May 2018

Fitting a Complex Markov Chain Model for Firm and Market Productivity

Julia Ruth Valder
University of Wisconsin-Milwaukee

Fitting a Complex Markov Chain Model for Firm and Market Productivity

by

Julia R. Valder

A Thesis Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Master of Science
in
Mathematics

at

The University of Wisconsin-Milwaukee
May 2018

ABSTRACT

Fitting a Complex Markov Chain Model for Firm and Market Productivity

by

Julia R. Valder

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Richard H. Stockbridge

This thesis develops a methodology of estimating parameters for a complex Markov chain model for firm productivity. The model consists of two Markov chains, one describing firm-level productivity and the other modeling the productivity of the whole market. If applicable, the model can be used to help with optimal decision making problems for labor demand. The need for such a model is motivated and the economical background of this research is shown. A brief introduction to the concept of Markov chains and their application in this context is given. The simulated data that is being used for the estimation is presented in detail. The underlying economical problem is described as a stochastic process. Available data for a single firm is limited, therefore a 2-step method is used to estimate the probability matrix for the firm Markov chain. Under a time homogeneity assumption, maximum likelihood estimation techniques are used to estimate the parameters of a Markov chain for one firm based on all firms in the market. These parameters are refined using a linear combination approach. The expectation and variance of the proposed estimator are analyzed. The method's validity is established using various goodness-of-fit tests. Theoretical explorations for the estimation of a market Markov chain are made. In the end, a summary of results and an outlook for further research directions is given.

Thank You.

To Professor Richard Stockbridge for giving me the opportunity to work on this thesis and for advising me throughout the last year. I highly valued his help and advice as he supported me in succeeding academically, especially as an international student at UWM.

To Dr. Atanas Christev from the Heriot-Watt University in Edinburgh for his trust in my abilities to work with the economical model he created. I could always count on his explanations of economical background while writing this thesis and for that I am truly grateful.

To Professor Gerhard Dikta from FH Aachen for bringing the dual degree program with UWM to my attention and advising me while taking part in it. The program highly relies on him and I appreciate how his constant guidance made the transfer process more transparent.

To Professor Chao Zhu and to Professor Gabriella Pinter for agreeing to be on my thesis committee. Their questions and comments helped in refining this thesis and I was able to submit it confidently thanks to their involvement.

Julia R. Valder

Table of Contents

List of Figures

List of Tables

Introduction and Motivation

This chapter motivates the thesis. The given economical model and a suggested mathematical model are described. The thesis goal is stated and distinction from previous work is established.

I.1 Optimal Decision Making Problems

In economics, a central question is: Is there an "optimal" action policy concerning hiring and firing of workers? This question is an example of the problem of optimal decision making under uncertainty. Firms face adjustment costs for hiring and firing and must therefore be careful not to act precipitously. Having to correct a mistake means increased costs. Firms can also be "paralyzed" by uncertainty and might lose productivity by not making necessary adjustments (see Rota (2004), Cooper et al. (2007), Elsby and Michaels (2014)). A model to predict productivity based on labor demand in a single firm or a market would be highly useful in order to avoid unnecessary adjustments or to make sure that adjustments happen in time. If such a model is given, it should explain real life data and be in accordance with empirical facts.

I.2 Labor Productivity Model

The proposed model considers workers and firms under non-convex adjustments costs in the presence of firm-level (also referred to as *idiosyncratic*) and market-level (also referred to as

aggregate) shocks. These shocks affect labor productivity and change the desired number of employees in order to achieve maximum profit. Therefore productivity should be maximized while costs are minimized. The non-convex adjustment costs create an environment where firms do not adjust continuously but keep their employment level constant for some intervals of time. Jumps in this level are observed occasionally.

At the beginning of a time period, firms observe their productivity level (Z) and the aggregate productivity (A). If, based on Z and A, the firm decides to adjusts its workforce (L), it faces adjustment costs denoted by τ. The firm's productivity is then given by the production function $y = AZH(L)$, where H is an increasing and concave function of workforce. We also consider a wage rate w. For the purpose of this thesis, it is an exogenously given constant. In the future, w could also be modeled by a Markov chain given wage negotiations can happen.

Using the aforementioned quantities, the total discounted profit of a firm during a period of time is given by the following equation:

$$\pi(L_{-1}, Z, A) = \sup_{L}\{AZH(L) - wL - \tau I_{\{L \neq L_{-1}\}} - \tau^{+}(L) I_{\{L > L_{-1}\}} - \tau^{-}(L) I_{\{L < L_{-1}\}} + \beta\pi_{+1}\}$$

where L_{-1} denotes the employment level in the previous time period, I is the indicator function that takes the value 1 if its argument is true and 0 otherwise. τ is the fixed adjustment cost, τ^{+} is the cost of hiring, τ^{-} is the cost of firing and $\beta\pi_{+1}$ is the expected profit of the next period, discounted for inflation, which is represented by β. This model is a version or extension of earlier models, such as in Cooper et al. (2007) and Elsby and Michaels (2014).

In order to maximize the profit over the employment level L, the productivity is maximized while all the costs are minimized. In this thesis, the focus lies on the productivity function $y = AZH(L)$. The profit can only be maximized using this equation if the parameters of the underlying processes are known.

I.3 Using Stochastic Processes

The described profit model works with discrete time periods, in each of which a realization of the production function can be observed; y can also be formulated as a stochastic process dependent on the time t. The time periods are individually labeled by an index set $T = \{0, 1, 2, ...\}$ so that we can refer to each time period individually. From here on, the production function will be considered a stochastic process on these time periods, given by

$$Y(t) = A(t)Z(t)H(L(t)), t \in T$$

The process $L(t)$ describes the level of employment that the firm can adjust at the beginning of each time period. It is therefore viewed as a control variable that only depends on past and present information, and doesn't anticipate future developments. The processes $A(t)$ and $Z(t)$ capture the aggregate and idiosyncratic productivity. It is proposed in Khan and Thomas (2004) that A and Z can be viewed as independent Markov chains that evolve randomly over time and that the process $P(t) = A(t)Z(t)$ is also a Markov Chain.

I.4 Thesis Goal

This thesis looks to create a method of estimating the time-homogeneous parameters of the Markov chains $A(t)$ and $Z(t)$. Appropriate state spaces for both Markov chains are proposed. Maximum likelihood estimation techniques are used on simulated data examples in order to find the transition probability matrices of A and Z. These matrices are refined using linear combination methods. All calculations are done with the a statistical software package R. The goal of this thesis is to showcase the methodology of estimating the transition probabilities and using goodness of fit tests to validate the estimation. In doing so, this thesis sets the stage for further research that may use real life data to fully validate this model and estimate its parameters, and use it to solve optimal decision making problems in economics.

I.5 Distinction and Limits

This thesis does not solve any optimal decision making problems, but it gives a solid basis for further research in this area. Unlike works which develop theoretical approaches to optimal decision problems, this thesis proposes a parameter estimation technique for the given model that can be replicated when using real data. However, the method described in this thesis cannot give a guaranteed proof of the model's validity. Rather, it creates highly suggestive data-based evidence that the estimated parameters are correct using established stochastic methods, given that the model is applicable. While it was originally planned to use real life data for this fitting, such data is highly proprietary and could not be obtained at this time. Therefore the method is illustrated using simulated data.

Mathematical Background

This chapter gives a brief introduction to Markov chains, maximum likelihood estimation and goodness of fit tests. These concepts are later applied to simulated data and show the method of estimating model parameters.

II.1 Markov Chains

As described in Karlin and Taylor (1975), a Markov process is a stochastic process where future behavior, given that the present state is known, does not depend on additional knowledge of the past. They were first studied by Andrey Markov in the early 20th century (see Gagniuc (2017)).

For the purpose of this thesis, the following notation from Karlin and Taylor (1975) is used. Markov chains are denoted as $X(t)$ with t being a value from the discrete time space $\{1, 2, ..., T\}$. The state space of a Markov chain is S, meaning $X(t)$ can attain values from the set S. $X(t)$ refers to the outcome of trial t. A Markov chain has "the property that given the value of $X(t)$, the values of $X(s), s > t$ do not depend on the values of $X(u), u < t$". When $X(t) = i$ it means that $X(t)$ is in state $i \in S$. The conditional probability of $X(t+1)$ being in state j given that $X(t)$ is in state i is denoted by $P_{ij}(t) = P(X(t+1) = j|X(t) = i)$. This is called a one-step-transition probability. If the transition probability does not depend on t, it is denoted by P_{ij} and is called time-homogeneous or stationary. The Markov chains in this thesis are discrete time, stationary and have a countable infinite or finite state space $S = \{1, 2, ..., N\}$.

For such Markov chains, the probabilities P_{ij} can be arranged in a matrix, as shown in Definition 1.1 which is given in Karlin and Taylor (1975).

Definition 1.1. *If P_{ij} denotes the transition probability of a Markov chain from state i to state j, the probabilities are arranged in the so-called transition probability matrix $P = (P_{ij})_{i,j=1,...,N}$ (also referred to as* transition matrix*), so that*

$$P = \begin{pmatrix} P_{11} & P_{12} & \ldots\ldots & P_{1N} \\ P_{21} & P_{22} & \ldots\ldots & P_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N1} & P_{N2} & \ldots\ldots & P_{NN} \end{pmatrix},$$

The i^{th} row of P is the conditional probability distribution of $X(t+1)$ given that $X(t) = i$. The probabilities P_{ij} satisfy

$$P_{ij} \geq 0, \qquad i, j = 0, 1, 2, ..., N$$

$$\sum_{j=1}^{N} P_{ij} = 1, \qquad i = 0, 1, 2, ..., N$$

In this thesis, there are multiple firms $k = \{1, 2, ..., F\}$ all moving according to their own Markov chain $Z^k(t)$ with distinct transition matrices P^k.

II.2 Maximum Likelihood Estimation

Maximum likelihood estimation can be used to estimate the parameters of a distribution. If data generated by a distribution with unknown parameters is given, the maximum likelihood method analyses under which parameters the given data sample is the most likely realization. Maximum likelihood techniques were first introduced by Ronald Fisher in the 1910s and later popularized by Samuel S. Wilks between 1938 and 1962 (see Pfanzagl and Hamboeker (1994) and Wilks (1962)). The following formal definitions of the likelihood function and a

maximum likelihood estimator are given in Casella and Berger (2001).

Definition 2.1. *Let $f(x|\tilde{\theta})$ denote the joint probability distribution function (pdf) or probability mass function (pmf) of the sample $\tilde{X} = (X_1, ..., X_n)$ with the parameter $\tilde{\theta} = (\theta_1, ..., \theta_k)$. Then, given that $\tilde{X} = \tilde{x}$ is observed, the function of $\tilde{\theta}$ defined by*

$$L(\tilde{\theta}|\tilde{x}) = \prod_{i=1}^{n} f(x_i|\tilde{\theta})$$

is called the likelihood function. If $\tilde{X}$ is a discrete random vector, then

$$L(\tilde{\theta}|\tilde{x}) = P_{\tilde{\theta}}(\tilde{X} = \tilde{x}).$$

Definition 2.2. *For each sample point $\tilde{x}$, let $\hat{\theta}(\tilde{x})$ be a parameter value at which $L(\tilde{\theta}|\tilde{x})$ attains its maximum as a function of $\tilde{\theta}$, with $\tilde{x}$ held fixed. A maximum likelihood estimator (MLE) of the parameter $\tilde{\theta}$ based on a sample $\tilde{X}$ is $\hat{\theta}(\tilde{X})$.*

In the context of Markov chains, the parameters $(\theta_1, ..., \theta_k)$ are the elements of the transition matrix P. The maximum likelihood estimator of P is the transition matrix, under which the observed data is the most likely realization of the Markov chain process. The following maximum likelihood estimation of the transition probabilities in a Markov chain is given in Bishop (1975). First, the notation is defined.

Definition 2.3. *Suppose the Markov chain has N possible states and we observe $T-1$ successive transitions (the first going from time 1 to time 2). Suppose there are $n_i(1)$ individuals in state i at time 1 and that the $\{n_i(1)\}_{i=1,...,N}$ are multinomially distributed with probabilities ν_i and sample size $F = \sum_{i=1}^{N} n_i(1)$. Finally, let $c_{ij}(t)$ be the number of individuals that were in state i at time $t-1$ and are in state j at time t and let $n_i(t)$ be the number of individuals in state i at time t.*

Then, the likelihood function and the maximum likelihood estimators are derived.

