EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

EUI Working Paper **ECO** No. 2002/22

Functional Weak Limit Theory for Rare Outlying Events

Iliyan Georgiev

BADIA FIESOLANA, SAN DOMENICO (FI)

Published in Italy in November 2002
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

Functional Weak Limit Theory for Rare Outlying Events

Iliyan Georgiev

October 2002

Abstract

The paper suggests a model of stochastic outliers whose number remains bounded as the sample size increases. A theorem for weak convergence to a compound Poisson process is combined with a standard FCLT to obtain the asymptotic distributions of statistics depending on both the ordinary and the outlying shocks that affect a time series. Results for deterministic models of outliers are derived as special cases by conditioning, and a specification of the outliers' size as a function of the sample size results in properties similar to those of asymptotically frequent stochastic outliers.

[0]European University Institute, Florence. E-mail: ilijan.georgiev@iue.it

I am very grateful to Søren Johansen who supervised my work and made detailed comments on all previous versions of the paper. Thanks are also due to N. Hargreaves whose linguistic expertise was helpful for improving the readability of the paper. The remaining deficiencies are to be blamed solely on me.

1 Introduction

A constant concern in time-series econometrics is the modelling of outlying events, both with permanent effect (structural breaks) and with temporary effect (generating temporary-change outliers). In this paper the models are classified according to two aspects of the timing of such events: deterministic versus stochastic timing, and finite versus infinite asymptotic frequency of occurrence.

Timing is deterministic if either the absolute or the relative (w.r.t. the sample size) dates at which outlying events occur are considered known. When timing is stochastic, however, the occurrence of an outlying event is modelled as random. As Perron [7] notes, the deterministic specification can be thought of as obtained from the stochastic one by conditioning on the dates of outlying events.

According to whether the (expected) number of outlying events remains bounded in asymptotic arguments or not, I will call these events respectively asymptotically rare and asymptotically frequent.

An example of deterministic asymptotically rare structural breaks is provided in Perron's work ([7], [8]), where the number and the relative position of breaks in the sample are fixed.

Franses and Haldrup [4], hereafter FH, work with a stochastic specification of outliers. With the notation of this paper, for $p \in (0, 1)$ let $\{\pi_t\}_{t=1}^T$ be a sequence of Bernoulli $B(1, p)$ random variables (rv's) and $\{\eta_t\}_{t=1}^T$ be a sequence of zero-mean rv's. Then $\pi_t\eta_t$ generates outliers: when π_t is one, an outlying shock with size η_t occurs. In contrast with Perron's approach, the outlying shocks of FH are asymptotically frequent since their expected number, Tp, grows at the rate of T. Regular shocks (η_t) and outliers ($\pi_t\eta_t$) are put asymptotically on equal footing in the sense that the same limit theorems apply to random elements constructed from η_t and to those constructed from $\pi_t\eta_t$.

The asymptotic theory suggested here, similarly to FH, is for stochastic outliers. However, as in Peron's work, the outliers are asymptotically rare. To the contrary of frequent-events asymptotics, the rare-events approach preserves the difference between ordinary shocks and outliers in the limit. It leads to natural results - first, the weak limit of a discrete-time jump process is a continuous-time jump process, not a continuous process as it would be in FH. The analogue of the FCLT, under the condition that the expected number of jumps is independent of T, is (with a slight abuse of notation) $\sum_{t=1}^{[Tu]} \pi_t \eta_t \xrightarrow{w} J(u) \overset{d}{=} \sum_{i=1}^{N(u)} \eta_i$. Here the limiting process is a compound Poisson process whose counting process $N(u)$ has on average as many jumps as $\sum_{t=1}^{[Tu]} \pi_t \eta_t$. Wiener asymptotics for the ordinary shocks can be combined with Poisson asymptotics for the rare shocks to get jump-diffusions in the limit.

Another natural aspect of rare-events asymptotics is that, by a conditioning argument, they have the results for deterministic asymptotically rare outliers as special cases. For example, consider a white noise with a single level shift, $x_t = \varepsilon_t + \gamma I(\{T\lambda < t\})$, where $\varepsilon_t \sim iid(0, \sigma^2)$. Then $\frac{1}{T}\sum_{t=1}^{T} x_t^2 \xrightarrow{P} \sigma^2 + \lambda\gamma^2$. Let the stochastic formulation $x_t = \varepsilon_t + \sum_{i=1}^{t} \pi_i \eta_i$ be adopted instead, with the assumption that level shifts are independent of the ε_t-s (an exogeneity assumption). If rare-events asymptotics are applied, the convergence $\frac{1}{T}\sum_{t=1}^{T} x_t^2 \xrightarrow{w} \sigma^2 + \int_0^1 J^2$ obtains. Conditionally on a single jump of size γ occurring at relative time λ, the latter result specializes to the deterministic one (see section 2.4).

The paper has the following structure. The next section contains the basic functional convergence theorem and some corollaries necessary for the analysis of econometric models. Applications are provided in section 3, where examples of Perron [8] and FH are rephrased and analyzed by means of the claims in section 2, and the results are compared with the original ones. Section 4 concludes. All proofs are collected in an appendix.

2 A functional limit theorem for jump processes

2.1 Definition and comments

For two vectors x and y of the same dimension, let $x \cdot y$ denote their Hadamard (component-wise) product.

Definition 1 *A k-dimensional random walk μ_t, $t = 1, ...,$ is called a* ***jump process*** *when it has the representation*

$$\mu_t = \sum_{i=1}^{t} \pi_i \cdot \eta_i,$$

where:

π_t is a k-dimensional iid sequence, the components of which are natural-valued random variables, each of them taking the value of 0 with a positive probability, and are independent both serially and contemporaneously;

$\eta_t \sim iid\,(0, \Sigma_\eta)$ *is a sequence of* k*-vectors*;

π_t *is independent of* η_s, $t, s = 1, 2, ...$

The name reflects the idea that μ_t "jumps" when π_t is different from zero. The independence properties are not related to the jump behavior, but are nonetheless included in the definition because throughout the whole subsequent argument they are assumed to hold.

In the discussion of asymptotics, sequences of jump processes indexed by T are considered: $\mu_{Tt} = \sum_{i=1}^{t} \pi_{Ti} \cdot \eta_i$. All processes μ_T are generated with the same sequence of jump sizes η_t but with different sequences of jump indicators π_{Tt}. The distributions of π_T are related by the requirement that $E\left(\Sigma_{t=1}^{T} \pi_{Tt}\right) \rightarrow \varkappa$, by which the number of jumps is stochastically bounded in T.

For the practical purposes of data analysis a sample of fixed size T is supposed to be available, and the data generating process is thought of as involving in one form or another μ_T for that particular T. Thus the observed data depend on only one term of the sequence of jump processes. The remaining terms only exist in the ideal world.

This setup is simply a device that allows the rarity of the jumps to be preserved in the limit. The asymptotics derived for the sequence μ_T will be referred to as asymptotics for jump processes with rare jumps. These are non-Wiener asymptotics for a particular sequence of random walks.

The setup can be regarded as a generalization of the following construct characterized by jumps at fixed relative dates. For $k = 1$, let $N(u)$ be a Poisson process defined on $[0, 1]$ and with intensity $\varkappa$. Define $\pi_{Tt} = N\left(\frac{t}{T}\right) - N\left(\frac{t-1}{T}\right)$, $t = 1, ..., T$. Then $E\left(\Sigma_{t=1}^{T} \pi_{Tt}\right) = E\left(N\left(1\right)\right) = \varkappa$, and

$$\mu_{T[Tu]} = \sum_{t=1}^{[Tu]} \pi_{Tt} \cdot \eta_t = \sum_{i=1}^{N(u)} \eta_{[\tau_i T]+1},$$

where τ_i are the times of jump of $N\left(u\right)$. Heuristically, for big T the probability of observing more than one jump in the intervals $[\frac{t-1}{T}, \frac{t}{T})$ becomes negligible, and asymptotically all addends on the RHS tend to be different terms of the sequence η_t. More precisely, it can be shown that $\mu_{T[Tu]} \overset{w}{\rightarrow} \sum_{i=1}^{N(u)} \eta_i$ (the meaning of this weak convergence is clarified in section 2.2).

In fact, the weak convergence result can be obtained under fewer assumptions, e.g. it will still hold if instead of $N(u)$ a sequence $N_T\left(u\right)$ of processes distributed as $N\left(u\right)$ is used to generate π_{Tt}. This corresponds to fixing the distribution of the jump times (and the expected number of jumps in particular) but without creating any dependence between jump times for different T.

This paper concentrates on weak limit theory for binary binomial π_{Tt}'s.

Notation 2 *In the rest of the text, the Hadamard product $\pi_{Tt} \cdot \eta_t$ is denoted by δ_{Tt}, or by δ_t when the dependence on T is subsumed.*

2.2 The theorem

The argument takes place in the set of all functions $[0,1] \to \mathbb{R}^k$ which are continuous from the right and with limits from the left (cadlag). When this set is endowed with the $k-$dimensional Skorohod metric (3), the resulting metric space will be denoted by $D_{\mathbb{R}^k}[0,1]$; it is complete and separable, [5]. The open balls w.r.t. d_{B^k} generate the Borel sigma algebra $\mathcal{D}^k$ on this space, which turns out to be the product of the sigma algebras generated in the coordinate spaces by the 1-dimensional Skorohod metric (although $D_{\mathbb{R}^k}[0,1]$ is not a product space, see the appendix).

Let π_{Tt} and η_t be defined on some probability space $(\Omega, \mathcal{B}, \mathbf{P})$. Consider for $u \in [0,1]$ the process $\mu_T(u)$ defined by

$$\mu_T(u) = \sum_{t=1}^{[Tu]} \delta_{Tt}.$$

Then, for a fixed $\omega \in \Omega, \mu_T(u) \in D_{\mathbb{R}^k}[0,1]$, while for a fixed $u \in [0,1]$ it is a random variable. Combined with the fact that the finite-dimensional sets form a determining class of $D_{\mathbb{R}^k}[0,1]$, this implies that $\mu_{[Tu]}$ is a random element on $D_{\mathbb{R}^k}[0,1]$,[1] p.57. It generates a sequence of probability measures on $(D_{\mathbb{R}^k}[0,1], \mathcal{D}^k)$ through $P_T(B) = \mathbf{P}(\mu_T(u) \in B)$ for every Borel set $B \in \mathcal{D}^k$. In the next theorem weak convergence of $\mu_T(u)$ means weak convergence of this sequence in the space of all probability measures on $\left(D_{\mathbb{R}^k}[0,1], \mathcal{D}^k\right)$.

