High-Frequency Lead-Lag Effects and Cross-Asset Linkages: a Multi-Asset Lagged Adjustment Model

Giuseppe Buccheri[*], Fulvio Corsi[†], Stefano Peluso[‡]

Online Appendix

[*]Buccheri: University of Rome Tor Vergata, giuseppe.buccheri@sns.it (corresponding author)
[†]Corsi: Department of Economics, University of Pisa and City University of London, fulvio.corsi@gmail.com
[‡]Peluso: Università Cattolica del Sacro Cuore, stefano.peluso@unicatt.it

1 EM algorithm

As a first step, we assume there are no missing observations. We will show how to handle missing observations in the next sub-section. We denote by $\mathcal{X}_n = \{Z_0, \ldots, Z_n\}$ the set of latent prices and by $\mathcal{Y}_n = \{Y_1, \ldots, Y_n\}$ the set of observed prices. Also, let us assume that $Z_0 \sim \mathrm{N}(\mu, \Sigma)$. Note that, since the knowledge of Z_{t-1} completely determines the last d components of Z_t, the density function $f(Z_t|Z_{t-1})$ can be written as:

$$f(Z_t|Z_{t-1}) = f(MZ_t|Z_{t-1}) \tag{1}$$

Therefore, denoting by $\log L = \log L(\mathcal{Y}_n, \mathcal{X}_n)$ the complete log-likelihood function, we have:

$$\begin{aligned} \log L = \mathrm{const} - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(Z_0 - \mu)'\Sigma^{-1}(Z_0 - \mu) \\ - \frac{n}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{n}(Z_t - \Phi Z_{t-1})'M'Q^{-1}M(Z_t - \Phi Z_{t-1}) \\ - \frac{n}{2}\log|H| - \frac{1}{2}\sum_{t=1}^{n}(Y_t - MZ_t)'H^{-1}(Y_t - MZ_t) \end{aligned} \tag{2}$$

The EM algorithm provides an iterative method for finding the MLE by successively maximizing the conditional expectation of the complete log-likelihood function. The latter can be computed using the Kalman filter and smoothing recursions.

Let us introduce the following quantities which can be recovered as an output of the Kalman filter and smoothing recursions in Section (1.3):

$$Z_t^s = \mathrm{E}[Z_t|\mathcal{Y}_s] \tag{3}$$

$$P_t^s = \mathrm{Cov}[Z_t|\mathcal{Y}_s] \tag{4}$$

$$P_{t,t-1}^s = \mathrm{Cov}[Z_t, Z_{t-1}|\mathcal{Y}_s] \tag{5}$$

With $s = t$, $s < t$ and $s > t$, the resulting conditional expectation is, respectively, an update filter, a predictive filter and a smoother. The Kalman filter is initialized with diffuse initial conditions, i.e. we set $\mathbb{E}[Z_1|Y_1] = 0$ and $\mathrm{Cov}[Z_1|Y_1] = \kappa\mathbb{I}_d$ with $\kappa \to \infty$. At iteration r, the expectation step in the EM algorithm consists in taking the conditional expectation of the complete log-likelihood given the observations $\mathcal{Y}_n$ and using the estimate of $\Omega = \{F, Q, H\}$ obtained at step $r - 1$:

$$\begin{aligned} \mathrm{E}[\log L|\mathcal{Y}_n, \hat{\Omega}_{r-1}] = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\mathrm{Tr}[\Sigma^{-1}[(Z_0^n - \mu)(Z_0^n - \mu)' + P_0^n] \\ - \frac{n}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}[M'Q^{-1}M(C - B\Phi' - \Phi B' + \Phi A\Phi')] \\ - \frac{n}{2}\log|H| - \frac{1}{2}\mathrm{Tr}[H^{-1}\sum_{t=1}^{n}[(Y_t - MZ_t^n)(Y_t - MZ_t^n)' + MP_t^nM'] \end{aligned} \tag{6}$$

where A, B and C are given by:

$$A = \sum_{t=1}^{n} (P_{t-1}^n + Z_{t-1}^n Z_{t-1}^{n\prime}) \tag{7}$$

$$B = \sum_{t=1}^{n} (P_{t,t-1}^n + Z_t^n Z_{t-1}^{n\prime}) \tag{8}$$

$$C = \sum_{t=1}^{n} (P_t^n + Z_t^n Z_t^{n\prime}). \tag{9}$$

In the maximization step, the function $Q(\Omega|\hat{\Omega}_{r-1}) = \mathrm{E}[\log L|\mathcal{Y}_n, \hat{\Omega}_{r-1}]$ is maximized with respect to Ω. Let us consider the following terms depending on F, Q and H:

$$G_1(F,Q) = -\frac{1}{2}\mathrm{Tr}[M'Q^{-1}M(C - B\Phi' - \Phi B' + \Phi A \Phi')]$$
$$G_2(F,Q) = -\frac{n}{2}\log|Q| + G_1(F,Q)$$
$$G_3(H) = -\frac{n}{2}\log|H| - \frac{1}{2}\mathrm{Tr}[H^{-1}[(Y_t - PZ_t)(Y_t - PZ_t)' + MP_t^n M']$$

We start by solving the first order condition $\nabla_F G_1(F,Q) = 0$. Let us write the matrices A and B in the following form:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{10}$$

where A_{ij} and B_{ij}, $i = 1,2$ are $d \times d$ submatrices of A and B. In Section (1.2) we prove the following:

Proposition 1. *The solution of the matrix equation $\nabla_F G_1(F,Q) = 0$ is:*

$$\hat{F}_r = \Gamma \Theta^{-1} \tag{11}$$

where $\Gamma = B_{11} - B_{12} - A_{11} + A_{12}$ and $\Theta = A_{11} + A_{22} - A_{12} - A_{21}$. The solution of the two matrix equations $\nabla_Q G_2(\hat{F}_r, Q) = 0$, $\nabla_H G_3(H) = 0$ are:

$$\hat{Q}_r = \frac{\hat{\Upsilon}}{n}, \quad \hat{H}_r = \frac{\mathrm{diag}(\Lambda)}{n} \tag{12}$$

where $\hat{\Upsilon} = M(C - B\hat{\Phi}_r' - \hat{\Phi}_r B' + \hat{\Phi}_r A \hat{\Phi}_r')M'$, $\Lambda = \sum_{t=1}^{n}[(Y_t - MZ_t^n)(Y_t - MZ_t^n)' + MP_t^n M']$ and

$$\hat{\Phi}_r = \begin{pmatrix} \mathbb{I}_d + \hat{F}_r & -\hat{F}_r \\ \mathbb{I}_d & 0_d \end{pmatrix} \tag{13}$$

