

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/168475/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Kim, Namhyun, Wongsart, Pipat and Xia, Yingcun 2024. A dynamic count process. *Journal of Statistical Planning and Inference* 233 , 106187. 10.1016/j.jspi.2024.106187

Publishers page: <https://doi.org/10.1016/j.jspi.2024.106187>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Journal Pre-proof

A dynamic count process

Namhyun Kim, Pipat Wongsart, Yingcun Xia

PII: S0378-3758(24)00044-2
DOI: <https://doi.org/10.1016/j.jspi.2024.106187>
Reference: JSPI 106187

To appear in: *Journal of Statistical Planning and Inference*

Received date : 11 April 2022
Revised date : 15 May 2023
Accepted date : 18 April 2024



Please cite this article as: N. Kim, P. Wongsart and Y. Xia, A dynamic count process. *Journal of Statistical Planning and Inference* (2024), doi: <https://doi.org/10.1016/j.jspi.2024.106187>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Elsevier B.V. All rights reserved.

A Dynamic Count Process

Namhyun Kim¹

*University of Exeter Business School
University of Exeter
United Kingdom*

Pipat Wongsart

*Cardiff Business School
Cardiff University
United Kingdom*

Yingcun Xia²

*Department of Statistics & Applied Probability
National University of Singapore
Singapore
University of Electronic Science and Technology of China
China*

Abstract

The current paper aims to complement the recent development of the observation-driven models of dynamic counts with a parametric-driven one for a general case, particularly discrete two parameters exponential family distributions. The current paper proposes a finite semiparametric exponential mixture of SETAR processes of the conditional mean of counts to capture the non-linearity and complexity. Because of the intrinsic latency of the conditional mean, the general additive state-space representation of dynamic counts is firstly proposed then stationarity and geometric ergodicity are established under a mild set of conditions. We also propose to estimate the unknown parameters by using quasi maximum likelihood estimation and establishes the asymptotic properties of the quasi maximum likelihood estimators (QMLEs), particularly \sqrt{T} -consistency and normality under the relatively mild set of conditions. Furthermore, the finite sample properties of the QMLEs are investigated via simulation exercises and an illustration of the proposed process is presented by applying the proposed method to the intraday transaction counts per minute of AstraZeneca stock.

JEL classification: C19, C22, C24, C25

Keywords: Time series of counts, Parameter-driven model, Mixture of distributions, SETAR process, Quasi maximum likelihood estimation

¹Corresponding author: n.kim@exeter.ac.uk

²Yingcun Xia is partially supported by the National Natural Science Foundation of China (72033002 and 11931014) and the Academic Research Fund Tier 1 (FY2022) of the Ministry of Education of Singapore.

34 1. Introduction

35 The studies of univariate time series of counts have received extensive attention
36 because of its applicabilities in many different disciplines (see Davis et al. (2021) for
37 a comprehensive list of applications in different areas). In the preceding decades, one
38 of the most well-accepted approaches to modelling the dynamic counts applied the
39 generalised linear model framework of Nelder and Wedderburn (1972) because of its
40 convenient interpretation of covariates on the observed counts and an easy extension
41 of the Gaussian linear regression to an exponential family distribution. For instance,
42 the approach is taken by Zeger (1988), Davis et al. (2000), Davis and Wu (2009),
43 and Samia and Chan (2011), to just name a few. The first three studies are closely
44 related to a parameter-driven model of the broad classification of Cox (1981) within
45 the generalised state-space framework, that of Samia and Chan (2011) is related
46 to an observation-driven one. Subsequently, observation-driven models have been
47 actively developed in the past two decades, particularly by Fokianos et al. (2009),
48 Neumann (2011), Fokianos and Tjøstheim (2011), Fokianos and Tjøstheim (2012),
49 Wang et al. (2014), and Doukhan et al. (2021) for a Poisson process, and Davis and
50 Liu (2016) for the general one parameter exponential family case, because of their
51 convenient accessibility of estimating these models (see Davis et al. (2021) for an
52 excellent review on the topic including more comprehensive references and a review
53 of other methodological approaches). While the dynamic evolution of the stochastic
54 conditional mean of counts is driven by the past observed counts in the case of an
55 observation-driven model, for instance Poisson integer-valued ARCH (INARCH) or
56 GARCH (INGARCH) processes (see Fokianos et al. (2009) for details and references
57 therein), it is driven by its own dynamic evolution in the case of a parameter-driven
58 one. The computation of the likelihoods of those parameter-driven models is, there-
59 fore, not straightforward, even for the simple AR(1) specification of the conditional
60 mean (see Davis et al. (2021) for a more comprehensive discussion and references
61 therein) because of its intrinsic latency. For instance, Harvey and Fernandes (1989)
62 and Jørgensen et al. (1999) required the specific conjugate prior distributions to
63 perform the linear filtering. Therefore, this paper aims to complement the recent
64 development of the observation-driven models with a parameter-driven one for a
65 general case in that the nonlinearity and complexity (see Doukhan et al. (2021) for
66 details and references therein) of dynamic counts are described by modelling the
67 latent stochastic conditional mean with the finite semiparametric mixture of self
68 exciting threshold autoregressive (SETAR) processes.

69 The current paper firstly proposes to represent the discrete two parameters ex-
70 ponential family distributions within the structural state-space representation (see
71 Harvey and Fernandes (1989) for details and references therein) by introducing a neg-
72 ligibly marginal tuning parameter. Jørgensen (1987) provided the extensive study
73 on the exponential family distributions and named these processes as exponential
74 dispersion processes (EDPs). Hence, the current study adopts his abbreviation of

75 discrete EDPs for referring the discrete two parameters exponential family distri-
76 butions. As a result of introducing the tuning parameter, a legitimately simple
77 additive state-space representation of the proposed count process can be achieved
78 via log-transformation. Although it seems to be quite similar to the case of the
79 generalised linear regression model framework, the simple additive state-space rep-
80 resentation of dynamic counts is nontrivial. Because it allows us straightforward
81 implementation of the linear filtering, namely Kalman filter and, hence, accessible
82 establishment of stationarity and geometric ergodicity for a mixture of nonlinear
83 dynamic counts under a mild set of conditions. In that the explicit forms of up to
84 the second moments are also presented. The unknown parameters in the proposed
85 process are then estimated by using quasi maximum likelihood (QML) estimation
86 and the asymptotic properties of quasi maximum likelihood estimators (QMLEs),
87 particularly \sqrt{T} -consistency and normality, are also obtained under relatively prim-
88 itive regularity conditions. Although one may advocate to apply other nonlinear
89 filters such as the Extended Kalman, Unscented Kalman or Particle filters of the
90 nonadditive form, it is not easy to establish the geometric convergence of these fil-
91 ters for the state estimation. A set of strict and meticulous stability conditions,
92 particularly the number of inequalities and random tuning parameters, needs to
93 be imposed for the stability of the Lyapunov functions of those filters (see Särkkä
94 (2013) for a comprehensive treatment of the nonlinear filters) compared to their
95 linear counterpart of relatively mild observability and controllability conditions of a
96 system (see Chapter 3 of Caines (1987) for details).

97 The current paper is particularly related to those observation-driven ones, namely
98 the work of Samia and Chan (2011), Wang et al. (2014), and Doukhan et al. (2021).
99 The first two studies modelled the dynamic evolution of counts with SETAR of
100 Chan (1993) for the generalised linear model framework of a discrete exponential
101 family and Poisson INGARCH processes, respectively. The nonlinearity of the dy-
102 namic counts in their studies were driven by modelling the conditional mean of
103 counts with the observed past counts following the discontinuous SETAR process.
104 Unlike these two studies, our proposed process attempts to model the nonlinearity
105 of dynamic counts by modelling the nonlinear dynamic mechanism of the stochas-
106 tic latent conditional mean of counts with the continuous SETAR of Chan and
107 Tsay (1998) (see Chan and Tsay (1998), and Xia et al. (2007) for details). On the
108 other hands, Doukhan et al. (2021) studied a mixture of nonlinear INARCH and
109 INGARCH Poisson processes with a time-homogenous hidden Markov switching
110 model and also provided the criterion for selecting the correct number of regimes.
111 More specifically, they proposed the mixture of the Poisson processes themselves,
112 not the conditional means. For our proposed case, the exponential mixture of the
113 conditional means of the discrete EDPs is proposed. Because of the simple addi-
114 tive state-space representation of the dynamic counts via log-transformation, the
115 finite semiparametric exponential mixture of count processes through the condi-

116 tional mean is easily deduced and the linear filtering is also easily implemented.
 117 Lindsay (1983a), Lindsay (1983b), and Van der Vaart (1996) studied the semipara-
 118 metric mixture of distributions including exponential family distributions without
 119 performing Kalman filtering.

120 The rest of the paper is structured as follows. Section 2 proposes the discrete
 121 finite semiparametric exponential mixture of SETAR dynamic count processes and
 122 the QML estimation procedure, and establishes the asymptotic properties of the
 123 proposed QMLEs. The finite sample performances of the proposed QMLEs with
 124 simple but interesting Monte Carlo designs and the details of the proposed estima-
 125 tion procedure are presented in Section 3. In addition, Section 3 also illustrates our
 126 proposed process by applying to the intraday transaction counts per minute of As-
 127 traZeneca stock. The paper then concludes with the summary. The mathematical
 128 proofs of the main theoretical results of the paper are presented in the Appendix.

129 2. Exponential Mixture of SETAR Count Processes

130 2.1. Exponential Mixture of SETAR Count Processes

131 In this section, a discrete finite semiparametric exponential mixture of the dis-
 132 crete EDPs is introduced. In particular, the exponential mixture of the stochastic
 133 conditional means of the discrete EDPs is proposed as follows

$$\text{Prob}(I_t = i; \mu_t) = \prod_{k=1}^K ED^*(\mu_{k,t}, \beta_k, \pi_k), \quad i = 0, 1, 2, \dots, \quad (2.1.1)$$

134 where μ_t is a stochastic latent process specified in (2.1.2) below, $\pi_k \in (0, 1]$ is
 135 a mixing parameter such that $\sum_{k=1}^K \pi_k = 1$ with K being assumed to be finite
 136 and known, and $\beta_k \equiv \frac{1}{\lambda_k}$ with λ_k being a dispersion parameter that varies in a
 137 subset of positive real values. Furthermore, $ED^*(\cdot, \cdot, \cdot)$ denotes a discrete EDP with
 138 the mixture parameter, which is specified with the conditional mean and variance
 139 of counts such that $E(I_{k,t}; \mu_{k,t}, \pi_k) = \mu_{k,t}^{\pi_k} \equiv \tau(\theta_{k,t})$, where $I_{k,t}$ takes nonnegative
 140 integers and $\tau(\theta_{k,t}) = \frac{\partial \kappa_k(\theta_{k,t})}{\partial \theta_{k,t}}$ with $\kappa_k(\cdot)$ and $\theta_{k,t}$ being a cumulant function and a
 141 canonical parameter, respectively, and $\text{Var}(I_{k,t}; \mu_{k,t}, \pi_k) = \beta_k \text{V}(I_{k,t}; \mu_{k,t}, \pi_k)$ where
 142 $\text{V}(I_{k,t}; \mu_{k,t}, \pi_k) = \frac{\partial^2 \kappa_k(\theta_{k,t})}{\partial \theta_{k,t}^2} \Big|_{\theta_{k,t} = \tau^{-1}(\mu_{k,t}^{\pi_k})}$. Importantly, the data generating processes
 143 of each clusters are assumed to be independent. Additionally, the conditional mean
 144 of $I_{k,t}$ is specified by the continuous SETAR process for a flexible dynamic evolution
 145 of counts (see (2.2.4) below). Hence, it is transpired that the conditional mean and
 146 variance of I_t in (2.1.1) are as follows. Firstly, the conditional mean is

$$E(I_t; \mu_t) \equiv \mu_t \quad (2.1.2)$$

$$= \prod_{k=1}^K \mu_{k,t}^{\pi_k}, \quad (2.1.3)$$

147 where

$$\mu_{k,t} = \alpha_{k,0} \prod_{l=1, \neq d_k}^{p_k} \mu_{k,t-l}^{a_{k,l}} \left\{ \prod_{j_k=1}^{m_k} \left(\frac{\mu_{k,t-d_k}}{r_{k,j_k}} \right)^{a_{k,d_k,j_k}} \epsilon_{k,j_k,t} \right\}^{\mathbb{I}_k(r_{k,j_k-1} < \mu_{k,t-d_k} \leq r_{k,j_k})} \quad (2.1.4)$$