Theorem 3 *Let π_{Tt} and η_t satisfy for every fixed T the assumptions of Definition 1 with $\pi_{Tt}^{(i)} \sim B(0, p_T^{(i)}), i = 1, ..., k$. Then the conditions $Tp_T^{(i)} \to \varkappa_i, i = 1, ..., k$, imply*

$$\mu_T(u) \xrightarrow{w} J(u),$$

where the components $J^{(i)}(u), i = 1, ..., k$, are jointly independent compound Poisson processes $J(u)^{(i)} = \sum_{j=1}^{N^{(i)}(u)} \eta_j^{(i)}$, $N^{(i)}(u)$ are Poisson process with jump intensity $\varkappa_i$, and $N^{(i)}(u)$ are independent of $\eta_\cdot^{(i)}$ for all i.

The independence of the components of $J(u)$ is due to the fact that as $T \to \infty$, the probability that two different components of $\mu_T(u)$ jump simultaneously tends to zero, and so the contemporaneous dependence between the components of η_t becomes irrelevant asymptotically.

The statement of the theorem is a functional analogue of Poisson convergence. The result is intuitive: $\sum_{t=1}^{T} \delta_{Tt}$ need not be normalized to achieve convergence; it is only compressed along the horizontal axis to obtain $\mu_T(u)$. The very mechanics of the transformation show that if the former process has only a few jumps (on average), so does the latter.

2.3 Corollaries to functional Poisson convergence

Corollaries needed for the analysis of linear models are derived from the statement of weak convergence to a compound Poisson process. These are based on the following extension of Theorem 3:

Claim 4 *Let $\varepsilon_t \sim iid(0, \Sigma_\varepsilon)$ be a $p-$variate random sequence independent of π_{Ts} and η_s for all t, s. Then*

$$A \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} \varepsilon_t \\ \mu_T(u) \end{pmatrix} \xrightarrow{w} A \begin{pmatrix} W(u) \\ J(u) \end{pmatrix}$$

holds in the topology of $D_{\mathbb{R}^n}[0,1]$, where A is an arbitrary $n \times (p+k)$ matrix and $W(u)$ is a Wiener process with covariance matrix Σ_ε.

In the case $p = 0$ this claim will be referred to as the functional Poisson convergence theorem (FPCT).

Work in the space $D_{\mathbb{R}^k}[0,1]$ is sometimes rendered difficult by the fact that it is not a topological vector space (i.e. $x_n \to x$ and $y_n \to y$ in the Skorohod topology do not imply $x_n + y_n \to x + y$). The formulation of the claim for an arbitrary A shows that this difficulty does not apply to the processes under study here.

A corollary of claim 4 for processes with linear structure is given next.

The notation $vecA = A^v$ is used for the vector obtained by stacking the columns of a matrix A under one another. The integral $\int_0^u X(s-)dY(s)$ is denoted by $\int_0^u XdY$. When X is continuous, for example $X = W$, this convention is not important, since $\int_0^u X(s-)dY(s) = \int_0^u X(s)dY(s)$. However, if $X = J$, then $\int_0^u X(s)dX(s)$ is not well-defined in the sense of Stieltjes (the integrand and the integrator are both left discontinuous at the points of jump), while $\int_0^u X(s-)dX(s)$ is.

The quadratic variation process $X(u)Y(u)' - \int_0^u XdY' - \int_0^u (dX)Y'$ is denoted by $[X,Y]_u$. If X is purely discontinuous, it holds that $[X,Y]_u = \sum_{s\leq u} \Delta X_s \Delta Y_s'$.

Finally, let $J^c(u) = \int_0^u J(s)ds$, $\int XdY = \int_0^1 X(s)\, dY(s)$, and $\int X = \int_0^1 X(s)\, ds$.

Corollary 5 *Let $\theta(z) = \sum_{i=0}^\infty \theta_i z^i$, $\phi(z) = \sum_{i=0}^\infty \phi_i z^i$ and $\tau(z) = \sum_{i=0}^\infty \tau_i z^i$ be convergent for $|z| < 1+\delta$ with some $\delta > 0$. For $t \leq T$, let $\nu_{Tt} = \sum_{j=0}^t \theta_j \mu_{Tt-j}$, $\gamma_{Tt} = \sum_{j=0}^t \phi_j \mu_{Tt-j}$, $\lambda_{Tt} = \Delta\nu_{Tt} = \sum_{j=0}^t \theta_j \delta_{Tt-j}$, $\omega_{Tt} = \Delta\gamma_{Tt} = \sum_{j=1}^t \phi_j \delta_{Tt-j}$, and $\psi_t = \sum_{i=0}^\infty \tau_i \varepsilon_{t-i}$. Then*

$$\begin{pmatrix} \left(\nu'_{T[Tu]}, \gamma'_{T[Tu]}\right)' \\ T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} \psi_t \end{pmatrix} \xrightarrow{w} \begin{pmatrix} \left(J(u)'\theta'(1), J(u)'\phi(1)'\right)' \\ \tau(1)W(u) \end{pmatrix} \tag{1}$$

in the Skorohod topology of $D_{\mathbb{R}^{2k+p}}[0,1]$, and the following converge jointly:

a. $vec\left(\sum_{t=1}^T \lambda_{Tt}\omega'_{Tt}\right) \xrightarrow{w} \left(\sum_{i=0}^\infty \phi_i \otimes \theta_i\right) vec\,[J,J]_1$;

b. $\sum_{t=1}^T \nu_{Tt}\omega'_{Tt} \xrightarrow{w} \theta(1)J(1)J(1)'\phi(1)' - \theta(1)\int (dJ)J'\phi(1)' - \sum_{i=0}^\infty \sum_{j=i+1}^\infty \theta_j\,[J,J]_1\,\phi_i'$
$= \theta(1)\int J(dJ)'\,\phi(1)' + \sum_{i=0}^\infty \sum_{j=0}^i \theta_j\,[J,J]_1\,\phi_i'$;

c. $T^{-\frac{1}{2}} \sum_{t=1}^T \left(\sum_{i=1}^{t-1} \lambda_{Ti}\right)\varepsilon_t' = T^{-\frac{1}{2}} \sum_{t=1}^T \nu_{Tt-1}\varepsilon_t' \xrightarrow{w} \theta(1)\int J(dW)'$;

d. $T^{-\frac{1}{2}} \sum_{t=1}^T \left(\sum_{i=1}^{t-1} \psi_i\right)\lambda'_{Tt} \xrightarrow{w} \tau(1)\int W(dJ)'\theta(1)'$;

e. $T^{-\frac{3}{2}} \sum_{t=1}^T \left(\sum_{i=1}^t \lambda_{Ti}\right)\left(\sum_{i=1}^t \psi_i\right)' = T^{-\frac{3}{2}} \sum_{t=1}^T \nu_{Tt}\left(\sum_{i=1}^t \psi_i\right)' \xrightarrow{w} \theta(1)\int JW'\tau(1)'$;

f. $T^{-1} \sum_{t=1}^T \nu_{Tt}\gamma'_{Tt} \xrightarrow{w} \theta(1)\int_0^1 JJ'\phi(1)'$;

g. $T^{-\frac{1}{2}} \sum_{t=1}^T \nu_{Tt}\psi_t' \xrightarrow{w} \theta(1)\int_0^1 J(dW)'\tau(1)'$;

h. $T^{-2} \sum_{t=1}^T \left(\sum_{i=1}^t \nu_{Ti}\right)\gamma'_{Tt} \xrightarrow{w} \theta(1)\int_0^1 J^c J'\phi(1)'$;

i. $T^{-3} \sum_{t=1}^T \left(\sum_{i=1}^t \nu_{Ti}\right)\left(\sum_{i=1}^t \gamma_{Ti}\right)' \xrightarrow{w} \theta(1)\int_0^1 J^c J^{c\prime}\phi(1)'$;

j. $T^{-\frac{3}{2}} \sum_{t=1}^T \left(\sum_{i=1}^{t-1} \nu_{Ti}\right)\psi_t' \xrightarrow{w} \theta(1)\int_0^1 J^c(dW)'\tau(1)' = \theta(1)\left[J^c(1)W(1)' - \int_0^1 JW'\right]\tau(1)'$.

From now on the index T to $\delta, \lambda, \nu, \omega$ and γ will be subsumed.

2.4 Convergence of some conditional measures

Here the weak convergence of $\mu_T(u)$ is considered conditionally on the location and the number of its jumps. Two types of conditions are introduced: conditions with zero limiting probability, indicating the precise location of some jumps, and conditions with non-vanishing probability, restricting the number of jumps.

In section 3.1 convergence conditional on the known location of jumps will allow asymptotic results derived for deterministic rare jumps (e.g. [8]) to be obtained as special cases of the specification with stochastic rare jumps.

Restricting the number of jumps is sometimes necessary for technical reasons. Consider for ease of exposition the univariate case. There are statistics depending on both ε and δ, which have different limiting distributions when $N_T(1) = \sum_{t=1}^{T} \pi_t$ is zero and when it is positive. Different distributions obtain if, given $N_T(1) > 0$, the jump components dominate the $\varepsilon-$components and thus determine the limiting distribution, while given $N_T(1) = 0$, the limiting distribution is determined by the ε-components.

As an example, consider the statistic

$$T\frac{\sum_{t=1}^{T} \varepsilon_t \sum_{i=1}^{t-1} \varepsilon_i + \sum_{t=1}^{T} \mu_t \sum_{i=1}^{t-1} \mu_i}{\sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \varepsilon_i\right)^2 + \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \mu_i\right)^2}.$$

The numerator normalized by T^2 converges to $\int JJ^c$, and the denominator normalized by T^3, to $\int J^2$. However, the statistic itself does not converge to $\frac{\int JJ^c}{\int J^2}$, since the probability that this expression is not well-defined is positive ($P(N(1) = 0) > 0$). Meaningful limit results are obtained by conditioning. Conditionally on $N_T(1) = 0$, the statistic converges to $\frac{\int W dW}{\int W^2}$, while conditionally on $N_T(1) > 0$, the limiting distribution is that of $\frac{\int J_+^c J_+}{\int (J_+^c)^2}$. Here $J_+(u)$ denotes a compound Poisson process conditioned on a positive number of jumps.[1] The unconditional limiting distribution can be expected to be (and is) that of $\frac{\int W dW}{\int W^2} I\{N(1) = 0\} + \frac{\int J_+^c J_+}{\int (J_+^c)^2} I\{N(1) > 0\}$.

