1 Descriptive information about the data

1.1 Basic information

In this section, we describe the 1st month of the OPRA data as well as some of their relevant characteristics.

In Table 1 some descriptive statistics are listed. The number of quotes and trades after applying the cleaning algorithm in Section 2 is provided.

Some examples of the traded contracts are shown in Figures 1 - 4. It can be seen that different contracts and underlyings vary significantly in terms of prevailing spreads, amount of transactions, dynamics of different exchanges and in time.

In the fourth panel of Table 1 it can be seen that the majority of traded contracts has maturity 1 week to 1 month in most of the cases, except for securities GOOG, XLE, K and BLK.

In the last panel of Table 1, trades' directions are determined using an algorithm similar to that of Lee and Ready (1991). It can be described as follows.

For each trade price P_t for a certain contract, do the following:

1. determine the prevailing bid and ask quotes Q_t^B and Q_t^A (1st BBO quotes at or immediately preceding time t);
2. compute the prevailing mid-quote $Q_t^M = (Q_t^B + Q_t^A)/2$;
3. then decide:
   (a) if $P_t > Q_t^M$, then the trade at time t is buyer-initiated (BI)
   (b) if $P_t < Q_t^M$, then the trade at time t is seller-initiated (SI)
   (c) if $P_t = Q_t^M$, then decide based on the following:
      i. if $P_t < P_{t-1}$, then the trade is SI
      ii. if $P_t > P_{t-1}$, then the trade is BI
      iii. if $P_t = P_{t-1}$, then one should sequentially check the differences between P_t and $P_{t-2}, \dots, P_1$ until the first such difference is strictly smaller (SI) or larger (BI) than zero;
4. if no condition from the above is satisfied, the the trade direction cannot be determined (UI).

As follows from Table 1, in most of the cases the number of buyer-initiated trades is larger than of those which are seller-initiated.

It must be noted that we do not include contracts which are "trades-only" into the calculation as these would make impossible the use of the algorithm above.

1.2 Liquidity measures

We further calculate a number of liquidity measures for contracts on the given underlying securities, proposed, e.g., in Huang and Stoll (1996), Hasbrouck (2009), Goyenko et al. (2009). The liquidity measures we consider are divided into 2 categories:

1. spread measures: Effective Spread, Realized Spread;
2. price impact measures: λ Price Impact, 5-Minute Price Impact.

The spread measures are intended to measure relationships either between trade prices and prevailing BBO mid-quotes at a given time or between trade prices at different time points. The price impact measures capture liquidity implications of transactions on the market.

The first measure, Effective Spread, is defined as follows:

$$ES_t \stackrel{\text{def}}{=} 2 \cdot |\log(P_t) - \log(Q_t^M)|, \tag{1}$$

where P_t, Q_t^M are as defined above.

The second measure, Realized Spread, is written in the following way:

$$RS_t \stackrel{\text{def}}{=} \begin{cases} 2 \cdot \{\log(P_t) - \log(P_{t+5})\}, & \text{if the } t\text{th trade is BI} \\ 2 \cdot \{\log(P_{t+5}) - \log(P_t)\}, & \text{if the } t\text{th trade is SI,} \end{cases} \tag{2}$$

where P_{t+5} is the price of the trade 5 minutes after the tth trade.

The next measure, λ Price Impact, is a a regression-based liquidity effect measure: it is defined as the slope coefficient of the following equation without a constant:

$$r_k = \lambda \cdot S_k + u_k, \tag{3}$$

where the log-return r_k is computed for the kth 5-minute period, u_k is the error term; S_k is the signed square-root dollar volume:

$$S_k \stackrel{\text{def}}{=} \sum_{t=1}^{T_k} Sign(V_t^k)\sqrt{V_t^k}, \tag{4}$$

where T_k is the number of trades in the kth 5-minute period, V_t^k is the dollar volume of the tth trade in the kth interval, and

$$Sign(V_t^k) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if the } t\text{th trade in the } k\text{th interval is BI,} \\ 0, & \text{if the } t\text{th trade in the } k\text{th interval is UI,} \\ -1, & \text{if the } t\text{th trade in the } k\text{th interval is SI.} \end{cases} \tag{5}$$

The last price impact measure, 5-Minute Price Impact, is written as follows:

$$PI_t \stackrel{\text{def}}{=} \begin{cases} 2 \cdot \left\{\log(Q_{t+5}^M) - \log(Q_t^M)\right\}, & \text{if the } t\text{th trade is BI} \\ 2 \cdot \left\{\log(Q_t^M) - \log(Q_{t+5}^M)\right\}, & \text{if the } t\text{th trade is SI,} \end{cases} \tag{6}$$

where Q_{t+5}^M is the midpoint of the BBO bid and ask quotes prevailing 5 minutes after the tth trade.

These liquidity measures are computed for each contract which has 2 or more trades in it. Contracts with trades only (without quotes) are also excluded due to the lack of quote entries. Descriptive statistics presented in Table 2, are calculated over 20 days and all eligible contracts for each underlying security in the 1st month of data. Statistics on the eligible contracts are given in the bottom panel of Table 2.

From Table 2 follows an obvious result that options on more actively traded stocks such as SPY, AAPL, GOOG produce lower values of spread liquidity indicators with average Effective Spread being equal to 0.053, 0.047 and 0.058, respectively which is still much higher than monthly and annual averages obtained by Goyenko et al. (2009) who used stock data in the NYSE TAQ database in the period from 1993 to 2005.

Interestingly enough, comparing a security with a relatively high amount of option trade and quote entries such as XLE, with a relatively illiquid one such as K, we find that their Effective and Realized spread measures are very close to each other. Also XLE has fewer trades than, e.g., XOM while having a higher amount of quote records.

Consistent with previous studies such as Goyenko et al. (2009), none of the spread or price impact indicators are of the same order of magnitude as the λ price impact measure. In terms of interpretation, the λ Price Impact for SPY at its median value 249 means that a \$10,000 buy order would move the log price by approximately $\sqrt{10,000} \times 249 \times 10^{-6} = 0.0249$, i.e., 2.5%. Median values of λ Price Impact obtained here are on average an order of magnitude larger than those presented in Goyenko et al. (2009) which is due to the fact that options normally have much smaller prices than the underlying stocks.

As the next step in this descriptive study, one could perform analysis similar to that of Goyenko et al. (2009), who tests the hypothesis that low-frequency liquidity measures can reliably estimate high-frequency benchmark indicators.

