# ВЫЧИСЛИТЕЛЬНО ЭФФЕКТИВНАЯ БИНОМИАЛЬНАЯ МОДЕЛЬ РАСПРЕДЕЛЕНИЯ ДЕФОЛТОВ

# A COMPUTATIONALLY EFFICIENT BINOMIAL MODEL OF DEFAULTS DISTRIBUTION

**Богданов Дмитрий Валериевич**

аспирант кафедры информационных технологий РЭУ им. Г. В. Плеханова. Адрес: ФГБОУ ВПО «Российский экономический университет имени Г. В. Плеханова», 117997, Москва, Стремянный пер., д. 36. E-mail: bogdv@rambler.ru.

**Bogdanov Dmitry Valerievich**

post-graduate of department of information technologies, PRUE. Address: Plekhanov Russian University of Economics, 36 Stremyanny Lane, Moscow, 117997, Russian Federation. E-mail: bogdv@rambler.ru.

В условиях нынешнего финансового кризиса становится особенно актуальным анализ корреляции дефолтов. В статье рассмотрены примеры применения биномиальной модели распределения дефолтов в некоторых частных случаях. Сформулирована и доказана теорема о виде соотношений для решения задачи о нахождении функции распределения вероятностей дефолтов в общем случае. Разработанная методика позволяет получать вектор элементарных вероятностей по известному набору предельных вероятностей дефолта каждой фирмы и предельным парным вероятностям дефолта взаимосвязанных фирм как решение систем линейных и линейно-логарифмических уравнений. Также предложены способы упрощения изучаемой математической модели системы. Полученные результаты могут быть использованы как для управления рисками, так и для кредитных деривативов в целом.

In the current financial crisis it is especially important to analyze possible correlation of defaults within a given branch. The paper deals with examples of how a binomial model of defaults distribution can be used in certain particular cases. We formulate and prove a theorem which gives the form of the probability distribution function of defaults in the general case. The developed methods allow one to obtain the vector of elementary probabilities for the same set of single default probabilities of each firm. We also compute the ultimate pair probabilities default interconnected firms as the solution of systems of linear and linear-logarithmic equations. The paper provides a method for simplifying the study of the mathematical model of the system. The results can be used for both risk management and credit derivative pricing.

Ключевые слова: биномиальная модель, корреляция дефолтов, торическая модель, модель Изинга, диверсификация рисков.

Key words: binomial model, correlation of defaults, toric model, Ising model, risk diversification.

Кредитный риск касается оценки и хеджирования ценных бумаг, имеющих возможность дефолта. Поскольку инвесторы практически всегда используют целый набор финансовых инструментов, связанных с разными фирмами, успешное моделирование взаимосвязи риска дефолта для нескольких фирм имеет решающее значение как для управления рисками, так и для кредитных деривативов в целом. Анализ корреляции дефолтов особенно актуален в условиях финансового кризиса.

Существуют различные подходы к моделированию корреляции дефолтов. В качестве примеров можно привести модели на основе корреляции процессов интенсивности [3] или так называемые «модели цепной реакции» [5]. Все эти модели не лишены недостатков [2]. Альтернативой может служить однопериодическая модель, в основе которой лежит суммирование случайных величин с биномиальным законом распределения [1][4]. Модель корректна в следующем смысле: она может представлять любое заданное предельное распределение вероятностей для отдельных фирм и попарную корреляционную матрицу. Для изучения предложенной модели, доказательства корректности и построения уравнений в явном виде используются методы алгебраической геометрии.

Рассмотрим ненаправленный граф $G = (V, E)$ с $M$ вершинами и определим его множество вершин $V = \{1,\dots,M\}$. Множество рёбер $E$ является подмножеством из $\binom{M}{2}$ возможных попарных связей между всеми парами вершин, то есть $E \subseteq \{(u,v): 1 \le u < v \le M\}$. Каждая вершина графа с номером $i \in V$ связана с соответствующей фирмой и имеет бинарное случайное значение $X_i$, где $X_i = 1$ для фирмы, объявившей дефолт и $X_i = 0$ для работающей фирмы. Совместное распределение вероятностей случайной величины $X = (X_1,\dots,X_M)$ имеет вид

$$p_\omega(\eta) = P(X = \omega) = \frac{1}{Z}\exp\left(\sum_{i\in V}\eta_i\omega_i + \sum_{(u,v)\in E}\eta_{uv}\omega_u\omega_v\right). \qquad (1)$$