Conditions under which the EM algorithm converges to a local maximum of the incomplete log-likelihood function are studied by Wu (1983). We check convergence by looking at the relative increase of the log-likelihood and stop the algorithm when it is lower than some small threshold ($\mu = 10^{-6}$ in our simulation and empirical study). The log-likelihood can be computed in the prediction error decomposition form:

$$\log L = \text{const} - \frac{1}{2}\sum_{t=1}^{n} \log |F_t| - \frac{1}{2}\sum_{t=1}^{n} v_t' F_t^{-1} v_t \tag{14}$$

where $v_t = Y_t - MZ_t^{t-1}$ is the prediction error and $F_t = MP_t^{t-1}M' + H$.

Once $\hat{F}$, $\hat{Q}$ and $\hat{H}$ have been estimated, the matrix of price adjustment Ψ and the covariance matrix of the efficient log-price process Σ can be computed as:

$$\hat{\Psi} = \mathbb{I}_d - \hat{F}, \quad \hat{\Sigma} = \hat{\Psi}^{-1}\hat{Q}\hat{\Psi}'^{-1} \tag{15}$$

The Kalman filter and smoothing recursions in Section (1.3) provide filtered and smoothed estimates of the lagged price X_t. From these, using Eq. (3), one also obtains as a byproduct filtered and smoothed estimates of the martingale efficient log-price process.

1.1 Missing value modification

The update formulas in the maximization step can be modified to take into account missing values. Let us assume that, at time t, d_1 components in the vector Y_t are observed while the remaining d_2 are not observed. We consider the d_1-dimensional vector $Y_t^{(1)}$ of observed components and the $d_1 \times d$ matrix $M_t^{(1)}$ whose lines are the lines of M corresponding to $Y_t^{(1)}$. Also, we consider the $d_1 \times d_1$ covariance matrix $H_t^{(11)}$ of observed components disturbances. Following Shumway and Stoffer (2015), the Kalman filter and smoothing recursions in Section (1.3) and the prediction error decomposition form of the log-likelihood, Eq. (14) are still valid, provided that one replaces Y_t, M and H with:

$$Y_{(t)} = \begin{pmatrix} Y_t^{(1)} \\ \underline{0} \end{pmatrix}, \quad M_{(t)} = \begin{pmatrix} M_t^{(1)} \\ \underline{0} \end{pmatrix}, \quad H_{(t)} = \begin{pmatrix} H_t^{(11)} & \underline{0} \\ \underline{0} & \mathbb{I}^{(22)} \end{pmatrix} \tag{16}$$

where $\mathbb{I}^{(22)}$ is the $d_2 \times d_2$ identity matrix and $\underline{0}$ generically denotes zero arrays of appropriate dimension. Note that the time dependence in $M_{(t)}$ and $H_{(t)}$ is only due to missing observations, while the matrices M and H are constant over time.

Taking the conditional expectation in Eq. (2) requires some modifications in case of missing observations. The second and the fourth term remain as in Eq. (6), provided that one runs Kalman filter and smoothing recursions as described in (16). The last term changes because one needs to evaluate expectations of Y_t conditioning to the incomplete data $\mathcal{Y}_n^{(1)} = \{Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_n^{(1)}\}$. If H is diagonal, as we are assuming here, Shumway and Stoffer (1982) showed that:

$$\begin{aligned}\mathbb{E}[(Y_t - MZ_t)(Y_t - MZ_t)'|\mathcal{Y}_n^{(1)}] &= (Y_{(t)} - M_{(t)}Z_t^n)(Y_{(t)} - M_{(t)}Z_t^n)' \\ &+ M_{(t)}P_t^n M_{(t)}' + \begin{pmatrix} \underline{0} & \underline{0} \\ \underline{0} & \hat{H}_{22,t,r-1} \end{pmatrix}\end{aligned} \tag{17}$$

where $\hat{H}_{22,r-1}$ is the $d_2 \times d_2$ covariance matrix of unobserved components disturbances at time t obtained using the estimate at step $r-1$ of the matrix H. Therefore, the update equation for H becomes:

$$\hat{H} = \frac{\mathrm{diag}(\Lambda^*)}{n} \tag{18}$$

where

$$\Lambda^* = \sum_{t=1}^{n} D_t \left[(Y_{(t)} - MZ_t^n)(Y_{(t)} - MZ_t^n)' + M_{(t)}P_t^n M_{(t)}' + \begin{pmatrix} \underline{0} & \underline{0} \\ \underline{0} & \hat{H}_{22,t,r-1} \end{pmatrix} \right] D_t', \tag{19}$$

D_t being a permutation matrix that rearranges the components of Y_t in their original order.

1.2 Proof of Proposition 1

We will use the following matrix differentiation rules:

$$\nabla_A \mathrm{tr}(AB) = B' \tag{20}$$

$$\nabla_A \mathrm{tr}(ABA'C) = CAB + C'AB' \tag{21}$$

$$\nabla_A |A| = |A|(A^{-1})' \tag{22}$$

where A, B and C are matrices of appropriate dimensions.

Let us re-write $G_1(F, Q)$ as:

$$G_1(F, Q) = -\frac{1}{2}\mathrm{Tr}[Q^{-1}(MCM' - \tilde{B}\tilde{\Phi}' - \tilde{\Phi}\tilde{B}' + \tilde{\Phi}A\tilde{\Phi}')] \tag{23}$$

where we have defined $\tilde{B} = MB$ and $\tilde{\Phi} = M\Phi$. Let us compute explicitly the terms in $G_1(F, Q)$ depending on F:

$$\begin{aligned}\tilde{B}\tilde{\Phi}' &= B_{11}(\mathbb{I} + F') - B_{12}F' \\ \tilde{\Phi}\tilde{B}' &= (\mathbb{I} + F)B_{11}' - FB_{12}' \\ \tilde{\Phi}A\tilde{\Phi}' &= (\mathbb{I} + F)A_{11}(\mathbb{I} + F') - FA_{21}(\mathbb{I} + F') \\ &- (\mathbb{I} + F)A_{12}F' + FA_{22}F'\end{aligned}$$

