Modeling Time-Varying Short-Term Dependence in FX

1 Introduction

Financial time series are well known to exhibit non-stationarity, volatility clustering, and structural change. While return autocorrelation at daily or lower frequencies is typically weak, there is some empirical evidence of significant serial dependence at very short horizons, particularly in high-frequency data.

In foreign exchange (FX) markets, tick-by-tick prices are influenced by market microstructure effects such as bid–ask bounce, order flow imbalance, and asynchronous trading. These effects can induce short-lived price runs and time-varying dependence structures. There is some empirical evidence that short-horizon return autocorrelation is not constant (see Lo and MacKinlay (1990)), and in Hamilton (1989) and Hamilton (1994) there is a general framework for modeling time variation via so-called regime-switching models.

The goal of this project is to model such time-varying short-term dependence in FX tick data using a regime-switching autoregressive model formulated as a Hidden Markov Model (HMM). This model will be compared to simple baseline models and also a fully adaptive online alternative procedure.

2 Data and Pre-Averaging

The FX tick data will be obtained from `https://www.histdata.com/`, which provides freely available historical tick and intraday foreign exchange data (start by looking at EUR/USD and other G7 currencies). Note that the market is very different at the start and end of the day; we may therefore clip each day and only consider data from the main London and New York trading hours.

Let P_τ denote the observed transaction price at tick time τ. The observed price can be decomposed as

$$P_\tau = P_\tau^* + \eta_\tau,$$

where P_τ^* is the latent efficient price and η_τ represents microstructure noise.

To mitigate the impact of noise, we apply pre-averaging as discussed in Jacod et al. (2017). In practice, for our purposes, we consider a simple local averaging scheme over k consecutive ticks. The pre-averaged price is defined as

$$\bar{P}_t = \frac{1}{k} \sum_{j=0}^{k-1} P_{t-j},$$

where t now indexes the last tick in each averaging block. Returns are then constructed as

$$r_t = \bar{P}_t - \bar{P}_{t-1}.$$

While more elaborate weight functions can be used in asymptotic theory, simple local averaging is sufficient here to reduce bid–ask bounce and attenuate noise-induced autocorrelation before inference, see Aït-Sahalia and Jacod (2014).

3 Baseline Autoregressive Model

As a benchmark, we consider a single-regime autoregressive model of order one,

$$r_t = \mu + \phi r_{t-1} + \varepsilon_t, \tag{1}$$

where μ is the unconditional mean return (in general $\mu = 0$ but we can experiment with various online/alternative estimation of μ), ϕ measures linear serial dependence and ε_t is an innovation term with $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$; we can potentially also consider non-Gaussian models. Note that this model assumes constant dependence and volatility over time, and therefore provides a useful reference point for assessing the gains from more flexible specifications.

4 Hidden Markov Model with Regime-Switching AR Dynamics

Let $S_t \in \{1, \ldots, K\}$ denote an unobserved discrete-time Markov chain representing the dependence regime at time t. The transition probabilities are defined as

$$\mathbb{P}(S_t = j \mid S_{t-1} = i) = p_{ij},$$

with $p_{ij} \geq 0$ and $\sum_{j=1}^{K} p_{ij} = 1$ for all i. For general treatments of HMM estimation and inference; see e.g. Rabiner (1989) and Cappé et al. (2005) for a general introduction to HMM.

Conditional on the regime $S_t = k$, returns follow an autoregressive process,

$$r_t = \mu_k + \phi_k r_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_k^2).$$

Here μ_k is the regime-specific mean (we will start by assuming $\mu = 0$), ϕ_k captures the strength of short-term dependence in regime k and σ_k^2 is the regime-specific innovation variance. Regimes with ϕ_k close to zero are interpreted as noise-dominated periods, while regimes with larger positive ϕ_k correspond to short-term momentum or price runs.

5 Modeling Volatility

Volatility plays a central role in financial time series, for example, it is known that volatility is persistent and clustered over time (Engle, 1982). Allowing σ_k^2 to vary across regimes captures abrupt changes in market uncertainty. As an extension, conditional heteroskedasticity can be introduced via a regime-dependent ARCH specification,

$$\sigma_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{t-1}^2,$$

where $\omega_k > 0$ and $\alpha_k \geq 0$. This approach builds on the ARCH framework of Engle (1982) while retaining computational tractability; see also Tsay (2010) for a comprehensive treatment of financial time series models.

6 Online Kernel-Based Dependence Estimation

As a contrast to discrete regime-switching models, we also consider a fully adaptive online estimator of serial dependence. Define (for example) the time-varying autocorrelation estimator

$$\hat{\rho}_t = \frac{\sum_{i<t} K_h(t-i)\, r_i r_{i-1}}{\sum_{i<t} K_h(t-i)\, r_{i-1}^2},$$

where $K_h(\cdot)$ is a kernel function with bandwidth parameter h. The bandwidth controls the trade-off between smoothness and responsiveness. This approach provides continuous adaptation without imposing a finite number of regimes.

References

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Lo, A. W., & MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1–2), 181–211.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2017). Between data cleaning and inference: Pre-averaging and robust estimators of the efficient price. *Journal of Econometrics*, 196(1), 1–22.

Aït-Sahalia, Y., & Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.

Cappé, O., Moulines, E., & Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). Wiley.