Supplemental material for the main submission entitled ”Expansion and Estimation of Lévy Process Functionals in Nonlinear and Nonstationary Time Series Regression”

Chaohua Dong^a and Jiti Gao^b

School of Economics,

Southwestern University of Finance and Economics^a, China

Department of Econometrics and Business Statistics,

Monash University^b, Australia

August 23, 2016

Appendix D: Basic lemmas and the proofs of Lemmas B.1 & B.2

D.1: Basic lemmas for orthogonal expansion

Suppose Lévy process $Z(t)$ admits a classical orthonormal polynomial system $\{Q_i(t, x)\}$ with weight $\rho(t, x)$. This means that the density of $Z(t)$ satisfies equation (E.1) and $\{Q_i(t, x)\}$ is given by either (E.15) or (E.31) depending on whether $Z(t)$ is continuous or discrete. Consider function space

$$L^2([0, T] \times I, \nu) = \left\{ f(t, x) : \int_I f^2(t, x) d\Psi_t(x) < \infty, \text{ for each } t \in [0, T], \right. \\ \left. \text{and } \int_0^T \int_I f^2(t, x) d\nu < \infty \right\},$$

where $\Psi_t(x)$ is the distribution function of $Z(t)$, I is the support of $\rho(t, x)$ in terms of x , and ν is the product measure of $\Psi_t(x)$ and Lebesgue measure on the real axis. Note that if $Z(t)$ is a discrete variable (e.g. Poisson random variable) for each t , the integral $\int_I f^2(t, x) d\Psi_t(x)$ would boil down to a sum $\sum_x f^2(t, x) \rho(t, x)$.

We abbreviate the notation of the space as $L^2([0, T] \times I)$. As a conventional L^2 space, $L^2([0, T] \times I)$ is a Hilbert space with $(f_1(t, x), f_2(t, x)) = \int_0^T \int_I f_1(t, x) f_2(t, x) d\nu$ as its inner product. Since $\{Q_i(t, x)\}$ is an orthonormal basis for $L^2(I, d\Psi_t(x))$ and $\{\varphi_{jT}(t)\}$, where $\varphi_{0T} = \sqrt{\frac{1}{T}}$ and $\varphi_{jT} = \sqrt{\frac{2}{T}} \cos \frac{j\pi t}{T}$ for $j \geq 1$, is an orthonormal basis in $L^2([0, T])$, according to Dudley (2003, p173), $\{Q_i(t, x) \varphi_{jT}(t)\}$ is an orthonormal basis in $L^2([0, T] \times I)$.

Construct a mapping $\mathcal{T}$ from $L^2([0, T] \times I)$ to a set of processes, $\mathcal{T} : f(t, x) \mapsto f(t, Z(t))$, for $f(t, x) \in L^2([0, T] \times I)$. Denote by Ξ the image of $\mathcal{T}$. Define $\langle f_1(t, Z(t)), f_2(t, Z(t)) \rangle_{\Xi} = \int_0^T E[f_1(t, Z(t))f_2(t, Z(t))]dt$ on Ξ . Obviously, $\langle \cdot, \cdot \rangle_{\Xi}$ is an inner product.

Lemma D.1. *The mapping $\mathcal{T}$ has the following properties: (1) $\mathcal{T}$ is linear; (2) $\mathcal{T}$ is an one-to-one mapping from $L^2([0, T] \times I)$ to Ξ ; and (3) $\mathcal{T}$ is an isomorphism.*

PROOF OF LEMMA D.1. (1) Straightforward verification. (2) $\mathcal{T}$ is an inner product preserving mapping since $\langle \mathcal{T}(f), \mathcal{T}(g) \rangle_{\Xi} = \int_0^T E[f(Z(t))g(Z(t))]dt = (f, g)_{L^2([0, T] \times I)}$. Therefore, $f \neq g \Leftrightarrow \mathcal{T}(f) \neq \mathcal{T}(g)$. Thus, $\mathcal{T}$ is one-to-one. (3) Since $\mathcal{T}$ is linear and $\|\mathcal{T}(f)\|_{\Xi} = \|f\|$, $\mathcal{T}$ is isomorphism. $\square$

Lemma D.2. *Ξ is a closed subspace of $L^2(\Omega)$, hence it is a Hilbert space.*

PROOF OF LEMMA D.2. Note that Ξ is a linear space due to the linearity of $\mathcal{T}$. Because $\mathcal{T}$ is one-to-one and inner product preserving, a Cauchy sequence is still of Cauchy under the mapping, so that Ξ is a closed subspace of $L^2(\Omega)$. Hence it is a Hilbert space. $\square$

Lemma D.3. *If $\{p_i(t, x)\}_{i=0}^{\infty}$ is any orthonormal basis in $L^2([0, T] \times I)$, then $\{\mathcal{T}(p_i)\}_{i=0}^{\infty}$ is an orthonormal basis in Ξ . Particularly, $\{\varphi_{jT}(t)Q_i(t, Z(t))\}_{i,j=0}^{\infty}$, $t > 0$, is an orthonormal basis in Ξ .*

PROOF OF LEMMA D.3. By virtue of the properties of $\mathcal{T}$ that $\mathcal{T}$ is one-to-one, inner product preserving, it is valid. $\square$

Lemmas D.1–D.3 assert that Ξ is a Hilbert space and $\{Q_i(t, Z(t))\varphi_{jT}(t)\}$ ($i, j = 0, 1, \dots$) is an orthonormal basis in Ξ . The following lemma is obtained from Hilbert space theory directly.

Lemma D.4. *Suppose that Lévy process $(Z(t), t > 0)$ admits a classical orthonormal polynomial system $\{Q_i(t, x)\}$ with weight $\rho(t, x)$. Suppose $f(t, x)$ is continuous in both t and x . Then, $f(t, Z(t)) \in \Xi$ admits a Fourier series expansion of the form*

$$f(t, Z(t)) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{ij} \varphi_{jT}(t) Q_i(t, Z(t)), \quad (\text{D.1})$$

where $c_{ij} = \langle f(t, Z(t)), Q_i(t, Z(t))\varphi_{jT}(t) \rangle_{\Xi}$. The convergence of (D.1) is in the sense of the norm induced by the inner product in Ξ .

PROOF OF LEMMA D.4. In view of the facts that Ξ is a Hilbert space and $\{\varphi_{jT}(t)Q_i(t, Z(t))\}$ is an orthonormal basis in Ξ , as well as the continuity of $f(t, x)$, (D.1) follows. $\square$

The expansion (D.1) can be regarded as a two-step expansion, that is, expand $f(t, Z(t))$ first in terms of $\{Q_i(t, Z(t))\}$ obtaining coefficient function $c_i(t, f)$, then expand $c_i(t, f)$ in terms of $\{\varphi_{jT}(t)\}$ on $[0, T]$, viz.,

$$c_i(t, f) = E[f(t, Z(t))Q_i(t, Z(t))] \quad \text{and} \quad c_{ij} = \int_0^T c_i(t, f) \varphi_{jT}(t) dt. \quad (\text{D.2})$$

Notice from Parseval equality that

$$\|f(t, Z(t))\|_{\Xi}^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{ij}^2 = \sum_{i=0}^{\infty} \|c_i(t, f)\|_{L^2[0, T]}^2, \quad (\text{D.3})$$

which indicates that $\|c_i(t, f)\|_{L^2[0, T]}^2 \rightarrow 0$ as $i \rightarrow \infty$.

Lemma D.4 shows that a time-inhomogeneous functional of $Z(t)$ can be expanded as an orthogonal series in $L^2(\Omega)$. Trivially, a time-homogeneous functional $f(Z(t))$ of $Z(t)$ can also be expanded via (D.1). We emphasize that the expansion (D.1) is completely different from the usual expansion in the literature. First, the expansion of (D.1) takes into account time variable t which is involved in the basis. Second, the expansion (D.1) takes place in the probability space since every element in $\{Q_i(t, Z(t))\varphi_{jT}(t)\}$ is a stochastic process of $L^2(\Omega)$ and all of them consist of an orthonormal basis of the Hilbert space that is a subspace of $L^2(\Omega)$. This entails that the partial sum of the expansion (D.1) has orthogonal increments. More precisely, every extra term added to a partial sum is orthogonal to all the previous terms in the sense of $L^2(\Omega)$, so that every term makes the partial sum approach the function $m(t, Z(t))$ in an effective way, implying that our approximation is accurate. A detailed discussion of such processes can be found in Chapter IX, Processes with orthogonal increments, of Doob (1953). It is worth pointing out that by virtue of the expansion (D.1), we can characterize the convergence of the partial sum sequence in $L^2(\Omega)$ sense to the functional $m(t, Z(t))$ by the coefficients

$$\int_0^T E \left[m(t, Z(t)) - \sum_{i=0}^{k-1} \sum_{j=0}^{p-1} c_{ij} \varphi_{jT}(t) Q_i(t, Z(t)) \right]^2 dt = \sum_{i=k}^{\infty} \sum_{j=p}^{\infty} c_{ij}^2 \rightarrow 0$$

as $k, p \rightarrow \infty$.

However, in the literature, to expand a functional of a process, $g(Z(t))$ say, researchers usually expand $g(x)$ first for $x \in \mathbb{R}$ in term of an orthonormal basis, $g_i(x)$ say, that is, $g(x) = \sum_{i=0}^{\infty} c_i g_i(x)$, and then plug in the process to have $g(Z(t)) = \sum_{i=0}^{\infty} c_i g_i(Z(t))$. The significant difference is that $\{g_i(Z(t))\}$ is generally not orthogonal in $L^2(\Omega)$ at all, so that the plug-in expansion is not an orthogonal expansion. It can only be understood as an expansion on each realization of the process. As a result, it is impossible to consider the convergence of $g(Z(t))$ in the sense of stochastic process in $L^2(\Omega)$ directly, though $g(Z(t)) \in L^2(\Omega)$ is a stochastic process, because in general $E[g(Z(t))^2] \neq \sum_{i=0}^{\infty} c_i^2$ and $E[g(Z(t)) - \sum_{i=0}^{k-1} c_i g_i(Z(t))]^2 \neq \sum_{i=k}^{\infty} c_i^2$. In this regard, our expansion method outperforms the conventional one.

Here, we provide a relationship with the coefficients $c_i(t, f) = E[f(t, Z(t))Q_i(t, Z(t))]$, where $Q_i(t, x) = \frac{1}{d_i(t)} y_i(t, x)$ defined in Appendix E. This is useful for the proof of the main result.

If $Z(t)$ is continuous with density function $\rho(t, x)$, the polynomials $y_i(t, x)$ satisfy the hypergeometric differential equations $s(t, x)y_i''(t, x) + v(t, x)y_i'(t, x) + \lambda_i(t)y_i(t, x) = 0$, where $s(t, x) > 0$, $v(t, x)$ and $\rho(t, x)$ satisfy conditions (E.1) in Appendix E and $\lambda_i(t) = -iv'(t, x)$ since $s''(t, x) = 0$ in the scope of this study (see Remark E.1 below). Throughout the differentiation is conducted with respect to x only.

The equation can be rewritten as $(s(t, x)\rho(t, x)y_i'(t, x))' + \lambda_i(t)\rho(t, x)y_i(t, x) = 0$. Multiplying both sides by $f(t, x)$, using condition (E.1), integrating by part on (a, b) with respect to x , we have $c_i(t, f) = \frac{d_{1i}(t)}{\lambda_i(t)d_i(t)} c_i(t, f')$, where d_i^2 and d_{1i}^2 are squared norms of y_i and y_i' , respectively. We can iterate the relation until r -th derivative, using the relation between d_i^2 and d_{ri}^2 (see Appendix E),

$$c_i(t, f) = \sqrt{\psi(t)}^r \sqrt{\frac{(i-r)!}{i!}} c_i(t, D^r f), \quad (\text{D.4})$$

where $\psi(t) > 0$ is determined by the process $Z(t)$ given in Remark E.1 below.

When $Z(t)$ is discrete, equation (D.4) is also true, in view of the conditions in (E.1) and the difference equation of (E.20). $\square$

D.2: Basic lemmas for asymptotic theory

We introduce for any $\epsilon > 0$ and $0 \leq r \leq 1$,

$$L_n^{(r)} = \frac{c_n}{n} \sum_{k=1}^{[nr]} f\left(\frac{k}{n}, c_n x_{k,n}\right),$$

$$L_{n,\epsilon}^{(r)} = \frac{c_n}{n} \sum_{k=1}^{[nr]} \int_{-\infty}^{\infty} f\left(\frac{k}{n}, c_n(x_{k,n} + z\epsilon)\right) \phi(z) dz,$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. For later use, we also define $\phi_\epsilon(z) = \frac{1}{\epsilon} \phi\left(\frac{z}{\epsilon}\right)$ for some $\epsilon > 0$.

Lemma D.5. *Suppose that Assumptions A.1(c) and A.3 hold. Then*

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{0 \leq r \leq 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0.$$

PROOF OF LEMMA D.5. The proof consists of two parts according to $x_{k,n}$ being continuous and discrete respectively in Assumption A.1 (c).

The following arguments about the continuous case naturally treat those used for the univariate case in Wang and Phillips (2009) as a special case.

