Issues in the Distribution Dynamics Approach to the Analysis of Regional Economic Growth and Convergence: Spatial Effects and Small Samples

by

Wei Kang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Sergio Rey, Chair
Stewart Fotheringham
Xinyue Ye

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

In the study of regional economic growth and convergence, the distribution dynamics approach which interrogates the evolution of the cross-sectional distribution as a whole and is concerned with both the external and internal dynamics of the distribution has received wide usage. However, many methodological issues remain to be resolved before valid inferences and conclusions can be drawn from empirical research. Among them, spatial effects including spatial heterogeneity and spatial dependence invalidate the assumption of independent and identical distributions underlying the conventional maximum likelihood techniques while the availability of small samples in regional settings questions the usage of the asymptotic properties. This dissertation is comprised of three papers targeted at addressing these two issues. The first paper investigates whether the conventional regional income mobility estimators are still suitable in the presence of spatial dependence and/or a small sample. It is approached through a series of Monte Carlo experiments which require the proposal of a novel data generating process (DGP) capable of generating spatially dependent time series. The second paper moves to the statistical tests for detecting specific forms of spatial (spatiotemporal) effects in the discrete Markov chain model, investigating their robustness to the alternative spatial effect, sensitivity to discretization granularity, and properties in small sample settings. The third paper proposes discrete kernel estimators with cross-validated bandwidths as an alternative to maximum likelihood estimators in small sample settings. It is demonstrated that the performance of discrete kernel estimators offers improvement when the sample size is small. Taken together, the three papers constitute an endeavor to relax the restrictive assumptions of spatial independence and spatial homogeneity, as well as demonstrating the difference between the small sample and asymptotic properties for conventionally adopted maximum likelihood estimators towards a more valid inferential

framework for the distribution dynamics approach to the study of regional economic growth and convergence.

To

My Husband, Hu Shao

and

My Parents, Jufang Dai & Qingcheng Kang

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Prof. Sergio Rey, for his enlightenment and support throughout my PhD study. It is a great fortune to have such a knowledgable yet modest advisor. My gratitude also goes to my committee member, Prof. Stewart Fotheringham, for his inspiration and generous support for my academic career. I also extend my gratitude to Prof. Xinyue Ye for his constructive advice on the dissertation.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Background and Purpose of the Research

The study of economic growth and convergence has been greatly developed both theoretically and methodologically since the publication of the seminal paper of Baumol (1986). Many studies are centered around the famous β-convergence hypothesis implicated by Solow growth theory (Solow, 1956; Barro and Sala-i Martin, 2003). The hypothesis refers to the situation where poorer economies are catching up with richer ones in per capita incomes and has been examined in numerous international and regional settings using various econometric techniques ranging from cross-sectional econometrics, time series econometrics to dynamic panel econometrics (Durlauf, 2001; Durlauf et al., 2005). The benchmark for testing for the β-convergence hypothesis is the convergence equation where the dependent variable is the average annual growth rate of per capita incomes and the independent variable the logarithm of the initial per capita income. A conclusion of *absolute convergence* could be drawn if the estimated coefficient is negative and statistically significant. Various extensions to the benchmark regression equation have been made to account for potential heterogeneity of steady states as well as exploring interactions among economies. The former may give rise to *conditional convergence* if independent variables determining the steady states are added to the right side of the equation (Mankiw et al., 1992), *club convergence* if there exist a multitude of steady state equilibria and convergence rates, or *local convergence* if convergence rates vary across economies but are more similar for closer economies (definition of closeness may vary). The latter could be addressed by relaxing the assumption of closed economies, that is, substituting them with open economies (Barro and Sala-i Martin, 2003). Though a general conditional β-convergence consensus seems to be reached, the perspective on the concrete rate of convergence varies from "Iron law of convergence" contending an about 2% cross-country conditional convergence rate (Barro, 2015, 2016) to the recognition of a wide range of empirical convergence rates (as high as 65.59%) produced in numerous empirical studies (Abreu et al., 2005).

Another important notion of convergence is the so-called σ-convergence which hypothesizes a diminishing tendency of cross-sectional variance in per capita incomes over time (Islam, 2003). σ-convergence is not necessarily implied by the aforementioned β-convergence since it is also impacted by the variance of random shocks occurring to individual economies. Compared with β-convergence, which is usually investigated in a confirmatory framework, σ-convergence is often examined in an exploratory setting.

Both of these convergence notions and approaches have their limitations. β-convergence implicitly assumes a steady-state growth path well approximated by a time trend for each economy, which is not born out by empirical data (Quah, 1993a). In addition, most coefficient estimates for the β-convergence regression are not robust to alternations in the conditioning variables (Levine and Renelt, 1992). As for σ-convergence, it is only concerned with the dynamics of the variance, which is the second moment of the cross-sectional distribution, and neglects the changes in all the other properties. In light of these limitations, Quah proposed the *distribution dynamics approach* in early 1990s aimed at revealing a more complete picture of the dynamics of the cross-sectional per capita income distribution while imposing fewer assumptions about the underlying dynamics (Quah, 1993a,b, 1996a, 1997).

Specifically, the *distribution dynamics approach* interrogates the evolution of the cross-sectional distribution as a whole and is concern with both external and internal dynamics of the distribution. The external dynamics refers to changes in the overall morphological properties of the distributions, such as shape, modality, variance and polarization, while the internal dynamics concerns about the mixing and transitions of individual economies from one part of the distribution to another over time, shedding light on the persistence or mobility of the economies in terms of per capita incomes (Quah, 1996b). Two main types of mobility notions are of interest, structural mobility and exchange mobility (Ruiz-Castillo, 2004). The former measures absolute income changes over time while the latter measures income changes relative to one another. When one is silent in some cases, the other might be able to identify some important mobility patterns. For example, if all the regions encounter the same level of economic growth, their income rank positions remain unchanged. In this case, the exchange mobility measures would not pick up anything while the structural mobility measures could. On the other hand, if the regions only exchange income values, the structural mobility measures would be silent while the exchange mobility measures would not. Thus, these two types serve as complements to one another.

The *distribution dynamics approach* projects the cross-sectional distribution at t to the future $t + s$. The projection operator could be a stochastic kernel (Villaverde and Maza, 2012), or a (k, k) transition probability matrix if the per capita income data are discretized into k income classes. The latter is related to the discrete Markov chain (DMC) model which has a well-developed theory and thus invites more empirical applications. One essential property of the DMC model is the limiting (or steady state) distribution in the long run which echoes the steady states implicated by the Solow growth theory. The difference is that the former implicates the long run in a stochastic sense, claiming the probability of falling into each income class stays the same over time and allowing for transitions between classes, while the latter is deterministic, leaving no space for leapfrogging.

Many empirical studies of economic growth and convergence have been conducted from a regional (subnational) point of view aimed at evaluating and guiding regional policies for eliminating or alleviating regional inequality (Bode and Rey, 2006). Adopting *region* as the unit of analysis invites new challenges as *regions* are usually classified into geographically connected groups at which different regional policies are targeted and they are characterized by profoundly higher degrees of openness than nations potentially inviting strong spatial interactions (Magrini, 2004). The former characteristic could lead to spatial heterogeneity which is related to the lack of stability over space in the growth/convergence process. More precisely, it implies that functional forms or parameters vary with location and are not homogeneous throughout the dataset. The latter leads to spatial dependence, which could be best summarized by Tobler's first law of geography: "everything depends on everything else, but closer things more so" (Tobler, 1970). It may be part of the growth process if originating from spatial interactions among economies such as knowledge flows through trade, foreign direct investment, technology transfers or human capital externalities - substantive spatial dependence. Or, it could be nuisance spatial dependence due to mismatched boundaries induced by data collection (Anselin, 1988). Since either spatial heterogeneity or spatial dependence invalidates the independently and identically distribution ($i.i.d$) assumption underlying the aforementioned classic approaches, these classic approaches need to be extended or adjusted to properly address either form of spatial effect in regional economic growth and convergence (Abreu et al., 2004; Rey and Janikas, 2005).

Various attempts have been made to address spatial effects in the three approaches. For testing for β convergence, two perspectives exist. From a model-driven (theoretical) perspective, the classical closed-form Solow growth theory is augmented to model spatial externalities based on which a convergence equation with spatial components could be derived and estimated using spatial econometric techniques (López-Bazo et al., 2004; Egger and Pfaffermayr, 2006; Ertur and Koch, 2007; Fischer, 2011, 2016). From a data-driven perspective, modern spatial econometrics techniques are performed on the benchmark (conditional) convergence equation where general-to-specific or specific-to-general specification search is conducted leading to the specification of a spatial autoregressive model (SAR), a spatial error model (SEM) or a spatial Durbin Model (SDM) (Rey and Montouri, 1999; Florax et al., 2003; Lesage and Fischer, 2008; LeSage and Pace, 2009). Substantive spatial dependence has been demonstrated in numerous studies, pointing to an explicit spatial econometric specification (Arbia, 2006). Evidence of discrete spatial heterogeneity in the form of spatially explicit club convergence has been found in European regions (Fischer and Stirböck, 2006; Piribauer, 2016) and Chinese counties (Qin et al., 2013). Here, *clubs* are comprised of contiguous regions close to one another. *Local convergence* which refers to the situation where convergence rates vary across regions but are more similar for geographically closer regions has also been found in European regions (Ertur et al., 2007).

It was demonstrated that the empirical sample variance is a combination of aspatial variance and a component capturing spatial dependence and/or spatial heterogeneity in a regional context for σ convergence (Rey and Dev, 2006; Egger and Pfaffermayr, 2006). In other words, the real cross-sectional variance could be overestimated because of potential spatial effects. Evidence was found that the trend in σ convergence of US states 1979-2000 was mostly driven by the dynamics of spatial dependence instead of the dynamics of real cross-sectional variance.

The distributional dynamics approach has also received attention in addressing potential spatial effects. The concept of a spatially conditioned stochastic kernel was proposed by Quah (1997) which maps the distribution of cross-sectional incomes to that of the weighted average of neighbors at the same period. In light of its ignorance of temporal dynamics of the income distribution, Rey (2001) proposed the spatial Markov chain model as a spatial extension of the classic Markov chain model which conditions the transitional dynamics on the regional context. It is said that the probability of transitioning to a specific income class at next period is not only dependent on the current income class but also on the current income levels of neighbors. Another extension to the classic Markov chain model which incorporates discrete spatial heterogeneity is to estimate m transition probability matrices from m subsamples (regimes) of regional time series (Bickenbach and Bode, 2003).

Many methodological issues remain to be resolved in incorporating spatial effects in the distributional dynamics approach (Rey, 2015). Focusing on the discrete Markov chain framework, underexplored issues include the choice of discretization strategy, the specification search, the tests for spatial effects, the properties of the maximum likelihood estimators conventionally used for estimating transition probabilities to the presence of spatial heterogeneity and/or spatial dependence as well as in small sample settings. This dissertation attempts to address some of the issues towards providing an improvement to the current distributional dynamics approach.

1.2 Significance and Contributions

The dissertation is comprised of three potentially publishable papers, each focusing on specific issues related to the spatial dimensions of the *distribution dynamics approach* to the study of regional economic growth and convergence.

The first paper focuses on several regional income mobility measures which are derived from the Markov transition probability matrix, and looks at whether there is a significant impact of spatial dependence on the statistical inference about these measures, and what the form of the impact would be if there is any. The nature of the issue is similar to that for σ convergence (Rey and Dev, 2006; Egger and Pfaffermayr, 2006). That is, properties of conventional estimators assuming $i.i.d$ might be impaired and thus the inference could be sabotaged if the spatial effects are left unattended. The difference is that the measures considered here have a temporal dimension, which complicates the issue. The issue is approached via a series of Monte Carlo experiments which requires the proposal of a novel data generating process (DGP) capable of generating spatially dependent time series given a transition probability matrix and the strength of the spatial dependence. An attempt towards the correction of these statistics to maintain proper size and power properties in the presence of spatial dependence is also made.

The second paper moves to the statistical tests for detecting specific forms of spatial (spatiotemporal) effects in the discrete Markov chain model, including two forms of spatiotemporal dependence, temporally lagged spatial dependence and contemporaneous spatial dependence, as well as spatial heterogeneity. Though the asymptotic properties can be constructed for these tests, the small sample properties remain unexplored. If the asymptotic properties significantly deviate from finite sample properties, invalid inference and conclusions would be made in empirical settings where the available regional income dataset usually spans a quite short period. In addition, Rey et al. (2016) provide evidence of the non-robustness of the test for temporally lagged spatial dependence to that for spatial heterogeneity and vice visa. It is unclear whether the test for contemporaneous spatial dependence suffers from the same issue. Further, the sensitivity of the tests to the discretization granularity of regional incomes is also unclear. A series of Monte Carlo experiments are conducted to shed light on these issues and provides guidance for employing these tests in empirical studies.

The third paper is devoted to addressing the poor behavior of conventionally used maximum likelihood estimators (MLEs) for transition probabilities in small sample settings. More precisely, MLEs could easily give rise to zero estimates of probabilities when sample size is small compared to the number of classes k, constituting a sparse transition matrix which has quite different properties from a non-sparse matrix. The sparsity issue becomes more relevant for the spatial Markov chain model which requires estimating a larger number of parameters. I find most empirical studies of regional income distribution dynamics employing this model produced a large portion of zero transition probabilities, such as the US (Rey, 2001), China (Pu et al., 2005) and Europe (Le Gallo, 2004; Maza et al., 2012). Therefore, estimators which avoid producing too many zero probability estimates and better recover the true underlying dynamics in small sample settings are desirable. The chapter follows Kullback et al. (1962) and views the (k, k) transition matrix for the classic Markov chain model and the (k, k, k) transition matrix for the spatial Markov chain model as two-way and three-way *conditional* contingency tables in the sense that all the cells are filled with conditional rather than joint probabilities. Then I modify the smoothing techniques for high-order contingency tables and the relevant cross-validation technique for smoothing parameter selection to suit the *conditional* contingency tables for Markov and spatial Markov chain models. Monte Carlo experiments are conducted for a comparison of the proposed smoothed estimators and MLEs.

Taken together, the three papers constitute an endeavor towards methodological improvements of the discrete Markov chain approach in studying regional economic growth and convergence. Spatial effects and small sample sizes which are commonly encountered in practice are the two main themes. By relaxing the restrictive assumptions of spatial independence and spatial homogeneity, as well as demonstrating the difference of small sample properties and asymptotic properties for conventionally adopted MLEs, the dissertation seeks to improve spatially explicit distributional dynamics approaches towards a more valid inferential framework for regional economic growth and convergence.

1.3 Organization

The rest of the dissertation is comprised of four chapters. Chapter 2 presents the paper focusing on impacts of cross-sectional spatial dependence on regional income mobility measures and attempts to correct for the dependence. Chapter 3 is the paper on exploring issues related to statistical tests for detecting spatiotemporal dependence and spatial heterogeneity in the discrete Markov chain model, including robustness to the the other form of spatial effect, small sample properties, and sensitivity to discretization granularity. Chapter 4 presents the paper on proposing discrete kernel estimators with cross validation-based smoothing parameters selection for producing less sparse transition probability matrices for the classic and spatial Markov chain models in small sample settings. Chapter 5 concludes with main findings, limitations and potential research directions.

Chapter 2

INFERENCE OF INCOME MOBILITY MEASURES IN THE PRESENCE OF SPATIAL DEPENDENCE

2.1 Introduction

Income inequality is an important subject of interest around the world. Many indices intended for measuring the income inequality of an economic system at a given time point have been developed and popularized, including but not limited to the Gini index, coefficient of variation and Theil's measure (Allison, 1978; Shorrocks, 1980). However, concern is not only about individuals'/households' current economic status, but also where they would end up and their lifetime welfare (Creedy and Wilhelm, 2002; Ruiz-Castillo, 2004; Khor and Pencavel, 2008). It is becoming increasingly recognized that a static view of the income distributions cannot reveal the whole picture, and that the dynamics of income distribution shapes social welfare as well (Schorrocks, 1978; Chakravarty, 1995; Maasoumi, 1998). Thus income mobility measures, which evaluate the changes in economic status over time or generations, serve as a complement to income inequality measures to reveal a fuller picture of income inequality dynamics and social welfare (Fields and Ok, 1996, 1999).

Similar issues arise when the focus shifts from the distribution of incomes taken over **individuals/households** in a society to the question of income distributions of **regions** (Rey, 2015). That is, in a national system what are the properties of the distribution of regional incomes, and how do these evolve over time? Similarly, regional income mobility measures offer a concise way to reveal the mobile nature of the regional income distribution and serves as a complement to regional income inequality measures. There are two main types of income mobility: structural mobility and exchange mobility (Ruiz-Castillo, 2004). The former measures absolute income changes over time while the latter measures income changes relative to one another. When one is silent in some cases, the other might be able to identify some important mobility patterns. For example, if all the regions encounter the same level of economic growth, their income rank positions remain unchanged. In this case, the exchange mobility measures would not pick up anything while the structural mobility measures could. On the other hand, if the regions only exchange income values, the structural mobility measures would be silent while the exchange mobility measures would not. Thus, these two types serve as complements to one another.

Statistical inference about regional income mobility measures is of great importance if a confidence interval is to be constructed for the estimate (Schluter, 1998), let alone when it comes to a comparison of two regional systems. Rey and Ye (2010) compared the regional income mobility 1978-1998 between the U.S. and China based on permutation-based sampling distributions. The theoretical inference framework has been built in Trede (1999) assuming regional time series are independently and identically distributed. However, spatial effects including spatial dependence and spatial heterogeneity are known as more of a rule than exception in a regional context, which poses a serious question: would the spatial effects impair classic inference so significantly that they could not be ignored? This question motivates the research presented in the chapter. Here, I focus on the so-called Markov-based mobility measures. I expect to expose the nature of the impact of spatial dependence on the inference through a series of Monte Carlo simulation experiments. To do this, I propose a novel data generating process (DGP) capable of generating spatially dependent Markov chains given a transition probability matrix and the strength of spatial dependence. Results suggest that spatial dependence does have a major influence on the properties of the mobility estimators and relevant test statistics. Though it does not bias the maximum likelihood estimators (MLEs) of the mobility measures, it dramatically increases the variances of their sampling distributions, raising the Type I error rate for one-sample tests. As for the two-sample tests, the size tends to become increasingly upward biased with stronger spatial dependence in either income system while the power decreases with stronger spatial dependence. The asymptotic properties originating from MLEs do not hold well for small sample sizes: not only the variance is underestimated, but also the MLEs are biased.

