Two questions on linear QBD's (quasi-birth-and-death) processes

Florin Avram

March 8, 2011

Contents

1 Quasi birth and death processes

Motivation. Many important stochastic models involve multidimensional random walks with discrete state space, whose coordinates split naturally into

1. **one infinite valued coordinate** $N(t) \in \mathbb{N}$ called **level**, and
2. the "rest of the information" $I(t)$, called **phase**, which takes a **finite number of possible values**.

Partitioned according to the level, the infinitesimal generator Q of such a Markov process $(N(t), I(t))$, is a **block tridiagonal** matrix, called **level-dependent quasi-birth-and-death** generator (LDQBD):

$$Q = \begin{bmatrix} B_0 & A_0 & & & \\ C_1 & B_1 & A_1 & & \\ & C_2 & B_2 & A_2 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{1}$$

(recall that Q is a matrix with nonnegative off-diagonal "rates/weights", and with row sums equal to 0).

QBD processes share the "skip free" structure of birth and death processes; however, the "**level transition weights**" A_n, B_n, C_n are now **matrices**, inviting one to enter the non-commutative world.

Level dependent QBD's. It seems fair to say that not much is known about level dependent QBD's, maybe due to the generality of the model. Note however that for the case of linear or polynomial dependence on the level, an important analytic-algebraic approach is available, namely the classic formal series approach, which motivates our interest in this particular case.

Stationary distributions. One challenging problem of great interest for **positive recurrent** level dependent QBD processes is that of computing the stationary distribution $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$ partitioned by level, where $\boldsymbol{\pi}_n = (\pi_{n,0}, \pi_{n,1}, \ldots, \pi_{n,s})$.

The equilibrium equations

$$\boldsymbol{\pi} Q = 0 \tag{2}$$

in partitioned form yield the second degree vector recursion:

$$\begin{cases} \boldsymbol{\pi}_{n-1} A_{n-1} + \boldsymbol{\pi}_n B_n + \boldsymbol{\pi}_{n+1} C_{n+1} = 0, \qquad n = 1, 2, \cdots \\ \boldsymbol{\pi}_0 B_0 + \boldsymbol{\pi}_1 C_1 = 0 \end{cases} \tag{3}$$

Note the absence of an initial value for $\boldsymbol{\pi}_0$.

2 The QBD of the linear retrial/overflow model

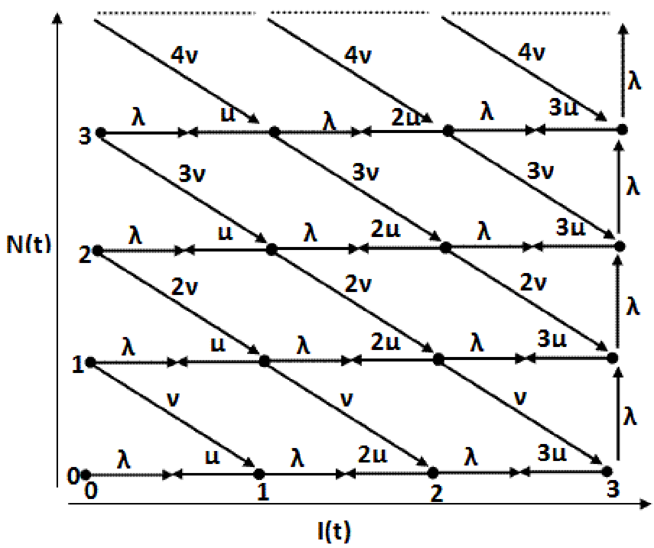

Figure 1: States and transitions of the classic retrial queue $M/M/3/3$ (all arrivals retry, retrials persist until served).

