-related events from Bloomberg, Reuters, and popular crypto websites CCN and CoinDesk

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

2. Measuring information content of different order types:

- Ideally: measure information content of limit orders at **all** 150 price levels on each side
- Challenge: not econometrically feasible
- Solution: group limit orders to 6 categories from most aggressive to most conservative: Tier 1, Tier 2-5, Tier 6-10, Tiers 11-50, Tiers 51-100, Tiers 101-150. Price of each order group = depth-weighted average price of orders within group
- Information content of market orders and 6 limit order categories: measured by how their prices (cointegrated) drive the underlying efficient price process

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

2. Measuring information content of different order types: VECM(10) estimated separately for each day on one-minute snapshot data

$$\Delta \mathbf{X}_t = \alpha z_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta \mathbf{X}_{t-j} + \epsilon_t,$$

where $\mathbf{X}_t \equiv \left[P^T,\, P^1,\, P^2,\, P^3,\, P^4,\, P^5,\, P^6\right]_t'$, and z_{t-1} is a 6×1 vector of correction terms:

$$z_{t-1} = \begin{bmatrix} P^T & - & \beta_2 P^1 \\ \dots & & \\ P^T & - & \beta_6 P^6 \end{bmatrix}_{t-1}$$

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

2. Measuring information content of different order types: Hasbrouck (1995)'s information shares

$$IS_j = \frac{\left[\sum_{i=j}^{n} \gamma_i m_{ij}\right]^2}{\left[\sum_{i=1}^{n} \gamma_i m_{i1}\right]^2 + \left[\sum_{i=2}^{n} \gamma_i m_{i2}\right]^2 + \ldots + \left[\gamma_n m_{nn}\right]^2},$$

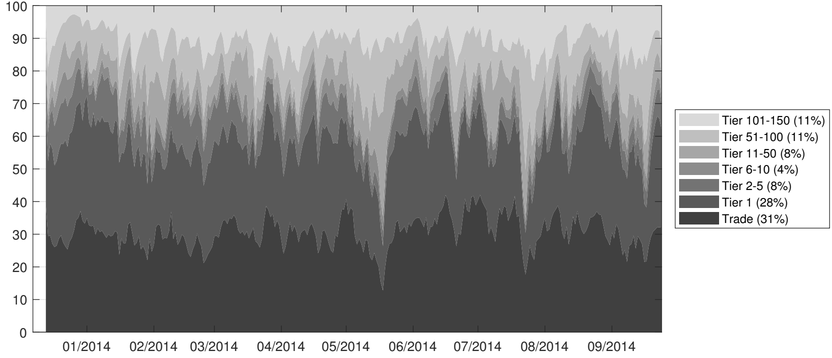
where:

- γ_i is the permanent price impact of shock i (from MA(∞) representation of VECM)
- m_{ij} is the (i, j) element of the lower triangular matrix M from Choleski decomposition of covariance matrix of residuals Ω $(MM' = \Omega)$
- IS in words: contribution of a price series' innovation variation to the variation of the underlying efficient price updates

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

2. Measuring information content of different order types: information share estimates over time

- Trade and then limit order at best quote: most informative
- Information content lowest in mid book, higher at far-away tiers



Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

3. Information shares on high vs. normal volatility environment

Statistic	Trade	Tier 1	Tiers 2-5	Tiers 6-10	Tiers 11-50	Tiers 51-100	Tiers 101-150
			A1. High volatility days (N=62)				
Mean	30.05	31.76	13.44	5.24	5.55	7.60	6.37
S.e.	1.11	1.03	0.70	0.39	0.51	0.90	0.88
			A2. Average volatility days (N=172)				
Mean	31.98	27.38	6.88	2.79	7.62	11.33	12.02
S.e.	0.90	0.79	0.37	0.22	0.63	0.92	0.94
			B1. Test of A1 $\neq$ A2				
t-stat	-1.357	**3.366**	**8.303**	**5.496**	**-2.552**	**-2.910**	**-4.381**
p-val	0.088	0.000	0.000	0.000	0.006	0.002	0.000