Здесь $\omega = (\omega_1,\dots,\omega_M) \in \{0,1\}^M$, скаляры $\eta_i \in \mathbf{R}$ и $\eta_{uv} \in \mathbf{R}$ – параметры, $Z$ – нормировочная константа:

$$Z = \sum_{\omega\in\{0,1\}^M}\exp\left(\sum_{i\in V}\eta_i\omega_i + \sum_{(u,v)\in E}\eta_{uv}\omega_u\omega_v\right).$$

Вероятностные модели вида (1) также известны как марковские модели случайных полей, как модели Изинга в физике или графические модели в информатике и статистике. В контексте финансов было доказано [1], что выбранная модель (1) является корректной для моделирования корреляции дефолтов: для каждого набора предельных вероятностей $P_i$ и парных линейных корреляций дефолтов $\rho_{uv}$ существует уникальный набор параметров $\eta_i$, $\eta_{uv}$, соответствующий этой информации. Данные о предельных вероятностях дефолта и парных линейных корреляций эквивалентны следующим наборам $M + |E|$ достаточной статистики:

- отдельные предельные узлы $P_i = P(X_i = 1)$ для всех $i \in V$;
- двойные предельные узлы $P_{uv} = P(X_u = X_v = 1)$ для всех $(u,v) \in E$.

Элементарные вероятности определяются следующим образом:

$$p_\omega = p_{\omega_1 \cdots \omega_M} = P(X_1 = \omega_1, \ldots, X_M = \omega_M), \qquad \omega = (\omega_1, \ldots, \omega_M) \in \{0,1\}^M .$$

Предельные вероятности дефолта в терминах элементарных вероятностей $p_\omega$ могут быть определены следующим образом:

$$P_i = \sum_{\omega \in \{0,1\}^M} a_{i\omega} p_\omega ,$$

где $a_{i\omega} \in \{0,1\}$. Определим бинарную матрицу $A_G$ с элементами $a_{i\omega}$, содержащую $2^M$ столбцов и $M + |E| + 1$ строк. Столбцы $A_G$ индексируются элементарными вероятностями $p_\omega$. Все строки, кроме последней, индексируются предельными вероятностями $P_i$ для $i \in V$ и $P_{uv}$ для $(u,v) \in E$. Записанные в этих строках коэффициенты используются в разложении предельных вероятностей в терминах $p_\omega$. Последняя строка $A_G$ имеет все элементы, равные единице, что соответствует вычислению тривиальной предельной вероятности $\sum_{\omega \in \{0,1\}^M} p_\omega = 1$

. Рассматриваемая графическая модель может быть представлена в виде торической модели, определяемой соответствующей матрицей $A_G$, где матрица $A_G$ является линейным отображением, преобразующим вектор элементарных вероятностей в вектор предельных вероятностей.

Для соответствия алгебраической литературе заменим параметры модели на их экспоненты, получая таким образом новые положительные параметры:

$$\theta_i = \exp(\eta_i) \text{ для } i \in V \quad \text{и} \quad \theta_{uv} = \exp(\eta_{uv}) \text{ для } (u,v) \in E .$$

Модель параметризации (1) имеет следующий мономиальный вид:

$$p_\omega = \frac{1}{Z} \cdot \prod_{i \in V} \theta_i^{\omega_i} \cdot \prod_{(u,v) \in E} \theta_{uv}^{\omega_u \omega_v} \tag{2}$$

где элементарные вероятности – это мономы, соответствующие столбцам из $A_G$. Последняя строка $A_G$ вносит множитель $\frac{1}{Z}$. Модель является подмногообразием, полученным как пересечение ($2^M$ – 1)-мерного вероятностного симплекса и гиперповерхностей, заданных биномиальными уравнениями вида

$$\prod_{\omega} p_{\omega}^{C_{\omega}} - \prod_{\omega} p_{\omega}^{D_{\omega}} = 0, \qquad (3)$$

где $C$ и $D$ пробегают попарно вектора в $\mathbf{N}^{2^M}$ так, что $A_G \cdot C = A_G \cdot D$.

