Supplemental Appendix for:

Modeling and Forecasting Persistent Financial Durations

Filip Žikeš Jozef Baruník Nikhil Shenai

April 19, 2014

A1. Derivation of the autocovariance functions and spectral densities

Defining $m_{j,i} = g(M_{j,i})$ for some function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}(g^2(M)) < \infty$, we start by showing that:

$$\mathrm{E}(m_{j,i}m_{j,i-h}) = \mathbb{E}(m_{j,i})^2 + \mathrm{Var}(m_{j,i})(1-\gamma_j)^h. \tag{0.1}$$

for $h \geq 0$. In the binomial MSMD model, the multiplier $M_{j,i}$, if it switches, takes the value of m_0 or $(2-m_0)$ with equal probability. To simplify notation, define $p_j := 1-\frac{1}{2}\gamma_j$, $m_{0,1} := g(m_0)$, $m_{0,2} := g(2-m_0)$, $\boldsymbol{m}_0 := (m_{0,1}, m_{0,2})'$. Then the transition matrix, $\boldsymbol{P}_j$, associated with the j-th multiplier can be written as:

$$\boldsymbol{P}_j = \begin{pmatrix} p_j & 1-p_j \\ 1-p_j & p_j \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2(p_j-1) \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix} = \boldsymbol{C}\boldsymbol{A}_j\boldsymbol{C}',$$

where $\boldsymbol{C}$ is the matrix of eigenvectors of the transition matrix and $\boldsymbol{A}_j$ holds the corresponding eigenvalues. Then by the Law of Iterated Expectations (LIE),

$$\begin{aligned}
\mathrm{E}(m_{j,i}m_{j,i-h}) &= \mathrm{E}(m_{j,i}m_{j,i-h}|m_{j,i-h} = m_{0,1})\mathbb{P}(m_{j,i-h} = m_{0,1}) \\
&\qquad + \mathrm{E}(m_{j,i}m_{j,i-h}|m_{j,i-h} = m_{0,2})\mathbb{P}(m_{j,i-h} = m_{0,2}), \\
&= \frac{1}{2}m_{0,1}\mathrm{E}(m_{j,i}|m_{j,i-h} = m_{0,1}) + \frac{1}{2}m_{0,2}\mathrm{E}(m_{j,i}|m_{j,i-h} = m_{0,2}), \\
&= \frac{1}{2}\boldsymbol{m}_0'\boldsymbol{P}_j^h\boldsymbol{m}_0, \\
&= \frac{1}{2}\boldsymbol{m}_0'\boldsymbol{C}\boldsymbol{A}_j^h\boldsymbol{C}'\boldsymbol{m}_0, \\
&= \frac{1}{4}(m_{0,1}+m_{0,2})^2 + \frac{1}{4}(1-\gamma_j)^h(m_{0,1}-m_{0,2})^2, \\
&= \mathbb{E}(m_{j,i})^2 + \mathrm{Var}(m_{j,i})(1-\gamma_j)^h.
\end{aligned}$$

When the multiplier $M_{j,i}$ is drawn from a continuous distribution upon switching, then the new value it takes is different from the current value with probability one. Then we have:

$$\begin{aligned}
\mathrm{E}(m_{j,i}m_{j,i-h}) &= \mathbb{E}[\mathbb{E}(m_{j,i}m_{j,i-h}|m_{j,i-h}), \\
&= \mathbb{E}[m_{j,i-h}\mathbb{E}(m_{j,i}|m_{j,i} \neq m_{j,i-h})\mathbb{P}(m_{j,i} \neq m_{j,i-h}) \\
&\qquad + \quad m_{j,i-h}\mathbb{E}(m_{j,i}|m_{j,i} = m_{j,i-h})\mathbb{P}(m_{j,i} = m_{j,i-h})] \\
&= \mathbb{E}[m_{j,i-h}(\mathbb{E}(m_{j,i})(1-(1-\gamma_j)^h) + m_{j,i-h}^2(1-\gamma_j)^h)] \\
&= \mathbb{E}(m_{j,i})^2(1-(1-\gamma_j)^h) + \mathbb{E}(m_{j,i}^2)(1-\gamma_j)^h, \\
&= \mathbb{E}(m_{j,i})^2 + \mathrm{Var}(m_{j,i})(1-\gamma_j)^h.
\end{aligned}$$

as claimed.