Denote $Y_{k,n}(z) = f\left(\frac{k}{n}, c_n x_{k,n}\right) - f\left(\frac{k}{n}, c_n(x_{k,n} + z\epsilon)\right)$. We have

$$\begin{aligned} \sup_{0 \leq r \leq 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| &= \sup_{0 \leq r \leq 1} E \left| \frac{c_n}{n} \int_{-\infty}^{\infty} \sum_{k=1}^{[nr]} Y_{k,n}(z) \phi(z) dz \right| \\ &\leq \frac{c_n}{n} \int_{-\infty}^{\infty} \sup_{0 \leq r \leq 1} E \left| \sum_{k=1}^{[nr]} Y_{k,n}(z) \right| \phi(z) dz, \end{aligned}$$

by the fact that $\int \phi(z) dz = 1$. Notice that, by Assumption A.1 (c),

$$\begin{aligned} E|Y_{k,n}(z)| &= \int_{-\infty}^{\infty} \left| f\left(\frac{k}{n}, c_n d_{k,0,n} x\right) - f\left(\frac{k}{n}, c_n d_{k,0,n} x + c_n z\epsilon\right) \right| h_{k,0,n}(x) dx \\ &\leq \frac{K}{c_n d_{k,0,n}} \left[\int_{-\infty}^{\infty} \left| f\left(\frac{k}{n}, x\right) \right| dx + \int_{-\infty}^{\infty} \left| f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| dx \right] \\ &= \frac{2K}{c_n d_{k,0,n}} g\left(\frac{k}{n}\right), \end{aligned} \tag{D.5}$$

where $g(\cdot) = \int_{-\infty}^{\infty} |f(\cdot, x)| dx$ and K is the uniform upper bound of the density $h_{l,k,n}$. Accordingly, for each $z \in \mathbb{R}$,

$$\frac{c_n}{n} \sup_{0 \leq r \leq 1} E \left| \sum_{k=1}^{[nr]} Y_{k,n}(z) \right| \leq \frac{c_n}{n} \sum_{k=1}^n \frac{2K}{c_n d_{k,0,n}} g\left(\frac{k}{n}\right) = 2K K_2 \frac{1}{n} \sum_{t=1}^n \frac{1}{d_{k,0,n}} < \infty$$

by Assumption A.1(c)(i), where $K_2 = \sup_{t \in [0,1]} g(t) < \infty$ due to the continuity of $g(t)$. It therefore follows from the dominated convergence theorem that, to prove the lemma, it suffices to show that for any fixed z , $\Lambda_n(\epsilon) = \frac{c_n^2}{n^2} \sup_{0 \leq r \leq 1} E \left[\sum_{k=1}^{[nr]} Y_{k,n}(z) \right]^2 \rightarrow 0$ as $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$.

Meanwhile, we have

$$\Lambda_n(\epsilon) \leq \frac{c_n^2}{n^2} \sum_{k=1}^n EY_{k,n}^2(z) + \frac{2c_n^2}{n^2} \sum_{k=1}^{n-1} \sum_{l=k+1}^n |E[Y_{k,n}(z)Y_{l,n}(z)]| = \Lambda_{1n}(\epsilon) + \Lambda_{2n}(\epsilon).$$

We next investigate $\Lambda_{1n}(\epsilon)$ and $\Lambda_{2n}(\epsilon)$ separately. In view of Assumption A.1 (c), we have as $n \rightarrow \infty$

$$\begin{aligned} \Lambda_{1n}(\epsilon) &= \frac{c_n^2}{n^2} \sum_{k=1}^n EY_{k,n}^2(z) = \frac{c_n^2}{n^2} \sum_{k=1}^n E \left[f \left(\frac{k}{n}, c_n x_{k,n} \right) - f \left(\frac{k}{n}, c_n(x_{k,n} + z\epsilon) \right) \right]^2 \\ &= \frac{c_n^2}{n^2} \sum_{k=1}^n \int_{-\infty}^{\infty} \left[f \left(\frac{k}{n}, c_n d_{k,0,n} x \right) - f \left(\frac{k}{n}, c_n d_{k,0,n} x + c_n z\epsilon \right) \right]^2 h_{k,0,n}(x) dx \\ &\leq \frac{c_n^2}{n^2} \sum_{k=1}^n \frac{K}{c_n d_{k,0,n}} \int_{-\infty}^{\infty} \left[f \left(\frac{k}{n}, x \right) - f \left(\frac{k}{n}, x + c_n z\epsilon \right) \right]^2 dx \\ &\leq \frac{4Kc_n}{n^2} \sum_{k=1}^n \frac{1}{d_{k,0,n}} \int_{-\infty}^{\infty} \left| f \left(\frac{k}{n}, x \right) \right|^2 dx = \frac{4Kc_n}{n^2} \sum_{k=1}^n \frac{1}{d_{k,0,n}} G_2 \left(\frac{k}{n} \right) \\ &\leq 4KK_3 \frac{c_n}{n} \frac{1}{n} \sum_{k=1}^n \frac{1}{d_{k,0,n}} \rightarrow 0, \end{aligned}$$

where $K_3 = \sup_{t \in [0,1]} G_2(t)$ and $G_2(\cdot) = \int f^2(\cdot, x) dx$ is continuous on the interval in question.

We then prove that $\Lambda_{2n}(\epsilon) \rightarrow 0$ as $n \rightarrow \infty$. Because

$$\Lambda_{2n}(\epsilon) = \frac{2c_n^2}{n^2} \sum_{k=1}^{n-1} \sum_{l=k+1}^n |E[Y_{k,n}(z)Y_{l,n}(z)]| = \frac{2c_n^2}{n^2} \sum_{k=1}^{n-1} \sum_{l=k+1}^n |E[Y_{k,n}(z)E(Y_{l,n}(z)|\mathcal{F}_{k,n})]|,$$

For $k < l$, we begin with the following calculation of the conditional expectation:

$$\begin{aligned} |E(Y_{l,n}(z)|\mathcal{F}_{k,n})| &= \left| E \left[f \left(\frac{l}{n}, c_n x_{l,n} \right) f \left(\frac{l}{n}, c_n(x_{l,n} + z\epsilon) \right) \middle| \mathcal{F}_{k,n} \right] \right| \\ &= \left| E \left[f \left(\frac{l}{n}, c_n x_{k,n} + c_n(x_{l,n} - x_{k,n}) \right) - f \left(\frac{l}{n}, c_n x_{k,n} + c_n(x_{l,n} - x_{k,n}) + c_n z\epsilon \right) \middle| \mathcal{F}_{k,n} \right] \right| \\ &= \left| \int_{-\infty}^{\infty} \left[f \left(\frac{l}{n}, c_n x_{k,n} + c_n d_{l,k,n} y \right) - f \left(\frac{l}{n}, c_n x_{k,n} + c_n d_{l,k,n} y + c_n z\epsilon \right) \right] h_{l,k,n}(y) dy \right| \\ &= \frac{1}{c_n d_{l,k,n}} \left| \int_{-\infty}^{\infty} \left[f \left(\frac{l}{n}, y \right) h_{l,k,n} \left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}} \right) - f \left(\frac{l}{n}, y \right) h_{l,k,n} \left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}} \right) \right] dy \right| \\ &= \frac{1}{c_n d_{l,k,n}} \left| \int_{-\infty}^{\infty} f \left(\frac{l}{n}, y \right) \left[h_{l,k,n} \left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}} \right) - h_{l,k,n} \left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}} \right) \right] dy \right| \\ &\leq \frac{1}{c_n d_{l,k,n}} \int_{-\infty}^{\infty} \left| f \left(\frac{l}{n}, y \right) \right| |V(y, c_n x_{k,n})| dy, \end{aligned}$$

where $V(y, c_n x_{k,n}) = h_{l,k,n} \left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}} \right) - h_{l,k,n} \left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}} \right)$.

Recall the definition of $\Omega_n(\epsilon)$ in Assumption A.1 (c) and note that a pair (l, k) ($l > k$) belongs to either $\Omega_n(\epsilon^{1/2m_0})$ or its complement. It follows that

$$\begin{aligned} &|E(Y_{l,n}(z)|\mathcal{F}_{k,n})| \\ &\leq \begin{cases} \frac{2K}{c_n d_{l,k,n}} \int_{-\infty}^{\infty} \left| f \left(\frac{l}{n}, y \right) \right| dy = \frac{2K}{c_n d_{l,k,n}} g \left(\frac{l}{n} \right), & \text{if } (l, k) \notin \Omega_n, \\ \frac{2K}{c_n d_{l,k,n}} \int_{|y| > \sqrt{c_n}} \left| f \left(\frac{l}{n}, y \right) \right| dy \\ \quad + \frac{1}{c_n d_{l,k,n}} \int_{|y| \leq \sqrt{c_n}} \left| f \left(\frac{l}{n}, y \right) \right| |V(y, c_n x_{k,n})| dy, & \text{if } (l, k) \in \Omega_n, \end{cases} \end{aligned}$$

where $\Omega_n = \Omega_n(\epsilon)$ is the same as in Assumption A.1 (c).

According to Assumption A.1 (c), $\inf_{(l,k) \in \Omega_n(\epsilon^{1/2m_0})} d_{l,k,n} \geq \sqrt{\epsilon}/C$, and at the same time we can choose n large enough such that $\sqrt{c_n}\epsilon > 1$. For $|y| \leq \sqrt{c_n}$ and $|x| \leq \sqrt{c_n} + c_n|z|\epsilon$, when $(l, k) \in \Omega_n(\epsilon^{1/2m_0})$, we have

$$\begin{aligned} |V(y, x)| &= \left| h_{l,k,n} \left(\frac{y-x}{c_n d_{l,k,n}} \right) - h_{l,k,n} \left(\frac{y-x-c_n \epsilon z}{c_n d_{l,k,n}} \right) \right| \\ &\leq \left| h_{l,k,n} \left(\frac{y-x}{c_n d_{l,k,n}} \right) - h_{l,k,n}(0) \right| + \left| h_{l,k,n} \left(\frac{y-x-c_n \epsilon z}{c_n d_{l,k,n}} \right) - h_{l,k,n}(0) \right| \\ &\leq 2 \sup_{|u| < 2C(1+|z|)\sqrt{\epsilon}} |h_{l,k,n}(u) - h_{l,k,n}(0)|. \end{aligned} \quad (D.6)$$

Therefore, when $|y| \leq \sqrt{c_n}$, n is large enough and $(l, k) \in \Omega_n(\epsilon^{1/2m_0})$, we have

$$\begin{aligned} &E|Y_{k,n}(z)||V(y, c_n x_{k,n})| \\ &= \int_{-\infty}^{\infty} \left| f \left(\frac{k}{n}, c_n d_{k,0,n} x \right) - f \left(\frac{k}{n}, c_n d_{k,0,n} x + c_n z \epsilon \right) \right| |V(y, c_n d_{k,0,n} x)| h_{k,0,n}(x) dx \\ &\leq \frac{K}{c_n d_{k,0,n}} \int_{-\infty}^{\infty} \left| f \left(\frac{k}{n}, x \right) - f \left(\frac{k}{n}, x + c_n z \epsilon \right) \right| |V(y, x)| dx \\ &\leq \frac{K}{c_n d_{k,0,n}} \left[\int_{-\infty}^{\infty} \left| f \left(\frac{k}{n}, x \right) \right| |V(y, x)| dx + \int_{-\infty}^{\infty} \left| f \left(\frac{k}{n}, x + c_n z \epsilon \right) \right| |V(y, x)| dx \right] \\ &= \frac{K}{c_n d_{k,0,n}} \int_{-\infty}^{\infty} \left| f \left(\frac{k}{n}, x \right) \right| [|V(y, x)| + |V(y, x - c_n z \epsilon)|] dx \\ &= \frac{K}{c_n d_{k,0,n}} \left[\int_{|x| > \sqrt{c_n}} + \int_{|x| \leq \sqrt{c_n}} \left| f \left(\frac{k}{n}, x \right) \right| [|V(y, x)| + |V(y, x - c_n z \epsilon)|] dx \right] \\ &\leq \frac{2K^2}{c_n d_{k,0,n}} \int_{|x| > \sqrt{c_n}} \left| f \left(\frac{k}{n}, x \right) \right| dx \\ &\quad + \frac{K}{c_n d_{k,0,n}} \int_{|x| \leq \sqrt{c_n}} \left| f \left(\frac{k}{n}, x \right) \right| [|V(y, x)| + |V(y, x - c_n z \epsilon)|] dx \\ &\leq \frac{2K^2}{c_n d_{k,0,n}} \int_{|x| > \sqrt{c_n}} \left| f \left(\frac{k}{n}, x \right) \right| dx \\ &\quad + \frac{4K}{c_n d_{k,0,n}} \sup_{|u| < 2C(1+|z|)\sqrt{\epsilon}} |h_{l,k,n}(u) - h_{l,k,n}(0)| \int_{|x| \leq \sqrt{c_n}} \left| f \left(\frac{k}{n}, x \right) \right| dx. \end{aligned}$$

