

# Testing the Markov property with high frequency data

**João Amaro de Matos**

Faculdade de Economia, Universidade Nova de Lisboa

Rua Marquês de Fronteira, 20, 1099-038 Lisbon, Portugal

E-mail: [amatos@fe.unl.pt](mailto:amatos@fe.unl.pt)

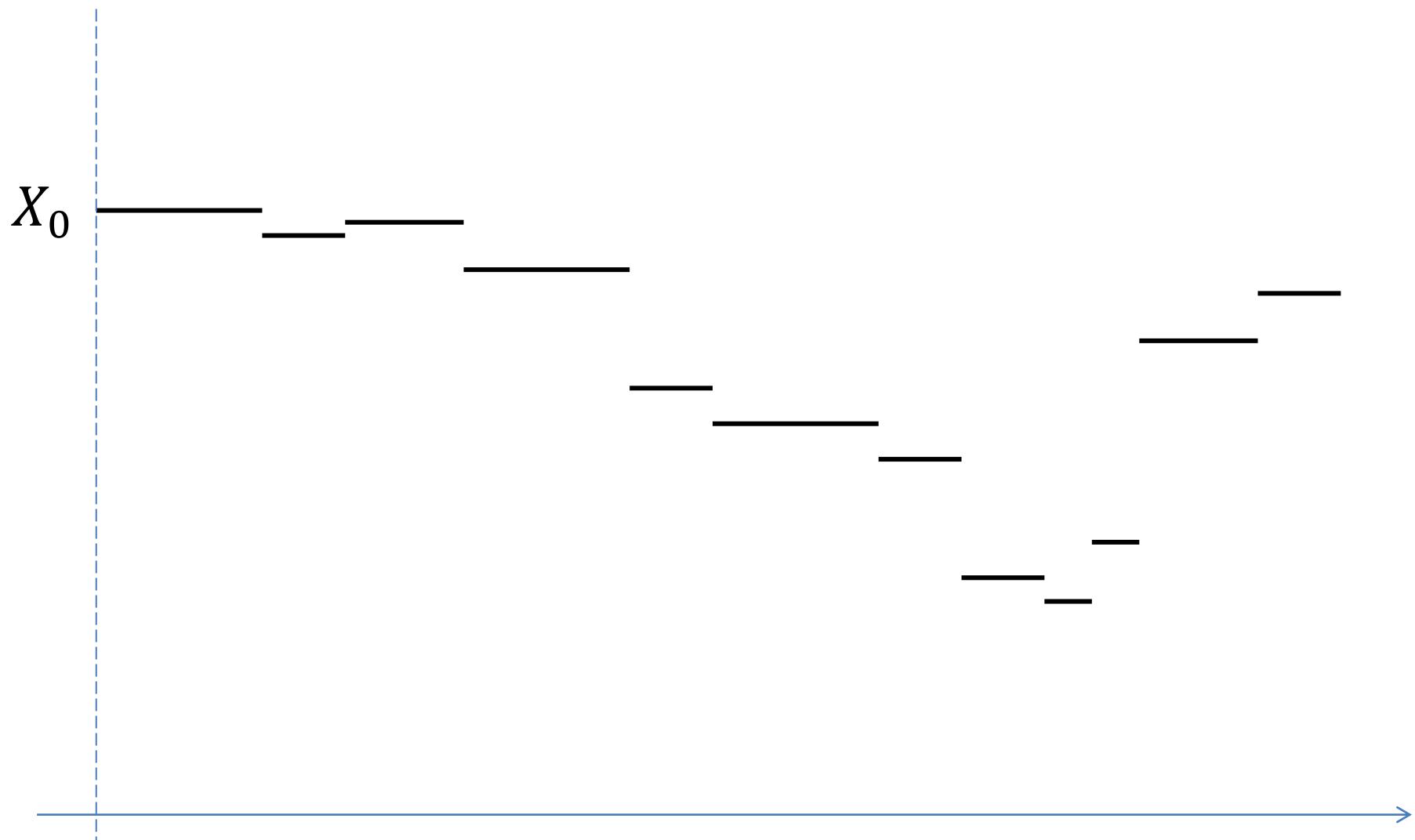
**Marcelo Fernandes**

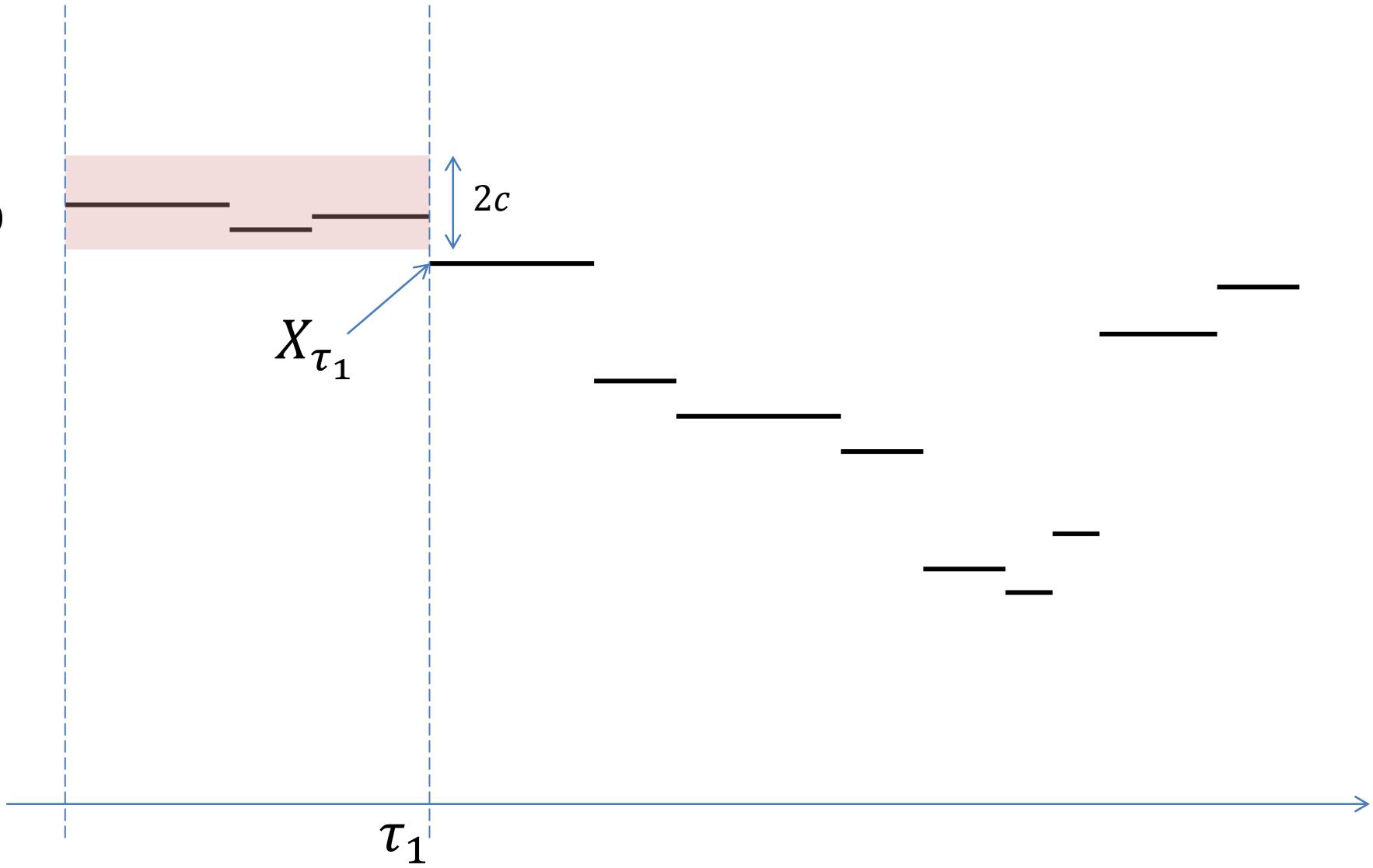
Economics Department, Queen Mary, University of London

