TIME-VARYING VECTOR POISSON PROCESSES WITH COINCIDENCES

Victor Solo, Boris Godoy

School of Electrical Engineering and Telecommunications
University of New South Wales
Sydney, AUSTRALIA
email: v.solo@unsw.edu.au

ABSTRACT

Three emerging applications are driving a renewed interest in vector point processes: neural coding, high frequency finance and genomics. This pressure has revealed a gross lack of models and system identification methods. In particular in at least the first two applications coincidences can occur i.e. more than one event can occur at the same time. Yet the models in common use exclude this possibility. In this paper we develop a class of time-varying vector Poisson models that allow coincident events and develop for the first time an hypothesis test for no coincidences. We show simulation results and an application to high frequency finance data.

***Index Terms*—** point process, genomics, finance, neural coding.

1. INTRODUCTION

Point processes have a long history beginning at least with queueing theory a century ago [1]. More recently the advent of communication networks in the 1980s and the web in the 1990s provided some impetus. But there is now a renewed interest driven by three application areas.

In neural coding [2] spike trains are recorded from the micro-electrodes inserted in the brains of animals such as rats and monkeys. Such studies date from the 1930s, but a radical change occurred in the 1990s when new surgical techniques, smaller electrodes and computing advances made possible recording from scores and then hundreds of electrodes.

In high frequency finance large numbers of trades on the same stock can take place within a second (but recorded only on a 1s time-scale) [3]. For financial modeling one must then keep track of trade times and following early work of [4] point process methods have been increasingly used [5],[6].

Genomics has undergone explosive growth due partly to the rapid improvements in sequencing technologies over the last 3 decades. Recently the advent of high throughput data from gene regulation studies has made it possible to model the interaction between transcriptional regulatory elements (TRE)s. Thus e.g. [7] have used recently available public data sets to model this interaction using Hawkes processes.

All three areas then are generating vector point process data which need new system identification methods for analysis. And in at least the first two applications simultaneous events or coincidences [1] can occur. In neural coding simultaneous events would be associated with synchrony between different brain regions which is believed to be fundamental to brain function. In finance simultaneous events could be e.g. evidence of coordinated speculative activity or of the very rapid spread of local (insider) information. Also mismodeling simultaneous events could lead to mispricing of financial instruments such as derivatives.

However the standard point process models such as the time-varying Poisson [1] and history dependent models such as the Hawkes model [9] assume no-simultaneity [2] i.e. in any small time interval of extent δ only one event of any kind can occur. And currently the array of techniques available for modeling vector point processes are restricted to processes obeying no-simultaneity.

To be sure the theoretical literature has long since described properties of point processes with simultaneity. Thus the seminal book [10] discusses such processes and the theory of their associated product intensities i.e. higher moments. But the book does not discuss statistical modeling. In particular it does not provide realization-wise descriptions of processes needed for system identification. And so does not discuss likelihood functions. Similarly the classic volume [1] develops some theory under simultaneity but again the only modeling and likelihood discussion refers to processes obeying no-simultaneity.

The first author [11] has already shown how to extend the classic scalar no-simultaneity likelihood function of [12] and vector no-simultaneity likelihood of [13] to the case of simultaneity. Also in [11] the first author described a class of models that extend the Hawkes process to admit simultaneity. But the issue pursued here, namely hypothesis testing and fitting methods, was not discussed there.

This work was partly supported by the Australian Research Council

[1] [8] uses this word in a different non-standard way

[2] The technical terminology for no-simultaneity is 'orderly' or 'simple'

Here we step back to begin the complex system identification task by concentrating on a simpler class of problems, namely vector[3] Poisson processes with simultaneity. We develop a test statistic for testing the null hypothesis of no-simultaneity between the components of a time-varying bivariate Poisson process and illustrate the test with simulations and analysis of real financial data. Specifically we develop a Lagrange multiplier (LM) test [14],[15].