Theorem 2.4. *Under the assumption of time-homogeneity, the likelihood functions of the probabilities* $\{\nu_i\}_{1\leq i\leq N}$ *and the* $\{P_{ij}\}_{1\leq i,j\leq N}$ *are*

$$L(\nu_1, ..., \nu_N \mid n_1(1), ..., n_N(1)) = \prod_{i=1}^{N} \nu_i^{n_i(1)}, \text{ and}$$

$$L(P \mid \{c_{ij}\}_{i,j=1,...,N}) = \prod_{i=1}^{N}\prod_{j=1}^{N} P_{ij}^{c_{ij}},$$

where $c_{ij} = \sum_{t=1}^{T-1} c_{ij}(t)$.

Theorem 2.5. *By maximizing the likelihood functions in Theorem 2.4 as functions of* ν_i *and* P_{ij}*, the following maximum likelihood estimates of the time-homogeneous transition probabilities are obtained.*

$$\hat{\nu_i} = \frac{n_i(1)}{F}$$

$$\hat{P}_{ij} = \frac{c_{ij}}{\sum_{t=1}^{T-1} n_i(t-1)}$$

Later, Theorems 2.4 and 2.5 are used for the first step of the estimation of the idiosyncratic Markov chains.

II.3 Goodness of Fit Tests

After estimating the parameters of a Markov chain, goodness-of-fit tests can be used to evaluate whether the estimated parameters make for a good fit of the model to the given data. Specifically, these tests evaluate how likely it is that an observed data sample arose by chance from the given distribution with the estimated parameters. Many different kinds of goodness-of-fit tests can be found in the literature; each test has its typical application. In this thesis, three kinds of goodness-of-fit tests are used, all applicable to different situations that arise during the estimation process. A goodness-of-fit test is a hypothesis test, which is described as follows in Casella and Berger (2001).

Definition 3.1. *The two complementary hypotheses in a hypothesis test about a parameter θ are called the null hypothesis H_0 and the alternative hypothesis H_1. Typically, $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1 = \Theta_0^C$. A hypothesis test is a rule that specifies*

1. *For which sample values the decision is made to not reject H_0,*

2. *For which sample values H_0 is rejected and H_1 is accepted as true.*

The subset of the sample space for which H_0 will be rejected is called rejection region. Typically, a hypothesis test is specified in terms of a test statistic, which is a function of the given sample.

In goodness-of-fit tests the null hypothesis $H_0 : \theta = \hat{\theta}$ is tested against the alternative $H_1 : \theta \neq \hat{\theta}$ are being tested. In general, goodness-of-fit test statistics are a measure of deviation. They calculate how much the predicted values differ from the observed values in a sample. If the parameter estimation is good, the predicted values should be close to the observed data. Goodness-of-fit test statistics have a known distribution or a known limiting distribution that can be used to evaluate whether the null hypothesis, claiming that the estimated parameters are the correct parameters, should be rejected or not (see Bishop (1975)). A limiting distribution is an approximation of the real distribution that is valid for large sample sizes.

II.3.1 Pearson-Chi-Square Test

The Pearson-Chi-Square test was first investigated by Karl Pearson (1900). It is one of many chi-square tests, which all have the feature that their test statistic has a limiting χ^2 distribution. As a goodness-of-fit test, it is defined in the following way:

Definition 3.2. *In a Pearson-Chi-Square goodness-of-fit test, the null hypothesis $H_0 : \tilde{\theta} = \tilde{\theta_0}$ is tested versus $H_1 : \tilde{\theta} \neq \tilde{\theta_0}$ as the alternative. $\tilde{\theta} = (\theta_1, ..., \theta_N)$ are the probabilities that an object falls into category $i = 1, ..., N$. The test statistic is given by*

$$\chi^2(\tilde{x}) = \sum_{i=1}^{N} \frac{(x_i - E_i)^2}{E_i}$$

where x_i is the number of times an object of category i is observed and E_i is the expected number of objects of type i. The expected number of objects is given by the null hypothesis that the probability of an object being in category i is θ_i. The test statistic has a limiting Chi-Square distribution with N degrees of freedom, where $n = N - p$ and p is the number of independent parameters used. The Pearson-Chi-Square test rejects the null hypothesis for all $\tilde{x} \in \{\tilde{x} : \chi^2(\tilde{x}) \geq \chi^2_{1-\alpha;n}\}$, in which $\chi^2_{1-\alpha;n}$ is the $1-\alpha$ quantile of a chi-square distribution with n degrees of freedom.

The Pearson-Chi-Square test is only applicable if the expected values are large enough, meaning the sample size is large. Otherwise, the assumption of a limiting χ^2 distribution is not reasonable.

II.3.2 Exact Multinomial Test

As just explained, the Pearson-Chi-Square test is only applicable for large sample sizes. Also, the limiting distribution is only known to be a χ^2-distribution when the maximum likelihood estimator for $\tilde{\theta}$ is used. When examining the idiosyncratic Markov chain for a single firm, both of these conditions are likely to be violated since data for a single firm is limited and a combined estimator is used (see Section III.3).

However, for a small number of states and small sample sizes an exact multinomial test can be used. Unlike Pearson-Chi-Square which only has a limiting distribution, this test calculates the exact probability for the observed data to occur under the estimated model parameters. An exact multinomial test is applicable when the underlying distribution of a sample is multinomial. This is the case for each row of a transition probability matrix that is time homogeneous. The multinomial distribution is defined as follows by Read and Cressie (1988).

Definition 3.3. *Consider observing random variables $Y_i, i = 1, ..., n$ which can have one of N possible outcomes $\{s_1, s_2, ..., s_N\}$ with probabilities $P_{i1}, P_{i2}, ..., P_{iN}$. The outcomes are mutually exclusive and $\sum_{j=1}^{N} P_{ij} = 1$. Let $\tilde{C} = (C_1, C_2, ..., C_N)$ where C_k is the number of Y_i's where $Y_i = s_k$, $k = 1, ..., N$. If the observed Y's are independent and identically distributed, then $\tilde{C}$ has a multinomial distribution with parameters n and $\tilde{P} = \{P_{i1}, ..., P_{iN}\}$. The probability of obtaining any particular sample $\tilde{c} = (c_1, ..., c_N)$ is then*

$$P(\tilde{C} = \tilde{c}) = N! \prod_{j=1}^{N} \frac{P_{ij}^{c_j}}{c_j!}.$$

The exact multinomial test calculates the probability of obtaining a particular multinomial sample or any sample with a smaller probability (a more extreme sample) under the null hypothesis that $\tilde{P} = \tilde{P}_0$. The null hypothesis is rejected if this probability is smaller than a defined significance level α, typically $\alpha = 0.05$ or $\alpha = 0.1$.

II.3.3 Likelihood Ratio Test and G-Test

While the exact multinomial test can only be feasibly calculated for a small number of states, a likelihood ratio test can be applied for small sample sizes and a large number of states. Thus, it closes the gap between Pearson-Chi-Square and the exact multinomial test. The likelihood ratio test is related to maximum likelihood estimation (see II.2). To test an estimation for the parameter θ is defined by Casella and Berger (2001) as follows.

Definition 3.4. *The likelihood ration test statistic for testing the null hypothesis $H_0 : \theta \in \Theta_0$ versus the alternative $H_1 : \theta \in \Theta_1$ is*

$$\lambda(\tilde{x}) = \frac{\sup_{\Theta_0} L(\theta|\tilde{x})}{\sup_{\Theta} L(\theta|\tilde{x})},$$

where $\Theta = \Theta_0 \cup \Theta_1$. A likelihood ratio test (LRT) is any test that has a rejection region of the form $\{\tilde{x} : \lambda(\tilde{x}) \leq c\}$, where $0 \leq c \leq 1$.

As explained in Casella and Berger (2001), the ratio $\lambda(\tilde{x})$ is small if there are parameter points in the alternative parameter space for which the given data sample is much more likely than for any parameter points in the null hypothesis. Therefore, when the ratio is small, H_0 should be rejected. Like in many hypothesis tests, c is selected so that the probability of a so-called false negative, rejecting H_0 when it is actually true, is small. Formally speaking, c is chosen such that $\sup_{\theta \in \Theta_0} P_\theta(\lambda(\tilde{X}) \leq c) = \alpha$ where α is a given constant, typically $\alpha = 0.1$ or $\alpha = 0.05$, where the latter is the slightly stricter criterion.

The G-Test is derived from the likelihood ratio test when the underlying model is multinomial. Each row of the transition matrix of a Markov chain represents a multinomial distribution; therefore the G-Test is applicable. Both the G-Test and the likelihood-ratio test can be used to test each row of the transition matrix individually when there is a large number of states. In accordance with Bishop (1975), the G-Test is defined as follows:

Definition 3.5. *The test statistic G of a G-Test is given by*

$$G = 2 \sum_{i=1}^{N} x_i \ln\left(\frac{x_i}{E_i}\right)$$

where x_i is the number of times an object of type i was observed and E_i was the expected number of objects of type i. The distribution of G is approximately a Chi-Square distribution with $N - p$ degrees of freedom, where p is the reduction in degrees of freedom as given in Definition 3.2.

A G-Test is any test that has a rejection region of the form $\{\tilde{x} : G \geq c\}$, where $0 \leq c < \infty$. When the null hypothesis is $H_0 : \theta = \theta_0$, the constant c is chosen such that $P_{\theta_0}(G(\tilde{X}) \geq c) = \alpha$, the probability of a false-negative.

The constant p is the reduction in degrees of freedom, which is the number of independent parameters used in fitting the distribution.

Model Fitting

III.1 Stochastic Process Formulation

As described in Chapter I, the productivity of a firm is given by the process Y with

$$Y(t) = A(t)Z(t)H(L(t)), \quad t = 1, 2, ..., T,$$

This thesis focuses on modeling the processes $A(t)$ and $Z(t)$. The process $A(t)$ captures the aggregate productivity in a certain market. For the purpose of this thesis, the following definition is used for the aggregate productivity process.

Definition 1.1. *Let there be firms $k \in \{1, 2, ..., F\}$ in a certain market and $S^k(t)$ be the total sales of firm number k in the time period t. Then, the process $\hat{A}(t)$ describes the aggregate productivity of a market as*

$$\hat{A}(t) = \frac{\sum_{k=1}^{F} S^k(t)}{\sum_{k=1}^{F} S^k(t-1)}.$$

As a consequence of this definition, $\hat{A}(t)$ is a dimensionless number which roughly describes the general growth trend of the market as a whole.

To turn the real-valued process $\hat{A}(t)$ into a Markov chain, its values are discretized into distinct intervals. The discretized version of $\hat{A}(t)$ is the aggregate Markov chain $A(t)$. Since a finite state space is needed for the estimation, the state space S_A of $A(t)$ is defined as a

finite number of intervals.

$$S_A = \{[a_i, b_i) \mid i = 1, 2, ..., N_A, 0 \leq a_1 < b_1 \leq a_2 < b_2 \leq ... < b_{N_A}\}$$

In a concrete application of the estimation method, the state intervals for $A(t)$ can be flexibly defined depending on the available data. It is expected that the values of $A(t)$ range around 1, since it is a relative growth factor.

The process $Z(t)$ gives a firm-specific measure of productivity. Each firm k in a certain market has its own firm-specific process $Z^k(t)$. In this thesis, $Z^k(t)$ is assumed to be a Markov chain (referred to as an idiosyncratic Markov chain), and is defined in the following way.

Definition 1.2. *For each firm $k \in \{1, 2, ..., F\}$ in a certain market, the process $\hat{Z}^k(t)$ describes the firm-specific productivity as*

$$\hat{Z}^k(t) = \frac{S^k(t)}{L^k(t)}.$$

where $S^k(t)$ is the total sales of firm k in the time period t and $L^k(t)$ is the number of employees for the firm k in the time period t.
To turn this real-values process into the idiosyncratic Markov chain $Z^k(t)$, its values are discretized into intervals. For all k, $Z^k(t)$ takes values from the same state space S_Z. Like the aggregate state space in Definition 1.1, the state space S_Z consists of a finite number of intervals.

$$S_Z = \{[a_i, b_i)] \mid i = 1, 2, ..., N_Z, 0 \leq a_1 < b_1 \leq a_2 < b_2 \leq ... < b_{N_Z}\}$$

Just like for $A(t)$, the intervals can be defined to fit the available data. The methodology of estimation is not dependent on the concrete intervals. It is assumed that both $A(t)$ and

all $Z^k(t)$ processes evolve randomly according to a Markov chain. That means that $A(t)$ has a transition probability matrix P_A and each $Z^k(t)$ process has its own transition probability matrix P_Z^k. In this thesis, the exponent of a matrix always represents the number of the firm to which it belongs. Matrices are never raised to a power during any calculations or examples.

The employment level $L(t)$ is a stochastic process, but it is not assumed to be a Markov chain. The firm determines its employment level at the beginning of each time period t, depending on multiple factors, for example the current firm productivity. It is a control variable and must only depend on the past and the present, not the future.