Therefore, we need to solve $\nabla_F \overline{G}_1(F) = 0$, where:

$$\overline{G}_1(F) = \text{Tr}[Q^{-1}(-B_{11}(\mathbb{I} + F') + B_{12}F' - (\mathbb{I} + F)B'_{11} + FB'_{12} \\ + (\mathbb{I} + F)A_{11}(\mathbb{I} + F') - FA_{21}(\mathbb{I} + F') - (\mathbb{I} + F)A_{12}F' + FA_{22}F')]$$

This can be done using Eq. (20) and (21). One obtains:

$$\nabla_F \overline{G}_1(F) = Q^{-1}[-2(B_{11} - B_{12} - A_{11} + A_{12}) \\ + 2F(A_{11} + A_{22} - A_{21} - A_{12})] \tag{24}$$

and therefore:

$$\hat{F} = \Gamma\Theta^{-1} \tag{25}$$

We now solve $\nabla_{Q^{-1}} G_2(\hat{F}_r, Q) = 0$. We obtain:

$$\begin{aligned} &\nabla_{Q^{-1}} G_2(\hat{F}_r, Q) = \\ &= \nabla_{Q^{-1}} \left[-\frac{n}{2} \log |Q| - \frac{1}{2} \text{Tr}(Q^{-1}\hat{\Upsilon}) \right] \\ &= \frac{n}{2} Q - \frac{1}{2} \hat{\Upsilon}' \end{aligned} \tag{26}$$

and therefore, since $\hat{\Upsilon}' = \hat{\Upsilon}$:

$$\hat{Q} = \frac{\hat{\Upsilon}}{n} \tag{27}$$

Finally, now solve $\nabla_H G_3(\hat{F}_r, Q) = 0$. Note that, since H is diagonal, we can write:

$$\begin{aligned} &\nabla_H G_3(H) = \\ &= \nabla_H \left[-\frac{n}{2} \log |H| - \frac{1}{2} \text{Tr}(H^{-1}\text{diag}(\Lambda)) \right] \\ &= \frac{n}{2} H - \frac{1}{2} \Lambda \end{aligned} \tag{28}$$

and therefore:

$$\hat{H} = \frac{\text{diag}(\Lambda)}{n} \tag{29}$$

1.3 Kalman filter and smoothing recursions

The set of Kalman filter recursions for the state-space model (9), (10) are given by:

$$Z_t^{t-1} = \Phi Z_{t-1}^{t-1} \tag{30}$$

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q \tag{31}$$

$$K_t = P_t^{t-1} M' (M P_t^{t-1} M' + H)^{-1} \tag{32}$$

$$Z_t^t = Z_t^{t-1} + K_t (Y_t - M Z_t^{t-1}) \tag{33}$$

$$P_t^t = P_t^{t-1} - K_t H P_t^{t-1} \tag{34}$$

for $t = 1, \dots, n$. The set of backward smoothing recursions are given by:

$$J_{t-1} = P_{t-1}^{t-1}\Phi'(P_t^{t-1})^{-1} \tag{35}$$

$$Z_{t-1}^n = Z_{t-1}^{t-1} + J_{t-1}(X_t^n - \Phi Z_{t-1}^{t-1}) \tag{36}$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}' \tag{37}$$

for $t = n, \dots, 1$. The covariance $P_{t,t-1}^n$ in Eq. (8) can be computed using the following backward recursion:

$$P_{t-1,t-2}^n = P_{t-1}^{t-1}J_{t-2}' + J_{t-1}(P_{t,t-1}^n - \Phi P_{t-1}^{t-1})J_{t-2}' \tag{38}$$

where $t = n, \dots 2$ and $P_{n,n-1}^n = (I - K_n M)\Phi P_{n-1}^{n-1}$.

1.4 Computation of lead-lag correlations

In order to compute lead-lag correlations, we first compute the j-th order autocovariance matrix, which is defined as:

$$S_j = \mathbb{E}[\Delta X_t \Delta X_{t-j}'] \tag{39}$$

It can be evaluated from the estimated matrices $\hat{F}$ and $\hat{Q}$ as:

$$\hat{S}_j = \hat{F}\hat{S}_{j-1}, \quad j = 1, 2, \dots \tag{40}$$

where the covariance matrix $S_0 = \mathbb{E}[\Delta X_t \Delta X_t']$ is estimated as:

$$\text{vec}(\hat{S}_0) = (\mathbb{I}_{d^2} - \hat{F} \otimes \hat{F})^{-1}\text{vec}(\hat{Q}) \tag{41}$$

see e.g. Hamilton (1994). Lead-lag correlations are finally obtained by normalizing the autocovariances with the diagonal elements of $\hat{S}_0$.

2 Arbitrage-linked securities

2.1 The MLA with cointegrated dynamics

In this section we discuss the case in which all or some of the assets are linked by non-arbitrage constraints. A paradigmatic example is given by a security traded in different exchanges. Due to non-arbitrage, the prices observed in the these exchanges cannot move "too far" from each others. In his pioneering work, Hasbrouck (1995) proposed a vector error correction model (VECM) approach and introduced the well-known information shares (IS), which quantify the fraction of the total variance of the (unique) efficient price process explained by individual exchanges. Alternative strategies have been proposed over the years by several researchers (including, amongst others, Booth et al. 1999, Chu et al. 1999, deB. Harris et al. 2002, De Jong and Schotman 2010).

