A Appendix (not for publication)

A.1 Maximum likelihood estimates of Markov chain transition probabilities

The ergodic distribution $\pi = (\pi_0, \pi_1)$ of the two-state Markov chain satisfies

$$\begin{pmatrix} 1-\pi_{01} & \pi_{10} \\ \pi_{01} & 1-\pi_{10} \end{pmatrix} \begin{pmatrix} \pi_0 \\ \pi_1 \end{pmatrix} = \begin{pmatrix} \pi_0 \\ \pi_1 \end{pmatrix}$$

with $\pi_0 + \pi_1 = 1$ and

$$\pi_0 = \frac{\pi_{10}}{A}, \quad \pi_1 = \frac{\pi_{01}}{A}, \quad A \equiv \pi_{01} + \pi_{10}$$

The part of log-likelihood that depends on the transition probabilities can be written

$$\begin{aligned} \ell(\pi_{jk}) &= \log p(s_1) + \sum_j \sum_k \delta_{jk} \log \pi_{jk} \\ &= \log p(s_1) + n_{00} \log(1-\pi_{01}) + n_{01} \log \pi_{01} + n_{10} \log \pi_{10} + n_{11} \log(1-\pi_{10}) \\ &= n_{00} \log(1-\pi_{01}) + (n_{01} + d_1^+) \log \pi_{01} + (n_{10} + 1 - d_1^+) \log \pi_{10} + n_{11} \log(1-\pi_{10}) - \log(A) \end{aligned}$$

where $n_{jk} = \sum_j \sum_k \delta_{jk}$ is the number of observations with $s_i = j$ followed by $s_{i+1} = k$ and $d_1^+ \equiv 1(s_1 > 0)$. The scores and hessian are

$$\begin{aligned} \frac{\partial \ell}{\partial \pi_{01}} &= -\frac{1}{A} - \frac{n_{00}}{1-\pi_{01}} + \frac{n_{01} + d_1^+}{\pi_{01}} \\ \frac{\partial \ell}{\partial \pi_{10}} &= -\frac{1}{A} + \frac{n_{10} + 1 - d_1^+}{\pi_{10}} - \frac{n_{11}}{1-\pi_{10}} \\ \frac{\partial^2 \ell}{\partial \pi_{01}^2} &= \frac{1}{A^2} - \frac{n_{00}}{(1-\pi_{01})^2} - \frac{n_{01} + d_1^+}{\pi_{01}^2} \\ \frac{\partial^2 \ell}{\partial \pi_{10}^2} &= \frac{1}{A^2} - \frac{n_{10} + 1 - d_1^+}{\pi_{10}^2} - \frac{n_{11}}{(1-\pi_{10})^2} \\ \frac{\partial^2 \ell}{\partial \pi_{01} \partial \pi_{10}} &= \frac{1}{A^2} \end{aligned}$$

To evaluate expected Hessian for Fisher information, use

$$\begin{aligned} & E[d_1^+] = \Pr(s_1 == 1) = \pi_1 \\ & E[n_{jk}] = n \Pr(s_i = j, s_{i+1} = k) = n \Pr(s_{i+1} = k | s_i = j) \Pr(s_i = j) = n \pi_{jk} \pi_j \\ & E[n_{00}] = n \frac{\pi_{10}(1-\pi_{01})}{A}, \quad E[n_{01}] = E[n_{10}] = n \frac{\pi_{10}\pi_{01}}{A}, \quad E[n_{11}] = n \frac{\pi_{01}(1-\pi_{10})}{A} \end{aligned}$$

Then expected Hessian

$$E[\frac{\partial^2 \ell}{\partial \pi_{01}^2}] = -\frac{\pi_{10}(1-\pi_{01}+nA)}{A^2 \pi_{01}(1-\pi_{01})}, \quad E[\frac{\partial^2 \ell}{\partial \pi_{10}^2}] = -\frac{\pi_{01}(1-\pi_{10}+nA)}{A^2 \pi_{10}(1-\pi_{10})}, \quad E[\frac{\partial^2 \ell}{\partial \pi_{01} \partial \pi_{10}}] = \frac{1}{A^2}$$

with inverse Fisher information $\mathcal{I} = -E[H]$

$$\mathcal{I}^{-1} = \frac{(1-\pi_{01})(1-\pi_{10})A}{n(2+(n-1)A)} \begin{pmatrix} \frac{\pi_{01}(1-\pi_{10}+nA)}{\pi_{10}(1-\pi_{10})} & 1 \\ 1 & \frac{\pi_{10}(1-\pi_{01}+nA)}{\pi_{01}(1-\pi_{01})} \end{pmatrix}$$