For the rest of the chapter I first introduce the definition of three mobility measures, as well as the respective estimators, one-sample and two-sample test statistics. Then a novel data generating process for producing spatially dependent Markov chains is proposed and adopted in a series of Monte Carlo simulation experiments intended for examining the properties of the aforementioned mobility estimators and test statistics. Next I discuss the experiment results and propose adjustments to the critical values of the tests on purpose of maintaining proper size and power properties. In the end I conclude and suggest some further research directions.

2.2 Regional Income Mobility Measures

In this chapter, I focus on Markov-based mobility measures. The motivation is that Discrete Markov Chain (DMC) theory has been widely applied in studying regional income dynamics and convergence (e.g. Quah (1996a); Le Gallo and Chasco (2008); Liao and Wei (2012); Rey and Gutiérrez (2015)) since the estimated transition probability matrix $\boldsymbol{P}$ can reveal abundant information on transition probabilities across states over time. However, the matrix $\boldsymbol{P}$, comprised of m^2 elements (m is the number of discrete states adopted to discretize the income dataset), is not as simple and straightforward as a single index especially when it comes to comparing two regional income systems. In this context, several Markov-based mobility measures have been proposed in the literature, all of which can be calculated from the estimated transition probability matrix.[1] Thus, I start by briefly introducing DMC theory and then proceed to derive the relevant mobility measures.

2.2.1 Discrete Markov Chains (DMC)

As mentioned before, the transition probability matrix $\boldsymbol{P}$, which is the core of DMC, contains information regarding mobility across discrete states over time. Equation (2.1) displays an example of such matrix in which p_{ij} represents the probability of transitioning from state i to state j over a given time interval.

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}, 0 \leq p_{ij} \leq 1, \sum_{j=1}^{m} p_{ij} = 1 \; \forall \; i, j \in \mathbb{S} = \{1, 2, \cdots, m\}. \quad (2.1)$$

Here m states are adopted to discretize the income data. Class boundaries, as well as preliminary transformations of incomes, are determined by the user. Cautions should be taken when making such decisions as different strategies might lead to different results and conclusions regarding income dynamics. For further discussion on the issue, please refer to Rey (2015).

Each row of $\boldsymbol{P}$ could be viewed as a multinomial distribution conditioned on the preceding state. For example, the second row of $\boldsymbol{P}$ represents the respective probabilities of transitioning to each of the m states at t given that an observation was in the second state at $t-1$. Since

---

[1] Refer to Shorrocks (1976), Formby et al. (2004), and Trede (1999) for a comprehensive survey of Markov-based mobility measures.

these multinomial distributions are conditionally independent, the maximum likelihood estimator for each individual transitional probability could be derived as shown in Equation (2.2) where n_{ij} is the number of transitions from state i to state j (Anderson and Goodman, 1957). Usually, a single transition probability matrix is estimated from the pooled income data across space and time. For the matrix to hold as the “ubiquitous” dynamic rule, several assumptions must be valid. Shorrocks (1976) presented three major assumptions:

1. First-order Markov: the income dynamic system has such a short memory that its current state is only influenced by the immediate past.
2. Population homogeneity: the same transition probabilities apply to all the regions being studied.
3. Time homogeneity: the transition probabilities remain constant over time.

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{q=1}^{m} n_{iq}}. \tag{2.2}$$

However, meticulous inspection of the above assumptions reveals its potential defect for applications in regional contexts. If there exists cross-sectional spatial dependence (Rey et al., 2016), which is very much likely, the assumption of random sampling that underlies the properties of the maximum likelihood estimators of the transition probabilities will be violated. As such the properties of these estimators and any mobility measure derived from them may be impaired.

2.2.2 Mobility Measures

A continuous real function $M(\cdot)$ is defined over the set of transition probability matrices to produce a real-value mobility measure. I concentrate on the following three mobility measures:

$$M_1(\boldsymbol{P}) = \frac{m - \sum_{i=1}^{m} p_{ii}}{m-1}, \tag{2.3}$$

$$M_2(\boldsymbol{P}) = 1 - |det(\boldsymbol{P})|, \tag{2.4}$$

$$M_3(\boldsymbol{P}) = 1 - |\lambda_2|. \tag{2.5}$$

where $det(\boldsymbol{P})$ is the determinant of $\boldsymbol{P}$ and λ_i represents the eigenvalue of $\boldsymbol{P}$ and $1 = |\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$. M_1 can be considered as the probability of leaving a class. As demonstrated in Prais (1955), the expected length of stay in class i is $\frac{1}{1-p_{ii}}$. Normalizing the reciprocal of the harmonic mean of the expected length of stay for every class by $\frac{n}{n-1}$ produces M_1 (Shorrocks, 1978). M_2 is the difference between 1 and the absolute value of the determinant of the transition probability matrix (Shorrocks, 1978). The final measure utilizes the absolute value of the second largest eigenvalue and deducts it from 1 (Sommers and Conlisk, 1979). Based on spectral theory, the largest eigenvalue of $\boldsymbol{P}$ is 1 ($\lambda_1 = 1$) and the remaining ones have absolute values less than 1. What is relevant here is that the absolute value of the second largest eigenvalue λ_2 determines the rate of the convergence of the Markov chain. That is, the smaller $|\lambda_2|$ is, the faster the chain converges. I will refer to these three mobility measures as Shorrocks mobility, determinant mobility, and eigenvalue mobility respectively in the rest of the chapter.

For any transition probability matrix with a quasi-maximal diagonal, all of the three mobility measures take values on $[0, 1]$[2]. 0 refers to immobility and 1 perfect mobility. Intuitively, if the transition probability matrix takes the form of the identity matrix, every region is stuck in its current state implying complete immobility. On the contrary, when each row of $\boldsymbol{P}$ is identical, current state is irrelevant to the probability of moving away to any class. Thus, the transition matrix with identical rows is considered perfect mobile. Although all three mobility measures have the same bounds, we should not expect that they are comparable to each other. As we shall see later, the mean and variance of these measures are rather different.

Another important property of mobility measures is monotonicity. Suppose that we increase one off-diagonal element at the expense of the diagonal element in the same row, we would expect the mobility measure to be able to pick up the change by raising its value. I will utilize this property in designing the Monte Carlo experiments.

2.2.3 Statistical Inference

2.2.3.1 Mobility Estimator

The natural estimators for the three mobility measures are $M_1(\hat{\boldsymbol{P}})$, $M_2(\hat{\boldsymbol{P}})$, and $M_3(\hat{\boldsymbol{P}})$ where $\hat{\boldsymbol{P}}$ is a maximum likelihood estimator whose elements are defined in Equation (2.2). Asymptotically, $\hat{\boldsymbol{P}}$ follows a multivariate normal distribution with the variance-covariance matrix $\Sigma_{\hat{\boldsymbol{P}}}$ defined in Equation (2.6). Here, n is the total number of transitions and $\hat{\pi}_i$ is the estimate of the probability of falling in state i.

[2]Please refer to Shorrocks (1978) regarding the definition of the transition probability matrix with a quasi-maximal diagonal.

$$cov(\hat{p}_{ij}, \hat{p}_{kl}) = \begin{cases} \frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n\hat{\pi}_i} & \text{if } i = k \text{ and } j = l, \\ -\frac{\hat{p}_{ij}\hat{p}_{il}}{n\hat{\pi}_i} & \text{if } i = k \text{ and } j \neq l, \\ 0 & \text{else,} \end{cases} \tag{2.6}$$

To derive the asymptotic variance for mobility measures, the delta-method could be utilized. Let $M(\hat{\boldsymbol{P}})$ represents any of the three measures.Then, the estimator of the asymptotic variance for $M(\hat{\boldsymbol{P}})$ is:

$$\sigma^2_{M(\hat{\boldsymbol{P}})} = D\Sigma_{\hat{\boldsymbol{P}}}D', \tag{2.7}$$

where D is the derivative of $M(\boldsymbol{P})$ with respect to $\boldsymbol{P}$ as shown in Equation (2.8) and D' is the transpose of D.

$$D = \frac{\partial M(\boldsymbol{P})}{\partial vec(\boldsymbol{P}')'}. \tag{2.8}$$

Here, vec converts a matrix into a column vector by stacking the columns on top of one another. For the three mobility measures studied in this chapter, the derivatives are obtained as follows (Trede, 1999):

$$D_{M_1} = -\frac{1}{m-1} vec(\boldsymbol{I})', \tag{2.9}$$

$$D_{M_2} = -sign(det(\boldsymbol{P}))vec(\tilde{\boldsymbol{P}}')', \tag{2.10}$$

$$D_{M_3} = -vec(\check{\boldsymbol{P}}'_{\lambda_2})', \tag{2.11}$$

where $\boldsymbol{I}$ is the $m \times m$ identity matrix, $\tilde{\boldsymbol{P}}$ is the cofactor matrix of $\boldsymbol{P}$, and $\check{\boldsymbol{P}}_{\lambda_2}$ is the derivative of the second absolute largest eigenvalue with respect to $\boldsymbol{P}$.

With these derivatives in hand, we are able to calculate the asymptotic variance of $M(\hat{\boldsymbol{P}})$. As shown in Trede (1999), the asymptotic sampling distribution of the estimator for each of the above three mobility measures follows a normal distribution with mean $M(\hat{\boldsymbol{P}})$ and variance $\sigma^2_{M(\hat{\boldsymbol{P}})}$. I am going to investigate how the contemporaneous spatial dependence across regional income time series impacts the properties of each of the three estimators.

2.2.3.2 One-sample test

It might be the case that researchers want to know whether the economic mobility of a regional system is equal to or lower/higer than a specific level. A one-sample test about the mobility measure could serve the purpose as shown in Equation (2.12):

$$z_1 = \frac{M - x}{\sigma_M}, \tag{2.12}$$

where z_1 is the test statistic, M is the observed mobility estimate (for Shorrocks mobility, determinant mobility, or eigenvalue mobility), x is a value between 0 and 1 representing the anticipated mobility level we want to test against, and σ_M is the analytical variance of M. Because M is asymptotically normally distributed, z_1 obeys the standard normal distribution asymptotically under the null hypothesis $H_0 : M = x$.

2.2.3.3 Two-Sample Test

For a mobility comparison of two income systems, such as US (System A) and China (System B), a two-sample test is required. Since it is known that the asymptotic sampling distribution of the estimator is a normal distribution, a two-sample z-test can be utilized to serve the purpose. The test statistic is defined in Equation (2.13).

$$z_2 = \frac{M^{(A)} - M^{(B)}}{\sqrt{\sigma^2_{M^{(A)}} + \sigma^2_{M^{(B)}}}}, \tag{2.13}$$

where $M^{(A)}$ and $M^{(B)}$ are mobility measures estimated from income dynamic systems A and B based on the same mobility function, such as M_1, M_2, or M_3. The null hypothesis is $H_0 : M^{(A)} = M^{(B)}$ while three alternatives can be specified as $H_a : M^{(A)} \neq M^{(B)}$, $H_{a1} : M^{(A)} > M^{(B)}$ and $H_{a2} : M^{(A)} < M^{(B)}$, leading to the two-tail test, upper-tail test, and lower-tail test. Under each null, the asymptotic sampling distribution of the test statistic is the standard normal distribution, that is, $z_2 \sim N(0, 1)$.

Various factors might impact the properties of this test statistic as it concerns two systems. Interaction between two income systems is one potential cause, though I am not going to investigate it in this chapter. I will always assume that the two systems being compared are independent of one another. Another factor concerns about the discretization strategy. Application of identical classification boundaries to the real income values of the two systems appears to be the natural way to proceed, but the possible unequal development status (such as China and US) will lead to an almost absolute rejection of the null. Normalizing the real incomes by the average and then using the quantile discretization strategy seems to be a better way to go. Here, the mobility comparison considered is more likely a relative mobility rather than an absolute one.

In addition to these two issues, contemporaneous spatial dependence across regional income time series in either system might impair the properties of the test statistic. I will investigate its impact via a series of Monte Carlo simulation experiments.

2.3 Monte Carlo Experiment

In this section, I introduce a series of Monte Carlo simulation experiments which are designed to examine the impact of contemporaneous spatial dependence between regional time series on the properties of mobility measure estimators and relevant test statistics. Here, the spatial dependence I consider is the so-called substantive spatial dependence rather than nuisance spatial dependence (Anselin, 1988). The former is part of the underlying process while the latter is not.

2.3.1 Data Generating Process

That all the three mobility measures are derived from the transition probability matrix $\boldsymbol{P}$ makes $\boldsymbol{P}$ the core of the data generating process (DGP). That is, a DGP generating time series mimicking the Markov chain governed by the transition matrix $\boldsymbol{P}$ needs to be proposed. The other significant factor to be incorporated in the DGP is the contemporaneous spatial dependence between time series. In the following sections, I first introduce a common approach to simulating a Markov chain given $\boldsymbol{P}$, followed by an extended approach to simulating a set of spatially dependent Markov chains given $\boldsymbol{P}$ and spatial dependence level ρ.

2.3.1.1 Generating a Markov Chain

The most common approach to producing a realization of a first-order Markov chain $\{X_0, X_1, \ldots, X_t\}, t > 0$ utilizes the continuous uniform distribution defined over the range $(0, 1)$. The cumulative distribution function (CDF) for the uniform distribution is a simple diagonal line $F(x) = x, x \in (0, 1)$. Starting with a simple two-state Markov chain with the transition probability matrix $\boldsymbol{P}_p$ defined in Equation (4.19), we need to transform $\boldsymbol{P}_p$ into a "cumulative probability matrix (CPM)" first. As mentioned before, each row of $\boldsymbol{P}_p$ is a multinomial distribution conditional on the preceding state. That is to say, if the region is in state 1 at t, then the probability of transitioning to state 1 and 2 at $t + 1$ are 0.7 and 0.3 respectively. Similarly, if the region is in state 2 at t, the probability of transitioning to state 1 and 2 at $t + 1$ are 0.5 and 0.5. To construct the CPM is to calculate cumulative probabilities for each row. Thus, the CPM for $\boldsymbol{P}_p$ would be $\boldsymbol{P}_c$ as shown in Equation (4.19).

$$\boldsymbol{P}_p = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix} \end{array} \quad \boldsymbol{P}_c = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.7 & 1.0 \\ 0.5 & 1.0 \end{bmatrix} \end{array} \tag{2.14}$$

Suppose we need to simulate a Markov chain with length $t > 3$ given the initial state $X_0 = 2$, t random numbers are generated from the continuous uniform distribution. Let's say they are $u = \{0.7, 0.2, 0.8, \ldots\}$. Because $X_0 = 2$, we pick the second row of $\boldsymbol{P}_c$ to determine the state at $t = 1$. As the cumulative probability of the random number 0.7 is 0.7, which is greater than the cumulative probability of the first state 0.5, and smaller than that of the second state 1.0, we assign 2 to the state at $t = 1$. The next two states would be determined in a similar fashion. In the end, we would end up with the simulated Markov chain $\{2, 2, 1, 2, \ldots\}$. With t large enough, the maximum likelihood estimation of the transition matrix would be very similar to the true matrix $\boldsymbol{P}_p$.

The **rule for determining the state of X_t** could be generalized as follows: compare the cumulative probability cp_t of the generated random number u_t and the cumulative probabilities of all m states conditional on X_{t-1}. That is to say, if $X_{t-1} = k, k \in \{1, 2, \ldots, m\}$, the kth row row of the CPM would be utilized. If $cp_t < CPM_{k1}$, assign 1 to X_t; if not, proceed to CPM_{k2}. If $cp_t < CPM_{k2}$, 2 is assigned to X_t; if not, proceed to the next state

CPM_{k3}. Since the cumulative probability of the last state is always 1, X_t should always be rightfully determined.

To summarize, the procedures of producing a T-long realization of a Markov chain given a initial state X_0 and a transition probability matrix $\boldsymbol{P}$ are:

1. Construct the CPM of $\boldsymbol{P}$.
2. Generate T random samples (Markov innovations) $\{u_1, u_2, \ldots u_T\}$ from the continuous uniform distribution. Set $j = 1$.
3. Use the above determination rule to find the state for X_j.
4. If $j < T$, repeat step (3); otherwise stop.

In the case of a collection of N regions, we can repeat this process N times to generate N independent Markov chains. If we collect the Markov innovations in the matrix $\boldsymbol{U}$ of size $N \times T$, we note that each pair of rows $i \neq j$ have pairwise 0 covariance $cov[\boldsymbol{U}_{i,.}, \boldsymbol{U}_{j,.}] = 0$. In other words, the innovation for region j in period t is independent of the innovation for region i in the same period.

2.3.1.2 Generating a Set of Spatially Dependent Markov Chains

In the regional setting, we are confronted with a number of time series each of which is the income trajectory of a specific region. Since common practice is to estimate one transition probability matrix $\boldsymbol{P}$ from the pooled dataset, the implicit assumption would be that $\boldsymbol{P}$ holds for every region. The complication here is that $\boldsymbol{P}$ would be a ubiquitous dynamic rule indeed, but the estimator (Equation 2.2) might be impaired if these time series are correlated to some degree. My interest lies in the influence of potential spatial dependence between time series. Thus, a DGP producing a set of spatially dependent time series each of which is governed by a common given transition probability matrix is required.

The approach is based on three steps:

1. Construct the CPM of $\boldsymbol{P}$.
2. Draw T samples from an N-dimensional joint normal distribution with a specified level of spatial dependence. Define this as a matrix $\boldsymbol{U}$ with size $N \times T$.
3. Derive N marginal univariate cumulative distribution functions based on which the cumulative probability of each element $i = 1, 2, \ldots, n$ in sample t ($t \in [1, T]$), u_{it}, could be obtained.
4. Apply the determination rule to the CPM of $\boldsymbol{P}$ and the cumulative probabilities from the previous step for selecting the next state in the Markov chain currently in state X_{it}.

For step (b), I employ the spatial lag model (SAR) to produce spatially dependent cross-sectional data:

$$\boldsymbol{U}_t = \rho \boldsymbol{W} \boldsymbol{U}_t + \boldsymbol{\epsilon}_t, \tag{2.15}$$

where $\boldsymbol{U}_t$ is a $(1, N)$ vector of random variates at time t, $\rho \in [0, 1)$ is the level of spatial dependence constant over time, $\boldsymbol{W}$ is the row-normalized spatial weight matrix indicating the interaction between regions, and $\boldsymbol{\epsilon}_t$ is a vector of random errors independently and identically distributed as a normal distribution $\epsilon_{ti} \sim N(\mu_\epsilon, \sigma_\epsilon^2), i \in \{1, 2, ..., N\}$ (N is the number of regions). Rewriting Equation (2.15) in reduced form, we acquire:

$$\boldsymbol{U}_t = (1 - \rho \boldsymbol{W})^{-1} \boldsymbol{\epsilon}_t. \tag{2.16}$$

Since $\boldsymbol{\epsilon}_t$ follows a multivariate normal distribution, $\boldsymbol{U}_t$ also follows a multivariate normal distribution with a variance-covariance matrix whose nondiagonal elements are not necessarily 0 when ρ is not equal to 0. More specifically,

$$\boldsymbol{U}_t \sim N(\mu_\epsilon, \sigma_\epsilon^2 (I - \rho \boldsymbol{W})^{-1} ((I - \rho \boldsymbol{W})^{-1})'). \tag{2.17}$$

I then convert the these series to the Markov States based on steps 3-4. Note that when $\rho = 0$ this approach collapses to the case of simulating N independent discrete Markov chains as in the previous section, since now the rows of the matrix $\boldsymbol{U}$ are pairwise independent. In contrast, when $\rho \neq 0$, the N rows of $\boldsymbol{U}$ are no longer independent and thus the N Markov chains are spatially correlated.