III.2 Simulation of Firm Productivity Data

The original intention of this thesis was to fit the model with real life data. However, real firm productivity data turned out to be proprietary and inaccessible at the time of publication of this thesis. Therefore, data is simulated in order to illustrate the methodology that has been developed. To simulate data, the statistical software package R is used. The full R source code for the simulation can be found in Appendix A.

III.2.1 Simulating Idiosyncratic Markov Chains

The idiosyncratic productivity, which is measured by sales per employee as stated in Definition 1.2, is simulated. The states of the process are given by intervals of productivity which are ordered from low to high productivity. In order to simplify the notation the intervals from Definition 1.2 are numbered from 1 to N and the state space is denoted as $S_Z = \{1, 2, ..., N_Z\}$. The distance between two states is defined as follows:

Definition 2.1. *Given a state space $S_Z = \{1, 2, 3, ..., N_Z\}$, the distance $d(i, j)$ between two states $i, j \in S$ is given by:*

$$d(i, j) = |i - j| + 1$$

The firms themselves are the individuals moving throughout the Markov chain. Some sensible assumptions are made about the firms and their productivity for the purpose of simulation.

Hypothesis 2.2. *The transition probabilities of the idiosyncratic productivity Markov chains are time-homogeneous. Therefore, all transition matrices are independent of the time t. This assumption is given by the economical model described in Chapter I.*

Hypothesis 2.3. *Large jumps in productivity in either direction are unlikely. It is much more likely that a firm will keep up their productivity or only slightly increase or decrease it. In terms of the Markov chains used to model productivity, this means that transitions to nearby states should be more likely than transitions to states that are more distant.*

Hypothesis 2.4. *All firms in the same market behave similarly but not quite the same. Hence, the models will be chosen in such a way that each firm's transition matrix is a slight perturbation of a "base matrix". This base matrix is constructed in accordance with Hypothesis 2.3.*

Hypothesis 2.5. *All firms follow a market-wide trend. The higher this trend is, the more likely firms are to increase their productivity, therefore transition to higher states are more likely.*

Hypothesis 2.6. *A firm has its own firm trend, which can be positive or negative. This firm trend is more likely to be positive the higher the market trend is. If the firm trend is positive, the firm's transition probabilities to higher states are slightly increased compared to the base matrix. If it is negative, the firm's transition probabilities to higher states are slightly decreased compared to the base matrix.*

The program first generates the base matrix that is later used to create each firm's individual transition matrix, according to Hypothesis 2.4. Definition 2.7 states how the base matrix is generated.

Definition 2.7. *Let $S_Z = \{1, 2, ..., N_Z\}$ be a state space of a Markov chain. The base matrix B is defined as follows:*

$$B = \left(\frac{b_{ij}}{\gamma_i}\right)_{1 \leq i,j \leq N_Z}$$

$$\textit{in which } b_{ij} = \frac{1}{2d(i,j)}$$

$$\textit{and } \gamma_i = \sum_{j=1}^{N} b_{ij}.$$

γ_i is a normalizing constant which ensures that the probabilities in each row sum up to 1. Otherwise B would not be a transition matrix. This definition can be adapted to create a higher concentration around same-state transitions, which are transitions from a state back to the same state. Figure III.1 shows the magnitudes of the b_{ij}'s using Definition 2.7, and Figure III.2 shows the magnitudes using a possible alternative definition $b_{ij} = \frac{1}{2^{d(i,j)}}$. In both cases $i = 10$ is used as an example. This thesis uses the distribution from Definition 2.7 for the simulation because otherwise some transitions might never occur. Karlin and Taylor (1975) refer to this phenomenon as "sample zeros", which means a transition is never observed although it has a probability that is larger than 0. To avoid sample zeros and to better illustrate the methodology on examples with few firms, a less concentrated distribution is desirable. Real life data would comprise samples from many enough firms, such that sample zeros are highly unlikely.

While the number of states and firms used in the simulation in general is adjustable, the following example is used in order to illustrate the program's method.

Example 2.8. *Suppose there are the firms $\{1, 2, ..., F\}$ where $F = 10$, T transitions, and there are $N = 4$ states the firms can be in. Using the source code in Appendix A, the following*

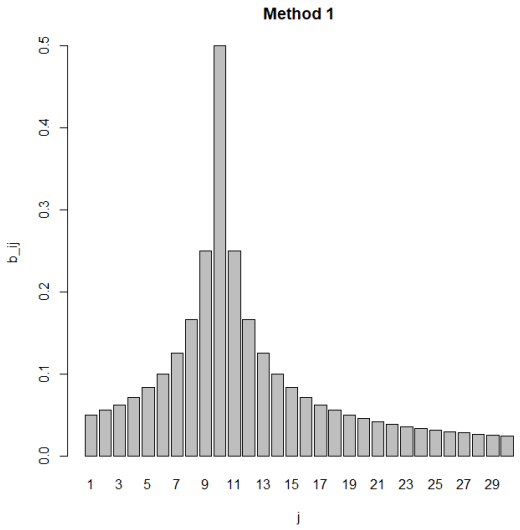

Figure III.1: Magnitude of base matrix elements using method 1

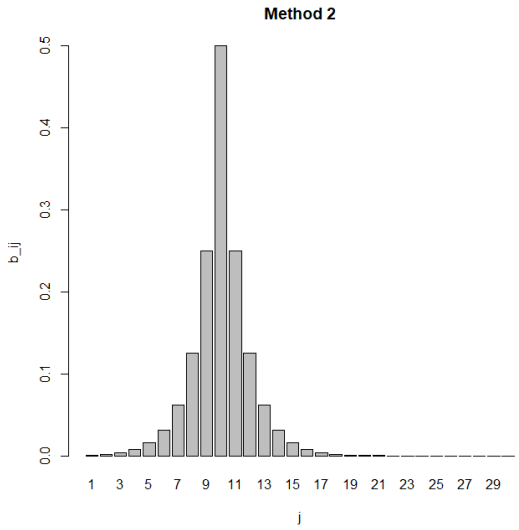

Figure III.2: Magnitude of base matrix elements using method 2

base matrix is generated for a 4-state Markov chain:

$$B_4 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.4800 & 0.2400 & 0.1600 & 0.1200 \\ 0.2143 & 0.4286 & 0.2143 & 0.1429 \\ 0.1429 & 0.2143 & 0.4286 & 0.2143 \\ 0.1200 & 0.1600 & 0.2400 & 0.4800 \end{pmatrix} \end{array}$$

As intended by Hypothesis 2.3, staying in the same state is always the most likely transition. The larger the distance between two states is, the less likely the transition becomes. The asymmetry of the base matrix is caused by using the normalization factor $\frac{1}{\gamma_i}$. This factor is different for different i's since the distance function is not linear.

After computing the base matrix, the program creates a random perturbation for each firm, which corresponds to Hypothesis 2.4. In order to also honor Hypothesis 2.5 and Hypothesis 2.6, a market trend and firm trends are created.

Definition 2.9. *Suppose there are F firms in the market. The market trend m is a realiza-*

tion of a uniform random variable $M \sim \mathcal{U}(0,1)$. For each firm $k = 1, 2, ..., F$, the firm trend f_k is also a realization of a random uniform random variable $F_k \sim \mathcal{U}(0,1)$. All firm trends and the market trend are independent.

Due to this definition, the higher the market trend is the more likely it is that the firm trend is smaller than the market trend. If the firm trend is smaller than the market trend, the program varies the base matrix so that transitions to higher states are more likely than transitions to smaller states. Thus, if the market trend is high, increases in productivity are more likely.

The perturbation of each probability is random and normally distributed. The firm trend dictates whether the mean of the normal distribution used is positive or negative.

Definition 2.10. *For each firm k, the time-homogeneous transition matrix P^k is defined as follows.*

$$P^k = \left(\frac{p_{ij}}{\beta_i}\right)_{1 \leq i,j \leq N},$$

$$\text{in which } p_{ij} = \begin{cases} b_{ij}x_{ij} + b_{ij} & , \text{ when } i \leq j \\ -b_{ij}x_{ij} + b_{ij} & , \text{ when } i > j \end{cases},$$

$$\text{and } \beta_i = \sum_{j=1}^{N} p_{ij}$$

where x_{ij} is the realization of a random variable $X_{ij} \sim \mathcal{N}(\mu sign(m - f_k), 2\mu), \mu \sim \mathcal{U}(0, 1/8)$ for $i \neq j$ and $X_{ij} \sim \mathcal{N}(\mu, 2\mu), \mu \sim \mathcal{U}(0, 1/8)$ for $i = j$.

The last part of the definition slightly boosts the transition probability to stay in the same state. This ensures that Hypothesis 2.3 is honored. The x_{ij}'s represent the random perturbations. Multiplying them with the base matrix probabilities makes them a relative perturbation rather than an absolute perturbation which could create too much variation from the base matrix.

Example 2.11. *Continuing Example 2.8, realizations of the transition matrices of the first*

two firms are given by:

$$P^1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.5244 & 0.2330 & 0.1408 & 0.1018 \\ 0.2030 & 0.4509 & 0.2133 & 0.1328 \\ 0.1448 & 0.2160 & 0.4317 & 0.2076 \\ 0.1018 & 0.1621 & 0.2187 & 0.5173 \end{pmatrix} \end{array}$$

$$P^2 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.5050 & 0.2328 & 0.1477 & 0.1144 \\ 0.2115 & 0.4569 & 0.1978 & 0.1339 \\ 0.1454 & 0.2129 & 0.4261 & 0.2156 \\ 0.1306 & 0.1594 & 0.2458 & 0.4642 \end{pmatrix} \end{array}$$

While they are similar in their properties, there are slight differences in the individual probabilities caused by the described randomization. In the simulation of the data, each firm uses its own transition matrix instead of all firms using the same matrix. This gives more credibility to the method of estimation shown in this thesis, since it better mimics the expected behavior of real life data.

Contingency tables contain information about how many firms have moved from one state to another in a given time period. To generate contingency tables, the program generates $T-1$ transitions for each firm according to its transition matrix. In each step, a uniform random number between 0 and 1 is generated for each firm. The distribution of this random number is given by the row in the transition matrix indexed by the current state of the firm. The resulting number presents the new state of the firm. Using the same conditions as before, this process is illustrated in Example 2.12.

Example 2.12. *Suppose at time* $t-1$*, firm* 1 *is in state* 3*. Firm* 1 *uses the transition matrix*

P^1 *as shown above.*

$$P^1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.5244 & 0.2330 & 0.1408 & 0.1018 \\ 0.2030 & 0.4509 & 0.2133 & 0.1328 \\ 0.1448 & 0.2160 & 0.4317 & 0.2076 \\ 0.1018 & 0.1621 & 0.2187 & 0.5173 \end{pmatrix} \end{array}$$

For the transition, a random number between 1 and 4 is generated according to the distribution given by row 3*. This determines the new state* j *for time* t*.*

$$P^1_{31} = 0.1448, P^1_{32} = 0.2160, P^1_{33} = 0.4317, P^1_{34} = 0.2076$$

Once the transition has been determined, it is recorded in the contingency table for this time step and the firm is placed in the new state.

Example 2.13. *This example continues Examples 2.8 and 2.12. At time* $t = 0$*, the firms* $\{1, 2, .., 10\}$ *are in the following states:*

$$\begin{array}{c} states \\ firms \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1, 2 & 3 & 4, 5 & 6, 7, 8, 9, 10 \end{pmatrix} \end{array}$$

The following contingency table is recorded for the transition from $t = 0$ *to* $t = 1$*.*

$$T_1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 3 & 1 \end{pmatrix} \end{array}$$

This means that, for example, 2 firms have moved from state 1 to state 1 and 3 firms

have moved from state 4 to state 3. After the transition occurs, the firms are in the following states at time $t = 1$.

$$\begin{array}{c} states \\ firms \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1,2,3 & 5,9 & 6,7,10 & 4,8 \end{pmatrix} \end{array}$$

This way, the program generates $T-1$ contingency tables. It also records each individual firm's state at every time step. This concludes the data simulation for the idiosyncratic Markov chains. The data is later used to show how the parameters of the idiosyncratic Markov chains can be estimated.

III.2.2 Simulating Employment Levels

While this thesis only gives a theoretical approach to estimating $A(t)$, simulated data might be needed for future methodology testing. In this case, the employment levels, that are often included in real life data, would be needed for every firm. This section shows how employment levels can be simulated as a discrete stochastic process $L(t)$. The following reasonable assumptions are made about $L(t)$.

Hypothesis 2.14. *The employment level for each firm* k *is decided at the beginning of a time period* t *and depends on the firms productivity from the last time period, given by* $Z^k(t-1)$.