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

3. Information shares on low vs. normal volatility environment

Statistic	Trade	Tier 1	Tiers 2-5	Tiers 6-10	Tiers 11-50	Tiers 51-100	Tiers 101-150
			A3. Low volatility days (N=58)				
Mean	27.94	23.60	5.88	3.97	13.45	12.62	12.53
S.e.	1.72	1.50	0.71	0.54	1.76	1.88	1.45
			A2. Average volatility days (N=172)				
Mean	31.98	27.38	6.88	2.79	7.62	11.33	12.02
S.e.	0.90	0.79	0.37	0.22	0.63	0.92	0.94
			B2. Test of A3 $\neq$ A2				
t-stat	**-2.085**	**-2.225**	-1.242	**2.025**	**3.129**	0.619	0.293
p-val	0.020	0.014	0.109	0.023	0.001	0.269	0.385

Testing Hypothesis 1: Does information content of orders increase with order aggressiveness?

- Yes if large value shock: informed traders $\uparrow$ aggressive limit orders
  - limit orders at or near best quote become more informative
  - far-away orders become less informative
- No if small value shock: informed traders $\downarrow$ market orders and most aggressive limit orders, and instead shift to more conservative (but not too conservative) limit orders
  - informativeness of market orders and best limit orders reduced,
  - informativeness of mid-book limit orders increased
  - no significant change in informativeness of far-away limit orders
- Results consistent with majority of theories

Hypothesis 2

Adverse selection worsens liquidity?

- Rosu (2016): **No** ($\uparrow$ fraction of informed traders $\rightarrow \uparrow$ information learning $\rightarrow \downarrow$ bid-ask spread)
- Ricco, Rindi, & Seppi (2018): **Depends!**
  - **Can be no** ($\uparrow$ value shock $\uparrow$ migration of informed liquidity to best quote, but $\uparrow$ outward migration of uninformed liquidity)
  - **Yes** ($\uparrow$ fraction of informed traders does not change informed's strategies but uninformed liquidity moves away from market)
- Goettler, Parlour, and Rajan (2009):
  - **Yes** (for liquidity at best quote, b/c informed agents use market orders instead)
  - **No** (for liquidity behind best quote, b/c agents submit more conservative limit orders)

Testing Hypothesis 2: Does Adverse Selection Worsen Liquidity?

Empirical strategy:

1. Measuring adverse selection at intraday frequency
2. Measuring movement of liquidity in limit order book
3. Multivariate regression of liquidity on adverse selection, distinguishing high and low value shock regimes

Testing Hypothesis 2: Does Adverse Selection Worsen Liquidity?

1. Measuring adverse selection at intraday frequency:

- Previous estimates of information content: feasible only for low frequency (lot of data needed for estimation)
- Need measure at intraday frequency to examine how it affects liquidity provision
- Previous estimates: information content concentrated at trades and inside limit orders $\rightarrow$ measure adverse selection by price impact of net order flow at best quote (Cont, Kukanov, & Stoikov, 2014):

$$\Delta P_{k,i} = \widehat{\text{Constant}}_i + \widehat{PI}_i \times OFI_{k,i} + \widehat{\epsilon}^{PI}_{k,i},$$

  - $\Delta P_{k,i}$: midquote change over minute k of hourly-interval i
  - $OFI_{k,i}$: order flow imbalance

Testing Hypothesis 2: Does Adverse Selection Worsen Liquidity?

2. Measuring movement of liquidity in the book:

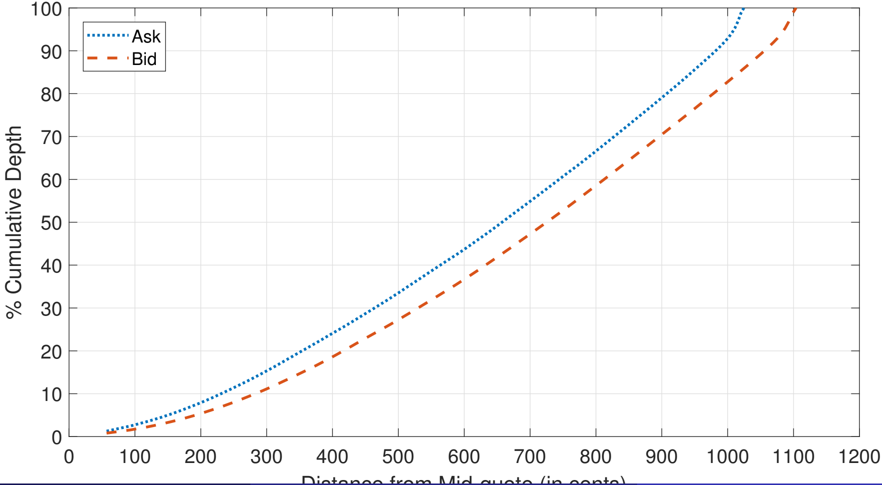
- Limit order book: high dimension
- Many facets of liquidity: spread, depth, distance of depth
- Slope: a comprehensive measure of liquidity distribution in the book
- Change in slope reflects movement of liquidity toward (steepening) or away from best quote (flattening)
- Slope estimated from regression of normalized cumulative depth on price distance from midquote