148 with $\alpha_{k,0} \geq 0$ ensuring the nonnegativeness of $\mu_{k,t}$, p_k being a nonnegative inte-
 149 ger, d_k being a positive integer such that $d_k \leq p_k$, r_{k,j_k} being a positive real value
 150 and $r_{k,0} = 0$, and $\epsilon_{k,j_k,t}$ being independently, identically, absolutely and continu-
 151 ously distributed (i.i.a.c.d.) over the positive real values with $E(\epsilon_{k,j_k,t}) = 1$ and
 152 $\text{Var}(\epsilon_{k,j_k,t}) = \sigma_{\epsilon,k,j_k}^2 < \infty$, and $\mathbb{I}_k(\cdot)$ denoting an indicator function. Note that the
 153 log-transformation of (2.1.4) is the standard continuous SETAR process (see (2.2.4)
 154 below). Then the conditional variance is

$$\text{Var}(I_t; \mu_t) = \prod_{k=1}^K \left\{ \sigma_{k,t}^2 + (\mu_{k,t}^{\pi_k})^2 \right\} - \prod_{k=1}^K (\mu_{k,t}^{\pi_k})^2, \quad (2.1.5)$$

155 where $\sigma_{k,t}^2$ denotes $\text{Var}(I_{k,t}; \mu_{k,t}, \pi_k)$ for the sake of notational simplicity, respectively.

156 The unconditional first two moments of counts are then obtained by applying
 157 the law of iterated expectations to (2.1.3) and (2.1.5), and they are as follows

$$E(I_t) = E[E(I_t; \mu_t)] \quad (2.1.6)$$

158 and

$$\text{Var}(I_t) = E[\text{Var}(I_t; \mu_t)] + \text{Var}[E(I_t; \mu_t)]. \quad (2.1.7)$$

159 The unconditional autocovariance is also obtained by applying the law of iterated
 160 expectation, similar to Davis and Wu (2009), as follows

$$\text{Cov}(I_t, I_{t+\tau}) = 0 + \text{Cov}[E(I_t; \mu_t), E(I_{t+\tau}; \mu_{t+\tau})]. \quad (2.1.8)$$

161 In addition, it is plausible to analyse the over or under-dispersion of the proposed
 162 process by applying Fisher's index to (2.1.6) and (2.1.7) as follows

$$\frac{E[\text{Var}(I_t; \mu_t)] + \text{Var}[E(I_t; \mu_t)]}{E[E(I_t; \mu_t)]} \leq 1. \quad (2.1.9)$$

163 According to (2.1.9), $E[\text{Var}(I_t; \mu_t)] + \text{Var}[E(I_t; \mu_t)] < E[E(I_t; \mu_t)]$ indicates the
 164 under-dispersed case and otherwise it is over-dispersed.

165 In order to obtain the unconditional first two moments in (2.1.6) to (2.1.8) and
 166 thus (2.1.9), $\mu_{k,t}$ in (2.1.4) is first rewritten in terms of $\epsilon'_{k,j_k,t}$'s under the following
 167 condition below

$$\max_{1 \leq k \leq K} \left(\max_{1 \leq j \leq m_k} \left(\sum_{l=1, \neq d_k}^{p_k} a_{k,l} + a_{k,d_k,j_k} \right) \right) < 1. \quad (2.1.10)$$

168 The condition in (2.1.10) is the necessary condition for the stationarity and geometric
 169 ergodicity of the conditional mean of the log-transformed μ_t (see Assumption 2.1 (i)
 170 below). Therefore, the condition in (2.1.10) leads to the stationarity and geometric
 171 ergodicity of μ_t . Now the law of iterated expectations is applied to obtain the first
 172 two unconditional moments of I_t . Therefore, they are given below

$$\begin{aligned} E(I_t) &= \prod_{k=1}^K E(\mu_{k,t}^{\pi_k}) \\ &= \prod_{k=1}^K \left[\left(\alpha_{k,0}^{\frac{\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \right) \left\{ \prod_{l=1}^{\infty} E \left(\epsilon_{k,j_k,t}^{\pi_k} \epsilon_{k,j_k,t-l}^{\pi_k b_{k,l}} \right) \right\} \right]^{\mathbb{I}_{k,j_k}}, \end{aligned} \quad (2.1.11)$$

173 where $\sum_{l=1}^{\infty} |b_{k,l}(L)| < \infty$ with L being a lag-operator, $\mathbb{I}_{k,j_k} = \mathbb{I}(r_{k,j_k-1} < \mu_{k,t-d_k} \leq$
 174 $r_{k,j_k})$ and $\sum_l^{p_k} a_{k,j,l}$ denotes $\sum_{l=1, \neq d_k}^{p_k} a_{k,l} + a_{k,d_k,j_k}$ for the sake of notational simplicity,
 175 and

$$\begin{aligned} \text{Var}(I_t) &= \prod_{k=1}^K E \{ \sigma_{k,t}^2 + \mu_{k,t}^{2\pi_k} \} - \prod_{k=1}^K \{ E(\mu_{k,t}^{\pi_k}) \}^2 \\ &= \prod_{k=1}^K \left[\left(E \{ \sigma_{k,t}^2 \} + \left\{ \alpha_{k,0}^{\frac{2\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{2\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \prod_{l=1}^{\infty} E \left(\epsilon_{k,j_k,t}^{2\pi_k} \epsilon_{k,j_k,t-l}^{2\pi_k b_{k,l}} \right) \right\} \right) \right. \\ &\quad \left. - \left(\prod_{k=1}^K \alpha_{k,0}^{\frac{\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \left\{ \prod_{l=1}^{\infty} E \left(\epsilon_{k,j_k,t}^{\pi_k} \epsilon_{k,j_k,t-l}^{\pi_k b_{k,l}} \right) \right\} \right)^2 \right]^{\mathbb{I}_{k,j_k}}. \end{aligned} \quad (2.1.12)$$

176 In addition, the unconditional autocovariance between I_t and $I_{t+\tau}$ is given below

$$\begin{aligned} \text{Cov}(I_t, I_{t+\tau}) &= \prod_{k=1}^K \left[\alpha_{k,0}^{\frac{2\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{2\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \left\{ \prod_{n=0}^{\tau-1} \prod_{l=0}^{\infty} E \left(\epsilon_{k,j_k,t+\tau-n}^{\pi_k b_{k,j,n}} \epsilon_{k,j_k,t-l}^{2\pi_k b_{k,j,t+l}} \right) \right. \right. \\ &\quad \left. \left. - \prod_{l=0}^{\infty} E \left(\epsilon_{k,j_k,t+l}^{\pi_k b_{k,j,l}} \right) E \left(\epsilon_{k,j_k,t+\tau-l}^{\pi_k b_{k,j,l}} \right) \right\} \right]^{\mathbb{I}_{k,j_k}}. \end{aligned} \quad (2.1.13)$$

177 The first two moments of (2.1.11) to (2.1.13) show that our proposed count process is
 178 weakly stationary under the condition of (2.1.10). Furthermore, the under and over-
 179 dispersion of the dynamic counts can be evaluated by applying the law of iterated
 180 expectations to (2.1.9) as follows

$$\begin{aligned}
 & \prod_{k=1}^K \left[\left(E \{ \sigma_{k,t}^2 \} + \left\{ \alpha_{k,0}^{\frac{2\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{2\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \prod_{l=1}^{\infty} E \left(\epsilon_{k,j_k,t}^{2\pi_k} \epsilon_{k,j_k,t-1}^{2\pi_k b_{k,l}} \right) \right\} \right) \right]^{\mathbb{I}_{k,j_k}} \\
 & \leq \left[\left(\prod_{k=1}^K \alpha_{k,0}^{\frac{\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \left\{ \prod_{l=1}^{\infty} E \left(\epsilon_{k,j_k,t}^{\pi_k} \epsilon_{k,j_k,t-l}^{\pi_k b_{k,l}} \right) \right\} \right) \right. \\
 & \quad \left. + \prod_{k=1}^K \alpha_{k,0}^{\frac{2\pi_k}{1-\sum_l^{p_k} a_{k,j,l}}} \left(\frac{1}{r_{k,j_k}} \right)^{\frac{2\pi_k a_{k,d_k,j_k}}{1-\sum_l^{p_k} a_{k,j,l}}} \left\{ \prod_{l=1}^{\infty} E \left(\epsilon_{k,j_k,t}^{\pi_k} \epsilon_{k,j_k,t-l}^{\pi_k b_{k,l}} \right) \right\}^2 \right]^{\mathbb{I}_{k,j_k}}.
 \end{aligned}$$

181 Moreover, the explicit forms of the above unconditional moments are obtained by
 182 considering the first few fractional moments of $\epsilon_{k,j_k,t}$. For example, consider a simple
 183 case where $K = 1$, $m = 1$, $p = 1$, $\alpha_0 = 1$, $a_1 = 0.5$, and $\epsilon_{1,t}$ is independently, identi-
 184 cally and exponentially distributed with $E(\epsilon_{1,t}) = 1$. The geometric approximation
 185 of $|a_1| < 1$ produces the unconditional mean of I_t such that $E(I_t) \approx E\left(\epsilon^{\frac{1}{1-a_1}}\right) = 2$;
 186 however, this is far from our expectations. In particular, the t-fold product of the
 187 fractional expectation of $\epsilon_{1,t}$ terms is expected to exponentially converge to 1 as
 188 the value of a_1 increases by the independence assumption. A fractional moment of
 189 $\epsilon_{1,t}$ can be obtained by applying the Riemann–Liouville fractional differ-integration
 190 technique to the moment generating function of $\epsilon_{1,t}$. Because the details of the
 191 Riemann–Liouville fractional differ-integration technique are easily found in Old-
 192 ham and Jerome (1974), only the brief application of the technique to the moment
 193 generating function is given below

$$\begin{aligned}
 \frac{d^q M_\epsilon(s=0)}{ds^q} &= \int_0^\infty (-\epsilon)^q \exp(-s\epsilon) f(\epsilon) d\epsilon \Big|_{s=0} \\
 &= (-)^q E(\epsilon^q),
 \end{aligned}$$

194 where q denotes a positive noninteger, $s \in \mathbb{R}$ is a neighbourhood value of 0, $M_\epsilon(\cdot)$ is
 195 a moment generating function of ϵ , and $f(\cdot)$ denotes a probability density function.
 196 The closed form of a fractional moment of ϵ can be obtained from the distributional
 197 assumption on ϵ .

198 2.2. Quasi Maximum Likelihood Estimation

199 The intrinsic difficulty of estimating our proposed discrete finite exponential
 200 mixture of the SETAR discrete EDPs is the presence of the latent stochastic con-
 201 ditional means and the unknown mixing parameters in the likelihood. Hence, a
 202 legitimate state-space representation of (2.1.1) is proposed. Before proceeding with
 203 the proposed state-space representation, it is imperative to discuss a few underlying
 204 remarks. Let us first introduce a stochastic process ζ_t , where ζ_t is defined below

205 (2.2.1), such that $E(I_t \zeta_t | \mu_t) = \mu_t$. Therefore, I_t can be rewritten by using ζ_t as
 206 follows

$$I_t \zeta_t = \mu_t, \quad (2.2.1)$$

207 where ζ_t is i.i.a.c.d. over the positive real values with $E(\zeta_t) = 1$ and $\text{Var}(\zeta_t) = \sigma_\zeta^2 <$
 208 ∞ .