The asymptotic results from the previous subsection remain valid after conditioning if $J(u)$ is replaced by $J_+(u)$. This is a corollary of the next claim (formulated for the univariate case).

Claim 6 *For $u \in [0,1]$ and $k = 1$, denote the process $\sum_{t=1}^{[Tu]} \pi_{Tt}$ by $N_T(u)$. Let $0 \le s_1 < s_2 < ... < s_k \le 1$ and $0 < u_1 < u_2 < ... < u_m < 1$ be given points. For a counting process $n(u)$ defined on $[0,1]$ consider the condition $E(n)$, given by $n(s_i) - n(s_{i-1}) = l_i \in \mathbb{N}$, $i = 2, ..., k$, and the condition $C(n)$ that jumps occur at relative times u_i, $i = 1, ..., m$, and possibly at other dates. Then:*

a. $\mu_T(u) | E(N_T) \xrightarrow{w} J(u) | E(N)$,

b. $\mu_T(u) | C(N_T) \xrightarrow{w} J(u) | C(N)$,

c. $\mu_T(u) | E(N_T) \& C(N_T) \xrightarrow{w} J(u) | E(N) \& C(N)$, *provided that the two conditions E and C are consistent with one another, i.e. between any s_{i-1} and s_i there are no more than l_i points of the u_j-s.*

Condition E restricts the number of jumps between certain dates. One example is $k = 2$, $s_1 = 0$ and $s_2 = 1$, when a condition on the total number of jumps in the interval $[0,1]$

[1] The conditioned process J_+ has a more complicated structure than J, since its increments are not independent.

obtains. Another example is an arbitrary k and a condition containing $n(s_i) - n(s_{i-1}) = 0$ for some i, specifying the absence of jumps between two relative dates.

Condition C fixes the location of some but not all jumps. When all jump dates are known, this can be formulated as a special case of $E\&C$.

Corollary 7 $\mu_T(u) \,|\, \{N_T(1) > 0\} \overset{w}{\to} J(u) \,|\, \{N(1) > 0\} =: J_+(u)$.

When a data set is a realization for which $N_T(1) = 0$, it is indistinguishable from a data set generated by a model without jump components, and should be analyzed as if it were generated by such a model. Therefore the assumption $N_T(1) > 0$ is the natural one to justify the analysis of jumps and will be made with no loss of generality from now on.

Corollary 5 remains valid for the conditional measures, since the same proofs apply.

3 Examples for univariate processes with AR(1) and jump components

This section provides examples of applications of the asymptotic theory derived above. Two of the examples concern issues related to the size and the power of the Dickey-Fuller (DF) test. Although these have already been studied in the literature, Perron [8] and FH [4], different methods were used. The goal here is to show how the existing results fit in the framework of stochastic asymptotically rare outliers. The third example aims to show that this framework is a natural bridge between discrete and continuous time.

The processes considered in the first two examples are of the class

$$y_t = \rho^t y_0 + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \text{ jump component}, \quad t = 1, ..., T,$$

with $\rho \in (-1, 1]$ and $\varepsilon_t \sim iid\left(0, \sigma^2\right)$. The jump component is specified either as $\sum_{i=0}^{t-1} \tau_i \mu_{t-i}$, to represent permanent shifts in the level of y_t, or as $\sum_{i=0}^{t-i} \tau_i \delta_{t-i}$, to represent outlying transitory effects on y_t. The series $\tau(z) = \sum_{i=0}^{\infty} \tau_i z^i$ satisfies the assumption of corollary 5.

A distinction is often made between additive and innovational jump components. Additive components correspond to $\tau_0 = 1, \tau_i = 0, i \geq 1$ and affect y_t instantaneously, with no transition period, while innovational ones obtain for $\tau_i = \rho^i$ and their effect on y_t follows the same dynamics as the effect of the innovations ε_t. For asymptotic arguments, however, the different transition is of little relevance. What may matter more is that an innovational outlier (level shift) for $\rho < 1$ becomes, as ρ approaches 1, an additive level shift (trend break). To the contrary, the order of integratedness of additive components does not change with ρ since they remain outliers or level shifts independently of whether ρ is 1 or below 1. In the following examples ρ will be fixed and innovational components will not be allowed to mutate, i.e. this aspect is also irrelevant. Thus the results for the additive and the innovational specification will only differ by the values of τ_i, which parameterize them.

Following Doornik et al. [3], a condition for a jump component not to influence the null asymptotic distribution of the Dickey-Fuller (DF) statistic is that $T^{-\frac{1}{2}} \sum_{t=1}^{T} component_t \to 0$. This condition is fulfilled by temporary change components since $\sum_{t=1}^{T} \sum_{i=0}^{t-i} \tau_i \delta_{t-i} = O_P(1)$ by (1) in corollary 5. Nevertheless, temporary change components affect finite-sample distributions when $\sum_{t=1}^{T} component_t$ is big relative to $\sqrt{T}$. In order to proxy this influence in the limit, I will consider jump components specified as $\sqrt{T} \sum_{i=0}^{t-i} \tau_i \delta_{t-i}$ and will call them asymptotically big. The alternative specification $\sum_{i=0}^{t-i} \tau_i \delta_{t-i}$ will be referred to as asymptotically small.

3.1 Example 1. Power of the DF test applied to a stable process with level shifts

Perron [8] shows that for an AR(1) process with a single level shift the OLS estimate of the autoregressive coefficient asymptotically overestimates the true parameter if the level shift is not accounted for, and thus induces a loss in the power of certain unit root tests. Here I address the same issue for an AR(1) process with stochastic level shifts and compare the results. The specification is the following:

$$y_t = \rho^t y_0 + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \sum_{i=0}^{t-1} \tau_i \mu_{t-i}.$$

To test the hypothesis $\rho = 1$ against $\rho < 1$, the autoregression $y_t = c + \rho y_{t-1} + u_t$ can be estimated by OLS ignoring the presence of jumps, and a test based on $T(\hat{\rho} - 1)$ can be conducted.

Product moments converge as follows:

$T^{-1} \sum y_{t-1}^2 \xrightarrow{w} \sigma^2 \left(1 - \rho^2\right)^{-1} + \tau(1)^2 \int J^2$ by applying a LLN together with corollary 5 (f), (g);

$T^{-1} \sum y_t y_{t-1} \xrightarrow{w} \rho\sigma^2 \left(1 - \rho^2\right)^{-1} + \tau(1)^2 \int J^2$ by applying the same statements;

$T^{-1}\sum y_t \xrightarrow{w} \tau(1)\int J$ by a LLN and (1) in corollary 5, combined with the continuous mapping theorem.

Thus

$$\hat{\rho} = \frac{T\sum y_t y_{t-1} - \sum y_t \sum y_{t-1}}{T\sum y_{t-1}^2 - \left(\sum y_{t-1}\right)^2} \xrightarrow{w} \frac{\frac{\rho\sigma^2}{1-\rho^2} + \tau(1)^2\left[\int J^2 - \left(\int J\right)^2\right]}{\frac{\sigma^2}{1-\rho^2} + \tau(1)^2\left[\int J^2 - \left(\int J\right)^2\right]}. \tag{2}$$

This limit is smaller than 1, and therefore a test based on $T(\hat{\rho}-1)$ is consistent. However, if $\int J^2 - \left(\int J\right)^2 > 0$, the limit is also greater than ρ and, in finite samples, power is likely to be lost when ρ is overestimated.

These are qualitatively the same conclusions as those in [8]. Now I proceed by showing that Perron's quantitative conclusions follow from above too.

First, in his model the level shift is additive and hence $\tau(1) = 1$. Second, he assumes a single level shift of size γ with relative position λ in the sample. By claim 6, the limit of $\hat{\rho}$ under this condition can be obtained by conditioning in the limit (2). This yields

$$\int J^2 - \left(\int J\right)^2 = \int_\lambda^1 \gamma^2 - \left(\int_\lambda^1 \gamma\right)^2 = (1-\lambda)\gamma^2 - \left[(1-\lambda)\gamma\right]^2 = \lambda(1-\lambda)\gamma^2.$$

Finally, he denotes $\frac{\rho\sigma^2}{1-\rho^2}$ by ρ_1, and $\frac{\sigma^2}{1-\rho^2}$ by σ_e^2, so that the limit of $\hat{\rho}$ becomes $\frac{\rho_1+\lambda(1-\lambda)\gamma^2}{\sigma_e^2+\lambda(1-\lambda)\gamma^2}$. This is exactly Perron's expression.

3.2 Example 2. Size of the DF test applied to an $I(1)$ process with temporary change outliers

FH demonstrate that the presence of temporary change outliers, "provided that they are sufficiently large or sufficiently frequent", may lead to overrejection of the correct unit root null. In both cases the danger is not that the outliers will not be found, but that they will be mismodelled as regular mean reverting observations. As already discussed, the outliers of FH are asymptotically frequent in the sense that their average number increases at the same rate as the sample size. Here I will show that asymptotically rare outliers, i.e. with expected number independent of the sample size, have the same effect as the frequent ones, provided that their size is specified as a fraction of $\sqrt{T}$.