We summarize that if $(l, k) \notin \Omega_n$, equation (D.5) yields

$$\begin{aligned} |E(Y_{k,n}(z)Y_{l,n}(z))| &= |E[Y_{k,n}(z)E(Y_{l,n}(z)|\mathcal{F}_{k,n})]| \\ &\leq \frac{2K}{c_n d_{l,k,n}} g \left(\frac{l}{n} \right) |EY_{k,n}(z)| \leq \frac{4K^2}{c_n^2 d_{l,k,n} d_{k,0,n}} g \left(\frac{l}{n} \right) g \left(\frac{k}{n} \right), \end{aligned}$$

while if $(l, k) \in \Omega_n$,

$$\begin{aligned} |E(Y_{k,n}(z)Y_{l,n}(z))| &= |E[Y_{k,n}(z)E(Y_{l,n}(z)|\mathcal{F}_{k,n})]| \\ &\leq E[|Y_{k,n}(z)| |E(Y_{l,n}(z)|\mathcal{F}_{k,n})|] \\ &\leq \frac{2K}{c_n d_{l,k,n}} \int_{|y| > \sqrt{c_n}} \left| f \left(\frac{l}{n}, y \right) \right| dy E|Y_{k,n}(z)| \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{c_n d_{l,k,n}} \int_{|y| \leq \sqrt{c_n}} \left| f\left(\frac{l}{n}, y\right) \right| E[|Y_{k,n}(z)| |V(y, c_n x_{k,n})|] dy \\
& \leq \frac{4K^2}{c_n^2 d_{l,k,n} d_{k,0,n}} g\left(\frac{k}{n}\right) \int_{|y| > \sqrt{c_n}} \left| f\left(\frac{l}{n}, y\right) \right| dy \\
& + \frac{2K^2}{c_n^2 d_{l,k,n} d_{k,0,n}} \int_{|y| \leq \sqrt{c_n}} \left| f\left(\frac{l}{n}, y\right) \right| dy \int_{|x| > \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| dx \\
& + \frac{4K}{c_n^2 d_{l,k,n} d_{k,0,n}} \int_{|y| \leq \sqrt{c_n}} \left| f\left(\frac{l}{n}, y\right) \right| dy \int_{|x| \leq \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| dx \\
& \times \sup_{|u| < 2C(1+|z|)\sqrt{\epsilon}} |h_{l,k,n}(u) - h_{l,k,n}(0)|.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
|\Lambda_{2n}(\epsilon)| & \leq \frac{2c_n^2}{n^2} \left(\sum_{l > k, (l,k) \notin \Omega_n} + \sum_{(l,k) \in \Omega_n} \right) E|Y_{s,n}(z)Y_{t,n}(z)| \\
& \leq \frac{2c_n^2}{n^2} \sum_{k=(1-\epsilon)n}^n \sum_{l=k+1}^n E|Y_{k,n}(z)Y_{l,n}(z)| + \frac{2c_n^2}{n^2} \sum_{k=1}^{n-1} \sum_{l=k+1}^{k+\epsilon n} E|Y_{k,n}(z)Y_{l,n}(z)| \\
& + \frac{2c_n^2}{n^2} \sum_{k=1}^{\epsilon n} \sum_{l=k+1}^n E|Y_{k,n}(z)Y_{l,n}(z)| + \frac{2c_n^2}{n^2} \sum_{k=1}^{n-1} \sum_{l=k+\epsilon n}^n E|Y_{k,n}(z)Y_{l,n}(z)| \\
& \leq 8K^2 K_2^2 \frac{1}{n} \sum_{k=(1-\epsilon)n}^n \frac{1}{d_{k,0,n}} \max_{1 \leq k \leq n-1} \frac{1}{n} \sum_{l=k+1}^n \frac{1}{d_{l,k,n}} \\
& + 8K^2 K_2^2 \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{d_{k,0,n}} \max_{1 \leq k \leq n-1} \frac{1}{n} \sum_{l=k+1}^{k+\epsilon n} \frac{1}{d_{l,k,n}} \\
& + 8K^2 K_2^2 \frac{1}{n} \sum_{k=1}^{\epsilon n} \frac{1}{d_{k,0,n}} \max_{1 \leq k \leq n-1} \frac{1}{n} \sum_{l=k+1}^n \frac{1}{d_{l,k,n}} \\
& + 8K^2 K_2 \int_{|y| > \sqrt{c_n}} c_f(y) dy \frac{1}{n^2} \sum_{k=1}^{n-1} \frac{1}{d_{k,0,n}} \max_{1 \leq k \leq n-1} \sum_{l=k+1}^n \frac{1}{d_{l,k,n}} \\
& + 4K^2 K_2 \int_{|x| > \sqrt{c_n}} c_f(x) dx \frac{1}{n^2} \sum_{k=1}^{n-1} \frac{1}{d_{k,0,n}} \max_{1 \leq k \leq n-1} \sum_{l=k+1}^n \frac{1}{d_{l,k,n}} \\
& + 8K K_2^2 \frac{1}{n^2} \sum_{k=1}^{n-1} \frac{1}{d_{k,0,n}} \max_{1 \leq k \leq n-1} \sum_{l=k+1}^n \frac{1}{d_{l,k,n}} \\
& \times \sup_{|u| < 2C(1+|z|)\sqrt{\epsilon}} |h_{l,k,n}(u) - h_{l,k,n}(0)|,
\end{aligned}$$

in which we have used Assumption A.3(c) that $|f(\frac{l}{n}, y)| \leq c_f(y)$ and the fact that

$$\int_{|y| \leq \sqrt{c_n}} \left| f\left(\frac{l}{n}, y\right) \right| dy \leq g\left(\frac{l}{n}\right) \leq K_2.$$

In view of Assumptions A.1 (c) and A.3, by virtue of the dominated convergence theorem, $\Lambda_{2n}(\epsilon) \rightarrow 0$ as $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$. This finishes the proof of the continuous case.

The proof of the discrete case is quite similar to that of the continuous case. Some critical steps are given as follows. Let $\mathcal{A}_{k,n}$ be the set of points that $x_{k,n}$ assumes. Suppose that the points are equally distributed on $\mathbb{R}$ with distance Δ . In what follows, define $\mathcal{B}_{k,n} := c_n d_{k,0,n} \mathcal{A}_{k,n} := \{c_n d_{k,0,n} a : a \in \mathcal{A}_{k,n}\}$. Then,

$$\begin{aligned}
E|Y_{k,n}(z)| &= E \left| f\left(\frac{k}{n}, c_n x_{k,n}\right) - f\left(\frac{k}{n}, c_n(x_{k,n} + z\epsilon)\right) \right| \\
&= \sum_{x \in \mathcal{A}_{k,n}} \left| f\left(\frac{k}{n}, c_n d_{k,0,n} x\right) - f\left(\frac{k}{n}, c_n(d_{k,0,n} x + z\epsilon)\right) \right| P_{k,0,n}(x) \\
&= \sum_{x \in \mathcal{B}_{k,n}} \left| f\left(\frac{k}{n}, x\right) - f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| P_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\
&\leq \sum_{x \in \mathcal{B}_{k,n}} \left| f\left(\frac{k}{n}, x\right) \right| + \sum_{x \in \mathcal{B}_{k,n}} \left| f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| \\
&= \frac{1}{c_n d_{k,0,n} \Delta} \sum_{x \in \mathcal{B}_{k,n}} \left| f\left(\frac{k}{n}, x\right) \right| c_n d_{k,0,n} \Delta \\
&\quad + \frac{1}{c_n d_{k,0,n} \Delta} \sum_{x \in \mathcal{B}_{k,n}} \left| f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| c_n d_{k,0,n} \Delta \\
&\leq \frac{1}{c_n d_{k,0,n} \Delta} \left(\int \left| f\left(\frac{k}{n}, x\right) \right| dx + \int \left| f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| dx \right) \\
&= \frac{2}{c_n d_{k,0,n} \Delta} \int \left| f\left(\frac{k}{n}, x\right) \right| dx = \frac{2}{c_n d_{k,0,n} \Delta} g\left(\frac{k}{n}\right) \leq \frac{2}{c_n d_{k,0,n} \Delta} K_2,
\end{aligned}$$

where we may modify the function f , e.g. $f^o(\cdot, x) = \max_{y \geq x} |f(\cdot, y)|$ for $x > 0$ to get the inequality in the derivation and note that the result above is similar to (D.5).

Following the same arguments as before, to complete the proof, it suffices to show both $\Lambda_{1n}(\epsilon)$ and $\Lambda_{2n}(\epsilon)$ converge to zero. While $\Lambda_{1n}(\epsilon) \rightarrow 0$ is easy to obtain, the key step in the proof of $\Lambda_{2n}(\epsilon) \rightarrow 0$ is the evaluation of the following conditional expectation:

$$\begin{aligned}
|E(Y_{l,n}(z)|\mathcal{F}_{k,n})| &= \left| E \left[f\left(\frac{l}{n}, c_n x_{l,n}\right) - f\left(\frac{l}{n}, c_n(x_{l,n} + z\epsilon)\right) \middle| \mathcal{F}_{k,n} \right] \right| \\
&= \left| E \left[f\left(\frac{l}{n}, c_n x_{k,n} + c_n(x_{l,n} - x_{k,n})\right) - f\left(\frac{l}{n}, c_n x_{k,n} + c_n(x_{l,n} - x_{k,n}) + c_n z\epsilon\right) \middle| \mathcal{F}_{k,n} \right] \right| \\
&= \left| \int \left[f\left(\frac{l}{n}, c_n x_{k,n} + c_n d_{l,k,n} y\right) - f\left(\frac{l}{n}, c_n x_{k,n} + c_n d_{l,k,n} y + c_n z\epsilon\right) \right] dF_{l,k,n}(y) \right| \\
&= \left| \int f\left(\frac{l}{n}, y\right) dF_{l,k,n}\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) - \int f\left(\frac{l}{n}, y\right) dF_{l,k,n}\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right) \right| \\
&= \left| \int f\left(\frac{l}{n}, y\right) d \left[F_{l,k,n}\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) - F_{l,k,n}\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right) \right] \right| \\
&= \left| \int f\left(\frac{l}{n}, y\right) dQ(y, c_n x_{k,n}) \right|,
\end{aligned}$$

where $Q(y, c_n x_{k,n}) = F_{l,k,n}\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) - F_{l,k,n}\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right)$.

Thus, we have

$$|E(Y_{l,n}(z)|\mathcal{F}_{k,n})|$$

$$\leq \begin{cases} \frac{2}{c_n d_{l,k,n} \Delta} \int |f(\frac{l}{n}, y)| dy = \frac{2}{c_n d_{l,k,n}} g(\frac{l}{n}), & \text{if } (l, k) \notin \Omega_n, \\ \frac{2}{c_n d_{l,k,n} \Delta} \int_{|y| > \sqrt{c_n}} |f(\frac{l}{n}, y)| dy + \left| \int_{|y| \leq \sqrt{c_n}} f(\frac{l}{n}, y) dQ(y, c_n x_{k,n}) \right|, & \text{if } (l, k) \in \Omega_n, \end{cases}$$

where $\Omega_n = \Omega_n(\epsilon)$ is the same as in Assumption A.1(c).

Then, an important step is to deal with the following expectation:

$$\begin{aligned} & E|Y_{k,n}| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, c_n x_{k,n}) \right| \\ &= E \left| f\left(\frac{k}{n}, c_n x_{k,n}\right) - f\left(\frac{k}{n}, c_n(x_{k,n} + z\epsilon)\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, c_n x_{k,n}) \right| \\ &= \int \left| f\left(\frac{k}{n}, c_n d_{k,0,n} x\right) - f\left(\frac{k}{n}, c_n(d_{k,0,n} x + z\epsilon)\right) \right| \\ &\quad \times \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, c_n d_{k,0,n} x) \right| dF_{k,0,n}(x) \\ &= \int \left| f\left(\frac{k}{n}, x\right) - f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\ &\leq \int \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\ &\quad + \int \left| f\left(\frac{k}{n}, x + c_n z\epsilon\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\ &= \int \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\ &\quad + \int \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x - c_n z\epsilon) \right| dF_{k,0,n}\left(\frac{x - c_n z\epsilon}{c_n d_{k,0,n}}\right) \\ &= \int_{|x| \leq \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\ &\quad + \int_{|x| > \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right) \\ &\quad + \int_{|x| \leq \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x - c_n z\epsilon) \right| dF_{k,0,n}\left(\frac{x - c_n z\epsilon}{c_n d_{k,0,n}}\right) \\ &\quad + \int_{|x| > \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x - c_n z\epsilon) \right| dF_{k,0,n}\left(\frac{x - c_n z\epsilon}{c_n d_{k,0,n}}\right) \\ &:= \sum_{i=1}^4 T_i(l, k; n). \end{aligned}$$

For $T_1(l, k; n)$ is similar to $T_3(l, k; n)$, and $T_2(l, k; n)$ is similar to $T_4(l, k; n)$, we need only to consider $T_1(l, k; n)$ and $T_2(l, k; n)$:

$$T_1(l, k; n) = \int_{|x| \leq \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| \left| \int_{|y| \leq \sqrt{c_n}} f\left(\frac{l}{n}, y\right) dQ(y, x) \right| dF_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right)$$

$$= \sum_{\substack{|x| \leq \sqrt{c_n} \\ x \in \mathcal{B}_{k,n}}} \left| f\left(\frac{k}{n}, x\right) \right| \left| \sum_{\substack{|y| \leq \sqrt{c_n} \\ y \in x + c_n z \epsilon \mathcal{B}_{l,k,n}}} f\left(\frac{l}{n}, y\right) P(y, x) \right| P_{k,0,n}\left(\frac{x}{c_n d_{k,0,n}}\right),$$

where $\mathcal{B}_{l,k,n} = \{c_n d_{l,k,n} a : a \in \mathcal{A}_{l,k,n}\}$, in which $\mathcal{A}_{l,k,n}$ is the set of points that $(x_{l,n} - x_{k,n})/d_{l,k,n}$ assumes.

Let $P(y, x) = P_{l,k,n}\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) - P_{l,k,n}\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right)$. Notice that when $|x| \leq \sqrt{c_n}$, $|y| \leq \sqrt{c_n}$ and $(l, k) \in \Omega_n(\epsilon)$, we have

$$\begin{aligned} |P(y, x)| &= \left| P_{l,k,n}\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) - P_{l,k,n}\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right) \right| \\ &= \left| F_{l,k,n}\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) - F_{l,k,n}^-\left(\frac{y - c_n x_{k,n}}{c_n d_{l,k,n}}\right) \right. \\ &\quad \left. - F_{l,k,n}\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right) + F_{l,k,n}^-\left(\frac{y - c_n x_{k,n} - c_n \epsilon z}{c_n d_{l,k,n}}\right) \right| \\ &\leq 4 \sup_{|u| < 2C(1+|z|)\sqrt{\epsilon}} |F_{l,k,n}(u) - F_{l,k,n}(0)|, \end{aligned}$$

where $F_{l,k,n}^-(\cdot)$ denotes the left limit of the function at the point.