209 It seems attractive to make an immediate logarithmic transformation of (2.2.1);
 210 however, this is not plausible because counts take nonnegative integers. This paper
 211 therefore proposes to introduce a tuning parameter, the so-called negligible marginal,
 212 such that $\Delta_T = O\left(c^{\frac{T}{2} + \delta}\right)$, where $c \in (0, 1)$ is unknown and $\delta > 0$ is arbitrarily
 213 small, in order to obtain the logarithmic transformation of (2.2.1). In fact, the
 214 tuning parameter in $\Delta_T = O\left(c^{\frac{T}{2} + \delta}\right)$ is $c \in (0, 1)$. The regularity condition on
 215 the rate of the tuning parameter, particularly $\frac{T}{2} + \delta$ where δ is fixed with a very
 216 small value, produces a fast enough convergence rate of Δ_T (faster than \sqrt{T}). This
 217 ensures the asymptotic uniform equivalence between $\ln \mu_t$ and $\ln(\mu_t + \Delta_T \zeta_t)$ (see
 218 (2.2.2) below) with a faster rate than \sqrt{T} , and ultimately generalises the existing
 219 state-space representation of dynamic count processes and consolidates the mixture
 220 of the discrete EDPs within a legitimately simple but general enough framework.

221 The state-space representation of (2.2.1) is thus as follows

$$y_t = \tilde{X}_t + \xi_t \text{ and } X_t = \ln \mu_t, \quad (2.2.2)$$

222 where $y_t = \ln(I_t + \Delta_T)$, $\tilde{X}_t = \ln(\mu_t + \Delta_T \zeta_t)$ and $\xi_t = -\ln \zeta_t$. At first glance, it seems
 223 implausible to apply the filtering to (2.2.2). However, \tilde{X}_t is shown to be asymptoti-
 224 cally equivalent to X_t uniformly over the parameter space of $(a's, r's, \beta's, c, \pi's, d's)^\top \in$
 225 D , where $a's$ includes $\ln \alpha'_{k,0}s$ hereafter, and D is a compact parameter space of
 226 $\mathbb{R}^{K + \sum_{k=1}^K m_k + p_k} \times \mathbb{R}_+^{\sum_{k=1}^K 2m_k - K} \times \mathbb{R}_+^K \times \mathbb{R}_{(0,1)} \times \mathbb{R}_{(0,1)}^K \times \mathbb{Z}_{0,1,\dots,p_K}^K$ with \mathbb{R}_+ , $\mathbb{R}_{(0,1)}$ and
 227 $\mathbb{Z}_{0,1,\dots,p_K}$ representing the positive real values, real values between 0 and 1 and
 228 integer values from 0 to p_K (p_K denotes $\max_{1 \leq k \leq K} (p_k)$), respectively, for the sake of
 229 notational simplicity. Additionally, the vectors of the parameters with a 0 subscript
 230 denote the vector of the true parameters hereafter. Before discussing the asymptotic
 231 equivalence between \tilde{X}_t and X_t uniformly over D , the geometric ergodicity of X_t
 232 is presented in Remark 2.1 under the appropriate regularity conditions as follows.

233 **Assumption 2.1.** (i) *The logarithmic transformed conditional mean of $I_{k,t}$, $X_{k,t}$,*
 234 *requires that $\max_{1 \leq j_k \leq m_k} \left| \sum_{l=1, \neq d_k}^{p_k} a_{k,l} + a_{k,d_k,j_k} \right| < 1$ for all $k = 1, \dots, K$.* (ii) *The*
 235 *stochastic process, $\epsilon_{k,j_k,t}$, is i.i.a.c.d. over positive real values with $E(\epsilon_{k,j_k,t}) = 1$,*
 236 *$\text{Var}(\epsilon_{k,j_k,t}) = \sigma_{\epsilon,k,j_k}^2$ and $E(\epsilon_{k,j_k,t}^{4+\delta}) < \infty$ for all $j_k = 1, \dots, m_k$ and $k = 1, \dots, K$. In*
 237 *addition, $\epsilon_{k,j_k,t}$ is independent to the initial state variable, $\mu_{k,1}$, for all $k = 1, \dots, K$*
 238 *and $t = 1, \dots, T$.*

239 **Remark 2.1.** Under Assumption 2.1, An and Huang (1996) showed that X_t is
 240 geometrically ergodic by using Tweedie's drift criterion (see Tweedie (1976)) and
 241 Tjøstheim's h -step criterion (see Tjøstheim (1990)).

242 The additional regularity conditions of the asymptotic equivalence between \tilde{X}_t
 243 and X_t are then as follows.

244 **Assumption 2.2.** (i) The negligible marginal, $\Delta_T = O\left(c^{\frac{T}{2}+\delta}\right)$, where $c \in (0, 1)$ is
 245 unknown and δ is arbitrarily small. (ii) The stochastic process, ζ_t , is i.i.a.c.d. over
 246 positive real values with $E(\zeta_t) = 1$, $\text{Var}(\zeta_t) = \sigma_\zeta^2$ and $E(\zeta_t^{4+\delta}) < \infty$. In addition,
 247 ζ_t is independent from the initial state variable, $\mu_{k,1}$, for all $k = 1, \dots, K$ and
 248 $t = 1, \dots, T$.

249 Firstly, the regularity condition on the rate of the tuning parameter is necessary
 250 to ensure the faster rate of convergence of \tilde{X}_t to X_t , particularly a faster rate than
 251 \sqrt{T} . Furthermore, the positive distributional assumption on ζ_t is necessary to ensure
 252 the positivity of μ_t and the finite moments of ζ_t up to fourth moments are necessary
 253 to apply the Cauchy–Schwartz inequality to establish the stochastic equi-continuity
 254 of the remainder term of $\tilde{X}_t - X_t$. The independence of ζ_t from the initial state
 255 variable $\mu_{1,k}$ for all $k = 1, \dots, K$ is also a necessary condition for the stability of
 256 the state-space representation in (2.2.2) (see Chapter 3 of Caines (1987) for details).
 257 The asymptotic equivalence between \tilde{X}_t and X_t uniformly over D is now ready to
 258 be presented.

259 **Lemma 2.1.** Under Assumptions 2.1 and 2.2, and where $\vartheta_0 = (a'_0 s, r'_0 s, \beta'_0 s, c_0, \pi'_0 s, d'_0 s)^\top \in$
 260 D , it can be shown that

$$\sup_{\vartheta \in D} \left| \tilde{X}_t - X_t \right| = o(T^{-1/2}) \text{ a.s.}, \text{ as } T \rightarrow \infty,$$

261 where a.s. denotes almost surely.

262 As a result of Lemma 2.1, the current paper finally proposes to represent the
 263 discrete finite semiparametric exponential mixture of the nonlinear dynamic count
 264 processes in (2.1.1) by the state-space representation below

$$y_t = \sum_{k=1}^K \pi_k X_{k,t} + \xi_t \quad (2.2.3)$$

265 and

$$X_{k,t} = a_{k,0} + \sum_{l=1, \neq d_k}^{p_k} a_{k,l} X_{k,t-l} + \left\{ \sum_{j_k}^{m_k} a_{k,d_k,j_k} (X_{k,t-d_k} - \mathbf{r}_{k,j_k}) + \eta_{k,j_k,t} \right\} \mathbb{I}_{k,j_k}, \quad (2.2.4)$$

266 where $X_{k,t} = \ln \mu_{k,t}$, $\mathbf{r}_{k,j_k} = \ln r_{k,j_k}$, and $\eta_{k,j_k,t} = \ln \epsilon_{k,j_k,t}$.

267 Let us now briefly discuss the estimation procedure of our proposed discrete finite
 268 exponential mixture of the discrete EDPs in (2.1.1). The first step is to apply the
 269 Expectation Maximisation (EM) algorithm of Shumway and Stoffer (1982) to (2.2.3)
 270 and (2.2.4), given the tuning and threshold parameters. The estimation procedure
 271 of the algorithm is similar to that of Chan and Tsay (1998). In particular, we
 272 firstly estimate $(a's, \beta's, d's)^\top$ given $\mathbf{r}'s$. $\mathbf{r}'s$ is then estimated by maximising the
 273 log-likelihood in (2.2.5) but substituting with $(\hat{a}'s, \hat{\beta}'s, \hat{d}'s)^\top$. The EM algorithm
 274 is the iterative estimation procedure of alternating between Kalman filtering and
 275 recursive smoothing, and QML estimation (see Shumway and Stoffer (1982) for
 276 details). Note that linear piecewise Kalman filtering and smoothing are required
 277 for the proposed procedure in our case. The tuning parameter is then estimated by
 278 maximising the Gaussian likelihood of $\xi_t s$ in (2.2.3). The second step is to perform
 279 the EM algorithm above with the estimated tuning parameter. The monotonicity of
 280 the sequence of the conditional log-likelihoods at each iterations of the EM algorithm
 281 ensures the convergence of the sequence of the conditional log-likelihoods to the one
 282 defined in (2.2.5) below (see Wu (1983) for details). Further details of the estimation
 283 procedures are presented in Section 3.1 below.

284 As a result of Lemma 2.1, $\hat{c} = c_0$ as $T \rightarrow \infty$, and the parameter space is modified
 285 accordingly such that $\psi_0 = (a'_0 s, \mathbf{r}'_0 s, \beta'_0 s, d'_0 s)^\top \in D_\psi$, where $D_\psi \subset D$ is a compact
 286 parameter space. The feasible conditional log-likelihood of I_t is then as follows

$$\mathcal{L}(\psi | \mathcal{F}_{t-1}) = \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \left(\sum_{k=1}^K \ln \widehat{ED}_{k,t}^* \right), \quad (2.2.5)$$

287 where $\widehat{ED}_{k,t}^*$ denotes $ED^*(\hat{\mu}_{k,t|t-1}, \beta_k, \hat{\pi}_k)$, and $\hat{\pi}_k$ and $\hat{\mu}_{k,t|t-1}$ are obtained by using
 288 the result of $\hat{X}_{k,t|t-1}$ which is the minimum conditional mean squared error estimate
 289 of $X_{k,t}$ given the sigma-field, \mathcal{F}_{t-1} , generated by $(I_1, I_2, \dots, I_{t-1})$ (see Chapters 3 and
 290 7 of Caines (1987) for details). The asymptotic properties of our proposed QMLEs
 291 are then established by showing the almost sure convergence of the feasible likeli-
 292 hood to the infeasible one uniformly over D_ψ with additional regularity conditions
 293 on EDPs. Hereafter, let us use $ED_{k,t}^*$ to denote $ED^*(\mu_{k,t}, \beta_k, \pi_k)$ for the sake of
 294 notational simplicity.

295 **Assumption 2.3.** (i) $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \ln ED_{k,t}^* \right)^2 < \infty$ and $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t}^*} \right)^4 <$
 296 ∞ for all $t = 1, \dots, T$, where $ED_{k,t,\pi}^{*(1)}$ denotes the first derivative of $ED_{k,t}^*$ with re-
 297 spect to π_k . (ii) $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \frac{ED_{k,t}^*}{ED_{k,t,\mu}^{*(1)}} \right)^4 < \infty$ and $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t,\mu}^{*(1)}} \right)^4 < \infty$
 298 for all $t = 1, \dots, T$, where $ED_{k,t,\mu}^{*(1)}$ denotes the first derivative of $ED_{k,t}^*$ with respect
 299 to $\mu_{k,t}$. (iii) $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \frac{ED_{k,t}^*}{ED_{k,t,\mu,\pi}^{*(2)}} \right)^4 < \infty$, $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t,\mu,\pi}^{*(2)}} \right)^4 < \infty$ and

300 $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k \in D_\psi} \frac{ED_{k,t,\mu}^{*(1)}}{ED_{k,t,\mu,\pi}^{*(2)}} \right)^4 < \infty$ for all $t = 1, \dots, T$, where $ED_{k,t,\mu,\pi}^{*(2)}$ denotes the
 301 second derivative of $ED_{k,t}^*$ with respect to $\mu_{k,t}$ and π_k .