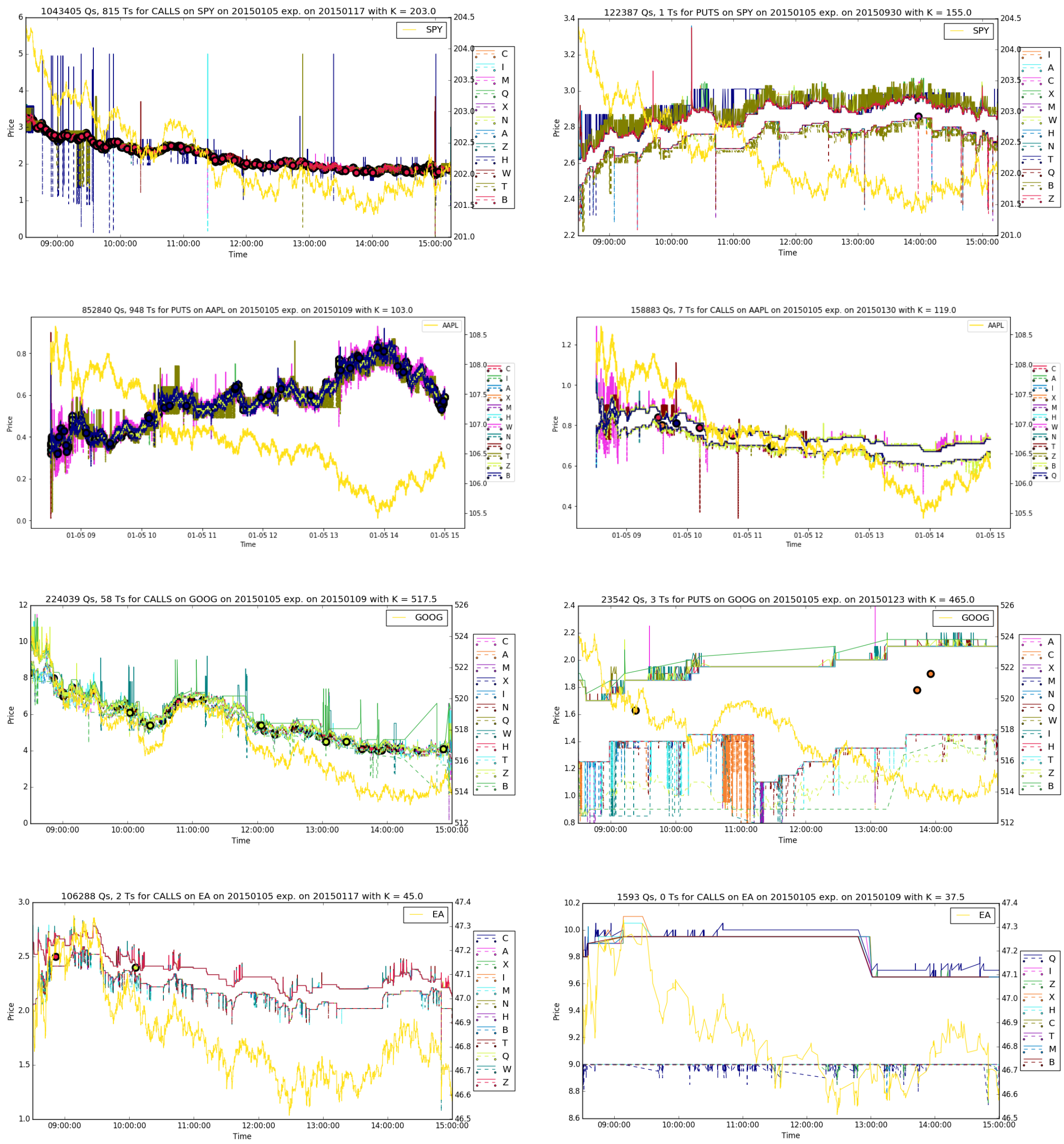

Figure 1: Dynamics of selected contracts on various underlyings on 20150105: bid prices as solid, ask prices as dashed lines, trades as dots; colors denote different exchanges; dynamics of the underlyings in yellow

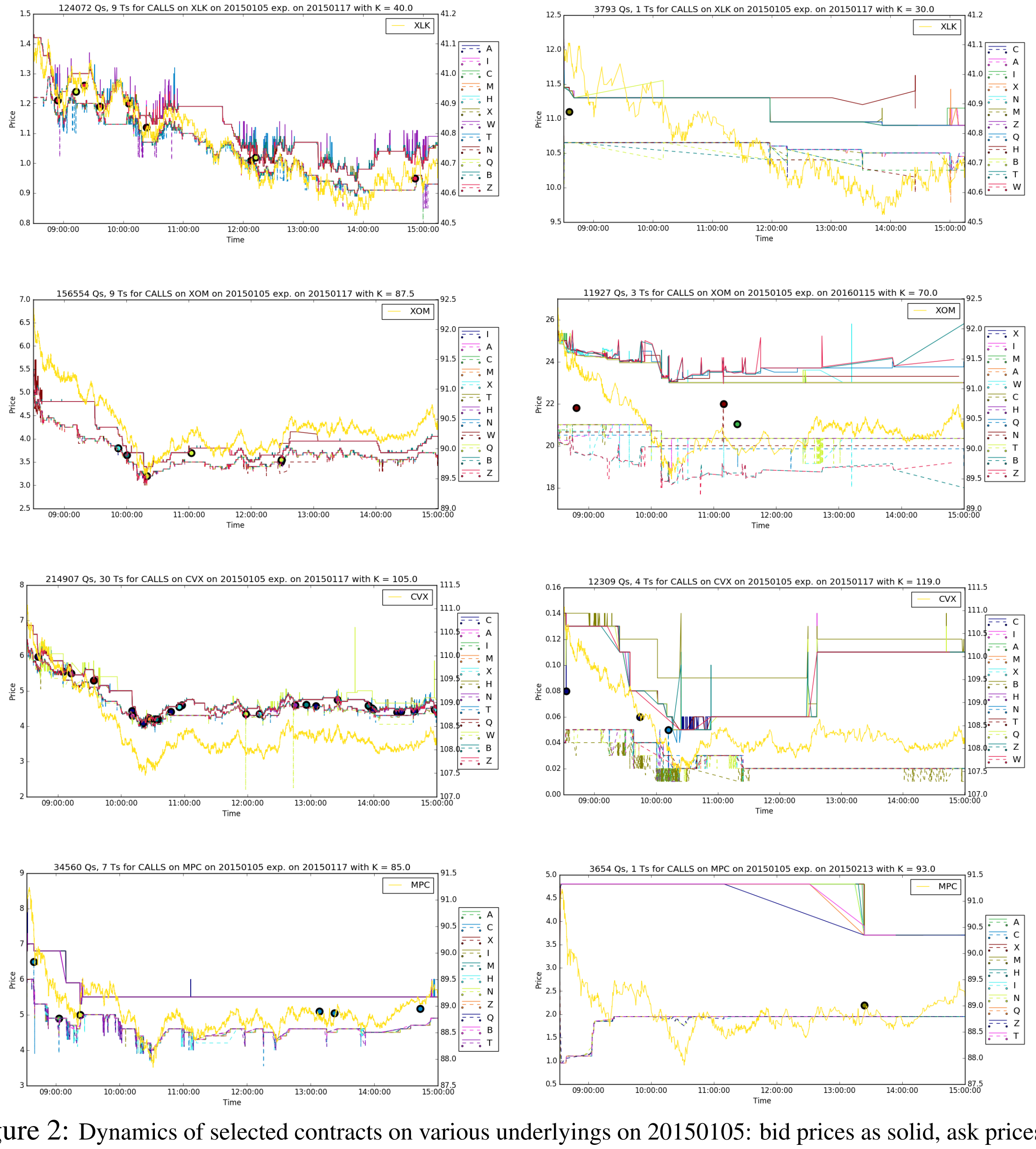

Figure 2: Dynamics of selected contracts on various underlyings on 20150105: bid prices as solid, ask prices as dashed lines, trades as dots; colors denote different exchanges; dynamics of the underlyings in yellow