$$QP_{\tau,i} = \widehat{\text{Constant}}_i + \widehat{SL}_i \times d_{\tau,i} + \widehat{\epsilon}^{SL}_{\tau,i},$$

  - $QP_{\tau,i}$: percent of cumulative depth up to Tier τ as of hour i
  - $d_{\tau,i}$: price distance from the midquote

Slope: a comprehensive measure of how liquidity is distributed

steeper slope = migration of liquidity **toward** best quote
flatter slope = migration of liquidity **away** from best quote



Testing Hypothesis 2: Does Adverse Selection Worsen Liquidity?

3. Multivariate regression of liquidity on adverse selection and controls:

	Dep. Variable = Ask Slope				Dep. Variable = Bid Slope			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PI	-0.40^{**}	-0.41^{**}	-0.46^{***}	-0.46^{***}	-0.48^{***}	-0.48^{***}	-0.51^{***}	-0.53^{***}
PI x hivol	-0.68^{**}	-0.68^{**}	-0.69^{**}	-0.68^{**}	-0.99^{***}	-1.05^{***}	-1.05^{***}	-0.99^{***}
PI x lovol	2.24^{***}	2.23^{***}	2.27^{***}	2.25^{***}	2.40^{***}	2.43^{***}	2.44^{***}	2.39^{***}
Control Variables:								
Realized Volatility	-2.90^{***}	-2.89^{***}	-2.86^{***}	-2.57^{***}	-2.68^{***}	-2.57^{***}	-2.58^{***}	-2.42^{***}
Opposite Slope	0.32^{***}	0.32^{***}	0.32^{***}	0.32^{***}	0.18^{***}	0.18^{***}	0.18^{***}	0.18^{***}
% Depth at Top Tier			-0.76^{***}	-0.75^{***}			-0.65^{***}	-0.65^{***}
Total Ask Depth (logged)	1.17^{*}	1.05	0.90	0.82	4.39^{***}	4.15^{***}	4.02^{***}	3.82^{***}
Total Bid Depth (logged)	-6.15^{***}	-6.16^{***}	-6.12^{***}	-6.25^{***}	-7.84^{***}	-7.95^{***}	-8.06^{***}	-8.04^{***}
Buyer-initiated Trade Volume (logged)	2.95^{***}	2.67^{***}	2.92^{***}	2.60^{***}	-7.27^{***}	-5.61^{***}	-5.50^{***}	-5.84^{***}
Buyer-initiated Trade Count (logged)		1.03	1.19	1.13		-2.92^{***}	-3.11^{***}	-3.04^{***}
Seller-initiated Trade Volume (logged)	-8.29^{***}	-7.46^{***}	-7.53^{***}	-7.67^{***}	1.78^{***}	2.04^{***}	2.10^{***}	2.06^{***}
Seller-initiated Trade Count (logged)		-1.62^{**}	-1.96^{**}	-1.71^{**}		0.52	0.63	0.65
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hourly Dummies	No	No	No	Yes	No	No	No	Yes
Nobs	7,007	7,007	7,007	7,007	7,007	7,007	7,007	7,007
Adjusted R^2	0.39	0.39	0.40	0.38	0.41	0.41	0.42	0.40

Testing Hypothesis 2: Does Adverse Selection Worsen Liquidity?