Now given that the multipliers and ϵ_i are all mutually independent, we obtain by LIE and (0.1) for $h > 0$:

$$\begin{aligned}
\mathrm{Cov}\,(X_i, X_{i-h}) &= \mathrm{Cov}(\psi_i\epsilon_i, \psi_{i-h}\epsilon_{i-h}), \\
&= \mathrm{E}(\psi_i\psi_{i-h})\mathrm{E}(\epsilon_i)\mathrm{E}(\epsilon_{i-h}) - \mathrm{E}(\psi_i)\mathrm{E}(\psi_{i-h})\mathrm{E}(\epsilon_i)\mathrm{E}(\epsilon_{i-h}), \\
&= \bar{\psi}^2\left[\prod_{j=1}^{k}\mathrm{E}(M_{j,i}M_{j,i-h}) - \left(\prod_{j=1}^{k}\mathrm{E}(M_{j,i})\right)^2\right], \\
&= \bar{\psi}^2\left(\prod_{j=1}^{k}[1+\mathrm{Var}(M)(1-\gamma_j)^h] - 1\right),
\end{aligned}$$

and for $h = 0$:

$$\begin{aligned}
\mathrm{Var}\,(X_i) &= \mathrm{E}(\psi_i^2)\mathrm{E}(\epsilon_i^2) - \mathrm{E}(\psi_i)^2\mathrm{E}(\epsilon_i)^2, \\
&= \bar{\psi}^2[\mathrm{E}(M^2)^k\mathrm{E}(\epsilon_i^2) - 1]
\end{aligned}$$

as claimed. Turning to the spectral density, take the discrete Fourier transform of the autocovariance function:

$$\begin{aligned}
&\frac{2\pi}{\bar{\psi}^2}f_X(\omega) \\
&= \sum_{h=-\infty}^{\infty}\frac{1}{\bar{\psi}^2}\mathrm{Cov}(X_i, X_{i-|h|})e^{-\imath\omega h}, \\
&= \mathrm{Var}(X_i) - [(1+\mathrm{Var}(M))^k - 1] + \sum_{h=-\infty}^{\infty}\left(\prod_{j=1}^{k}(1+\mathrm{Var}(M)(1-\gamma_j)^{|h|}) - 1\right)e^{-\imath\omega h},
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}(M^2)\mathrm{Var}(\epsilon_1) + \sum_{h=-\infty}^{\infty} \sum_{\substack{p_1=0 \\ (p_1,...,p_k)\neq(0,...,0)}}^{1} \sum_{p_2=0}^{1} \cdots \sum_{p_k=0}^{1} \left(\mathrm{Var}(M)^{\sum_{j=1}^k p_j} \prod_{j=1}^{k} (1-\gamma_j)^{|h|p_j} \right) e^{-\imath\omega h}, \\
&= \mathrm{E}(M^2)\mathrm{Var}(\epsilon_1) + \sum_{\substack{p_1=0 \\ (p_1,...,p_k)\neq(0,...,0)}}^{1} \sum_{p_2=0}^{1} \cdots \sum_{p_k=0}^{1} \left(\mathrm{Var}(M)^{\sum_{j=1}^k p_j} \right) \sum_{h=-\infty}^{\infty} \left(\prod_{j=1}^{k} (1-\gamma_j)^{p_j} \right)^{|h|} e^{-\imath\omega h} \\
&= \mathrm{E}(M^2)\mathrm{Var}(\epsilon_1) \qquad\qquad (0.2) \\
&+ \sum_{\substack{p_1=0 \\ (p_1,...,p_k)\neq(0,...,0)}}^{1} \sum_{p_2=0}^{1} \cdots \sum_{p_k=0}^{1} \left(\frac{\mathrm{Var}(M)^{\sum_{j=1}^k p_j} \left[1 - \left(\prod_{j=1}^k (1-\gamma_j)^{p_j} \right)^2 \right]}{1 + \left(\prod_{j=1}^k (1-\gamma_j)^{p_j} \right)^2 - 2 \left(\prod_{j=1}^k (1-\gamma_j)^{p_j} \right) \cos\omega} \right),
\end{aligned}
$$