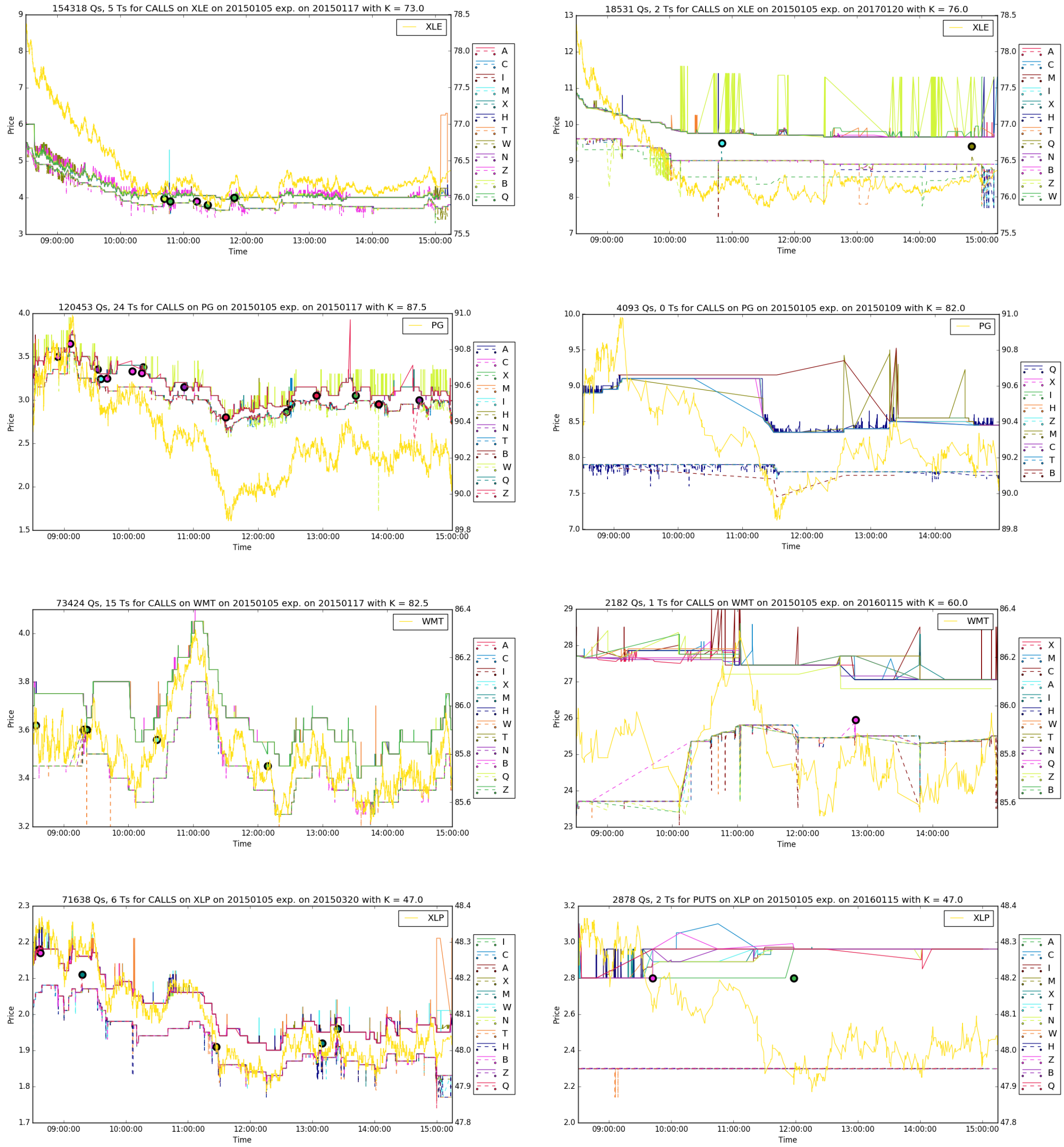

Figure 3: Dynamics of selected contracts on various underlyings on 20150105: bid prices as solid, ask prices as dashed lines, trades as dots; colors denote different exchanges; dynamics of the underlyings in yellow

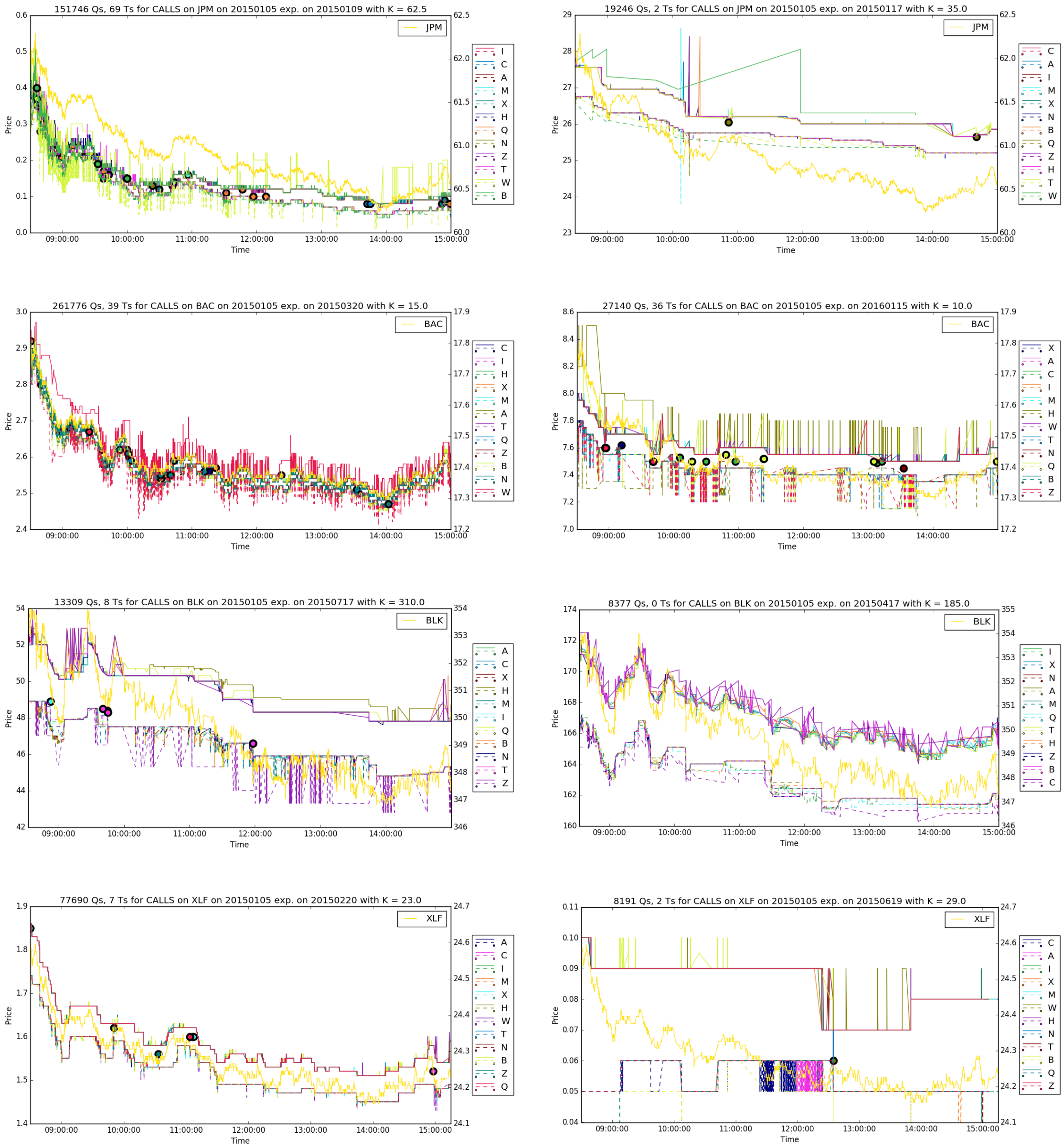

Figure 4: Dynamics of selected contracts on various underlyings on 20150105: bid prices as solid, ask prices as dashed lines, trades as dots; colors denote different exchanges; dynamics of the underlyings in yellow