Пусть имеются наборы векторов $\{\omega^{(a)}\}_{a\in A}$ и $\{\omega^{(b)}\}_{b\in B}$. Сформулируем и докажем необходимые и достаточные условия существования уравнения (3) для произвольного значения $M$.

**Теорема 1.** Биномиальное уравнение гиперповерхности в ($2^M$ – 1)-мерном вероятностном симплексе, описывающее биномиальную модель распределения дефолтов,

$$\prod_{a\in A} p_{\omega^{(a)}} - \prod_{b\in B} p_{\omega^{(b)}} = 0, \qquad (4)$$

должно удовлетворять следующим условиям:

$$\sum_{a\in A} \omega_i^{(a)} = \sum_{b\in B} \omega_i^{(b)}, \qquad (5)$$

$$\sum_{a\in A} \omega_u^{(a)}\omega_v^{(a)} = \sum_{b\in B} \omega_u^{(b)}\omega_v^{(b)}, \qquad (6)$$

для всех $1 \le i \le M$ и $1 \le u < v \le M$.

**Доказательство.** Согласно модели параметризации (2), каждому мультиндексу $\omega_i$ в уравнении (4) соответствует сомножитель вида $\theta_i^{\omega_i}$, а каждой паре мультииндексов ($\omega_u$, $\omega_v$) соответствует сомножитель вида $\theta_{uv}^{\omega_u\omega_v}$ для всех $1 \le i \le M$ и $1 \le u < v \le M$. Учитывая, что не имеется иных сомножителей, равенство выполняется лишь при равных наборах рассмотренных сомножителей. ■

Определим вектор $\xi = (\xi_1, \ldots, \xi_M) \in \mathbf{C}^M$, при этом $\xi^{\omega} = \xi_1^{\omega_1} \cdots \xi_M^{\omega_M}$. Тогда условия (5) и (6) эквивалентны следующим дифференциальным условиям:

$$\frac{\partial}{\partial \xi_i} \sum_{a\in A} \xi^{\omega^{(a)}} \Bigg|_{\xi=(1,\ldots,1)} = \frac{\partial}{\partial \xi_i} \sum_{b\in B} \xi^{\omega^{(b)}} \Bigg|_{\xi=(1,\ldots,1)},$$

$$\frac{\partial^2}{\partial \xi_u \partial \xi_v} \sum_{a\in A} \xi^{\omega^{(a)}} \Bigg|_{\xi=(1,\ldots,1)} = \frac{\partial^2}{\partial \xi_u \partial \xi_v} \sum_{b\in B} \xi^{\omega^{(b)}} \Bigg|_{\xi=(1,\ldots,1)},$$

для всех $1 \le i \le M$ и $1 \le u < v \le M$.

**Замечание 1.** Для задания полилинейного уравнения, содержащего все $2^M$ элементарные вероятности $p_{\omega}$, необходимо и достаточно, чтобы все сомножители $p_{\omega}$ с чётной и нечётной суммой мультииндексов $\sum_{i=1}^{M} \omega_i$ располагались в разных мономах уравнения (4).

Используя данное замечание, на языке Visual C++ автором была разработана программа, позволяющая представить модель как в параметрическом виде (2), так и в терминах элементарных вероятностей (4) для произвольного значения $M$.

**Пример 1.** Пусть $G$ – граф с вершинами $V = \{1, 2, 3\}$ и сторонами $E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Матрица $A_G$ имеет следующий вид:

$$\begin{array}{ccccccccc} & p_{000} & p_{001} & p_{010} & p_{011} & p_{100} & p_{101} & p_{110} & p_{111} \\ P_1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ P_2 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ P_3 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ P_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ P_{13} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ P_{23} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \Sigma & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \tag{7}$$

Модель параметризации (2) приводит к следующему:

$$(p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) =$$
$$=(1, \theta_3, \theta_2, \theta_2\theta_3\theta_{23}, \theta_1, \theta_1\theta_3\theta_{13}, \theta_1\theta_2\theta_{12}, \theta_1\theta_2\theta_3\theta_{12}\theta_{13}\theta_{23}).$$