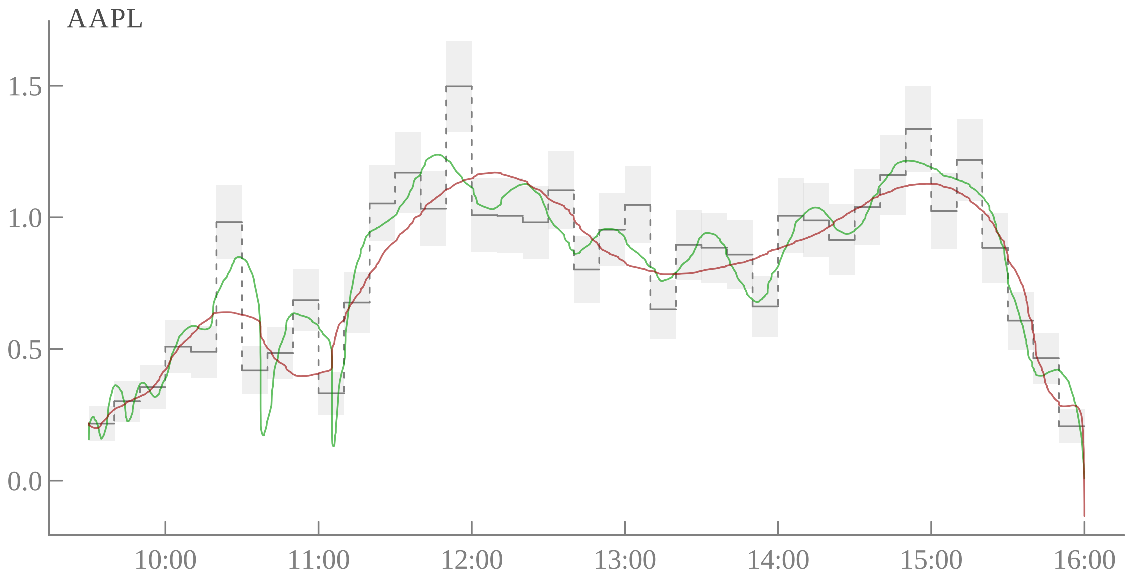
If we condition on the first observation, the log-likelihood and scores are

$$
\begin{aligned}
\ell(s_2,\ldots,s_n|s_1) &= n_{00}\log(1-\pi_{01}) + n_{01}\log\pi_{01} + n_{10}\log\pi_{10} + n_{11}\log(1-\pi_{10}) \\
\frac{\partial\ell}{\partial\pi_{01}} &= -\frac{n_{00}}{1-\pi_{01}} + \frac{n_{01}}{\pi_{01}}, \quad \frac{\partial\ell}{\partial\pi_{10}} = \frac{n_{10}}{\pi_{10}} - \frac{n_{11}}{1-\pi_{10}} \\
\frac{\partial^2\ell}{\partial\pi_{01}^2} &= -\frac{n_{00}}{(1-\pi_{01})^2} - \frac{n_{01}}{\pi_{01}^2}, \quad \frac{\partial^2\ell}{\partial\pi_{10}^2} = -\frac{n_{10}}{\pi_{10}^2} - \frac{n_{11}}{(1-\pi_{10})^2}, \quad \frac{\partial^2\ell}{\partial\pi_{01}\partial\pi_{10}} = 0
\end{aligned}
$$

with MLE $\widehat{\pi}_{01} = n_{01}/(n_{00}+n_{01})$, $\widehat{\pi}_{10} = n_{10}/(n_{10}+n_{11})$ and

$$
\begin{aligned}
E[\frac{\partial^2\ell}{\partial\pi_{01}^2}] &= -\frac{n\pi_{10}}{A\pi_{01}(1-\pi_{01})}, \quad E[\frac{\partial^2\ell}{\partial\pi_{10}^2}] = -\frac{n\pi_{01}}{A\pi_{10}(1-\pi_{10})} \\
\mathrm{Var}(\widehat{\pi}_{01}) &= \frac{A\pi_{01}(1-\pi_{01})}{n\pi_{10}}, \quad \mathrm{Var}(\widehat{\pi}_{10}) = \frac{A\pi_{10}(1-\pi_{10})}{n\pi_{01}}
\end{aligned}
$$