Table 1: Entries' statistics

	SPY	AAPL	GOOG	EA	XLK	XOM	CVX	MPC	XLE	PG	WMT	K	XLP	JPM	BAC	BLK	XLF
Underlying stocks																	
Number of records (before cleaning)																	
Avg	700,838,009	301,575,942	73,543,546	15,184,625	9,489,613	36,198,297	28,728,799	5,215,386	59,890,142	23,723,427	22,802,939	745,353	8,778,407	43,001,824	36,051,422	3,166,838	11,930,495
Std Dev	139,414,541	78,014,816	17,064,430	4,269,292	2,406,511	7,548,864	5,085,071	1,290,710	11,537,807	8,134,120	6,288,729	120,562	2,341,311	8,747,138	6,725,070	724,910	2,642,094
Median	694,355,725	310,310,810	71,363,687	14,629,507	8,729,688	34,021,556	27,206,324	5,066,729	58,010,337	24,989,843	23,184,688	718,706	8,477,310	41,790,218	34,030,353	3,139,884	11,100,126
Min	409,607,244	159,094,428	45,479,752	9,744,870	5,399,868	24,735,171	21,143,683	3,245,001	45,196,390	9,543,493	12,224,635	574,871	4,851,901	25,718,345	23,281,419	1,968,863	6,549,215
Max	984,010,543	417,942,352	113,190,636	23,388,291	13,105,051	49,712,308	38,714,736	8,314,240	84,261,758	38,230,503	33,970,789	981,853	13,133,224	59,398,468	49,911,031	5,340,371	17,999,368
Number of trades (after cleaning)																	
Avg	80,476	78,298	8,078	607	462	3,846	2,979	405	3,635	1,728	1,699	127	375	4,117	9,895	157	1,388
Std Dev	15,585	27,043	3,932	643	125	1,028	1,084	261	1,223	742	623	59	179	1,941	3,724	79	449
Median	80,004	72,670	6,637	424	456	3,820	2,828	301	3,277	1,513	1,515	122	337	3,443	8,869	115	1,321
Min	44,787	45,928	4,947	185	224	1,847	1,568	143	1,933	912	858	52	164	1,617	3,649	55	685
Max	104,560	162,313	20,738	2,898	706	5,809	6,149	936	6,542	3,714	2,911	243	826	9,552	19,100	350	2,299
Number of quotes (after cleaning)																	
Avg	628,611,886	279,290,504	64,487,534	13,617,224	7,987,132	32,616,857	25,518,930	3,958,422	54,090,705	21,348,654	20,543,501	508,302	7,325,981	38,322,047	30,050,269	2,532,815	9,503,469
Std Dev	125,112,503	72,431,205	14,818,791	3,903,640	2,323,915	7,434,977	4,757,349	1,572,582	11,359,835	7,476,082	5,899,002	136,216	2,196,516	7,838,284	5,886,635	699,342	2,421,343
Median	622,338,806	286,858,000	61,416,002	13,364,134	7,488,202	29,583,046	24,582,347	4,098,120	51,518,109	22,057,529	20,700,550	465,697	7,268,446	366,82,359	28,184,448	2,534,352	9,194,951
Min	368,646,694	147,174,936	40,321,726	8,347,573	3,747,030	22,156,109	19,131,646	1,293,294	37,783,596	8,222,172	10,856,256	318,438	3,309,087	22,522,639	19,449,233	1,336,222	4,186,745
Max	872,278,526	389,048,310	97,324,007	20,947,630	11,382,152	46,440,632	34,812,831	7,343,634	77,700,703	34,932,632	31,134,228	838,467	11,548,186	51,724,816	43,635,724	4,508,035	14,354,507
Contracts' information (daily averages)																	
$\leq 7d$	196	159	146	86	58	98	93	88	111	80	81	6	59	91	60	13	56
%	6	12	9	11	10	13	13	12	9	12	11	4	10	12	12	5	10
$7d$-$1m$	546	350	452	295	182	279	262	281	322	256	273	17	192	266	179	31	165
%	17	27	27	39	32	37	37	37	25	38	37	9	32	36	35	11	29
$1m$-$3m$	796	302	499	210	161	201	189	180	402	186	190	44	171	185	141	68	149
%	24	24	29	28	28	27	27	23	31	28	26	23	28	25	28	25	26
$3m$-$1y$	1,044	326	395	111	95	94	86	109	355	72	111	64	104	120	99	120	123
%	32	25	23	15	16	13	12	15	27	11	15	33	17	17	19	44	22
$\geq 1y$	713	154	210	50	79	80	78	95	116	73	73	61	80	68	34	44	68
%	22	12	13	7	14	11	11	13	9	11	10	32	13	9	7	15	12
Trades' direction																	
BI	713,701	717,628	67,618	4,262	3,569	30,588	24,988	2,745	29,279	13,894	13,452	907	2,731	33,972	84,673	816	11,127
SI	737,074	710,349	68,155	4,159	3,019	34,012	25,081	2,581	28,195	13,544	13,499	790	2,436	36,107	85,566	1,007	10,895
UI	1,789	2,015	387	66	29	254	130	43	163	103	176	27	77	221	1,057	13	220

1st month data are used. "Trades-only" contracts are not considered for the calculation.

Table 2: Liquidity measures' summary statistics

	Underlying stocks																
	SPY	**AAPL**	**GOOG**	**EA**	**XLK**	**XOM**	**CVX**	**MPC**	**XLE**	**PG**	**WMT**	**K**	**XLP**	**JPM**	**BAC**	**BLK**	**XLF**
Effective spread																	
Avg	0.053	0.047	0.058	0.104	0.085	0.123	0.169	0.232	0.086	0.108	0.118	0.090	0.098	0.083	0.065	0.124	0.088
Std Dev	0.144	0.111	0.096	0.131	0.089	0.205	0.396	0.376	0.115	0.207	0.223	0.073	0.138	0.126	0.157	0.192	0.159
Median	0.019	0.012	0.030	0.061	0.066	0.060	0.070	0.110	0.046	0.051	0.050	0.065	0.050	0.045	0.015	0.077	0.046
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000
Max	4.725	1.230	1.468	1.014	0.663	2.035	6.477	3.116	0.971	2.197	2.055	0.323	1.079	1.381	2.197	1.435	2.144
Realized spread																	
Avg	0.005	0.009	0.010	0.031	0.004	0.029	0.009	0.007	0.015	-0.002	0.016	0.037	-0.010	0.005	0.025	-0.016	0.017
Std Dev	0.065	0.071	0.129	0.196	0.091	0.151	0.142	0.152	0.140	0.135	0.180	0.064	0.096	0.094	0.089	0.080	0.117
Median	0.001	0.003	0.002	0.000	0.000	0.010	0.000	0.007	0.004	0.000	0.000	0.037	-0.008	0.000	0.003	0.000	0.000
Min	-0.489	-0.232	-1.386	-0.450	-0.258	-0.760	-0.811	-0.615	-0.696	-0.673	-0.825	-0.047	-0.198	-0.470	-0.100	-0.359	-0.191
Max	0.801	1.371	1.486	1.833	0.453	1.854	0.811	0.538	1.022	0.811	1.833	0.206	0.322	0.437	0.606	0.080	0.971
5-Minute Price Impact																	
Avg	0.018	0.014	0.028	0.053	0.020	0.035	0.057	0.057	0.038	0.053	0.055	0.036	0.052	0.033	0.018	0.055	0.028
Std Dev	0.055	0.045	0.062	0.136	0.054	0.076	0.137	0.139	0.095	0.208	0.151	0.051	0.162	0.087	0.057	0.129	0.056
Median	0.010	0.006	0.015	0.025	0.018	0.021	0.026	0.027	0.017	0.020	0.020	0.027	0.010	0.020	0.006	0.016	0.012
Min	-0.815	-0.341	-0.211	-0.186	-0.308	-0.513	-0.613	-0.285	-0.212	-0.325	-0.309	-0.084	-0.102	-0.334	-0.236	-0.085	-0.170
Max	0.840	0.575	0.816	1.751	0.209	0.484	1.608	1.649	1.085	2.778	1.622	0.174	1.212	0.973	0.575	0.713	0.358
λ Price Impact ($\lambda \times 10^6$)																	
Avg	353	405	1388	NA	425	1085	1571	NA	278	-67	839	NA	165	537	218	NA	708
Std Dev	823	2371	2409	NA	0	1280	1643	NA	142	1480	1238	NA	457	1325	375	NA	1021
Median	249	209	1341	NA	425	1104	1173	NA	280	520	337	NA	165	144	156	NA	141
Min	-2202	-15823	-6385	NA	425	-519	105	NA	75	-2576	-464	NA	-291	-112	-114	NA	73
Max	6878	10810	5388	NA	425	2558	5822	NA	475	1266	3777	NA	622	4492	1834	NA	2474
Statistics on eligible contracts																	
Avg	1111	799	386	72	52	230	195	47	268	152	146	14	36	226	242	25	94
Std Dev	80	66	83	42	12	38	32	20	49	37	27	4	8	43	29	8	16
Median	1103	825	385	62	50	226	197	40	264	155	149	13	33	221	244	24	92
Min	993	672	267	32	34	174	149	22	164	96	106	10	27	160	176	14	63
Max	1259	895	580	205	80	314	264	90	378	221	204	22	50	327	292	39	121

1st month data are used. "Trades-only" contracts are not considered for the calculation.