Answer: Yes (high value shock), No (low value shock)

- High value shock: slope flattens after controlling for depth at Tier 1 and total depth $\rightarrow$ movement of liquidity **away** from market
  - Informed traders $\uparrow$ market orders and $\downarrow$ limit orders
  - Uninformed traders move away from market due to increased adverse selection
- Low value shock: slope steepens $\rightarrow$ liquidity moving **toward** market
  - Informed traders $\downarrow$ market orders and $\uparrow$ limit orders
  - Less adverse selection concern for uninformed traders
  - In low value shock environment: increased adverse selection more likely due to increased fraction of informed traders $\rightarrow$ improve information learning for uninformed $\rightarrow$ improve liquidity – Rosu (2016)

Hypothesis 3

Leaning in market is non-Markovian

- Rosu (2016) and Goettler, Parlour, & Rajan (2009): Markovian learning (traders condition their strategies on current state of market $\rightarrow$ price discovery depends on current market observables)
- Ricco, Rindi, & Seppi (2018): non-Markovian learning (traders condition their strategies on order history, not just current state $\rightarrow$ price discovery depends not only on current market observables but also the path leading to current state)
- How important is it assumption in practice?

Testing Hypothesis 3: Is Price Discovery Non-Markovian?

Empirical strategy:

- If price discovery is non-Markovian, lagged market variables should have explanatory power in addition to current state variables:

$$PI_t = c + \beta_0' Z_t + \sum_{l=1}^{24} \theta_{j,t} Z(j)_{t-l} + \epsilon_t.$$

  - PI_t: price impact of order flow over hour t
  - Z_t: collects variables that capture the state of the order book at beginning of hour t

- Estimate baseline regression containing current state variables only

- Add to baseline specification the 24-hour history of each state variable one at a time to identify which history more important

- Caveat: linear form of dependency $\rightarrow$ rejection of null only tells us: not **linear** history-dependence

Testing Hypothesis 3: Is Price Discovery Non-Markovian?

Baseline regression of price impact on current state variables only

Explanatory Variable	Model 1	Model 2
Ask Slope	-0.000	-0.000
Bid Slope	0.001	0.000
Total Ask Depth (logged)	-0.141^{***}	-0.147^{***}
Total Bid Depth (logged)	0.099^{*}	0.095^{*}
Buy Volume (logged)	-0.214^{***}	-0.231^{***}
Sell Volume (logged)	-0.064^{**}	-0.067^{**}
Realized Volatility	0.642^{***}	0.659^{***}
% Ask Depth at Top Tier	-0.007	-0.007
% Bid Depth at Top Tier	-0.010	-0.010
% Ask Depth at Top 5 Tiers	-0.008^{***}	-0.008^{**}
% Bid Depth at Top 5 Tiers	-0.011^{**}	-0.010^{**}
Hourly Dummies	No	Yes
Adjusted R^2	12.61	11.49

Testing Hypothesis 3: Is Price Discovery Non-Markovian?

Regression of price impact on current state variables and 24-hour history of each state variable

	# Significant Lag Coefficients			
	1% Level	5% Level	10% Level	Adj. R^2
Ask Slope	0	0	0	10.04
Bid Slope	0	0	1	10.08
Total Ask Depth (logged)	0	0	0	9.95
Total Bid Depth (logged)	1	1	3	10.27
Buy Volume (logged)	0	1	4	10.24
Sell Volume (logged)	0	1	1	10.27
Realized Volatility	1	2	3	9.69
% Ask Depth at Top Tier	0	2	3	10.15
% Bid Depth at Top Tier	0	1	2	10.08
% Ask Depth at Top 5 Tiers	0	0	1	10.07
% Bid Depth at Top 5 Tiers	0	0	0	9.98

Testing Hypothesis 3: Is Price Discovery Non-Markovian?

- No evidence to support **linear** dependence of price discovery on history of **individual** state variable
- Other plausible scenarios: history dependence could be of some non-linear form, on some combination of all state variables
- As a first pass: results indicate the Markovian assumption of market learning might be reasonable $\rightarrow$ important because this assumption allows theorists to simplify the state space significantly

Conclusion

- Study price discovery & liquidity in a bitcoin limit order market
- Important results:
  - Information content of aggressive limit orders increases in high value shock environment, but reduces in low value shock environment while information content moves to mid-book orders $\rightarrow$ empirical support to theoretical/experimental studies of dynamic limit order markets
  - Liquidity flows toward the market in low value shock environment but away from market in high value shock environment $\rightarrow$ adding empirical evidence to help reconcile different theories
  - No supportive evidence of non-Markovian learning in linear sense $\rightarrow$ scope for additional work
- Work in progress: further tests of history dependence of price discovery, explore if technical trading rules can deliver valuable trading signals