Hypothesis 2.15. *The higher the productivity in the last time period was, the more employees the firms will hire for this time period. The lower the productivity was the more employees will be laid off. If the productivity is average, the net employment change should be near 0.*

While these assumptions have an effect on the way $L(t)$ is simulated, they do not affect the methodology of the parameter estimation in Section III.4. Therefore, the method remains valid when used with real life data.

The simulation program has all the information from the simulation of the idiosyncratic data. In every time step, it uses the productivity of the firm in the previous time step to generate the net employment change, which is defined as follows.

Definition 2.16. *Let $L^k(t)$ be the employment level of firm k at time $t \in \{1, ..., T\}$. Let M be the average productivity across all states. When the states are equidistant, $M = (N+1)/2$ can be used. Let $D(s) = s - M$ be the distance of a state s from the average productivity M. Then the employment level for the firm k for the time period $t+1$ is defined as follows.*

$$\textit{For } t = 1\textit{: } L^k(1) = \textit{round}(|E^k|),$$

$$E^k \sim \mathcal{N}(1000, \sqrt{1000}),$$

$$L^k(t+1) = \begin{cases} L^k(t) + |\textit{round}(X^k)|, & \textit{when } D(Z^k(t)) > 0 \\ L^k(t) - |\textit{round}(X^k)|, & \textit{when } D(Z^k(t)) < 0 \\ L^k(t) + \textit{round}(X^k), & \textit{when } D(Z^k(t)) = 0 \end{cases},$$

$$X^k \sim \mathcal{N}\Big(D(Z^k(t))\frac{L^k(t)}{50}, 2\Big)$$

where round is the function that rounds its input to an integer number.

The factor $\frac{L^k(t)}{50}$ makes sure the mean of the change is relative to the previous employment levels. This avoids huge, unrealistic jumps in employment. It can be adjusted as needed.

III.3 Two-Step Estimation of the Idiosyncratic Markov Chain

This section develops a two-step method of estimating the transition matrix for the idiosyncratic Markov chains introduced in Definition 1.2. Real life data for firm productivity is often only available for a limited number of time periods. When looking at the contingency table for a single firm, there is only one entry in each table that is equal to 1 and all other

entries are equal to 0, since there is only one firm moving through the Markov chain. A naive approach would be to use maximum likelihood estimation. Under the time homogeneity assumption given by the model, the $(T-1)$ contingency tables can be reduced to a single contingency table by element-wise addition of all of the contingency tables. But even then, the number of entries is small and thus a simple maximum likelihood estimation does not yield a reasonable estimation. This problem is illustrated in Example 3.1.

Example 3.1. *For $N = 4$ states, the data simulation from section III.2 creates the following transition matrix for firm number 1.*

$$P^1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.5086 & 0.2222 & 0.1540 & 0.1152 \\ 0.2189 & 0.4258 & 0.2110 & 0.1444 \\ 0.1335 & 0.2051 & 0.4617 & 0.1996 \\ 0.1082 & 0.1463 & 0.2157 & 0.5298 \end{pmatrix} \end{array}$$

After simulating the idiosyncratic Markov chains for $T = 12$ time periods, this is the reduced contingency table C for firm number 1.

$$C^1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 3 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix} \end{array} \tag{3.2}$$

The simple maximum likelihood estimation shown in Theorem II.2.5 yields the following

estimated transition matrix $\hat{P}^1_{MLE}$ *for firm 1.*

$$\hat{P}^1_{MLE} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0 & 0.5 & 0 \\ 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0 & 0.67 \end{pmatrix} \end{array}$$

Clearly, the estimation is imprecise because of the small sample size for each state. Instead of using this naive maximum likelihood approach for a single firm, a new method is developed which is mainly based on Hypothesis 2.4: All firms in a single market behave in an inherently similar way with slight individual variations. Therefore, the hypothesized base matrix will be estimated using a maximum likelihood approach and for an individual firm, the transition matrix will be refined by increasing the probability of those transitions that have actually occurred. The method is presented in detail in the following two sections.

III.3.1 Initial Maximum Likelihood Estimation

It is proposed that all firms in a single market follow a similar base behavior. We can use the maximum likelihood approach from Theorem II.2.5 to obtain an estimate for a transition matrix under the assumptions that all firms followed the same Markov chain. While this assumption is generally not regarded as true, the resulting estimate can be used to create a better estimate in Section III.3.2. Continuing Example 3.1 the following base matrix is estimated using the R source code in Appendix B.

Example 3.3. *For* $F = 120$ *firms, the following realization of the reduced contingency table is obtained.*

$$C = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 142 & 58 & 51 & 40 \\ 62 & 117 & 84 & 45 \\ 59 & 77 & 187 & 88 \\ 28 & 37 & 81 & 164 \end{pmatrix} \end{array}$$

Using maximum likelihood estimation results in the following estimate for the base matrix.

$$\hat{B}_4 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.4879 & 0.1993 & 0.1753 & 0.1375 \\ 0.2013 & 0.3799 & 0.2727 & 0.1461 \\ 0.1436 & 0.1873 & 0.4550 & 0.2141 \\ 0.0903 & 0.1194 & 0.2613 & 0.5290 \end{pmatrix} \end{array}$$

A Pearson-Chi-Square test can be used to assess the goodness of fit for this estimation for all firms. Since the sample size is large, this test is applicable. The categories for the test are the states in each time period except the first, so there are a total of $N(T-1)$ categories. Since we are looking at the number of firms in each state and not the number of transitions from one state to another, the number of parameters is N since we have N rows of the transition matrix. However, if the number of firms in the first $N-1$ states are known, the number of firms in the last state is determined since the total has to sum up to the number of firms F. Therefore, the limiting Chi-Square distribution has $N(T-1)-(N-1) = (N-1)(T-1)+1$ degrees of freedom.

A first Pearson-Chi-Square goodness-of-fit test for Example 3.3 results in a test statistic $\chi^2 = 30.70$. The probability of obtaining a test statistic at least this large is approximately 53.2%, which means the fit is fairly good but can be improved. This coincides with the way the data was simulated, since all firms actually followed their own Markov chains, which

were similar but slightly different.

To test the fit of this estimation for a single firm in this example, an exact multinomial test can be used since the sample size and the number of states are both small. If the number of states is large, a likelihood ratio test can be used instead. The following example shows the result of the exact multinomial test when fitting the base matrix estimation to the contingency table of firm number 1 from Example 3.1.

Example 3.4. *The reduced contingency table for firm number 1 is given in Equation 3.2 in Example 3.1. An exact multinomial test is used to test the fit of the estimated base matrix to firm number 1 specifically. This results in one p-value for each row of the estimated matrix. Let $\hat{B}_i$ denote the i^{th} row of the estimated transition matrix and let c_i be the i^{th} row of the observed reduced contingency table for firm number 1. In this example, the following p-values are obtained:*

$$P_{\hat{B_1}}(c_1) = 0.0947, P_{\hat{B_2}}(c_2) = 0.3846, P_{\hat{B_3}}(c_3) = 0.1436, P_{\hat{B_4}}(c_4) = 0.3250.$$

This means that, for example, the probability of the first row of the contingency table to look like this or be more extreme, when the true distribution is $\hat{B}_1$, is 9.47%. The null hypothesis, that $\hat{B}_1$ is the true distribution of the first row would therefore be rejected to a significance level of $\alpha = 0.1$. The other rows have a probability that is higher than the significance level but still have a seemingly bad fit with low p-values. The estimation for any firm can be improved by using the method shown in Section III.3.2.

III.3.2 Linear Combination Refinement

The naive estimation in Example 3.1 was imprecise given the true transition matrix but the multinomial test yields that it is a perfect fit for the single firm contingency table, meaning that the probability of obtaining this or a more extreme contingency table under the

maximum likelihood estimation for a single firm is 1, because by definition of the maximum likelihood estimator the given contingency table is the most likely sample. The maximum likelihood estimation based on all firms in Example 3.3 is much closer to the actual transition matrix but its fit was not as good. A logical conclusion from this can be that a good estimate lies between the maximum likelihood estimate based on the single firm and the maximum likelihood estimate based on all firms. Therefore, the following estimator is proposed to compromise between a realistic estimation and a good fit.

Definition 3.5. *Let $\hat{B}$ be the maximum likelihood estimate transition matrix based on all firms. Let $\hat{P}^k_{MLE}$ be the maximum likelihood estimate matrix based on the single firm k. When estimating the transition probability P^k_{ij} for a single firm Markov chain $Z^k(t)$, define the estimator $\hat{P}^k_{ij}$ to be*

$$\hat{P}^k_{ij} = (\hat{B}_{ij} + \beta \hat{P}^k_{MLE;ij}) \frac{1}{\gamma_i}$$

where β is any real number that satisfies $\beta \geq 0$. All the transition probabilities can be arranged in a matrix to create an estimator for the transition matrix P^k as a whole, such as

$$\hat{P}^k = (\hat{P}^k_{ij})_{i,j=1,\ldots,N}$$

The factor γ_i is a normalizing factor to ensure that all the probabilities in a row sum up to 1. It is given by

$$\gamma_i = \sum_{l=1}^{N} \left(\hat{B}_{il} + \beta \hat{P}^k_{MLE;il} \right) = 1 + \beta.$$

The intuition behind this theorem is to start with the base matrix estimate and then to slowly increase the probabilities of the transitions that have actually been observed while simultaneously lowering the probabilities of the transitions that have not been observed. Goodness-of-fit tests can be used to evaluate the fit of the estimate for different values of β. Continuing the example in Section III.3, the following example illustrates how different values of β affect the goodness of fit.

Example 3.6. *Consider a sequence $\beta_n = 0.01n$ for $n = 1, ..., 20$. Table III.1 shows the results of multinomial goodness-of-fit tests, which are the probabilities of obtaining this or a more extreme sample under $\hat{P}^1$ given $\beta = \beta_n$. The following observations on this table can be made.*

- *The average fit is always increasing but sometimes experiences larger jumps.*
- *The fit of row 1 rises above the significance level of $\alpha = 0.1$ when $\beta = 0.02$.*
- *The fit of row 3 has a significant jump when $\beta = 0.05$*
- *Row 4 shows a large increase in goodness of fit when $\beta = 0.06$.*
- *Row 2 has a large jump in probability when $\beta = 0.07$.*
- *Row 1 increases its fit by a large amount when $\beta = 0.14$.*

Since $\hat{P}^k_{MLE}$ is the perfect fit, larger values of β will always increase the goodness of fit. Therefore additional criteria must be used to keep β within reasonable bounds. The following proposition gives some sensible ways of choosing β.

Remark 3.7. *When choosing the factor β for the estimator $\hat{P}^k$, one of the following methods can be considered.*

1. *Choose β such that all p-values for the exact multinomial test are larger than a self-defined satisfactory level.*
2. *Choose the largest possible β for which Hypothesis 2.3 still holds, meaning that transitions to nearby states are more likely than transitions to states that are far away.*
3. *Choose the smallest β such that every row has experienced a significant jump in goodness of fit.*
4. *Choose β to minimize the variance of $\hat{P}^k$ (see Section III.3.3).*

β	Row 1	Row 2	Row 3	Row 4	Average
0.01	0.0961	0.4981	0.152	0.3289	0.2688
0.02	0.1094	0.5005	0.1603	0.3328	0.2758
0.03	0.1229	0.503	0.1685	0.3366	0.2828
0.04	0.1244	0.5055	0.1765	0.3403	0.2867
0.05	0.126	0.508	0.3628	0.344	0.3352
0.06	0.1277	0.5105	0.3688	0.437	0.361
0.07	0.1294	0.639	0.3747	0.4391	0.3955
0.08	0.1312	0.6391	0.5787	0.4411	0.4475
0.09	0.133	0.6393	0.5826	0.6322	0.4968
0.10	0.1348	0.6396	0.5864	0.6322	0.4982
0.11	0.1366	0.6399	0.5901	0.6322	0.4997
0.12	0.1385	0.6402	0.5938	0.6323	0.5012
0.13	0.1405	0.6406	0.5974	0.6323	0.5027
0.14	0.2207	0.641	0.6009	0.6323	0.5237
0.15	0.2424	0.6415	0.6044	0.6324	0.5302
0.16	0.2637	0.642	0.6078	0.6325	0.5365
0.17	0.286	0.6426	0.6111	0.6325	0.5431
0.18	0.3089	0.6431	0.6144	0.6326	0.5498
0.19	0.3328	0.6437	0.6177	0.6327	0.5567
0.20	0.3335	0.6444	0.6208	0.6328	0.5579

Table III.1: Multinomial sample probabilities for different values of β

III.3.3 Expectation and Variance

When evaluating the estimator in Section III.3.2, it is of interest to know its expectation and variance. This is especially important for checking unbiasedness and in regard of method 4 in Remark 3.7.