2 Volatility and risk-neutral distribution analysis

2.1 Definitions

In this section, we estimate risk-neutral densities from the OPRA option data. Risk-neutral density estimation is a relevant topic of financial data analysis, as it provides insights about the future risk-neutral pricing measures. It also makes analysis of investor attitudes toward market risk possible through the study of pricing kernels, see Aït-Sahalia and Lo (2000).

Following Breeden and Litzenberger (1996), the risk-neutral density can be defined as the compounded second partial derivative of the option price C with respect to strike K with the terminal condition $K = S_T$:

$$f_Q(S_T|S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}, \sigma_{t,\tau}) \stackrel{\text{def}}{=} e^{r_{t,\tau}} \frac{\partial^2 C(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}, \sigma_{t,\tau})}{\partial K^2}\bigg|_{K=S_T}, \tag{7}$$

where the implied volatility $\sigma_{t,\tau} = \sigma(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau})$ is a non-parametric function of the state variables S_t (price of the underlying security at time t), K, time to maturity $\tau = T - t$, the risk-free interest rate $r_{t,\tau}$, the asset yield $\delta_{t,\tau}$ at time t for time to maturity τ, respectively.

Following Aït-Sahalia and Lo (1998), we reduce the number of state variables by assuming that the option price and the risk-neutral-density depend on $S_t, r_{t,\tau}, \delta_{t,\tau}$ jointly through the forward price $F_{t,\tau} = S_t \exp\{(r_{t,\tau} - \delta_{t,\tau})\tau\}$. Also assume that the option function C is homogeneous of degree 1 in $F_{t,\tau}$ and K, then it follows:

$$\begin{aligned} C(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}, \sigma_t) &= F_{t,\tau} \cdot C\left(F_{t,\tau} \cdot \frac{1}{F_{t,\tau}}, \frac{K}{F_{t,\tau}}, \tau, \sigma_{t,\tau}\right) \\ &= F_{t,\tau} \cdot \overline{C}(m_{t,\tau}, \tau, \sigma_{t,\tau}), \end{aligned} \tag{8}$$

where $m_{t,\tau} \stackrel{\text{def}}{=} K/F_{t,\tau}$ is the moneyness of the option.

From (8), it follows that

$$\frac{\partial^2 C(F_{t,\tau}, K, \tau, \sigma_{t,\tau})}{\partial K^2}\bigg|_{K=S_T} = \frac{\partial^2 \overline{C}(m, \tau, \sigma_{t,\tau})}{\partial m^2}\bigg|_{m=\frac{S_T}{S_t}}, \tag{9}$$

and the final expression for risk-neutral density takes the form:

$$f_Q(S_T|S_t, K, \tau, r_{t,\tau}, \sigma_t) \stackrel{\text{def}}{=} e^{r_{t,\tau}} \frac{\partial^2 \overline{C}(m, \tau, \sigma_{t,\tau})}{\partial m^2}\bigg|_{m=\frac{S_T}{S_t}}. \tag{10}$$

2.2 Estimation

2.2.1 Local linear regression

We need to estimate the second derivative of $\overline{C}$ in (10). For this purpose, we use local polynomial regression and numeric finite difference methods.

In Aït-Sahalia and Lo (1998), a local constant estimator, also called Nadaraya-Watson estimator, was used to infer state-price densities. We use the local constant version of the estimator because of its convenient properties: first, as noted by Li and Racine (2011), Song and Xiu (2016), it has better bias and variance performance; second, it provides estimates of derivatives along with predictions of dependent variables; third, it yields tighter asymptotic confidence bands.

Local (multivariate) linear regression is a special case of local polynomial regression. Let us define the regressors $X_i = (m_i, \tau_i, \widehat{\sigma}_i)^\top$: forward moneyness, as defined above, time to maturity and implied volatility (see below). This regresison model fits a first-degree polynomial $M \stackrel{\text{def}}{=} (M(X_1), \dots, M(X_n))^\top$

$$M \approx M(x)1_n + L(x) \tag{11}$$

in the neighborhood of the point x, see Härdle et al. (2004). Here 1_n is an n-dimensional vector of ones and

$$L(x) \stackrel{\text{def}}{=} \begin{pmatrix} (X_1 - x)^\top \nabla_M(x) \\ (X_2 - x)^\top \nabla_M(x) \\ \vdots \\ (X_n - x)^\top \nabla_M(x) \end{pmatrix}, \tag{12}$$

where $\nabla_M(x)$ is the gradient of M estimated at x.

Then for each $x = (m, \tau, \widehat{\sigma})^\top$ the following optimization problem is solved:

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \left\{ \overline{C}_i - \alpha - \beta^\top (X_i - x) \right\}^2 \mathrm{K}_H(X_i - x), \tag{13}$$

where $H \stackrel{\text{def}}{=} (h_m, h_\tau, h_{\widehat{\sigma}})^\top$, K_H is the trivariate kernel:

$$\mathrm{K}_H(X_i - x) \stackrel{\text{def}}{=} K_{h_m}(m_i - m) K_{h_\tau}(\tau_i - \tau) K_{h_{\widehat{\sigma}}}(\widehat{\sigma}_i - \widehat{\sigma}), \tag{14}$$

with $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(x)$ a kernel function. We take $K(x) \stackrel{\text{def}}{=} (2\pi)^{-1/2}\exp(-0.5x^2)$ to be the Gaussian kernel.

Implied volatility $\widehat{\sigma}$ is obtained as the solution to the equation:

$$C^{BS}(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}, \widehat{\sigma}) - \tilde{C}^{Mkt}_{t,\tau} = 0, \tag{15}$$

where $\tilde{C}^{Mkt}_{t,\tau}$ is the market price of the option at time t with time to maturity τ, C^{BS} is the Black-Scholes model price given the option variables.

For the purposes of risk-neutral density estimation, we obtain the implied volatility surface (IVS) by employing the bivariate Nadaraya-Watson smoother:

$$\widehat{\sigma}(\tau, m) = \frac{\sum_{i=1}^{n} K_{h_\tau}(\tau_i - \tau) K_{h_m}(m_i - m)\widehat{\sigma}_i}{\sum_{i=1}^{n} K_{h_\tau}(\tau_i - \tau) K_{h_m}(m_i - m)} \tag{16}$$

An example of the IVS estimated on 20150105 for call options on GOOG, is shown in Figure 5. A certain "dip" in volatility is experienced for at-the-money options for longer times-to-maturity.