The MLA with cointegration restrictions can be employed to determining the contribution of individual exchanges to the price discovery of arbitrage-linked securities. Let us first consider a simple case with two distinct securities. The efficient log-prices of the two securities evolve as a random walk:

$$P_{t+1}^{(i)} = P_t^{(i)} + u_{t+1}^{(i)}, \quad i = 1, 2 \tag{42}$$

where $\text{Var}[u_{t+1}^{(i)}] = q_i$ and $\text{Cov}[u_{t+1}^{(1)}, u_{t+1}^{(2)}] = c$. Let us assume that the first security is traded in $d_1 \geq 1$ different markets and the second security is traded in $d_2 \geq 1$ different markets. We denote by $Y_t^{(1)} \in \mathbb{R}^{d_1}$ the vector of observations of the first asset log-price in the d_1 markets and, similarly, we denote by $Y_t^{(2)} \in \mathbb{R}^{d_2}$ the vector of observations of the second asset log-price in the d_2 markets. We define $d = d_1 + d_2$ and write $Y_t = [Y_t^{(1)\prime}, Y_t^{(2)\prime}]' \in \mathbb{R}^d$ as:

$$Y_t = X_t + \epsilon_t \tag{43}$$

where $\text{Cov}[\epsilon_t] = H$ and X_t is the lagged log-price process. The latter is decomposed as $X_t = [X_t^{(1)\prime}, X_t^{(2)\prime}]'$, where $X_t^{(1)} \in \mathbb{R}^{d_1}$ and $X_t^{(2)} \in \mathbb{R}^{d_2}$ are given by:

$$X_{t+1}^{(i)} = X_t^{(i)} + \Psi^{(i)}(\iota_{d_i} P_{t+1}^{(i)} - X_t^{(i)}), \quad i = 1, 2 \tag{44}$$

with $\Psi^{(i)} \in \mathbb{R}^{d_i \times d_i}$ and $\iota_{d_i} \in \mathbb{R}^{d_i}$ is a vector of ones. The difference with respect to Eq. (3) in the paper is that $X_t^{(i)} \in \mathbb{R}^{d_i}$ pertain the same security and are therefore driven by the same scalar log-price $P_{t+1}^{(i)}$.

It is immediate to see that the log-return process $\Delta X_t^{(i)} = X_t^{(i)} - X_{t-1}^{(i)}$ follows the VAR(1) process:

$$\Delta X_{t+1}^{(i)} = (\mathbb{I}_{d_i} - \Psi^{(i)}) \Delta X_t^{(i)} + \Psi^{(i)} \iota_{d_i} \omega_{t+1}^{(i)}, \quad i = 1, 2 \tag{45}$$

As before, the difference with respect to Eq. (5) in the paper is that $\Delta X_{t+1}^{(i)} \in \mathbb{R}^{d_i}$ are now driven by the same scalar innovation $\omega_t^{(i)}$. We then consider the whole vector of log-returns $\Delta X_t = [\Delta X_t^{(1)\prime}, \Delta X_t^{(2)\prime}]' \in \mathbb{R}^d$, which follows:

$$\Delta X_{t+1} = (\mathbb{I}_d - \Psi) \Delta X_t + \Psi \omega_{t+1} \tag{46}$$

where:

$$\Psi = \begin{pmatrix} \Psi^{(1)} & 0_{d_1 \times d_2} \\ 0_{d_2 \times d_1} & \Psi^{(2)} \end{pmatrix}, \quad \omega_{t+1} = \begin{pmatrix} \iota_{d_1} \omega_{t+1}^{(1)} \\ \iota_{d_2} \omega_{t+1}^{(2)} \end{pmatrix} \tag{47}$$

Note that the covariance matrix $Q = \text{Cov}[\omega_t]$ has rank equal to two.

It is not difficult to generalize the previous equations to the case in which there are k distinct securities, with the i-th efficient log-price observed in $d_i \geq 1$ different markets, $i = 1, \dots, k$ and $\text{Cov}[u_{t+1}^{(i)}, u_{t+1}^{(j)}] = c_{ij}$. We recover the standard MLA considered in the paper when $d_i = 1$ for each i. If $d_i > 1$ for at least one i, some of the log-prices are cointegrated. Defining $d = \sum_{i=1}^{k} d_i$, the vector of log-returns $\Delta X_t = [\Delta X_t^{(1)\prime}, \dots, \Delta X_t^{(k)\prime}]' \in \mathbb{R}^d$ follows:

$$\Delta X_{t+1} = (\mathbb{I}_d - \Psi)\Delta X_t + \Psi \omega_{t+1} \tag{48}$$

where:

$$\Psi = \begin{pmatrix} \Psi^{(1)} & \dots & 0_{n_1 \times n_k} \\ \vdots & \ddots & \vdots \\ 0_{n_k \times n_1} & \dots & \Psi^{(k)} \end{pmatrix}, \quad \omega_{t+1} = \begin{pmatrix} \iota_{n_1} \omega_{t+1}^{(1)} \\ \vdots \\ \iota_{n_k} \omega_{t+1}^{(k)} \end{pmatrix} \tag{49}$$

The rank of the covariance matrix $Q = \text{Cov}[\omega_{t+1}]$ is always equal to k.

Similarly to what we have done in the standard MLA, let us introduce the augmented state vector $Z_t = [X_t', X_{t-1}']' \in \mathbb{R}^{2d}$. We can re-write Eq. (43), (46) as:

$$Y_t = MZ_t + \epsilon_t, \quad \text{Cov}[\epsilon_t] = H \tag{50}$$

$$Z_{t+1} = \Phi Z_t + R\xi_{t+1}, \quad \text{Cov}[\xi_t] = W \tag{51}$$

where:

$$\Phi = \begin{pmatrix} 2\mathbb{I}_d - \Psi & -\mathbb{I}_d + \Psi \\ \mathbb{I}_d & 0_d \end{pmatrix}, \quad R = \begin{pmatrix} \Psi & 0_d \\ 0_d & 0_d \end{pmatrix}, \quad W = \begin{pmatrix} Q & 0_d \\ 0_d & 0_d \end{pmatrix} \tag{52}$$

and $M = [\mathbb{I}_d, 0_d]$. This is a linear-Gaussian state-space representation that is susceptible of treatment through the Kalman filter. Due to the singularity of Q, the complete log-likelihood in Eq. (2) does not exist. The complete log-likelihood exists in a k-dimensional subspace of $\mathbb{R}^d$ generated by linear combinations of the components of X_t. We thus estimate the model by standard maximum-likelihood, i.e. by numerically optimizing the log-likelihood in Eq. (14). Compared to the standard MLA described in the paper, numerical optimization is feasible here, because the number of parameters of Q is $\mathcal{O}(k^2)$ rather than $\mathcal{O}(d^2)$. For instance, if there is only one security traded in d exchanges, we only need to estimate the variance of the unique innovation driving the dynamics in the d exchanges, regardless the value of d.