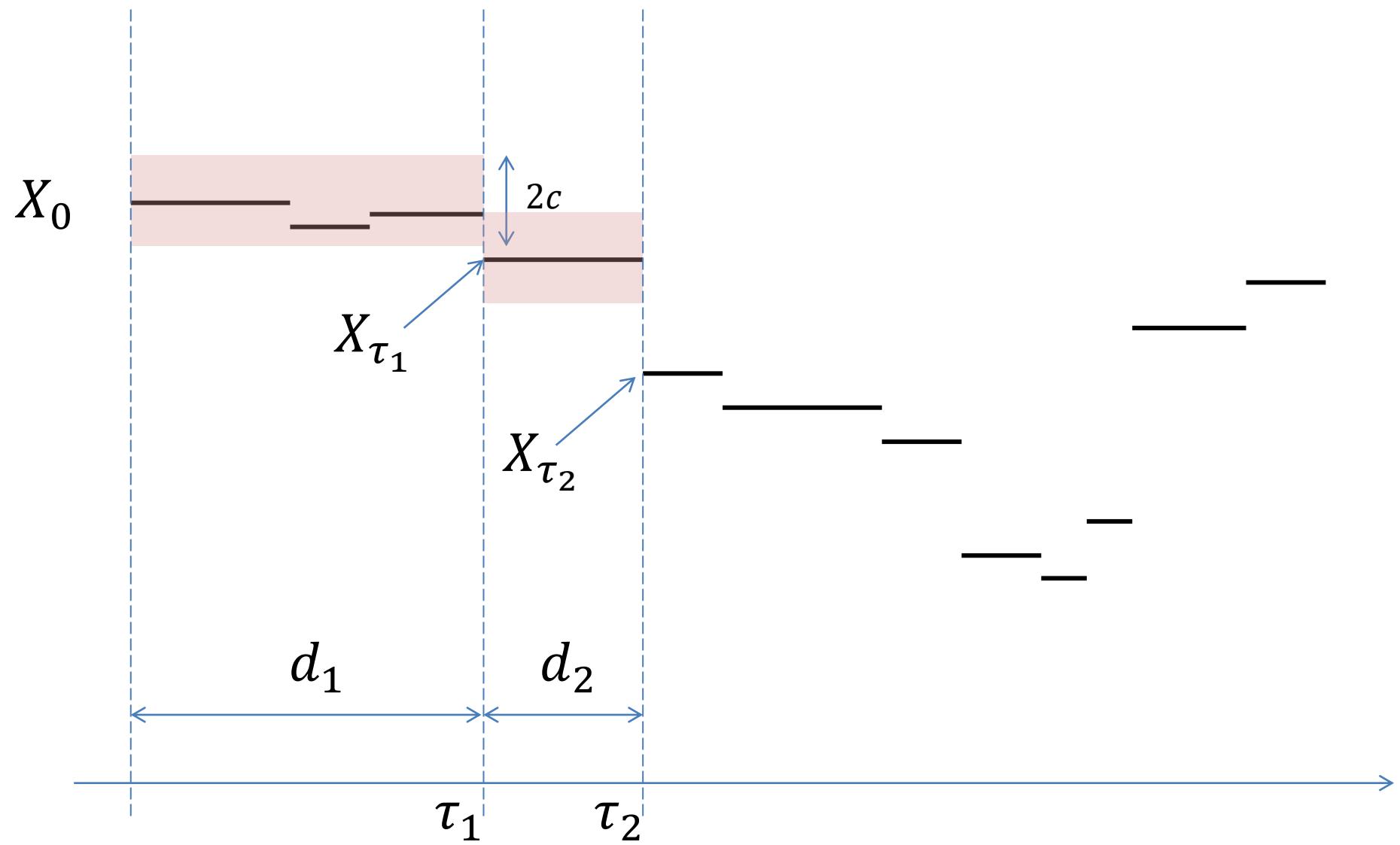
Mile End Road, E1 4NS, London, UK

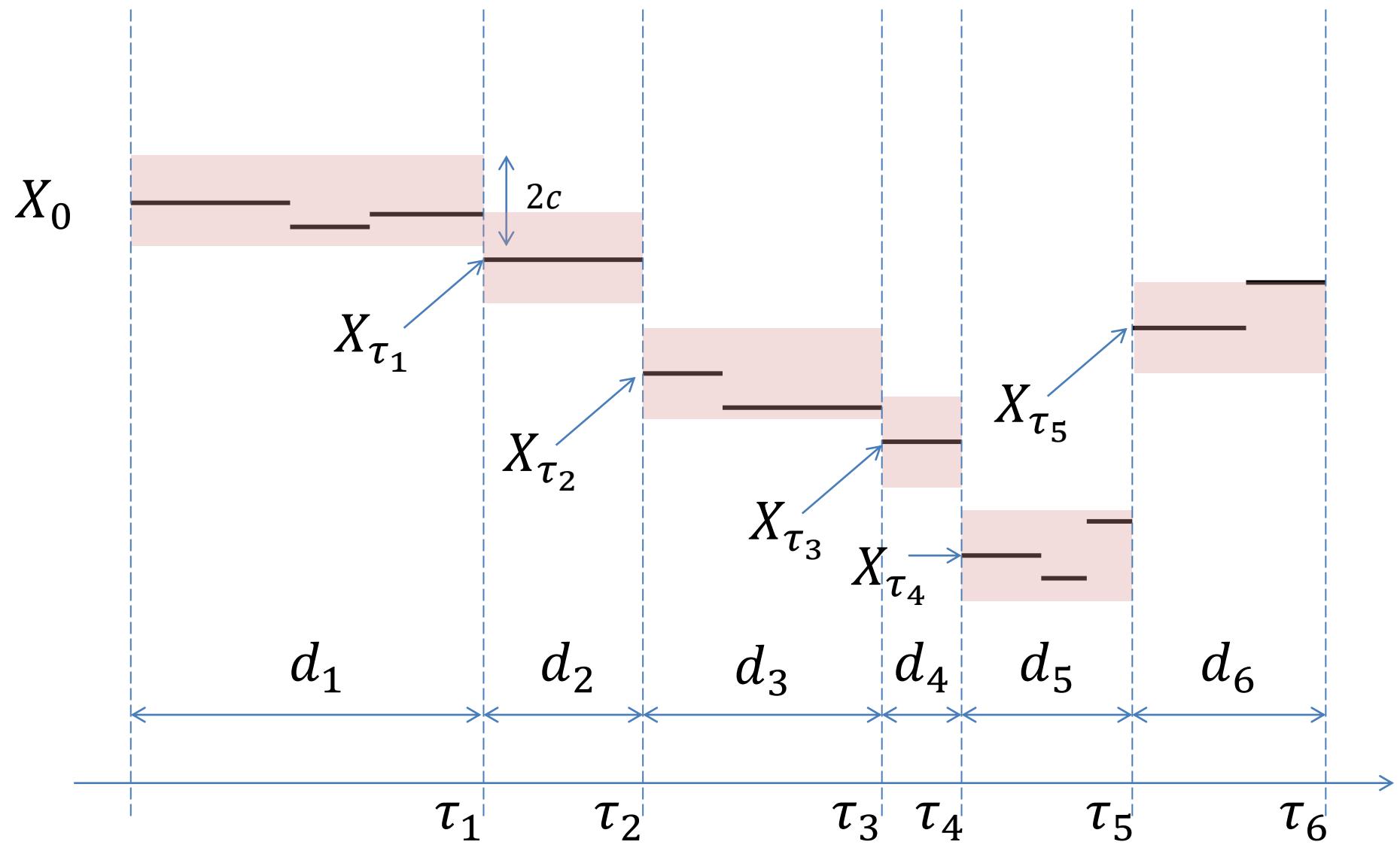
E-mail: [m.fernandes@qmul.ac.uk](mailto:m.fernandes@qmul.ac.uk)

$P_t$   $\xrightarrow{\text{log transform}}$   $\log(P_t)$   $\xrightarrow{\text{Z transform}}$   $X_t$