Comparing the Taylor expansion (11) with the regression problem (23), it is readily seen that $\widehat{\overline{C}}(x) = \widehat{\alpha}$, and

$$\widehat{\frac{\partial \overline{C}}{\partial m}} = \widehat{\beta}^{(1)}, \tag{17}$$

from where it follows that we can estimate the option price and derivatives by solving (23). Conveniently enough, the solution to (23) can be written in closed form:

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \left(X^\top W X\right)^{-1} X^\top W \overline{C}, \tag{18}$$

where

$$X = \begin{pmatrix} 1 & (X_1 - x)^\top \\ 1 & (X_2 - x)^\top \\ \vdots & \vdots \\ 1 & (X_n - x)^\top \end{pmatrix}, \overline{C} = \begin{pmatrix} \overline{C}_1 \\ \overline{C}_2 \\ \vdots \\ \overline{C}_n \end{pmatrix}, W = \begin{pmatrix} \mathrm{K}_H(X_1 - x) & 0 & \dots & 0 \\ 0 & \mathrm{K}_H(X_2 - x) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathrm{K}_H(X_n - x) \end{pmatrix}, \tag{19}$$

from where it follows that

$$\widehat{\overline{C}}(x) = e_1(\widehat{\alpha}, \widehat{\beta})^\top, \tag{20}$$

where $e_1 = (1,0,0,0)$, and

$$\widehat{\frac{\partial \overline{C}}{\partial m}} = e_4(\widehat{\alpha}, \widehat{\beta})^\top, \tag{21}$$

where $e_4 = (0,0,0,1)$.

To estimate the second partial derivative in (10), we use numeric differentiation. In fact, it holds that

$$\widehat{\frac{\partial^2 \overline{C}}{\partial m^2}} \approx \frac{\widehat{\beta}^{(1)}(m+\varepsilon) - \widehat{\beta}^{(1)}(m-\varepsilon)}{2\varepsilon}, \tag{22}$$

where ε is a small change in moneyness m. An estimation error of order $\mathscr{O}(10^{-6})$ is introduced if we take $\varepsilon = 0.001$.

2.2.2 Bandwidth selection

A key issue for the estimation of α and β in (23) is model selection or the selection of bandwidth vector H.

A standard practice is leave-one-out cross-validation approach which selects an optimal bandwidth as a solution to the following problem:

$$\min_H \frac{1}{n} \sum_{i=1}^{n} \left\{ \overline{C}_i - \widehat{\overline{C}}_{H,-i}(m_i, \tau_i, \widehat{\sigma}_i) \right\}^2 \omega(m_i, \tau_i, \widehat{\sigma}_i), \tag{23}$$

where $\widehat{\overline{C}}_{H,-i}$ is the estimate of $\overline{C}_i$ obtained from the data given a bandwidth vector H and with observation i left out; $\omega(m_i, \tau_i, \widehat{\sigma}_i)$ is a weighting function. More details on the procedure can be found in, e.g., Hastie et al. (2008), Härdle et al. (2004).

2.3 Asymptotic theory

We can analyse the uncertainty in the estimated risk-neutral density by means of confidence bands' construction. Let us denote by $R_T = \log(S_T/S_t)$ the logarithmic market returns of the underlying asset. Then, following Aït-Sahalia and Lo (1998), Song and Xiu (2016), we can adapt the formula for the asymptotic confidence bands for $f_Q(r|m,\tau,\sigma)$ as follows:

$$\sqrt{nh_m h_\tau h_\sigma} h_m \left\{ \widehat{f}_Q(R|m,\tau,\sigma) - f_Q(R|m,\tau,\sigma) \right\} \xrightarrow{\mathscr{L}} \mathbb{N}\left(0, m^2 \cdot c_K \cdot \frac{s^2(m,\tau,\sigma)}{p(m,\tau,\sigma)}\right), \tag{24}$$

where c_K is a constant depending on the choice of the kernel:

$$c_K = \frac{\left\{\int_{-\infty}^{+\infty} K^2(t)dt\right\} \cdot \left\{\int_{-\infty}^{+\infty} \left(t\dot{K}^2(t) + K(t)\right)^2 dt\right\}}{\left(\int_{-\infty}^{+\infty} K(t)t^2 dt\right)^2}. \tag{25}$$

For the Gaussian kernel, it is readily calculated that $c_k = 3/64\pi^2$.

An estimator for s^2 can be constructed using the local linear regression approach in (23) on the squared residuals $\widehat{r}_i \stackrel{\text{def}}{=} (C_i - \widehat{\widehat{C}}_i)^2$ on X_i. The joint density p of X_i is computed using a kernel smoothing Nadaraya-Watson method.

2.4 Empirical results

The results of risk neutral density calculation for options on GOOG on 20150105 are shown in Figures 6 - 9. For each density estimate for a given time-to-maturity, a corresponding implied volatility smile is presented.

It can be observed that the obtained densities have significant mass for negative returns for given time-to-maturities. There is also a pronounced "hump" in the right tail which tends to smooth out fo longer maturities.

As regards uncertainty in densities' estimation, the asymptotic confidence bands are extremely narrow, with the exception of the far right tail. Therefore the amount of confidence in the obtained estimates is rather high. This is a natural improvement over the use of daily data in this situation which tend to provide significantly wider confidence bands.

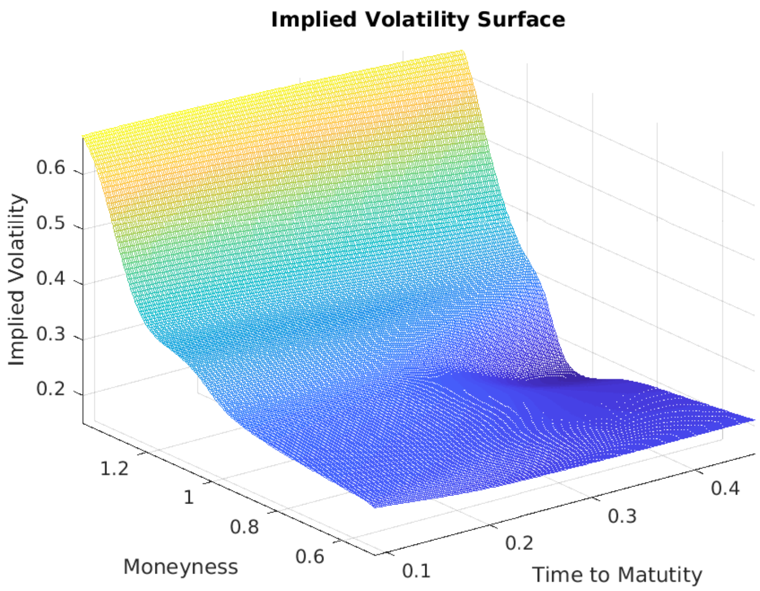

Figure 5: Example of the fitted implied volatility surface on GOOG call options, 20150105