Figure 3: Diurnal pattern of duration between trades. The step function is the fitted values from the regression on ten minute interval indicators between 9:30 and 16:00 on 2017-08-30. Shaded area is the two standard error bands. The two curves are the interpolated values from a smoothing cubic spline with knots every 10 minutes (less smooth) and 30 minutes (more smooth).



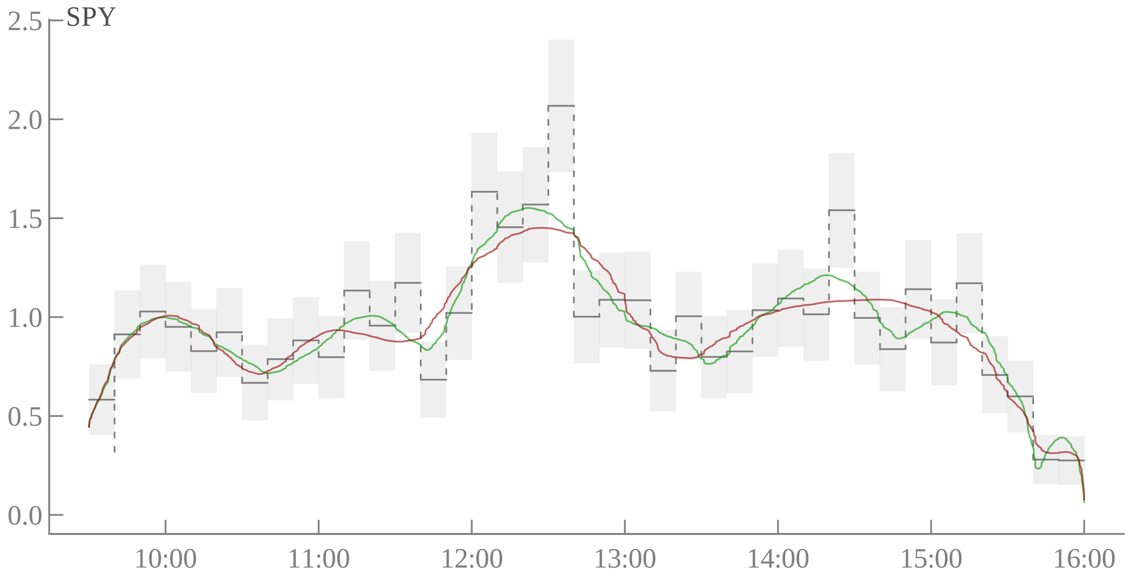


Figure 4: Autocorrelation functions for diurnally adjusted duration between trades for two tickers AAPL and SPY in Nasdaq ITCH. The dashed lines are adjusted by a smoothing cubic spline (with knots every ten minutes) and the solid lines by the regression method. x_i (top two panels) for the level of adjusted durations and x_i^2 (bottom two panels) for squared adjusted durations. The shaded area is the $\pm 2/\sqrt{n}$ bands.