Модель является гиперповерхностью в семимерном вероятностном симплексе, заданной уравнением (все условия теоремы 1 выполнены):

$$p_{000}\,p_{011}\,p_{101}\,p_{110} = p_{001}\,p_{010}\,p_{100}\,p_{111}. \tag{8}$$

Матрица $A_G$ имеет rank $A_G = 7$. Разрешим линейную подсистему (7) относительно предельных вероятностей $P_i$, $P_{uv}$ и элементарной вероятности дефолта трёх фирм $p_{111}$:

$$p_{000} = 1 - P_1 - P_2 - P_3 + P_{12} + P_{13} + P_{23} - p_{111},$$
$$p_{001} = P_3 - P_{13} - P_{23} + p_{111},\ p_{010} = P_2 - P_{12} - P_{23} + p_{111},\ p_{100} = P_1 - P_{12} - P_{13} + p_{111},$$
$$p_{011} = P_{23} - p_{111},\ p_{101} = P_{13} - p_{111},\ p_{110} = P_{12} - p_{111}.$$

Справедливость полученных соотношений легко проверить с помощью диаграмм Эйлера-Венна. Из соотношения (8) выразим $p_{111}$:

$$p_{111} = \frac{p_{000}p_{011}p_{101}p_{110}}{p_{001}p_{010}p_{100}}.$$

Подставляя найденные соотношения для других элементарных вероятностей, окончательно получим кубическое уравнение относительно $p_{111}$:

$$p_{111} = \frac{(1 - P_1 - P_2 - P_3 + P_{12} + P_{13} + P_{23} - p_{111})(P_{23} - p_{111})(P_{13} - p_{111})(P_{12} - p_{111})}{(P_3 - P_{13} - P_{23} + p_{111})(P_2 - P_{12} - P_{23} + p_{111})(P_1 - P_{12} - P_{13} + p_{111})}. \tag{9}$$

Аналитическое решение данного уравнения может быть получено с помощью известных формул Кардано, однако ввиду его громоздкости в данной работе не приводится.

Вычислим элементарные вероятности $p_\omega$ и построим функцию распределения случайной величины $Y = \{0, 1, 2, 3\}$, равной количеству обанкротившихся фирм, в случае следующего набора исходных данных: $P_1 = 0{,}1$; $P_2 = 0{,}2$; $P_3 = 0{,}3$; $P_{12} = 0{,}05$; $P_{13} = 0{,}07$; $P_{23} = 0{,}12$. Решая уравнение (9), получим следующие значения $p_{111}$:

$$p_{111} = 0{,}0644385 \pm 0{,}0472122i; \quad p_{111} = 0{,}042123.$$

В контексте данной задачи имеет смысл лишь третье решение $p_{111} = 0{,}042123$. Подставляя $p_{111}$ в выражения для других элементарных вероятностей, получим:

$$p_{000} = 0{,}597877;\ p_{001} = 0{,}152123;\ p_{010} = 0{,}072123;\ p_{100} = 0{,}022123;$$

$$p_{011} = 0{,}077877;\ p_{101} = 0{,}027877;\ p_{110} = 0{,}00787703.$$

Условие нормировки выполнено, так как

$$0{,}597877+0{,}152123+0{,}072123+0{,}022123+0{,}077877+0{,}027877+0{,}00787703+0{,}042123 = 1.$$

Вычислим значения функции распределения $f(Y)$:

$$f(0) = p_{000} = 0{,}597877;$$

$$f(1) = p_{001} + p_{010} + p_{100} = 0{,}152123 + 0{,}072123 + 0{,}022123 = 0{,}246369;$$

$$f(2) = p_{011} + p_{101} + p_{110} = 0{,}077877 + 0{,}027877 + 0{,}00787703 = 0{,}113631;$$