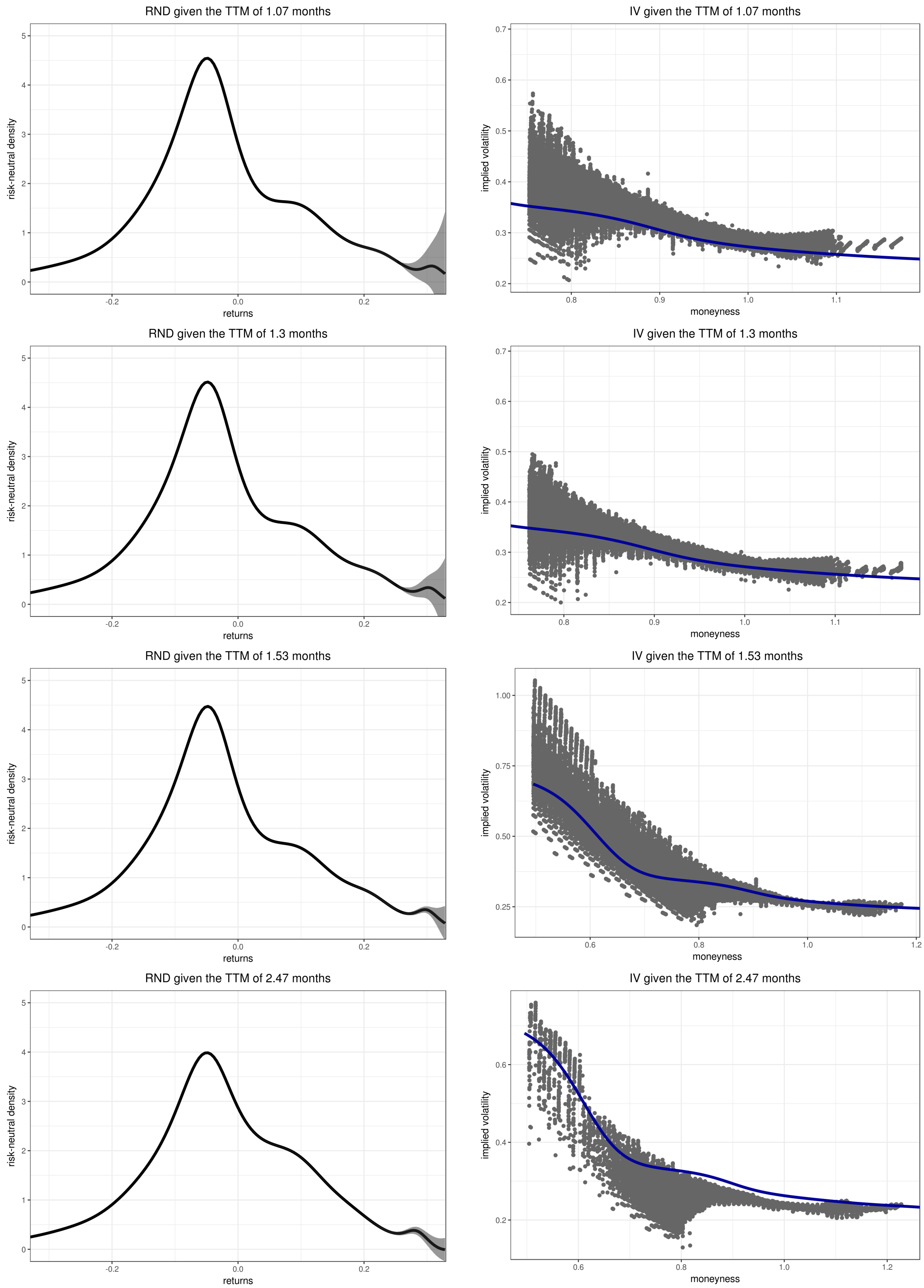

Figure 6: Risk-neutral densities and the corresponding implied volatility smiles obtained from call options on GOOG on 20150105, 1st hour of trading (9:30 - 10:30)

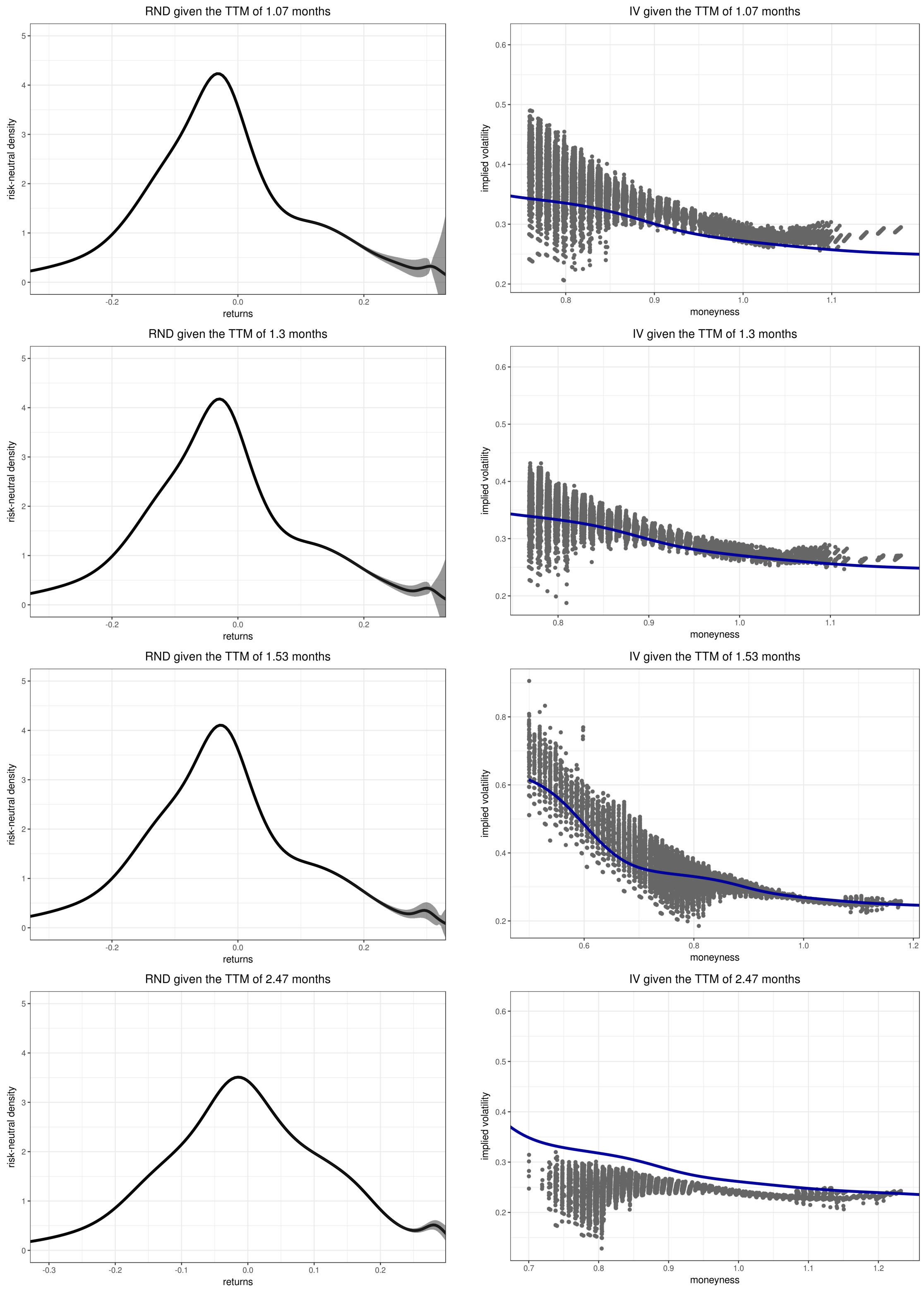

Figure 7: Risk-neutral densities and the corresponding implied volatility smiles obtained from call options on GOOG on 20150105, 3rd hour of trading (11:30 - 12:30)