302 The above conditions are required to establish the almost sure convergence of the
 303 feasible log-likelihood to the infeasible one uniformly over the compact parameter
 304 space. The main strategy of showing the almost sure convergence is to show the
 305 almost sure negligibility of the remainder term of the difference between the two
 306 log-likelihoods over the parameter space by using the Taylor expansion arguments
 307 of a logarithmic function. This produces a number of first and second derivatives of
 308 $ED_{k,t}^*$ with respect to π_k and $\mu_{k,t}$ for all $k = 1, \dots, K$. Therefore, the finite moments
 309 of the suprema of their derivatives up to fourth moment over the parameter space are
 310 required to the apply Cauchy–Schwartz inequality. With these regularity conditions
 311 on EDPs, the strong convergence of the feasible likelihood to the infeasible one
 312 uniformly over the parameter space is established as follows.

313 **Lemma 2.2.** Under Assumptions 2.1 to 2.3, and with $\psi_0 = (a_0's, \mathbf{r}'_0s, \beta'_0s, d'_0s)^\top \in$
 314 D_ψ , it is shown that

$$\sup_{\psi \in D_\psi} |\mathcal{L}(\psi | \mathcal{F}_{t-1}) - \mathcal{L}^*(\psi)| = O(T^{-1/2}) \text{ a.s., as } T \rightarrow \infty,$$

315 where

$$\mathcal{L}^*(\psi) = \frac{1}{T} \sum_{t=1}^T \left(\sum_{k=1}^K \ln ED_{k,t}^* \right).$$

316 Next, the almost sure convergence and asymptotic normality of the QMLEs are
 317 discussed below with additional regularity conditions as follows.

318 **Assumption 2.4.** (i) $\mu_{k,t}$ is an α -mixing process such that $\alpha(T) = O\left(T^{-\frac{(2+\delta)}{2}}\right)$
 319 for all $k = 1, \dots, K$. (ii) $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k^* \in D_{\psi^*}} \left| \frac{\partial \ln ED_{k,t}^*}{\partial \psi^*} \right|^{2+\delta} \right) < \infty$ for all $t = 1, \dots, T$,
 320 where $\psi^* = (a's, \mathbf{r}'s, \beta's)^\top \in D_{\psi^*}$ and $D_{\psi^*} \subset D_\psi$ is the compact parameter space.
 321 (iii) $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k^* \in D_{\psi^*}} \frac{\partial^2 \ln ED_{k,t}^*}{\partial \psi^* \partial \psi^{*\top}} \right)^2 < \infty$ and $\max_{1 \leq k \leq K} E \left(\sup_{\psi_k^* \in D_{\psi^*}} \frac{\partial^3 \ln ED_{k,t}^*}{\partial \psi^* \partial \psi^{*\top} \partial \psi^*} \right)^2 < \infty$ for
 322 all $t = 1, \dots, T$.

323 The first condition of Assumption 2.4, particularly the mixing condition, is the
 324 least restrictive serial dependence of the time series and the regularity condition of
 325 the rate on the mixing coefficient ensures the convergence rate \sqrt{T} (see Chapter 2
 326 of Fan and Yao (2008) for more details). The rest of the regularity conditions are
 327 necessary to establish the strong consistency uniformly over the parameter space and

328 the asymptotic normality of our proposed QMLEs (see Chapter 4 of Amemiya (1985)
 329 for an example). In particular, these conditions are used to apply the Chebyshev
 330 inequality and the Borel–Cantelli lemma.

331 **Theorem 2.1.** *Under Assumptions 2.1 to 2.4, and with $\limsup_{T \rightarrow \infty} \left(E \max_{\psi \in \bar{D}_\delta(\psi_0) \cap D_\psi} \mathcal{L}^*(\psi) \right) \neq$
 332 $\limsup_{T \rightarrow \infty} E \mathcal{L}^*(\psi_0)$ for any $\psi \in D_\psi$, where $\bar{D}_\delta(\psi_0)$ is the complement of an open δ -
 333 neighbourhood of ψ_0 , ψ_0 is uniquely identified and $\hat{\psi} = \psi_0 + O(T^{-1/2})$ a.s., as
 334 $T \rightarrow \infty$.*

335 As a result of Theorem 2.1 and the discreteness of $d's$, $\hat{d}'s = d'_0s$ as $T \rightarrow \infty$
 336 (see Chan and Tsay (1998) for details), the parameter space is modified accordingly.
 337 The asymptotic normality of our proposed QMLEs can then be as follows.

338 **Theorem 2.2.** *Under Assumptions 2.1 to 2.4, and when ψ_0^* is an interior of D_{ψ^*} ,*

$$\sqrt{T}(\hat{\psi}^* - \psi_0^*) \sim N(0, \Sigma),$$

339 where $\Sigma = B_0^{-1}(\psi_0^*)A_0(\psi_0^*)B_0^{-1}(\psi_0^*)$ with $B_0(\psi_0^*) = \lim_{T \rightarrow \infty} E \frac{\partial^2 \mathcal{L}^*(\psi^*)}{\partial \psi^* \partial \psi^{*\top}} \Big|_{\psi^* = \psi_0^*}$ and

$$340 A_0(\psi_0^*) = \lim_{T \rightarrow \infty} E \left(\sqrt{T} \frac{\partial \mathcal{L}^*(\psi^*)}{\partial \psi^*} \right) \left(\sqrt{T} \frac{\partial \mathcal{L}^*(\psi^*)}{\partial \psi^*} \right)^\top \Big|_{\psi^* = \psi_0^*}, \text{ as } T \rightarrow \infty.$$

341 3. Simulation and Illustration

342 In this section, the finite sample performance of our proposed QMLEs is in-
 343 vestigated with Monte Carlos simulation exercises. In addition, we illustrate the
 344 proposed process and estimation procedure by applying those to the intraday trans-
 345 action counts of AstraZeneca stock.

346 3.1. Simulation Study

347 The finite sample performances of the QMLEs are investigated with the most
 348 fundamental data generating process of counts, namely a Poisson process, and its
 349 extensions. This simulation exercise also focuses on how to implement the proposed
 350 estimation procedure presented in Section 2.2. The number of replications in all the
 351 simulation exercises is 10,000.

352 Let us firstly define the Poisson process with the conditional mean specified by
 353 the rudimentary SETAR process below

$$\text{Prob}(I_t = i; \mu_t) = \frac{\mu_t^{I_t} \exp(-\mu_t)}{I_t!}, \quad i = 0, 1, 2, \dots, \quad (3.1.1)$$

354 where $\mu_t = \begin{cases} \mu_{t-1}^{0.5} \epsilon_{1t} & \text{if } \mu_t \leq 1 \\ \mu_{t-1}^{-0.5} \epsilon_{2t} & \text{if } \mu_t > 1 \end{cases}$ with ϵ_{1t} and $\epsilon_{2t} \sim \text{Lognormal}(0, 1)$. The vector
 355 of parameters, namely $(\alpha_{0,1,1} = 0.5, \alpha_{0,1,2} = -0.5, r_0 = 1, c_0)'$, is estimated by the

356 proposed estimation procedure as follows. First, the EM algorithm with linear
 357 piecewise Kalman filtering and smoothing is applied, given the initial tuning and
 358 threshold parameters below

$$y_t = X_t + \xi_t \quad \text{and} \quad X_t = \begin{cases} 0.5(X_{t-1} - 0)_- + \eta_{1t} \\ -0.5(X_{t-1} - 0)_+ + \eta_{2t}, \end{cases}$$

359 where $y_t = \ln\left(I_t + c^{\frac{T}{2} + \delta}\right)$ with $0 < c < 1$ and $0 < \delta < 1$ being arbitrary, and
 360 $(X_{t-1} - 0)_-$ and $(X_{t-1} - 0)_+$ denote $X_{t-1} \leq 0$ and $X_{t-1} > 0$, respectively, until the
 361 sequence of the conditional likelihoods converges to the below

$$\mathcal{L}(\alpha_{1,1}, \alpha_{1,2} | \mathcal{F}_{t-1}, r) = \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \frac{\hat{\mu}_{t|t-1}^{I_t} \exp(-\hat{\mu}_{t|t-1})}{I_t!},$$

362 where $\hat{\mu}_{t|t-1} = \exp(\hat{X}_{t|t-1})$ with $\hat{X}_{t|t-1}$ being obtained by implementing Kalman
 363 filtering with $(\hat{\alpha}_{1,1}, \hat{\alpha}_{1,2})'$ from the last iteration of the EM algorithm. We then
 364 estimate r by maximising the following

$$\mathcal{L}(r | \mathcal{F}_{t-1}, \hat{\alpha}_{1,1}, \hat{\alpha}_{1,2}) = \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \frac{\hat{\mu}_{t|t-1}^{I_t} \exp(-\hat{\mu}_{t|t-1})}{I_t!},$$

365 where $\hat{\mu}_{t|t-1} = \exp(\hat{X}_{t|t-1})$, and $\hat{X}_{t|t-1} = \begin{cases} \hat{\alpha}_{1,1} \hat{X}_{t-1|t-1} & \text{if } \hat{X}_{t-1|t-1} \leq 0 \\ \hat{\alpha}_{2,1} \hat{X}_{t-1|t-1} & \text{otherwise} \end{cases}$. The tuning

366 parameter is then estimated by maximising the Gaussian likelihood of $\hat{\xi}_t$ s, where
 367 $\hat{\xi}_t = \log\left(I_t + c^{\frac{T}{2} + \delta}\right) - \hat{X}_{t|t-1}$. The next step is to apply the EM algorithm with the
 368 estimated tuning parameter to the below

$$\hat{y}_t = X_t + \xi_t \quad \text{and} \quad X_t = \begin{cases} 0.5(X_{t-1} - 0)_- + \eta_{1t} \\ -0.5(X_{t-1} - 0)_+ + \eta_{2t}, \end{cases}$$

369 where $\hat{y}_t = \ln\left(I_t + \hat{c}^{\frac{T}{2} + \delta}\right)$. The estimation procedure is similar to that of the two-
 370 step estimation procedure. First, estimate $(\alpha_{1,1}, \alpha_{1,2})'$ then r . The results for (3.1.1)
 371 are presented in Table 1.

372 Next, the Type II Negative Binomial (NB) process is considered. Although
 373 it is a simple extension of a Poisson process, it is one of the most popular count
 374 processes in practice because of its over-dispersion property. The Type II NB process
 375 is commonly obtained by mixing a Poisson process with Gamma distribution as
 376 follows

$$\text{Prob}(I_t = i; \mu_t) = \frac{\gamma_t^{I_t} \exp(-\gamma_t)}{I_t!}, \quad i = 0, 1, 2, \dots, \quad (3.1.2)$$

377 where $\gamma_t = \mu_t \nu_t$ with $\mu_t = \begin{cases} \mu_{t-1}^{0.2} \epsilon_{1t} & \text{if } \mu_t \leq 1 \\ \mu_{t-1}^{-0.2} \epsilon_{2t} & \text{if } \mu_t > 1 \end{cases}$ with ϵ_{1t} and $\epsilon_{2t} \sim \text{Exp}(1)$, and
 378 $\nu_t \sim \text{Gamma}(\beta, 1/\beta)$ with $\beta = 1$. The count process in (3.1.2) can be rewritten as

Table 1: *Simulation results for (3.1.1)*

T	$\hat{\alpha}_{1,1}$	$s.e.(\hat{\beta}_{1,1})$	$mse(\hat{\beta}_{1,1})$	$\hat{\alpha}_{1,2}$	$s.e.(\hat{\beta}_{1,2})$	$mse(\hat{\beta}_{1,2})$
100	0.4925	0.0665	0.0046	-0.4870	0.1007	0.0111
250	0.4976	0.0378	0.0015	-0.4951	0.0570	0.0033
500	0.4984	0.0258	0.0006	-0.4977	0.0383	0.0015
700	0.4989	0.0214	0.0005	-0.4986	0.0317	0.0010
1000	0.4996	0.0177	0.0003	-0.4990	0.0263	0.0007
T	\hat{r}	$s.e.(\hat{r})$	$mse(\hat{r})$	\hat{c}	$s.e.(\hat{c})$	
100	1.0096	0.1155	0.0125	0.9963	0.0027	.
250	1.0051	0.0670	0.0044	0.9985	0.0006	.
500	1.0012	0.0464	0.0021	0.9992	0.0002	.
700	1.0013	0.0389	0.0014	0.9994	0.0001	.
1000	1.0012	0.0323	0.0010	0.9996	0.0001	.