2.3.2 Simulation Design

A set of simulation experiments which are designed to examine the impact of contemporaneous spatial dependence on the sampling distribution of the (three) estimator(s), as well as the size and power of the (three) test statistic(s) are introduced in this section.

2.3.2.1 Monotone Markov Matrix

As illustrated in before, the DGP requires a specification of a transition probability matrix $\boldsymbol{P}$. I restrict the research to the so-called monotone Markov matrix, which is usually encountered in empirical economic analysis. A transition matrix is considered monotone if each row stochastically dominate the row above it (Conlisk, 1990). As a consequence, the probability of any region transitioning to better-off states would be higher next period if it is currently in state $i + 1$ than i. One important implication of the monotone transition matrix is given in Dardanoni (1995) as **Lemma 1**, which states that if two regions are faced with a common monotone transition probability matrix, the income distribution for region l would always stochastically dominate that for region h if the initial income distribution for region l

stochastically dominate that for region h though both regions would converge to a common steady state distribution in the long run. This echoes the neoclassical economic growth theory (Barro and Sala-i Martin, 2003) in the sense of all regions monotonically converging to a common steady state. A major difference to be noticed here is that the neoclassical economic growth theory describes the income trajectory in a more deterministic sense while the monotone Markov chain is a stochastic model. Thus, the monotone Markov chain leaves more space for intradistributional dynamics such as leapfrogging.

2.3.2.2 Experiments for Mobility Estimator and one-sample test

I adopted a 5×5 transition probability matrix $\boldsymbol{P5}$ which was estimated from the discretized (quantiles) relative US state income time series 1929-2010 for the DGP. It is obvious that $\boldsymbol{P5}$ is a monotone transition matrix:

$$\boldsymbol{P5} = \begin{bmatrix} 0.915 & 0.075 & 0.009 & 0.001 & 0.000 \\ 0.066 & 0.827 & 0.105 & 0.001 & 0.001 \\ 0.005 & 0.103 & 0.794 & 0.095 & 0.003 \\ 0.000 & 0.009 & 0.094 & 0.849 & 0.048 \\ 0.000 & 0.000 & 0.000 & 0.062 & 0.938 \end{bmatrix}. \tag{2.18}$$

In addition to the transition matrix, the DGP also requires the specification of sample size (N, T), a spatial weighting matrix $\boldsymbol{W}$, a level of spatial dependence ρ, initial states and the parameters $(\mu_\epsilon, \sigma_\epsilon^2)$ of the normal distribution for the error term. To investigate whether the asymptotic properties of the three estimators hold in small sample settings, I incorporated $N = 25, 169$ and $T = 50, 200$ in the simulation experiments. The spatial configuration was a $N^{\frac{1}{2}} \times N^{\frac{1}{2}}$ regular grid based on which a rook contiguity weight matrix is constructed and used in the DGP. I varied spatial dependence levels $\rho = 0, 0.2, 0.5, 0.7, 0.9, 0.98$ to investigate the pattern of impacts imposed by dependence and whether there was a threshold value above which the impact could not be readily ignored. The initial states were randomly assigned and $\mu_\epsilon = 0, \sigma_\epsilon^2 = 0.5$ throughout the experiments.

For each combination of parameters, I simulated the DGP $1,000$ times and built the empirical sampling distribution for each of the three mobility estimators. Since the "true" transition probability matrix $\boldsymbol{P5}$ is given, I could analytically derive the asymptotic sampling distribution under the circumstances of no spatial dependence. Comparing the empirical and analytical asymptotic distributions would shed light on the influence of contemporaneous spatial dependence in small and large sample settings.

2.3.2.3 Experiments for Two-Sample Test Statistic

To investigate the properties of the two-sample test statistic, I need to simulate two dynamic systems which requires two transition probability matrices $\boldsymbol{P}^{(A)}$ and $\boldsymbol{P}^{(B)}$. $\boldsymbol{P}^{(A)}$ serves as the dynamic rule for system A and $\boldsymbol{P}^{(B)}$ for system B. As the null hypothesis is that both systems share a common mobility value, I used the same transition matrix $\boldsymbol{P5}$ for both systems. That is, $\boldsymbol{P}^{(A)} = \boldsymbol{P}^{(B)} = \boldsymbol{P5}$.

To examine the power of the two-sample tests for three different alternatives H_a, H_{a1} and H_{a2}, I need to come up with another transition probability matrix which is different from the baseline matrix $\boldsymbol{P5}$. The intuitive approach is to adjust the elements of $\boldsymbol{P5}$ in a systematic way so that I have control over the direction and magnitude of the difference in terms of mobility.

As I have mentioned earlier, all of the three mobility measures have an important property, monotonicity. Dardanoni (1995) discussed a type of perturbation to a transition matrix called “diagonalising shift” which decreases mobility by shifting probability mass towards the main diagonal. Here, I slightly adjust the approach to make it more systematic and operable. Instead of shifting towards the main diagonal, I shift from it. In order to control the magnitude of the shifting, I shift a certain portion $\beta \in [0, 1)$ at a time. As shown in Equation (2.19), the shifted mass is proportionally assigned to the nondiagonal elements in each row. For example, if I am to investigate the power of the tests when the mobility difference between two systems is small, I can adopt a small portion $\beta = 0.01$ in the adjusted diagonalising shift method. Thus, the new transition probability matrix $\boldsymbol{P}5_{0.01}$ is acquired as shown in Equation (2.20). By assigning $\boldsymbol{P}^{(A)} = \boldsymbol{P}5$ and $\boldsymbol{P}^{(B)} = \boldsymbol{P}5_{0.01}$ in the DGP, I could simulate two regional income systems governed by two different transition probability matrices.

$$\begin{aligned} p_{ii}^{new} &= (1-\beta)p_{ii}, i \in \{1, \dots, m\}, \\ p_{ij}^{new} &= p_{ij}^{new} + \frac{\beta p_{ii}}{m-1}, j \in \{1, \dots, m\}, j \neq i. \end{aligned} \tag{2.19}$$

$$\boldsymbol{P}5_{0.01} = \begin{bmatrix} 0.906 & 0.077 & 0.011 & 0.003 & 0.002 \\ 0.068 & 0.819 & 0.107 & 0.003 & 0.003 \\ 0.007 & 0.105 & 0.786 & 0.097 & 0.005 \\ 0.002 & 0.011 & 0.096 & 0.841 & 0.050 \\ 0.002 & 0.002 & 0.002 & 0.064 & 0.929 \end{bmatrix}. \tag{2.20}$$

When $\beta = 0$, the new transition probability matrix would be the same as $\boldsymbol{P}5$. To examine the power of the two sample test, I also varied $\beta = 0.01, 0.03, 0.05$ to investigate the sensitivity of the tests to contemporaneous spatial dependence under different circumstances. The “true” mobility differences for varied β based on the three measures are shown in Table

Table 1. True Mobility Differences

		Difference $M^{(A)} - M^{(B)}$		
Mobility measure	$M^{(A)}$	$\beta = 0.01$	$\beta = 0.03$	$\beta = 0.05$
M_1	0.169	-0.011	-0.032	-0.054
M_2	0.540	-0.024	-0.068	-0.110
M_3	0.041	-0.011	-0.034	-0.057

1. The determinant mobility measure tends to give the largest difference. It is almost twice the difference obtained from the other two measures.

Besides the two transition matrices, $\boldsymbol{P}^{(A)}$ and $\boldsymbol{P}^{(B)}$, the other parameters needed for the DGP were the same as that used in the experiments for mobility estimators. That is, $N = 25, 169$, $T = 50, 200$, a rook contiguity weight matrix for regular lattice, $\rho^{(A)} = 0, 0.2, 0.5, 0.7, 0.9, 0.98$, $\rho^{(B)} = 0, 0.2, 0.5, 0.7, 0.9, 0.98$, $\mu_\epsilon = 0$ and $\sigma_\epsilon^2 = 0.5$. For each combination of parameters, I simulated from the DGP $2,000$ times ($1,000$ for $\boldsymbol{P}^{(A)}$ and 1000 for $\boldsymbol{P}^{(B)}$). For each set of simulated datasets, I calculated three test statistics, each for one type of mobility measures, and recorded rejection ratios at the $\alpha = 0.05$ significance level.

2.4 Results

2.4.1 Sampling Distributions of Mobility Estimators

Let us start with looking at the sampling distributions of three mobility estimators $M_1(\hat{\boldsymbol{P}})$, $M_2(\hat{\boldsymbol{P}})$ and $M_3(\hat{\boldsymbol{P}})$. As discussed earlier, when the regional time series are free of spatial dependence, the asymptotic analytical sampling distribution for each measure is a normal distribution with the mean and variance determined by the underlying dynamic rule (the transition probability matrix $\boldsymbol{P}$) and the sample size N, T. Since the "true" transition probability matrix is given, we could easily derive the analytical sampling distribution. By comparing it with the empirical sampling distribution constructed from 1000 simulated samples under various circumstances, we could observe the impact of contemporaneous spatial dependence as well as sample size.

For Shorrocks mobility estimator $M_1(\hat{\boldsymbol{P}})$, Figure 1 shows the asymptotic analytical and empirical sampling distributions. The red curve depicts the former, while the grey curves the latter. The darker the grey curve, the higher the level of spatial dependence. Each subplot represents a different sample size. The subplots in the upper row display the sampling distributions when $T = 50$, while those in the lower row $T = 200$. The subplots in the left column display the sampling distributions when $N = 25$, while those in the right column $N = 169$. Thus the upper-left subplot shows the case when sample size is fairly small $N = 25, T = 50$ and the lower-right one shows a large sample case $N = 169, T = 200$.

We can observe from the lower-right subplot that when $\rho = 0$ the empirical distribution fits quite well with the asymptotic analytical distribution. As ρ increases, it is still a normal distribution though the variance increases dramatically. The normality of the distribution to the presence of spatial autocorrelation has been validated by conducting several normality tests including Kolmogorov-Smirnov test, Shapiro-Wilk test (Shapiro and Wilk, 1965) and D'Agostino and Pearson's normality test (D'Agostino and Pearson, 1973), none of which rejects the null hypothesis of a normal distribution. When spatial dependence is very strong $\rho = 0.98$, it can reach 28 times the analytical variance. On the other hand, the mean doesn't seem to deviate from the analytical mean until $\rho = 0.98$.

Moving to the upper-left subplot where sample size is small, the pattern is a little different. Even when $\rho = 0$ the empirical distribution doesn't seem to fit well with the asymptotic analytical distribution [3]. It is a little more dispersed and slightly shifts to the

[3] Two conservative normality tests, Shapiro-Wilk test and D'Agostino and Pearson's normality test, reject the null, though Kolmogorov-Smirnov normality test fails to reject the null.

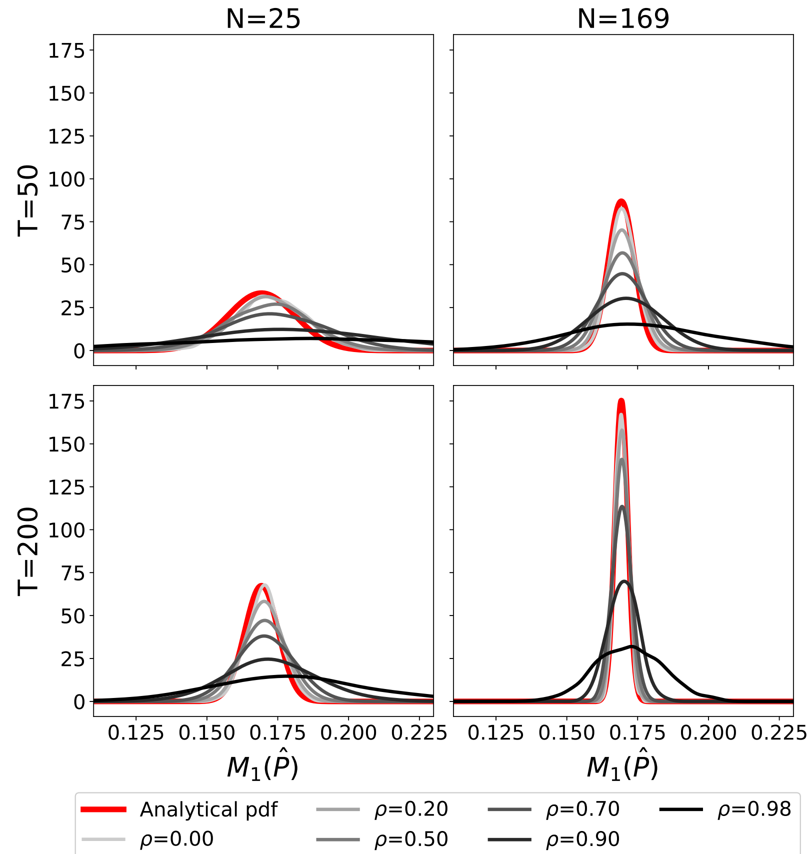

Figure 1. Asymptotic Analytical and Empirical Sampling Distributions of the Shorrocks Mobility Estimator $M_1(\hat{\boldsymbol{P}})$.

right of the latter. In other words, the asymptotic properties do not hold for small sample sizes: not only the variance is underestimated, but also the estimator is biased. Therefore the actual significance level would be larger than 0.05 leading to a higher Type I error rate even the regional economic system is exempt from spatial dependence. When there is spatial dependence between time series at work, both the variance and mean grow dramatically with ρ increasing.

Figures 2 and 3 show the asymptotic analytical and empirical sampling distributions for different sample sizes and under varied spatial dependence levels for mobility estimators $M_2(\hat{\boldsymbol{P}})$ and $M_3(\hat{\boldsymbol{P}})$. The general pattern is quite similar to Shorrocks mobility estimator

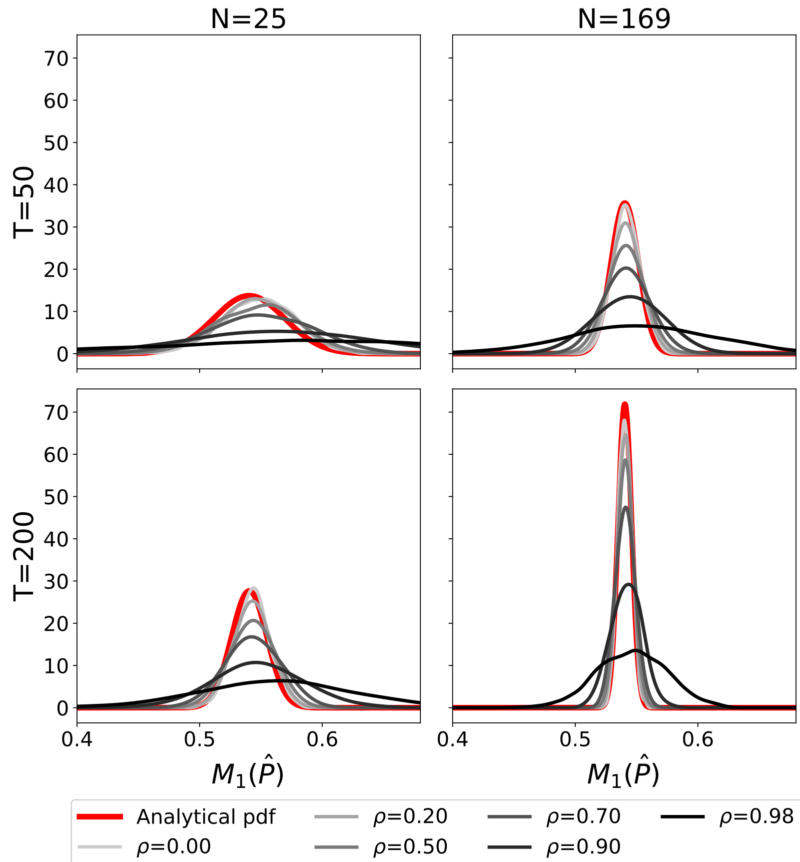

Figure 2. Asymptotic Analytical and Empirical Sampling Distributions of the Determinant Mobility Estimator $M_2(\hat{\boldsymbol{P}})$.

$M_1(\hat{\boldsymbol{P}})$. That is, as the spatial dependence strength becomes stronger, the empirical sampling distribution would still stay as a normal distribution though the variance grows dramatically and the mean grows mildly. In addition, the asymptotic properties do not seem to hold well in small sample settings, at least not when $N = 25, T = 50$.

The dramatic inflation of the variance makes sense. The contemporaneous spatial dependence existing in the regional income systems invalidates the $i.i.d$ assumption. The effective sample size for the transition probability estimator $\hat{\boldsymbol{P}}$ is less than $N \times T$. Thus the actual variance of each element of $\hat{\boldsymbol{P}}$ should be larger than what is indicated in Equation

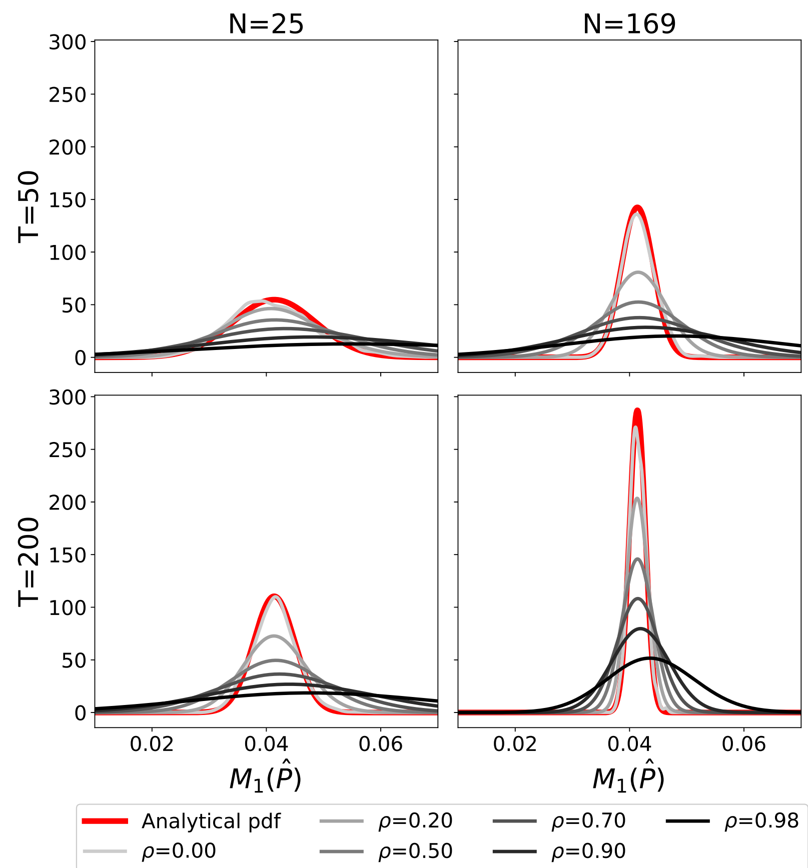

Figure 3. Asymptotic Analytical and Empirical Sampling Distributions of the Eigenvalue Mobility Estimator $M_3(\hat{\boldsymbol{P}})$.