Therefore, we have

$$\begin{aligned} T_1(l, k; n) &\leq \frac{4}{c_n d_{k,0,n} \Delta} \frac{1}{c_n d_{l,k,n} \Delta} g\left(\frac{k}{n}\right) g\left(\frac{l}{n}\right) \\ &\quad \times \sup_{|u| < 2C(1+|z|)\sqrt{\epsilon}} |F_{l,k,n}(u) - F_{l,k,n}(0)|. \end{aligned}$$

Regarding $T_2(l, k; n)$, we directly have

$$\begin{aligned} T_2(l, k; n) &\leq \frac{2}{c_n d_{k,0,n} \Delta} \frac{1}{c_n d_{l,k,n} \Delta} \int_{|x| \geq \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| dx \int_{|y| \leq \sqrt{c_n}} \left| f\left(\frac{k}{n}, y\right) \right| dy \\ &\leq \frac{2}{c_n d_{k,0,n} \Delta} \frac{1}{c_n d_{l,k,n} \Delta} g\left(\frac{l}{n}\right) \int_{|x| \geq \sqrt{c_n}} \left| f\left(\frac{k}{n}, x\right) \right| dx. \end{aligned}$$

As can be seen, the derivation for each term in $\Lambda_{2n}(\epsilon)$ in the discrete case has almost the same evaluation as for the corresponding term in the continuous case. So that we have shown that $\Lambda_{2n}(\epsilon)$ is asymptotically negligible. As a sequence, we have completed the whole proof. $\square$

Lemma D.6. *Let Assumption A.3 hold. Then we have for any fixed $\epsilon > 0$,*

$$L_{n,\epsilon}^{(r)} - \int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{[nr]} f\left(\frac{k}{n}, y\right) \phi_{\epsilon}(x_{k,n}) dy \rightarrow_{a.s.} 0$$

uniformly in $r \in [0, 1]$ as $n \rightarrow \infty$.

PROOF OF LEMMA D.6. Observe that

$$L_{n,\epsilon}^{(r)} = \frac{c_n}{n} \sum_{k=1}^{[nr]} \int_{-\infty}^{\infty} f\left(\frac{k}{n}, c_n(x_{t,n} + z\epsilon)\right) \phi(z) dz$$

$$\begin{aligned}
&= \frac{c_n}{n} \sum_{k=1}^{[nr]} \int_{-\infty}^{\infty} f\left(\frac{k}{n}, y\right) \phi\left(\frac{y - c_n x_{t,n}}{c_n \epsilon}\right) \frac{1}{c_n \epsilon} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{[nr]} f\left(\frac{k}{n}, y\right) \phi_{\epsilon}\left(\frac{y}{c_n} - x_{t,n}\right) dy.
\end{aligned}$$

It follows that for any $M > 0$,

$$\begin{aligned}
&\left| L_{n,\epsilon}^{(r)} - \int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{[nr]} f\left(\frac{k}{n}, y\right) \phi_{\epsilon}(x_{k,n}) dy \right| \\
&\leq \int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^n \left| f\left(\frac{k}{n}, y\right) \right| \left| \phi_{\epsilon}\left(\frac{y}{c_n} - x_{t,n}\right) - \phi_{\epsilon}(x_{k,n}) \right| dy \\
&= \int_{|y|>M} + \int_{|y|\leq M} \frac{1}{n} \sum_{k=1}^n \left| f\left(\frac{k}{n}, y\right) \right| \left| \phi_{\epsilon}\left(\frac{y}{c_n} - x_{t,n}\right) - \phi_{\epsilon}(x_{k,n}) \right| dy \\
&:= \Gamma_{1n} + \Gamma_{2n}.
\end{aligned}$$

Notice that,

$$\Gamma_{1n} \leq \frac{2}{\sqrt{2\pi}\epsilon} \int_{|y|>M} \frac{1}{n} \sum_{k=1}^n \left| f\left(\frac{k}{n}, y\right) \right| dy \leq \frac{2}{\sqrt{2\pi}\epsilon} \int_{|y|>M} c_f(y) dy,$$

using Assumption A.3 (c). Due to the integrability of $c_f(y)$ on $\mathbb{R}$, one can choose large enough M such that $\Gamma_{1n} < \varepsilon$ for any given $\varepsilon > 0$.

Moreover, since $\phi'_{\epsilon}(x) = -\frac{x}{\sqrt{2\pi}\epsilon^3} e^{-x^2/2\epsilon^2}$ and $|\phi'_{\epsilon}(x)|$ is bounded by $\frac{1}{\sqrt{2\pi}\epsilon^2}$ on $\mathbb{R}$, we have

$$\left| \phi_{\epsilon}\left(\frac{y}{c_n} - x_{t,n}\right) - \phi_{\epsilon}(x_{t,n}) \right| = \left| \phi'_{\epsilon}(\xi) \left(-\frac{y}{c_n}\right) \right| \leq \frac{|y|}{\sqrt{2\pi}\epsilon^2 c_n},$$

where ξ is between $x_{t,n} - \frac{y}{c_n}$ and $x_{t,n}$. Therefore,

$$\begin{aligned}
\Gamma_{2n} &\leq \int_{|y|\leq M} \frac{1}{n} \sum_{k=1}^n \left| f\left(\frac{k}{n}, y\right) \right| \frac{|y|}{\sqrt{2\pi}\epsilon^2 c_n} dy \\
&\leq \frac{M}{\sqrt{2\pi}\epsilon^2 c_n} \frac{1}{n} \sum_{k=1}^n \int_{|y|\leq M} \left| f\left(\frac{k}{n}, y\right) \right| dy \leq \frac{M}{\sqrt{2\pi}\epsilon^2 c_n} \frac{1}{n} \sum_{k=1}^n g\left(\frac{k}{n}\right),
\end{aligned}$$

where $g(t) = \int_{-\infty}^{\infty} |f(t, x)| dx$ continuous on $[0, 1]$.

As $\frac{1}{n} \sum_{k=1}^n g\left(\frac{k}{n}\right) \leq \sup_{0 \leq t \leq 1} g(t) < \infty$ and $c_n \rightarrow \infty$ as $n \rightarrow \infty$, $\Gamma_{2n} \rightarrow 0$. The assertion follows. $\square$

D.3: Proofs of Lemmas B.1 & B.2

PROOF OF LEMMA B.1.

In view of Lemmas D.5 and D.6, we first investigate $\int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{[nr]} f\left(\frac{k}{n}, y\right) \phi_{\epsilon}(x_{k,n}) dy$. It follows from Assumptions A.1 (a) and A.3, the continuous mapping theorem and the occupation time formula Revuz and Yor (1999, p. 232) that

$$\int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{[nr]} f\left(\frac{k}{n}, y\right) \phi_{\epsilon}(x_{k,n}) dy = \frac{1}{n} \sum_{k=1}^{[nr]} \phi_{\epsilon}(x_{k,n}) G_1\left(\frac{k}{n}\right)$$

$$\begin{aligned}
&= \int_0^r G_1 \left(\frac{[nt]}{n} \right) \phi_\epsilon(x_{[nt],n}) dt - \frac{1}{n} G_1(0) \phi_\epsilon(0) + \frac{1}{n} G_1 \left(\frac{[nr]}{n} \right) \phi_\epsilon(x_{[nr],n}) \\
&\rightarrow_D \int_0^r G_1(t) \phi_\epsilon(W(t)) dt = \int_{-\infty}^{\infty} dy \int_0^r G_1(t) \phi_\epsilon(y) dL_W(t, y) \\
&= \int_{-\infty}^{\infty} dy \int_0^r G_1(t) \phi(y) dL_W(t, \epsilon y) \quad \text{as } \epsilon \rightarrow 0 \\
&\rightarrow_{a.s.} \int_{-\infty}^{\infty} dy \int_0^r G_1(t) \phi(y) dL_W(t, 0) = \int_0^r G_1(t) dL_W(t, 0).
\end{aligned}$$

This finishes the proof of (B.1). To prove (B.2), we first have

$$\begin{aligned}
&\sup_{0 \leq r \leq 1} \left| \int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{[nr]} f \left(\frac{k}{n}, y \right) \phi_\epsilon(x_{k,n}) dy - \int_0^r G_1(t) \phi_\epsilon(W(t)) dt \right| \\
&= \sup_{0 \leq r \leq 1} \left| \frac{1}{n} \sum_{k=1}^{[nr]} \phi_\epsilon(x_{k,n}) G_1 \left(\frac{k}{n} \right) - \int_0^r G_1(t) \phi_\epsilon(W(t)) dt \right| \\
&\leq \int_0^1 |G_1(t)| |\phi_\epsilon(x_{[nt],n}) - \phi_\epsilon(W(t))| dt + \frac{A}{n},
\end{aligned} \tag{D.7}$$

where A stands for the maximum value of the bounds of G_1 on $[0, 1]$ and ϕ_ϵ on $\mathbb{R}$. It follows that

$$\begin{aligned}
&\int_0^1 |G_1(t)| |\phi_\epsilon(x_{[nt],n}) - \phi_\epsilon(W(t))| dt \\
&\leq K_1 \int_0^1 |\phi_\epsilon(x_{[nt],n}) - \phi_\epsilon(W(t))| dt \leq \frac{K_1}{\sqrt{2\pi}\epsilon\epsilon^2} \sup_{0 \leq t \leq 1} |x_{[nt],n} - W(t)|,
\end{aligned}$$

where $K_1 = \max_{0 \leq t \leq 1} |G_1(t)|$.

Hence, using the dominated convergence theorem and Assumption A.1 (b), as $n \rightarrow \infty$, (D.7) converges in probability to zero. Then the assertion follows as $\epsilon \rightarrow 0$.

Now we turn to prove (B.3). To this end, it suffices to verify Assumptions 1–3 for Theorem 2.1 of Wang (2014, p. 512-513).

Assumptions 1 and 2 of Theorem 2.1 are fulfilled due to Assumption A.1(c) and Assumption A.2 (a-b). Precisely, because $(e_s, \mathcal{F}_{n,s})$ is a martingale difference sequence satisfying that $E(e_s^2 | \mathcal{F}_{n,s}) = \sigma_e^2$, Assumption 1 for e_s is fulfilled with 1 replaced by σ_e^2 . In addition, let $\eta_{s,n} = d_{s,s-1,n}^{-1}(x_{s,n} - x_{s-1,n})$, then it is readily to see from Assumption A.1(c) that $\eta_{s,n}$ satisfies the condition in Assumption 1. Assumption 2 is easily verified because $x_{s,n}$ is adapted to $\mathcal{F}_{n,s-1}$. For Assumption 3, define

$$G_n^2(r) := \frac{c_n}{n} \sum_{s=1}^{[nr]} f^2 \left(\frac{s}{n}, c_n x_{s,n} \right)$$

and it follows from (B.1) and Assumption A.2(c) that $(W_n(r), G_n^2(r)) \Rightarrow (W(r), \int_0^r G_2(t) dL_W(t, 0))$, where $G_2(t) = \int f^2(t, x) dx$.

Then, the assertion (B.3) holds directly by Theorem 2.1 of Wang (2014). $\square$

PROOF OF LEMMA B.2.

We begin to prove the result in (B.4). Write

$$\frac{1}{n} \sum_{s=1}^n f\left(\frac{s}{n}, x_{s,n}\right) = \int_0^1 f\left(\frac{[nr]}{n}, W_n(r)\right) dr + \frac{1}{n} [f(1, x_{nn}) - f(0, 0)],$$

since $x_{0n} = 0$. Thus, it suffices to show $\int_0^1 f\left(\frac{[nr]}{n}, W_n(r)\right) dr \rightarrow_D \int_0^1 f(r, W(r)) dr$.

Because of the condition (b) in the regularity definition, there exists a constant $c > 0$ such that $f(t, x)$ is continuous in x whenever $|x| > c$. Let $J = [-c - 1, c + 1]$. For any given $\epsilon > 0$, it follows from the regularity of f that there exist continuous functions $\underline{f}_\epsilon(r, x)$ and $\bar{f}_\epsilon(r, x)$ in x and $\delta > 0$ such that whenever $|x - y| < \delta$ on J , for each $r \in [0, 1]$,

$$\underline{f}_\epsilon(r, x) \leq f(r, y) \leq \bar{f}_\epsilon(r, x).$$

Note that when $x = y \in J$, we always have $\underline{f}_\epsilon(r, x) \leq f(r, x) \leq \bar{f}_\epsilon(r, x)$.

With the condition (c) in Assumption A.2, it follows from the so-called Skorohod-Dudley-Wichura representation theorem that there is a common probability space $(\Omega, \mathcal{F}, P)$ supporting (U_n^0, W_n^0) and (U_n, W_n) such that

$$(U_n^0, W_n^0) \stackrel{D}{=} (U_n, W_n) \quad \text{and} \quad (U_n^0, W_n^0) \rightarrow_{a.s.} (U, W), \quad (\text{D.8})$$

in $D[0, 1]^2$ with uniform topology.

Thus, under this probability space we may have $\sup_{0 \leq r \leq 1} |W_n(r) - W(r)| = o_P(1)$. Consequently, we have $\sup_{0 \leq r \leq 1} |W_n(r) - W(r)| < \delta$ almost surely for n large enough and small $\delta > 0$.