$X_0$  $X_{\tau_1}$  $2c$  $\tau_1$ 





$X_t$  is strong Markov



$X_i := X_{\tau_i}$  is Markov

and

$d_i$  and  $d_{i+1}$  are independent if we know  $X_i$

$$H_0 : f_{iXj}(a_1, x, a_2) = f_{i|X}(a_1)f_{Xj}(x, a_2) \text{ a.e. for a fixed } j.$$

$$\Lambda_f = \mathbb{E} \left\{ w_{iXj} \left[ f_{iXj}(d_i, X_j, d_j) - f_{i|X}(d_i|X_j)f_{Xj}(X_j, d_j) \right]^2 \right\}, \quad (3)$$

where  $w_{iXj} = w(d_i, X_j, d_j) = \mathbf{1}[(d_i, X_j, d_j) \in \mathcal{S}]$  trims the data down to a compact support  $\mathcal{S}$

Letting  $j = i - j$  and  $n = N - j$ , the sample analog of (3) is

$$\Lambda_{\hat{f}} = \frac{1}{n} \sum_{k=1}^n w(d_{k+j}, X_k, d_k) \left[ \hat{f}_{iXj}(d_{k+j}, X_k, d_k) - \hat{g}_{iXj}(d_{k+j}, X_k, d_k) \right]^2$$

where  $\hat{g}_{iXj}(d_{k+j}, X_k, d_k) = \hat{f}_{i|X}(d_{k+j}|X_k) \hat{f}_{Xj}(X_k, d_k) = \frac{\hat{f}_{iX}(d_{k+j}, X_k)}{\hat{f}_X(X_k)} \hat{f}_{Xj}(X_k, d_k)$ ,

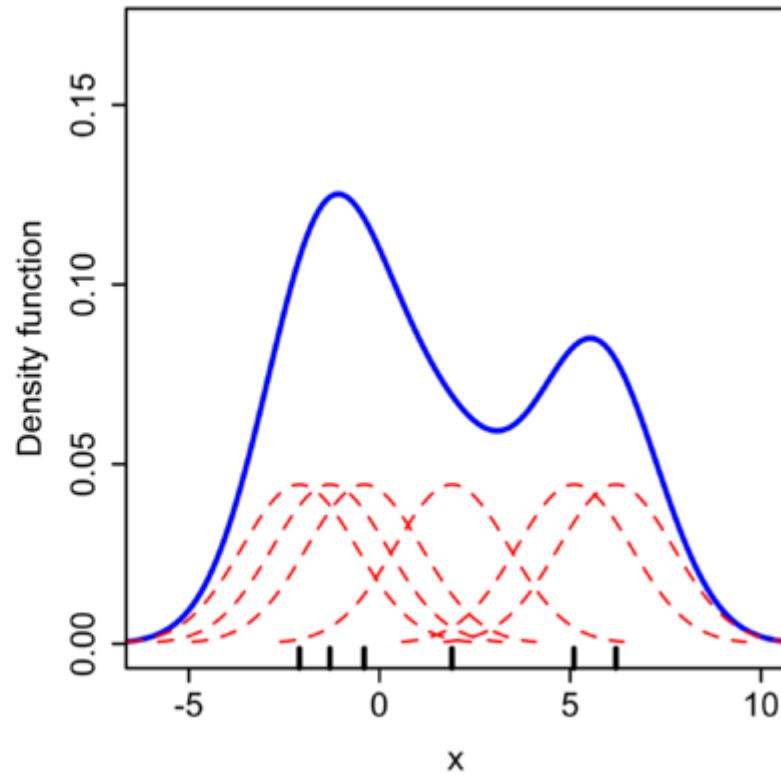
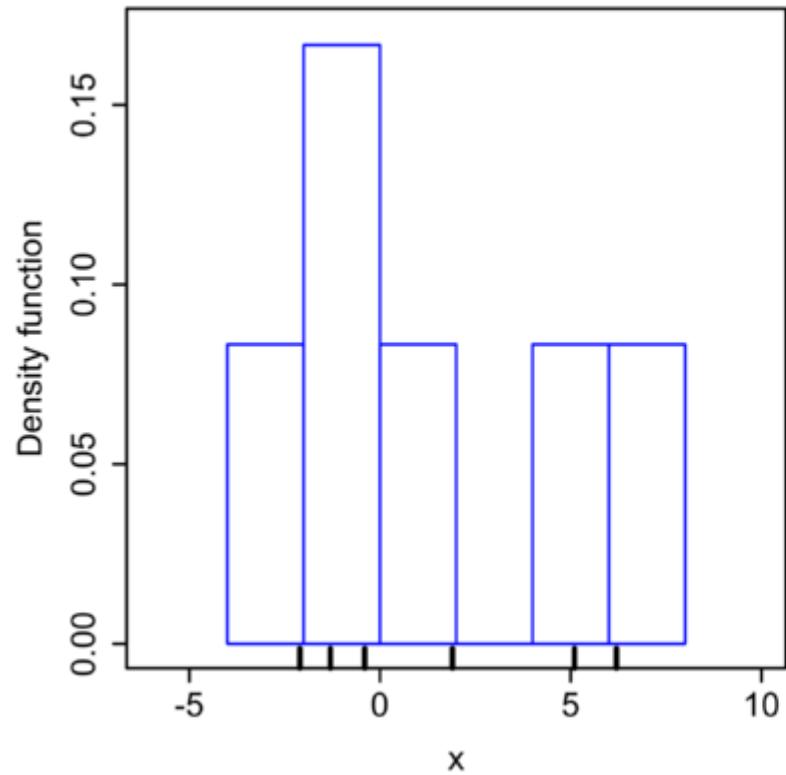
$$\hat{f}_{iXj}(a_1, x, a_2) = \frac{1}{n b_n^3} \sum_{k=1}^n K\left(\frac{a_1 - d_{k+j}}{b_n}\right) K\left(\frac{x - X_k}{b_n}\right) K\left(\frac{a_2 - d_k}{b_n}\right),$$