(2.6). Since all of the three mobility estimators are derived from $\boldsymbol{P}$, their variances would also be inflated.

2.4.2 Properties of Two-Sample Test Statistics

Results regarding the properties of the two-sample test statistics for the three mobility measures are discussed in this section.

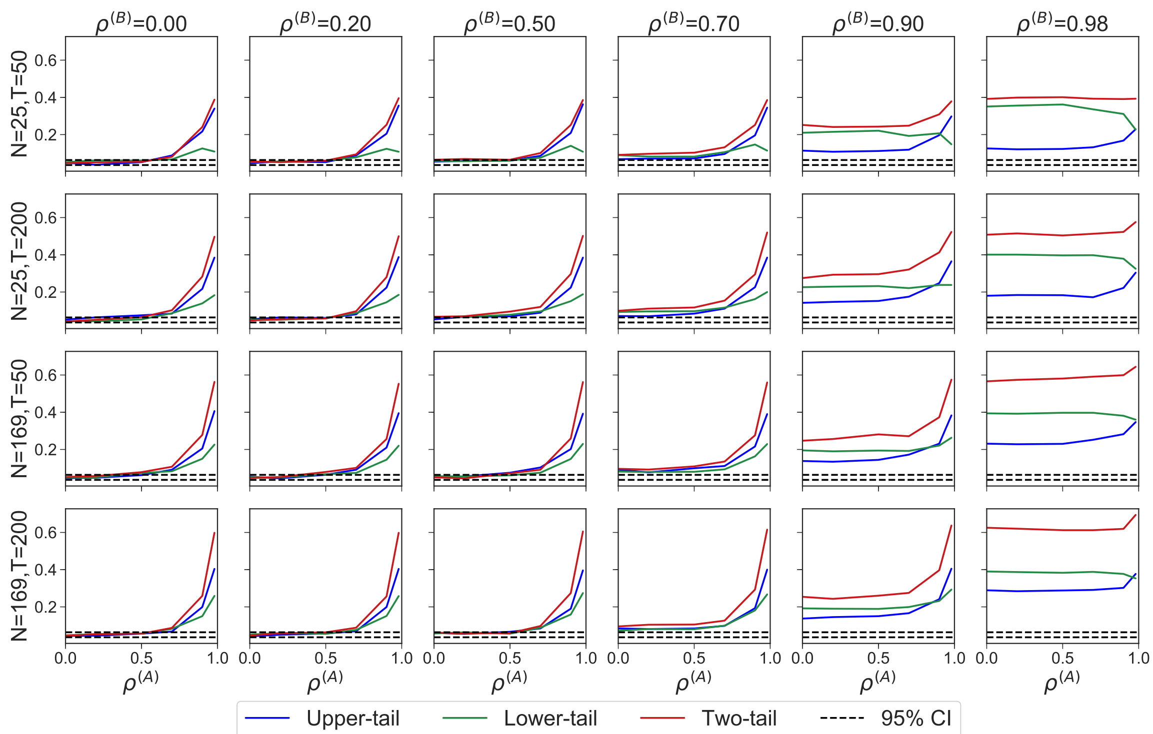

Figure 4. Size Properties of the Upper-, Lower- and Two-Tail Two-Sample Tests for $M_1(\boldsymbol{P})$.

2.4.2.1 *Size*

The general pattern for the impact of the contemporaneous spatial dependence on the size properties for three mobility measures are quite similar. Thus, we focus only on discussing the results for Shorrocks mobility measure.

The rejection ratios of the null under various circumstances for the two-sample test statistics for the Shorrocks mobility measure are displayed in Figure 4. The x-axis of each subplot is indexed by $\rho^{(A)}$, the level of contemporaneous spatial dependence in System A, and the y-axis indexes the rejection ratio of the null. The upper and lower bounds of the 95% confidence interval $(0.0365, 0.0635)$ are shown by two black horizontal dashed lines. The upper-, lower- and the two-tail test are symbolized in blue, green and red lines respectively.

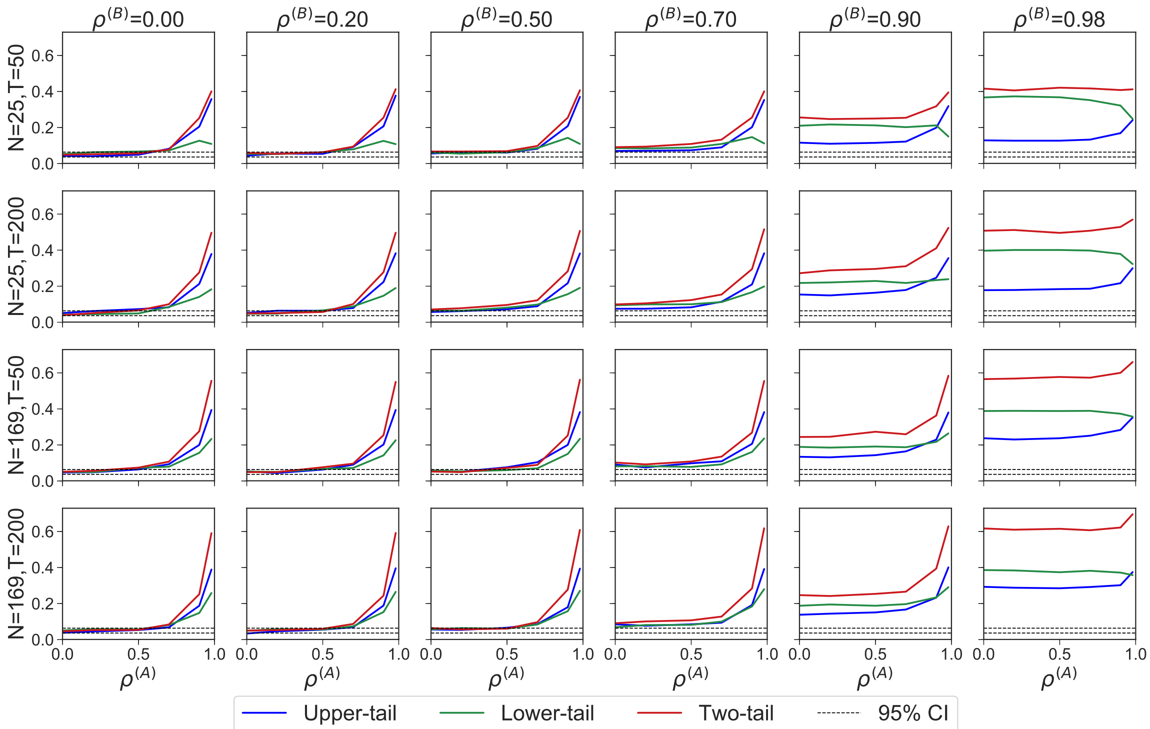

Figure 5. Size Properties of the Upper-, Lower- and Two-Tail Two-Sample Tests for $M_2(\boldsymbol{P})$.

$\rho^{(B)}$, the contemporaneous spatial dependence in System B becomes stronger from the left to the right subplot. From the top subplot to the bottom, the sample size increases. We can easily observe that relatively strong spatial dependence in either distribution (such as $\rho^{(A)} = 0.7$ or $\rho^{(B)} = 0.7$) has an significant influence on the size properties. It tends to make the size biased upward. As the level of spatial dependence in either system becomes higher, the upward bias tendency becomes stronger. It also seem to be the case that larger sample size is companied with more upward biased size. Comparing three different alternatives, the upper- and lower- tests seem to be more robust to spatial dependence than the two-tail test. This is especially true when $\rho^{(A)}$ or $\rho^{(B)}$ is quite large.

Figure 5 shows the impact of contemporaneous spatial dependence of varied levels on the size properties of the two-sample test statistics for the determinant mobility measure,

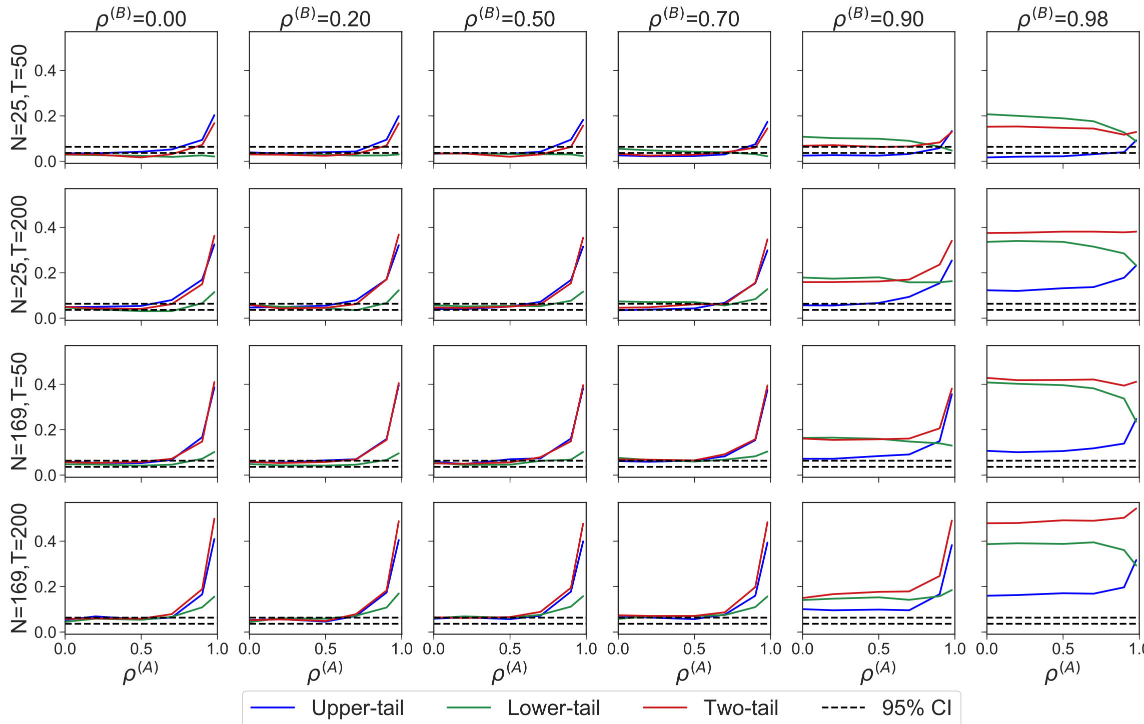

Figure 6. Size Properties of the Upper-, Lower- and Two-Tail Two-Sample Tests for $M_3(\boldsymbol{P})$.

while Figure 6 for the eigenvalue mobility measure. The patterns are rather similar to what we have observed for the Shorrocks mobility measure.

2.4.2.2 *Power*

Turning to the power properties of the test statistics, it turns out that they are also similar among three mobility measures. To save the space, I am only going to discuss results for the Shorrocks mobility measure in detail[4].

Figure 7 displays the rejection ratios when the mobility difference between two income systems is small ($\beta = 0.01$). Since the **true** mobility difference is negative, rejection ratios

[4]Results for determinant mobility measure and eigenvalue mobility measure are available upon request.

of the lower- and two-tail tests shed light on their power properties, while the ratios of the upper-tail test indicate its robustness as it is not supposed to pick up the negative difference. The power for the lower- and two-tail tests tends to grow with the sample size: for the lower-tail test, the rejection ratio increases from 0.146 all the way to 0.957 when both systems do not suffer from spatial dependence. The reason is that the variance for each of the mobility in the z test statistic decreases with the sample size N, T. Therefore, the denominator, which is the difference between the standard deviations for mobilities measured for two economic systems, decreases with the sample size. Thus facing the same mobility difference, the test with a larger set of observations tends to reject more. The general pattern for the impacts of spatial dependence also varies between small and large sample size. Looking at the first row where sample size is fairly small $N = 25, T = 50$, it seems that the power for the two-tail test increases with the spatial dependence level in either system, while the power for the lower-tail test increases with the spatial dependence level in income system B and decreases with the spatial dependence level in A. This is also true for some larger sample cases $N = 25, T = 200$ and $N = 169, T = 50$. However, when sample size is quite large as shown in the bottom row, the power decreases with stronger spatial dependence in either system. For the upper-tail test, the rejection ratios are always close to 0 except when spatial dependence is strong in either system and sample size is relatively small.

Increasing the difference between two transition probability matrices ($\beta = 0.03$) results in a stronger mobility difference of -0.068 for Shorrocks mobility measure. As shown in Figure 8, the power for the both of the lower- and two-tail tests mildly increases with the spatial dependence level in income system B and decreases with the spatial dependence level in A when sample size is very small $N = 25, T = 50$. For larger sample size, both tests have good power properties. They become less powerful in detecting the mobility difference when the spatial dependence is stronger in either system. However, as the sample size

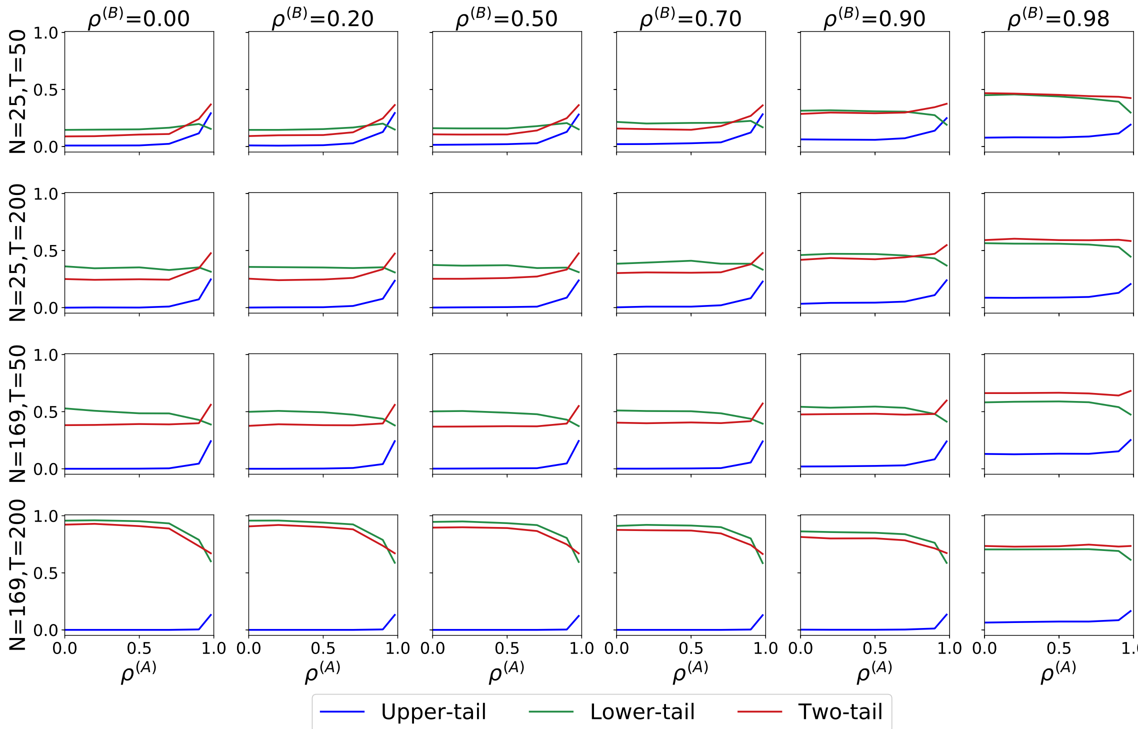

Figure 7. Power Properties of the Upper-, Lower- and Two-Tail Two-Sample Tests for M_1 ($\beta = 0.01$).

becomes larger, the decreasing trend is more and more negligible. Looking at the third row, it is clear that the power does not decrease until the dependence is very strong ($\rho^{(A)} = 0.9$ or $\rho^{(B)} = 0.9$).

Turning to the power properties of the tests when the mobility difference is much larger (-0.110), the patterns are more consistent as shown in Figure 9. Only when the sample size is quite small does the power decreases as the spatial dependence level in either system increases. This decreasing trend can be readily ignored when sample size is large: the power is quite close to 1 even when spatial dependence is strong. The impact of the spatial dependence is very similar for the other two mobility measures.