Theorem 3.8. *Consider the estimator $\hat{P}^k$ for the transition matrix of single firm k's Markov chain. The estimator is given in Definition 3.5. The estimator for a single matrix element is $\hat{P}^k_{ij}$. Its expectation and variance are given by*

$$\mathbb{E}\big[\hat{P}^k_{ij}\big] = \frac{B_{ij} + \beta P^k_{ij}}{1+\beta},$$

$$Var\big[\hat{P}^k_{ij}\big] = \frac{1}{(1+\beta)^2} Var\Big[\hat{B}_{ij}\Big] + \frac{\beta^2}{(1+\beta)^2} Var\Big[\hat{P}^k_{MLE;ij}\Big] + \frac{2\beta}{(1+\beta)^2} Cov\Big[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\Big]$$

where P^k_{ij} is the true transition probability from state i to state j for firm k. B_{ij} is the true transition probability from state i to state j in a theoretical Markov chain that all firms follow. While this Markov chain does not actually exist, it can be seen as the average Markov chain of all firms, so B_{ij} can be interpreted as the average probability of going from state i to state j.

Proof.

$$\begin{aligned}
\mathbb{E}\big[\hat{P}^k_{ij}\big] &= \mathbb{E}\left[\frac{1}{1+\beta}(\hat{B}_{ij} + \beta \hat{P}^k_{MLE;ij})\right] \\
&= \frac{1}{1+\beta}\mathbb{E}\Big[\hat{B}_{ij} + \beta \hat{P}^k_{MLE;ij}\Big] \\
&= \frac{1}{1+\beta}\Big(\mathbb{E}\Big[\hat{B}_{ij}\Big] + \beta \mathbb{E}\Big[\hat{P}^k_{MLE;ij}\Big]\Big) \\
&= \frac{B_{ij} + \beta P^k_{ij}}{1+\beta}.
\end{aligned}$$

The last step can be done because maximum likelihood estimators are inherently unbiased.

$$
\begin{aligned}
\mathrm{Var}\big[\hat{P}^k_{ij}\big] &= \mathrm{Var}\left[\frac{1}{1+\beta}(\hat{B}_{ij} + \beta \hat{P}^k_{MLE;ij})\right] \\
&= \frac{1}{(1+\beta)^2}\mathrm{Var}\Big[\hat{B}_{ij} + \beta \hat{P}^k_{MLE;ij}\Big] \\
&= \frac{1}{(1+\beta)^2}\left(\mathrm{Var}\Big[\hat{B}_{ij}\Big] + \mathrm{Var}\Big[\beta \hat{P}^k_{MLE;ij}\Big] + 2\mathrm{Cov}\Big[\hat{B}_{ij}, \beta \hat{P}^k_{MLE;ij}\Big]\right) \\
&= \frac{1}{(1+\beta)^2}\mathrm{Var}\Big[\hat{B}_{ij}\Big] + \frac{\beta^2}{(1+\beta)^2}\mathrm{Var}\Big[\hat{P}^k_{MLE;ij}\Big] + \frac{2\beta}{(1+\beta)^2}\mathrm{Cov}\Big[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\Big]
\end{aligned}
$$

□

For the estimator to be unbiased, the expectation would have to be equal to the true transition probability P^k_{ij}. However this only happens when $\hat{P}^k_{ij} = B_{ij}$, meaning that P^k_{ij} would have to be exactly equal to the average probability of going from state i to state j. Given this criterion, unbiasedness cannot be achieved by choosing a certain β rather it has to be inherent in the given data. However, even though the estimator is not unbiased, it is better than the maximum likelihood estimator given the small number of entries in a contingency table for a single firm (see Example 3.1).

Since unbiasedness cannot be generally achieved, the following theorem takes a closer look at the result of Theorem 3.8 for the variance in order to test the applicability of method 4 in Remark 3.7.

Theorem 3.9. *Let S_i be the set of all firms that have ever been in state i over all time periods (firms can be present in S_i multiple times) and let n_i be the size of S_i. Let n^k_i be the total number of times firm number k has been in state i. Then:*

$$
Var\Big[\hat{B}_{ij}\Big] = \left(\frac{1}{n_i}\right)^2 \sum_{l \in S_i} P^l_{ij}(1 - P^l_{ij}),
$$

$$
Var\Big[\hat{P}^k_{MLE;ij}\Big] = \frac{1}{n^k_i} P^k_{ij}(1 - P^k_{ij}), \ \textit{and}
$$

$$Cov\Big[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\Big] = \frac{1}{n_i} P^k_{ij}(1 - P^k_{ij}) = \frac{n^k_i}{n_i} Var\Big[\hat{P}^k_{MLE;ij}\Big]$$

Proof. Let c_{ij} be the total number of times a firm has ever moved from state i to state j. Let c^k_{ij} be the number of times firm number k has moved from state i to state j. Let $\mathbb{I}_{\{k,j,a\}}$ denote the indicator of firm k moving to state j at the a^{th} time it is in state i and let $\mathbb{I}_{\{k,j\}}$ denote the indicator of firm k moving from state i to state j at a non-specified time. Then the variance of $\hat{B}_{ij}$ is determined as follows.

$$\begin{aligned}
\mathrm{Var}\Big[\hat{B}_{ij}\Big] &= \mathrm{Var}\left[\frac{c_{ij}}{n_i}\right] \\
&= \left(\frac{1}{n_i}\right)^2 \mathrm{Var}\big[c_{ij}\big] \\
&= \left(\frac{1}{n_i}\right)^2 \mathrm{Var}\left[\sum_{l \in S_i} \mathbb{I}_{\{l,j\}}\right] \\
&= \left(\frac{1}{n_i}\right)^2 \sum_{l \in S_i} \mathrm{Var}\Big[\mathbb{I}_{\{l,j\}}\Big] \\
&= \left(\frac{1}{n_i}\right)^2 \sum_{l \in S_i} \mathrm{Var}\Big[\mathbb{I}_{\{l,j\}}\Big] \\
&= \left(\frac{1}{n_i}\right)^2 \sum_{l \in S_i} P^l_{ij}(1 - P^l_{ij})
\end{aligned}$$

Next, the evaluation of $\mathrm{Var}\Big[\hat{P}^k_{MLE;ij}\Big]$ is obtained.

$$
\begin{aligned}
\mathrm{Var}\Big[\hat{P}^k_{MLE;ij}\Big] &= \mathrm{Var}\left[\frac{c^k_{ij}}{n^k_i}\right] \\
&= \left(\frac{1}{n^k_i}\right)^2 \mathrm{Var}\big[c^k_{ij}\big] \\
&= \left(\frac{1}{n^k_i}\right)^2 \mathrm{Var}\left[\sum_{a=1}^{n^k_i} \mathbb{I}_{\{k,j,a\}}\right] \\
&= \left(\frac{1}{n^k_i}\right)^2 \sum_{a=1}^{n^k_i} \mathrm{Var}\left[\mathbb{I}_{\{k,j,a\}}\right] \\
&= \left(\frac{1}{n^k_i}\right)^2 \sum_{a=1}^{n^k_i} P^k_{ij}(1-P^k_{ij}) \\
&= \left(\frac{1}{n^k_i}\right)^2 n^k_i P^k_{ij}(1-P^k_{ij}) \\
&= \left(\frac{1}{n^k_i}\right) P^k_{ij}(1-P^k_{ij})
\end{aligned}
$$

Finally, observe, that c^l_{ij} and c^k_{ij} are independent when $k \neq l$. Then the covariance is evaluated as follows.

$$
\begin{aligned}
\mathrm{Cov}\Big[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\Big] &= \mathbb{E}\Big[\hat{B}_{ij}\hat{P}^k_{MLE;ij}\Big] - \mathbb{E}\Big[\hat{B}_{ij}\Big]\mathbb{E}\Big[\hat{P}^k_{MLE;ij}\Big] \\
&= \mathbb{E}\left[\frac{c_{ij}c^k_{ij}}{n_i n^k_i}\right] - B_{ij}P^k_{ij} \\
&= \mathbb{E}\left[\sum_{l=1}^{F}\Big(\frac{c^l_{ij}c^k_{ij}}{n_i n^k_i}\Big)\right] - B_{ij}P^k_{ij} \\
&= \sum_{l=1}^{F}\mathbb{E}\left[\frac{c^l_{ij}c^k_{ij}}{n_i n^k_i}\right] - B_{ij}P^k_{ij}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{l=1;l\neq k}^{F} \left(\mathbb{E}\Big[\frac{c_{ij}^{l}}{n_i}\Big]\mathbb{E}\Big[\frac{c_{ij}^{k}}{n_i^k}\Big] \right) + \mathbb{E}\Big[\frac{(c_{ij}^{k})^2}{n_i n_i^k}\Big] - B_{ij}P_{ij}^{k} \\
&= \sum_{l=1}^{F} \left(\mathbb{E}\Big[\frac{c_{ij}^{l}}{n_i}\Big]\mathbb{E}\Big[\frac{c_{ij}^{k}}{n_i^k}\Big] \right) - \mathbb{E}\Big[\frac{c_{ij}^{k}}{n_i}\Big]\mathbb{E}\Big[\frac{c_{ij}^{k}}{n_i^k}\Big] + \frac{1}{n_i n_i^k}\mathbb{E}\Big[(c_{ij}^{k})^2\Big] - B_{ij}P_{ij}^{k} \\
&= B_{ij}P_{ij}^{k} - \mathbb{E}\Big[\frac{c_{ij}^{k}}{n_i}\Big]P_{ij}^{k} + \frac{1}{n_i n_i^k}\mathbb{E}\Big[(c_{ij}^{k})^2\Big] - B_{ij}P_{ij}^{k} \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\mathbb{E}\Big[(c_{ij}^{k})^2\Big] \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\mathbb{E}\Big[\sum_{a=1}^{n_i^k}\mathbb{I}_{\{k,j,a\}}\sum_{b=1}^{n_i^k}\mathbb{I}_{\{k,j,b\}}\Big] \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\mathbb{E}\Big[\sum_{a=1}^{n_i^k}\Big(\mathbb{I}_{\{k,j,a\}}\sum_{b=1}^{n_i^k}\mathbb{I}_{\{k,j,b\}}\Big)\Big] \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\sum_{a=1}^{n_i^k}\Big(\mathbb{E}\Big[\mathbb{I}_{\{k,j,a\}}\sum_{b=1}^{n_i^k}\mathbb{I}_{\{k,j,b\}}\Big]\Big) \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\sum_{a=1}^{n_i^k}\Big(\mathbb{E}\Big[\mathbb{I}_{\{k,j,a\}}\sum_{b=1;b\neq a}^{n_i^k}\mathbb{I}_{\{k,j,b\}} + \mathbb{I}_{\{k,j,a\}}\Big]\Big) \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\sum_{a=1}^{n_i^k}\Big(\mathbb{E}\Big[\mathbb{I}_{\{k,j,a\}}\Big]\mathbb{E}\Big[\sum_{b=1;b\neq a}^{n_i^k}\mathbb{I}_{\{k,j,b\}}\Big] + \mathbb{E}\Big[\mathbb{I}_{\{k,j,a\}}\Big]\Big) \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}\sum_{a=1}^{n_i^k}\Big(P_{ij}^{k}(n_i^k - 1)P_{ij}^{k} + P_{ij}^{k}\Big) \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i n_i^k}n_i^k P_{ij}^{k}\Big(n_i^k P_{ij}^{k} - P_{ij}^{k} + 1\Big) \\
&= -\frac{n_i^k P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{n_i^k P_{ij}}{n_i}P_{ij}^{k} - \frac{P_{ij}^{k}}{n_i}P_{ij}^{k} + \frac{1}{n_i}P_{ij}^{k} \\
&= \frac{1}{n_i}P_{ij}^{k}(1 - P_{ij}^{k}).
\end{aligned}
$$

□

When comparing the two variances, it can be easily seen that $\mathrm{Var}\big[\hat{B}_{ij}\big] < \mathrm{Var}\big[\hat{P}_{MLE;ij}^{k}\big]$ if and only if

$$
\frac{n_i^k}{n_i} < \frac{P_{ij}^{k}(1 - P_{ij}^{k})}{\frac{1}{n_i}\sum_{l\in S_i}P_{ij}^{l}(1 - P_{ij}^{l})}
$$

meaning that the variance of the base matrix estimation is smaller as long as its sample size is sufficiently large. The expression $\frac{1}{n_i}\sum_{l\in S_i} P_{ij}^l(1-P_{ij}^l)$ can be interpreted as the average variance of c_{ij}^l across all firms l in state i. Therefore, the right side of the expression will usually be near 1 while the left side decreases when increasing the number of firms. This means, $\mathrm{Var}\big[\hat{B}_{ij}\big]$ is usually smaller than $\mathrm{Var}\big[\hat{P}^k_{MLE;ij}\big]$. Also, the covariance of the two estimators is less than $\mathrm{Var}\big[\hat{P}^k_{MLE;ij}\big]$, since $n_i^k \leq n_i$. The condition of the covariance being less than $\mathrm{Var}\big[\hat{B}_{ij}\big]$ is

$$\begin{aligned}
\mathrm{Cov}\Big[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\Big] &< \mathrm{Var}\big[\hat{B}_{ij}\big] \\
\frac{n_i^k}{n_i}\mathrm{Var}\Big[\hat{P}^k_{MLE;ij}\Big] &< \mathrm{Var}\big[\hat{B}_{ij}\big] \\
1 &< \frac{\mathrm{Var}\big[\hat{B}_{ij}\big]}{\frac{n_i^k}{n_i}\mathrm{Var}\Big[\hat{P}^k_{MLE;ij}\Big]} \\
1 &< \frac{\left(\frac{1}{n_i}\right)^2 \sum_{l\in S_i} P_{ij}^l(1-P_{ij}^l)}{\frac{1}{n_i}P_{ij}^k(1-P_{ij}^k)} \\
1 &< \frac{\frac{1}{n_i}\sum_{l\in S_i} P_{ij}^l(1-P_{ij}^l)}{P_{ij}^k(1-P_{ij}^k)} \\
P_{ij}^k(1-P_{ij}^k) &< \frac{1}{n_i}\sum_{l\in S_i} P_{ij}^l(1-P_{ij}^l).
\end{aligned}$$

This means that the covariance is less than the variance of $\hat{B}_{ij}$ if and only if the variance of and indicator for firm k going from i to j is less than an expression that can be interpreted as the average variance of such indicators for all firms.