The arrivals and "loops" are:

$$A = \begin{bmatrix} 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 0 & \lambda_s \end{bmatrix},$$

$$B = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \dots & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots & 0 \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \lambda \\ 0 & \cdots & \cdots & \cdots & s\mu & -s\mu \end{bmatrix}$$

The departures matrix for the linear retrial model with **abandon probability** $\bar{p} = 1 - p$ is given by $C_n = nC$, where:

$$C = \begin{bmatrix} \nu\bar{p} & \nu p & 0 & \dots & 0 \\ 0 & \nu\bar{p} & \nu p & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \nu\bar{p} & \nu p \\ 0 & \dots & \dots & \dots & \nu_s \end{bmatrix}.$$

Remark 1. *The A, B, C entries correspond respectively to upward, level and downward arrows in the figure 1.*

Definition 1. *An* **affine death QBD** *is a QBD with affine dependence on the level*

$$A_n = A, \quad C_n = nC + C^{(0)} 1_{\{n>0\}}, \tag{4}$$
$$B_n = B - \tilde{A} - n\tilde{C} - \tilde{C}^{(0)} 1_{\{n>0\}}$$

where $\tilde{A}, \tilde{C}^{(0)}, \tilde{C}$ *contain on their diagonals the sum of the rows of* $A, C^{(0)}, C$.

The two terms in C_n correspond respectively to "**infinite server**" and "**one server**" in the secondary service area/orbit, and we are mainly interested here in the linear case with $C_n = nC, C^{(0)} = 0$.

A classic approach for tackling the recursion (3) is via the **generating functions**

$$p_i(z) = \sum_{n=0}^{\infty} \pi_{i,n} z^n, i = 0, ..., s, \; \Leftrightarrow \mathbf{p}(z) = \sum_{n=0}^{\infty} \boldsymbol{\pi}_n z^n, \tag{5}$$

where $\mathbf{p}(z) = (p_0(z), \cdots, p_s(z))$.

For affine QBD's (4), the recursion

$$\begin{cases} \boldsymbol{\pi}_{n-1} A + \boldsymbol{\pi}_n (B - \tilde{A} - n\tilde{C} - C^{(0)}) + \boldsymbol{\pi}_{n+1}((n+1)C + C^{(0)}) = 0, \; n = 1, 2 \\ \boldsymbol{\pi}_0 (B - \tilde{A}) + \boldsymbol{\pi}_1 (C + C^{(0)}) = 0 \end{cases} \tag{6}$$

results in a linear differential system [§]

$$\boxed{\mathbf{p}'(z) V(z) = \mathbf{p}(z) U(z) + \boldsymbol{\pi}_0 (\tilde{C}^{(0)} - z^{-1} C^{(0)})} \tag{7}$$

[§]Multiplying the equilibrium equations (6) by z^n and summing up gives rise to a linear first order differential system:

$$\sum_{n=0}^{\infty} z^n \boldsymbol{\pi}_{n-1} A + z^n \boldsymbol{\pi}_n (B - \tilde{A}) - \boldsymbol{\pi}_n z^n (n\tilde{C} + \tilde{C}^{(0)} 1_{\{n>0\}}) + z^n \boldsymbol{\pi}_{n+1}((n+1)C + C^{(0)})$$
$$= z\boldsymbol{p}(z) A + \boldsymbol{p}(z)(B - \tilde{A} - \tilde{C}^{(0)}) + \boldsymbol{\pi}_0 \tilde{C}^{(0)} - z\boldsymbol{p}'(z)\tilde{C} + \boldsymbol{p}'(z) C + z^{-1}(\boldsymbol{p}(z) - \boldsymbol{\pi}_0) C^{(0)} = 0$$

yielding (7)

where

$$\boxed{U(z) = B + zA - \tilde{A} + z^{-1}C^{(0)} - \tilde{C}^{(0)}}, \tag{8}$$

$$\boxed{V(z) = z\tilde{C} - C} = \nu \begin{bmatrix} z-\bar{p} & -p & 0 & \dots & 0 \\ 0 & z-\bar{p} & -p & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & z-\bar{p} & -p \\ 0 & \dots & \dots & \dots & \bar{\alpha}(z-1) \end{bmatrix} \tag{9}$$

are square matrices of order $(s+1)$, and where $\bar{\alpha} = 1 - \alpha = \nu_s/\nu$.