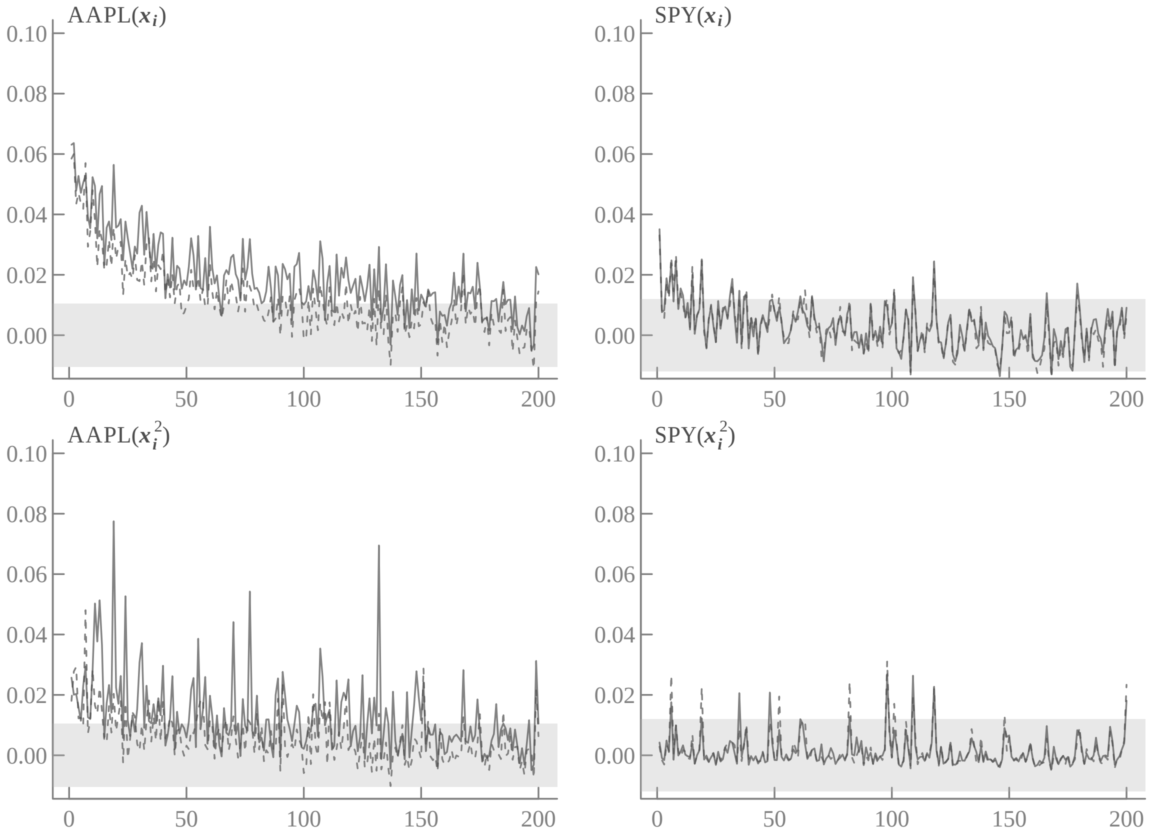


Table 5: Summary of least squares projections of x_i and x_i^2 for AAPL. The diurnally adjusted duration x_i and its square x_i^2 is regressed on a constant and lags of x_i, x_i^2 up to $p = 10$ lags for non-overlapping ten minute intervals between 9:30 and 16:00 on 2017-08-30. The time is the mid-point of the ten minute interval, n is the number of observations in the interval, F is the regression F-statistic with p-values in squared brackets to its right. $E_{i-1}[x_i] = \sum \widehat{x}_i/n$ is the estimated conditional mean and $\sqrt{\mathrm{Var}_{i-1}(x_i)} = \sqrt{\sum \widehat{x}_i^2/n - (\sum \widehat{x}_i/n)^2}$ the estimated conditional standard deviation.