The diagonal elements of $\Psi^{(i)}$ induce a delay between the d_i exchanges and the i-th efficient log-price. For this reason they can be regarded as a measure of the informativeness of each exchange. For instance, if we find that one market anticipates another market, we conclude that the former is more informative. In principle one can study "cross-market" effects among different exchanges generated by non-diagonal coefficients in Ψ_i. This is computationally feasible when d_i is not too large, since the number of parameters of a non-diagonal Ψ_i scales as $\mathcal{O}(d_i^2)$. Similarly, to include cross-asset effects among prices corresponding to different securities, we need non-diagonal coefficients in Ψ. Even in this case one should pay attention that the total number of coefficients in Ψ does not grow too fast with d.

2.2 Comparison with other methodologies

As we have seen in the Monte-Carlo analysis of Section (3) in the paper, one of the main advantages of the MLA is that parameter estimates are not affected by differences in the level of trading activity among different assets. Similarly, in the case of arbitrage-linked securities, parameter estimates are not affected by differences in the level of trading activity among different markets. From an empirical point of view, this circumstance is of particular relevance in presence of informed traders. As predicted by classical models of price formation (Glosten and Milgrom 1985, Kyle 1985), the informed traders buy or sell securities if their trade guarantees a profit net of transaction costs, i.e. when their size relative to fundamental values is large. According to this logic, the informed traders tend to buy or sell in the market with larger mispricing between the efficient log-price and the mid-quote price.

To illustrate this concept, we consider the price formation model in Bandi et al. (2017) and adapt it to our framework with one security traded in several exchanges. In this model of price formation we have three components: an efficient log-price process, the midquote adjustments and the observed log-prices. The efficient log-price process follows a random walk:

$$P_{t+1} = P_t + u_{t+1} \tag{53}$$

where $\mathrm{Var}[u_t] = q$. We assume that the security described by P_t is traded in two exchanges. The midquote log-price is related to the efficient log-price by a lagged adjustment process:

$$X_{t+1}^{(i)} = X_t^{(i)} + \delta^{(i)}(P_{t+1} - X_t^{(i)}), \quad i = 1, 2 \tag{54}$$

The observed log-price depends on the trader type. Let us denote by $\mathcal{I}$ the probability of arrival of informed traders. For simplicity, we assume that $\mathcal{I}$ is the same in both markets. The informed trader knows the value of the efficient log-prices P_t and decides whether to trade or not by comparing the mispricing $|X_t^{(i)} - P_t|$ to the transaction cost $c^{(i)}$. More specifically, if $|X_t^{(i)} - P_t| > c^{(i)}$ the informed trader decides to trade and the observed log-price is:

$$Y_t^{(i)} = X_t^{(i)} + c \cdot 1_{\{P_t - X_t^{(i)} > c\}} - c \cdot 1_{\{P_t - X_t^{(i)} < -c\}}, \quad i = 1, 2 \tag{55}$$

In contrast, if $|X_t^{(i)} - P_t| \leq c^{(i)}$, the informed trader decides not to trade. Noise traders behave randomly. They simply toss a coin and decide whether to trade or not. When a noise trader arrives on the market, the observed log-price is:

$$Y_t^{(i)} = X_t^{(i)} + \nu_t^{(i)} c^{(i)} \tag{56}$$

where $\nu_t^{(i)}$ is a sequence of independent Bernoulli variables taking the values ± 1 with likelihood 50%.

If there are informed traders ($\mathcal{I} > 0$), the level of trading activity depends on the transaction cost $c^{(i)}$ and on the speed $\delta^{(i)}$ of adjustment to the efficient price. In particular, if $c^{(i)}$ is large, informed traders cannot reward themselves net of transaction costs, and decide not to trade. Similarly, if the market rapidly adjusts to the efficient price ($\delta^{(i)} \approx 1$), the mispricing $|X_t^{(i)} - P_t|$ is small and the prices are not updated. If $\mathcal{I}$ is sufficiently large, the latter case leads to a highly informative market (i.e. one with a high speed of adjustment to the efficient price) being less traded than an inefficient market (i.e. one which slowly adapts to the efficient price). In reality this situation might represent a transition to an equilibrium where the price in the less efficient market gradually reverts to the martingale process. This simple microstructure model can be extended into several directions, e.g. introducing a time-varying $\delta_t^{(i)}$ or allowing noise traders to take into account transaction costs (see Bandi et al. 2017). For simplicity we examine here the basic specification, however nothing prevents adding other features to simulate more and more realistic scenarios.

The MLA, though misspecified in this setting (note that the measurement noise in Eq. (56) is not normal), has a clear advantage in being robust to differences in the level of trading activity. To see this, let us assume that we have two markets, one which rapidly adjusts to new information ($\delta^{(1)} = 0.9$), and one which adapts slowly ($\delta^{(2)} = 0.2$). For simplicity, transaction costs are assumed to be the same in the two markets. We set the remaining parameters as $c^{(1)} = c^{(2)} = 0.2\,\$$, $q = 0.5$. As a time-horizon we consider a trading day of 6.5 hours, from 9:30 to 16:00. We thus simulate 23400

	MLA		IS bounds	
	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	1^{st} market	2^{nd} market
$\mathcal{I} = 0$	0.8456 (0.051)	0.2108 (0.027)	$(0.6187, 0.9969)$	$(0.0031, 0.3813)$
$\mathcal{I} = 0.3$	0.9175 (0.054)	0.1975 (0.022)	$(0.3777, 0.9136)$	$(0.0863, 0.6222)$
$\mathcal{I} = 0.5$	0.9181 (0.075)	0.1850 (0.042)	$(0.2717, 0.7692)$	$(0.2307, 0.7282)$
$\mathcal{I} = 0.7$	0.8713 (0.078)	0.2234 (0.051)	$(0.0835, 0.3679)$	$(0.6320, 0.9164)$

Table 1: MLA estimates of parameters $\delta^{(i)}$ in Eq. (54) with standard errors reported in parenthesis and IS bounds.

one-second realizations of the efficient log-price P_t in Eq. (53) and of the two lagged adjustment log-prices $X_t^{(1)}$ and $X_t^{(2)}$ in Eq. (54).

We start by assuming that there are no informed traders ($\mathcal{I} = 0$). The observed prices are determined only by noise traders who do not know the value of the true efficient log-price P_t. The level of trading activity in the two markets is thus the same. Table (1) shows the parameters $\hat{\delta}^{(i)}$, $i = 1, 2$ estimated by the MLA and corresponding to the diagonal elements of the matrix Ψ in Eq. (44). We also report the IS measure of Hasbrouck (1995). The IS is generally not uniquely defined since it depends on the order of the assets in the underlying VECM. For this reason, we report for each market the two bounds obtained by reversing the order of the two time-series. The MLA estimates are very close to the true parameters, and thus we conclude that the first market is the one where price discovery occurs. Using the IS we get to the same conclusion since the first market has larger information share and the bounds are relatively narrow.