$$\hat{f}_{iX}(a_1, x) = \frac{1}{n b_n^2} \sum_{k=1}^n K\left(\frac{a_1 - d_{k+j}}{b_n}\right) K\left(\frac{x - X_k}{b_n}\right),$$

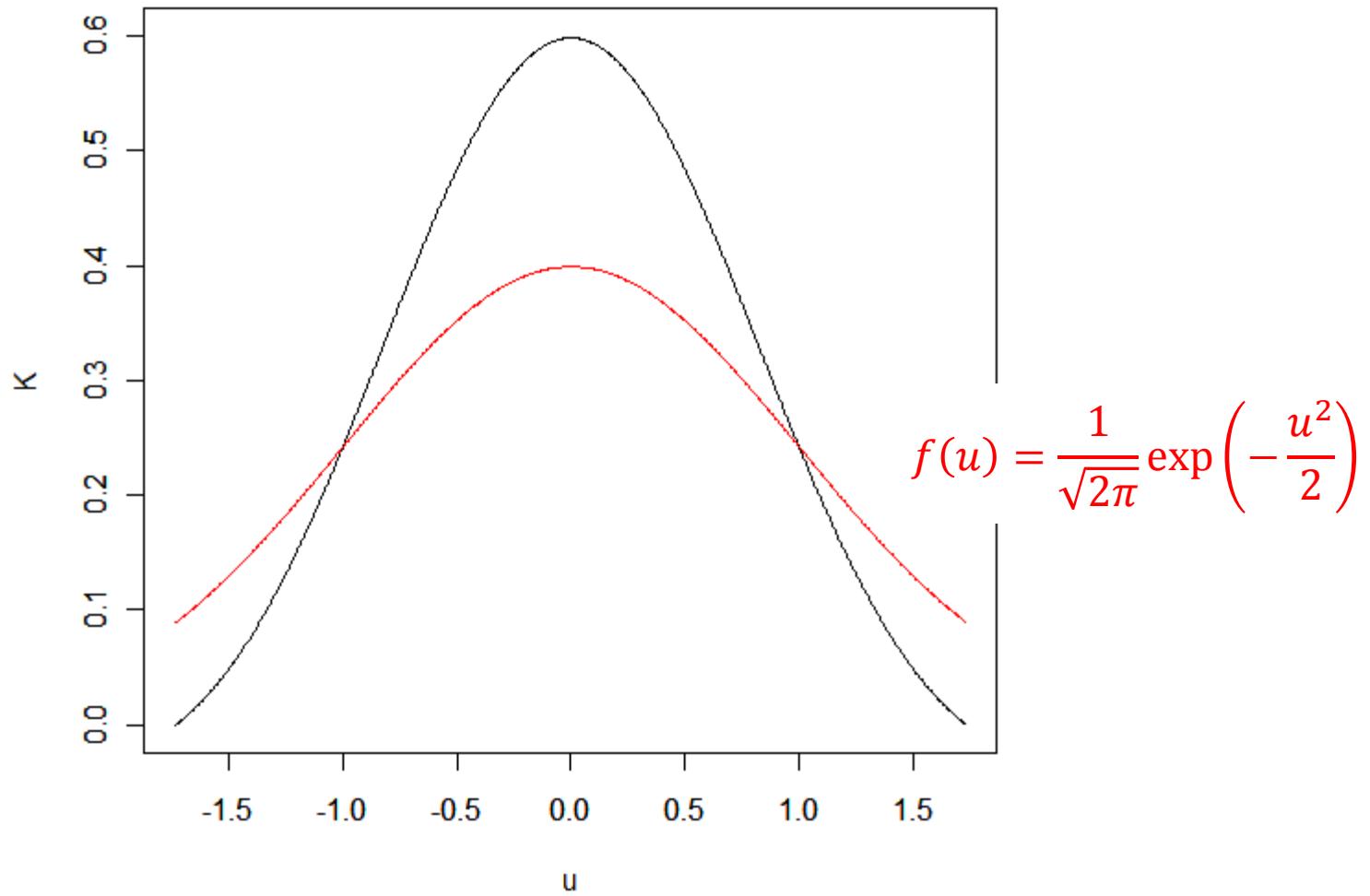
$$\hat{f}_{Xj}(x, a_2) = \frac{1}{n b_n^2} \sum_{k=1}^n K\left(\frac{x - X_k}{b_n}\right) K\left(\frac{a_2 - d_k}{b_n}\right),$$