				x_i				x_i^2	
	n	R^2	F		$E_{i-1}[x_i]$	R^2	F		$\sqrt{\mathrm{Var}_{i-1}(x_i)}$
09:35	2,756	0.02	3.25	[0.000]	1.00	0.01	1.74	[0.022]	2.80
09:45	1,983	0.02	2.02	[0.005]	1.00	0.01	0.75	[0.774]	2.51
09:55	1,679	0.02	1.68	[0.029]	1.00	0.02	1.30	[0.165]	2.21
10:05	1,164	0.02	1.32	[0.153]	1.00	0.02	1.02	[0.431]	2.12
10:15	1,221	0.03	1.76	[0.020]	0.99	0.01	0.86	[0.634]	2.44
10:25	591	0.05	1.38	[0.127]	1.00	0.05	1.57	[0.055]	2.10
10:35	1,431	0.09	7.16	[0.000]	0.98	0.03	2.44	[0.000]	3.12
10:45	1,242	0.04	2.58	[0.000]	0.99	0.02	1.49	[0.074]	2.31
10:55	865	0.08	3.66	[0.000]	0.99	0.05	2.27	[0.001]	2.14
11:05	1,805	0.14	14.35	[0.000]	0.99	0.07	6.93	[0.000]	3.51
11:15	874	0.04	1.60	[0.046]	0.99	0.02	0.83	[0.679]	2.27
11:25	561	0.04	1.00	[0.455]	0.98	0.03	0.85	[0.648]	2.14
11:35	502	0.08	2.19	[0.002]	0.98	0.08	2.04	[0.005]	2.31
11:45	572	0.05	1.33	[0.153]	0.99	0.04	1.11	[0.329]	2.06
11:55	388	0.09	1.75	[0.024]	1.00	0.07	1.32	[0.161]	1.95
12:05	587	0.09	2.78	[0.000]	0.99	0.06	1.84	[0.015]	2.38
12:15	589	0.05	1.45	[0.094]	0.99	0.02	0.63	[0.892]	2.30
12:25	602	0.06	1.79	[0.019]	0.96	0.04	1.34	[0.146]	2.23
12:35	534	0.10	2.92	[0.000]	1.01	0.06	1.66	[0.036]	2.59
12:45	738	0.05	1.80	[0.018]	1.01	0.04	1.56	[0.055]	2.51
12:55	620	0.05	1.44	[0.096]	1.01	0.05	1.48	[0.083]	2.35
13:05	552	0.05	1.28	[0.189]	1.00	0.05	1.41	[0.111]	2.18
13:15	927	0.07	3.18	[0.000]	0.95	0.04	1.72	[0.025]	2.69
13:25	661	0.02	0.60	[0.917]	0.99	0.02	0.75	[0.777]	2.67
13:35	668	0.04	1.48	[0.081]	1.01	0.03	0.99	[0.478]	2.96
13:45	687	0.06	2.27	[0.001]	1.00	0.06	2.15	[0.003]	2.55
13:55	901	0.04	1.82	[0.015]	0.97	0.03	1.23	[0.219]	2.86
14:05	586	0.05	1.51	[0.071]	1.00	0.04	1.30	[0.171]	2.20
14:15	595	0.05	1.57	[0.056]	1.02	0.05	1.44	[0.097]	2.20
14:25	647	0.03	1.04	[0.417]	0.97	0.03	0.92	[0.560]	2.44
14:35	567	0.05	1.37	[0.133]	1.01	0.02	0.51	[0.963]	2.72
14:45	509	0.04	0.96	[0.515]	0.98	0.04	0.99	[0.476]	2.14
14:55	438	0.08	1.80	[0.019]	0.98	0.06	1.39	[0.123]	2.13
15:05	573	0.07	2.20	[0.002]	1.00	0.07	1.93	[0.009]	2.38
15:15	475	0.03	0.76	[0.760]	0.96	0.03	0.73	[0.792]	2.17
15:25	682	0.05	1.76	[0.022]	0.98	0.08	2.89	[0.000]	2.76
15:35	980	0.08	4.07	[0.000]	1.00	0.06	3.21	[0.000]	2.64
15:45	1,275	0.03	1.82	[0.015]	0.99	0.03	1.77	[0.019]	2.77
15:55	2,914	0.03	5.10	[0.000]	1.00	0.02	3.36	[0.000]	3.51

Table 6: Summary of least squares projections of x_i and x_i^2 for SPY. The diurnally adjusted duration x_i and its square x_i^2 is regressed on a constant and lags of x_i, x_i^2 up to $p = 10$ lags for non-overlapping ten minute intervals between 9:30 and 16:00 on 2017-08-30. The time is the mid-point of the ten minute interval, n is the number of observations in the interval, F is the regression F-statistic with p-values in squared brackets to its right. $E_{i-1}[x_i] = \sum \widehat{x}_i/n$ is the estimated conditional mean and $\sqrt{\mathrm{Var}_{i-1}(x_i)} = \sqrt{\sum \widehat{x}_i^2/n - (\sum \widehat{x}_i/n)^2}$ the estimated conditional standard deviation.