$$f(3) = p_{111} = 0{,}042123;$$

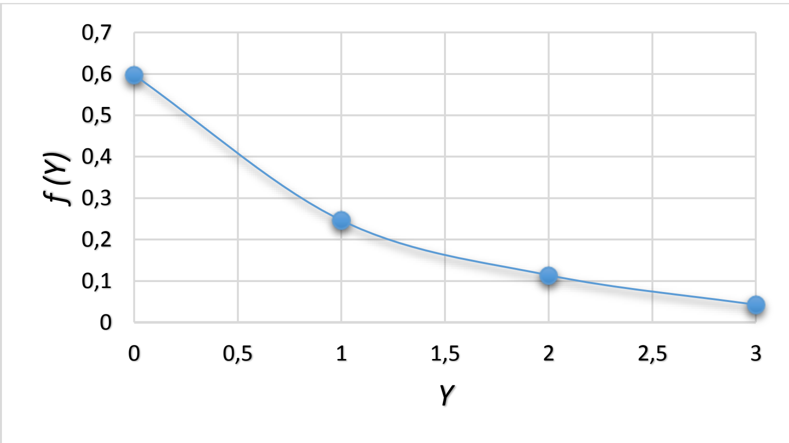


Рис. 1 – Функция распределения $f(Y)$

Математическое ожидание $M[Y]$ и дисперсия $D[Y]$ соответственно равны:

$$M[Y] = 0{,}597877 \cdot 0 + 0{,}246369 \cdot 1 + 0{,}113611 \cdot 2 + 0{,}042123 \cdot 3 = 0{,}6;$$

$$D[Y] = 0{,}597877 \cdot (0 - 0{,}6)^2 + 0{,}246369 \cdot (1 - 0{,}6)^2 +$$

$$+ 0{,}113611 \cdot (2 - 0{,}6)^2 + 0{,}042123 \cdot (3 - 0{,}6)^2 = 1{,}53.\ \blacksquare$$

**Пример 2.** Пусть $G$ – граф (рис. 2) с вершинами $V = \{1, 2, 3, 4\}$ и сторонами $E = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$.

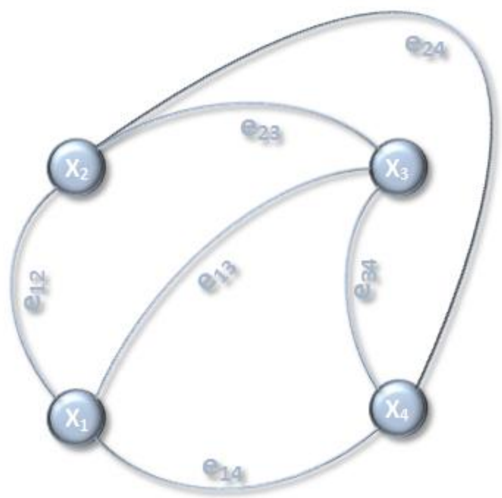


Рис. 2 – Графическая модель системы

Матрица $A_G$ имеет 16 столбцов и 11 строк (из-за громоздкости в явном виде не приводится). Параметризация (2) приводит к гиперповерхности в пятнадцатимерном вероятностном симплексе, заданной уравнением

$$p_{0000}\, p_{0011}\, p_{0101}\, p_{0110}\, p_{1001}\, p_{1010}\, p_{1100}\, p_{1111} = p_{0001}\, p_{0010}\, p_{0100}\, p_{0111}\, p_{1000}\, p_{1011}\, p_{1101}\, p_{1110}.$$

Учитывая, что линейная подсистема, заданная матрицей $A_G$, содержит только 10 линейно-независимых строк, для нахождения шестнадцати неизвестных $p_\omega$ их недостаточно. Легко убедиться, что приведенные ниже уравнения также удовлетворяют условиям теоремы 1:

$$p_{0001}\, p_{0010}\, p_{0100}\, p_{0111} - p_{0011}\, p_{0101}\, p_{0110} = 0;$$

$$p_{0001}^2 p_{0010} p_{0111} p_{1011} p_{1100} - p_{0011}^2 p_{0101} p_{1001} p_{1110} = 0\,.$$

Аналогичным образом можно записать необходимое количество уравнений для определения неизвестных $p_\omega$. При помощи логарифмирования предложенная торическая подсистема приводится к линейной системе относительно $\ln p_\omega$, задаваемой матрицей