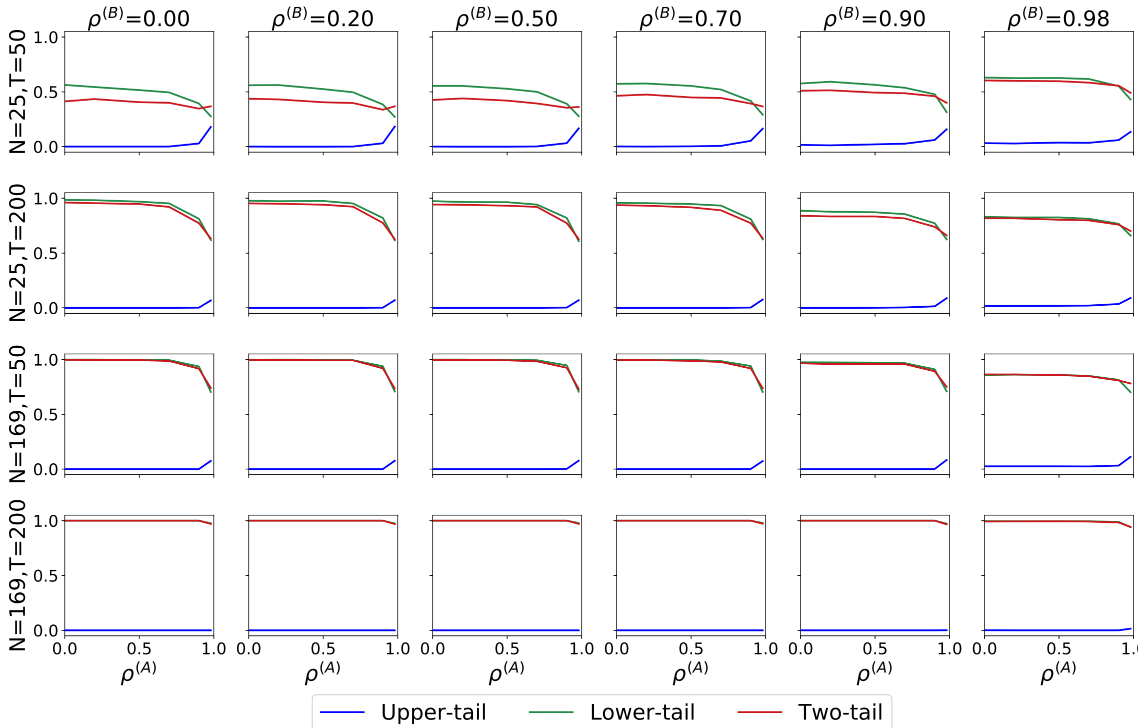

Figure 8. Power Properties of the Upper-, Lower- and Two-Tail Two-Sample Tests for M_1 ($\beta = 0.03$).

2.5 Adjusting Critical Values

As shown in the last section, contemporaneous spatial dependence inflates variances of sampling distributions of mobility estimators and raises the Type I error rates for both one-sample and two-sample tests. I resort to adjusting critical values to their "true" levels in order to maintain a proper size for the tests. Since I adopted Monte Carlo simulations to simulate the null where (1) mobility level equals a given level for the one-sample test, and (2) two regional system are equally mobile for the two-sample test, the empirical sampling distribution of estimates could be considered as the "true" sampling distribution to the presence of spatial autocorrelation of varying levels. Thus, the "true" critical values at

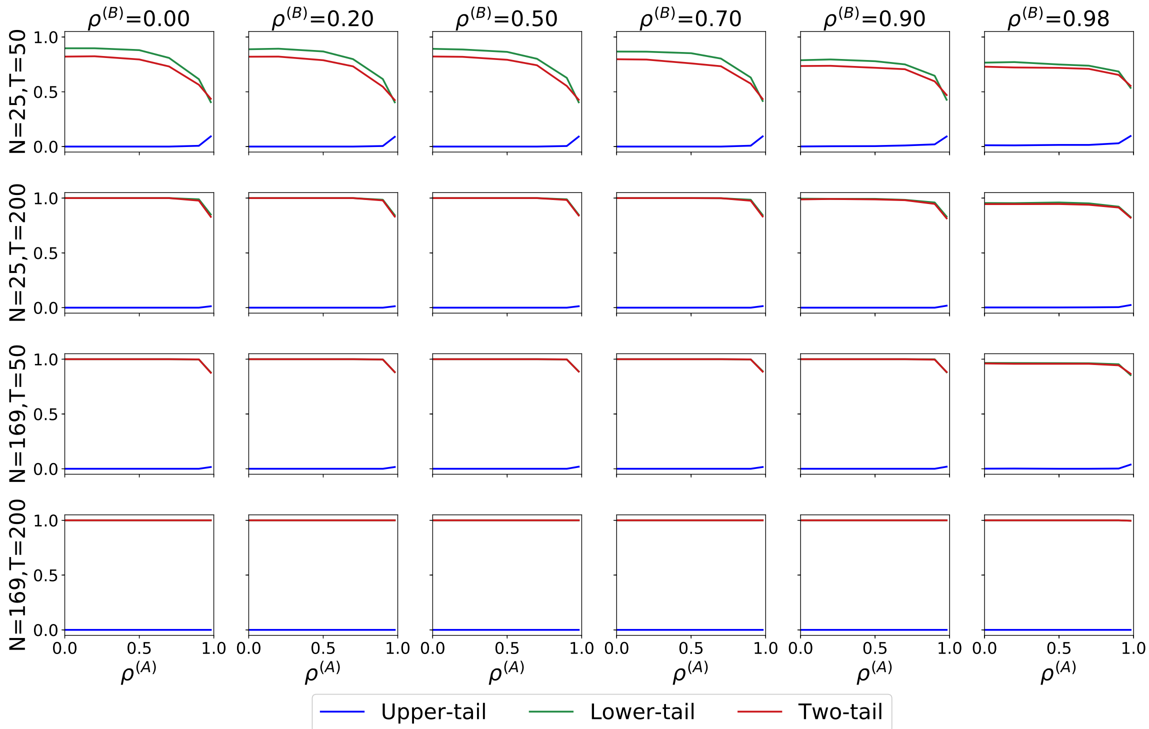

Figure 9. Power Properties of the Upper-, Lower- and Two-Tail Two-Sample Tests for M_1 ($\beta = 0.05$).

the 5% significance level for a two-sided test are the 25th and 975th of the ordered 1000 estimated test statistics.

2.5.1 One-sample test

For the one-sample test in Equation (2.12), assigning the “true" mobility level which is used as a simulation parameter (as shown in second column ($M^{(A)}$) of Table 1) to x would give z_1 estimates under the null. Therefore, the z_1 statistics estimated from $1,000$ realizations should follow the standard normal distribution $N(0,1)$. By testing those estimates against $N(0,1)$, we could know whether the empirical distribution deviates significantly from $N(0,1)$ and thus whether adjustments are needed.

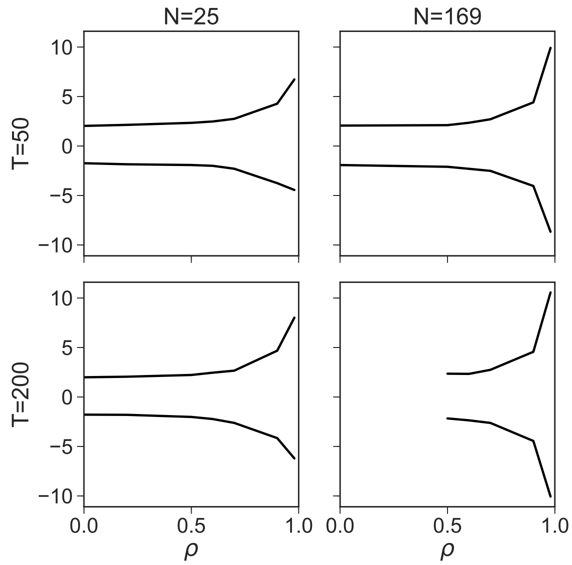

Figure 10. Empirical Critical Values of a One-Sample Two-Tail Test for $M_1(\boldsymbol{P})$.

Focusing on the Shorrock mobility measure, I plot the upper and lower empirical critical values for its one-sample test where testing for $N(0, 1)$ is rejected in Figure 10. Similar to before, each subplot represents a specific sample size and the x-axis indexes contemporaneous spatial autocorrelation level (ρ). From the plot, we could discern that adjustment is needed for all cases when sample size is small. On the opposite, for a large sample size as shown in the lower-right subplot, the critical values -1.96 and 1.96 obtained from $N(0, 1)$ could well serve the purpose for regional systems which are not highly spatially autocorrelated ($\rho < 0.5$). However, strong spatial autocorrelation inflates critical values more severely for larger sample sizes. Results for the other two mobility measures are similar and are available upon request.

2.5.2 Two-sample test

Turning to the two-sample test (Equation (2.13)), since the test statistic z_2 follows a standard normal distribution asymptotically, I adopt a similar approach. That is, I test for the standard normal distribution and obtain empirical critical values for cases where the tests are rejected. Those empirical critical values are visualized in Figure (11). The plots suggest that when both regional systems are strongly spatially autocorrelated, the critical values have to be increased for the comparison to be statistically valid. What's more, the inflation of critical values gets more severe with the increasing spatial autocorrelation level in either system. If both regional systems are weakly spatially autocorrelated, there is no need to make adjustment [5].

2.6 Conclusion

Regional income mobility measures are a useful complement to the inequality measures as they allow for a fuller understanding of regional income systems and their dynamics. However, the potential interactions between regions invalidate the $i.i.d$ assumption of classic statistical inference, posing a significant challenge to the statistical inference regarding mobility measures. This challenge is rather pertinent in the regional context as the notion of spatial dependence being a rule instead of an exception is widely acknowledged. This chapter takes up the challenge and explores the impacts of spatial dependence on the mobility inference via a series of Monte Carlo simulation experiments.

I focused on three Markov-based mobility measures, and found that the impacts from

[5] Results for the other two mobility measures are available upon request.

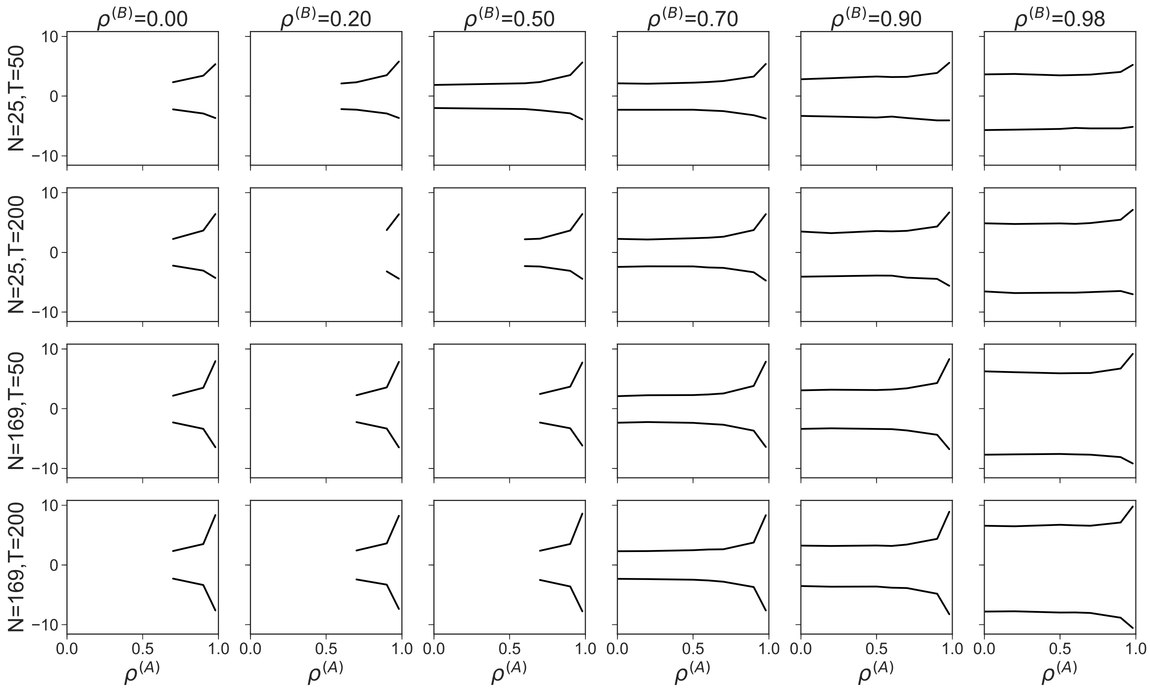

Figure 11. Empirical Critical Values of a Two-Sample Two-Tail Test for $M_1(\boldsymbol{P})$.

spatial dependence are rather similar. Dependence does have a major influence on the properties of the mobility estimators, one-sample and two-sample test statistics. It does not bias the mobility estimators when the spatial dependence is not extreme, but does dramatically increase the variances, leading to a inflated Type I error rate for a one-sample test. As for the two-sample test, the size tends to become more and more upward biased with increasing spatial dependence in either income system, which indicates that conclusions about differences in mobility between two different regional systems need to drawn with caution as the presence of spatial dependence can lead to false positives. The reason for the size distortion is due to the inflated variance of the test statistics. For the power properties, the impact has a mixed pattern in small sample settings, while when sample size is large the power decreases with stronger spatial dependence. Since the size is upward biased when there is spatial dependence in either income system, the power acquired based on the theoretical critical value would be inflated. Therefore, the actual power under the impact of spatial dependence is quite low.

Having found that spatial dependence impacts on the properties of mobility estimators and related tests, I attempted to account for the dependence by making adjustments to the critical values based on the results acquired from the Monte Carlo experiments. I have also tested the empirical distributions of the test statistics against its analytical asymptotic distribution - $N(0, 1)$ to differentiate cases where the impact of spatial autocorrelation is so trivial that an adjustment is not needed. It turns out that there is no need to make adjustment under the circumstance of a relatively large sample size and weak spatial dependence. Further research could be directed to the generalization of the adjustments to incorporate a wider range of cases. Empirical applications of the adjusted one-sample and two-sample tests are of great potential once a general formula is readily available.

Other approaches to accounting for spatial dependence could also be promising. Among them, parametric and nonparametric spatial filtering methods (Anselin, 1988; Getis and Griffith, 2002; Griffith and Chun, 2014) are tractable and commonly used. They treat the spatial dependence as nuisance and attempt to filter out spatially correlated components while leaving the independent components as the input for classic inference. We could also resort to the spatial bootstrap technique (Nordman et al., 2007; Cavaliere et al., 2015) which extends the conventional bootstrap to take account of dependence structure in the resampling process.

Chapter 3

CONDITIONAL AND JOINT TESTS FOR SPATIAL EFFECTS IN DISCRETE MARKOV CHAIN MODELS OF REGIONAL INCOME DISTRIBUTION DYNAMICS

3.1 Introduction

Discrete Markov Chain (DMC) models have been widely applied to the study of regional income distribution dynamics and convergence for the past 20 years (Quah, 1996a). A vast number of studies apply a first-order time-homogeneous DMC to the discretized per capita income data measured for a set of regional units and for a number of years with the implicit assumption that time series are pairwise independent and obey the same transitional dynamics rule (i.i.d. assumption). However, in the regional context, spatial effects including spatial heterogeneity and spatial dependence, if present, will invalidate the assumption. Ignoring space may give rise to misleading conclusions regarding transitional dynamics and convergence (Arbia et al., 2006).

Amongst the literature of regional income growth and convergence, the definition of spatial heterogeneity is similar to that in a cross-sectional context - underlying mechanisms are different across space due to differences in structural characteristics, giving rise to spatial regimes or spatial convergence clubs (Ertur et al., 2006). In the DMC framework, spatial heterogeneity means that different transitional dynamics rules hold across spatial regimes (Rey and Gutiérrez, 2015). Obviously, spatial heterogeneity invalidates the i.i.d. assumption.

In the spatiotemporal context, spatial dependence can take more complex forms than in a pure cross-sectional context. Two types of spatial dependence can be differentiated, contemporaneous spatial dependence and temporally lagged spatial dependence. The former is similar to the spatial dependence in a cross-sectional context in the sense of nonzero covariance between incomes of regions and their neighbors at the same time point (Rey et al., 2012). The latter only exists in the spatiotemporal setting as it describes the phenomenon that current income of a region is influenced by that of its neighbors at the preceding time point (Rey, 2001). Both of these two types of spatial dependence can be reasonably anticipated because of potential trade, migration, and technological spillovers among regions (Hammond, 2004; Le Gallo, 2004; Pu et al., 2005; Liao and Wei, 2012). Either type impairs the i.i.d. assumption.

In light of this, several test frameworks for spatial effects in DMC have been proposed. The Conditional Spatial Markov Chains (CSMC) test framework consists of two test statistics, a likelihood ratio test statistic and a χ^2 test statistic (Bickenbach and Bode, 2003; Anderson and Goodman, 1957). Each can be used to test for temporally lagged spatial dependence and spatial heterogeneity by specifying a particular form of conditioning. If temporally lagged spatial dependence is to be detected, neighbors' preceding income level serves as the conditioning. Similarly, to test for spatial heterogeneity, conditioning is formed through a spatial regime, which is a group of regions governed by the same transitional dynamics. The Joint Spatial Markov Chains (JSMC) test framework, consisting of χ^2 test statistic of independence, can be used to detect contemporaneous spatial dependence (Rey et al., 2012). Instead of conditioning on neighbors' preceding income levels, it is aimed at testing the (in)dependence of simultaneous spatial dynamics, that is, whether the transitional dynamics a region faces are independent of that faced by its neighboring regions during the same time interval.

All of these test statistics asymptotically obey χ^2 distributions with appropriate degrees of freedom. However, the small sample properties are of great significance for practice. Indeed, available regional income time series span 100 years at most and the number of regional units is not large (for example $N = 48$ if US contiguous states are to be considered). Rey et al. (2016) investigated this issue by simulating Vector Autoregressive (VAR) models with and without spatial effects, based on which the size, power and robustness properties of CSMC tests (likelihood ratio and χ^2 test statistics) were examined and evaluated. It turns out that all four test statistics (two for spatial dependence and two for spatial heterogeneity) display good size properties and have good power in terms of picking up the spatial effect they are designed for, though the robustness properties demonstrate mixed patterns.

This chapter extends Rey et al. (2016) in three aspects:

1. I investigate the performance of all tests in the presence of small sample size. Rey et al. (2016) considered only one temporal span $T = 100$. Here, I am interested in how a shortened temporal span impacts on the performance of the tests. Indeed, if temporal heterogeneity is detected, subsamples of shorter temporal spans should thus be considered, let alone that available regional income dataset itself might span a quite short period. Under which circumstances can asymptotics be considered to hold is an unsolved issue.
2. I evaluate the performance of JSMC test and compare it with CSMC test. Rey et al. (2016) considers the CSMC tests for temporally lagged spatial dependence. I differentiate them from another test that considers contemporaneous spatial dependence, namely the JSMC test. I am interested in how these two testing frameworks differ in terms of size, power and robustness properties. As the CSMC tests for temporally lagged spatial dependence are not robust to process mean heterogeneity, I examine whether the JSMC test can serve as a complement.
3. I examine implications of granularity of discretization of regional incomes. The application of DMC to regional income time series is not straightforward - continuous income data need to be discretized first. In most studies, global quintiles are used

as the cutoff to discretize income data into five classes. Thus, a 5×5 transition matrix is estimated to represent the rule for transitional dynamics. However, the Markov property possessed by original time series might not be preserved after inappropriate discretization (Rey, 2015; Guihenneuc-Jouyaux and Robert, 1998; Bulli, 2001; Magrini, 1999). Wolf and Rey (2015) finds that low-level granularity will lead to the loss of Markov property for 48 contiguous US states per capita incomes from 1920 to 2010. Specifically, if we discretize income series into a small number of classes, say less than 5, the time series acquired from quantile discretization will not be Markovian. On the other hand, though a better approximation to original continuous series can be acquired by raising granularity, it might compromise the estimator properties if the sample size is small, as it requires a larger set of parameters to be estimated. As I will demonstrate later, all test statistics require a much larger set of parameters to be estimated for the alternative. If sample size is not large enough, estimation would be problematic. How the level of discretization granularity impacts on the properties of tests is to be investigated.