Remark 2. *The singularities are the roots of* $Det(V(z)) = \bar{\alpha}(z-\bar{p})^s(z-1)$. *It turns out however that the singularity at* 1 *may be simplified after replacing the last equation by the sum of the equations, arriving thus at a system* $\mathbf{p}'(z)V(z) = \mathbf{p}(z)U(z)$ *with:*

$$V(z) = \begin{bmatrix} z-\bar{p} & -p & 0 & \dots & 1 \\ 0 & z-\bar{p} & -p & \dots & 1 \\ \vdots & \ddots & \ddots & -p & \vdots \\ 0 & \dots & \dots & z-\bar{p} & 1 \\ 0 & \dots & \dots & \dots & \bar{\alpha} \end{bmatrix} \tag{10}$$

$$U(z) = \begin{bmatrix} \tilde{\lambda}(z\bar{p}-1) & \tilde{\lambda}p & 0 & \dots & \tilde{\lambda}\bar{p} \\ \tilde{\mu} & \tilde{\lambda}(z\bar{p}-1)-\tilde{\mu} & \tilde{\lambda}p & \dots & \tilde{\lambda}\bar{p} \\ 0 & 2\tilde{\mu} & \tilde{\lambda}(z\bar{p}-1)-2\tilde{\mu} & \dots & \vdots \\ \vdots & \ddots & \ddots & \tilde{\lambda}p & \tilde{\lambda}\bar{p} \\ \vdots & \ddots & \ddots & \tilde{\lambda}(z\bar{p}-1)-(s-1)\tilde{\mu} & \tilde{\lambda}\bar{p} \\ 0 & \dots & \dots & s\tilde{\mu} & \tilde{\lambda}_s \end{bmatrix}$$

One needs to distinguish further two cases:

1. $\bar{\alpha} = 0$, when the determinant is 0 identically, and the system may be replaced by **an equivalent system of dimension smaller by one**. The determinant of the reduced system is

   $$Det(V(z)) = (z - \bar{p})^{s-1}(z - \bar{p} - \tilde{\rho}) = 0, \text{ where } \tilde{\rho} = \frac{s\mu}{\lambda_s}.$$

   For $s \geq 3$, $\bar{p}$ is an irregular singularity.

   The ergodicity condition is precisely $\bar{p} + \tilde{\rho} > 1$, and the asymptotic behavior may be determined by expanding around the (regular) singularity $\bar{p} + \tilde{\rho}$.

2. $\bar{\alpha} > 0$, in which case the system is always ergodic, and the asymptotic behavior is different.

Remark 3. *Sometimes the system* (7) *may be put into the "canonical Okubo" form, characterized by*

$$V(z) = zI - T, \quad U(z) = U.$$

When T is diagonal, the solutions are classical hypergeometric functions, but this is not the case for the retrial queue with $s \geq 3, p > 0$.

Remark 4. *The existence of a* **unique** *"non-negative Perron-Frobenius/Krein-Rutman"/***minimal solution $\boldsymbol{\pi}$** *of the recursion* (3) *for chains with a unique recurrence class is easily argued via probabilistic arguments -recall the formula*

$\pi_{n,i} = T_{n,i}^{-1}$. *It follows that the corresponding generating function solving* (7) *must be unique, the only "boundary conditions" being the analyticity in the unit circle (in the ergodic case), the nonnegativity of the coefficients of the power series expansion of the generating function* $\mathbf{p}(z)$, *and the normalization* $\mathbf{p}(1)\mathbf{1} = 1$.

It is not easy to find in the literature an explanation of how this crucial uniqueness fits with the theory of formal series solving an ODE with polynomial coefficients and an irregular singularity. In fact, Maple reveals that the space of formal power series solving (7) *has dimension larger than* 1 *when* $s \geq 3$.

Q 1: An intriguing question is whether a simple constructive approach may yield the "Perron-Frobenius/Krein-Rutman" one dimensional subspace of generating functions corresponding to the stationary distribution.

3 The factorial moments

The generating function must be studied at several points: the irregular singularity, the dominant one, for asymptotics (when $\alpha = 0$), and at 1, which yields a practical approximation.