Denote $A(r) = \{|W(r)| < c + 1\}$. It follows that on $A(r)$, when n is large, $W_n(r), W(r) \in J$; while on $\bar{A}(r)$, the complement set of $A(r)$, $|W_n(r)| > c$, $|W(r)| > c$. Notice by Definition B.1 that

$$\begin{aligned} & \left| \int_0^1 \left[f\left(\frac{[nr]}{n}, W_n(r)\right) - f(r, W(r)) \right] dr \right| \\ & \leq \left| \int_0^1 \left[f\left(\frac{[nr]}{n}, W_n(r)\right) - f(r, W_n(r)) \right] dr \right| + \left| \int_0^1 [f(r, W_n(r)) - f(r, W(r))] dr \right| \\ & \leq \left| \frac{[nr]}{n} - r \right| \int_0^1 L(W_n(r)) dr + \int_0^1 |f(r, W_n(r)) - f(r, W(r))| I(A(r)) dr \\ & \quad + \left| \int_0^1 [f(r, W_n(r)) - f(r, W(r))] I(\bar{A}(r)) dr \right| \\ & \leq \frac{1}{n} \int_0^1 L(W_n(r)) dr + \int_0^1 |\bar{f}_\epsilon(r, W(r)) - \underline{f}_\epsilon(r, W(r))| I(A(r)) dr \\ & \quad + \left| \int_0^1 [f(r, W_n(r)) - f(r, W(r))] I(\bar{A}(r)) dr \right| =: \Delta_1 + \Delta_2 + \Delta_3, \end{aligned}$$

where $I(\cdot)$ is the indicator function.

Moreover, it follows from the occupation time formula that

$$\begin{aligned} \Delta_2 &= \int_{-c-1}^{c+1} da \int_0^1 [\bar{f}_\epsilon(r, a) - \underline{f}_\epsilon(r, a)] dL_W(r, a) \\ &\leq \int_J \sup_{0 \leq r \leq 1} [\bar{f}_\epsilon(r, a) - \underline{f}_\epsilon(r, a)] da \int_0^1 dL_W(r, a) \end{aligned}$$

$$\leq \sup_{a \in J} L_W(1, a) \int_J \sup_{0 \leq r \leq 1} [\bar{f}_\epsilon(r, a) - \underline{f}_\epsilon(r, a)] da \rightarrow_{a.s.} 0,$$

as $\epsilon \rightarrow 0$, due to the regularity condition of f .

Furthermore, because $f(r, x)$ is continuous on $|x| > c$, the continuous mapping theorem implies that $\Delta_3 \rightarrow 0$ a.s.. Regarding Δ_1 , since $L(\cdot)$ satisfies Conditions (b) and (c) in the regularity definition, a similar derivation to Δ_2 yields that it approaches $\int_0^1 L(W(r))dr$ almost surely. Hence, the proof of (B.4) is completed.

We are ready to prove (B.5). Once again the embedding schedule described in the first part permits us to derive it under a stronger condition that $(U_n, W_n) \rightarrow_{a.s.} (W, U)$. Write

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{s=1}^n f\left(\frac{s}{n}, x_{s,n}\right) e_s &= \sum_{s=1}^n f\left(\frac{s}{n}, x_{s,n}\right) \frac{1}{\sqrt{n}} e_s \\ &= \int_0^1 f\left(\frac{[nr]}{n}, W_n(r)\right) dU_n(r) + \frac{1}{\sqrt{n}} (f(1, x_{nn})e_n - f(0, 0)e_0) := \sum_{k=1}^4 \Pi_k + o_P(1), \end{aligned}$$

where

$$\begin{aligned} \Pi_1 &= \int_0^1 \left[f\left(\frac{[nr]}{n}, W_n(r)\right) - f(r, W_n(r)) \right] dU_n(r), & \Pi_2 &= \int_0^1 [f(r, W_n(r)) - f_\epsilon(r, W_n(r))] dU_n(r), \\ \Pi_3 &= \int_0^1 f_\epsilon(r, W_n(r)) dU_n(r) - \int_0^1 f_\epsilon(r, W(r)) dU(r), & \Pi_4 &= \int_0^1 f_\epsilon(r, W(r)) dU(r), \end{aligned}$$

in which we denote $f_\epsilon(r, x) = \bar{f}_\epsilon(r, x)$ or $\underline{f}_\epsilon(r, x)$ for notational convenience.

Observe that $(f_\epsilon(r, W_n(r)), U_n(r)) \rightarrow (f_\epsilon(r, W(r)), U(r))$ almost surely due to continuity in x of f_ϵ . It follows from Theorem 2.2 in Kurtz and Protter (1991) that $\Pi_3 \rightarrow_P 0$ as $n \rightarrow \infty$.

Therefore, in order to finish the proof, we need to show: i) $\Pi_1 \rightarrow_P 0$ as $n \rightarrow \infty$; and ii) for all large n , $\Pi_2 \rightarrow_P 0$ and $\Pi_4 \rightarrow_P \int_0^1 f(r, W(r)) dU(r)$ as $\epsilon \rightarrow 0$. Let us investigate them term by term.

It follows from Assumption A.2 and the regularity condition that

$$\begin{aligned} E[\Pi_1]^2 &= \sigma_\epsilon^2 E \int_0^1 \left[f\left(\frac{[nr]}{n}, W_n(r)\right) - f(r, W_n(r)) \right]^2 dr \\ &\leq \frac{1}{n^2} \sigma_\epsilon^2 E \int_0^1 L^2(W_n(r)) dr \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$, because we have $\int_0^1 L^2(W_n(r)) dr \rightarrow_{a.s.} \int_0^1 L^2(W(r)) dr$, and by virtue of the regularity, $L^2(W_n(r))$ can be dominated by $L_\epsilon^2(W(r))$ when n is large for some $\epsilon > 0$ and $L_\epsilon(\cdot)$ is continuous, $E \int_0^1 L^2(W_n(r)) dr \rightarrow E \int_0^1 L^2(W(r)) dr < \infty$. This finishes the proof of $\Pi_1 = o_P(1)$.

The convergence of Π_2 and Π_4 can be proven at the same time if we show

$$\int_0^1 [f(r, W_n(r)) - f_\epsilon(r, W_n(r))] dU_n(r) \rightarrow_P 0,$$

as $\epsilon \rightarrow 0$ and for all large enough n , including the case of $n = \infty$ that means conventionally $(U_\infty(r), V_\infty(r)) = (U(r), V(r))$.

All notations $c, \epsilon, \delta, J, A(r), \bar{f}_\epsilon(t, x)$ and $\underline{f}_\epsilon(t, x)$ remain the same as in the first part. In view of regularity condition (b), we may find $\bar{f}_\epsilon(r, x)$ and $\underline{f}_\epsilon(r, x)$ such that they are continuous in x on $\mathbb{R}$ for

each $r \in [0, 1]$, since beyond $[-c, c]$, we can take $\bar{f}_\epsilon(r, x) = \underline{f}_\epsilon(r, x) = f(t, x)$ and due to this reason, $\sup_{r \in [0, 1]} (\bar{f}_\epsilon(r, x) - \underline{f}_\epsilon(r, x))$ is bounded on $\mathbb{R}$. Let C be the upper bound of $\sup_{r \in [0, 1]} [\bar{f}_\epsilon(r, x) - \underline{f}_\epsilon(r, x)]$.

By the adaptivity of $(U_n(r), W_n(r))$, for large n ,

$$\begin{aligned} & E \left\{ \int_0^1 [f(r, W_n(r)) - f_\epsilon(r, W_n(r))] dU_n(r) \right\}^2 \\ &= \sigma_e^2 E \int_0^1 [f(r, W_n(r)) - f_\epsilon(r, W_n(r))]^2 dr \\ &= \sigma_e^2 E \int_0^1 [f(r, W_n(r)) - f_\epsilon(r, W_n(r))]^2 I(A(r)) dr \\ &\leq \sigma_e^2 E \int_0^1 [\bar{f}_\epsilon(r, W_n(r)) - \underline{f}_\epsilon(r, W_n(r))]^2 I(A(r)) dr \\ &\rightarrow \sigma_e^2 E \int_0^1 [\bar{f}_\epsilon(r, W(r)) - \underline{f}_\epsilon(r, W(r))]^2 I(A(r)) dr, \end{aligned}$$

by virtue of continuity and boundedness of $\bar{f}_\epsilon(t, x) - \underline{f}_\epsilon(t, x)$ in x as $n \rightarrow \infty$. Observe that by the occupation formula, as $\epsilon \rightarrow 0$,

$$\begin{aligned} & \int_0^1 [\bar{f}_\epsilon(r, W(r)) - \underline{f}_\epsilon(r, W(r))]^2 I(|W(r)| \leq c+1) dr \\ &= \int_{-\infty}^{\infty} da \int_0^1 [\bar{f}_\epsilon(r, a) - \underline{f}_\epsilon(r, a)]^2 I(|a| \leq c+1) dL_W(r, a) \\ &\leq C \sup_a L_W(1, a) \int_J \sup_{0 \leq r \leq 1} [\bar{f}_\epsilon(r, a) - \underline{f}_\epsilon(r, a)] da \rightarrow_{a.s.} 0. \end{aligned}$$

It follows from the dominated convergence theorem that $\Pi_2 \rightarrow_P 0$ and Π_4 converges to the desired variable in probability as $\epsilon \rightarrow 0$. $\square$

Appendix E: Existence of orthogonal polynomial systems associated with Lévy processes

Let $Z(t)$ be a Lévy process and $\rho(t, x)$ be the density or probability distribution of $Z(t)$. Suppose that for every $t > 0$, $\rho(t, x)$ is defined on $x \in I \subset \mathbb{R}$. Let a and b be the boundary points of I (they may be infinite). Suppose further that there exist real polynomials $s(t, x)$ and $v(t, x)$ in x of degree at most 2 and 1, respectively, such that $\rho(t, x)$ satisfies for any $t > 0$,

$$\begin{aligned} (i) \quad & (s(t, x)\rho(t, x))' = v(t, x)\rho(t, x), \quad \text{if } Z(t) \text{ is continuous, or} \\ (i') \quad & \Delta(s(t, x)\rho(t, x)) = v(t, x)\rho(t, x), \quad \text{if } Z(t) \text{ is discrete, and} \\ (ii) \quad & s(t, x)\rho(t, x)x^k|_a^b = 0, \quad k = 0, 1, \dots, \end{aligned} \tag{E.1}$$

where and hereafter differentiation (prime) and difference (Δ) operators are conducted with respect to x only (not to t).

There are indeed several Lévy processes whose densities or probability distributions satisfy (E.1). For example, Brownian motion has density $\rho(t, x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t})$, $x \in \mathbb{R}$. Hence, this $\rho(t, x)$, along with $s(t, x) = t$ and $v(t, x) = -x$, satisfies the conditions. Gamma process has density $\rho(t, x) = \frac{1}{\Gamma(1+\alpha t)} x^{\alpha t} e^{-x}$, $x > 0$ and $\alpha > -1$. For Gamma process $s(t, x) = x$, $v(t, x) = 1 + \alpha t - x$. Moreover, a Poisson process

has probability distribution $\rho(t, x) = e^{-\mu t} \frac{(\mu t)^x}{x!}$, $x = 0, 1, 2, \dots$. This $\rho(t, x)$ satisfies (i') and (ii) in (E.1) with $s(t, x) = x$ and $v(t, x) = \mu t - x$.

Additionally, in the scope of application one may consider compound Poisson process. By definition, a compound Poisson process is given by

$$Z(t) = \sum_{i=1}^{N(t)} X_i$$

where $\{X_i\}$ is an i.i.d. sequence and $N(t)$ is a Poisson process independent of $\{X_i\}$. It is clear that $Z(t)$ describes the coming of $N(t)$ incidental events X_i sequentially over time period $[0, t]$, and accordingly is often used in insurance industry. Notice that Poisson process is a special case of the compound Poisson process, because if $X_i \equiv 1$, $Z(t) = N(t)$; In general, by the law of total probability the distribution function of $Z(t)$ is

$$\begin{aligned} P(Z(t) < z) &= \sum_{n=1}^{\infty} P(X_1 + \dots + X_n < z | N(t) = n) P(N(t) = n) \\ &= \sum_{n=1}^{\infty} P(X_1 + \dots + X_n < z) P(N(t) = n) \\ &= \sum_{n=1}^{\kappa} P(X_1 + \dots + X_n < z) P(N(t) = n) \\ &\quad + \sum_{n=\kappa+1}^{\infty} P(X_1 + \dots + X_n < z) P(N(t) = n) \\ &= \sum_{n=1}^{\kappa} P(X_1 + \dots + X_n < z) P(N(t) = n) \\ &\quad + \sum_{n=\kappa+1}^{\infty} P\left(\frac{1}{\sqrt{n}}(X_1 + \dots + X_n) < \frac{1}{\sqrt{n}}z\right) P(N(t) = n) \\ &= \sum_{n=1}^{\kappa} G^{*n}(z) P(N(t) = n) + \Phi(0) P(N(t) > \kappa), \end{aligned}$$

for a sufficient large κ , where $G^{*n}(z)$ is the n -times convolution of the distribution $G(\cdot)$ of X_1 , $\Phi(\cdot)$ is the distribution function of a standard normal variable, in which we have used the central limit theorem to imply that $\frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$ is approximately a normal random variable and $\frac{1}{\sqrt{n}}z \approx 0$ for fixed z and $n > \kappa$. Noting that $P(N(t) > \kappa)$ is the tail of the distribution of $N(t)$, and the distribution of $Z(t)$ is mainly determined by the first term. In particular, if X_1 has either an exponential distribution or a gamma distribution, the first term on the above follows a gamma distribution. This discussion shows the possibility of using the theoretical results in this paper for functionals of the compound Poisson process.