		x_i				x_i^2			
	n	R^2	F		$E_{i-1}[x_i]$	R^2	F		$\sqrt{\mathrm{Var}_{i-1}(x_i)}$
09:35	1,018	0.04	1.88	[0.011]	1.01	0.04	1.88	[0.011]	2.13
09:45	648	0.03	0.84	[0.661]	1.01	0.01	0.27	[1.000]	2.23
09:55	574	0.02	0.58	[0.929]	1.02	0.01	0.30	[0.999]	2.38
10:05	617	0.05	1.68	[0.032]	0.99	0.04	1.18	[0.269]	2.61
10:15	718	0.05	2.02	[0.005]	1.01	0.03	1.10	[0.344]	3.35
10:25	639	0.02	0.74	[0.790]	0.99	0.02	0.58	[0.930]	3.51
10:35	888	0.01	0.55	[0.942]	1.01	0.01	0.41	[0.991]	3.13
10:45	752	0.02	0.68	[0.851]	0.99	0.01	0.49	[0.972]	2.67
10:55	673	0.02	0.59	[0.920]	0.99	0.02	0.60	[0.916]	2.58
11:05	742	0.02	0.61	[0.909]	1.01	0.01	0.31	[0.998]	2.84
11:15	514	0.04	1.04	[0.415]	1.02	0.03	0.80	[0.711]	2.98
11:25	622	0.04	1.41	[0.111]	0.98	0.05	1.74	[0.023]	2.73
11:35	502	0.08	2.03	[0.005]	1.01	0.07	1.83	[0.016]	2.72
11:45	869	0.03	1.27	[0.189]	1.00	0.01	0.62	[0.899]	3.31
11:55	573	0.02	0.71	[0.821]	1.01	0.02	0.55	[0.943]	2.80
12:05	357	0.06	1.03	[0.424]	1.01	0.07	1.25	[0.208]	2.56
12:15	402	0.09	1.91	[0.011]	1.00	0.08	1.57	[0.057]	2.91
12:25	375	0.05	0.96	[0.508]	1.01	0.05	0.85	[0.657]	2.94
12:35	281	0.14	2.08	[0.005]	0.97	0.12	1.74	[0.028]	3.16
12:45	578	0.05	1.42	[0.107]	1.02	0.03	0.94	[0.533]	3.70
12:55	552	0.08	2.39	[0.001]	0.98	0.12	3.52	[0.000]	5.25
13:05	531	0.07	1.85	[0.014]	1.01	0.06	1.55	[0.060]	3.92
13:15	755	0.03	1.17	[0.277]	0.98	0.07	2.89	[0.000]	3.64
13:25	642	0.04	1.18	[0.266]	0.91	0.03	0.81	[0.708]	3.38
13:35	738	0.05	1.98	[0.007]	0.98	0.04	1.66	[0.035]	3.66
13:45	718	0.07	2.68	[0.000]	1.00	0.08	3.02	[0.000]	3.43
13:55	571	0.03	0.76	[0.762]	1.00	0.02	0.50	[0.967]	3.13
14:05	532	0.05	1.35	[0.144]	0.97	0.06	1.51	[0.073]	2.86
14:15	588	0.04	1.33	[0.153]	1.00	0.04	1.25	[0.206]	2.77
14:25	378	0.07	1.36	[0.138]	0.98	0.05	0.87	[0.620]	3.02
14:35	586	0.04	1.26	[0.203]	0.98	0.03	0.99	[0.477]	3.51
14:45	714	0.03	1.11	[0.331]	0.99	0.03	1.22	[0.234]	2.70
14:55	517	0.06	1.58	[0.052]	1.01	0.03	0.73	[0.796]	3.14
15:05	679	0.08	2.87	[0.000]	1.00	0.11	4.08	[0.000]	3.38
15:15	499	0.06	1.51	[0.074]	0.98	0.05	1.22	[0.229]	3.02
15:25	846	0.03	1.45	[0.091]	0.88	0.02	0.84	[0.666]	3.35
15:35	991	0.03	1.34	[0.142]	1.01	0.02	1.16	[0.278]	3.53
15:45	2,123	0.02	1.94	[0.008]	0.98	0.01	1.35	[0.139]	3.88
15:55	2,185	0.02	1.90	[0.009]	1.00	0.01	0.94	[0.532]	3.82

A.2 Maximum likelihood estimates for cubic spline (knots every 10 minutes) diurnally adjusted series