The LM test (aka the score test) has been very popular in econometrics over the last 3 decades and has the enormous advantage that while the model must be formulated under the alternative hypothesis it only requires model fitting under the null hypothesis. This translates into computational savings which with large data sets can be huge. The disadvantage is a potential loss of power compared to the likelihood -ratio test.

The remainder of the paper is organised as follows. In section 2 we review the classic vector time-varying Poisson (TVP) and then develop the extension that allows simultaneity. In section 3 we develop the LM test statistic. Section 4 contains simulations. Section 5 applies the method to real high frequency finance data. and conclusions are in section 6.

Notation & Elementary Properties.
N_t^r is the counting process of the rth point process and is the number of events up to time $t, r = 1, \cdots, d$.
We denote by λ_t^r the intensity of the rth point process. We introduce the counting increments. $N_{a,b}^r = N^r(a,b] = N_{(b)}^r - N_{(a)}^r$ number of events after time a up to and including time b. We use the big o, little o notation. So e.g. $f(\delta) = o(\delta)$ means $f(\delta)/\delta \to 0$ as $\delta \to 0$.

2. MODELS WITH COINCIDENCES

The issues can be best explained with the bivariate case and then can be easily generalised.

In 2.1 we review the classic vector TVP process. In 2.2 we construct a bivariate TVP process having simultaneity. In 2.3 we get a latent process representation useful for modelling.

2.1. No-simultaneity

So consider two point processes $N_t^{(1)}, N_t^{(2)}$ obeying the classic time-varying Poisson assumptions are [16].

NoSEA. No simultaneous events allowed.
$P(max_{r \in 1,2} dN_t^r > 1) = o(dt)$.

IndInc. Independent Increments.
For any finite number m of non-overlapping subintervals $(a_i, b_i]$ of $[0,T]$ the corresponding counting increments N_{a_i,b_i} are independent.

MI. Marginal Intensities.
$P(dN_t^r = 1) = \lambda_t^r dt + o(dt), r = 1, 2$.

Since each point process obeys the assumptions they are each scalar TVPs.

[3] aka multi-channel; aka multiple; aka multivariate

The following result is well known in point process folklore.

Result I. $N_t^{(1)}, N_t^{(2)}$ are independent.

Proof. Straightforward but omitted for lack of space.

2.2. Simultaneity

Since IndInc is fundamental to the time-varying Poisson process we would like to keep it. This means the marginal point processes will be scalar Poisson processes. So we need to drop the joint aspect of NoSEA. So we replace NoSEA with,

NoMSEA. No Marginal Simultaneous Events Allowed. $P(dN_t^r > 1) = o(dt), r = 1, 2$.

IndInc and NoMSEA will ensure the point processes are scalar TVPs. But we must respecify the intensities.

JI. Joint Intensities.

$$\begin{aligned} P(dN_t^{(1)} = 1, dN_t^{(2)} = 0) &= \lambda_t^a dt + o(dt) \\ P(dN_t^{(1)} = 0, dN_t^{(2)} = 1) &= \lambda_t^b dt + o(dt) \\ P(dN_t^{(1)} = 1, dN_t^{(2)} = 1) &= \lambda_t^c dt + o(dt). \end{aligned}$$

The third joint intensity shows the joint simultaneity.

An event time where: only one event occurs is called a singleton; two events occur is called a doublet.

Using NoMSEA and JI the joint mass function is now

$$\begin{aligned} P(dN_t^{(1)} = 1) &= P(dN_t^{(1)} = 1, dN_t^{(2)} = 0) \\ &+ P(dN_t^{(1)} = 1, dN_t^{(2)} = 1) + o(dt) \\ &= (\lambda_t^a + \lambda_t^c)dt + o(dt) \end{aligned}$$

Similarly $P(dN_t^{(2)} = 1) = (\lambda_t^b + \lambda_t^c)dt + o(dt)$.

2.3. Latent Process Representation

We can rewrite this in terms of three independent latent processes M_t^a, M_t^b, M_t.