Indeed, consider the power series expansion around $y = 0$ of

$\mathbf{m}(y) := \mathbf{p}(y+1)$, whose coefficients are the factorial moments

$$\mathbf{m}_k := (m_{k,0}, ..., m_{k,s}),\ m_{k,i} = E[\frac{N(N-1)...(N-k+1)}{k!} 1_{\{I=i\}}]$$

of the orbit. These satisfy the recurrence

$$\begin{aligned} &(k+1)\mathbf{m}_{k+1}(\tilde{C} - C) + k\mathbf{m}_k \tilde{C} = \mathbf{m}_k(B + A - \tilde{A}) + \mathbf{m}_{k-1}A, k \geq 1 \\ &\mathbf{m}_1(\tilde{C} - C) = \mathbf{m}_0(B + A - \tilde{A}) \end{aligned} \quad (11)$$

With the exception of the case "Markov modulated M/M/∞ queue" when A, C are diagonal, and the moments recursion may further be reduced to:

$$\begin{aligned} &k\mathbf{m}_k C\ F = \mathbf{m}_{k-1} A\ F, k \geq 1 \\ &\mathbf{m}_0 B = 0, \end{aligned}$$

where F is a matrix with a first column of 1's and 0 else, the recursion (11) is of second degree.

Q 2: Provide converging approximations for the factorial moments (by a perturbation approach with respect to p –cf. O'Cinneide and Purdue, the factorial moments are explicit for the Markov modulated M/M/∞ queue obtained when $p = 0$, or by vector Pade methods, or?).

In fact, the essential quantities of interest for retrial/overflow queues are

1. the marginal stationary distribution of the primary service

area

$$\mathbf{p}(1) = \sum_{n=0}^{\infty} \boldsymbol{\pi}_n = \mathbf{m}_0 \tag{12}$$

2. the expected number in the orbit/secondary area, joint with a given occupancy

$$\mathbf{p}'(1) = \sum_{n=0}^{\infty} n\boldsymbol{\pi}_n = \mathbf{m}_1 \tag{13}$$

   (which determine the expected conditional arrival rates from the orbit into the main service area).

Indeed, as noted by Fredericks and Reisner, it follows from the equilibrium equations that the stationary distribution of the primary service area is proportional to the solutions of the recurrence

$$(i+1)\mu p_{i+1} = (\lambda + r_i)p_i, i = 0, ...s-1 \quad \Longrightarrow \quad p_i \sim \frac{\prod_{j=0}^{i-1}(\lambda + r_i)}{i!\mu^i}$$

see (2.27) [17], [50], where r_i, the conditional rates of retrial

$$r_i = E[\nu N(t)/I(t) = i] = \nu m_{1,i}/m_{0,i},$$

are however unknown.

Thus, the marginal stationary distribution of the number of busy servers is identical to that of a finite birth-and-death process with unknown birth rates $\lambda + r_k$ and death rates $k\mu$.

References

[1] Daniel Adelman (2008). A Simple Algebraic Approximation to the Erlang Loss System. Operations Research Letters, 36(4), pp. 484-91

[2] J.R. Artalejo (1999). Accessible bibliography on retrial queues, *Mathematical and Computer Modelling*, 30(3-4), 1–6.

[3] JR Artalejo, A Gomez-Corral, MF Neuts (2001). Analysis of multiserver queues with constant retrial rate. European Journal of Operational,

[4] J.R. Artalejo, A. Economou and A. Gómez-Corral (2008). Algorithmic analysis of the $Geo/Geo/c$ retrial queue, *European J. of OR*, 189(3), 1042–1056.

[5] J.R. Artalejo and A. Gómez-Corral (2008). *Retrial Queueing Systems*, Springer-Verlag, Berlin Heidelberg.

[6] J.R. Artalejo and M. Pozo (2002). Numerical calculation of the stationary distribution of the main multiserver retrial queue, *Annals of Operations Research*, 116, 41–56.

[7] W. Balser, C. Röscheisen and F. Steiner, Systems of linear ordinary differential equations - a review of three solution methods, Ulmer Seminare ber Funktionalanalysis und Differentialgleichungen, no. 11, 2006, 53-62

[8] L. Bright and P.G. Taylor (1995). Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes, *Stochastic Models*, 11, 497–525.