As $\mathcal{I}$ increases, informed traders arrive on the market. They can decide not to trade should the mispricing $|X_t^{(i)} - P_t|$ be smaller than the transaction costs. The absence of trading leads to missing values in the two time-series. The MLA estimates remain close to the true parameters, and we still conclude that the first market is the one where price discovery occurs. In order to estimate the VECM, we fill the missing values by previous-tick interpolation, as commonly done in the financial econometric literature. Such procedure leads to erroneous conclusions on the degree of informativeness of the two markets. As $\mathcal{I}$ increases, the IS bounds widen. When $\mathcal{I} = 0.5$, the two markets have very similar bounds, and it is impossible to discern where price discovery occurs. For $\mathcal{I} = 0.7$, the bounds narrow but we wrongly conclude that the second market is more informative than the first. This is due to the first market being less traded than the second because of the large amount of informed traders who exploit arbitrage opportunities in the less efficient market.

More generally, asynchronous trading leads to missing values in the observed time-series which are typically filled by previous-tick interpolation. This leads, in turn, to a large number of "artificial" zero-returns which are misspecified under the common semimartingale assumption (in continuous-time) or under the VAR/VECM assumption (in discrete-time). The most known distortions of zero-returns is the Epps effect (cf. e.g. Hayashi and Yoshida 2005 and references therein). Thus, the fact that we find misleading results when applying the VECM to asynchronous data is not surprising. The main advantage of the MLA is that it can handle missing observations without introducing artificial zero returns.

There are other differences between IS and the MLA with cointegrated dynamics. The IS of the i-th exchange is defined as the fraction of the long-run variance of the common trend imputable to that exchange. The market contributing with the largest fraction of variance is the most informative, i.e. the one where price discovery occurs. In the MLA with cointegrated dynamics, the most informative market is the one with highest speed of adjustment to the efficient log-price. In other words the market that leads all the other markets is the one where price discovery occurs.

Using these two approaches one may achieve a different conclusion on which market is the most informative. The main reason is that IS does not depend on the matrices of autoregressive coefficients of the underlying VECM, as formally shown by Buccheri et al. (2019b). This implies that,in some circumstances, a market with a substantial delay from the common trend may be judged as equally informative or even more informative than a faster market (see the examples in Buccheri et al. 2019b). In the MLA this cannot happen since the most informative market is, by definition, the one which adapts with highest speed to the common trend represented by the efficient log-price process. We refer to Buccheri et al. (2019b) for further discussions on these aspects.

Finally, we point out that, if quote data are available, X_t becomes observable and thus IS is not affected by asynchronous trading. However, compared to the MLA, it is still true that it may lead to misleading results in the presence of significant lags between the observed prices and the underlying martingale process (cf. Buccheri et al. 2019b).

3 Robustness checks

3.1 Model invariance to the choice of the assets

In Section (4.2) in the paper we estimated the MLA on a cross-section of 10 NYSE stocks and showed the average cross-autocorrelations in Figures (10), (11). The question naturally arises whether these lead-lag correlations depend on the specific choice of the dataset and whether by selecting a subset of these assets one would obtain the same result. This is a standard issue in the specification of VAR models (see e.g. Kilian and Lütkepohl 2017).

In order to investigate if the recovered lead-lag correlations are robust with respect to the choice of the assets, we perform the same analysis of Section (4.2) in the paper but here we estimate a different MLA for each couple of assets. Specifically, in the first case, the lead-lag correlations of two assets are computed using the log-prices of all the 10 NYSE assets, while in the second case they are computed based only on the log-prices of the two assets. This comparison is interesting because we are considering the scenario in which the discrepancy between the two kinds of lead-lag correlations is largest: estimating the MLA on cross-sections of growing dimensions provides lead-lag correlations which become more and more similar to those obtained using the entire cross-section.

The results are reported in Figures (1), (2), where we show in blue the correlogram obtained in Section (4.2) in the paper and in red the new lead-lag correlations. The lead-lag correlations recovered using the whole dataset of 10 assets are very similar to those obtained by estimating the MLA pairwisely. We only observe that, due to data reduction, pairwise correlations are slightly lower in some cases. However, our conclusions on which among two assets is the leader is invariant with respect to the choice of the dataset.

In Section (4.3) in the paper we assess the effect of the inclusion of the SPY in the sample of assets used for the estimation. The SPY differs significantly from the 10 NYSE assets in terms of liquidity, as it features a larger number of trades per day. Even in that case we find that the recovered cross-autocorrelation structure is not altered significantly.

3.2 Statistical significance at sparser modelling frequencies

All the empirical results reported in Section (4) have been recovered at the sampling frequency of one second, which is the highest frequency achievable in our dataset. It is interesting to investigate how these results would change at sparser frequencies. As a result of data reduction, the use of

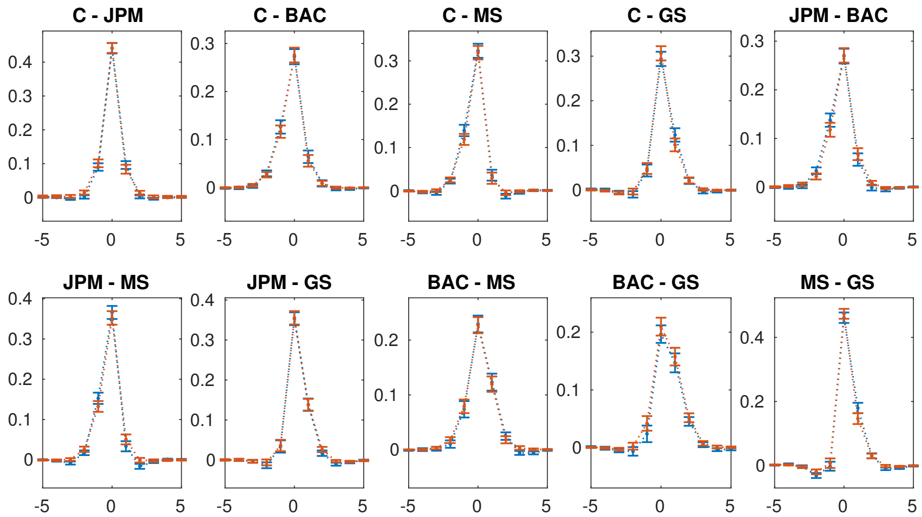

Figure 1: Average cross-autocorrelations of all the couples of assets in Group I. Averages are computed over all the business days of 2014. We show in blue the lead-lag correlations computed as in Section (4.2) in the paper and in red the lead-lag correlations computed pairwisely. Error bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset displayed in the title leads the first asset and the other way around for negative lags.