Each of these processes obeys IndInc, NoSEA and their marginal intensities are as follows

$$\begin{aligned} P(dM_t^a = 1) &= P(dN_t^{(1)} = 1, dN_t^{(2)} = 0) \\ &= \lambda_t^a dt + o(dt) \\ \text{Similarly } P(dM_t^b = 1) &= \lambda_t^b dt + o(dt) \\ P(dM_t = 1) &= \lambda_t^c dt + o(dt). \end{aligned}$$

So these processes are TVPs. Finally by construction these processes are independent. Then we have

$$N_t^{(1)} = M_t^a + M_t \text{ and } N_t^{(2)} = M_t^b + M_t$$

After this paper was completed we became aware of the unpublished technical report [17] where a construction (different to ours) of a time-varying Poisson process is given which leads to the latent process representation. However [17] does not recognise the fundamental feature that the process violates NoSEA i.e. exhibits simultaneity.

Also they do not discuss model fitting.

3. LAGRANGE MULTIPLIER TEST

To keep the discussion manageable we consider only the bivariate case; but the approach is general.

Given observations of $N_t = (N_t^{(1)}, N_t^{(2)}), 0 \le t \le T$ consider testing H_0: $N_t^{(1)}, N_t^{(2)}$ are independent versus H_1 : $N_t^{(1)}, N_t^{(2)}$ are dependent. Equivalently we have $H_0 : \lambda_t^c = 0$, for all t against $H_1 : \lambda_t^c > 0$ for all t.

If we observe the event times the problem is trivial since a single doublet is enough to reject H_0. However in high frequency finance and neural coding it is often the case that the data comes binned; in that case the problem is non-trivial.

This is because, since we do not observe event times, we cannot construct the ground process and so cannot construct the latent processes which are now unobservable.

Suppose then that we have n equispaced observation bins of width Δ so that $n\Delta = T$ and $t = k\Delta$. For $r = 1, 2, a, b, c$ and $k = 1, \cdots, n$ denote

$$N_k^{r,\Delta} = \#\text{ events in } (k\Delta - \Delta, k\Delta] = N_{k\Delta}^r - N_{k\Delta-\Delta}^r$$

Also introduce the corresponding integrated intensities

$$\Lambda_t^r = \int_0^t \lambda_t^r dt, r = a, b, c, 1, 2$$

and the increments $\Lambda_k^{r,\Delta} = \Lambda_{k\Delta}^r - \Lambda_{k\Delta-\Delta}^r$.

To develop an LM test we must first parameterise the intensities and then restate the hypotheses in terms of the parameters. Also the likelihood under the alternative must be found as a function of the parameters. Thus suppose it becomes $L(\theta) = L(\theta^c, \theta^{ab})$ where the hypothesis test is now $H_0 : \theta^c = 0$ versus $H_1 : \theta^c \neq 0$. The likelihood under the null is $L(0, \theta^{ab})$. Let $\tilde{\theta}$ be the maximum likelihood estimator (MLE) of θ under the null hypothesis; i.e. $\tilde{\theta} = arg.max_{\theta:\theta^c=0} L(\theta)$. Then the LM test is given by [14],[15]

$$V = \frac{1}{2} L_c^T (\mathcal{I}_{cc} - \mathcal{I}_{cx}\mathcal{I}_{xx}^{-1}\mathcal{I}_{cx}^T)^{-1} L_c$$

where $x = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $L_c = \frac{\partial L(\theta^c, \theta^{ab})}{\partial \theta^c}|_{\tilde{\theta}}$ and $\mathcal{I}$s are information matrices e.g. $\mathcal{I}_{cc} = -E(\frac{\partial^2 L(\theta)}{\partial\theta^c \partial\theta^{c,T}})|_{\tilde{\theta}}$. Note that evaluation of the LM statistic requires expressions for first and second derivatives of the full likelihood $L(\theta)$. But these derivatives are evaluated only under the null. In practice we usually replace the information quantities with sample information e.g. $\mathcal{I}_{cc}$ is replace by $-\mathcal{L}_{cc}$.

We now sketch the assembly of the LM statistic. In subsection 3.1 we specify intensity models; in 3.2 we exhibit the likelihood under the null; in 3.3 we mention an EM algorithm. In 3.4 we exhibit the full likelihood and mention computation of derivatives needed for the LM test.