Given that $\rho(t, x)$ satisfies the conditions in (E.1), this section is to show the existence of orthogonal polynomial system associated $Z(t)$ and the explicit expressions of the polynomials. Note that the following derivations extend the case $\rho(x)$ in Nikiforov and Uvarov (1988) to $\rho(t, x)$. Also, the discussion has to be divided into the continuous case and the discrete case in order to be concert with the classification of Lévy processes.

E.1: Orthogonal polynomials of a continuous variable

Let $(Z(t), t \geq 0)$ be a Lévy process. Suppose that $Z(t)$ is a continuous random variable with distribution function $\Psi_t(x)$ for every $t > 0$, and $\frac{d}{dx}\Psi_t(x) = \rho(t, x)$. Let us now consider differential equation of hypergeometric type with parameter $t > 0$:

$$s(t, x)y''(t, x) + v(t, x)y'(t, x) + \lambda(t)y(t, x) = 0. \quad (\text{E.2})$$

where $s(t, x)$ and $v(t, x)$ are polynomials in x of degree at most 2 and 1 respectively, while $\lambda(t)$ is independent of x . We shall refer to the solutions of (E.2) as functions of hypergeometric type. Note that in (E.2), and also in the sequel, all derivatives are conducted with respect to x , not to t .

Lemma E.1. *All derivatives of the solutions of (E.2) are still of hypergeometric type; meanwhile, any function of hypergeometric type is a derivative of some function of hypergeometric type with $\lambda(t) \neq 0$ in the differential equation.*

Proof. Differentiating (E.2) and denoting $z_1(t, x) = y'(t, x)$ entail that

$$s(t, x)z_1''(t, x) + v_1(t, x)z_1'(t, x) + \eta_1(t)z_1(t, x) = 0 \quad (\text{E.3})$$

where $v_1(t, x) = v(t, x) + s'(t, x)$ and $\eta_1(t) = \lambda(t) + v'(t, x)$.

Since $v_1(t, x)$ is a polynomial in x of degree at most 1 and η_1 is independent of x , equation (E.3) is of hypergeometric type.

Now let $z_1(t, x)$ be a solution of (E.3). We then have $v(t, x) = v_1(t, x) - s'(t, x)$ and $\lambda(t) = \eta_1(t) - v_1'(t, x) + s''(t, x) \neq 0$. Construct a function $y(t, x) = -\frac{1}{\lambda(t)}(s(t, x)z_1'(t, x) + v(t, x)z_1(t, x))$. Then

$$\lambda(t)y'(t, x) = -(s(t, x)z_1'(t, x) + v(t, x)z_1(t, x))' = \lambda(t)z_1(t, x),$$

by virtue of (E.3). Hence, $y'(t, x) = z_1(t, x)$ and the construction of $y(t, x)$ implies that $y(t, x)$ is a solution of equation (E.2).

It follows from induction that $z_i(t, x) = y^{(i)}(t, x)$ is a solution of

$$s(t, x)z_i''(t, x) + v_i(t, x)z_i'(t, x) + \eta_i(t)z_i(t, x) = 0 \quad (\text{E.4})$$

where $v_i(t, x) = v(t, x) + is'(t, x)$ and $\eta_i(t) = \lambda(t) + iv'(t, x) + \frac{i(i-1)}{2}s''(t, x)$.

Moreover, every solution of (E.4) for $\eta_k \neq 0$ ($k = 0, 1, \dots, i-1$) can be represented as $z_i(t, x) = y^{(i)}(t, x)$, where $y(t, x)$ is a solution of (E.2). $\square$

Observe that Lemma E.1 provides us with a possibility to find out a family of particular solutions of (E.2) according to a given $\lambda(t)$. Indeed, if $\eta_i(t) = \lambda(t) + iv'(t, x) + \frac{i(i-1)}{2}s''(t, x) = 0$, it is evident that equation (E.4) has a solution $z_i(t, x) = z_i(t)$ independent of x . Since $z_i(t) = y^{(i)}(t, x)$, we assert that when $\lambda(t) \equiv \lambda_i(t) = -iv'(t, x) - \frac{i(i-1)}{2}s''(t, x)$, $y(t, x) = y_i(t, x)$, as a solution of (E.2), is a polynomial in x of degree exactly i .

Notice that (E.4) is valid for $k = 0, 1, \dots, i$, that is, i can be substituted by any such k . Also, note that $\eta_0 = \lambda(t) \equiv \lambda_i(t)$, $\eta_k(t) = \lambda_i(t) - \lambda_k(t)$ for $k = 1, \dots, i$.

Denote $\rho_0(t, x) = \rho(t, x)$ and $\rho_k(t, x) = s^k(t, x)\rho(t, x)$ for $k \geq 1$. The relation $(s(t, x)\rho(t, x))' = v(t, x)\rho(t, x)$ implies $(s(t, x)\rho_k(t, x))' = v_k(t, x)\rho_k(t, x)$ where $v_k(t, x) = ks'(t, x) + v(t, x)$, and these enable us to write (E.2) and (E.4) in self-adjoint form

$$(s(t, x)\rho(t, x)y'(t, x))' + \lambda(t)\rho(t, x)y(t, x) = 0, \quad (\text{E.5a})$$

$$(s(t, x)\rho_k(t, x)z'_k(t, x))' + \eta_k(t)\rho_k(t, x)z_k(t, x) = 0. \quad (\text{E.5b})$$

Observing that $s(t, x)\rho_k(t, x) = \rho_{k+1}(t, x)$ and $z'_k(t, x) = z_{k+1}(t, x)$, it follows from (E.5b) that

$$\rho_k(t, x)z_k(t, x) = -\frac{1}{\eta_k}(\rho_{k+1}(t, x)z_{k+1}(t, x))'. \quad (\text{E.6})$$

Hence, when $k < i$, we obtain successively

$$\begin{aligned} \rho_k(t, x)z_k(t, x) &= -\frac{1}{\eta_k}(\rho_{k+1}(t, x)z_{k+1}(t, x))' = \dots = (-1)^{i-k} \frac{1}{\eta_k \dots \eta_{i-1}} (\rho_i(t, x)z_i(t, x))^{(i-k)} \\ &= \frac{A_k}{A_i} (\rho_i(t, x)z_i(t, x))^{(i-k)}, \end{aligned}$$

where we denote $A_i = (-1)^i \prod_{j=0}^{i-1} \eta_j$ for $i \geq 1$ and $A_0 = 1$.

Since $z_k = y^{(k)}$ and $z_i = y^{(i)} = 1$ (without loss of generality), we finally have

$$y_i^{(k)}(t, x) = \frac{A_k}{A_i} \frac{1}{\rho_k(t, x)} [\rho_i(t, x)]^{(i-k)}. \quad (\text{E.7})$$

In particular, when $k = 0$, we have an explicit representation of $y_i(t, x)$ of hypergeometric type

$$y_i(t, x) = \frac{1}{A_i} \frac{1}{\rho(t, x)} \frac{d^i}{dx^i} [s^i(t, x)\rho(t, x)]. \quad (\text{E.8})$$

The expressions (E.8) of polynomials are called the Rodrigues formula.

Lemma E.2. Suppose that polynomials $y_i(t, x)$ in x are the solutions of equation of hypergeometric type

$$s(t, x)y''(t, x) + v(t, x)y'(t, x) + \lambda_i(t)y(t, x) = 0, \quad (\text{E.9})$$

where $\lambda_i(t) = -iv'(t, x) - \frac{i(i-1)}{2}s''(t, x)$. In addition, suppose a density function $\rho(t, x)$, along with $(s(t, x)$ and $v(t, x)$, satisfies conditions (i) and (ii) in (E.1). Then the $y_i(t, x)$ are orthogonal on (a, b) with respect to $\rho(t, x)$, where a and b are boundary points (maybe infinity) of the support of $\rho(t, x)$.

Proof. Observe that $y_m(t, x)$ and $y_i(t, x)$ satisfy the following differential equations in self-adjoint form respectively

$$(s(t, x)\rho(t, x)y'_i(t, x))' + \lambda_i(t)\rho(t, x)y_i(t, x) = 0, \quad (\text{E.10a})$$

$$(s(t, x)\rho(t, x)y'_m(t, x))' + \lambda_m(t)\rho(t, x)y_m(t, x) = 0. \quad (\text{E.10b})$$

Operation (E.10b) $\times y_i$ - (E.10a) $\times y_m$ yields

$$(\lambda_m(t) - \lambda_i(t))\rho(t, x)y_m(t, x)y_i(t, x)$$

$$=(s(t, x)\rho(t, x)y_i'(t, x))'y_m(t, x) - (s(t, x)\rho(t, x)y_m'(t, x))'y_i(t, x).$$

Integration by parts gives

$$\begin{aligned} & (\lambda_m(t) - \lambda_i(t)) \int_a^b \rho(t, x)y_m(t, x)y_i(t, x)dx \\ &= s(t, x)\rho(t, x)y_i'(t, x)y_m(t, x)|_a^b - s(t, x)\rho(t, x)y_m'(t, x)y_i(t, x)|_a^b = 0 \end{aligned}$$

due to the boundary condition. Hence, if $\lambda_m(t) \neq \lambda_i(t)$, then $y_i(t, x)$ are of orthogonality with weight $\rho(t, x)$. $\square$

Lemma E.3. *For orthogonal polynomials $y_i(t, x)$ of hypergeometric type with weight $\rho(t, x)$*

$$s(t, x)y''(t, x) + v(t, x)y'(t, x) + \lambda_i(t)y(t, x) = 0, \quad (\text{E.11})$$

where $\lambda_i(t) = -iv'(t, x) - \frac{1}{2}i(i-1)s''(t, x)$, the derivatives $y_i^{(k)}(t, x)$ are orthogonal with respect to $\rho_k(t, x) = s^k(t, x)\rho(t, x)$ on (a, b) .

Proof. According to Lemma E.1, because $y_i(t, x)$ satisfies equation (E.11), $y_i^{(k)}(t, x)$ is a solution of the following differential equation

$$s(t, x)[y_i^{(k)}(t, x)]'' + v_k(t, x)[y_i^{(k)}(t, x)]' + \eta_{ki}y_i^{(k)}(t, x) = 0, \quad (\text{E.12})$$

where $v_k(t, x) = v(t, x) + ks'(t, x)$ and $\eta_{ki}(t) = \lambda_i(t) - \lambda_k(t)$.

Similar to (E.5), we have a self-adjoint form of equation (E.12)

$$\{s(t, x)\rho_k(t, x)[y_i^{(k)}(t, x)]'\}' + \eta_{ki}\rho_k(t, x)y_i^{(k)}(t, x) = 0. \quad (\text{E.13})$$

Suppose, on the other hand, that $y_m^{(k)}(t, x)$ is the k -th derivative of $y_m(t, x)$ of hypergeometric type corresponding to λ_m , then we also have a similar equation to (E.13). Aligning them together gives

$$\{s(t, x)\rho_k(t, x)[y_i^{(k)}(t, x)]'\}' + \eta_{ki}\rho_k(t, x)y_i^{(k)}(t, x) = 0, \quad (\text{E.14a})$$

$$\{s(t, x)\rho_k(t, x)[y_m^{(k)}(t, x)]'\}' + \eta_{km}\rho_k(t, x)y_m^{(k)}(t, x) = 0. \quad (\text{E.14b})$$

It follows from (E.14) that $\int_a^b \rho_k(t, x)y_m^{(k)}(t, x)y_i^{(k)}(t, x)dx = 0$ when $\lambda_m \neq \lambda_i$. In other words,

$$\int_a^b \rho_k(t, x)y_m^{(k)}(t, x)y_i^{(k)}(t, x)dx = \delta_{mi}d_{ki}^2$$

, where $d_{ki}^2(t) := \int_a^b \rho_k(t, x)[y_i^{(k)}(t, x)]^2dx$. $\square$

For Lévy process $Z(t)$, if $\rho(t, x)$ satisfies the conditions in (E.1), there is an orthogonal polynomial system $y_i(t, x)$ with weight $\rho(t, x)$. Let us define $Q_i(t, x)$ and $Q_{ki}(t, x)$ by

$$Q_i(t, x) = \frac{1}{d_i(t)}y_i(t, x) \quad \text{and} \quad Q_{ki}(t, x) = \frac{1}{d_{ki}(t)}D^k y_i(t, x), \quad (\text{E.15})$$

each of which is an orthonormal polynomial system with $\rho(t, x)$ or $\rho_k(t, x)$ being its weight. In such a situation, we say that $Z(t)$ admits a *classical orthonormal polynomial system* $Q_i(t, x)$.

Example E.1

(1) If $Z(t) = B(t)$ is a Brownian motion with $\rho(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$, the corresponding hypergeometric differential equations are

$$ty''(t, x) - xy'(t, x) + iy(t, x) = 0.$$

Since $A_i = (-1)^i i!$, it follows from (E.8) that the polynomial solutions are

$$y_i(t, x) = \frac{(-1)^i t^i}{i!} \frac{1}{\rho(t, x)} \rho^{(i)}(t, x) = \frac{1}{\sqrt{i!}} H_i(x/\sqrt{t})$$

for $t > 0$ where $H_i(\cdot)$ are the Hermite polynomial sequence orthogonal with respect to $\exp(-x^2/2)$.