[9] L. Bright and P.G. Taylor (1997). Equilibrium distribution for level-dependent quasi-birth-and-death processes, in *Matrix-Analysis Methods in Stochastic Models*, edited by S.R. Chakravarthy and A.S. Alfa, CRC Press, 359–375.

[10] J.E. Diamond and A.S. Alfa (1995). Matrix analytical method for M/PH/1 retrial queues, *Stochastic Models*, 11(3), 447–470.

[11] J.W. Cohen (1957). Basic problems of telephone traffic theory and the influence of repeated calls, *Philips Telecommunication Review*, 18, 49–100.

[12] BD Choi, YC Kim, YW Lee (1998). The M/M/c retrial queue with geometric loss and feedback. Computers & Mathematics with Applications, Vol. 36, No. 6, pp. 41-52

[13] Tuan Phung-Duc, Hiroyuki Masuyama, Shoji Kasahara and Yutaka Takahashi (2009). /M/M/3/3/ AND /M/M/4/4/ RETRIAL QUEUES, JOURNAL OF INDUSTRIAL AND MANAGEMENT OPTIMIZATION Volume 5, Number 3, August 2009 pp. 431451

[14] Economou, A. (2005) Generalized product-form stationary distributions for Markov chains in random environments with queueing applications. Advances in Applied Probability 37, 185-211

[15] Falin. On Sufficient Conditions for Ergodicity of Multichannel Queueing Systems with Repeated Calls. Advances in Applied Probability, Vol. 16, No. 2 (Jun., 1984), pp. 447-448

[16] G.I. Falin (1990). A survey of retrial queues, *Queueing Systems*, 7, 127–167.

[17] G.I. Falin and J.G.C. Templeton (1997). *Retrial Queues*, Chapman & Hall, London.

[18] F Fauvet, F Richard-Jung, J Thomann (2009). Automatic computation of Stokes matrices - Numerical Algorithms, Volume 50, Number 2, 179-213

[19] A.A. Fredericks and G.A. Reisner (1979). Approximations to stochastic service systems, with an application to a retrial model. *The Bell System Technical Journal*, 58, 557–576.

[20] DP Gaver, PA Jacobs, G Latouche (1984). Finite birth-and-death models in randomly changing environments - Advances in Applied Probability, 16, 715-731

[21] A. Gómez-Corral (2006). A bibliographical guide to the analysis of retrial queues through matrix-analytic techniques, *Annals of Operations Research*, 141(1), 163–191.

[22] A. Gómez-Corral and M.F. Ramalhoto (1999). The stationary distribution of a Markovian process arising in the theory of multiserver retrial queueing systems, *Mathematical and Computer Modelling*, 30(3–4), 141–158.

[23] Differential equations and group theory from Riemann to Poincare

[24] Nathaniel Grier, William A. Massey, Tyrone McKoy and Ward Whitt (1997). The time-dependent Erlang loss model with retrials. Telecommunication Systems Volume 7, Numbers 1-3, 253-265.

[25] Hajek, B., Birth-and-death processes on the integers with phases and general boundaries", Journal of Applied Probability, 19 (1982), 488-499.

[26] T. Hanschke, Explicit Formulas for the caracterictics of the M/M/2/2 queue with Repeated Attempts Author(s). Journal of Applied probability, Vol. 24, No. 2 (Juin., 1987), pp. 486-494

[27] T. Hanschke, A matrix continued fraction algorithm for the multiserver repeated order queue Mathematical and Computer Modelling Volume 30, Issues 3-4, August 1999, Pages 159-170

[28] K. Hiroe (2010). Twisted Euler transform of differential equations with an irregular singular point

[29] Kawanishi, K (2005). On the Counting Process for a Class of Markovian Arrival Processes with an Application to a Queueing System. Queueing Systems, 49, 93-122

[30] J. Keilson, J. Cozzolino and H. Young, A service system with unfilled request repeated, Oper. Res. 16, 1126-1137 (1968).