Table 2: *Simulation results for (3.1.3)*

T	$\hat{\alpha}_{1,1}$	$s.e.(\hat{\alpha}_{1,1})$	$mse(\hat{\alpha}_{1,1})$	$\hat{\alpha}_{1,2}$	$s.e.(\hat{\alpha}_{1,2})$	$mse(\hat{\alpha}_{1,2})$	\hat{c}	$s.e.(\hat{c})$
100	0.2208	0.1200	0.0105	-0.2276	0.2510	0.0196	0.9923	0.0024
250	0.2070	0.0754	0.0049	-0.2120	0.1579	0.0114	0.9969	0.0006
500	0.2024	0.0529	0.0026	-0.2014	0.1030	0.0067	0.9984	0.0002
700	0.2010	0.0447	0.0020	-0.1989	0.0849	0.0052	0.9989	0.0001
1000	0.2011	0.0376	0.0014	-0.1976	0.0698	0.0039	0.9992	0.0001
T	\hat{r}	$s.e.(\hat{r})$	$mse(\hat{r})$	$\hat{\beta}$	$s.e.(\hat{\beta})$	$mse(\hat{\beta})$		
100	0.9360	0.8529	0.4956	0.9467	0.3660	0.0378	.	.
250	0.9774	0.5135	0.1751	0.9723	0.2355	0.0209	.	.
500	0.9966	0.3500	0.0770	0.9869	0.1704	0.0134	.	.
700	0.9963	0.2917	0.0539	0.9900	0.1446	0.0110	.	.
1000	1.0008	0.2410	0.0365	0.9579	0.1136	0.0054	.	.

379 follows

$$\text{Prob}(I_t = i; \mu_t) = \frac{\Gamma(\beta + I_t)}{\Gamma(\beta)\Gamma(I_t + 1)} \left(\frac{\beta}{\beta + \mu_t}\right)^\beta \left(\frac{\mu_t}{\beta + \mu_t}\right)^{I_t}, \quad i = 0, 1, 2, \dots, \quad (3.1.3)$$

380 where $\Gamma(\cdot)$ denotes a gamma function. The vector of the parameters, $(\alpha_{0,1,1} =$
 381 $0.2, \alpha_{0,1,2} = -0.2, r_0 = 1, \beta_0 = 1, c_0)'$, are estimated by the similar procedure as
 382 above, and the results are presented in Table 2.

383 The last exercise involves a mixture of a Poisson process with the conditional
 384 means specified by a SETAR and a simple autoregressive (AR(1)) processes as fol-
 385 lows

$$\text{Prob}(I_t = i; \mu_t) = \frac{\mu_t^{I_t} \exp(-\mu_t)}{I_t!}, \quad i = 0, 1, 2, \dots, \quad (3.1.4)$$

386 where $\mu_t = \prod_{k=1}^2 \mu_{k,t}^{\pi_k}$ with $\pi_1 = \pi_2 = 0.5$,

$$\mu_{1,t} \begin{cases} \mu_{1,t-1}^{0.5} \epsilon_{11,t} & \text{if } \mu_{1,t} \leq 1 \\ \mu_{1,t-1}^{-0.5} \epsilon_{12,t} & \text{if } \mu_{1,t} > 1 \end{cases} \quad \text{with } \epsilon_{11,t} \text{ and } \epsilon_{12,t} \sim \text{Lognormal}(0, 1)$$

Table 3: *Simulation results for (3.1.4)*

T	$\hat{\alpha}_{1,1,1}$	$s.e.(\hat{\alpha}_{1,1,1})$	$mse(\hat{\alpha}_{1,1,1})$	$\hat{\alpha}_{1,1,2}$	$s.e.(\hat{\alpha}_{1,1,2})$	$mse(\hat{\alpha}_{1,1,2})$	\hat{c}	$s.e.(\hat{c})$
100	0.4689	0.1672	0.0333	-0.4511	0.2143	0.0643	0.9977	0.0028
250	0.4864	0.0931	0.0096	-0.4826	0.1319	0.0187	0.9992	0.0007
500	0.4946	0.0621	0.0042	-0.4934	0.0878	0.0080	0.9996	0.0002
700	0.4969	0.0516	0.0029	-0.4956	0.0727	0.0056	0.9997	0.0001
1000	0.4979	0.0428	0.0020	-0.4967	0.0602	0.0038	0.9998	0.0001
T	\hat{r}_1	$s.e.(\hat{r}_1)$	$mse(\hat{r}_1)$	$\hat{\alpha}_{2,1}$	$s.e.(\hat{\alpha}_{2,1})$	$mse(\hat{\alpha}_{2,1})$		
100	0.9270	0.3858	0.0148	0.1847	0.1870	0.0375	.	.
250	0.9073	0.1749	0.0106	0.1950	0.1123	0.0130	.	.
500	0.9050	0.1098	0.0100	0.1984	0.0774	0.0064	.	.
700	0.9048	0.0900	0.0100	0.1989	0.0648	0.0044	.	.
1000	0.9048	0.0741	0.0100	0.1997	0.0540	0.0030	.	.
T	$\hat{\pi}_1$	$s.e.(\hat{\pi}_1)$	$mse(\hat{\pi}_1)$	$\hat{\pi}_2$	$s.e.(\hat{\pi}_2)$	$mse(\hat{\pi}_2)$		
100	0.490	0.080	0.005	0.481	0.085	0.007	.	.
250	0.496	0.049	0.002	0.489	0.052	0.003	.	.
500	0.498	0.034	0.001	0.493	0.037	0.001	.	.
700	0.498	0.029	0.001	0.495	0.031	0.001	.	.
1000	0.498	0.024	0.001	0.496	0.026	0.001	.	.

387 and

$$\mu_{2,t} = \mu_{2,t-1}^{0.2} \epsilon_{2,t} \text{ with } \epsilon_{2,t} \sim \text{Lognormal}(0, 1).$$

388 The vector of the parameters, $(\alpha_{0,1,1,1} = 0.5, \alpha_{0,1,1,2} = -0.5, r_{0,1} = 1, \alpha_{0,2,1} =$
389 $0.2, \pi_{0,1} = 0.5, \pi_{0,2} = 0.5, c_0)'$, are estimated by a similar procedure to (3.1.1) and
390 (3.1.3). The results for (3.1.4) are presented in Table 3.

391 For all these cases, the simulation exercise shows the satisfactory finite sam-
392 ple performance of our proposed QML estimation procedure. The estimates of the
393 tuning parameters are close to the value of 1, as we expected. Additionally, notice
394 that our proposed process is the special cases of those parametric-driven specifica-
395 tions within the generalised linear regression framework of Nelder and Wedderburn
396 (1972), particularly Zeger (1988) for a Poisson process and Davis and Wu (2009) for
397 a negative binomial process, where there is no covariate. The simulation results for
398 the simple Poisson and negative binomial processes of Zeger (1988) and Davis and
399 Wu (2009) without a covariate can be obtained by requesting the authors. In the
400 following, we apply the proposed process and estimation procedure to the intraday
401 transaction counts of AstraZeneca stock.

402 3.2. Illustration of Real Data Analysis

403 We now illustrate our proposed count process by applying it to analyse the
404 number of transactions per minute for AstraZeneca stock, closely following Fokianos
405 et al. (2009). The randomly selected trading day is 30 July 2019. There are about
406 500 observations after eliminating the first and last minutes of transactions for about

Table 4: Estimation results of (3.2.1)

	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}$	\hat{c}
estimates	3.2034	-0.1511	0.7650	0.9503
s.e.	(0.1664)	(0.0559)	(0.0459)	(0.0028)

407 8 trading hours. The autocorrelation function of this data (see Figure 1 (b)) shows
 408 the moderate dependence between transactions not as strong as the case of Ericsson
 409 B stock in Fokianos et al. (2009). Furthermore, it is an over-dispersed case. The
 410 value of the sample mean is 10.0266 with the variance of 65.3661. The over-dispersion
 411 in this data might be caused by the frequent zero transactions and a few large number
 412 of transactions (see Figure 1(a)). Therefore, the Type II NB process discussed in
 413 Section 3.1 is considered in this analysis.

414 Applying first the simple AR(1) process to the conditional mean of the transac-
 415 tion counts, the Type II NB process is

$$\text{Prob}(I_t = i; \mu_t) = \frac{\Gamma(\beta + I_t)}{\Gamma(\beta)\Gamma(I_t + 1)} \left(\frac{\beta}{\beta + \mu_t} \right)^\beta \left(\frac{\mu_t}{\beta + \mu_t} \right)^{I_t}, \quad i = 0, 1, 2, \dots, \quad (3.2.1)$$

416 where $\mu_t = \alpha_0 \mu_{t-1}^{\alpha_1} \epsilon_t$. The results of this estimation are reported in Table 4. For
 417 examining the adequacy of the fit, an analysis of the Pearson residuals is performed.
 418 The Pearson residuals are defined as $e_t = \frac{I_t - \hat{\mu}_t}{\sqrt{\hat{\mu}_t \left(1 + \frac{\hat{\mu}_t}{\beta}\right)}}$ in the case of Type II NB
 419 process, and e_t is an white noise process under a correct specification of I_t . The
 420 cumulative periodogram plot of e_t s (see Figure 1(d)) indicates that (3.2.1) is not
 421 adequate enough to model the intraday transactions of this data. The prediction of
 422 I_t of (3.2.1) is also shown in Figure 1(c). Furthermore, the mean squared error of
 423 (2.2.3) for the case of (3.2.1) is 6.6976.

424

425 Now let us apply the continuous two-regime SETAR with $p = d = 1$ to the
 426 conditional mean of the transaction counts, the Type II NB process is

$$\text{Prob}(I_t = i; \mu_t) = \frac{\Gamma(\beta + I_t)}{\Gamma(\beta)\Gamma(I_t + 1)} \left(\frac{\beta}{\beta + \mu_t} \right)^\beta \left(\frac{\mu_t}{\beta + \mu_t} \right)^{I_t}, \quad i = 0, 1, 2, \dots, \quad (3.2.2)$$

427 where $\mu_t = \alpha_0 \prod_{k=1}^2 \left\{ \left(\frac{\mu_{t-k}}{r} \right)^{\alpha_{1,k}} \epsilon_{k,t} \right\}^{\mathbb{I}_k}$. The results of this estimation are reported in
 428 Table 5. The cumulative periodogram plot of e_t s and the prediction of I_t of (3.2.2)
 429 are shown in Figure 1(e) and (f), respectively. The improvement on the Pearson's
 430 residuals (see Figure 1(f)) supports the use of (3.2.2), namely the nonlinear Type
 431 II NB process, instead of the linear one. There is also improvement on the mean
 432 squared error for the case of (3.2.2), it is 6.3317.