The chapter is organized as follows. In Section 3.2, I introduce the CSMC and JSMC test statistics as well as the formation of alternatives. I present the design of a series of Monte Carlo experiments that are intended for examining and comparing properties of the tests in Section 3.3. I report the results in Section 4.3. Section 3.5 concludes with some key findings and directions for future research.

3.2 Tests for Spatial Effects in Discrete Markov Chain Models

All tests rely on the definition of a classic first-order time-homogeneous Discrete Markov Chains model (DMC) with the core of a $m \times m$ stationary transition probability matrix $\boldsymbol{P}$:

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}, 0 \leq p_{ij} \leq 1 \text{ and } \sum_{j=1}^{m} p_{ij} = 1, \forall i, j \in \mathbb{S}. \tag{3.1}$$

Here, regional income time series are classified into m states according to some discretization scheme, resulting in the state space $\mathbb{S} = \{1, 2, \cdots, m\}$. p_{ij} represents the probability of a region transitioning to state j at the immediate subsequent period given that it is currently in state i.

Two important assumptions are imposed on $\boldsymbol{P}$. The first is the first-order Markov property, which can be illustrated by Equation (3.2), that is, current state only impacts on its immediate future.

$$\boldsymbol{P}(x_t = j | x_{t-1} = i, \cdots, x_0 = l) = \boldsymbol{P}(x_t = j | x_{t-1} = i). \tag{3.2}$$

The second is time-homogeneous property, as shown in Equation (3.3), which implies that the transition rule $\boldsymbol{P}$ holds throughout the entire study time span T.

$$\boldsymbol{P}_0 = \boldsymbol{P}_1 = \boldsymbol{P}_2 = \cdots = \boldsymbol{P}_{T-1} = \boldsymbol{P}. \tag{3.3}$$

Assuming that N regional time series are independently and identically distributed - meaning that they share a common transitional dynamic rule, a maximum likelihood estimator for each element of $\boldsymbol{P}$ could be conveniently derived by considering all rows of the transition probability matrix as pairwise independent multinomial distributions:

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}}. \tag{3.4}$$

Here, n_{ij} is the total number of transitions from state i to state j across successive time points over the entire study time span for N regions.

3.2.1 Conditional Spatial Markov Chains (CSMC) Test Statistics

The CSMC test statistics are formed by comparing the transition probability matrix estimated from the whole sample with each of k matrices estimated from k exhaustive and mutually exclusive subsamples. Each of these k matrices is said to govern the dynamics of one subsample and can thus be estimated using the same estimator as Equation (3.4). Subsamples are obtained by dividing the entire sample based on spatial effects impacting the underlying transitional dynamics. Two CSMC test statistics are:

1. Likelihood ratio test statistic

$$LR^{(k)} = 2\sum_{l=1}^{k}\sum_{i=1}^{m}\sum_{j\in C_{i|j}} n_{ij|l} \ln \frac{\hat{p}_{ij|l}}{\hat{p}_{ij}} \sim \ asy\, \chi^2\left(\sum_{i=1}^{m}(c_i-1)(d_i-1)\right), \tag{3.5}$$

2. χ^2 test statistic

$$Q^{(k)} = 2\sum_{l=1}^{k}\sum_{i=1}^{m}\sum_{j\in C_{i|j}} n_{ij|l} \frac{(\hat{p}_{ij|l}-\hat{p}_{ij})^2}{\hat{p}_{ij}} \sim \ asy\, \chi^2\left(\sum_{i=1}^{m}(c_i-1)(d_i-1)\right), \tag{3.6}$$

where $\hat{p}_{ij|l}$ is the estimate of the probability transitioning from state i to state j across successive time points in the lth subsample, $C_{i|l} = \{j : \hat{p}_{ij|l} > 0\}$ is the set of nonzero probability estimates for the lth subsample transitioning from state i, $C_i = \{j : \hat{p}_{ij} > 0\}$ is the set of nonzero probability estimates transitioning from state i for the whole sample, $c_i = \#C_i$ is the number of elements in C_i and $d_i = \#D_i$ is the number of elements in $D_i = \{l : n_{i|l} > 0\}$ which is the set of nonzero transitions for the lth subsample transitioning from state i. Clearly, both tests attempt to deal with the potential sparsity of the estimated transition probabilities for the whole sample or the subsamples by ignoring the zero estimates and adjusting the degrees of freedom accordingly. Both of these test statistics are asymptotically distributed as χ^2 under the null hypothesis (Bickenbach and Bode, 2003):

$$p_{ij} = p_{ij|l} \,\forall\, l \in \{1, 2, \ldots, k\}. \tag{3.7}$$

As was mentioned earlier, CSMC tests can be used to detect spatial heterogeneity, as well as temporally lagged spatial dependence. Specifically, when spatial heterogeneity is present, that is, transitional dynamics are different across k spatial regimes, k subsamples should be acquired, each of which contains time series of regions belonging to a spatial regime. Similarly, when every region's current income level is impacted by local context at the immediate preceding time point, constituting the so-called temporally lagged spatial dependence, subsamples should be acquired by classifying each region's transition pairs $(t, t + 1)$ based on income level of local context at time point t. Here, the local context is formalized by the spatial lag z_t, which is the weighted average of neighbors' income level. The spatial lag of all regions at time point t can be calculated by:

$$\boldsymbol{Z}_t = \boldsymbol{W}_t \boldsymbol{Y}_t, \tag{3.8}$$

where $\boldsymbol{Y}_t$ is the regional income vector at time point t, and $\boldsymbol{W}_t$ is a row-normalized spatial weight matrix expressing the spatial interactions among regions at time point t.

Since the spatial lag is a continuous variable, we need to discretize it into k classes in defining the k subsamples. Usually, we am prone to adopt the same discretization strategy as regional time series, resulting in $k = m$. That is, m, $m \times m$ transition probability matrices, each of which is conditional on a specific income level of local context at the preceding time point, will be estimated. [6]

[6] For the rest of the chapter, I use CSMC Het LR and CSMC Het Q to represent CSMC likelihood ratio test

3.2.2 Joint Spatial Markov Chains (JSMC) Test Statistic

The JSMC test statistic is designed for detecting contemporaneous spatial dependence, that is, whether the transitional dynamics of each region and the transitional dynamics of its simultaneous local context are dependent (Rey et al., 2012). Similar to CSMC tests, local context is also formalized by spatial lag. In order to implement the test, I need to define two classic DMCs: own-chain O and neighbor-chain N. The former is the DMC for discretized regional time series while the latter is the DMC for discretized spatial lags of the regional time series. Following the same notation as before, regional time series are discretized into m classes and spatial lags are discretized into k classes. Thus, a $k \times k$ transition probability matrix $\boldsymbol{P}(O)$ for the own-chain is estimated and similarly a $m \times m$ matrix $\boldsymbol{P}(N)$ is estimated for the neighbor-chain .

The joint chain ON is a DMC on the extended state space $S_{ON} = \{(1,1),(1,2),\ldots,(k,m)\}$ where the first element in each tuple represents the state of a region and the second element represents its contemporaneous spatial lag state. The estimator for joint transition probability matrix $\boldsymbol{P}(ON)$ can be defined similar to equation (3.4).

Under the null hypothesis of independence between O and N, the following equation should hold:

$$\boldsymbol{P}(ON) = \boldsymbol{P}(O) \otimes \boldsymbol{P}(N), \tag{3.9}$$

where $\otimes$ is the Kronecker product operator.

JSMC χ^2 test statistic for contemporaneous spatial dependence is defined as

statistic and CSMC Q test statistic for spatial heterogeneity. Similarly, I use CSMC Dep LR and CSMC Dep Q to represent CSMC likelihood ratio test statistic and CSMC Q test statistic for temporally lagged spatial dependence.

$$
\begin{aligned}
\chi^2 &= \sum_{i=1}^{m}\sum_{j=1}^{k}\sum_{s=1}^{m}\sum_{q=1}^{k} n_{(i^o,j^n)} \frac{(\hat{p}_{(i^o,j^n),(s^o,q^n)} - \hat{p}^o_{is}\hat{p}^n_{jq})^2}{\hat{p}^o_{is}\hat{p}^n_{jq}} \\
&\sim asy\ \chi^2\ (k * m * (k * m - 1) - k * (k-1) - m * (m-1)),
\end{aligned} \tag{3.10}
$$

where superscript o and n index the region itself and its spatial lag respectively, $\hat{p}^o_{is}$ is the estimate of the probability of transitioning from class i to s for the own-chain, $\hat{p}^n_{jq}$ is the estimate of the probability of transitioning from class j to q for the neighbor-chain, and $\hat{p}_{(i^o,j^n),(s^o,q^n)}$ is the estimate of the joint probability of a region's income level transitioning from class i to s and its spatial lag from class j to q across the same successive time points.[7]

3.3 Design of the Experiments

The spatial layout for all the Monte Carlo simulation experiments consists of a regular $N^{1/2} \times N^{1/2}$ spatial lattice. In this section, I introduce the experimental design: the Data Generating Process (DGP) for the null and alternatives, and various factors which might influence the size, power and robustness properties of the five test statistics introduced in Section 3.2.

3.3.1 Data Generating Process (DGP)

A common scenario for the null hypothesis of all the five tests is one where N discretized regional time series are i.i.d and possess the first-order Markov property. Since a finite state Markov chain could well approximate both univariate and vector autoregressive (VAR)

[7] Similar to before, I use JSMC Dep to represent JSMC χ^2 test statistic for contemporaneous spatial dependence.

models (Tauchen, 1986; Silos, 2006), I can simulate one univariate autoregressive model N times to generate N independently and identically distributed regional time series which will turn into N Markov Chains obeying the same transition rule after discretization. Equation (3.11) is a univariate autoregressive process where region i's current income level $y_{i,t}$ only depends on its immediate preceding income level $y_{i,t-1}$. A generalization of a univariate autoregressive model to allow full interrelationship is a VAR model where every region's income level is determined by all the regions' past incomes as shown in Equation (3.12) where $\boldsymbol{Y}_t$ is the $n \times 1$ vector of regional incomes at time point t, $\boldsymbol{Y}_{t-1}$ is the vector of regional incomes at the immediate preceding time point $t-1$, $\boldsymbol{\nu}$ is a vector of constant terms, $\boldsymbol{M}$ is coefficient matrix and $\boldsymbol{\epsilon}_t \sim N(0, \sigma^2 I)$ is a temporally non-autocorrelated white noise error vector $E[\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_s] = 0 \; \forall s \neq t$ (Lütkepohl, 2005).

$$y_{i,t} = \nu + b y_{i,t-1} + \epsilon_t. \tag{3.11}$$

$$\boldsymbol{Y}_t = \boldsymbol{\nu} + \boldsymbol{M}\boldsymbol{Y}_{t-1} + \boldsymbol{\epsilon}_t. \tag{3.12}$$

A spatial extension of the VAR model is one where space is an important factor in determining parameters of Equation (3.12), either the coefficient matrix $\boldsymbol{M}$ or the constant terms $\boldsymbol{\nu}$. The spatial VAR model is shown in Equation (3.13) where both of $\boldsymbol{B}$ and $\boldsymbol{\Lambda}$ are diagonal matrices, and $\boldsymbol{W}$ is the same row-normalized spatial weight matrix as is defined in Equation (3.8) (I assume that it is constant over time) (LeSage and Krivelyova, 1999; LeSage and Cashell, 2015).

$$\boldsymbol{Y}_t = \boldsymbol{\nu} + \boldsymbol{B}\boldsymbol{Y}_{t-1} + \boldsymbol{\Lambda}\boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{\epsilon}_t. \tag{3.13}$$

Equation (3.13) indicates that each region's income at time point t is determined by its income as well as its neighbors' incomes at the preceding time point $t-1$. The strength of the impact of the former is indicated by the diagonal elements of $\boldsymbol{B}$. The diagonal elements of $\boldsymbol{\Lambda}$ represent the strength of the spatiotemporal diffusion, that is, the temporally lagged spatial dependence.

Rewriting (3.13):

$$\boldsymbol{Y}_t = \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{Y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{3.14}$$

where $\boldsymbol{A} = \boldsymbol{B} + \boldsymbol{\Lambda}\boldsymbol{W}$. The stability condition requires that all eigenvalues of $\boldsymbol{A}$ have modulus less than 1. A stable first-order VAR has a long run process mean vector $\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\nu}$.

By imposing restrictions on $\boldsymbol{\Lambda}$, $\boldsymbol{B}$ and $\boldsymbol{\mu}$, we can generate sets of times series with and without spatial heterogeneity or spatial dependence at work. Since the properties of these simulated times series are under complete control, I can evaluate the size, power and robustness properties of the 5 test statistics in a rigorous way.

It should be noted that the DGP (spatial VAR model) is a special case of a more general specification, the simultaneous dynamic space-time panel model (Elhorst, 2001; Debarsy et al., 2012; Parent and LeSage, 2012), which further allows for contemporaneous spatial dependence in addition to temporally lagged spatial dependence and spatial heterogeneity. The specification is shown in Equation (3.15) where $\boldsymbol{R}$ is a diagonal matrix. As the spatial VAR model is a natural extension of Tauchen's VAR model and corresponds to the logic of the CSMC dependence test, I restrict the attention to the spatial VAR model in this chapter. Future research could be directed to using dynamic space-time panel model as DGP.

$$\boldsymbol{Y}_t = \boldsymbol{\nu} + \boldsymbol{R}\boldsymbol{W}\boldsymbol{Y}_t + \boldsymbol{B}\boldsymbol{Y}_{t-1} + \boldsymbol{\Lambda}\boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{\epsilon}_t. \tag{3.15}$$

3.3.1.1 The Null

As has been discussed before, the null scenario is a first-order stable VAR process exempt from any spatial effect. To exclude spatial dependence, $\mathbf{\Lambda}$ should be a zero matrix. To exclude spatial heterogeneity, the DGP for every region should be the same. To be more specific, $\boldsymbol{B}_{hh} = \bar{\beta}, \boldsymbol{\mu}_h = \bar{\mu}, \forall\, h \in \{1, 2, \ldots, n\}$. In the experiments, I set $\bar{\beta} = 0.5, \bar{\mu} = 1, \sigma_t^2 = 0.5\, \forall\, t$.

3.3.1.2 Spatial Heterogeneity

Departing from the null, two forms of spatial heterogeneity can be introduced in the VAR: mean heterogeneity and lag heterogeneity. The former is reflected in k different long run process means $\boldsymbol{\mu}_h$ for regions belonging to corresponding k spatial regimes. The latter refers to varied temporal own-lag coefficients $\boldsymbol{B}_{hh}$ across k spatial regimes. In the experiments, n regions were equally and randomly assigned to k regimes to avoid introducing global spatial dependence. The spatial configuration might look like Figure 12 where 3 regimes are present.

Because it is pointed out by Bickenbach and Bode (2003) that CSMC Het Q is sensitive to the number of subsamples defined under the alternative, and over-rejecting the null hypothesis tends to be more severe as the number of subsamples increases for a given sample size, I designed three circumstances each having a different number of spatial regimes $k = 2, 3, 4$ to investigate this issue. The assignment of k different values to process means for mean heterogeneity and that to temporal own-lag coefficients for lag heterogeneity are given in Table 2 and Table 3 respectively. Take the second column where number of regimes is 3 in Table 2 as an example. Regions belonging to the first regime have the long run process

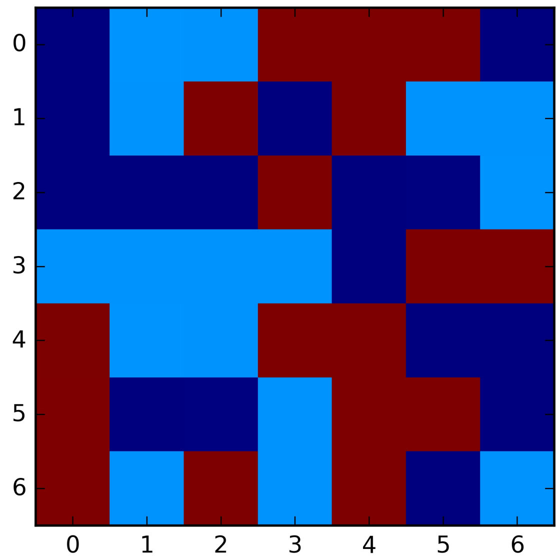

Figure 12. Spatial Configuration of 3 Regimes for $N = 49$.

Table 2. Process Means ($\boldsymbol{\mu}_h$) by Regime

	Number of Regimes		
Regime	2	3	4
1	0.50	0.25	0.25
2	1.50	1.00	1.00
3		3.00	2.00
4			3.00

Table 3. Temporal Own-Lag Coefficients ($\boldsymbol{B}_{hh}$) by Regime.

	Number of Regimes		
Regime	2	3	4
1	0.25	0.25	0.25
2	0.50	0.50	0.42
3		0.75	0.59
4			0.75

mean 0.25, that is, $\boldsymbol{\mu}_h = 0.25$ if $h \in$ Regime1. Similarly, $\boldsymbol{\mu}_h = 1.00$ if $h \in$ Regime2 and $\boldsymbol{\mu}_h = 3.00$ if $h \in$ Regime3.

For all the spatial heterogeneity simulations, spatial dependence was absent $\boldsymbol{\Lambda}_{hh} = 0 \, \forall \, h$.

As for mean heterogeneity, temporal own-lag coefficients were set to $\boldsymbol{B}_{hh} = \bar{\beta} = 0.50\,\forall\, h$. Similarly, for lag heterogeneity, long run process means were set to $\boldsymbol{\mu}_h = \bar{\mu} = 1\,\forall\, h$. I did so to examine the performance of the test statistics in the presence of either form of heterogeneity.

3.3.1.3 Spatial Dependence

Temporally lagged spatial dependence can be conveniently introduced in the VAR model by assigning nonzero value to the diagonal of $\mathbf{\Lambda}$. As is shown in Rey et al. (2016), its introduction will create contemporaneous cross-regional correlation, which could be detected by JSMC χ^2 test. In addition, the fact that $\mathbf{\Lambda}$ being nonzero matrix induces heteroscedasticity creates the possibility for CSMC Het tests to pick up something though the spatial regimes might not be properly defined.