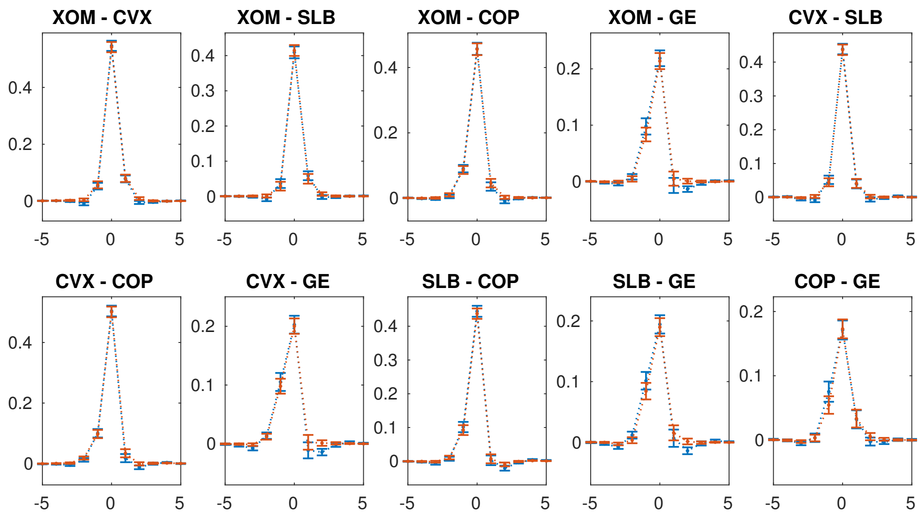

Figure 2: Average cross-autocorrelations of all the couples of assets in Group II. Averages are computed over all the business days of 2014. We show in blue the lead-lag correlations computed as in Section (4.2) in the paper and in red the lead-lag correlations computed pairwisely. Error bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset displayed in the title leads the first asset and the other way around for negative lags.

sparser sampling frequencies naturally leads to a lower statistical efficiency. However, the loss of efficiency is not the only source of concern when subsampling prices at sparser frequencies. Another aspect that must be taken into account is the fact that, being generated by high-frequency trading strategies, lead-lad dependencies exist at small time-scales and tend to decay at longer time-scales.

To illustrate this phenomenon, we compare in Figure (3) the average cross-autocorrelations of the assets of Group I computed at the sampling frequency of one second (in red) and the cross-autocorrelations of the same assets computed at the sampling frequency of 10 seconds (in blue). We use the subsample of Section (4.3) which includes SPY data. Figure (4) shows a similar comparison for the lead-lag correlations between the SPY and the five assets of Group I. We immediately note that, as a consequence of modelling prices at sparser sampling frequencies, the contemporaneous correlations increase and the lead-lag correlations decrease. In other words, the lead-lag correlations detected at higher resolutions are "averaged out" and are eventually seen as contemporaneous correlations when observing the market at longer time-scales. The cross-autocorrelation structure of the market thus emerges at small time-scales, where algorithmic trading strategies are more likely to operate. We obtain the same result when looking at the cross-autocorrelations of the assets of Group II.

The result of this analysis serves as a guideline for the empirical implementation of the MLA. In order to exploit more and more information sets, it is preferable to choose the highest available sampling frequency. In certain circumstances (e.g. when data are available at the millisecond precision) the computational power required for the estimation might be onerous, depending on the cross-section dimension. As underlined in Section (2.2) in the paper, a feasible solution is to reduce the intraday estimation window. Given the large amount of high-frequency data available even on sub-periods of the trading day, the choice of a shorter window does not affect significantly the quality of the inference.

3.3 Potential diurnal effects in cross-asset trading

Intraday covariances are characterized by well-known diurnal effects (cf. e.g. Andersen and Bollerslev, 1997, Tsay, 2005, Bibinger et al., 2014, Buccheri et al., 2019a). For instance, volatilities are larger at the beginning and at the end of the trading day while correlations tend to increase throughout the day. We thus wonder whether cross-asset trading exhibits similar intraday patterns. Of course, lead-lag effects are naturally influenced by the intraday pattern of covariances.

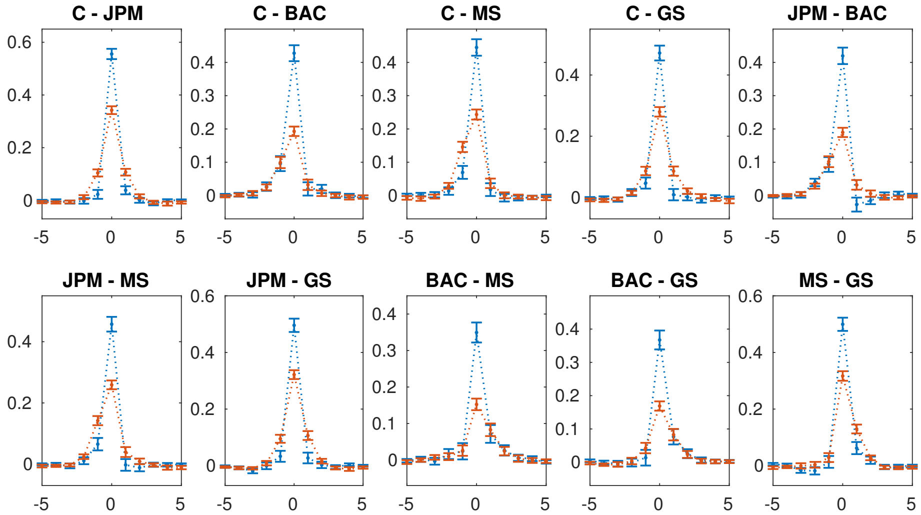

Figure 3: Average cross-autocorrelations of all the couples of assets in Group I. Averages are computed over all the business days of 2012. We show in blue the lead-lag correlations computed at the sampling frequency of 10 seconds and in red the lead-lag correlations computed at the sampling frequency of one second. Error bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset displayed in the title leads the first asset and the other way around for negative lags.