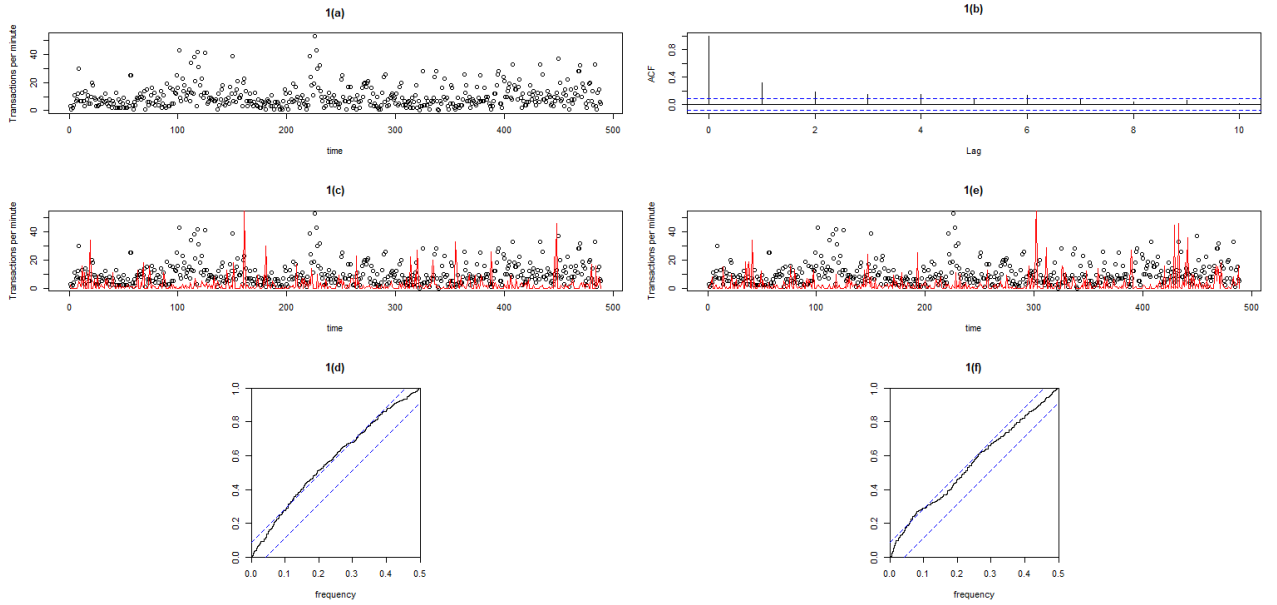
433

Table 5: Estimation results of (3.2.2)

	$\hat{\alpha}_0$	$\hat{\alpha}_{1,1}$	$\hat{\alpha}_{1,2}$	\hat{r}	$\hat{\beta}$	\hat{c}
estimates	3.3202	0.2700	-0.2023	1.2825	0.7493	0.9486
s.e.	(0.1498)	(0.0977)	(0.0490)	(0.3375)	(0.0455)	(0.0027)

434 Figure 1: Intraday transaction counts of AstraZeneca stock on 30 July, 2019

435



436

437 (a) Number of transactions per minute for AstraZeneca stock on 30 July, 2019. (b) Autocorrelation
438 function of the transaction data. (c) Observed and predicted (red) number of transactions per
439 minute calculated by using (3.2.1). (d) Cumulative periodogram plot of the Pearson residuals
440 calculated by using (3.2.1). (e) Observed and predicted (red) number of transactions per minute
441 calculated by using (3.2.2). (f) Cumulative periodogram plot of the Pearson residuals calculated by
442 using (3.2.2)

443 **4. Summary**

444 This paper aims to complement the recent development of the observation-driven
445 models of dynamic counts with a parameter-driven one for the general case, specif-
446 ically the discrete two parameters exponential family distributions. In particular,
447 we propose to model the mixture of nonlinear dynamic counts by representing a dy-
448 namic count process with a simple additive state-space representation. As a result
449 of this, a more flexible dynamic evolution than a stationary AR(p) process of the
450 conditional mean, particularly continuous SETAR process, and the discrete finite
451 semiparametric exponential mixture of dynamic count processes are analysed with
452 the well established linear filtering in that stationarity and geometric ergodicity of

453 the process are obtained under a mild set of conditions. Furthermore, the unknown
454 parameters are proposed to be estimated with quasi maximum likelihood estima-
455 tion and the asymptotic properties of the QMLEs, particularly \sqrt{T} -consistency and
456 normality, are established under a relatively primitive set of conditions.

Journal Pre-proof

457 **References**

- 458 Amemiya, T., 1985. *Advanced Econometrics*. Harvard University Press.
- 459 An, H., Huang, F., 1996. The geometrical ergodicity of nonlinear autoregressive
460 models. *Statistica Sinica* , 943–956.
- 461 Caines, P.E., 1987. *Linear Stochastic Systems*. John Wiley & Sons, Inc.
- 462 Chan, K.S., 1993. Consistency and limiting distribution of the least squares estima-
463 tor of a threshold autoregressive model. *The Annals of Statistics* 21, 520–533.
- 464 Chan, K.S., Tsay, R.S., 1998. Limiting properties of the least squares estimator of
465 a continuous threshold autoregressive model. *Biometrika* 85, 413–426.
- 466 Cox, D.R., 1981. Statistical analysis of time series: Some recent developments.
467 *Scandinavian Journal of Statistics* , 93–115.
- 468 Davis, R.A., Dunsmuir, W.T., Wang, Y., 2000. On autocorrelation in a poisson
469 regression model. *Biometrika* 87, 491–505.
- 470 Davis, R.A., Fokianos, K., Holan, S.H., Joe, H., Livsey, J., Lund, R., Pipiras, V.,
471 Ravishanker, N., 2021. Count time series: A methodological review. *Journal of*
472 *the American Statistical Association* 116, 1533–1547.
- 473 Davis, R.A., Liu, H., 2016. Theory and inference for a class of nonlinear models
474 with application to time series of counts. *Statistica Sinica* , 1673–1707.
- 475 Davis, R.A., Wu, R., 2009. A negative binomial model for time series of counts.
476 *Biometrika* 96, 735–749.
- 477 Doukhan, P., Fokianos, K., Rynkiewicz, J., 2021. Mixtures of nonlinear poisson
478 autoregressions. *Journal of Time Series Analysis* 42, 107–135.
- 479 Fan, J., Yao, Q., 2008. *Nonlinear Time Series: Nonparametric and Parametric*
480 *Methods*. Springer Science & Business Media.
- 481 Fokianos, K., Rahbek, A., Tjøstheim, D., 2009. Poisson autoregression. *Journal of*
482 *the American Statistical Association* 104, 1430–1439.
- 483 Fokianos, K., Tjøstheim, D., 2011. Log-linear poisson autoregression. *Journal of*
484 *Multivariate Analysis* 102, 563–578.
- 485 Fokianos, K., Tjøstheim, D., 2012. Nonlinear poisson autoregression. *Annals of the*
486 *Institute of Statistical Mathematics* 64, 1205–1225.
- 487 Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative
488 observations. *Journal of Business & Economic Statistics* 7, 407–417.

- 489 Jørgensen, B., 1987. Exponential dispersion models. *Journal of the Royal Statistical*
490 *Society: Series B (Methodological)* 49, 127–145.
- 491 Jørgensen, B., Lundbye-Christensen, S., Song, P.K., Sun, L., 1999. A state space
492 model for multivariate longitudinal count data. *Biometrika* 86, 169–181.
- 493 Lindsay, B.G., 1983a. The geometry of mixture likelihoods: a general theory. *The*
494 *Annals of Statistics* , 86–94.
- 495 Lindsay, B.G., 1983b. The geometry of mixture likelihoods, part ii: the exponential
496 family. *The Annals of Statistics* 11, 783–792.
- 497 Nelder, J.A., Wedderburn, R.W., 1972. Generalized linear models. *Journal of the*
498 *Royal Statistical Society: Series A (General)* 135, 370–384.
- 499 Neumann, M.H., 2011. Absolute regularity and ergodicity of poisson count processes.
500 *Bernoulli* 17, 1268–1284.
- 501 Oldham, K.B., Jerome, S., 1974. *The Fractional Calculus*. Academic Press, nc.
- 502 Samia, N.I., Chan, K.S., 2011. Maximum likelihood estimation of a generalized
503 threshold stochastic regression model. *Biometrika* 98, 433–448.
- 504 Särkkä, S., 2013. *Bayesian Filtering and Smoothing*. Cambridge University Press.
- 505 Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and
506 forecasting using the em algorithm. *Journal of Time Series Analysis* 3, 253–264.
- 507 Tjøstheim, D., 1990. Non-linear time series and markov chains. *Advances in Applied*
508 *Probability* 22, 587–611.
- 509 Tweedie, R., 1976. Criteria for classifying general markov chains. *Advances in*
510 *Applied Probability* 8, 737–771.
- 511 Van der Vaart, A., 1996. Efficient maximum likelihood estimation in semiparametric
512 mixture models. *The Annals of Statistics* 24, 862–878.
- 513 Wang, C., Liu, H., Yao, J.F., Davis, R.A., Li, W.K., 2014. Self-excited threshold
514 poisson autoregression. *Journal of the American Statistical Association* 109, 777–
515 787.
- 516 Wu, C.J., 1983. On the convergence properties of the em algorithm. *The Annals of*
517 *Statistics* , 95–103.
- 518 Xia, Y., Li, W.K., Tong, H., 2007. Threshold variable selection using nonparametric
519 methods. *Statistica Sinica* 17, 265–S57.
- 520 Zeger, S.L., 1988. A regression model for time series of counts. *Biometrika* 75,
521 621–629.

522 **Appendix**

523 In this section, the mathematical proofs of the main theoretical results of the
 524 paper, particularly Lemmas 2.1 and 2.2, and Theorems 2.1 and 2.2, are presented.
 525 The proofs of these are mainly shown within the conventional QML estimation
 526 literature, particularly the two main steps. The first step is to show the almost sure
 527 pointwise convergence, then to establish the almost sure stochastic equi-continuity.
 528 Hereafter, we omit \mathbb{I}_{j_k} for notational simplicity.

529 *Proof of Lemma 2.1*

530 This proof shows the almost sure equivalence between X_t and \tilde{X}_t uniformly over
 531 $\vartheta \in D$ in the two steps mentioned above, under Assumptions 2.1 and 2.2. Firstly,
 532 let us approximate \tilde{X}_t by using the Taylor expansion of a logarithmic function such
 533 that $\tilde{X}_t = X_t + R_t$, where $R_t = \frac{\Delta_T \zeta_t}{\mu_t} + \sum_{l=2} \frac{(-1)^{(l+1)}}{l} \left(\frac{\Delta_T \zeta_t}{\mu_t} \right)^l$. The first main term
 534 in R_t , particularly $\frac{\Delta_T \zeta_t}{\mu_t} \equiv R_{1,t}$, is then easily shown to be $o(T^{-1/2})$ *a.s.* as follows.
 535 By using the Cauchy–Schwartz inequality,

$$\begin{aligned} E \left(\frac{\Delta_T \zeta_t}{\mu_t} \right) &\leq \Delta_T \left\{ E \left(\frac{1}{\mu_t^2} \right)^{1/2} E(\zeta_t^2)^{1/2} \right\} \\ &= o(c^{T/2}) \end{aligned}$$

536 then apply the Markov inequality and Borel–Cantelli Lemma.