The variable y_t is assumed to follow

$$y_t = y_0 + \sum_{i=0}^{t-1}\varepsilon_t + \sqrt{T}\sum_{i=0}^{t-1}\tau_i\delta_{t-i},$$

and again the autoregression $y_t = \rho y_{t-1} + u_t$ is estimated and a DF unit root test based on the t-statistics of $\hat{\rho}$ is conducted. It holds $T^{-2}\sum y_{t-1}^2 \overset{w}{\to} \int W^2$ and $T^{-1}\sum y_{t-1}\Delta y_t \overset{w}{\to} \int WdW + \sum_{i=0}^{\infty}\tau_i\Delta\tau_{i+1}[J,J]_1$ (by corollary 5 (a) among others), so that

$$T(\hat{\rho}-1) \overset{w}{\to} \frac{\int WdW}{\int W^2} + \sum_{i=0}^{\infty}\tau_i\Delta\tau_{i+1}\frac{[J,J]_1}{\int W^2}.$$

Hence $\hat{\rho}$ is still superconsistent but the distribution of $T(\hat{\rho}-1)$ is shifted due to the jumps and a unit root test based on the percentiles of $\int WdW\left[\int W^2\right]^{-1}$ will asymptotically have the wrong size. The direction of the size distortion (over or underrejection) will depend on the sign of $\sum_{i=0}^{\infty}\tau_i\Delta\tau_{i+1}$, which is seen to be negative:

$$\sum_{i=0}^{\infty}\tau_i\tau_{i+1} - \sum_{i=0}^{\infty}\tau_i^2 \leq \left(\sum_{i=0}^{\infty}\tau_i^2\right)^{\frac{1}{2}}\left(\sum_{i=0}^{\infty}\tau_{i+1}^2\right)^{\frac{1}{2}} - \sum_{i=0}^{\infty}\tau_i^2 < 0$$

for $\tau_0 \neq 0$. Therefore, the unit root test will overreject, as in the case of additive outliers analyzed by FH. In that case the limit distribution of $T(\hat{\rho}-1)$ specializes to $\left\{\int WdW - [J,J]_1\right\}\left\{\int W^2\right\}^{-1}$, which, compared to the distribution of FH, $\left\{\int WdW - E[J,J]_1\right\}\left\{\int W^2\right\}^{-1}$, has higher variance.

Since

$$\hat{\sigma}^2 = T^{-1}\sum\left(\Delta y_t - (\hat{\rho}-1)y_{t-1}\right)^2 = T^{-1}\sum\Delta y_t^2 + o_p(1) \overset{w}{\to} \sigma^2 + \left\{\tau_0^2 + \sum_{i=1}^{\infty}\Delta\tau_i^2\right\}[J,J]_1,$$

it follows that

$$t_{\hat{\rho}} = (\hat{\rho}-1)\frac{\left(\sum y_{t-1}^2\right)^{\frac{1}{2}}}{\hat{\sigma}} \overset{w}{\to} \frac{\int WdW + \sum_{i=0}^{\infty}\tau_i\Delta\tau_{i+1}[J,J]_1}{\left(\int W^2\right)^{\frac{1}{2}}\left(\sigma^2 + \left\{\tau_0^2 + \sum_{i=1}^{\infty}\Delta\tau_i^2\right\}[J,J]_1\right)^{\frac{1}{2}}}.$$

Again, a comparison with the quantiles of $\int WdW\left[\sigma^2\int W^2\right]^{-\frac{1}{2}}$ will lead to overrejection of the unit root null. In the additive outlier case the difference from the limit distribution of FH is, as before, that in their result $[J,J]_1$ is replaced by its expectation.

3.3 Example 3. Local-to-unity asymptotics

The final example is of a continuous-time jump-diffusion that occurs as the weak limit of an AR(1) process with local-to-unity autoregressive root and asymptotically big innovational outliers. The specification is

$$\Delta y_t = \left(1 - \frac{\beta}{T}\right)y_{t-1} + \varepsilon_t + \sqrt{T}\delta_{Tt},\ \ t = 1, ..., T,$$

and y_0 is assumed fixed. Then

$$\begin{aligned} T^{-\frac{1}{2}} y_{[Tu]} &= \sum_{i=1}^{[Tu]} \left(1-\frac{\beta}{T}\right)^{[Tu]-i} \left[\frac{\varepsilon_i}{\sqrt{T}} + \delta_{Ti}\right] + T^{-\frac{1}{2}} \left(1-\frac{\beta}{T}\right)^{[Tu]} y_0 \\ &= \int_0^u \left(1-\frac{\beta}{T}\right)^{[Tu]-[Ts-]-1} d\sum_{t=1}^{[Ts]} \left[\frac{\varepsilon_i}{\sqrt{T}} + \delta_{Ti}\right] + o(1). \end{aligned}$$

Since $\left(1-\frac{\beta}{T}\right)^{[Tu]-[Ts-]-1} = e^{-\beta(u-s)} + o(1)$ uniformly in $u, s \in [0,1]$, it follows that

$$T^{-\frac{1}{2}} y_{[Tu]} = \int_0^u e^{-\beta(u-s)} d\sum_{j=1}^{[Ts]} \left[\frac{\varepsilon_i}{\sqrt{T}} + \delta_{Ti}\right] + o_P(1).$$

Next, the integral $d\sum_{j=1}^{[Ts]} \frac{\varepsilon_i}{\sqrt{T}}$ converges weakly to an integral dW, as can be seen, for example, by partial integration, and the integral $d\sum_{j=1}^{[Ts]} \delta_{Ti}$ converges weakly to an integral dJ by Theorem 2.7 in [6] (see its conditions in the proof of corollary 10). Thus $T^{-\frac{1}{2}} y_{[Tu]} \xrightarrow{w} \int_0^u e^{-\beta(u-s)} d(W(s) + J(s))$, a jump-diffusion satisfying the SDE

$$dX = -\beta X dt + dW + dJ$$

with the initial condition $X(0) = 0$. It can be viewed as an Ornstein-Uhlenbeck process with innovational outliers.

4 Conclusions

An apparatus for the asymptotic analysis of econometric models with stochastically specified outliers and structural breaks was developed. The asymptotic distributions of statistics related to discrete-time processes with autoregressive and jump components turn out to be functionals of continuous-time jump-diffusions. The results are consistent with the deterministic formulation of outliers and structural breaks with fixed relative position in the sample.

5 Proofs and intermediate results

5.1 Proofs of the main functional limit statements

Let

$$\Lambda = \{\lambda : [0,1] \to [0,1] : \lambda(0) = 0, \lambda(1) = 1, \lambda \text{ strictly increasing and continuous}\},$$

and let

$$\|\lambda\| = \sup_{0 \le t < s \le 1} \left| \frac{\lambda(t) - \lambda(s)}{t - s} \right|.$$

The Skorohod metric in k dimensions on the space of cadlag functions is defined by

$$d_{B^k}(x, y) = \inf \left\{ \varepsilon > 0 : \exists \lambda \in \Lambda \text{ s.t. } \|\lambda\| \le \varepsilon, \max_i \sup_{t \in [0,1]} |x_i(\lambda(t)) - y_i(t)| \le \varepsilon \right\}. \tag{3}$$

The resulting metric space was denoted by $D_{\mathbb{R}^k}[0, 1]$.

A different metric on the space of cadlag functions is the product metric

$$d(x, y) = \max \{ d_B(x_i, y_i) : i = 1, ..., k \},$$

where the metric in the coordinate spaces is the 1-dimensional Skorohod metric defined by (3) with $k = 1$. The resulting metric space will be denoted by $D^k[0, 1]$; it is complete and separable, [1]. The topology on $D_{\mathbb{R}^k}[0, 1]$ is strictly finer than the topology on $D^k[0, 1]$, but the open balls w.r.t. d generate the same Borel sigma algebra $\mathcal{D}^k$ as d_{B^k} does on $D_{\mathbb{R}^k}[0, 1]$.

With E standing for either $D^k[0, 1]$ or $D_{\mathbb{R}^k}[0, 1]$, let $\mathcal{P}(E)$ be the space of all probability measures on $(E, \mathcal{D}^k)$. It is first proved that the convergence statement of Theorem 3 holds on $\mathcal{P}(D^k[0, 1])$, and then it is extended to $\mathcal{P}(D_{\mathbb{R}^k}[0, 1])$. The extension is necessary for obtaining convergence to certain stochastic integrals later on.

The proof that a sequence belonging to $\mathcal{P}(D^k[0, 1])$ converges weakly can be based on the verification of two properties.

The first one is *relative compactness*. A sequence $\{P_T\}$ of probability measures is relatively compact if from each subsequence it is possible to extract a weakly converging subsubsequence. If all converging subsubsequences have the same weak limit, then it is also the limit of the original sequence.

In order to demonstrate that this limit is some prespecified probability measure, the concept of *determining class* is used. A determining class is a family of Borel sets such that whenever two probability measures coincide on this family, they are necessarily the same measure. One determining class of $D^k[0, 1]$ is the family of finite-dimensional sets, which consists of all sets $\pi^{-1}_{t_1,...,t_n} H$, where $n \ge 1, t_1, ..., t_n \in [0, 1]$, $\pi_{t_1,...,t_n}$ are the natural projections from $D^k[0, 1]$ to $\mathbb{R}^{nk}$, and H are Borel sets of $\mathbb{R}^{nk}$, [2]. This means that two probability measures on $(D^k[0, 1], \mathcal{D}^k)$ whose finite-dimensional distributions coincide, coincide themselves. The space $D^k[0, 1]$ has smaller determining classes as well, and the determining class to be used in a proof is chosen such that its sets are generated by projections a.s. continuous w.r.t. the hypothesized limiting measure. Then verifying that the latter is a true limiting measure reduces to verifying that the finite-dimensional distributions $P_T\pi^{-1}_{t_1,\ldots,t_n}$, corresponding to the a.s. continuous projections, converge to its respective finite-dimensional distributions.

Prohorov's theorem states that on a complete and separable metric space relative compactness is equivalent to tightness: $\{P_T\}$ is *tight* if for each $\varepsilon > 0$ there exists a compact K such that $P_T(K) > 1-\varepsilon$ for all T. If $\{P_T\}$ is defined on a product space, it is tight if and only if the sequences of marginal probability measures on the coordinate spaces are tight, [1] p.41. Let a marginal probability measure satisfy $P^{(i)}_T(B) = \mathbf{P}_T(X_T \in B)$ for every Borel set $B \in \mathcal{D}$, where X_T is a random element defined on some probability space $(\Omega_T, \mathcal{B}_T, \mathbf{P}_T)$ and with values in $D[0,1]$. Let also the finite-dimensional distributions of $P^{(i)}_T$ converge to those of a process a.s. left continuous at 1. Then a sufficient condition for $P^{(i)}_T$ to be tight is that for some constant A it holds that

$$E_T\left\{(X_T(u)-X_T(u_1))^2(X_T(u_2)-X_T(u))^2\right\} \le A(u_2-u_1)^2, \quad u_1 \le u \le u_2, \tag{4}$$

[1] Th. 15.6.

Lemma 8 *The convergence statement in theorem 3 holds on* $\mathcal{P}\left(D^k\,[0,1]\right)$.