When looking to minimize the variance of the estimator $\hat{P}^k$ as a function of β, the minimizing β will therefore most likely be small since $\frac{1}{(1+\beta)^2}$ is decreasing in β and $\frac{\beta^2}{(1+\beta)^2}$ is increasing in β and converges to 1. However, because of the covariance term, a minimizing

$\beta \geq 0$ cannot always be found. The following example shows the structure of the variance as a function of β.

Example 3.10. *Let* $Var\left[\hat{B}_{ij}\right] = 1$, $Var\left[\hat{P}^k_{MLE;ij}\right] = 2$ *and* $Cov\left[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\right] = 0.5$. *Figure III.3 shows the plot of* $Var\left[\hat{P}^k_{ij}\right]$ *as a function of* β. *Similar plots can be found for different example values of* $Var\left[\hat{B}_{ij}\right]$ *and* $Var\left[\hat{P}^k_{MLE;ij}\right]$ *given that* $Cov\left[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\right] < Var\left[\hat{B}_{ij}\right] < Var\left[\hat{P}^k_{MLE;ij}\right]$. *Figure III.4 shows the shape of the variance function when* $Cov\left[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\right] > Var\left[\hat{B}_{ij}\right]$.

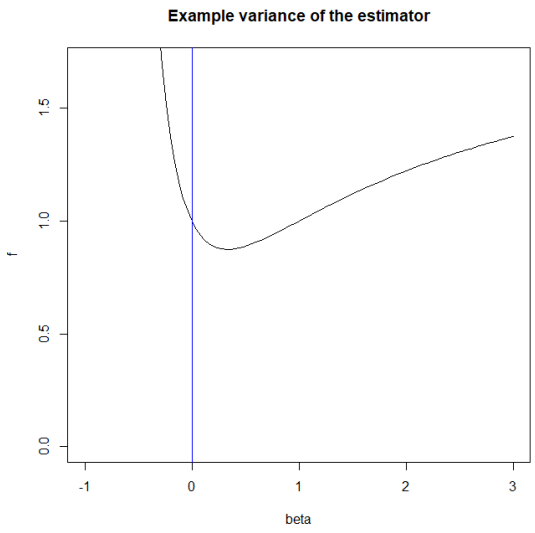

Figure III.3: Example variance of $\hat{P}^k_{ij}$ with small covariance

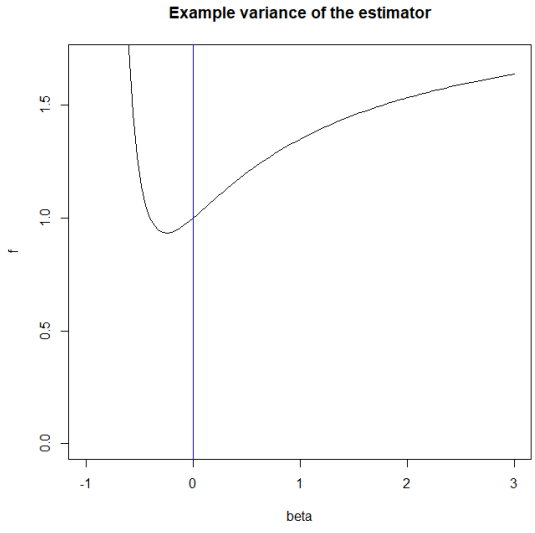

Figure III.4: Example variance of $\hat{P}^k_{ij}$ with large covariance

Since the variance for every element of $\hat{P}^k$ is different but uses the same β, one β would have to minimize all variances at the same time. This is hardly possible. But a value of β that minimizes the sum of all element variances can be found. It is given by the following theorem.

Theorem 3.11. *Let* β_{min} *be the value of* β *in* $\hat{P}^k$ *that minimizes*

$$V(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{N} Var(\hat{P}^k_{ij}).$$

It is given by

$$\beta_{min} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} Var(\hat{B}_{ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N} Cov(\hat{B}_{ij}, \hat{P}_{MLE;ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N} Var(\hat{P}_{MLE;ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N} Cov(\hat{B}_{ij}, \hat{P}_{MLE;ij})}$$

as long as

$$\sum_{i=1}^{N}\sum_{j=1}^{N} Var(\hat{B}_{ij}) > \sum_{i=1}^{N}\sum_{j=1}^{N} Cov(\hat{B}_{ij}, \hat{P}_{MLE;ij}).$$

Proof. Let $\mathrm{Var}(\hat{B}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}\left[\hat{B}_{ij}\right]$, let $\mathrm{Var}(\hat{P}^k_{MLE}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}\left[\hat{P}^k_{MLE;ij}\right]$ and let $\mathrm{Cov}(\hat{B}, \hat{P}^k_{MLE}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(\hat{B}_{ij}, \hat{P}_{MLE;ij})$ Then:

$$\begin{aligned}
V(\beta) &= \sum_{i=1}^{N}\sum_{j=1}^{N} Var(\hat{P}^k_{ij}) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \left[\frac{1}{(1+\beta)^2}\mathrm{Var}\left[\hat{B}_{ij}\right] + \frac{\beta^2}{(1+\beta)^2}\mathrm{Var}\left[\hat{P}^k_{MLE;ij}\right] + \frac{2\beta}{(1+\beta)^2}\mathrm{Cov}\left[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\right] \right] \\
&= \frac{1}{(1+\beta)^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}\left[\hat{B}_{ij}\right] + \frac{\beta^2}{(1+\beta)^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}\left[\hat{P}^k_{MLE;ij}\right] + \frac{2\beta}{(1+\beta)^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}\left[\hat{B}_{ij}, \hat{P}^k_{MLE;ij}\right] \\
&= \frac{1}{(1+\beta)^2}\mathrm{Var}(\hat{B}) + \frac{\beta^2}{(1+\beta)^2}\mathrm{Var}(\hat{P}^k_{MLE}) + \frac{2\beta}{(1+\beta)^2}\mathrm{Cov}(\hat{B}, \hat{P}_{MLE})
\end{aligned}$$

To minimize this function of β, the derivative is taken and set equal to 0 to obtain a critical value of β.

$$
\begin{aligned}
V'(\beta) &= -\frac{2}{(1+\beta)^3}\mathrm{Var}(\hat{B}) + \frac{2\beta(1+\beta)^2 - \beta^2 2(1+\beta)}{(1+\beta)^4}\mathrm{Var}(\hat{P}^k_{MLE}) \\
&\quad + \frac{2(1+\beta)^2 - 4\beta(1+\beta)}{(1+\beta)^4}\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) \\
&= -\frac{2}{(1+\beta)^3}\mathrm{Var}(\hat{B}) + \frac{2\beta(1+\beta) - 2\beta^2}{(1+\beta)^3}\mathrm{Var}(\hat{P}^k_{MLE}) \\
&\quad + \frac{2(1+\beta) - 4\beta}{(1+\beta)^3}\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) \\
&= -\frac{2}{(1+\beta)^3}\mathrm{Var}(\hat{B}) + \frac{2\beta}{(1+\beta)^3}\mathrm{Var}(\hat{P}^k_{MLE}) + \frac{2-2\beta}{(1+\beta)^3}\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) \\
&= \frac{-2\mathrm{Var}(\hat{B}) + 2\beta\mathrm{Var}(\hat{P}^k_{MLE}) + 2\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) - 2\beta\mathrm{Cov}(\hat{B}, \hat{P}_{MLE})}{(1+\beta)^3}
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{-2\mathrm{Var}(\hat{B}) + 2\beta\mathrm{Var}(\hat{P}^k_{MLE}) + 2\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) - 2\beta\mathrm{Cov}(\hat{B}, \hat{P}_{MLE})}{(1+\beta)^3} \\
0 &= -2\mathrm{Var}(\hat{B}) + 2\beta\mathrm{Var}(\hat{P}^k_{MLE}) + 2\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) - 2\beta\mathrm{Cov}(\hat{B}, \hat{P}_{MLE}) \\
\beta &= \frac{\mathrm{Var}(\hat{B}) - \mathrm{Cov}(\hat{B}, \hat{P}_{MLE})}{\mathrm{Var}(\hat{P}^k_{MLE}) - \mathrm{Cov}(\hat{B}, \hat{P}_{MLE})} \\
\beta &= \frac{\sum_{i=1}^N \sum_{j=1}^N Var(\hat{B}_{ij}) - \sum_{i=1}^N \sum_{j=1}^N \mathrm{Cov}(\hat{B}_{ij}, \hat{P}_{MLE;ij})}{\sum_{i=1}^N \sum_{j=1}^N Var(\hat{P}_{MLE;ij}) - \sum_{i=1}^N \sum_{j=1}^N \mathrm{Cov}(\hat{B}_{ij}, \hat{P}_{MLE;ij})}
\end{aligned}
$$

To prove that this is in fact a minimum, the second derivative test is used.

$$\begin{aligned}
V''(\beta) &= \frac{6}{(1+\beta)^4}\text{Var}(\hat{B}) + \frac{2(1+\beta)^3 - 6\beta(1+\beta)^2}{(1+\beta)^6}\text{Var}(\hat{P}^k_{MLE}) \\
&\quad + \frac{(-2)(1+\beta)^3 - 3(2-2\beta)(1+\beta)^2}{(1+\beta)^6}\text{Cov}(\hat{B}, \hat{P}_{MLE}) \\
&= \frac{6}{(1+\beta)^4}\text{Var}(\hat{B}) + \frac{2(1+\beta) - 6\beta}{(1+\beta)^4}\text{Var}(\hat{P}^k_{MLE}) \\
&\quad + \frac{(-2)(1+\beta) - (6-6\beta)}{(1+\beta)^4}\text{Cov}(\hat{B}, \hat{P}_{MLE}) \\
&= \frac{6\text{Var}(\hat{B}) + (2-4\beta)\text{Var}(\hat{P}^k_{MLE}) + (4\beta-8)\text{Cov}(\hat{B}, \hat{P}_{MLE})}{(1+\beta)^4}.
\end{aligned}$$

Let $\beta_0 = \dfrac{\text{Var}(\hat{B}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})}{\text{Var}(\hat{P}^k_{MLE}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})}$.