(2) If $Z(t) = G(t)$ is a Gamma process with $\rho(t, x) = \frac{1}{\Gamma(1+\alpha t)} x^{\alpha t} e^{-x}$, the corresponding hypergeometric differential equations are

$$xy''(t, x) + (\alpha t + 1 - x)y'(t, x) + iy(t, x) = 0.$$

It is clear that $A_i = (-1)^i i!$ and hence it follows from (E.8) that the polynomial solutions of the equations are

$$y_i(t, x) = \frac{(-1)^i}{i!} \frac{1}{\rho(t, x)} [x^i \rho(t, x)]^{(i)}$$

which in spirit are Laguerre polynomials (due to the involvement of t).

Let us find out the relationship between the squared norm $d_{ki}^2(t)$ of $y_i^{(k)}(t, x)$ and the squared norm $d_i^2(t) := d_{0i}^2(t)$ of $y_i(t, x)$. Rewrite equation (E.14a) as

$$[\rho_{k+1}(t, x) y_i^{(k+1)}(t, x)]' + \eta_{ki} \rho_k(t, x) y_i^{(k)}(t, x) = 0. \quad (\text{E.16})$$

Multiplying (E.16) by $y_i^{(k)}(t, x)$ and integrating by parts over (a, b) give

$$\begin{aligned} \eta_{ki} d_{ki}^2(t) &= \eta_{ki} \int_a^b \rho_k(t, x) [y_i^{(k)}(t, x)]^2 dx \\ &= - \int_a^b [\rho_{k+1}(t, x) y_i^{(k+1)}(t, x)]' y_i^{(k)}(t, x) dx = d_{k+1, i}^2(t). \end{aligned} \quad (\text{E.17})$$

Whence, by induction, we obtain

$$d_{ki}^2(t) = d_i^2(t) \prod_{j=0}^{k-1} \eta_{ji}(t), \quad (\text{E.18})$$

where $\eta_{0i}(t) = \lambda_i(t)$, $\eta_{ji}(t) = \lambda_j(t) - \lambda_i(t)$ and $\lambda_i(t) = -iv'(t, x) - \frac{i(i-1)}{2} s''(t, x)$.

E.2: Orthogonal polynomials of a discrete variable

First, we study the solutions of difference equation of hypergeometric type with parameter $t > 0$. Define difference operation $\triangle f(x) = f(x+1) - f(x)$ and $\nabla f(x) = f(x) - f(x-1)$. In what follows, the following identities are frequently utilised

$$\left\{ \begin{array}{l} \triangle f(x) = \nabla f(x+1), \\ \triangle \nabla f(x) = \nabla \triangle f(x) = f(x+1) - 2f(x) + f(x-1) \\ \triangle [f(x)g(x)] = f(x)\triangle g(x) + g(x+1)\triangle f(x) \end{array} \right. \quad (\text{E.19})$$

Note that in the sequel all difference operations are applied with respect to x only.

The difference equation of hypergeometric type takes the form

$$s(t, x)\Delta\nabla y(t, x) + v(t, x)\Delta y(t, x) + \lambda(t)y(t, x) = 0 \quad (\text{E.20})$$

where $s(t, x)$ and $v(t, x)$ are polynomials in x at most second and first degree respectively, and $\lambda(t)$ is a constant relative to t .

Lemma E.4. *If $y(t, x)$ is a solution to (E.20), then $z_1(t, x) = \Delta y(t, x)$ is also a solution of some difference equation of hypergeometric type. If $\lambda(t) \neq 0$, the converse is also true.*

Proof. Applying Δ on (E.20) yields

$$\Delta[s(t, x)\nabla z_1(t, x)] + \Delta[v(t, x)z_1(t, x)] + \lambda(t)z_1(t, x) = 0. \quad (\text{E.21})$$

Moreover, by (E.19),

$$\begin{aligned} \Delta[s(t, x)\nabla z_1(t, x)] &= s(t, x)\Delta\nabla z_1(t, x) + \Delta s(t, x)\nabla z_1(t, x+1) \\ &= s(t, x)\Delta\nabla z_1(t, x) + \Delta s(t, x)\Delta z_1(t, x), \\ \Delta[v(t, x)z_1(t, x)] &= z_1(t, x)\Delta v(t, x) + v(t, x+1)\Delta z_1(t, x). \end{aligned}$$

It thus follows that

$$s(t, x)\Delta\nabla z_1(t, x) + v_1(t, x)\Delta z_1(t, x) + \eta_1(t)z_1(t, x) = 0, \quad (\text{E.22})$$

where $v_1(t, x) = v(t, x+1) + \Delta s(t, x)$ and $\eta_1(t) = \lambda(t) + \Delta v(t, x)$.

Apparently, $v_1(t, x)$ is a polynomial in x of degree at most 1 and $\eta_1(t)$ is independent of x . Therefore, (E.22) is of the same form as (E.20).

If $\lambda(t) \neq 0$ (which can be represented by $\eta_1(t)$, $v_1(t, x)$ and $s(t, x)$), for each solution $z_1(t, x)$ of (E.22), construct

$$y(t, x) = -\frac{1}{\lambda(t)}[s(t, x)\nabla z_1(t, x) + v(t, x)z_1(t, x)].$$

Then, it is easy to verify that $\Delta y(t, x) = z_1(t, x)$ and $y(t, x)$ satisfies equation (E.20). In fact, since (E.22) is equivalent to (E.21), $-\lambda(t)\Delta y(t, x) = \Delta[s(t, x)\nabla z_1(t, x) + v(t, x)z_1(t, x)] = -\lambda(t)z_1(t, x)$, hence, $\Delta y(t, x) = z_1(t, x)$. Moreover,

$$\begin{aligned} &s(t, x)\Delta\nabla y(t, x) + v(t, x)\Delta y(t, x) + \lambda(t)y(t, x) \\ &= s(t, x)\nabla z_1(t, x) + v(t, x)z_1(t, x) + \lambda(t)y(t, x) = -\lambda(t)y(t, x) + \lambda(t)y(t, x) = 0 \end{aligned}$$

by the definition of $y(t, x)$. The proof is finished. $\square$

It follows from induction that $z_i(t, x) = \Delta^i y(t, x)$ satisfies a difference equation of hypergeometric type:

$$s(t, x)\Delta\nabla z_i(t, x) + v_i(t, x)\Delta z_i(t, x) + \eta_i(t)z_i(t, x) = 0, \quad (\text{E.23})$$

where

$$v_i(t, x) = v_{i-1}(t, x+1) + \Delta s(t, x), \quad v_0(t, x) = v(t, x) \quad (\text{E.24a})$$

$$\eta_i(t) = \eta_{i-1}(t) + \Delta v_{i-1}(t, x), \quad \eta_0(t) = \lambda(t). \quad (\text{E.24b})$$

The converse is also correct, viz., every solution $z_i(t, x)$ of (E.23) with $\eta_k \neq 0$ ($k = 0, 1, \dots, i-1$) can be rephrased as $z_i(t, x) = \Delta^i y(t, x)$ where $y(t, x)$ is a solution of (E.20).

Since the first part of (E.24a) may be rewritten as

$$v_i(t, x) + s(t, x) = v_{i-1}(t, x+1) + s(t, x+1), \quad (\text{E.25})$$

it is clear that $v_i(t, x) = v(t, x+i) + s(t, x+i) - s(t, x)$.

Observe that $\Delta v_i(t, x)$ and $\Delta^2 s(t, x)$ are independent of x . Thus, it follows from once again the first part of (E.24a) that $\Delta v_i(t, x) = \Delta v_{i-1}(t, x) + \Delta^2 s(t, x) = \dots = \Delta v(t, x) + i\Delta^2 s(t, x)$. Whence, the first part of (E.24b) implies $\eta_i(t) = \eta_{i-1}(t) + \Delta v_{i-1}(t, x) = \eta_{i-1}(t) + \Delta v(t, x) + (i-1)\Delta^2 s(t, x)$, which gives

$$\begin{aligned} \eta_i(t) &= \eta_0(t) + \sum_{k=1}^i [\eta_k(t) - \eta_{k-1}(t)] = \lambda(t) + \sum_{k=1}^i [\Delta v(t, x) + (k-1)\Delta^2 s(t, x)] \\ &= \lambda(t) + i\Delta v(t, x) + \frac{1}{2}i(i-1)\Delta^2 s(t, x) = \lambda(t) + iv'(t, x) + \frac{1}{2}i(i-1)s''(t, x). \end{aligned} \quad (\text{E.26})$$

Note that apparently if $\eta_i(t) = 0$ in equation (E.23), then $z_i(t, x) = \text{const.}$ is a solution of equation (E.23). Note also that $z_i(t, x) = \Delta^i y(t, x)$. That means when $\lambda(t) = \lambda_i(t) = -iv'(t, x) - \frac{1}{2}i(i-1)s''(t, x)$, equation (E.20) has a solution $y(t, x) = y_i(t, x)$ which is a polynomial in x of degree exactly i provided that $\eta_k \neq 0$ for $k = 0, 1, \dots, i-1$. Denote $\rho_0(t, x) = \rho(t, x)$ and $\rho_k(t, x) = \rho(t, x+k) \prod_{j=1}^k s(t, x+j)$.

Then, the relation $\Delta(s(t, x)\rho(t, x)) = v(t, x)\rho(t, x)$ implies that $\Delta(s(t, x)\rho_k(t, x)) = v_k(t, x)\rho_k(t, x)$. In order to obtain explicit solution $y_i(t, x)$, we rewrite equations (E.20) and (E.23) in a self-adjoint form

$$\Delta(s(t, x)\rho(t, x)\nabla y(t, x)) + \lambda(t)\rho(t, x)y(t, x) = 0, \quad (\text{E.27})$$

$$\Delta(s(t, x)\rho_k(t, x)\nabla z_k(t, x)) + \eta_k(t)\rho_k(t, x)z_k(t, x) = 0. \quad (\text{E.28})$$

It follows from equation (E.28) that

$$\begin{aligned} \rho_k(t, x)z_k(t, x) &= -\frac{1}{\eta_k(t)}\Delta(s(t, x)\rho_k(t, x)\nabla z_k(t, x)) \\ &= -\frac{1}{\eta_k(t)}\nabla(s(t, x+1)\rho_k(t, x+1)\Delta z_k(t, x)) \\ &= -\frac{1}{\eta_k(t)}\nabla(\rho_{k+1}(t, x)z_{k+1}(t, x)). \end{aligned}$$

For $k < i$, we obtain successively

$$\begin{aligned} \rho_k(t, x)z_k(t, x) &= -\frac{1}{\eta_k(t)}\nabla(\rho_{k+1}(t, x)z_{k+1}(t, x)) \\ &= \dots \\ &= \frac{A_k}{A_i}\nabla^{i-k}(\rho_i(t, x)z_i(t, x)), \end{aligned} \quad (\text{E.29})$$

where $A_i = (-1)^i \prod_{j=0}^{i-1} \eta_j(t)$ for $i \geq 1$ and $A_0 = 1$.

Without loss of generality let $z_i(t, x) = \Delta^i y_i(t, x) = 1$. We have

$$z_k(t, x) = \Delta^k y_i(t, x) = \frac{A_k}{A_i} \frac{1}{\rho_k(t, x)} \nabla^{i-k} [\rho_i(t, x)]. \quad (\text{E.30})$$

Particularly, an explicit expression, Rodrigues formula, of $y_i(t, x)$ is

$$y_i(t, x) = \frac{1}{A_i} \frac{1}{\rho(t, x)} \nabla^i [\rho_i(t, x)]. \quad (\text{E.31})$$

Lemma E.5. *Given that $\rho(t, x) > 0$, $s(t, x)$ and $v(t, x)$ satisfy the conditions (i') and (ii) in (E.1), the polynomial solutions $y_i(t, x)$ of the difference equation (E.20) with $\lambda(t) = \lambda_i(t) = -iv'(t, x) - \frac{1}{2}i(i-1)s''(t, x)$ are orthogonal on $[a, b-1]$ with weight $\rho(t, x)$*

$$\sum_{x_j=a}^{b-1} y_m(t, x_j) y_i(t, x_j) \rho(t, x_j) = \delta_{mi} d_i^2(t). \quad (\text{E.32})$$

Similarly, $\Delta^k y_i(t, x)$ are orthogonal with respect to $\rho_k(t, x)$:

$$\sum_{x_j=a}^{b-k} \Delta^k y_m(t, x_j) \Delta^k y_j(t, x_j) \rho_k(t, x_j) = \delta_{mi} d_{ki}^2(t). \quad (\text{E.33})$$

The orthogonal polynomial system $\{y_i(t, x)\}$ with $\rho(t, x)$ satisfying conditions in Lemma E.5 is called *classic orthogonal polynomial system* of discrete variable.

Proof. The equations for $y_i(t, x)$ and $y_m(t, x)$ in self-adjoint form are

$$\Delta(s(t, x)\rho(t, x)\nabla y_i(t, x)) + \lambda_i(t)\rho(t, x)y_i(t, x) = 0, \quad (\text{E.34})$$

$$\Delta(s(t, x)\rho(t, x)\nabla y_m(t, x)) + \lambda_m(t)\rho(t, x)y_m(t, x) = 0. \quad (\text{E.35})$$

Multiply (E.34) by $y_m(t, x)$ and (E.35) by $y_i(t, x)$, subtract the second from the first. We have

$$\begin{aligned} & (\lambda_m(t) - \lambda_i(t))\rho(t, x)y_m(t, x)y_i(t, x) \\ &= y_m(t, x)\Delta(s(t, x)\rho(t, x)\nabla y_i(t, x)) - y_i(t, x)\Delta(s(t, x)\rho(t, x)\nabla y_m(t, x)) \\ &= \Delta[s(t, x)\rho(t, x)(y_m(t, x)\nabla y_i(t, x) - y_i(t, x)\nabla y_m(t, x))], \end{aligned}$$

on account of (E.19).