Proof. The scheme outlined above is followed. Both the compound Poisson and the Wiener processes are a.s. left continuous at 1. *Tightness* of the vector sequence reduces to tightness of the component sequences. It is ensured by condition (4) for the case $k=1$. *Convergence of the finite-dimensional distributions* follows since, first, for all $x \in \mathbb{R}^k$ and $u_1, u_2 \in [0,1]$, $x'\left(\mu_T\left(u_2\right)-\mu_T\left(u_1\right)\right) \stackrel{w}{\to} x'(J(u_2)-J(u_1))$, as can be shown by considering the characteristic function. Second, by the Cramer-Wold device, this implies $\mu_T\left(u_2\right)-\mu_T\left(u_1\right) \stackrel{w}{\to} J(u_2)-J(u_1)$ for every $u_1, u_2 \in [0,1]$. Finally, since the process considered has independent increments, convergence of all finite-dimensional distributions follows ([5], p.11, lemma 1.3). ■

The extension to $\mathcal{P}(D_{\mathbb{R}^k}\,[0,1])$ is prepared next.

Let $S_{\mathbb{R}^k}[0,1]$ be the set of all k-dimensional vectors whose components are step cadlag functions defined on $[0,1]$, and (i) each has no more than a finite number of jumps, (ii) none of them jumps at 1, and (iii) at each point in $(0,1)$ at most one of them has a jump. This set is interesting because of

Remark 9 *a.* $P(J \in S_{\mathbb{R}^k}[0,1]) = 1$, *i.e.* $S_{\mathbb{R}^k}[0,1]$ *supports the probability measure generated by* J*;*

b. If $f_n \in D^k[0,1]$ *and* $f_n \to f \in S_{\mathbb{R}^k}[0,1]$ *in the product Skorohod topology, then for each* $u \in (0,1]$ *there exists a sequence* u_n *such that* $\Delta f_n(u_n) \to \Delta f(u)$*. Here* $\Delta f(u) := f(u) - f(u-)$*.*

The first part follows from the independence of the Poisson processes governing the components of $J(u)$.

As to the second part, if only the component $f^{(j)}$ is discontinuous at u, set $i = j$; if all components are continuous at u, set $i = 1$. By the definition of $S_{\mathbb{R}^k}[0,1]$ there are no other possibilities. By proposition 2.1 in [5], Vol. 1, Chapter 6, the convergence $f_n^{(i)} \to f^{(i)}$ in $D[0,1]$ implies the existence of a sequence u_n such that $\Delta f_n^{(i)}(u_n) \to \Delta f^{(i)}(u)$. However the same convergence also holds for the other components of f, since they are constant in a neighborhood of u and convergence to them is uniform on that neighborhood. Thus the remark is correct.

This remark is the basis of the next proof.

Proof of theorem 3. Thanks to the continuous mapping theorem (Theorem 5.1 in [1]) and to remark 9 (a), the convergence in lemma 8 can be extended from $\mathcal{P}\left(D^k[0,1]\right)$ to $\mathcal{P}(D_{\mathbb{R}^k}[0,1])$ if the identity $id : D^k[0,1] \to D_{\mathbb{R}^k}[0,1]$ is continuous on $S_{\mathbb{R}^k}[0,1]$. This means that the convergence $f_n \to f \in S_{\mathbb{R}^k}[0,1]$ in the product topology should imply the same convergence in the topology of $D_{\mathbb{R}^k}[0,1]$. Indeed, remark 9 (b) ensures that the conditions of proposition 2.2b in [5], Vol. 1, Chapter 6, are satisfied, and so the desired implication is true. ■

Proof of claim 4. By Donsker's invariance principle, $T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} \varepsilon_t \xrightarrow{w} W(u)$ on $D_{\mathbb{R}^p}[0,1]$. Since $D_{\mathbb{R}^{p+k}}[0,1]$ is separable, $\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} \varepsilon_t',\ \mu_T(u)'\right)'$ and $(W(u)',\ J(u)')'$ are random elements of $D_{\mathbb{R}^{p+k}}[0,1]$, [1] p.225. Due to the assumed independence, the convergences stated by Theorem 3 and by Donsker's invariance principle are joint, [1] p.26, i.e. $\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} \varepsilon_t',\ \mu_T(u)'\right)' \xrightarrow{w}$ $(W(u)',\ J(u)')'$ in the topology of $D_{\mathbb{R}^{p+k}}[0,1]$. This proves the claim for $A = I$.

For an arbitrary A, the functional $(x \to Ax)$: $S_{\mathbb{R}^{p+k}}[0,1] \to D_{\mathbb{R}^n}[0,1]$ is continuous by remark 9 (b) and proposition 2.2b in [5], Vol. 1, Chapter 6, implying the validity of the claim. ■

5.2 Corollaries to functional Poisson convergence

A special case of corollary 5 is proved first.

Corollary 10 *Under the assumptions of the FPCT, the following converge jointly:*

a. $\sum_{t=1}^{T} \mu_{Tt-1}\delta'_{Tt} \xrightarrow{w} \int J(dJ)'$;

b. $T^{-\frac{1}{2}} \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \varepsilon_i\right) \delta'_{Tt} \xrightarrow{w} \int W(dJ)'$;

c. $\sum_{t=1}^{[Tu]} \delta_{Tt}\delta'_{Tt} \xrightarrow{w} [J,J]_u \overset{a.s.}{=} \frac{1}{2} diag(J_1^2(u) - \int_0^u J_1 dJ_1, ..., J_k^2(u) - \int_0^u J_k dJ_k)$;

d. $T^{-1} \sum_{t=1}^{T} \mu_{Tt}\mu'_{Tt} \xrightarrow{w} \int JJ'$;

e. $T^{-2} \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \mu_{Ti}\right) \mu'_{Tt} \xrightarrow{w} \int J^c J'$;

f. $T^{-3} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} \mu_{Ti}\right) \left(\sum_{i=1}^{t} \mu_{Ti}\right)' \xrightarrow{w} \int J^c J^{c\prime}$;

g. $T^{-\frac{1}{2}} \sum_{t=1}^{T} \mu_{Tt}\varepsilon'_t \xrightarrow{w} J(1)W(1)' - \int (dJ) W' = \int J(dW)'$;

h. $T^{-\frac{3}{2}} \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \varepsilon_i\right) \mu'_{Tt} \xrightarrow{w} \int WJ'$;

i. $T^{-\frac{3}{2}} \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \mu_{Ti}\right) \varepsilon'_t \xrightarrow{w} J^c(1)W(1)' - \int JW' = \int J^c(dW)'$.

Proof of corollary 10. Theorem 2.7 in [6] provides a sufficient condition for convergence of integrals whose integrators are constant except for finitely many discontinuities (here $\mu_T(u)$). First, it requires that the sequence $\sum_{t=1}^{T} \pi_t$, counting the jumps of μ_T, be stochastically bounded. Since this sequence converges in distribution, the requirement is fulfilled. The second condition is convergence in the Skorohod topology, and in the present context it reduces to $\left(T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} \varepsilon'_t, \mu_T(u)', \mu_T(u)', \mu_T(u)'\right)' \xrightarrow{w} (W(u)', J(u)', J(u)', J(u)')'$ in the topology of $D([0,1]^{p+3k})$. It is implied by claim 4 and the continuity of $(x,y) \to (x,y,y,y)$ from $D([0,1]^{p+k})$ to $D([0,1]^{p+3k})$. Therefore it holds that

$$\int \left(T^{-\frac{1}{2}} \sum_{t=1}^{[Tu-]} \varepsilon_t, \mu_T(u-)\right) d\mu'_T(u) \xrightarrow{w} \int (W,J)\, dJ',$$

which contains (a) and (b).

(c) follows from

$$\sum_{t=1}^{[Tu]} (\pi_{Tt} \cdot \varepsilon_t)(\pi_{Tt} \cdot \varepsilon_t)' = [\mu_T, \mu_T]_u = \mu_T(u)\mu_T(u)' - 2\int_0^u \mu_T d\mu'_T \xrightarrow{w} J(u)J(u)' - 2\int_0^u J dJ' = [J,J]_u$$

and the latter is a.s. a diagonal matrix since for $i \neq j$, J_i and J_j a.s. do not have points of jump in common.

(d), (e) and (f) follow respectively from the continuity of the functionals $x \to \int_0^1 x(u)x(u)'du$, $x \to \int_0^1 \int_0^u x(s)ds\, x(u)'du$ and $x \to \int_0^1 \int_0^u x(s)ds \left[\int_0^u x(s)ds\right]' du$, all from $D_{\mathbb{R}^k}[0,1]$ to R^{k^2}, and (h) follows from the continuity of $(x,y) \to \int_0^1 x(u)y(u)'du$ from $D([0,1]^{2k})$ to R^{k^2}.

(g) follows from the partial summation

$$\sum_{t=1}^{T} \mu_t \varepsilon'_t = \mu_T \sum_{t=1}^{T} \varepsilon'_t - \sum_{t=1}^{T} \delta_t \sum_{i=1}^{t-1} \varepsilon'_i,$$

and from (b). Similarly, the partial summation

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\mu_i\right)\varepsilon_t' = \sum_{t=1}^{T}\mu_t\sum_{t=1}^{T}\varepsilon_t' - \sum_{t=1}^{T}\mu_t\sum_{i=1}^{t}\varepsilon_i'$$

and the continuity of the functional $(x,y) \to \int_0^1 x(u)du\, y(1)' - \int_0^1 x(u)y(u)'du$ imply (i).

Convergence is joint because the vector functional with components the individual ones, is continuous. ■

The proof of corollary 5 makes use of some preliminary results.

Remark 11 (on representation) *For $\theta(z)$ defined in corollary 5, let $\theta_i^* = \sum_{j=i+1}^{\infty}\theta_j,\ i = 0,1,...$ Let also α_{Tt}, $t \leq T$, be a generic notation for an element of a triangular array. Then*

$$\sum_{t=1}^{T}\sum_{i=0}^{t-1}\theta_i\alpha_{Tt-i} = \theta(1)\sum_{t=1}^{T}\alpha_{Tt} - \sum_{i=0}^{t-1}\theta_i^*\alpha_{Tt-i},$$

and therefore $\sum_{i=0}^{t-1}\theta_i^*\alpha_{Tt-i}$ has the same convergence properties as $\sum_{i=0}^{t-1}\theta_i\alpha_{Tt-i}$.