537 The proof is completed by showing the almost sure stochastic equi-continuity of
 538 $R_{1,t}$ as follows

$$\begin{aligned} \sup_{\|\vartheta - \tilde{\vartheta}\| < \delta} \left| R_{1,t}(\vartheta) - R_{1,t}(\tilde{\vartheta}) \right| &\leq \sup_{\|\vartheta - \tilde{\vartheta}\| < \delta} \left\{ \left\| R_{1,t}^{(1)}(\bar{\vartheta}_{-d}) \right\| + \left\| R_{1,t}(d) - R_{1,t}(\tilde{d}) \right\| \right\} \cdot \|\vartheta - \tilde{\vartheta}\| \\ &= o(1) \text{ a.s.}, \end{aligned} \tag{A.1.1}$$

539 where $\tilde{\vartheta}$ is an δ -neighbourhood of ϑ such that $\lim_{\delta \rightarrow 0} \sup_{\|\vartheta - \tilde{\vartheta}\| < \delta} \|\vartheta - \tilde{\vartheta}\| \rightarrow 0$, $\bar{\vartheta}_{-d}$ lies on the

540 line segment of $\{\rho\vartheta + (1-\rho)\tilde{\vartheta}; \rho \in (0, 1)\}$, $R_{1,t}^{(1)}(\bar{\vartheta}_{-d})$ denotes the gradients of $R_{1,t}$ with
 541 respect to the vector of the parameters, $\bar{\vartheta}_{-d} = (\bar{a}'s, \bar{c}'s, \bar{\beta}'s, \bar{c}, \bar{\pi}'s)^\top$. (A.1.1) can be
 542 established by showing that $\left\| R_{1,t}^{(1)}(\bar{\vartheta}_{-d}) \right\| = O(1)$ *a.s.* because $\|R_{1,t}(d) - R_{1,t}(\tilde{d})\| =$
 543 $o(1)$ *a.s.*, given the discreteness of d 's. Now let us consider $R_{1,t}^{(1)}(\bar{\vartheta}_{-d})$ as follows

$$\begin{aligned} E \left(\left\| R_{1,t}^{(1)}(\bar{\vartheta}_{-d}) \right\| \right) &\leq E \|R_{a_{k,0,1,t}}^{(1)}\| + E \|R_{a_{j_k,1,t}}^{(1)}\| + E \|R_{\tau_{j_k,1,t}}^{(1)}\| + E \|R_{\pi_{k,1,t}}^{(1)}\| \\ &\quad + E \|R_{\beta_k}^{(1)}\| + E |R_{c,t}^{(1)}| \\ &= O \left(\left(\frac{T}{2} + \delta \right) c^{\frac{T-2}{2} + \delta} \right) \end{aligned}$$

544 with the similar arguments to those. The each components of $R_{1,t}^{(1)}(\bar{\vartheta}_{-d})$ are $R_{a_{k,0,1,t}}^{(1)} =$
 545 $-\frac{\Delta_T \zeta_t \pi_k}{\mu_t}$, $R_{a_{j_k,1,t}}^{(1)} = -\frac{\Delta_T \zeta_t \{\pi_k X_{k,t-l}\}}{\mu_t}$, $R_{\tau_{j_k,1,t}}^{(1)} = -\frac{\Delta_T \zeta_t}{\mu_t} \pi_k a_{k,d_k,j_k}$, $R_{\pi_{k,t}}^{(1)} = -\frac{\Delta_T \zeta_t \ln \mu_{k,t}}{\mu_t}$,

546 $R_{\beta_k}^{(1)} = \frac{\Delta_T \zeta_t}{2\mu_t} \pi_k V(I_{k,t}; \mu_{k,t})$ and $R_{c,1,t}^{(1)} = \frac{\zeta_t}{\mu_t} O\left(\left(\frac{T}{2} + \delta\right) c^{\frac{T-2}{2} + \delta}\right)$. Note that the deriva-
 547 tive of $X_{k,t}$ with respect to τ_{j_k} is not well defined, so we set $\frac{\partial X_{k,t}}{\partial \tau_{j_k}} = a_{k,d_k,j_k}$, following
 548 the arguments in Chan and Tsay (1998). By applying the Markov inequality and
 549 Borel–Cantelli Lemma, $\|R_{1,t}^{(1)}(\bar{\vartheta}_{-d})\| = o(1)$ *a.s.* \square

550 *Proof of Lemma 2.2*

551 This proof establishes (A.2.1) below, under Assumptions 2.1 to 2.3 and the
 552 independence assumption of the data generating processes of each cluster.

$$\begin{aligned} \sup_{\psi \in D_\psi} |\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi)| &= \sup_{\psi \in D_\psi} \left| \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \left\{ \sum_{k=1}^K \left(\ln \widehat{ED}_{k,t}^* - \ln ED_{k,t}^* \right) \right\} \right| \\ &= O(T^{-1/2}) \text{ a.s.} \end{aligned} \quad (\text{A.2.1})$$

553 Let us firstly consider the strong pointwise convergence of $\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi)$. By
 554 using the Taylor expansion argument of a logarithmic function and the results in
 555 Chapters 3 and 7 of Caines (1987), particularly $\hat{\mu}_{k,t|t-1} = \mu_{k,t} + o(c^{T/2})$ *a.s.* and
 556 $\hat{\pi}_k = \pi_k + O(T^{-1/2})$ *a.s.*, and hence

$$\ln \widehat{ED}_{k,t}^* = \ln ED_{k,t}^* + \frac{ED_{k,t,\pi}^{*(1)}(\hat{\pi} - \pi_k)}{ED_{k,t}^*} + \frac{ED_{k,t,\mu}^*(\hat{\mu}_{k,t|t-1} - \mu_{k,t})}{ED_{k,t}^* + ED_{k,t,\pi}^{*(1)}(\hat{\pi} - \pi)} + o(c^{T/2}) \text{ a.s.}$$

557 We can then rewrite $\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi)$ as follows

$$\begin{aligned} &\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi) \\ &= \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \sum_{k=1}^K \left\{ \frac{ED_{k,t,\pi}^{*(1)}(\hat{\pi} - \pi_k)}{ED_{k,t}^*} + \frac{ED_{k,t,\mu}^{*(1)}(\hat{\mu}_{k,t|t-1} - \mu_{k,t})}{ED_{k,t}^* + ED_{k,t,\pi}^{*(1)}(\hat{\pi} - \pi)} \right\} + o(T^{-1/2}) \text{ a.s.} \end{aligned}$$

558 By using the Cauchy–Schwartz inequality,

$$\begin{aligned} &E(\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi))^2 \\ &= \sum_{k=1}^K \left(\frac{1}{T^2} \sum_{t=p_K+1}^{T+p_K} E \left\{ \left((\hat{\pi}_k - \pi_k)^2 \left(\frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t}^*} \right)^2 + \left(\frac{ED_{k,t,\mu}^{*(1)}(\hat{\mu}_{k,t|t-1} - \mu_{k,t})}{ED_{k,t}^* + ED_{k,t,\pi}^{*(1)}(\hat{\pi}_k - \pi_k)} \right)^2 \right. \right. \\ &\quad \left. \left. + 2(\hat{\pi}_k - \pi_k)(\hat{\mu}_{k,t|t-1} - \mu_{k,t}) \frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t}^*} \frac{ED_{k,t,\mu}^{*(1)}}{ED_{k,t}^* + ED_{k,t,\pi}^{*(1)}(\hat{\pi}_k - \pi_k)} \right) \right\} \\ &\quad + 2 \frac{1}{T^2} \sum_{t=p_K+1}^{T+p_K} \sum_{\ell \neq t}^{T+p_K} E \left\{ (\hat{\pi}_k - \pi_k)^2 \left(\frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t}^*} \right) \left(\frac{ED_{k,\ell,\pi}^{*(1)}}{ED_{k,\ell}^*} \right) \right. \\ &\quad \left. + \left(\frac{ED_{k,t,\mu}^{*(1)}(\hat{\mu}_{k,t|t-1} - \mu_{k,t})}{ED_{k,t}^* + ED_{k,t,\pi}^{*(1)}(\hat{\pi}_k - \pi_k)} \right) \left(\frac{ED_{k,\ell,\mu}^{*(1)}(\hat{\mu}_{k,\ell|\ell-1} - \mu_{k,\ell})}{ED_{k,\ell}^* + ED_{k,\ell,\pi}^{*(1)}(\hat{\pi}_k - \pi_k)} \right) \right. \\ &\quad \left. + 2(\hat{\pi}_k - \pi_k)(\hat{\mu}_{k,\ell|\ell-1} - \mu_{k,\ell}) \left(\frac{ED_{k,t,\pi}^{*(1)}}{ED_{k,t}^*} \right) \left(\frac{ED_{k,\ell,\mu}^{*(1)}}{ED_{k,\ell}^* + ED_{k,\ell,\pi}^{*(1)}(\hat{\pi}_k - \pi_k)} \right) \right\} \\ &= O(T^{-1}), \end{aligned}$$

559 particularly under Assumption 2.3. Then, by applying the Chebyshev inequality
560 and Borel–Cantelli Lemma,

$$\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi) = O(T^{-1/2}) \text{ a.s.}$$

561 The next step is then to show the stochastic equi-continuity of $\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi)$.
562 Hereafter, let us denote $\mathcal{L}(\psi|\mathcal{F}_{t-1}) - \mathcal{L}^*(\psi)$ as $\mathcal{L}_1(\psi)$ for the sake of notational
563 simplicity.

$$\begin{aligned} \sup_{\|\psi - \tilde{\psi}\| < \delta} \left| \mathcal{L}_1(\psi) - \mathcal{L}_1(\tilde{\psi}) \right| &\leq \sup_{\|\psi - \tilde{\psi}\| < \delta} \left\{ \|\mathcal{L}_1^{(1)}(\bar{\psi}_{-d})\| + |\mathcal{L}_1(d) - \mathcal{L}_1(\tilde{d})| \right\} \cdot \|\psi - \tilde{\psi}\| \\ &= o(1) \text{ a.s.}, \end{aligned}$$

564 where $\mathcal{L}_1^{(1)}(\bar{\psi}_{-d})$ denotes the first gradients of $\mathcal{L}_1(\bar{\psi})$ with respect to $\bar{\psi}_{-d} = (\bar{a}'s, \bar{c}'s, \bar{\beta}'s)^\top$.
565 Hence, it is

$$\frac{\partial \mathcal{L}_1(\bar{\psi}_{-d})}{\partial \bar{\psi}_{-d}} = \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \left\{ \sum_{k=1}^K \left(\frac{\widehat{ED}_{k,t,\mu}^{*(1)}(\hat{\mu}_{k,t|t-1})'_{\bar{\psi}_{-d}}}{\widehat{ED}_{k,t}^*} - \frac{ED_{k,t,\mu}^{*(1)}(\mu_{k,t|t-1})'_{\bar{\psi}_{-d}}}{ED_{k,t}^*} \right) \right\}, \quad (\text{A.2.2})$$

566 where $\widehat{ED}_{k,t,\mu}^{*(1)}$ is the first derivative of $\widehat{ED}_{k,t}^*$ with respect to $\hat{\mu}_{k,t|t-1}$, and $(\mu_{k,t})'_{\bar{\psi}_{-d}}$
567 and $(\hat{\mu}_{k,t|t-1})'_{\bar{\psi}_{-d}}$ denote the first gradients of $\mu_{k,t}$ and $\hat{\mu}_{k,t|t-1}$ with respect to $\bar{\psi}_{-d}$,
568 respectively, and which are as follows

$$(\mu_{k,t})'_{\bar{\psi}_{-d}} = \left(\mu_{k,t} \quad \mu_{k,t} X_{k,t-l} \quad \mu_{k,t} a_{k,j,d-} \quad \mu_{k,t} (X_{k,t})'_{\bar{\beta}_k} \right)^\top$$

569 and

$$(\hat{\mu}_{k,t|t-1})'_{\bar{\psi}_{-d}} = \begin{pmatrix} \hat{\mu}_{k,t|t-1} \\ \hat{\mu}_{k,t|t-1} \hat{X}_{k,t-l|t-1} \\ \hat{\mu}_{k,t|t-1} a_{k,j,d-} \\ \hat{\mu}_{k,t|t-1} \left\{ (X_{k,t})'_{\bar{\beta}_k} + (X_{k,t})''_{\bar{\beta}_k, \bar{\mu}_{k,t}} (\hat{X}_{k,t|t-1} - X_{k,t}) \right\} \end{pmatrix}$$