[31] J.G. Kemeny, J.L. Snell and A.W. Knapp (1966). *Denumerable Markov Chains*, 2nd ed., Van Nostrand, Princeton, NJ.

[32] Y.C. Kim (1995). On $M/M/3/3$ retrial queueing system, *Honam Mathematical Journal*, 17, 141–147.

[33] B. Kim and J. Kim (2009). Tail asymptotics for the queue size distribution in a discrete-time $Geo/G/1$ retrial queue, *Queueing Systems*, 61, 243–254.

[34] B. Kim and J. Kim (2009). Queue size distribution in a discrete-time D-BMAP/G/1 retrial queue, *Computers & Operations Research*, in press.

[35] J. Kim, B. Kim and S.-S. Ko (2007). Tail asymptotics for the queue size distribution in an $M/G/1$ retrial queue, *Journal of Appllied Probabability*, 44(4), 1111–1118.

[36] Kosten, L. (1947). On the influence of repeated calls in the theory of probabilities of blocking. De Ingenieur 59.

[37] V.G. Kulkarni and H.M. Liang (1996). Retrial queues revisited, in *Frontiers in queueing: Models and Applications in Science and Engineering*, edited by J.H. Dshalalow, CRC Press, 19–34.

[38] G. Latouche and V. Ramaswami (1995). Expected passage times in homogeneous quasi-birth-and-death processes", Stochastic Models, 11, 103-122

[39] G. Latouche and V. Ramaswami (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM, Philadelphia.

[40] Quan-Lin Li and Jinhua Cao (2006). Two Types of RG-Factorizations of Continuous-Time Level-Dependent Quasi-Birth-and-Death Processes. Quan-Lin Li and Jinhua Cao

[41] B Liu, YQ Zhao (2010). Analyzing retrial queues by censoring. Queueing Systems

[42] A Mandelbaum, WA Massey, MI Reiman, A. Stolyar, B. Rider (2002). Queue lengths and waiting times for multiserver queues with abandonment and retrials. Telecommunication Systems 21, 149-171

[43] CD Meyer (1989). Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems SIAM review,

[44] Neuts, M. F. (1981) Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach John Hopkins University Press, 254-275.

[45] M.F. Neuts and B.M. Rao (1990). Numerical investigation of multi-server retrial model, *Queueing systems*, 7(2), 169–190.

[46] P.R. Parthasarathy and R. Sudhesh (2007). Time-dependent analysis of a single-server retrial queue with state-dependent rates, *Operations Research Letters*, 35(5), 601–611.

[47] On the problem of re-attempted calls in teletraffic C. E. M. Pearce Stochastic Models, 1532-4214, Volume 3, Issue 3, 1987, Pages 393 407

[48] C. E. M. Pearce (1989). Extended Continued Fractions, Recurrence Relations and Two-Dimensional Markov Processes. Advances in Applied Probability, Vol. 21, pp. 357-375

[49] J. Riordan, Stochastic Service Systems (SIAM Series in Applied Mathe- matics). John Wiley and Sons, New York, 1962

[50] Yang Woo Shin and Dug Hee Moon (2010). Approximations of retrial queue with limited number of retrials. Computers & Operations Research Volume 37, Issue 7, July 2010, Pages 1262-1270

[51] S.N. Stepanov (1999). Markov models with retrials: The calculation of stationary performance measures based on the concept of truncation, *Mathematical and Computer Modelling*, 30(3–4), 207–228.

[52] J. Wimp & D. Zeilberger (1985). Resurrecting the asymptotics of linear recurrences. Journal of Mathematical Analysis and Applications Volume 111, Issue 1, October 1985, Pages 162-176

[53] R.W. Wolff (1989). *Stochastic Modeling and the Theory of Queues*, Prentice-Hall.

[54] T. Yang and J.G.C. Templeton (1987). A survey on retrial queues, *Queueing Systems*, 2, 201–233.

[55] Y.Q. Zhao (2000). Censoring technique in studying block-structured Markov chains, in *Advances in Algorithmic Methods for Stochastic Models*, edited by G. Latouche and P. Taylor, Notable Publications Inc., 417–433.