where we use the multi-binomial theorem and the well-known fact that for any $\rho \in (-1,1)$,

$$
\sum_{h=-\infty}^{\infty} \rho^{|h|} e^{-\imath\omega h} = \frac{1-\rho^2}{1+\rho^2-2\rho\cos\omega},
$$

since $(p_1,...,p_k) \neq (0,...,0)$ implies that $|\prod_{j=1}^k \delta_j^{p_j}| < 1$.

Turning to the autocovariance function for logarithmic durations, given that the multipliers and ϵ_i are all independent, we obtain by (0.1) for $h \neq 0$:

$$
\mathrm{Cov}(x_i, x_{i-h}) = \sum_{j=1}^{k} \mathrm{Cov}(m_{j,i}, m_{j,i-h}) = \mathrm{Var}(\log M) \sum_{j=1}^{k} (1-\gamma_j)^h,
$$

and for $h = 0$:

$$
\mathrm{Var}(x_i) = \sum_{j=1}^{k} \mathrm{Var}(m_{j,i}) + \mathrm{Var}(\log\epsilon_i) = k\mathrm{Var}(\log M) + \mathrm{Var}(\log\epsilon_i),
$$

as claimed. The spectral density then follows directly by calculating the discrete Fourier transform of the autocovariance function:

$$
\begin{aligned}
2\pi f(\omega) &= \sum_{h=-\infty}^{\infty} \mathrm{Cov}(x_i, x_{i-|h|}) e^{-\imath\omega h}, \\
&= \mathrm{Var}(\log\epsilon_1) + \sum_{h=-\infty}^{\infty} \left(\mathrm{Var}(\log M) \sum_{j=1}^{k} (1-\gamma_j)^{|h|} \right) e^{-\imath\omega h}, \\
&= \mathrm{Var}(\log\epsilon_1) + \mathrm{Var}(\log M) \sum_{j=1}^{k} \frac{1-(1-\gamma_j)^2}{1+(1-\gamma_j)^2 - 2(1-\gamma_j)\cos\omega}.
\end{aligned}
$$

It is interesting to note that the autocovariance function of the logarithmic MSMD process is equivalent to that of a signal-plus-noise process $\{z_i\}$, in which the signal is a sum of k independent AR(1) processes:

$$
\begin{aligned}
z_i &= \sum_{j=1}^{k} y_{j,i} + \eta_i, \\
y_{j,i} &= \rho_j y_{j,i-1} + \eta_{j,i}
\end{aligned}
$$

parametrized by $\rho_j = 1 - \gamma_j, \sigma^2_{\eta_{j,i}} = \sigma^2_m(1 - (1 - \gamma_j)^2)$, $j = 1, .., k$, and $\sigma^2_{\eta_i} = \sigma^2_e$. In view of the seminal work of Granger (1980) on aggregation of short-memory processes of heterogenous persistence, it is hardly surprising to find that as $k \to \infty$ the MSMD process can generate highly persistent logarithmic durations.