Observe that $\beta_0 > 0$ if $\text{Var}(\hat{B}) > \text{Cov}(\hat{B}, \hat{P}_{MLE})$. Then

$$V''\Big(\beta_0\Big) = \frac{6\text{Var}(\hat{B}) + \Big(2-4\beta_0\Big)\text{Var}(\hat{P}^k_{MLE}) + (4\beta_0-8)\text{Cov}(\hat{B}, \hat{P}_{MLE})}{\Big(1+\beta_0\Big)^4}.$$

$0 < V''\Big(\beta_0\Big)$ when the following equivalent statements hold

$$0 < 6\text{Var}(\hat{B}) + \Big(2-4\beta_0\Big)\text{Var}(\hat{P}^k_{MLE}) + (4\beta_0-8)\text{Cov}(\hat{B}, \hat{P}_{MLE})$$

$$0 < 6\text{Var}(\hat{B}) + 2\text{Var}(\hat{P}^k_{MLE}) - 4\beta_0\text{Var}(\hat{P}^k_{MLE}) + 4\beta_0\text{Cov}(\hat{B}, \hat{P}_{MLE}) - 8\text{Cov}(\hat{B}, \hat{P}_{MLE})$$

$$0 < 6\Big(\text{Var}(\hat{B}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})\Big) + (2-4\beta_0)\Big(\text{Var}(\hat{P}^k_{MLE}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})\Big)$$

$$(4\beta_0-2)\Big(\text{Var}(\hat{P}^k_{MLE}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})\Big) < 6\Big(\text{Var}(\hat{B}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})\Big)$$

$$\frac{2}{3}\beta_0 - \frac{1}{3} < \frac{\text{Var}(\hat{B}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})}{\text{Var}(\hat{P}^k_{MLE}) - \text{Cov}(\hat{B}, \hat{P}_{MLE})}$$

$$\frac{2}{3}\beta_0 - \frac{1}{3} < \beta_0$$

$$\beta_0 > -1.$$

Since this proves that the second derivative is larger than 0 for the critical value, it is in fact

a minimum. Therefore:

$$\beta_{min} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} Var(\hat{B}_{ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(\hat{B}_{ij}, \hat{P}_{MLE;ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N} Var(\hat{P}_{MLE;ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(\hat{B}_{ij}, \hat{P}_{MLE;ij})}$$

□

The problem is that the optimal, variance minimizing value of β, β_{min}, is dependent on the true transition probabilities of the firms, which are unknown. However, there are two possible methods of estimating them in order to estimate β_{min}. The obvious approach is to use the base matrix estimation from Section III.3.1. This approach might not be optimal in the sense that it replaces all P_{ij} with the same probability even though it is known that they are different for different firms. However, the actual estimations $\hat{P}_{ij}$ are dependent on β. So how can a variance minimizing β be found before the estimates $\hat{P}_{ij}$ are known? A reasonable approach would be to choose a first β according to the first three methods in Remark 3.7 and then use those estimates to estimate a new, variance minimizing β iteratively.

III.4 Estimating the Aggregate Markov Chain

When estimating the transition probability matrix for the aggregate Markov chain, the method from the idiosyncratic Markov chain cannot be used. There is only one market and there is no "base estimation". The problem is, again, that the number of data points will be limited in most real life data sets. However, the way the sales of the firms within the market move is known, or at least can be properly estimated using the method in Section III.3. This knowledge can be used to estimate the transition probabilities of the aggregate Markov chain. The methods in this section can be used when sales data is available. In this thesis, however, data is simulated and sales would have to be calculated from the simulated productivity and employment data. Using simulated sales data significantly complicates the process of estimating the aggregate Markov chain. It also introduces the problem that employment decisions have to be made during the simulation while in real life data these

decisions have already been made. Therefore, this section explores a theoretical approach to how estimation would work if real life data were available, rather than showing the estimation with simulated data.

As described in Section III.1, the aggregate productivity is measured by the ratio of the sales in time period t against the sales in time period $t-1$ and its state space is given by discrete intervals. Like the idiosyncratic Markov chain, it is time-homogeneous. A special property of the aggregate Markov chain is, that not only the value of $A(t-1)$ is known but all firms' sales in the time period $t-1$ and their transition probabilities are known or estimated. The transition probabilities of the Markov chain $\{A(t)\}_{t=1,..,T}$ can then be reformulated as shown in the following theorem.

Theorem 4.1. *Let A_{ij} be the probability that the Markov chain $\{A(t)\}_{t=1,..,T}$ is in state i at time $t-1$ and in state j at time t. Let $S_A = \{I_1, I_2, ..., I_{N_A}\}$ be the state space of $\{A(t)\}_{t=1,..,T}$ and let $I_n = (a_n, b_n]$, where $a_n < b_n$ for all $n = 1, ..., N_A$, be the intervals in that state space. Let F be the number of firms in the market. Let $S(t)$ denote the total sales in the market and let $S^k(t)$ denote the sales at firm k in time period t. Then, by the definition of $A(t)$,*

$$\begin{aligned} A_{ij} = {} & P(S(t) \leq b_j S(t-1) \mid S^k(t-1) = s^k_{t-1} \forall k = 1, ..., F) \\ & - P(S(t) \leq a_j S(t-1) \mid S^k(t-1) = s^k_{t-1} \forall k = 1, ..., F). \end{aligned}$$

Proof.

$$\begin{aligned} A_{ij} &= P(A(t) \in I_j \mid A(t-1) \in I_i) \\ &= P(A(t) \in (a_j, b_j] \mid S^k(t-1) = s^k_{t-1} \forall k = 1, ..., F) \\ &= P(S(t) \in (a_j S(t-1), b_j S(t-1)] \mid S^k(t-1) = s^k_{t-1} \forall k = 1, ..., F) \\ &= P(S(t) \leq b_j S(t-1) \mid S^k(t-1) = s^k_{t-1} \forall k = 1, ..., F) \\ &\quad - P(S(t) \leq a_j S(t-1) \mid S^k(t-1) = s^k_{t-1} \forall k = 1, ..., F). \end{aligned}$$

□

However, to move forward the sales and their transition probabilities have to be known. The sales are usually available in real life data. When divded into intervals, they can be seen as a Markov chain, similar to $Z^k(t)$.

Definition 4.2. *For each firm $k \in \{1, 2, ..., F\}$ in a certain market, the process $S^k(t)$ describes the firm's sales in time period t. The process satisfies*

$$S^k(t) \in (a_t^k L^k(t), b_t^k L^k(t)],$$

where $L^k(t)$ is the number of employees for the firm k in the time period t and $(a_t^k, b_t^k]$ is the state of $Z^k(t)$, the firm-specific productivity process, at time t.

Calculating these transition probabilities directly from Z and L is difficult since the two processes are not independent and L is not a Markov chain. Instead, a reasonable approach is to assume that $S^k(t)$ moves according to a Markov chain with transition probabilities S_{ij} and for all k, $S^k(t)$ takes values from the same state space I which consists of a finite number of intervals.

$$I = \{I_i = [c_i, d_i)] \mid i = 1, 2, ..., M, 0 \leq a_1 < b_1 \leq a_2 < b_2 \leq ... < b_M\}.$$

If $S^k(t)$ is a Markov chain, its transition probabilities, denoted as S_{ij},can be estimated in the exact same way described for Z^k in Section III.3, given that real life sales data is available. The next theorem shows what the probability for the total sales in time period t being less than a certain number is, while the sales of all firms in time period $t-1$ are given.

Theorem 4.3. *Let $P(I_j = x \mid S_{t-1})$ denote the conditional probability that there are x or less firms in state $I_j \in I$ at time t. Let S_{t-1} denote the knowledge that $S^k(t-1) = s_{t-1}^k \forall k =$*

$1, ..., F$. *Then*

$$P(S(t) \leq s \mid S_{t-1})$$
$$\geq \sum_{\tilde{x} \in X_U} P(I_1 = x_1 \mid S_{t-1})\ P(I_t = x_2 \mid S_{t-1})\ \ldots\ P(I_{M-1} = x_{M-1} \mid S_{t-1})$$

where

$$X_U = \{\tilde{x} = (x_1, ..., x_M) \mid x_1 + x_2 + ... + x_M = F, 0 \leq x_j \leq U_j\ \ \forall j = 1, ..., M\}$$

and

$$U_j = min\left(\left\lfloor \frac{1}{d_j}\Big(s - \sum_{n=1}^{j-1} x_n d_n\Big) \right\rfloor, F - \sum_{i=1}^{j-1} x_i \right).$$

Also

$$P(S(t) \leq s \mid S_{t-1})$$
$$\leq \sum_{\tilde{x} \in X_O} P(I_1 = x_1 \mid S_{t-1})\ P(I_t = x_2 \mid S_{t-1})\ \ldots\ P(I_{M-1} = x_{M-1} \mid S_{t-1})$$

where

$$X_O = \{\tilde{x} = (x_1, ..., x_M) \mid x_1 + x_2 + ... + x_M = F, 0 \leq x_j \leq O_j\ \ \forall j = 1, ..., M\}$$

and

$$O_j = min\left(\left\lfloor \frac{1}{d_j}\Big(s - \sum_{n=1}^{j-1} x_n d_n\Big) \right\rfloor + 1, F - \sum_{i=1}^{j-1} x_i \right).$$

d_j is the upper bound for the interval I_j from the state space I from the sales Markov chain.

Only the probabilities from 1 to $M-1$ need to be multiplied since the last probability is 1 if all other x's are known. This comes from the constraint that all x's have to sum up to the total number of firms. The intuition of this theorem is, that it sums up the probabilities of all possible firm state combinations that would sum up to total market sales less than s. This probability interval becomes more accurate the finer the state space is since s is a number and the states of the sales process are intervals. By using the upper bound of those intervals in the first inequality, it is made sure that the sales are never greater than s, but

if the sales are actually in the lower region of an interval, some possible scenarios might be missed in the summation. But if an additional summand is added to every sum, the sales might be slightly greater than s. To further evaluate the expression on the right side of the inequalities, the following theorem can be used.

Theorem 4.4. *Let $P((i,j) = x \mid S_{t-1})$ denote the conditional probability that x firms go from state i to state j given S_{t-1}. Then, given the history,*

$$P(I_j = x \mid S_{t-1}) = \sum_{\tilde{x} \in X_R} P((1,j) = x_1 \mid S_{t-1})\ P((2,j) = x_2 \mid S_{t-1})\ \ldots\ P(M,j) = x_M \mid S_{t-1}),$$

where

$$X_R = \{\tilde{x} = (x_1, ..., x_M) \mid x_1 + x_2 + ... + x_M = x, 0 \leq x_j \leq R_j\ \forall j = 1, ..., M\}$$

and

$$R_j = \min\left(n_j, x - \sum_{k=1}^{j-1} x_k\right).$$

Next, $P((i,j) = x \mid S_{t-1})$ is investigated. If all firms had the same transition probabilities B_{ij}, then $P((i,j) = x \mid S_{t-1}) = \binom{n_i}{t}(B_{ij})^x(1 - B_{ij})^{n_i - x}$. However, every firm has different transition probabilities S_{ij}^k for the sales Markov chain. The following theorem gives a more accurate representation of $P((i,j) = x \mid S_{t-1})$.

Theorem 4.5. *Let $S_{x,a,i}$ denote the a^{th} subset of size x of the firms that were in state i in the last time period (firms can appear in this set multiple times). Then*

$$P((i,j) = x \mid S_{t-1}) = \sum_{a=1}^{\binom{n_i}{x}} \prod_{k \in S_{x,a,i}} S_{ij}^k \prod_{k \notin S_{x,a,i}} (1 - S_{ij}^k).$$

The following simple example better illustrates the whole procedure.

Example 4.6. *Let the sales process have $M = 3$ states and $F = 3$ firms moving in it through*

$T = 4$ time periods. The state space of the sales process is $I = \{(0, 100], (100, 200], (200, 300]\}$. Let $s = 500$. The following reduced contingency table was recorded.

$$C = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & 1 \end{pmatrix} \end{array}$$

To set up the first two inequalities, the index sets X_U and X_O must be defined. In this case

$$X_U = \{(0, 2, 1), (1, 2, 0), (2, 0, 1), (2, 1, 0), (3, 0, 0)\},$$

$$X_O = \{(0, 1, 2), (0, 2, 1), (0, 3, 0), (1, 0, 2), (1, 1, 1), (1, 2, 0), (2, 0, 1), (2, 1, 0), (3, 0, 0)\}.$$

Then, the first inequality resolves to

$$\begin{aligned} P(S(t) \leq 500 \mid S_{t-1}) = {} & P(I_1 = 0 \mid S_{t-1})P(I_2 = 2 \mid S_{t-1}) + P(I_1 = 1 \mid S_{t-1})P(I_2 = 2 \mid S_{t-1}) \\ & + P(I_1 = 2 \mid S_{t-1})P(I_2 = 0 \mid S_{t-1}) + P(I_1 = 2 \mid S_{t-1})P(I_2 = 1 \mid S_{t-1}) \\ & + P(I_1 = 3 \mid S_{t-1}). \end{aligned}$$

Then, for example, for the equation of $P(I_1 = 2 \mid S_{t-1})$, the set X_R must be defined. In this case

$$X_R = \{(0, 0, 2), (0, 1, 1), (0, 2, 0), (1, 1, 0), (1, 0, 1), (2, 0, 0)\}.$$

This means that for example in the first tuple, 0 firms move from state 1 to state 1, 0 firms move from state 2 to state 1, but 2 firms move from state 3 to state 1. Now, the equation of $P(I_1 = 2 \mid S_{t-1})$ resolves to

$$P(I_1 = 2 \mid S_{t-1}) = P((1,1) = 0 \mid S_{t-1})P((2,1) = 0 \mid S_{t-1})P((3,1) = 2 \mid S_{t-1}) + \ldots$$
$$+ P((1,1) = 2 \mid S_{t-1})P((2,1) = 0 \mid S_{t-1})P((3,1) = 0 \mid S_{t-1}).$$

Next, the individual probabilities in this equation are resolved. The n_i needed for this can be taken from the contingency table C. For example,

$$P((1,2) = 1 \mid S_{t-1}) = \sum_{a=1}^{2} \prod_{k \in S_{1,a,1}} P_{ij}^k \prod_{k \notin S_{1,a,1}} (1 - P_{ij}^k).$$

So, if the firms in state 1 *were Firm 1 and Firm 2, then there are only two possible subsets of size 1:*

$$S_{1,1,1} = \{1\}, S_{1,2,1} = \{2\}.$$

Then, the probability resolves to

$$P((1,2) = 1 \mid S_{t-1}) = S_{ij}^1(1 - S_{ij}^2) + S_{ij}^2(1 - S_{ij}^1).$$

The transition probabilities S_{ij}^k must be estimated using the methods in Section III.3. In further research, the transition probabilities of A might be directly calculated from the processes Z and L. Once the transition probabilities A_{ij} are estimated, either an exact multinomial test or a likelihood ratio test can be used to test the goodness of fit of A to the market movement.