This follows from

$$\begin{aligned}\sum_{i=0}^{t-1}\theta_i\alpha_{Tt-i} &= \left(\sum_{i=0}^{t-1}\theta_i\right)\alpha_{Tt} - \sum_{i=0}^{t-1}\left(\sum_{j=i+1}^{t-1}\theta_j\right)\Delta\alpha_{Tt-i} \\ &= \left(\theta(1) - \theta_{t-1}^*\right)\alpha_{Tt} - \sum_{i=0}^{t-1}\left(\theta_i^* - \theta_{t-1}^*\right)\Delta\alpha_{Tt-i} = \theta(1)\alpha_{Tt} - \sum_{i=0}^{t-1}\theta_i^*\Delta\alpha_{Tt-i}\end{aligned}$$

by summation over t. Since $-\theta_i^*$ are the coefficients of the power series of the function $(\theta(z) - \theta(1))/(z-1)$, which satisfies the same convergence hypothesis as $\theta(z)$, $\sum_{i=0}^{t-1}\theta_i^*\alpha_{Tt-i}$ has the same convergence properties as $\sum_{i=0}^{t-1}\theta_i\alpha_{Tt-i}$.

Remark 12 *Under the assumptions of the FPCT it holds:*

a. if ξ_t are serially uncorrelated random vectors with common VCM $\Sigma_\xi < \infty$, $\sqrt{T}\sum_{i=0}^{t-1}\theta_i\pi_{Tt-i}\cdot\xi_{t-i}$ are L_2- bounded uniformly in T and $t \leq T$;

b[2]*. $\sum_{t=1}^{T}\lambda_{Tt}\psi_t'$ are L_2-bounded uniformly in T and $t \leq T$;*

c. $\sum_{t=1(2)}^{T}\sum_{0\leq i<j\leq t-1}\theta_i\delta_{Tt-i}\delta_{Tt-j}'\phi_j' = o_P(1)$.

[2] In fact, $\sum_{t=1}^{T}\lambda_t\eta_t'$ converges to a distribution. For example, if $n = 1$, $\psi(z) = 1$ and $\tau(z) = 1$, claim 1 implies that $\sum_{t=1}^{T}\pi_t(\varepsilon_{2t}\varepsilon_{1t}) \overset{w}{\to} \sum_{i=1}^{N(r)}\varsigma_i$, where $\varsigma_i \sim Nid(0,\sigma_1^2\sigma_2^2)$. A derivation for the general case requires a stronger statement than claim 1. For future reference it will be sufficient to have (b) as in the remark.

Proof. Since $\sum_{j=0}^{\infty} \|\theta_j\|^2 < \infty$ and $T \|Var(\pi_{T\cdot} \cdot \xi_{\cdot})\| \to const$ (due to $Tp_{1,T}^{(i)} \to \varkappa_i, i = 1, ..., k$), it holds that

$$\left\| Var\left(\sqrt{T}\sum_{i=0}^{t-1} \theta_i \pi_{Tt-i} \cdot \xi_{t-i}\right)\right\| \leq T\|Var(\pi_{T\cdot} \cdot \xi_{\cdot})\| \sum_{j=0}^{t} \|\theta_j\|^2 \to const,$$

which confirms (a).

(b) can be approached similarly. It holds that $\sum_{t=1}^{T} \lambda_{Tt}\psi_t' = \sum_{t=1}^{T}\sum_{i=0}^{T-t} \theta_i \delta_{Tt} \psi_{t+i}'$, and hence

$$vec\left(\sum_{t=1}^{T} \lambda_t \psi_t'\right) = \sum_{t=1}^{T}\left(\sum_{i=0}^{T-t} \psi_{t+i} \otimes \theta_i\right)\delta_{Tt},$$

as $vec(\delta_{Tt}) = \delta_{Tt}$. Denote $A_{Tt} = \sum_{i=0}^{T-t} \psi_{t+i} \otimes \theta_i$. Then, by the convergence assumptions on τ and θ, $E(A_{Tt} \otimes A_{Tt}) = O(1)$. By the serial independence of δ_{Tt},

$$Var\left\{\left(\sum_{t=1}^{T} \lambda_t \psi_t'\right)^v\right\} = \sum_{t=1}^{T} Var(A_{Tt}\delta_{Tt}),$$

and by the independence of δ_{Tt} and ε_s,

$$\left[Var\left\{\left(\sum_{t=1}^{T} \lambda_t \psi_t'\right)^v\right\}\right]^v = \left[\sum_{t=1}^{T} E(A_{Tt} \otimes A_{Tt})\right][Var\delta_T]^v = \left[\sum_{t=1}^{T} O(1)\right] O(T^{-1}) = O(1).$$

Statement (c) is proved using the definition of convergence in probability. Fix η, $\delta > 0$. Let T_0 be an integer to be fixed later but arbitrary at the moment, and let $T > T_0$. Let $\|.\|$ denote the max matrix norm. Then

$$P\left(\left\|\sum_{t=2}^{T}\sum_{0\leq i<j\leq t-1} \theta_i \delta_{Tt-i}\delta_{Tt-j}'\phi_j'\right\| > \eta\right) \leq$$

$$\underbrace{P\left(\left\|\sum_{t=2}^{T}\sum_{i=0}^{t-2}\sum_{j=i+1}^{T_0-1} \theta_i \delta_{Tt-i}\delta_{Tt-j}'\phi_j'\right\| > \frac{\eta}{2}\right)}_{p_1} + \underbrace{P\left(\left\|\sum_{t=2}^{T}\sum_{i=0}^{t-2}\sum_{j=(i+1)\vee T_0}^{t-1} \theta_i \delta_{Tt-i}\delta_{Tt-j}'\phi_j'\right\| > \frac{\eta}{2}\right)}_{p_2}.$$

In its turn,

$$p_1 \leq \sum_{t=2}^{T} P\left(\left\|\sum_{i=0}^{t-2}\sum_{j=i+1}^{T_0-1} \theta_i \delta_{Tt-i}\delta_{Tt-j}'\phi_j'\right\| \neq 0\right).$$

The upper limit for the sum in i is effectively $T_0 - 2$, since for bigger values of i the sum in j is empty. The realization of each of the events indexed by t implies that at least two different vectors among $\pi_{t-T_0+1}, ..., \pi_t$ are nonzero. Let $\Pi_{Tt} = \max_{i=1,...,k} \pi_{Tt}^{(i)}$; it holds $P(\Pi_{Tt} = 1) = P(\Pi_{T1} \neq 0) = O(T^{-1})$. Thus

$$p_1 \leq \sum_{t=2}^{T}\sum_{s=t-T_0+1}^{t-1}\sum_{u=s+1}^{t} P(\pi_{Ts} \neq 0 \;\&\; \pi_{Tu} \neq 0) = \frac{(T-1)T_0(T_0-1)}{2}[P(\Pi_{T1} \neq 0)]^2 = \frac{T_0(T_0-1)}{2} O(T^{-1}).$$

The other addend satisfies

$$p_2 \leq P\left(\sum_{t=2}^{T}\sum_{i=0}^{t-2}\|\theta_i\delta_{Tt-i}\|\sum_{j=(i+1)\vee T_0}^{t-1}\|\phi_j\eta_{t-j}\| > \frac{\eta}{2}\right) \leq P\left(\sum_{j=T_0}^{\infty}\|\phi_j\|\,\|\eta_{t-j}\|\sum_{t=1}^{T}\sum_{i=0}^{t-1}\|\theta_i\|\,\|\delta_{Tt-i}\| > \frac{\eta}{2}\right).$$

Since $\sum_{j=T_0}^{\infty}\|\phi_j\|\,\|\eta_{t-j}\|$ is the tail of $\sum_{j=0}^{\infty}\|\phi_j\|\,\|\eta_{t-j}\| < \infty$ a.s., it follows that $\sum_{j=T_0}^{\infty}\|\phi_j\|\,\|\eta_{t-j}\| = o_P(1)$ as $T_0 \to \infty$. Consider now the second factor in the parentheses. It holds by remark 11 that

$$\sum_{t=1}^{T}\sum_{i=0}^{t-1}\|\theta_i\|\,\|\delta_{Tt-i}\| = \left(\sum_{i=0}^{\infty}\|\theta_i\|\right)\sum_{t=1}^{T}\|\delta_{Tt}\| - \sum_{i=0}^{T-1}\left(\sum_{j=i+1}^{\infty}\|\theta_i\|\right)\|\delta_{TT-i}\|.$$

With $\Omega_t = \max_{i=1,\ldots,k}\left|\varepsilon_t^{(i)}\right|$,

$$\sum_{t=1}^{T}\|\delta_{Tt}\| \leq \sum_{t=1}^{T}\Pi_{Tt}\left(\Omega_t - E\Omega_1\right) + E\Omega_1\sum_{t=1}^{T}\Pi_{Tt} = O_P(1)$$

by the FPCT and by usual Poisson convergence, and

$$\sum_{i=0}^{T-1}\left(\sum_{j=i+1}^{\infty}\|\theta_i\|\right)\|\delta_{TT-i}\| \leq \sum_{i=0}^{T-1}\left(\sum_{j=i+1}^{\infty}\|\theta_i\|\right)\Pi_{Tt}\Omega_t = o_P(1)$$

by remark 12 (a). Hence $\sum_{t=1}^{T}\sum_{i=0}^{t-1}\|\theta_i\|\,\|\delta_{Tt-i}\| = O_P(1)$. Summarizing,

$$p_2 \leq P\left(o_P(1)\,O_P(1) > \frac{\eta}{2}\right) \text{ as } T_0 \to \infty.$$

Thus T_0 can be chosen such that $p_2 \leq \frac{\delta}{2}$ and independent of T. Choose and fix such a T_0.

Next $T_1 > T_0$ can be found such that $p_1 \leq \frac{T_0(T_0-1)}{2}O\left(T^{-1}\right) < \frac{\delta}{2}$ for $T > T_1$. Therefore, for $T > T_1$,

$$P\left(\left\|\sum_{t=2}^{T}\sum_{0\leq i<j\leq t-1}\theta_i\delta_{Tt-i}\delta_{Tt-j}'\phi_j'\right\| > \eta\right) \leq \frac{\delta}{2} + \frac{\delta}{2},$$

which completes the proof of (c). ■

Proof of corollary 5. Let $t \leq T$. Subsuming the index T, by remark 11 it holds that

$$\nu_t = \sum_{i=1}^{t}\lambda_i = \theta(1)\,\mu_t - \sum_{i=0}^{t-1}\theta_i^*\delta_{t-i} = \theta(1)\,\mu_t - \lambda_t^*,$$

where λ_t^* satisfies the same assumption as λ_t does, and is in particular $o_P(1)$ by remark 12 (a). By the FPCT and Th. 4.1 in [1], $\nu_{[Tu]} \xrightarrow{w} \theta(1)J(u)$. Analogously $\gamma_{[Tu]} \xrightarrow{w} \phi(1)J(u)$.