If we put $x = x_j$ and sum them over j

$$\begin{aligned} & (\lambda_m(t) - \lambda_i(t)) \sum_j y_m(t, x_j) y_i(t, x_j) \rho(t, x_j) \\ &= s(t, x)\rho(t, x)(y_m(t, x)\nabla y_n(t, x) - y_n(t, x)\nabla y_m(t, x))|_a^b = 0 \end{aligned}$$

by virtue of the boundary condition. Hence, (E.32) is valid.

We are now in a position to demonstrate (E.33). We use induction. Notice that $\Delta y_i(t, x)$ is a solution of difference equation which can be written in self-adjoint form with $\rho_1(t, x) = s(t, x+1)\rho(t, x+1) = (v(t, x) + s(t, x))\rho(t, x)$ by the condition that $\rho(t, x)$ satisfies.

Because $\rho(t, x)$ satisfies condition (ii) in (2.1), $\rho_1(t, x)$ satisfies a similar condition, viz.,

$$s(t, x)\rho_1(t, x)x^k|_{x=a,b} = 0, \quad \text{for } k = 0, 1, \dots.$$

Whence the polynomials $\Delta y_i(t, x)$ have orthogonality

$$\sum_{x_j=a}^{b-2} \Delta y_m(t, x_j) \Delta y_i(t, x_j) \rho_1(t, x_j) = \delta_{mi} d_{1i}^2(t).$$

Similarly, we have the relationship of orthogonality for $\Delta^k y_i(t, x)$,

$$\sum_{x_j=a}^{b-k-1} \Delta^k y_m(t, x_j) \Delta^k y_i(t, x_j) \rho_k(t, x_j) = \delta_{mi} d_{ki}^2(t).$$

□

Example E.2

If $Z(t) = N(t)$ is a Poisson process with $\rho(t, x) = e^{-\mu t} \frac{(\mu t)^x}{x!}$, $x = 0, 1, 2, \dots$, the corresponding hypergeometric difference equations are

$$x \Delta \nabla y(t, x) + (\mu t - x) \Delta y(t, x) + i y(t, x) = 0.$$

Since $A_i = (-1)^i i!$, the polynomial solutions are

$$y_i(t, x) = \frac{(-1)^i}{i!} \frac{1}{\rho(t, x)} \nabla^i \left[\rho(t, x+i) \prod_{j=1}^i (x+j) \right] = \frac{1}{\sqrt{i! (\lambda t)^i}} C_i^{(\lambda t)}(x),$$

in which $C_i^{(\mu)}(x)$ is the Charlier polynomial system orthogonal with Poisson density $\rho(\mu, x) = e^{-\mu} (\mu)^x / x!$ possessing expression

$$C_i^{(\mu)}(x) = \sum_{k=0}^i \binom{i}{k} \binom{x}{k} k! (-\mu)^{i-k}. \quad (\text{E.36})$$

In order to obtain the squared norm $d_i^2(t)$, we first establish the connection between $d_{ki}^2(t)$ and $d_{k+1,i}^2(t)$ where

$$d_{ki}^2(t) = \sum_{x_j=a}^{b-k-1} z_{ki}^2(t, x_j) \rho_k(t, x_j), \quad d_{0i}^2(t) = d_i^2(t), \quad z_{ki}(t, x) = \Delta^k y_i(t, x).$$

The self-adjoint equation for $z_{ki}(t, x)$ is

$$\Delta(s(t, x) \rho_k(t, x) \nabla z_{ki}(t, x)) + \eta_{ki}(t) \rho_k(t, x) z_{ki}(t, x) = 0,$$

where $\eta_{ki}(t) = \lambda_i(t) - \lambda_k(t)$.

Multiply by $z_{ki}(t, x)$, sum up over the values $x = x_j$ for which $a \leq x_j \leq b - k - 1$:

$$\sum_j z_{ki}(t, x_j) \Delta(s(t, x_j) \rho_k(t, x_j) \nabla z_{ki}(t, x_j)) + \eta_{ki}(t) d_{ki}^2(t) = 0.$$

Note that $\Delta z_{ki}(t, x) = z_{k+1,i}(t, x)$, $s(t, x+1)\rho_k(t, x+1) = \rho_{k+1}(t, x)$. Using difference identity for product in (E.19) gives

$$\begin{aligned}
& \sum_j z_{ki}(t, x_j) \Delta(s(t, x_j) \rho_k(t, x_j) \nabla z_{ki}(t, x_j)) \\
&= \sum_j [\Delta(s(t, x_j) \rho_k(t, x_j) z_{ki}(t, x_j) \nabla z_{ki}(t, x_j)) \\
&\quad - s(t, x_j+1) \rho_k(t, x_j+1) \nabla z_{ki}(t, x_j+1) \Delta z_{ki}(t, x_j)] \\
&= \sum_j [\Delta(s(t, x_j) \rho_k(t, x_j) z_{ki}(t, x_j) \nabla z_{ki}(t, x_j)) - \rho_{k+1}(t, x_j) z_{k+1,i}^2(t, x_j)] \\
&= s(t, x) \rho_k(t, x) z_{ki}(t, x) \nabla z_{ki}(t, x) \big|_a^{b-k} - d_{k+1,i}^2(t) = -d_{k+1,i}^2(t).
\end{aligned}$$

We thus have

$$d_{ki}^2(t) = \frac{1}{\eta_{ki}} d_{k+1,i}^2(t). \quad (\text{E.37})$$

And iterating the formula gives

$$\begin{aligned}
d_i^2(t) &= d_{0i}^2(t) = \frac{1}{\eta_{0i}} d_{1,i}^2(t) = \frac{1}{\eta_{0i}\eta_{1i}} d_{2,i}^2(t) = \dots \\
&= \frac{1}{\prod_{k=0}^{i-1} \eta_{ki}} d_{ii}^2(t) = \frac{1}{\prod_{k=0}^{i-1} \eta_{ki}} \sum_{x_j=a}^{b-i-1} \rho_i(t, x_j)
\end{aligned} \quad (\text{E.38})$$

provided that $z_{ii}^2(x, t) = \Delta^i y_i(t, x) = 1$.

Remark E.1: It follows from (E.18) and (E.38) that for any i , $\lambda_i(t) > 0$. That entails that $v'(t, x) < 0$ and $s''(t, x) = 0$ or $v'(t, x) < 0$ and $s''(t, x) < 0$. The former includes at least three processes: Brownian motion, Gamma process and Poisson process, while in the latter, after a transformation $s(t, x)$ can be written as $c^2 - x^2$ with fixed $c > 0$. However, this scenario is beyond the scope of this paper since we are interested in that $Z(t)$ assumes values on infinite interval or set, specifically, $\mathbb{R}$, $\mathbb{R}^+$ or $\mathbb{N}$. Therefore, our development will focus on the case where $v'(t, x) < 0$ and $s''(t, x) = 0$.

Although in our examples $v'(t, x) = -1$, in order to keep the framework as general as possible, we shall always treat $v'(t, x)$ as a negative function of t . Denote $\psi(t) := -[v'_x(t, x)]^{-1} > 0$.

Remark E.2: We may also need some asymptotic properties about the orthogonal polynomials. In the sequel, the following inequalities for Hermite polynomials and Laguerre polynomials are useful, which can be found in Nikiforov and Uvarov (1988, p. 54) for large i ,

$$\frac{1}{d_i} |H_i(x)| \leq C_1 i^{-\frac{1}{4}} \quad \text{and} \quad \frac{1}{d_i} |L_i^{(\alpha)}(x)| \leq C_2 i^{-\frac{1}{4}},$$

where d_i 's are the norm of Hermite and Laguerre polynomials in different inequalities respectively; C_1 and C_2 only depend on fixed x .

In addition, in view of the relation $c_i(\mu, x) = c_x(\mu, i) = x! L_x^{(i-x)}(\mu)$, the above inequality is true for Charlier polynomials as well. Thus, we may assert that within the ambit of our study, all classical orthonormal polynomials $Q_i(t, x)$ satisfy that $|Q_i(t, x)| \leq C i^{-\frac{1}{4}}$ for fixed t and x , where C is independent of i .

Appendix F: Finite horizons

We can do the same thing for the case that $t \in [0, T]$ with fixed T as in Section 4 of the main submission. We only state the assumptions and the results for this case.

Let $t \in [0, T]$ with T fixed. Suppose that we have n observations $(Y_{s,n}, Z_{s,n})$ at $t_{s,n} = T \frac{s}{n}$ for $s = 1, 2, \dots, n$. At the sampling points, we have the following model

$$Y_{s,n} = m(t_{s,n}, Z_{s,n}) + e_s, \quad s = 1, \dots, n, \quad (\text{F.1})$$

where $Y_{s,n} = Y(T \frac{s}{n})$, $Z_{s,n} = Z(T \frac{s}{n})$ and $e_s = \varepsilon(T \frac{s}{n})$. Let $x_{s,n} = \frac{1}{\sqrt{n}} \sum_{j=1}^s w_j$ where $w_j = \frac{\sqrt{n}}{\sqrt{T}\sigma_z}(Z_{j,n} - Z_{j-1,n} - \frac{\mu T}{n})$ with $\mu = E[Z(1)]$ and $\sigma_z^2 = \text{Var}(Z(1))$ form an i.i.d.(0,1) sequence, and then $Z_{s,n} = \frac{\mu T}{n}s + \sqrt{T}\sigma_z x_{s,n}$.

Assumption F.1

(a) Let $D^r m(t, x) \in L^2(I, \rho_r(t, x))$ for any $t \in [0, T]$ and $r = 0, 1, 2$. Moreover, the expansion of $D^2 m(t, Z(t))$ in terms of $Q_{2i}(t, Z(t))$ converges in the sense of mean square uniformly on $[0, T]$.

(b) For each i , $b_i(t, m) = E[m(t, Z_t)Q_i(t, Z_t)]$ and its derivatives of up to third order belong to $L^2[0, T]$.

REMARK F.1 Condition (a) is a routine requirement on the smoothness of the regression function which entails the application of the proposed orthogonal expansion and the negligibility of the residues after truncation. Condition (b) may be implied from Condition (a) if we impose some condition on the density of $Z(t)$. We simply put Condition (b) here to avoid any deviation from our main purpose.

Expanding m function at each sampling point gives n equations from the model which are written in matrix form $Y = X\beta + \delta + \gamma + \varepsilon$. By SLS, $\hat{\beta} = (X'X)^{-1}X'Y$. We then can estimate $m(\tau, x)$ by $\hat{m}(\tau, x) = A'(\tau, x)\hat{\beta}$ for $\tau \in [0, T]$ and x on the path of $Z(\tau)$. Let $\mathcal{S}$ and α define similarly as in Section 4 and after reshuffling $\mathcal{S}$ by α and $\frac{1}{\|A(\tau, x)\|}A(\tau, x)$, we obtain two functionals $F(t, Z(t))$ and $G(t, Z(t))$ for $t \in [0, T]$.

Assumption F.2

(a) Let $k = [n^{\kappa_1}]$ and $\frac{1}{2} < \kappa_1 < 1$.

(b) Let $p_{\min} = [n^{\kappa_2}]$, $p_{\max} = [n^{\bar{\kappa}_2}]$ with $0 < \kappa_2 \leq \bar{\kappa}_2 < 1$ and $0 \leq \bar{\kappa}_2 - \kappa_2 < 3\kappa_2 - \kappa_1 - 1$.

Assumption F.3 Both $F(t, x)$ and $G(t, x)$ are continuous in t and x .

REMARK F.2 Both Assumptions F.2 and F.3 are much simpler than their counterparts Assumptions A.6 and A.7 due to fixed T . This is because the estimator will converge given the continuity, while when $T = T_n$ we have to normalize this factor in the regressor and therefore the asymptotic theory is more complicated.

Theorem F.1: Suppose that $\{x_{s,n}\}_1^n$ and $\{e_s\}_1^n$ satisfy Assumption A.2. Under Assumptions F.1–F.3 we have as $n \rightarrow \infty$,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \frac{\alpha' X' X A(\tau, x)}{\sqrt{p_{\max}} \|A(\tau, x)\|^2} (\hat{m}(\tau, x) - m(\tau, x)) \\ & \rightarrow_D \int_0^1 F(Tr, T\mu r + \sqrt{T}\sigma_z W(r)) dU(r), \end{aligned} \quad (\text{F.2})$$

where $(U(r), W(r))$ is the vector of Brownian motions introduced in Assumption A.2.

REMARK F.3: The rate of convergence of $\widehat{m}(\tau, x) - m(\tau, x)$ is about $\frac{\sqrt{np_{\max}}}{\|A(\tau, x)\|}$. In view of the estimation of $\|A(\tau, x)\|$, the rate is between $n^{\frac{1}{2}(1-\kappa_1)}$ and $n^{\frac{1}{2}(1-\kappa_1)+\frac{1}{2}(\bar{\kappa}_2-\kappa_2)}$. The minimum order is smaller than $\frac{1}{4}$, while the maximum order is slightly bigger than the minimum.

References

- Doob, J. L. (1953). *Stochastic Processes*. Wiley Publications in Statistics. John Wiley & Sons, Inc., New York.
- Dudley, R. M. (2003). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics 74. Cambridge University Press, Cambridge, U.K.
- Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, 19(3):1035–1070.
- Nikiforov, A. F. and Uvarov, V. B. (1988). *Special Functions of Mathematical Physics*. Birkhauser, Boston.
- Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*. A Series of Comprehensive Studies in Mathematics 293. Springer-Verlag.
- Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegration regression. *Econometric Theory*, 30:509–535.
- Wang, Q. and Phillips, P. C. B. (2009). Asymptotic theory for local time density estimation and nonparametric cointegreting regression. *Econometric Theory*, 25:710–738.