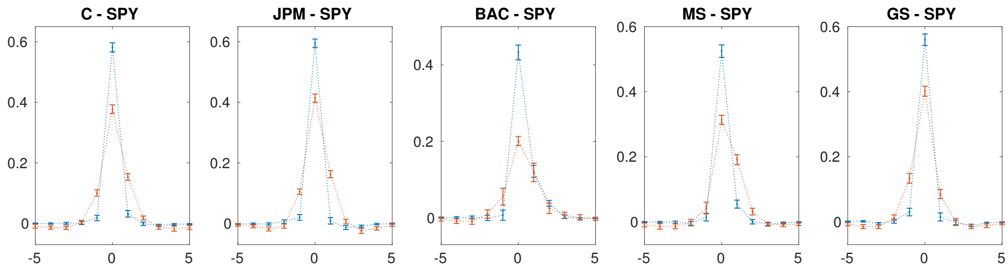

Figure 4: Average cross-autocorrelations between SPY and the stocks of Group I. Averages are computed over all the business days of 2012. We show in blue the lead-lag correlations computed at the sampling frequency of 10 seconds and in red the lead-lag correlations computed at the sampling frequency of 1 second. Error bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset displayed in the title leads the first asset and the other way around for negative lags.

	9:30−11:00	11:01−14:30	14:31−16:00
N. of coefficients of $\theta_t^{(i)}$ for which H_0 is not rejected (out of 121)	108	114	98
Avg. p-value	0.85	0.87	0.84

Table 2: We report for each of the three sub-periods of the trading day the number of coefficients of $\theta_t^{(i)}$ for which H_0 is not rejected at the 5% confidence level. We also show the average p-value of the 121 t-tests performed on each time series.

We are instead interested in potential diurnal effects induced by the lagged matrix Ψ of lagged price adjustment.

In order to investigate the behavior of Ψ at the intraday level, we consider the same subsample used in the previous analysis and divide the trading day into three sub-periods, the first from 9:30 to 11:00, the second from 11:01 to 14:30 and the third from 14:31 to 16:00. In each of these sub-periods we estimate the MLA and recover three lagged adjustment matrices, $\Psi_t^{(1)}$, $\Psi_t^{(2)}$, $\Psi_t^{(3)}$, where t is a daily index going from 03-01-2012 to 28-12-2012. We compare each of these matrices with the matrix Ψ_t estimated in the entire day. Specifically, for $i = 1, 2, 3$, we consider the differences $\theta_t^{(i)} = \text{vec}(\Psi_t^{(i)} - \Psi_t)$ and test the null hypothesis H_0 that the $d^2 = 121$ elements of $\theta_t^{(i)}$ have mean equal to zero.

Table (2) shows the results of the one-sample t-test performed for each of the 121 time-series corresponding to the coefficients of $\theta_t^{(i)}$, for $i = 1, 2, 3$. In the first line we report the number of coefficients for which the null hypothesis H_0 is not rejected at the 5% confidence level. In the second line we report the average p-value of the 121 t-tests. It is immediate to note that the vast majority of the coefficients of $\Psi_t^{(i)}$ estimated in the three sub-periods of the trading day are statistically indistinguishable from the coefficients of Ψ_t estimated in the entire period. The coefficients for which H_0 is rejected have p-values that are slightly smaller than 5% and always larger than 1%. This result corroborates the assumption of a constant lagged adjustment matrix Ψ in Section (2) in the paper and shows that cross-asset trading does not change significantly during the trading day.

References

Andersen, T., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4 (2-3), 115–158.

Bandi, F. M., Pirino, D., Reno, R., 2017. EXcess Idle Time. Econometrica 85 (6), 1793–1846.

Bibinger, M., Hautsch, N., Malec, P., Reiss, M., 2014. Estimating the spot covariation of asset prices: Statistical theory and empirical evidence. CFS Working Paper Series 477, Center for Financial Studies (CFS).

Booth, G. G., So, R. W., Tse, Y., 1999. Price discovery in the german equity index derivatives markets. Journal of Futures Markets 19 (6), 619–643.

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2019a. A score-driven conditional correlation model for noisy and asynchronous data: an application to high-frequency covariance dynamics. Working Paper, Available at https://ssrn.com/abstract=2912438.

Buccheri, G., Bormetti, G., Lillo, F., Corsi, F., 03 2019b. Comment on: Price Discovery in High Resolution. Journal of Financial Econometrics.

Chu, Q. C., liang Gideon Hsieh, W., Tse, Y., 1999. Price discovery on the s&p 500 index markets: An analysis of spot index, index futures, and spdrs. International Review of Financial Analysis 8 (1), 21 – 34.

De Jong, F., Schotman, P. C., Winter 2010. Price Discovery in Fragmented Markets. Journal of Financial Econometrics 8 (1), 1–28.

deB. Harris, F. H., McInish, T. H., Wood, R. A., 2002. Security price adjustment across exchanges: an investigation of common factor components for dow stocks. Journal of Financial Markets 5 (3), 277 – 308, price Discovery.

Glosten, L. R., Milgrom, P. R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14 (1), 71 – 100.

Hamilton, J., 1994. Time series analysis. Princeton Univ. Press, Princeton, NJ.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. The Journal of Finance 50 (4), 1175–1199.

Hayashi, T., Yoshida, N., 04 2005. On covariance estimation of non-synchronously observed diffusion processes. Bernoulli 11 (2), 359–379.

Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1335.

Shumway, R. H., Stoffer, D. S., 1982. An approach to time series smoothing and forecasting using the em algorithm. Journal of Time Series Analysis 3 (4), 253–264.

Shumway, R. H., Stoffer, D. S., 2015. Time series analysis and its applications : with R examples. Springer texts in statistics. Springer, New York.

Tsay, R. S., 2005. Analysis of financial time series. Wiley series in probability and statistics. Wiley-Interscience, Hoboken (N.J.).

Wu, C. F. J., 03 1983. On the convergence properties of the em algorithm. Ann. Statist. 11 (1), 95–103.