The result $T^{-\frac{1}{2}}\sum_{t=1}^{[Tu]}\psi_t \xrightarrow{w} \tau(1)W(u)$ is known from e.g. [6].

The three convergences are joint in $D([0,1]^{2k+p})$ similarly to claim 4.

To show (a), start from

$$\sum_{t=1}^{T}\lambda_t\omega_t' = \sum_{t=1}^{T}\left\{\sum_{i=0}^{t-1}\theta_i\delta_{t-i}\delta_{t-i}'\phi_i' + 2\sum_{0\le i<j\le t-1}\theta_i\delta_{t-i}\delta_{t-j}'\phi_j'\right\} = \sum_{t=1}^{T}\sum_{i=0}^{t-1}\theta_i\delta_{t-i}\delta_{t-i}'\phi_i' + o_P(1),$$

which follows from remark 12 (c).

Thus,

$$\begin{aligned} vec\left(\sum_{t=1}^{T}\lambda_t\omega_t'\right) &= \sum_{t=1}^{T}\sum_{i=0}^{t-1}(\phi_i\otimes\theta_i)\,vec\left(\delta_{T-i}\delta_{T-i}'\right) \\ &= \sum_{i=0}^{\infty}(\phi_i\otimes\theta_i)\,vec\left(\sum_{t=1}^{T}\delta_t\delta_t'\right) - \sum_{t=0}^{T-1}\left[\sum_{i=t+1}^{\infty}(\phi_i\otimes\theta_i)\right]vec\left(\delta_{T-t}\delta_{T-t}'\right) + o_P(1). \end{aligned}$$

From here (a) follows by corollary 10 (c) applied to the first addend, since the second one is $o_P(1)$. Indeed,

$$\begin{aligned} \left\|\sum_{t=0}^{T-1}\left[\sum_{i=t+1}^{\infty}(\phi_i\otimes\theta_i)\right]vec\left(\delta_{T-t}\delta_{T-t}'\right)\right\| &\le \sum_{t=0}^{T-1}\left(\sum_{i=t+1}^{\infty}\|\phi_i\|\,\|\theta_i\|\right)\|\delta_{T-t}\|^2 \\ &\le \sum_{t=0}^{T-1}\left(\sum_{i=t+1}^{\infty}\|\phi_i\|\,\|\theta_i\|\right)\Pi_{Tt}\Omega_t^2 \end{aligned}$$

and if η_t have finite fourth moments, convergence to 0 follows from remark 12 (a). More generally it follows from a consideration of the characteristic function.

The LHS of (b) can be written as $\sum_{t=1}^{T}\nu_t\omega_t' = \theta(1)\sum_{t=1}^{T}\mu_t\omega_t' - \sum_{t=1}^{T}\lambda_t^*\omega_t'$, and the asymptotic distribution of the two addends can be established separately.

Summing partially gives

$$\begin{aligned} \sum_{t=1}^{T}\mu_t\omega_t' = \mu_T\sum_{t=1}^{T}\omega_t' - \sum_{t=1}^{T}\delta_t\sum_{i=1}^{t-1}\omega_i' &= \mu_T\gamma_T - \sum_{t=1}^{T}\delta_t\gamma_{t-1} \\ &= \mu_T\mu_T'\phi(1)' - \mu_T\omega_T^{*\prime} - \sum_{t=1}^{T}\delta_t\mu_{t-1}'\phi(1)' + \sum_{t=1}^{T}\delta_t\omega_{t-1}^{*\prime}, \end{aligned}$$

where ω_t^* is defined analogously to λ_t^*. Then by the FPCT, remark 12 (a), corollary 10 (a) and corollary 5 (a) applied in this order to the respective members of the above expression, it follows that $\sum_{t=1}^{T}\mu_t\omega_t' \xrightarrow{w} \left[J(1)J(1)' - \int(dJ)J'\right]\phi(1)'$.

For $\sum_{t=1}^{T}\lambda_t^*\omega_t'$, corollary 5 (a) implies $vec\left(\sum_{t=1}^{T}\lambda_t^*\omega_t'\right) \xrightarrow{w} \sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}(\phi_i\otimes\theta_j)\,vec\,[J,J]_1$, which completes the proof of the first line in (b).

To obtain the expression in the second line, $J(1)J(1)' - \int(dJ)J'$ should be substituted by $[J,J]_1 + \int J\,(dJ)'$, and then $\theta(1)\,[J,J]_1\,\phi(1)' - \sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\theta_j\,[J,J]_1\,\phi_i'$ by $\sum_{i=0}^{\infty}\sum_{j=0}^{i}\theta_j\,[J,J]_1\,\phi_i'$.

Statement (c) follows from $T^{-\frac{1}{2}}\sum_{t=1}^{T}\nu_{t-1}\varepsilon_t' = T^{-\frac{1}{2}}\theta(1)\sum_{t=1}^{T}\mu_{t-1}\varepsilon_t' + T^{-\frac{1}{2}}\sum_{t=1}^{T}\lambda_t^*\varepsilon_t'$ by applying corollary 10 (g) to the first term and remark 12 (b) to the second one.

To prove (d), note that by remark 11

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\psi_i\right)\lambda_t' = \tau(1)\,T^{-\frac{1}{2}}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\lambda_t' - T^{-\frac{1}{2}}\sum_{t=1}^{T}\psi_{t-1}^*\lambda_t',$$

and by remark 12 (b) $T^{-\frac{1}{2}}\sum_{t=1}^{T}\psi_{t-1}^*\lambda_t' = o_P(1)$. Next, from the partial summation

$$\begin{aligned} T^{-\frac{1}{2}}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\lambda_t' &= T^{-\frac{1}{2}}\sum_{i=1}^{T}\varepsilon_t\left(\sum_{i=1}^{T}\lambda_t\right)' - T^{-\frac{1}{2}}\sum_{t=1}^{T}\varepsilon_t\left(\sum_{i=1}^{t}\lambda_i\right)' \\ &= T^{-\frac{1}{2}}\sum_{i=1}^{T}\varepsilon_t\mu_T'\theta(1)' + T^{-\frac{1}{2}}\sum_{i=1}^{T}\varepsilon_t\lambda_t^* - T^{-\frac{1}{2}}\sum_{t=1}^{T}\varepsilon_t\nu_t', \end{aligned}$$

(d) obtains by applying (1) to the first term, remark 12 (b) to the second term, and corollary 5 (c) to the last one.

Statements (e), (f), (g), (h), (i) and (j) follow from (1) and the continuity of functionals of obvious choice (where applicable, integrals $d\frac{\sum_{t=1}^{[Tu]}\varepsilon_t}{\sqrt{T}}$ after integration by parts are transformed into integrals $d\mu_T(u)$, whose convergence was discussed in the proof of the previous corollary).
■

Proof of claim 6. The notation $P(G\in(.))$ is used for the probability measure corresponding to a random element G. The sign $\overset{d}{=}$ between two random elements denotes that these generate the same probability measure. The same notation is applied to conditional measures.

a. **Convergence conditionally on a known number of jumps between certain dates.** The distribution of $\{n(s_i)-n(s_{i-1})\}_{i=1}^{k}$ is a distribution on the discrete set $\mathbb{N}^k$. Hence the sets $\{n(s_i)-n(s_{i-1})=l_i\}$, being simultaneously open and closed, have empty boundaries and are thus continuity sets of $P(n(s_i)-n(s_{i-1})\in(.))$. Let B be a continuity set of $P(J\in(.)\,|E(N))$, so that $P(J\in\partial B|E(N))=0$. Since $\partial\{l_i\}=\phi$, it holds that

$$\partial\left(B\times\prod_{i=1}^{k}\{l_i\}\right)\subset\partial B\times\prod_{i=1}^{k}\{l_i\}\cup B\times\prod_{i=1}^{k}\partial\{l_i\}=\partial B\times\prod_{i=1}^{k}\{l_i\}.$$

Then, introducing the notation $\vec{\Delta}n=(n(s_2)-n(s_1),...,n(s_k)-n(s_{k-1}))$, it follows that

$$P\left((J,\vec{\Delta}N)\in\partial\left(B\times\prod_{i=1}^{k}\{l_i\}\right)\right)\leq P(J\in\partial B|E(N))\,P(E(N)),$$

which is zero by the choice of B. Therefore the probability on the LHS is zero, i.e. $B\times\prod_{i=1}^{k}\{l_i\}$ is a continuity set of $P\left((J,\vec{\Delta}N)\in(.)\right)$. By the continuous mapping theorem, $\mu_T(u)\overset{w}{\to}J(u)$

implies $\left(\mu_T, \vec{\Delta} N_T\right) \xrightarrow{w} \left(J, \vec{\Delta} N\right)$. Therefore,

$$P\left(\left(\mu_T, \vec{\Delta} N_T\right) \in B \times \prod_{i=1}^{k} \{l_i\}\right) \to P\left(\left(J, \vec{\Delta} N\right) \in B \times \prod_{i=1}^{k} \{l_i\}\right).$$

Furthermore,

$$\begin{aligned} P\left(\mu_T \in B \,|E\left(N_T\right)\right) &= \frac{P\left(\left(\mu_T, \vec{\Delta} N_T\right) \in B \times \prod_{i=1}^{k} \{l_i\}\right)}{P\left(E\left(N_T\right)\right)} \to \frac{P\left(\left(J, \vec{\Delta} N\right) \in B \times \prod_{i=1}^{k} \{l_i\}\right)}{P\left(E\left(N\right)\right)} \\ &= P\left(J \in B | E\left(N\right)\right), \end{aligned}$$

which by the arbitrariness of B shows that

$$\mu_T\left(u\right) | E\left(N_T\right) \xrightarrow{w} J\left(u\right) | E\left(N\right). \tag{5}$$

b1. **Convergence conditionally on known jump dates.** Let jumps be known to occur at relative times u_i, $0 < u_1 < u_2 < ... < u_m < 1$, and at no other dates. Then, in $D\left[0,1\right]$,

$$m_{T,\omega}\left(u\right) = \sum_{i:[Tu_i] \le [Tu]} \eta_i\left(\omega\right) \underset{T\to\infty}{\longrightarrow} \sum_{i:u_i \le u} \eta_i\left(\omega\right)$$

for all ω in the sample space on which η_i are defined.[3] Next, for big T any interval of length $\frac{1}{T}$ contains at most one jump point u_i, so that

$$m_T\left(u\right) \stackrel{d}{=} \mu_T\left(u\right) | \left\{\pi_t = I\left\{t = [Tu_i] \text{ for some } i \in 1,..,m\right\}, t = 1, ..., T\right\}.$$