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1k} \\ t_{21} & t_{22} & \cdots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{12} & \cdots & t_{mk} \end{pmatrix}, \qquad (10)$$

где каждый столбец соответствует уравнению вида

$$\prod_{i:t_{ij}>0} p_{\omega_i}^{t_{ij}} - \prod_{i:t_{ij}<0} p_{\omega_i}^{-t_{ij}} = 0, \qquad j = 1, \ldots, k.$$

Предложенные выше уравнения логарифмируются следующим образом:

$$\ln p_{0001} + \ln p_{0010} + \ln p_{0100} + \ln p_{0111} - \ln p_{0011} - \ln p_{0101} - \ln p_{0110} = 0;$$

$$2\ln p_{0001} + \ln p_{0010} + \ln p_{0111} + \ln p_{1011} + \ln p_{1100} - 2\ln p_{0011} - \ln p_{0101} - \ln p_{1001} - \ln p_{1110} = 0\,.$$

Соответствующая матрица $T$ имеет шестнадцать строк и два столбца (из-за громоздкости в явном виде не приводится). Решение подсистем, задаваемых матрицами $A_G$ и $T$, может

Как наглядно показывают вышеприведенные примеры, вычисление вектора элементарных вероятностей $p_\omega$ по известным наборам предельных вероятностей $P_i$ и $P_{uv}$ сопряжено со значительными вычислительными трудностями. Поэтому имеет смысл рассматривать не каждую фирму в отдельности, а лишь взаимосвязанные секторы экономики. Также необходимо отметить, что на практике количество ребёр графа $G$ значительно меньше $\binom{M}{2}$, то есть граф $G$ не является полным. При моделировании корреляции дефолтов могут вычисляться не элементарные вероятности $p_\omega$, а некоторые их линейные комбинации. Например, с помощью предложенной методики можно определить вероятность дефолта не менее половины всех фирм. При разработке специального математического аппарата для решения систем вида (10) и соответствующего программного обеспечения моделирование корреляции дефолтов на основе биномиальной модели может оказаться полезным инструментом для решения широкого круга практических задач математической экономики.

**Список литературы**


1. Graphical models for correlated defaults / I. O. Filiz, X. Guo, J. Morton, B. Sturmfels // Mathematical Finance. – 2012 – Vol. 22, №4. – P. 621 – 644.
2. Bielecki T., Credit risk Frontiers: Subprime Crisis, Pricing and Hedging, CVA, MBS, Ratings, and Liquidity. / T. Bielecki, D. Brigo, F. Patras, – Hardcover, 2011. – 754 p.
3. Common failings: how corporate defaults are correlated / S. R. Das, D. Duffie, N. Kapadia, L. Saita // The Journal of Finance. – 2007 – Vol. 62, №2. – P. 93 – 117.
4. Evans, S. N. Nonexistence of Markovian time dynamics for graphical models of correlated default. / S. N. Evans, A. Hening. // Queueing Systems. – 2011 – Vol. 69, №3-4. – P. 293-312.
5. Gu, J.-W. A Markovian infectious model for dependent default risk / J.-W. Gu, W.-K. Ching, T.-K. Siu. // International Journal of Intelligent Engineering Informatics – 2011 – Vol. 1, №2 – P. 174-195.


**References**


1. I. O. Filiz, X. Guo, J. Morton, B. Sturmfels, *Graphical models for correlated defaults*, Mathematical Finance, 22:4 (2012), 621-644.

2. T. Bielecki, D. Brigo, F. Patras, *Credit risk Frontiers: Subprime Crisis, Pricing and Hedging, CVA, MBS, Ratings, and Liquidity*, Hardcover, 2011, 754 p.

3. S. R. Das, D. Duffie, N. Kapadia, L. Saita, *Common failings: how corporate defaults are correlated*, The Journal of Finance, 62:2 (2007), 93-117.

4. S. N. Evans, A. Hening, *Nonexistence of Markovian time dynamics for graphical models of correlated default*, Queueing Systems, 69:3-4 (2011), 293-312.

5. J.-W. Gu, W.-K. Ching, T.-K. Siu, *A Markovian infectious model for dependent default risk*, International Journal of Intelligent Engineering Informatics, 1:2 (2011), 174-195.