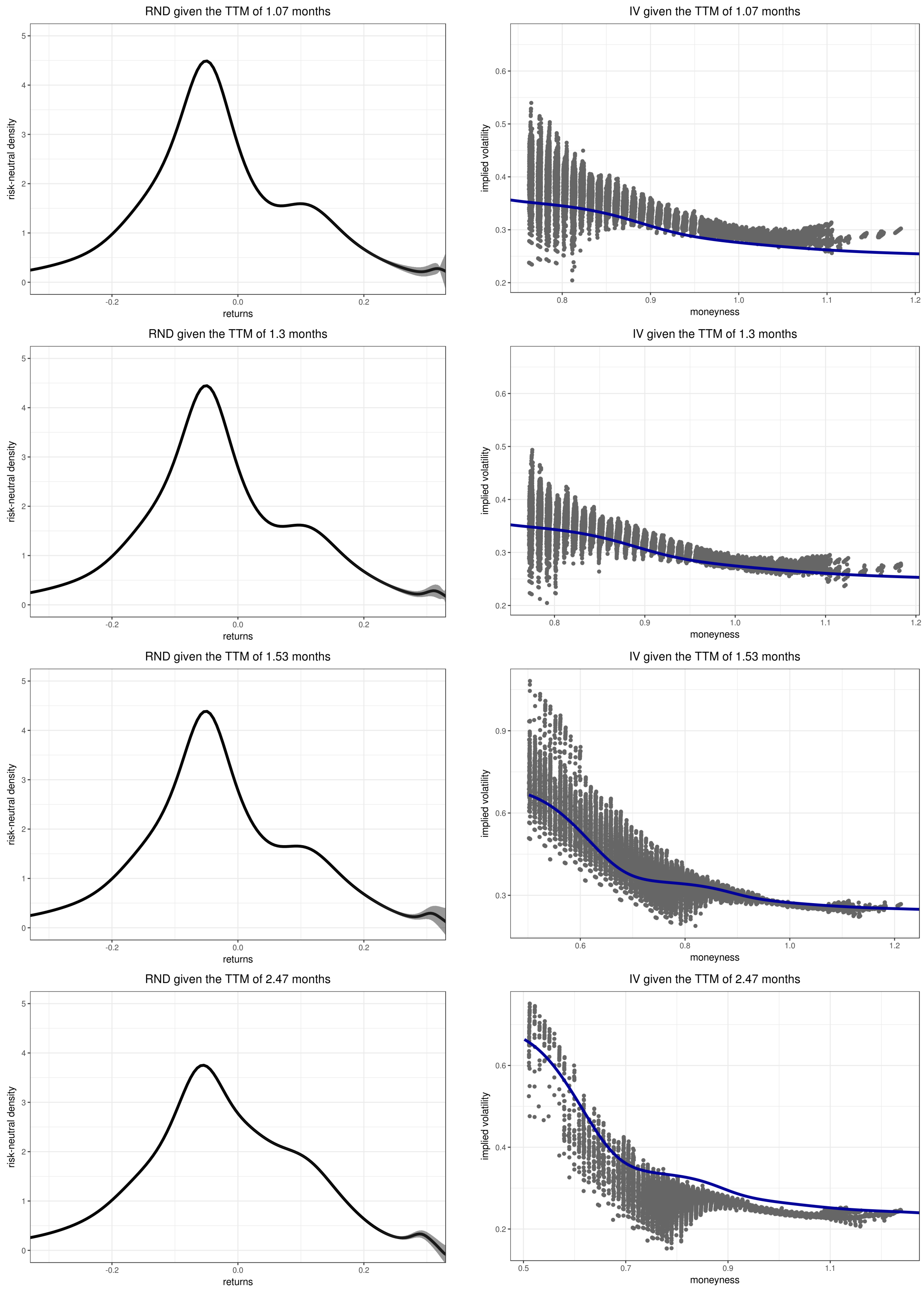

Figure 8: Risk-neutral densities and the corresponding implied volatility smiles obtained from call options on GOOG on 20150105, 5th hour of trading (13:30 - 14:30)

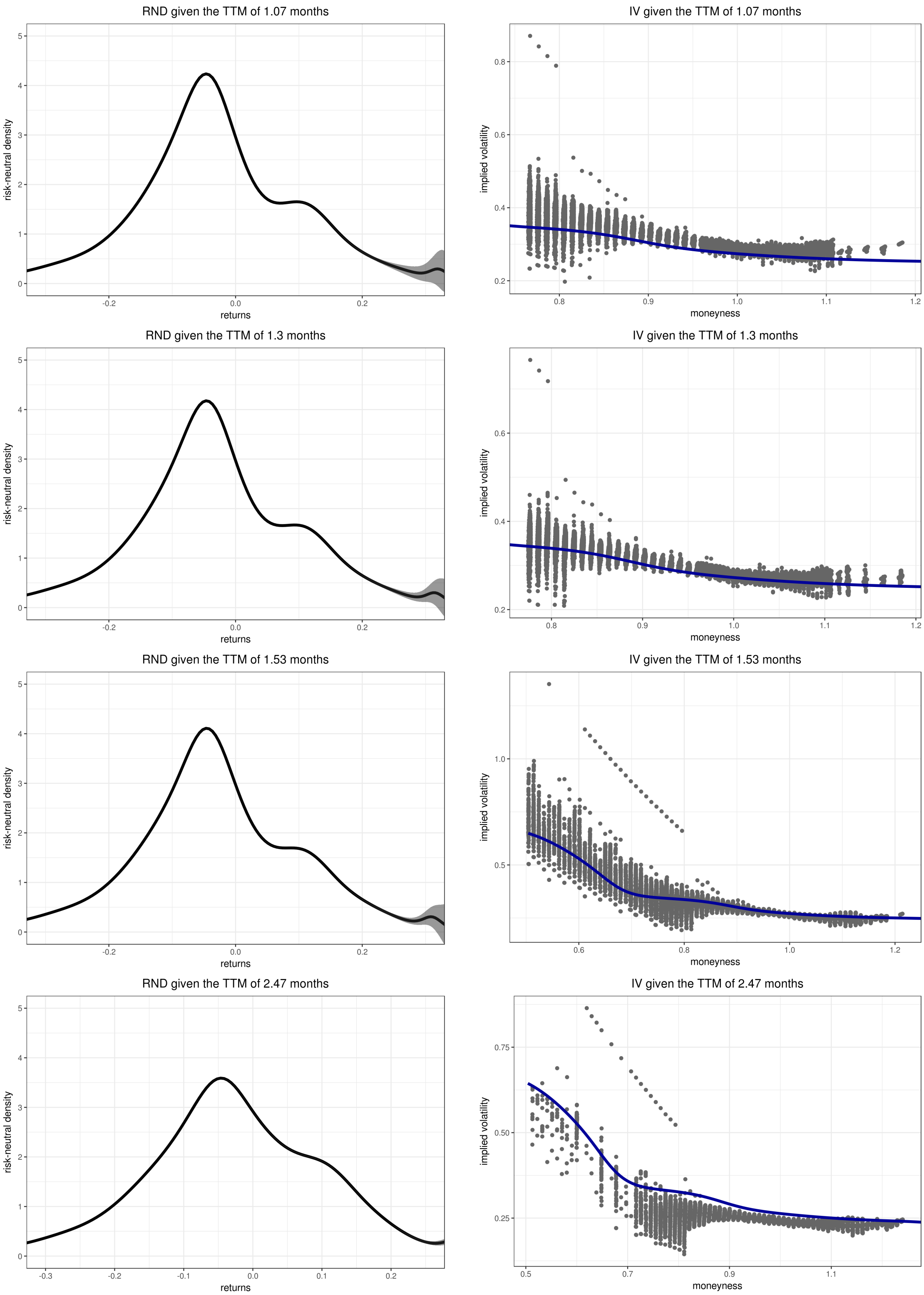

Figure 9: Risk-neutral densities and the corresponding implied volatility smiles obtained from call options on GOOG on 20150105, 7th hour of trading (15:30 - 16:30)

References

Y. Aït-Sahalia and A. Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 3(2):499–547, 1998.

Y. Aït-Sahalia and A. Lo. Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94:9–51, 2000.

D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1996.

R. Y. Goyenko, C. W. Holden, and C. A. Trzcinka. Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92:153–181, 2009.

W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models: An Introduction*. Springer, 2004.

J. Hasbrouck. Trading costs and returns for u.s. equities: Estimating effective costs from daily data. *Journal of Finance*, 64(3):1445–1477, 2009.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2008.

R. Huang and H. Stoll. Dealer versus auction markets: a paired comparison of execution costs on nasdaq and the nyse. *Journal of Financial Economics*, 41:313–357, 1996.

C. M. C. Lee and M. J. Ready. Inferring trade direction from intraday data. *Journal of Finance*, 46(2):733–746, 1991.

Q. Li and J. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2011.

Z. Song and D. Xiu. A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190:176–196, 2016.