It also holds that $\sum_{i:u_i \le u} \eta_i \stackrel{d}{=} J\left(u\right) | N$ (with conditioning on a sample path of N with jumps at and only at u_i). Then

$$m_T\left(u\right) \xrightarrow{w} J\left(u\right) | N.$$

b. **Convergence conditionally on partially known jump dates.** As above, let jumps be known to occur at relative times $u_1 < u_2 < ... < u_m$, but let the possibility of other jumps not be precluded (condition C). Under this condition,

$$\mu_T\left(u\right) | C\left(N_T\right) \stackrel{d}{=} \sum_{t=1}^{[Tu]} \delta_t - \sum_{t=[Tu_i] \le [Tu]} \delta_t + \sum_{t=[Tu_i] \le [Tu]} \eta_t,$$

where the uncertain jumps at relative times u_i have been subtracted from the first sum on the RHS and have been replaced by certain jumps. Since $\left\|\sum_{t=[Tu_i] \le [Tu]} \delta_t\right\| \le \sum_{i=1}^{k} \left\|\delta_{[Tu_i]}\right\| = o_P\left(1\right)$, it holds that

$$\sum_{t=1}^{[Tu]} \delta_t - \sum_{t=[Tu_i] \le [Tu]} \delta_t = \sum_{t=1}^{[Tu]} \delta_t + o_P\left(1\right) \xrightarrow{w} J\left(u\right)$$

[3] This holds in spite of the fact that the difference $m_{T,\omega}\left(u\right) - \sum_{i:u_i \le u} \vartheta_i\left(\omega\right) = \sum_{i:u < u_i < ([Tu]+1)/T} \vartheta_i\left(\omega\right)$ fails to converge to 0 in $D\left[0,1\right]$ (the latter is not a topological vector space).

by applying FPCT to the first term. By the argument in (b1), it holds further that

$$\sum_{t=[Tu_i]\leq[Tu]} \eta_t \overset{w}{\to} \sum_{u_i<u} \eta_i \overset{d}{=} \sum_{u_i<u} \eta_{N(1)+i}.$$

The latter distribution does not depend on $N(1)$, the inclusion of $N(1)$ in the subscript is only a notational device for indicating that the jump sizes of J and those of the certain jumps are independent.

Next, the processes $\sum_{t=1}^{[Tu]} \delta_t - \sum_{t=[Tu_i]\leq[Tu]} \delta_t$ and $\sum_{t=[Tu_i]\leq[Tu]} \eta_t$ are independent, and therefore their weak convergence is joint on $D^2[0,1]$. The limiting processes almost surely have no jump points in common, so that the mapping $(x,y) \to x+y$ is continuous on a support of $\left(\sum_{t=1}^{[Tu]} \delta_t - \sum_{t=[Tu_i]\leq[Tu]} \delta_t,\ \sum_{t=[Tu_i]\leq[Tu]} \eta_t\right)$. Hence

$$\mu_T(u)\,|C(N_T) \overset{w}{\to} J(u) + \sum_{u_i<u} \eta_{N(1)+i}.$$

It remains to show that $J(u)\,|C(N) \overset{d}{=} J(u) + \sum_{u_i<u} \eta_{N(1)+i}$. Introduce $u_0 = 0$, and for a $u \in [0,1]$ let j_u be the maximal index such that $u_j \leq u$. Then

$$J(u)\,|C(N) \overset{d}{=} \sum_{i=1}^{j_u} (J(u_i-) - J(u_{i-1}))\,|C(N) + (J(u) - J(u_{j_u}))\,|C(N) + \sum_{i=1}^{j_u} \Delta J(u_i)\,|C(N),$$

where the probability measures corresponding to the separate addends are independent. Since $J(u) - J(u_j)$ and $J(u_i-) - J(u_{i-1})$ are independent of $\Delta J(u_i)$, it follows that

$$(J(u) - J(u_j))\,|C(N) \overset{d}{=} J(u) - J(u_j)$$

and

$$(J(u_i-) - J(u_{i-1}))\,|C(N) \overset{d}{=} J(u_i-) - J(u_{i-1}) \overset{d}{=} J(u_i) - J(u_{i-1}).$$

The last equality of distributions is true because the unconditional probability for a jump at at least one u_i is zero. Thus,

$$J(u)\,|C(N) \overset{d}{=} \sum_{i=1}^{j} (J(u_i) - J(u_{i-1})) + J(u) - J(u_j) + \sum_{i=1}^{j} \Delta J(u_i)\,|C(N) \overset{d}{=} J(u) + \sum_{u_i<u} \eta_{N(1)+i},$$

so that

$$\mu_T(u)\,|C(N_T) \overset{w}{\to} J(u)\,|C(N).$$

c. **Convergence conditionally both on known number of jumps between certain dates and on partially known jump dates.** As in (b1), start from

$$\mu_T(u)\,|E(N_T)\,\&C(N_T) \overset{d}{=} \left(\sum_{t=1}^{[Tu]}\delta_t - \sum_{t=[Tu_i]\le[Tu]}\delta_t + \sum_{i:[Tu_i]\le[Tu]}\eta_{T+i}\right)\Bigg|\, E(N_T)\,\&C(N_T)$$

$$\overset{d}{=} \left(\sum_{t=1}^{[Tu]}\delta_t - \sum_{t=[Tu_i]\le[Tu]}\delta_t\right)\Bigg|\, E(N_T)\,\&C(N_T) + \sum_{i:[Tu_i]\le[Tu]}\eta_{T+i}\Bigg|\, E(N_T)\,\&C(N_T),$$

where the measures corresponding to the two addends are independent. Eliminating, by independence, irrelevant conditions, allows us to simplify the above distribution to that of

$$\left(\sum_{t=1}^{[Tu]}\delta_t - \sum_{t=[Tu_i]\le[Tu]}\delta_t\right)\Bigg|\, \tilde{E}(N_T) + \sum_{i:[Tu_i]\le[Tu]}\eta_{T+i},$$

where $\tilde{E}(N_T)$ is the condition $N_T(s_i) - N_T(s_{i-1}) = l_i - |\{u_j : s_{i-1} < u_j \le s_i\}|$. The distribution can finally be written as

$$\sum_{t=1}^{[Tu]}\delta_t\Bigg|\, \tilde{E}(N_T) + \sum_{i:[Tu_i]\le[Tu]}\eta_{T+i} + o_P(1),$$

again with the first two addends independent. From above, using the result from (a) for the convergence of $\sum_{t=1}^{[Tu]}\delta_t\Big|\,\tilde{E}(N_T)$, by the same argument as in (b1), it follows that

$$\mu_T(u)\,|E(N_T)\,\&C(N_T) \xrightarrow{w} J(u)\,|\tilde{E}(N) + \sum_{u_i<u}\eta_{N(1)+i} \overset{d}{=} J(u)\,|E(N)\,\&C(N).$$

Hence,

$$\mu_T(u)\,|E(N_T)\,\&C(N_T) \xrightarrow{w} J(u)\,|E(N)\,\&C(N).$$

Corollary. Let $f : D_{\mathbb{R}^k}[0,1] \to \mathbb{R}$ be an arbitrary bounded continuous functional. The FPCT can be rewritten as

$$P(N(1)=0)\int f dP(J\in(\cdot)\,|N(1)=0) + P(N(1)>0)\int f dP(J\in(\cdot)\,|N(1)>0) \qquad (6)$$

$$= \int f dP(J\in(\cdot)) \underset{T\to\infty}{\leftarrow} \int f dP^T(\mu_T\in(\cdot))$$

$$= P(N_T(1)=0)\int f dP(\mu_T\in(\cdot)\,|N_T(1)=0) + P(N_T(1)>0)\int f dP(\mu_T\in(\cdot)\,|N_T(1)>0).$$

By the law of rare events for Bernoulli rv's, $P(N_T(1)=0) \to P(N(1)=0)$ and $P(N_T(1)>0) \to P(N(1)>0) \neq 0$. By $P(\mu_T\in(\cdot)\,|N_T(1)=0) \xrightarrow{w} P(J\in(\cdot)\,|N(1)=0)$, it holds $\int f dP(\mu_T\in(\cdot)\,|N_T(1)=0) \to \int f dP(J\in(\cdot)\,|N(1)=0)$. Thus, three terms on the RHS of (6) converge to their counterparts on the LHS, and so the fourth term must also converge: $\int f dP(\mu_T\in(\cdot)\,|N_T(1)>0) \to \int f dP(J\in(\cdot)\,|N(1)>0)$, and hence, $P(\mu_T\in(\cdot)\,|N_T(1)>0) \xrightarrow{w} P(J\in(\cdot)\,|N(1)>0)$. ■

References

[1] Billingsley P., "Convergence of Probability Measures", John Wiley and Sons, 1968

[2] Davidson J., "Stochastic Limit Theory", Oxford University Press, 1994

[3] Doornik J, D. Hendry and B. Nielsen, "Inference in Cointegrated Models: UK M1 Revisited", Journal of Economic Surveys, 1998, Vol. 12 No. 5

[4] Franses P.H., N. Haldrup, "The Effects of Additive Outliers on Tests for Unit Roots and Cointegration", Journal of Business and Economic Statistics, 1994, Vol. 12, No. 4

[5] Jacod J., A.N. Shiryaev, "Predel'nuye teoremuy dlya sluchaynuyh processov", Volumes 1 and 2, Moscow, Fizmatlit, 1994 (in Russian, translated from "Limit Theorems for Stochastic Processes", Springer-Verlag, 1987)

[6] Kurtz T., P. Protter, "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations", The Annals of Probability, 1991, 19(3): 1035-1070

[7] Perron P., "The Great Crash, the Oil Price Shock, and the Unit Root Hypotheses", Econometrica, 1989, Vol.57, No. 6

[8] Perron P., "Testing for a Unit Root in a Time Series With a Changing Mean", Journal of Business and Economic Statistics, 1990, Vol.8, No. 2