$$\hat{f}_X(x) = \frac{1}{n b_n} \sum_{k=1}^n K\left(\frac{x - X_k}{b_n}\right),$$

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$



$$K(u) = \frac{3}{\sqrt{8\pi}} \left(1 - \frac{u^2}{3}\right) \exp\left(-\frac{u^2}{2}\right)$$



PROPOSITION 1: Under the null hypothesis  $H_0$  and Assumptions A1 to A4, the statistic

$$\hat{\lambda}_n = \frac{n b_n^{3/2} \Lambda_{\hat{f}} - b_n^{-3/2} \hat{\sigma}_{\Lambda}}{\hat{\sigma}_{\Lambda}} \xrightarrow{d} N(0, 1), \quad (4)$$

where  $\hat{\delta}_{\Lambda}$  and  $\hat{\sigma}_{\Lambda}^2$  are consistent estimates of  $\delta_{\Lambda} = e_K \mathbb{E}(w_{iXj} f_{iXj})$  and  $\sigma_{\Lambda}^2 = v_K \mathbb{E}(w_{iXj}^2 f_{iXj}^3)$ , respectively.

## Finite sample properties

- S1 Draw the initial observation  $X_0^{(b)}$  from the kernel-based nonparametric estimate of the stationary distribution of the bid-ask spreads and then draw the remaining artificial sample  $\{d_k^{(b)}, X_k^{(b)}\}_{k=1}^m$  from the kernel estimates  $\hat{F}\left(X_k, d_k \mid X_{k-1} = X_{k-1}^{(b)}\right)$  of the conditional distribution of  $(d_k, X_k)$  given the previous realization of the bid-ask spread. This is the bootstrap sample, for which the null hypothesis in (2) holds conditional on the original sample.
- S2 Compute the test statistic  $\hat{\lambda}_m^{(b)}$  as in (4) using the bootstrap sample.
- S3 Repeat the steps S1 and S2 for a large number of time, say  $B$ , and obtain the empirical distribution function of  $\{\hat{\lambda}_m^{(b)}\}_{b=1}^B$ .