3.1. Intensity Models

We model the intensity functions non-parametrically using e.g. B-splines which allow positivity provided the coefficients are non-negative. So we put

$$\lambda_t^r = \mu^r + \Sigma_1^m \psi_{j,(t)} \beta_j^r = \mu^r + \psi_{(t)}^T \beta^r$$

where μ^r are positive background rates; $\psi_{j,(t)}$ are the B-spline basis elements; β_j^r are the coefficients. Thus

$$\lambda_k^{r,\Delta} = \frac{\Lambda_k^{r,\Delta}}{\Delta} = \mu^r + \int_{k\Delta-\Delta}^{k\Delta} \frac{\psi_{(t)}^T}{\Delta} dt \beta^r = \mu^r + \psi_k^T \beta^r$$

Provided $\beta_j^r \ge 0$ the itensities are guaranteed to be positive.

We abuse the notation by denoting $\boxed{\lambda_k^r = \lambda_k^{r,\Delta}}$. This should not cause confusion with λ_t^r since $t = k\Delta$ is continuous time while k is discrete time.

3.2. Likelihood Under the Null

Under H_0 we have $M_t = 0$ and so $N_t^{(1)} = M_t^a, N_t^{(2)} = M_t^b$. Since M_t^a, M_t^b are independent and by IndInc the likelihood is $P^a P^b$ where for $r = a, b$

$$P^r = \Pi_1^n P(M_k^{r,\Delta} = n_k^r) = \Pi_1^n e^{-\Lambda_k^{r,\Delta}} (\Lambda_k^{r,\Delta})^{n_k^r} / n_k^r!$$

Computationally and theoretically it is well known that we need to normalize the likelihood with a reference likelihood namely that of a unit rate Poisson. This yields

$$\frac{P^r}{P_{r,o}} = \Pi_1^n e^{-\Delta\lambda_k^r} (\lambda_k^r)^{n_k^r}$$

Putting $y_k^r = \frac{n_k^r}{\Delta}$ the relative log-likelihood is then

$$\mathcal{L}^{r,\Delta} = \frac{\mathcal{L}_r}{\Delta} = -\Sigma_1^n(\lambda_k^r - 1) + \Sigma_1^n y_k^r ln \lambda_k^r$$

3.3. EM Algorithm

Lack of space prevents inclusion of the derivation of the EM algorithm but it is straightforward.

3.4. Likelihood & Derivatives under the Alternative

In view of IndInc the likelihood under the alternative is $\Pi_1^n \frac{P_k}{P_{u,k}}$ where $P_k = P(N_k^{1,\Delta} = n_k^1, N_k^{2,\Delta} = n_k^2)$ and $P_{u,k}$ is the reference mass function of the same form but using a unit rate Poisson model.

Now set $n_k = min(n_k^1, n_k^2)$ and $M_k = N_k^{c,\Delta}$. Then using the latent process representation we get $P_k = \Sigma_{m=0}^{n_k} P_k^m$ and similarly for $P_{u,k}$ where

$$P_k^m = P(M_k^{a,\Delta} = n_k^1 - m) P(M_k^{b,\Delta} = n_k^2 - m) P(M_k = m)$$

We thus find $R_k^m = \frac{P_k^m}{P_{u,k}^m}$ is given by

$$\begin{aligned} R_k^m &= e^{-\Delta\lambda_k^a} (\lambda_k^a)^{n_k^1 - m_k} e^{-\Delta\lambda_k^b} (\lambda_k^b)^{n_k^2 - m_k} \\ &\times e^{-\Delta\lambda_k^c} (\lambda_k^c)^{m_k} \end{aligned}$$

Derivatives can now be computed in a straightforward way but details are long and tedious and omitted for lack of space.