Table 7: Maximum likelihood parameter estimates for AAPL duration ($n = 36,331$ observations). The (common to all models) estimated transition probabilites are $\widehat{\pi}_{01} = 0.539$ (0.004) and $\widehat{\pi}_{10} = 0.361$ (0.003) with standard errors in parentheses. MEM is the multiplicative error model (1) with unit exponential ($v = 1$), MEMZ is the zero augmented MEM, LNZ is the log duration model (5)–(6) with $\delta = 0$ and constant conditional variance, LNZ-G is LNZ with GARCH(1,1) (7), LNZ-E is LNZ with EGARCH(1,1) (8). QML standard errors in parentheses. ℓ/n is the log-likelihood value per observation and AIC, BIC, HQ are the Akaike, Bayesian (Schwarz), Hannan-Quinn information criterion, respectively.

	MEM	MEMZ		LNZ	LNZ-G		LNZ-E
ω_0	0.012	0.069	α_0	-0.270	-0.237		-0.202
	(0.003)	(0.007)		(0.044)	(0.042)		(0.042)
ω_1	0.082	0.103	α_1	0.692	0.670		0.681
	(0.008)	(0.009)		(0.041)	(0.043)		(0.043)
ω_2	0.915	0.894	α_2	-0.405	-0.381		-0.401
	(0.009)	(0.009)		(0.036)	(0.035)		(0.035)
			α_3	0.851	0.872		0.878
				(0.017)	(0.016)		(0.016)
			β_0	15.358	1.220	γ_0	0.066
				(0.115)	(0.453)		(0.028)
			β_1		0.089	γ_1	0.206
					(0.017)		(0.028)
			β_2		0.310	γ_2	0.080
					(0.179)		(0.006)
			β_3		0.857	γ_3	0.110
					(0.042)		(0.021)
						γ_4	0.920
							(0.016)
ℓ/n	-0.884	-1.528		-0.611	-0.606		-0.603
AIC/n	1.768	3.057		1.223	1.213		1.208
BIC/n	1.769	3.058		1.224	1.215		1.210
HQ/n	1.768	3.057		1.223	1.213		1.208

Table 8: Maximum likelihood parameter estimates for SPY duration ($n = 27,877$ observations). The (common to all models) estimated transition probabilites are $\widehat{\pi}_{01} = 0.384$ (0.004) and $\widehat{\pi}_{10} = 0.449$ (0.004) with standard errors in parentheses. MEM is the multiplicative error model (1) with unit exponential ($v = 1$), MEMZ is the zero augmented MEM, LNZ is the log duration model (5)–(6) with $\delta = 0$ and constant conditional variance, LNZ-G is LNZ with GARCH(1,1) (7), LNZ-E is LNZ with EGARCH(1,1) (8). QML standard errors in parentheses. ℓ/n is the log-likelihood value per observation and AIC, BIC, HQ are the Akaike, Bayesian (Schwarz), Hannan-Quinn information criterion, respectively.

	MEM	MEMZ		LNZ	LNZ-G		LNZ-E
ω_0	0.026	0.183	α_0	−0.142	−0.138		−0.113
	(0.007)	(0.030)		(0.038)	(0.037)		(0.037)
ω_1	0.037	0.136	α_1	0.933	0.868		0.879
	(0.007)	(0.020)		(0.046)	(0.047)		(0.047)
ω_2	0.939	0.855	α_2	−0.434	−0.396		−0.409
	(0.012)	(0.021)		(0.039)	(0.038)		(0.038)
			α_3	0.837	0.852		0.855
				(0.015)	(0.015)		(0.015)
			β_0	13.424	2.183	γ_0	0.167
				(0.163)	(0.375)		(0.051)
			β_1		0.153	γ_1	0.351
					(0.014)		(0.032)
			β_2		1.779	γ_2	0.110
					(0.265)		(0.012)
			β_3		0.696	γ_3	0.250
					(0.035)		(0.031)
						γ_4	0.832
							(0.026)
ℓ/n	−0.959	−1.466		−0.833	−0.828		−0.827
AIC/n	1.918	2.932		1.667	1.657		1.654
BIC/n	1.918	2.933		1.669	1.660		1.658
HQ/n	1.918	2.932		1.667	1.658		1.655