In the spatial dependence simulations, I set $\boldsymbol{B}_{hh} = \bar{\beta} = 0.5$ and $\boldsymbol{\mu} = \bar{\mu} = 1$ to exclude both forms of spatial heterogeneity. The diagonal elements of $\mathbf{\Lambda}$ were set to be the same nonzero value $\mathbf{\Lambda}_{hh} = \bar{\lambda} \neq 0\,\forall\, h$. I also varied the level of spatial dependence $\bar{\lambda} = 0.125, 0.250, 0.375, 0.499$ to investigate how it impacted the properties of the tests.

3.3.2 Monte Carlo Experiments

3.3.2.1 Sample Size

The experiments cover small sample sizes, as well as large sample sizes: spatial grids $N = 25, 49, 81, 121, 169$, and temporal span $T = 50, 100, 150, 200$.

3.3.2.2 Discretization Granularity

As mentioned before, the level of discretization granularity might impact the properties of the tests. Here, I adopt the common discretization strategy, that is, global quantile classification. I varied the number of classes $m = 3, 5, 7$ to look at how the higher level of granularity differed from lower level regarding the properties of the tests. For all the 3 test statistics for spatial dependence, CSMC Dep LR, CSMC Dep LR and JSMC Dep, I applied the same discretization strategy and granularity to spatial lags as that to regional time series. In other words, if I use quintile cutoffs to discretize the simulated time series, 5, 5×5 conditional transition probability matrices will be estimated for CSMC Dep LR and CSMC Dep LR, while a $5^2 \times 5^2$ joint matrix will be estimated for JSMC Dep.

3.3.2.3 Monte Carlo Experiments Design

Using various parameter values in equation (3.13) I simulated the null, the alternative of mean heterogeneity, the alternative of lag heterogeneity and the alternative of spatial dependence. For each DGP, I simulated 1000 realizations, with each realization generating N time series of length T.

For each realization, I applied m global quintile classification. Then, I applied 5 test statistics CSMC Het LR, CSMC Het Q, CSMC Dep LR, CSMC Dep Q and JSMC Dep to each discretized realization. The decision of rejecting the null was based on a 5% significance level. Thus, rejection frequency was calculated by dividing the number of all the rejection cases by 1000 for each case. When the DPG is for the null, the rejection frequency would provide insights into size properties. When the DPG is for the alternative of mean heterogeneity or lag heterogeneity, it would provide insights into the power of

CSMC Het LR and CSMC Het Q, as well as the robustness of CSMC Dep LR, CSMC Dep Q and JSMC Dep to the corresponding spatial heterogeneity. Similarly, when the DPG is for the alternative of spatial dependence, it would reveal the power of CSMC Dep LR, CSMC Dep Q and JSMC Dep, as well as the robustness of CSMC Het LR and CSMC Het Q, to spatial dependence.

3.4 Empirical Results

In this section, I present the results of the Monte Carlo experiments. Rejection frequencies of 5 test statistics to the absence or presence of spatial effects are reported to shed light on the size, power and robustness properties.

3.4.1 Size Properties

I find that when sample size is relatively large, two types of CSMC test statistics, $LR^{(k)}$ and $Q^{(k)}$ have quite similar size properties, either used for testing spatial dependence or spatial heterogeneity, which points to the argument that they are asymptotically equivalent (Anderson and Goodman, 1957). Figure 13 displays the sampling distribution of all the 5 test statistics for the case of $N = 169, T = 200$ under the null. The level of discretization granularity was low $m = 3$, which means that 3, 3×3 conditional transition probability matrices were estimated for the construction of CSMC Dep LR and CSMC Dep Q, while a $3^2 \times 3^2$ joint matrix was estimated to construct JSMC Dep. Simulated time series were randomly divided into 3 spatial regimes for CSMC Het LR and CSMC Het Q, resulting in estimating 3, 3×3 transition probability matrices. The red line is the theoretical χ^2 distribution. The grey histogram is the empirical distribution of test statistics calculated

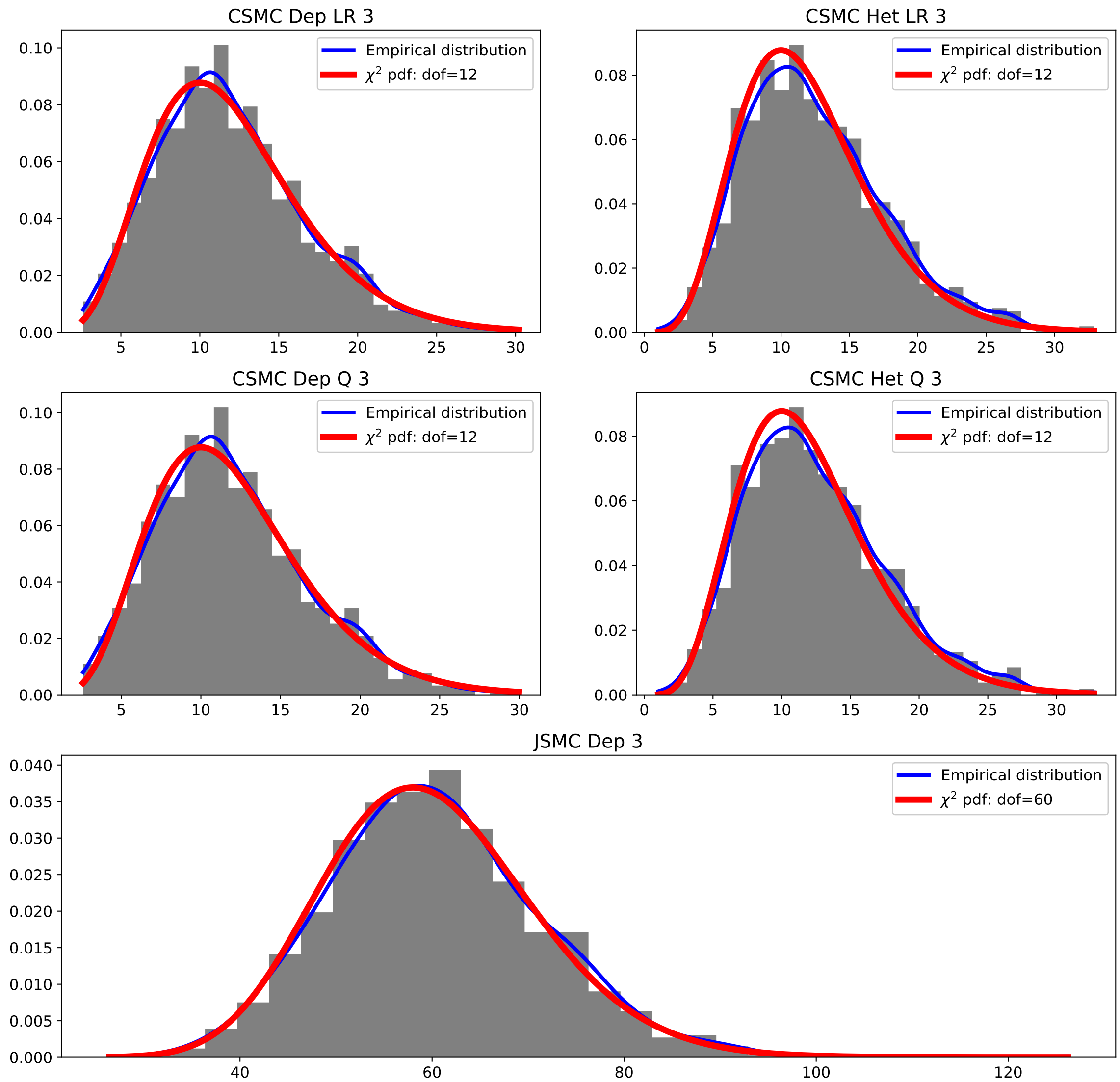

Figure 13. Sampling Distribution of 5 Test Statistics Under the Null, $N = 169, T = 200, m = 3$.

for 1000 simulated realizations under the null and blue line is the estimated nonparametric kernel. It is quite obvious that empirical and theoretical sampling distributions under the null are very similar.

When the sample size $N * T$ is small, and the level of discretization granularity m is high, CSMC Dep LR and CSMC Het LR are prone to upward bias while the other 3 still display

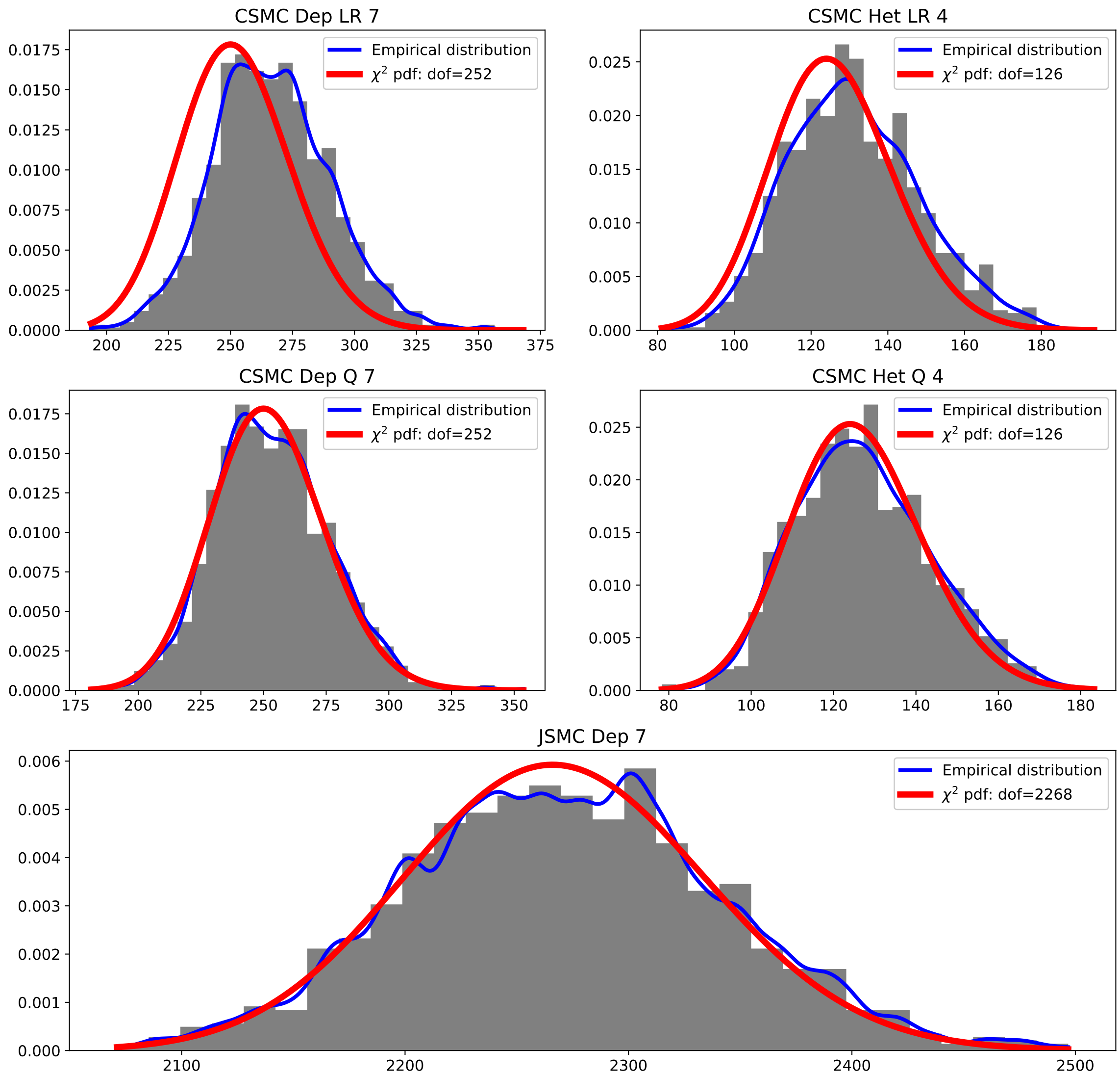

Figure 14. Sampling Distribution of 5 Test Statistics Under the Null,
$N = 25, T = 100, m = 7$.

good size properties. Figure 14 displays the sampling distributions under the null for the case of $N = 25, T = 100, m = 7$. Empirical distribution (blue line and grey histogram) shifts to the right of the theoretical distribution (red line) for CSMC Dep LR and CSMC Het LR, which will result in false rejections of the null.

In most cases, the empirical size falls within the 95% confidence interval $(0.0365, 0.0635)$

as shown in Table 4. This is especially true for CSMC Dep Q which displays good size property in almost all cases. However, CSMC Dep LR is very sensitive to the increased level of discretization granularity, especially when sample size is relatively small. Its rejection frequency could rise to as high as 0.315 when $m = 7$ and $N = 25, T = 50$. The rejection frequencies for the other 3 test statistics, JSMC Dep, CSMC Het LR and Het Q don't fall within the confidence interval as often as CSMC Dep Q, but never exceed 0.1.

3.4.2 Presence of Spatial Heterogeneity

I now examine the power of all 5 test statistics in terms of picking up mean or lag spatial heterogeneity when it is present.

I start with the presence of mean heterogeneity. In general, both of CSMC Het LR and Het Q display good power as rejection frequencies reach 1 for all cases (see Table 5). As for JSMC Dep, CSMC Dep LR and Q, they are not robust to mean heterogeneity since their rejection frequencies easily exceed 0.05.

Figure 15 visualizes parallel coordinates of rejection frequencies for all sample sizes using low level of discretization granularity $m = 3$. For all the 20 plots, temporal span increases from left to right, and spatial coverage increases from top to bottom. For each plot, 5 vertical lines show rejection frequencies for CSMC Dep LR, CSMC Dep Q, JSMC Dep, CSMC Het LR and CSMC Het Q from left to right. The blue, red and green lines represent 2, 3 and 4 regimes used for DGP respectively. Clearly, rejection frequencies for both of CSMC Het LR and CSMC Het Q for all cases are 1, indicating high power in detecting mean heterogeneity, the spatial effect they are designed for. For the 3 dependence test statistics, they are also able to pick up this spatial effect to some extent, though they are not expected to. The performance of CSMC Dep LR and CSMC Dep Q are quite similar. JSMC Dep always has a higher rejection frequency, indicating that it is more sensitive to the presence of mean heterogeneity. There is a clear trend that as temporal span increases, rejection frequencies of these 3 test statistics increase. However, this is not the case when spatial coverage increases where the pattern is more mixed. The number of spatial regimes doesn't impact the power of CSMC Het LR and CSMC Het Q, but as it increases, all 3 dependence tests seem to become more sensitive to mean heterogeneity.

Increasing the level of discretization granularity to $m = 7$ as is shown in Figure 16, rejection frequencies for JSMC Dep drop down sharply. Since they are almost always smaller than 0.1, except for when there are 2 regimes, I can consider it robust. The increase of granularity level also lowers the rejection frequencies of CSMC Dep LR and CSMC Dep Q, though not as much as that of JSMC Dep. Rejection frequencies for CSMC Het LR and CSMC Het Q are still as high as 1.

Turning to the presence of lag heterogeneity, the two heterogeneity test statistics, CSMC Het LR and Het Q, display good power except for small samples. CSMC Dep LR, CSMC Dep Q, JSMC Dep are robust to presence of this form of spatial heterogeneity. More details can be seen in Table 6.

Figure 17 shows parallel coordinates of rejection frequencies for all sample sizes using low level of discretization granularity $m = 3$. Rejection frequencies for CSMC Het LR and Het Q are always high except when $N = 25, T = 50$, while for CSMC Dep LR, CSMC Dep Q and JSMC Dep, most of them fall within the 95% confidence interval $(0.0365, 0.0635)$. Increasing the level of discretization granularity to $m = 7$ as is shown in Figure 18, CSMC Dep Q and CSMC Dep LR become sensitive when sample size is small, for example, when $N = 25, T = 50$ and $N = 100, T = 50$.

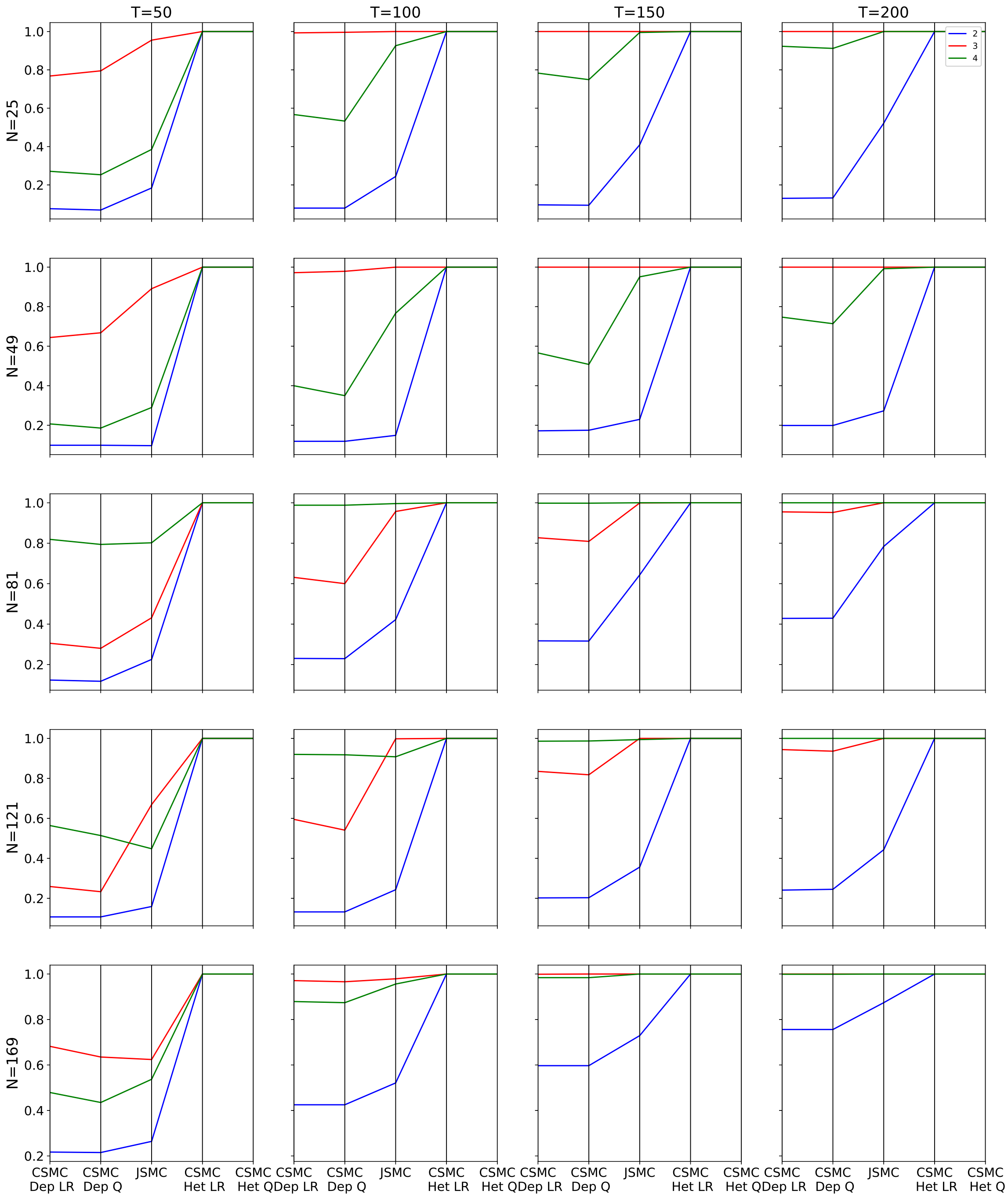

Figure 15. Parallel Coordinates of Rejection Frequencies of 5 Test Statistics for All Sample Sizes in the Presence of Mean Heterogeneity (2, 3 or 4 Spatial Regimes). The Level of Discretization Granularity is Low: $m = 3$.