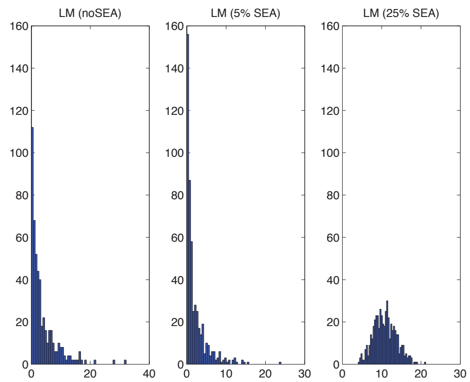


Fig. 1. LM Histograms: NoSEA, 5% SEA, 25% SEA

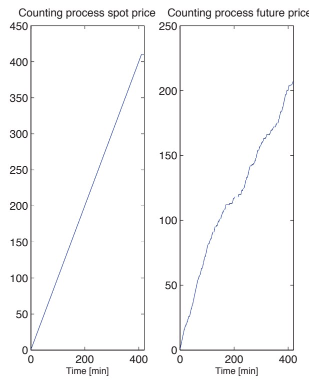


Fig. 2. Spot and Futures Counting Processes

4. SIMULATIONS

Using thinning [1] we simulated two independent point processes ($r = a, b$) with intensity functions of the form:

$$\lambda^r(t) = \mu^r + \beta_1^r \cos(\omega_1 t) + \beta_2^r \cos(\omega_2 t)$$

and with associated parameters $\mu^a = 2.0$, $\beta^a = [1\ 1]^T$, and $\mu^b = 0.5$, $\beta^b = [0.1\ 0.1]^T$. Also $\omega_1 = 0.1\omega_0$ and $\omega_2 = 100\omega_0$ with $\omega_0 = \pi/25$.

To generate simultaneity we simulate a third point process with same intensity function structure as above with parameters $\mu^c = 0.1$, and $\beta^c = [0.05\ 0.05]$. This will generate data with around 5% SEA. Choosing $\mu^c = 0.49$, and $\beta^c = [0.05\ 0.05]$, we can generate 25% SEA.

We now carry out model fitting as described above. We did 500 replicates of this to produce the LM histograms shown for the two cases: (i) 5 % SEA, (ii) 25% SEA.

Fig.1. compares the noSEA LM statistic with the 5% SEA LM statistic and with the 25% SEA LM statistic.

5. APPLICATION TO HIGH FREQUENCY FINANCE

One of the simplest but most pervasive examples of simultaneity is that between spot prices and futures prices. In particular we consider, the Standard & Poor 500 (S& P 500) spot price index, and the S& P 500 futures price index [18][19].

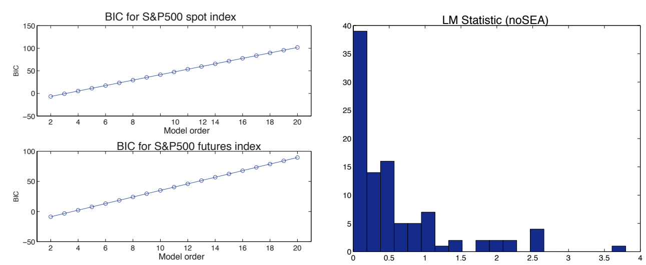
Using intraday data for September 8th, 2014 for these two indices, at a 1min sampling rate, gave 411 spot price samples and 207 futures prices samples.

In Fig.2. we show plots of the counting processes of the spot price and futures price data.

Thus, for the intensity functions $\lambda_a(t)$ and $\lambda_b(t)$ we propose the following models (with the help of BIC)

$$\lambda_a(t) = \mu^a + \beta^a \cos(\omega_0 t), \quad \lambda_b(t) = \mu^b + \beta^b \cos(\omega_0 t)$$

where $\omega_0 = \pi/T$, $T = 24481$. To generate the null distribution, we estimated (using the EM algorithm) the parameters for the models above assuming noSEA. Then we simulated 100 replicates using these parameter values followed by 100 model fits and LM computations. The resulting (bootstrapped) null distribution is shown in Fig.3. along with BIC plots for the initial model fit.