Summary and Outlook

IV.1 Summary of Results

In this thesis, a methodology of estimating the parameters of a complex Markov model for firm and aggregate productivity was developed. A single firm's productivity is assumed to follow a Markov chain with its own transition probability matrix. Maximum likelihood estimation often yields unsatisfying results with a large variance since data for a single firm is limited. However, if it is assumed that all firms in a market follow a similar behavior, a variation of a base estimation can be used to find an estimate for a single firm's productivity Markov chain. This estimator starts with a maximum likelihood estimate based on all firms in the market and slightly increases the probabilities of the transitions that have actually occurred for this firm. The level of increase depends on a factor β. Under certain conditions this factor can also be chosen to minimize the variance of the new estimator. Different goodness-of-fit tests can be used to test the fit of the new estimate for a single firm. If the sample size and number of states are small, a multinomial test can be used, otherwise a likelihood ratio test is applicable. A Pearson-Chi-Square test can be used for the base estimation or when a lot of single firm data is available. While increasing the factor β always increases the goodness-of-fit, it also increases the estimators variance if it cannot be minimized. Therefore, different methods of choosing β are proposed, for example a goodness-of-fit condition.

While this thesis uses simulated data to illustrate this methodology, applications would use real life data which often includes employment levels and sales. If sales data is available, it can be used to estimate the transition probabilities of the aggregate productivity Markov chain. This estimation is complex and depends on sales processes from all firms in the market. This thesis explores a theoretical approach to this estimation and breaks down the formulas in an example. Both estimation methods, for the firm and aggregate Markov chains, can be used on real life data to help with solving optimal decision making problems for employment levels.

IV.2 Outlook

Many different steps can be taken to incorporate the results of this thesis in further research. The obvious next step is to use the methods developed in this thesis on real life data to estimate idiosyncratic and aggregate Markov chains. When these Markov chains have been estimated, the rest of the profit formula that was shown in the introduction could be analyzed. Ultimately, an approach to solving labor demand problems can be developed. A possible direct extension of this thesis would be to analyze the expectation and variance of the aggregate estimator. Another way of refining the model is to incorporate the wage rate as a stochastic process and estimate it using similar methods as in this thesis. If the approach of using Markov chains is in question, methods as proposed by Eggar (2002) could be employed to validate the Markov model itself.

Bibliography

Bishop, Y. M. M. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Massachusetts.

Casella, G. and Berger, R. L. (2001). *Statistical Inference*. Cengage Learning, 2 edition.

Cooper, R., Haltiwanger, J., and Willis, J. (2007). Search frictions: Matching aggregate and establishment observations. *Journal of Monetary Economics*, 54:56–78.

Eggar, M. H. (2002). Validity of fitting a first-order markov chain model to data. *The Statistician*, 51(2):259–265.

Elsby, M. and Michaels, R. (2014). Fixed adjustment costs and aggregate fluctuations. *American Economic Review*.

Gagniuc, P. (2017). *Markov Chains: From Theory to Implementation and Experimentation*. Wiley.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*. Academic Press, 2 edition.

Khan, A. and Thomas, J. (2004). Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment. *Econometrica*, 76(2):395–436.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated systems of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical magazie*, 50(302):157–175.

Pfanzagl, J. and Hamboeker, R. (1994). *Parametric Statistical Theory*. Walter de Gruyter.

Read, T. R. and Cressie, N. A. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York.

Rota, P. (2004). Estimating labor demand with fixed costs. *International Economic Review*, 45:24–48.

Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley and Sons, New York.

Appendix

A Data Simulation in R

```
library(expm)
library(pracma)


#Simulating F N-state markov chain over T time periods,
#each having slightly different transition matrices

F=10
N=4
T=3

stateNames = seq(10000,10000*N,10000)
for(k in 1:N){
 stateNames[k]=toString(stateNames[k])
}


#base transition matrix that makes it less likely
#to transition to states that are far away
normal = 0
basematrix = matrix(0L,nrow=N,ncol=N)
for (i in 1:N){
 vec = seq(0,0,length.out=N)
 normal = 0
 for (j in 1:N){
  p = 1/2*1/(abs(i-j)+1)
  normal = normal+p
  vec[j]=p
 }
 basematrix[i,]=vec*1/normal
}
```

```
#vary the transition matrix for each firm -- STARTING HERE
matrixList = list()

#Randomly create market trend
mtrend = runif(1)

for(k in 1:F){
 matrixList[[k]]=matrix(0L,nrow=N,ncol=N)

 #Does the firm tend to do better or worse?
 #The better the market trend is, the more likely the firm is to do well
 f = runif(1)
 ftrend = runif(1)
 ftrend = sign(mtrend-f)*ftrend*1/8


 for(i in 1:N){
  vec = seq(0,0,length.out=N)
  normal = 0
  for (j in 1:N){

   if (i-j<0){
    #if the trend is positive the probability of getting
    #into a higher sales class in increased
    pInc = rnorm(1,ftrend,2*abs(ftrend))*basematrix[i,j]
      +basematrix[i,j]
   } else if (i-j>0) {
    #if the trend is positive the probability of falling
    #into a lower sales class is decreased
    pInc = rnorm(1,(-1)*ftrend,2*abs(ftrend))*basematrix[i,j]
     +basematrix[i,j]
   } else {
    pInc = rnorm(1,1/16,1/8)*basematrix[i,j]+basematrix[i,j]
   }
   normal = normal+pInc
   vec[j]=pInc
  }
  matrixList[[k]][i,]=vec*1/normal
 }

}

#vary the transition matrix for each firm -- ENDING HERE
```

```
#Randomized starting vector assigns firms to the states (state boundaries)

n = matrix(0,nrow=T,ncol=N+1)
n[1,] = c(0,sort(round(runif(N-1)*F)),F)


#Simulate the transitions and create contingiency tables -- STARTING HERE
firms = array(rep(c(1:F),T),dim=c(F,T))

contin = list()

for(k in 1:(T-1)){
 contin[[k]]=matrix(0,ncol=N,nrow=N)
}

statesNew = list()
#Create T-1 contingiency tables
for (k in 1:(T-1)){

 #Create Matrix for new state assignments
 for(i in 1:N){
  statesNew[[i]]=vector()
 }

 for( j in 1:N){
  for (i in firms[(n[k,j]+1):n[k,j+1],k]){
   #For all firms in this state

   #Generate random number to decide new state
   x1 = runif(1)


   #Find the index of the first cumulative P where x1<=P using the
   #transition matrix for this firm
   ns = 1
   while(x1>sum(matrixList[[i]][j,1:ns])){
    ns = ns+1
   }

   statesNew[[ns]]=c(statesNew[[ns]],i)
   contin[[k]][j,ns] = contin[[k]][j,ns]+1
  }

 }
```

```
 #assign new state boundaries
 for(i in 2:N){
  length = 0
  for(j in 1:(i-1)){
   length = length + length(statesNew[[j]])
  }
  n[k+1,i]=length

 }
 n[k+1,N+1]=F

 #assign new order of the firms
 newFirms=vector()
 for(i in 1:N){
  newFirms = c(newFirms,statesNew[[i]])
 }
 firms[,k+1] = newFirms
}

#Simulate the transitions and create contingiency tables -- ENDING HERE


#Simulate employment levels -- STARTING HERE

L = matrix(0,ncol=T,nrow=F)

L[,1]=round(1000*abs(rnorm(F,0,1)))

Av = (N+1)/2

for(k in 2:T){
 for(f in 1:F){
  index = match(f,firms[,k-1])
  s = 0
  while(index>n[k-1,s+1]){
   s = s + 1
  }

  #At time k-1, firm f was in state s

  if(s-Av>0){
   newemp = abs(round(rnorm(1,(s-Av)*L[f,k-1]/50,2)))
  } else if (s-Av<0){
   newemp = -abs(round(rnorm(1,(s-Av)*L[f,k-1]/50,2)))
  } else{
```

```
    newemp = round(rnorm(1,0,2))
   }

   L[f,k]=L[f,k-1]+newemp

  }
}

#Simulate employment levels -- ENDING HERE
```

B Parameter Estimation in R

```
library(XNomial)
#Estimating base-matrix -- STARTING HERE

P_hat=matrix(0,nrow=N,ncol=N)

#sum up all contingency matrices, because of time-homogenuity
sumcontin = Reduce('+',contin)


for(i in 1:N){

        for(j in 1:N){
                P_hat[i,j] = sumcontin[i,j]/sum(sumcontin[i,])
        }

}

basematrix-P_hat

#Estimating base-matrix -- ENDING HERE


#Goodness of fit test, Pearson Chi Square for the base matrix

X2 = 0

for (k in 1:(T-1)){
for (i in 1:N){

Oi = sum(contin[[k]][,i])
```

```
u=vector()
for (j in 1:N){
u = c(u,sum(contin[[k]][j,]))
}

Ei = (u%*%P_hat)[i]

X2 = X2 + (Oi-Ei)^2/Ei

}
}

X2
1-pchisq(X2, N*(T-N))

#Estimate transition probabilities for single firm -- START HERE

#Finding contingency table of single firm f

f=2

continsf = matrix(0,nrow=N,ncol=N)

for(k in 2:T){

        continsf[states[f,k-1],states[f,k]]
                = continsf[states[f,k-1],states[f,k]]+1

}
continsf

#Use maximum likelihood estimation

P_hatsf = matrix(0,nrow=N,ncol=N)

for(i in 1:N){

        for(j in 1:N){
                P_hatsf[i,j] = continsf[i,j]/sum(continsf[i,])
        }

}

#Estimate transition probabilities for single firm -- ENDING HERE
```

```
#Goodness of fit test (Multinomial) for single firm with base
#matrix estimation

M=vector()

for(i in 1:N){

        observed = continsf[i,]
        if(sum(observed)>0){
                expected = P_hat[i,]

                V=N
                j=1
                while(j<=V){
                        if(expected[j]==0 && observed[j]==0){
                                observed=observed[-j]
                                expected=expected[-j]
                                j=j-1
                                V=V-1
                        }
                        j=j+1
                }

                mul = xmulti(observed,expected,detail=2);

                M = c(M,mul$pProb)
        }
}

M

sum(M)/length(M)

#Improve goodness of fit by linear combination approach

beta = seq(0.01,0.2,by=0.01)

P_hat_refined = list()

for(k in 1:(length(beta))){

        P_hat_refined[[k]] = P_hat+beta[k]*P_hatsf

        normal = vector()
        for(i in 1:N){
```

```
                normal[i]=sum(P_hat_refined[[k]][i,])
        }

        for(i in 1:N){
                P_hat_refined[[k]][i,]=P_hat_refined[[k]][i,]*1/normal[i]
        }

}


#Test goodness of fit for the new refined P_hat estimation

k_test = 2

M=matrix(0,nrow=length(beta),ncol=N)
sink("output.txt")
for(k in 1:length(beta)){
for(i in 1:N){

        observed = continsf[i,]
        if(sum(observed)>0){
                expected = P_hat_refined[[k]][i,]

                V=N
                j=1
                while(j<=V){
                        if(expected[j]==0 && observed[j]==0){
                                observed=observed[-j]
                                expected=expected[-j]
                                j=j-1
                                V=V-1
                        }
                        j=j+1
                }

                mul = xmulti(observed,expected,detail=2);

                M[k,i]=mul$pProb
        }
}
}
sink()
M

M_avg = vector()
```

```
for(k in 1:length(beta)){
        M_avg[k]=sum(M[k,])/length(M[k,])
}
```