570 with $\hat{X}_{k,t-l|t-1}$ denoting the minimum conditional mean-squared error estimate of
571 $X_{k,t-l}$ given \mathcal{F}_{t-1} , and $(X_{k,t})'_{\bar{\beta}_k} = \frac{V(I_{k,t}; \mu_{k,t})}{2\mu_{k,t}^2}$ and $(X_{k,t})''_{\bar{\beta}_k, \bar{\mu}_{k,t}} = -\frac{V(I_{k,t}; \mu_{k,t})}{\mu_{k,t}^3}$. We next
572 use the Taylor expansion argument below

$$\widehat{ED}_{k,t,\mu}^{*(1)} = ED_{k,t,\mu}^{*(1)} + ED_{k,t,\mu}^{*(2)}(\hat{\mu}_{k,t|t-1} - \mu_{k,t}) + o(T^{-1/2}) \text{ a.s.},$$

573 (A.2.2) is then rewritten as follows

$$\mathcal{L}_1^{(1)}(\bar{\psi}_{-d}) = \mathcal{L}_{11}^{(1)}(\bar{\psi}_{-d}) + \mathcal{L}_{12}^{(1)}(\bar{\psi}_{-d}) + O(T^{-1}) \text{ a.s.},$$

574 where

$$\begin{aligned} & \mathcal{L}_{11}^{(1)}(\bar{\psi}_{-d}) \\ &= \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \left\{ \sum_{k=1}^K \frac{\left(ED_{k,t}^* ED_{k,t,\mu}^{*(2)} (\hat{\mu}_{k,t|t-1})'_{\bar{\psi}_{-d}} - ED_{k,t,\mu}^{*(1)} ED_{k,t,\pi}^{*(1)} (\mu_{k,t|t-1})'_{\bar{\psi}_{-d}} \right) (\hat{\pi}_k - \pi_k)}{ED_{k,t}^* \left(ED_{k,t}^* + ED_{k,t,\pi}^{*(1)} (\hat{\pi}_k - \pi_k) + ED_{k,t,\mu}^{*(1)} (\hat{\mu}_{k,t|t-1} - \mu_{k,t}) \right) + o(T^{-1})} \right\} \end{aligned}$$

575 and

$$\begin{aligned} & \mathcal{L}_{12}^{(1)}(\bar{\psi}_{-d}) \\ &= \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \sum_{k=1}^K \left\{ \frac{ED_{k,t}^* ED_{k,t,\mu}^{*(1)} \left\{ (\hat{\mu}_{k,t|t-1})'_{\bar{\psi}_{-d}} - (\mu_{k,t|t-1})'_{\bar{\psi}_{-d}} \right\}}{ED_{k,t}^* \left(ED_{k,t}^* + ED_{k,t,\pi}^{*(1)} (\hat{\pi}_k - \pi_k) + ED_{k,t,\mu}^{*(1)} (\hat{\mu}_{k,t|t-1} - \mu_{k,t}) \right) + o(T^{-1})} \right. \\ & \quad \left. + \frac{\left(ED_{k,t}^* ED_{k,t,\mu}^{*(2)} (\hat{\mu}_{k,t|t-1})'_{\bar{\psi}_{-d}} - \left(ED_{k,t,\mu}^{*(1)} \right)^2 (\mu_{k,t|t-1})'_{\bar{\psi}_{-d}} \right) (\hat{\mu}_{k,t|t-1} - \mu_{k,t})}{ED_{k,t}^* \left(ED_{k,t}^* + ED_{k,t,\pi}^{*(1)} (\hat{\pi}_k - \pi_k) + ED_{k,t,\mu}^{*(1)} (\hat{\mu}_{k,t|t-1} - \mu_{k,t}) \right) + o(T^{-1})} \right\}. \end{aligned}$$

576 By applying the Cauchy–Schwartz and Chebyshev inequalities, and the Borel–Cantelli
577 lemma to $\mathcal{L}_{11}^{(1)}(\bar{\psi}_{-d})$, and the Markov inequality and Borel–Cantelli lemma to $\mathcal{L}_{12}^{(1)}(\bar{\psi}_{-d})$,
578 respectively, we obtain $\mathcal{L}_1^{(1)}(\bar{\psi}_{-d}) = o(1)$ *a.s.* \square

579 *Proof of Theorem 2.1*

580 This proof can be shown in the two steps under Assumptions 2.1 to 2.4, with the
581 independence assumption on the data generating process of each cluster. The first
582 step is to show the almost sure convergence of $\hat{\psi}$ to ψ uniformly over D_ψ by using
583 similar arguments to those in Lemma 2.2. We can then verify the identification
584 condition of ψ_0 .

585 The first step can be shown by establishing (A.3.1) below

$$\sup_{\psi \in D_\psi} |\mathcal{L}^*(\psi) - E\mathcal{L}^*(\psi)| = O(T^{-1/2}) \text{ a.s.} \quad (\text{A.3.1})$$

586 Firstly, it is

$$\begin{aligned} E(\mathcal{L}^*(\psi) - E\mathcal{L}^*(\psi))^2 &= \frac{1}{T^2} \sum_{t=p_K+1}^{T+p_K} \left\{ \sum_{k=1}^K E(\ln ED_{k,t}^* - E \ln ED_{k,t}^*)^2 \right\} \\ & \quad + 2 \frac{1}{T^2} \sum_{t=p_K+1}^{T+p_K} \sum_{\iota \neq t}^{T+p_K} \left\{ \sum_{k=1}^K E(\{\ln ED_{k,t}^* - E \ln ED_{k,t}^*\} \{\ln ED_{k,\iota}^* - E \ln ED_{k,\iota}^*\}) \right\} \end{aligned}$$

587 then apply the Chebyshev inequality and Borel–Cantelli lemma. We thus obtain
588 that $\mathcal{L}^*(\psi) = E\mathcal{L}^*(\psi) + O(T^{-1/2})$ *a.s.* The next step is to show the stochas-
589 tic equi-continuity of $\mathcal{L}^*(\psi) - E\mathcal{L}^*(\psi)$. This can be established by showing that

590 $E \left(\frac{\partial \mathcal{L}^*(\psi_{-d})}{\partial \psi_{-d}} \right)^2 = O(T^{-1})$ as follows

$$\begin{aligned}
 & E \left(\frac{\partial \mathcal{L}^*(\psi_{-d})}{\partial \psi_{-d}} \right)^2 \\
 &= \sum_{k=1}^K \frac{1}{T^2} \left(\sum_{t=1}^T \left\{ E \left(\frac{ED_{k,t}^{*(1)}}{ED_{k,t}^*} \right)^2 + \sum_{t=1}^T \sum_{\iota \neq t} E \left(\frac{ED_{k,t}^{*(1)}}{ED_{k,t}^*} \frac{ED_{k,\iota}^{*(1)}}{ED_{k,\iota}^*} \right) \right\} \right) \\
 &= O(T^{-1}) \tag{A.3.2}
 \end{aligned}$$

591 particularly under Assumptions 2.4 (i) and (ii). By applying the Chebyshev inequality and Borel–Cantelli lemma to (A.3.2), (A.3.1) is shown.

592 The identification condition of ψ_0 is then verified by using the counter argument
 593 as follows. Consider the Jensen’s inequality in (A.3.3), taking the expectation with
 594 ψ_0 as follows

$$\begin{aligned}
 E\mathcal{L}^*(\psi) - E\mathcal{L}^*(\psi_0) &\leq \frac{1}{T} \sum_{t=p_K+1}^{T+p_K} \left\{ \sum_{k=1}^K \ln E \left(\frac{ED^*(\mu_{k,t}, \beta_k, \pi_k)}{ED^*(\mu_{0,k,t}, \beta_{0,k}, \pi_k)} \right) \right\} \\
 &\leq 0. \tag{A.3.3}
 \end{aligned}$$

595 The equality of (A.3.3) holds when $\psi \rightarrow \psi_0$. Hence, ψ_0 is not uniquely identified
 596 when there is a sequence such that $\psi_T \in D_\delta(\psi^*)$ converges to $\psi^* \in \bar{D}_\delta(\psi_0) \cap D_\psi$,
 597 where $D_\delta(\cdot)$ and $\bar{D}_\delta(\cdot)$ represent an open δ -neighbourhood and its complement,
 598 respectively, and $\lim_{T \rightarrow \infty} E\mathcal{L}^*(\psi^*) \rightarrow \lim_{T \rightarrow \infty} E\mathcal{L}^*(\psi_0)$. Hence, the unique identification

599 condition requires that $\limsup_{T \rightarrow \infty} \left(\max_{\psi \in \bar{D}_\delta(\psi_0) \cap D_\psi} E\mathcal{L}^*(\psi) \right) \neq E\mathcal{L}^*(\psi_0)$ for any ψ . \square

600 *Proof of Theorem 2.2*

601 The asymptotic normality of our proposed QMLEs can be obtained by consid-
 602 ering the extension of the mean value theorem of $\frac{\partial \mathcal{L}^*(\hat{\psi}^*)}{\partial \psi^*}$ as follows

$$\frac{\partial \mathcal{L}^*(\hat{\psi}^*)}{\partial \psi^*} = \frac{\partial \mathcal{L}^*(\psi_0^*)}{\partial \psi^*} + \frac{\partial^2 \mathcal{L}^*(\bar{\psi}^*)}{\partial \psi^* \partial \psi^{*\top}} (\hat{\psi}^* - \psi_0^*),$$

603 where $\bar{\psi}^*$ is between $\hat{\psi}^*$ and ψ_0^* . Therefore, the asymptotic normality of $\hat{\psi}^*$ can be ob-
 604 tained by showing that $\sqrt{T} \frac{\partial \mathcal{L}^*(\psi^*)}{\partial \psi^*} \Big|_{\psi^*=\psi_0^*} \sim N(0, A_0(\psi_0^*))$ and $\frac{\partial^2 \mathcal{L}^*(\psi^*)}{\partial \psi^* \partial \psi^{*\top}} = \lim_{T \rightarrow \infty} E \frac{\partial^2 \mathcal{L}^*(\psi^*)}{\partial \psi^* \partial \psi^{*\top}} +$
 605 $O_P(T^{-1/2})$ uniformly over D_{ψ^*} . Particularly under Assumptions 2.4 (i) and (iii),
 606 $\sqrt{T} \frac{\partial \mathcal{L}^*(\psi^*)}{\partial \psi^*} \sim N(0, A_0(\psi_0^*))$ can be easily shown by using the small and large blocks
 607 arguments (see Chapter 2 of Fan and Yao (2008) for details).

608 The last step of this proof is to establish below

$$\sup_{\psi^* \in D_{\psi^*}} \|B_T(\psi^*) - B_0(\psi^*)\|_F = O_P(T^{-1/2}), \tag{A.4.1}$$

610 where $\|\cdot\|_F$ denotes the Frobenius norm, $B_T(\psi^*) = \frac{\partial^2 \mathcal{L}^*(\psi^*)}{\partial \psi^* \partial \psi^{*\top}}$ and $B_0(\psi^*) = \lim_{T \rightarrow \infty} E \frac{\partial^2 \mathcal{L}^*(\psi^*)}{\partial \psi^* \partial \psi^{*\top}}$.

611 The result of (A.4.1) is obtained by applying the Chebyshev inequality. Next, the
 612 stochastic equi-continuity of $B_T(\psi^*)$ can be established by showing that

$$\|C_T(\bar{\psi}^*) - C_0(\bar{\psi}^*)\|_F \leq \|C_T(\bar{\psi}^*)\|_F = O_p(T^{-1/2}), \quad (\text{A.4.2})$$

613 where $C_T(\bar{\psi}^*) = \frac{\partial B_T(\bar{\psi}^*)}{\partial \bar{\psi}^*}$ and $C_0(\bar{\psi}^*) = \lim_{T \rightarrow \infty} E C_T(\bar{\psi}^*)$. The result of (A.4.2) is then
 614 obtained by applying the Chebyshev inequality. \square