A2. Densities of innovations ϵ_i

Imposing a unit mean, the densities corresponding to the exponential and Weibull distributions are:

$$
\begin{aligned}
f_E(\epsilon) &= \exp(-\epsilon), \\
f_W(\epsilon; \kappa) &= \kappa \xi_W^{\kappa} \epsilon^{\kappa - 1} \exp(-\xi_W^{\kappa} \epsilon^{\kappa}), \quad \xi_W = \Gamma(1 + 1/\kappa).
\end{aligned}
$$

For $\kappa = 1$, the Weibull distribution reduces to the exponential distribution with unit mean.

A3. Estimators of the asymptotic variance of the Whittle estimator

The plug-in estimators of $\boldsymbol{M}(\boldsymbol{\theta}_0)$ and $\boldsymbol{V}(\boldsymbol{\theta}_0)$ are given by

$$
\begin{aligned}
\boldsymbol{M}(\hat{\boldsymbol{\theta}}) &= \frac{1}{n} \sum_{i=1}^{n-1} \left[\frac{\partial^2 f(\omega_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \left(\frac{1}{f(\omega_i; \boldsymbol{\theta})} - \frac{I_n(\omega_i)}{f^2(\omega_i; \boldsymbol{\theta})} \right) \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}, \\
\boldsymbol{V}(\hat{\boldsymbol{\theta}}) &= \frac{2}{n} \sum_{i=1}^{n-1} [g(\omega_i; \boldsymbol{\theta}) g(\omega_i; \boldsymbol{\theta})']_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \\
&\quad + \frac{2\pi}{n^2} \sum_{i_1=1}^{n-1} \sum_{i_2=1}^{n-1} \left[\frac{g(\omega_{i_1}; \boldsymbol{\theta})}{f(\omega_{i_1})} \frac{g(\omega_{i_2}; \boldsymbol{\theta})'}{f(\omega_{i_2})} S(-\omega_{i_1}, \omega_{i_2}, -\omega_{i_2}; \boldsymbol{\theta}) \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}.
\end{aligned}
$$

The Newey-West estimator of $\boldsymbol{V}(\boldsymbol{\theta}_0)$ reads

$$
\widehat{\boldsymbol{V}(\theta_0)} = \frac{1}{n} \sum_{i=1}^{n-1} \left[\frac{\partial q_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial q_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}
$$

$$+\frac{1}{n}\sum_{b=1}^{B}\sum_{i=b+1}^{n-1}\left(1-\frac{b}{B+1}\right)\left[\frac{\partial q_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial q_{i-b}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\frac{\partial q_{i-b}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial q_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $q_i(\boldsymbol{\theta}) = \log f(\omega_i;\boldsymbol{\theta}) + \frac{I_n(\omega_i)}{f(\omega_i;\boldsymbol{\theta})}$.

A4. Centering and scaling terms in the specification test in Section 3.4

The centering and scaling terms are given by:

$$\begin{aligned}
C_n(k) &= \frac{1}{n\pi}\sum_{l=1}^{n-1}(1-l/n)k^2(l/p_n)+\frac{1}{2\pi}, \\
D_n(k) &= \frac{2}{\pi^2}\sum_{l=1}^{n-2}(1-l/n)(1-(l+1)/n)k^4(l/p_n).
\end{aligned}$$

A5. Competing duration models

The ACD model

Engle and Russell (1998) suggest that the durations, x_i, obey the following process abbreviated as ACD(p,q):

$$\begin{aligned}
x_i &= \psi_i\varepsilon_i, \\
\psi_i &= \omega+\sum_{j=1}^{q}\alpha_j x_{i-j}+\sum_{l=1}^{p}\beta_l\psi_{i-l}
\end{aligned}$$

where ω, α_i and β_i are parameters to be estimated, ψ_i is the conditional duration, the conditional mean of x_i i.e. $\mathrm{E}_{i-1}(x_i) = \psi_i$, and ε_i is the *iid* duration innovation having a distribution with positive support. Sufficient conditions for positive durations are that $\omega > 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$. Weak stationarity is guaranteed by $\sum_{j=1}^{q}\alpha_j + \sum_{j=1}^{p}\beta_j < 1$. Overall, the model specification is similar to a GARCH model, except that the conditional mean is being modelled as opposed to the conditional volatility. The autocovariance function of the ACD model decays exponentially, thereby not enabling long memory which is signified by hyperbolic decay.