Spot & Futures BICs   LM Null Density

Fig. 3. BIC and LM Null Distribution

The LM statistic for the real data was 3.67. This is in the tail of the values seen in the histogram in Fig.3, so we can reject the hypothesis of no coincidences.

6. CONCLUSIONS

In this work we have developed, for the first time, a test of the hypothesis that a bivariate time-varying Poisson process has coincidences.

This required the construction, for the first time, of a bivariate time-varying Poisson process model that allows coincidences. This built on earlier work of the first author [11].

We developed a Lagrange multiplier test which has the advantage of only requiring model fitting under the null hypothesis and so can lead to huge computational savings. In the current case it just involves estimating two time varying rate functions and computing some model derivatives.

The method has been illustrated with a simulation and an application to real high frequency finance data.

7. REFERENCES

[1] D.J. Daley and D. Vere-Jones, *An introduction to the Theory of Point Processes, Volume I (2nd. ed.)*, Springer-Verlag, New York, 2003.

[2] F. Rieke, D. Warland, R. de Ruyter van Stevenink, and W. Bialek, *Spikes: Exploring the Neural Code*, MIT Press, Boston, 1997.

[3] L. Bauwens and N. Hautsch, "Modeling financial high frequency data using point processes," in *Handbook of Financial Time Series*. Springer, 2009, pp. 953–979.

[4] R.F. Engle and J.R. Russell, "Autoregressive conditional duration: A new model for irregularly spaced transaction data.," *Econometrica*, vol. 66, pp. 1127–1162, 1998.

[5] C.G. Bowsher, "Modelling security markets in continuous time: Intensity based, multivariate point process models," *Jl Econometrics*, vol. 141, pp. 876–912, 2007.

[6] E. Bacry, S. Delattre, M. Hoffman, and JF. Muz, "Modelling microstructure noise with mutually exciting point processes," *Quantitative Finance*, vol. 13, pp. 65–77, 2013.

[7] L Carstensen, A Sandelin, O Winther, and NR Hansen, "Multivariate hawkes process models of the occurrence of regulatory elements," *BMC Bioinformatics*, vol. 11, pp. 456–474, 2010.

[8] O. Macchi, "Stochastic processes and multicoincidences," *IEEE Trans. Inf. Thy.*, vol. 17, pp. 1–7, 1971.

[9] A.G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, pp. 83–90, 1971.

[10] S.K. Srinivasan, *Stochastic Point Processes and their Applications*, Griffin, London, UK, 1974.

[11] V Solo, "Likelihood functions for multivariate point processes with coincidences," in *IEEE Conference on Decision and Control, New Orleans, December 2007*. IEEE, 2007, pp. 4245–4250.

[12] I. Rubin, "Regular point processes and their detection," *IEEE Trans. Inf. Thy.*, vol. 18, pp. 547–557, 1972.

[13] J. Jacod, "Multivariate point processes: Predictiable projection, Radon Nikodym derivatives, representation of martingales," *Z. Wahrsch. verw. Geb.*, vol. 31, pp. 235–253, 1975.

[14] TS Breusch and AR Pagan, "The lagrange multiplier test and its applications to model specication in econometrics," *Review of Economic Studies*, vol. 47, pp. 239–254, 1980.

[15] R Engle, "A general formulation of wald, likelihood ratio, and lagrange multiplier tests in econometrics," in *Handbook of econometrics*. Elsevier Science Publishers, 1984, pp. 775–826.

[16] D.L. Snyder, *Random Point Processes*, J. Wiley, New York, 1975.

[17] DH. Johnson and IN. Goodman, "Jointly poisson processes," Tech. Rep. arXiv:0911.2524, Dept. ECE, Rice Univ.,, 2009.

[18] P.D. Koch, "Reexamining intrady simultaneity in stock index futures markets," *Journal of Banking and Finances*, vol. 17, pp. 1191–1205, 1993.

[19] B. Arshanapalli and J. Doukas, "The linkages of S&P 500 stock index and S&P 500 stock index futures prices during october 1987," *Journal of Economics and Business*, vol. 49, pp. 253–266, 1997.