The ACD models can be estimated using maximum likelihood, given the distribution of the disturbance term. Engle & Russell (1998) propose the exponential and Weibull distributions, while Grammig & Maurer (2000) suggest the Burr distribution and Lunde (1999) the generalized gamma distributions. An attractive property of the exponential distribution is that the maximum likelihood estimator has a QMLE interpretation, akin to the MLE of GARCH model under normality. Forecasting in the ACD model proceeds via the ARMA representation (see Engle & Russell, 1998 for details).

The LMSD model

Bauwens & Veredas (2004) propose the the Stochastic Conditional Duration (SCD) model given by:

$$\begin{aligned} x_i &= \varepsilon_i e^{\psi_i}, \\ \psi_i &= \omega + \beta\psi_{i-1} + u_i, \end{aligned}$$

where ε_i and u_i are mutually independent *iid* innovations and ω and β are parameters to be estimated. Unlike the ACD model, no conditions on parameters are required to ensure positive durations. Also, weak stationarity is guaranteed as long as β is less than 1, which is a simpler condition than for the ACD model. Overall, the model specification is similar to a stochastic volatility model.

While the ACD has only one, observable random variable driving the system dynamics, the SCD model has an observable random variable driving the observed duration and a latent random variable, u_i, driving the conditional duration (now e^{ψ_i}) via an AR(1) process. The extra random variable enables a richer dynamics structure: Bauwens & Veredas (2004) point out that the parameters governing dispersion (σ) and persistence (β) are separated under the SCD model, whereas they are the same in the ACD model ($\alpha + \beta$), so enabling the SCD model to fit a greater variety of persistence-dispersion profiles.

As with the ACD model, the SCD model is only capable of generating geometric decay in the autocovariance function. In order to enable long memory, Deo, Hsieh & Hurvich (2006) introduce the Long Memory Stochastic Duration (LMSD) process, in which the logged conditional duration equation is replaced with:

$$\psi_i = \omega + \beta\psi_{i-1} + (1-L)^{-d}u_i$$

Here there is more persistence because the logged conditional duration equation has changed from an AR(1) process to an ARFIMA process.

Estimation of the SCD and LMSD models is less straightforward owing to the unobservable factor. Bauwens & Veredas (2004) advocate employing the Kalman Filter, while Deo et al. (2006) suggest QMLE using the Whittle approximation. We adopt the latter approach here. The Whittle estimator of the parameters is consistent and asymptotically normal. Forecasting the SCD and LMSD models is possible either through calibration of the best linear predictor, as advocated by Deo, Hsieh & Hurvich (2010), or via the

Kalman Filter. While the LMSD process contains an infinite series of coefficients, it is still possible to create a state-space form as observed by Chan & Palma (1998) and we adopt their approach here.

References

Bauwens, L. & Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations, *Journal of Econometrics* **119**(2): 381–412.

Chan, N. & Palma, W. (1998). State space modeling of long-memory processes, *The Annals of Statistics* **26**(2): 719–740.

Deo, R., Hsieh, M. & Hurvich, C. (2006). The persistence of memory: From durations to realized volatility. Technical report.

Deo, R., Hsieh, M. & Hurvich, C. M. (2010). Long memory in inter trade durations, counts and realized volatility of NYSE stocks, *Journal of Statistical Planning and Inference* **140**: 3715–3733.

Engle, R. F. & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced time series data, *Econometrica* **66**(5): 1127–1162.

Grammig, J. & Maurer, K. O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model, *Econometrics Journal* **3**: 16–38.

Granger, C. W. J. (1980). Long memory relationships and aggregation of dynamic models, *Journal of Econometrics* **14**: 227–238.

Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. Discussion Paper, Aalborg University.