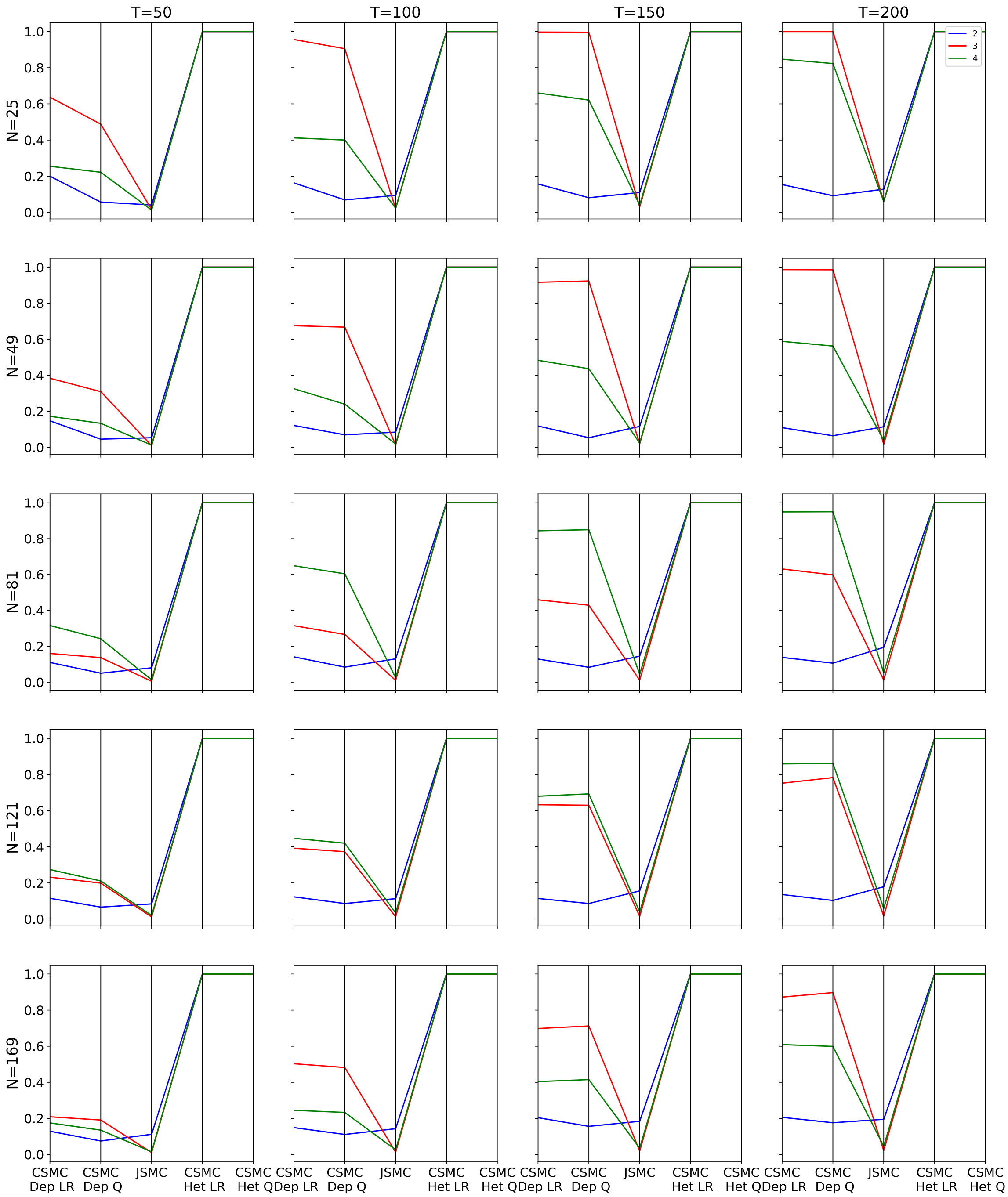

Figure 16. Parallel Coordinates of Rejection Frequencies of 5 Test Statistics for All Sample Sizes in the Presence of Mean Heterogeneity (2, 3 or 4 Spatial Regimes). The Level of Discretization Granularity is High: $m = 7$.

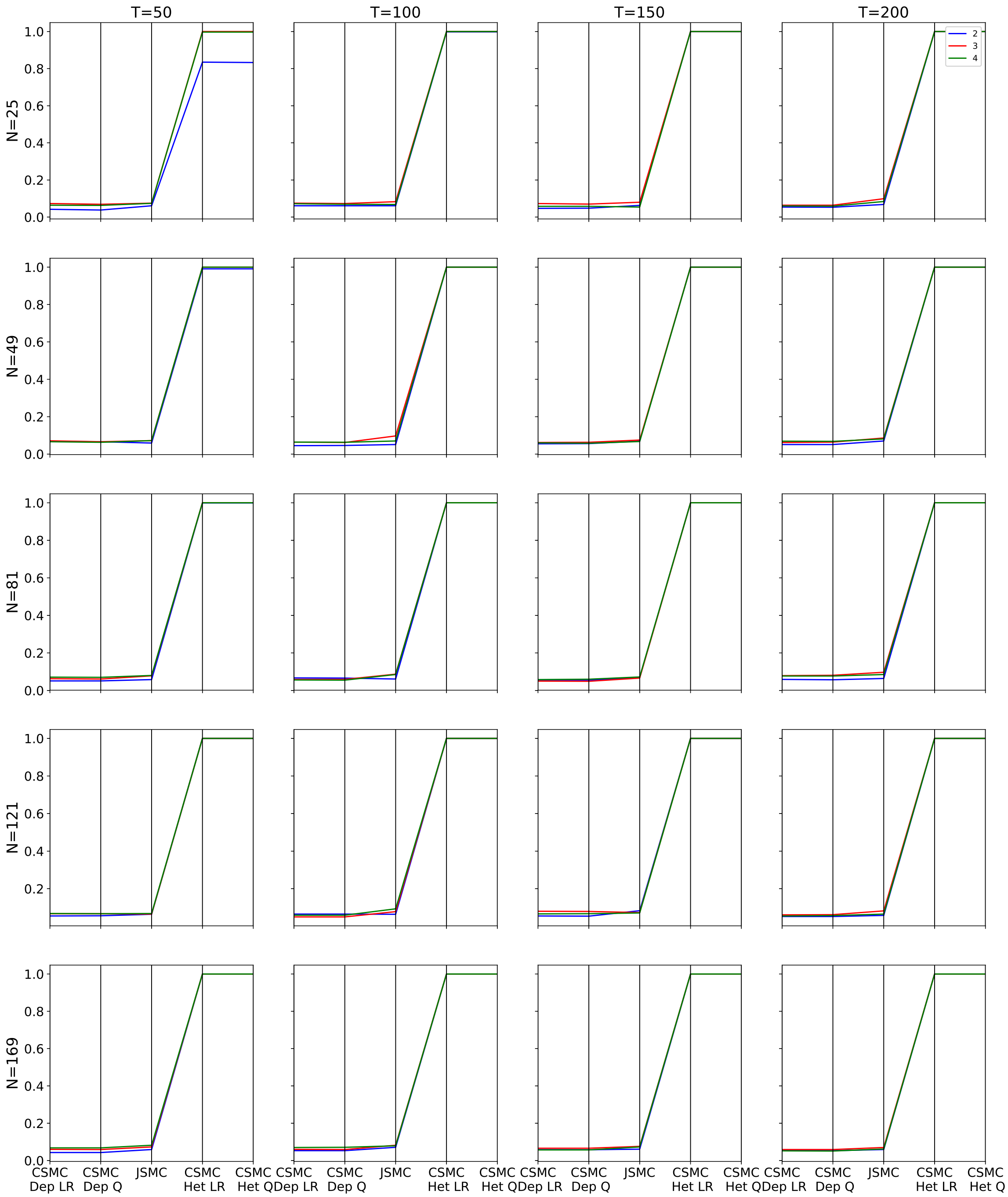

Figure 17. Parallel Coordinates of Rejection Frequencies of 5 Test Statistics for All Sample Sizes in the Presence of Lag Heterogeneity (2, 3 or 4 Spatial Regimes). The Level of Discretization Granularity is Low: $m = 3$.

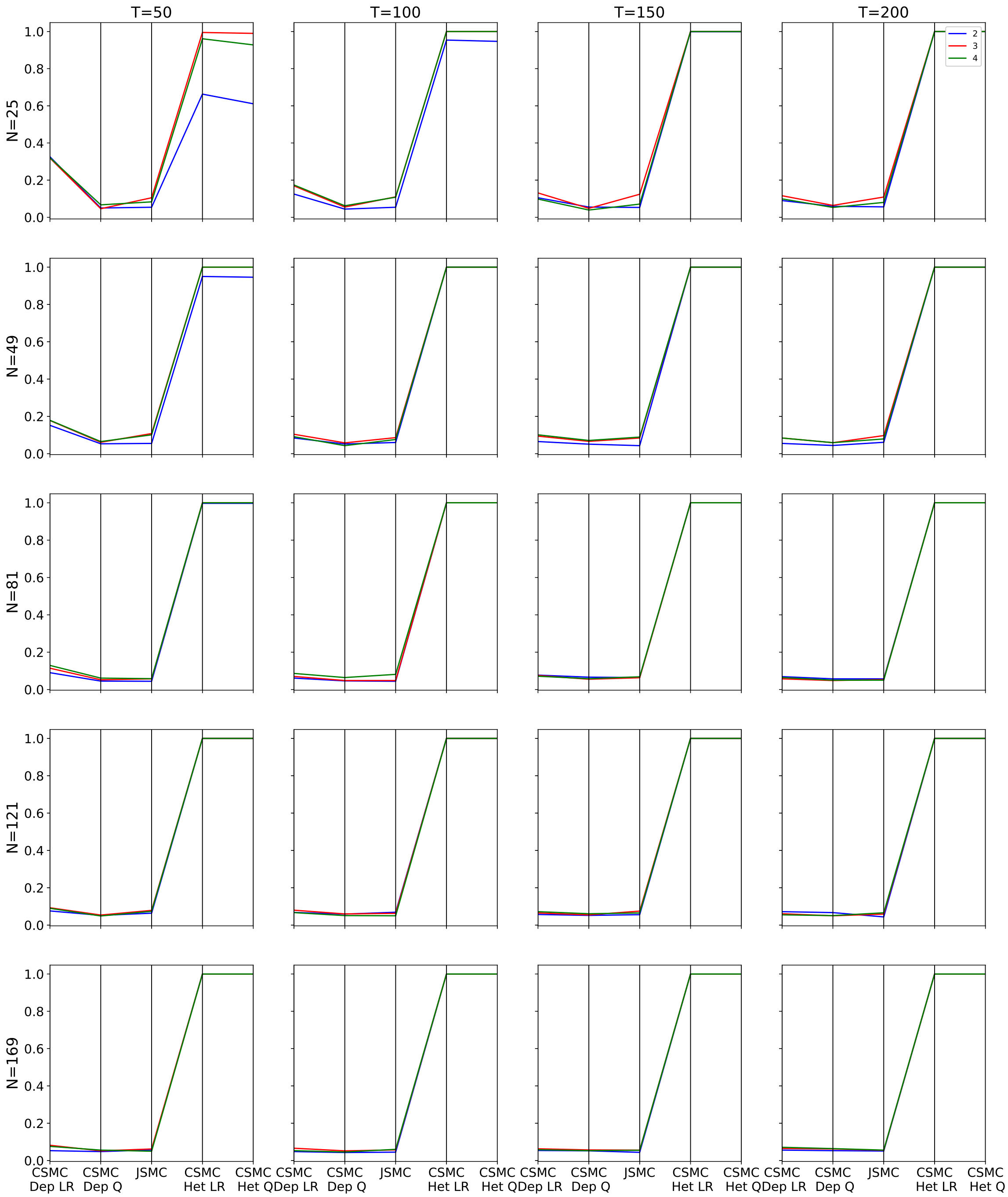

Figure 18. Parallel Coordinates of Rejection Frequencies of 5 Test Statistics for All Sample Sizes in the Presence of Lag Heterogeneity (2, 3 or 4 Spatial Regimes). The Level of Discretization Granularity is High: $m = 7$.

3.4.3 Presence of Spatial Dependence

The final alternative examined is spatial dependence. As I mentioned before, two forms of spatiotemporal dependence can be introduced, namely, temporally lagged spatial dependence and contemporaneous spatial dependence. CSMC and JSMC test statistics can be used to detect these two types respectively.

Figure 19 shows rejection frequencies of 5 test statistics for all sample sizes. Here, the level of discretization granularity is low $m = 3$. The x axis indexes $\bar{\lambda}$, the strength of spatiotemporal spillover. For all 5 test statistics, rejection frequencies have a positive relationship with $\bar{\lambda}$. CSMC Dep LR and CSMC Dep Q are very similar in terms of detecting spatial dependence, except for very small sample sizes and low levels of dependence. Increasing the sample size, through either spatial coverage or temporal spans, will raise the power. CSMC Het LR and CSMC Het Q can be considered robust to the presence of spatial dependence except when the dependence is strong.

The impact of discretization granularity level can be discerned by comparing Figure 19 and Figure 20, where the granularity level is raised to $m = 7$. For 3 dependence test statistics, the power of detecting spatial dependence is lower only when sample size is relatively small and dependence strength is week (for example $N = 81, T = 50, \bar{\lambda} = 0.25$). As for CSMC Het LR and CSMC Het Q, they become more sensitive only when dependence is very strong. CSMC LR test statistics for either spatial dependence or spatial heterogeneity have a inferior performance compared to CSMC Q test statistics. Full results are given in Table 7.

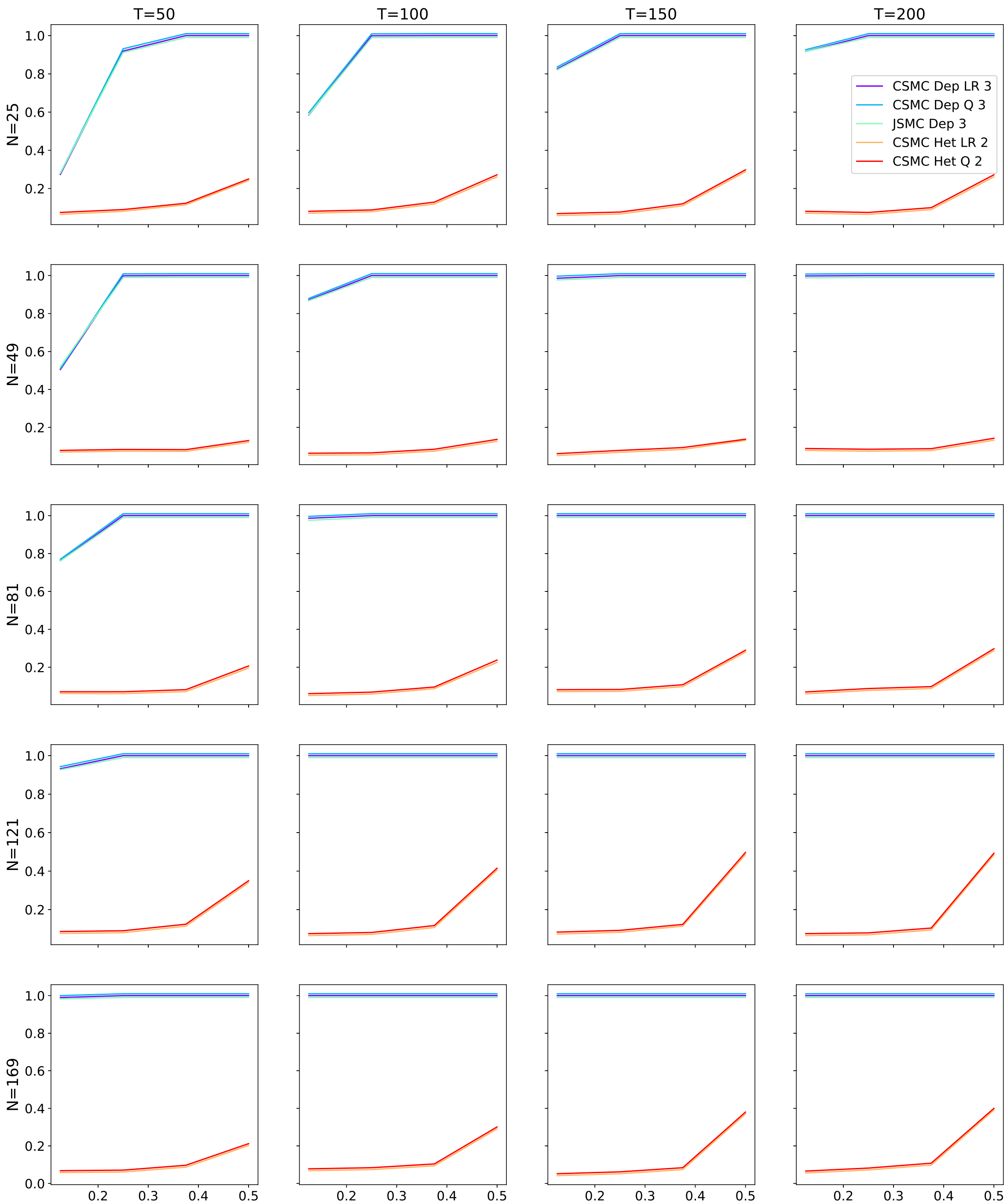

Figure 19. Rejection Frequencies of 5 Test Statistics for All Sample Sizes. The Level of Discretization Granularity is Low: $m = 3$. 2 Spatial Regimes are Constructed for the Alternative of CSMC Het LR and CSMC Het Q.

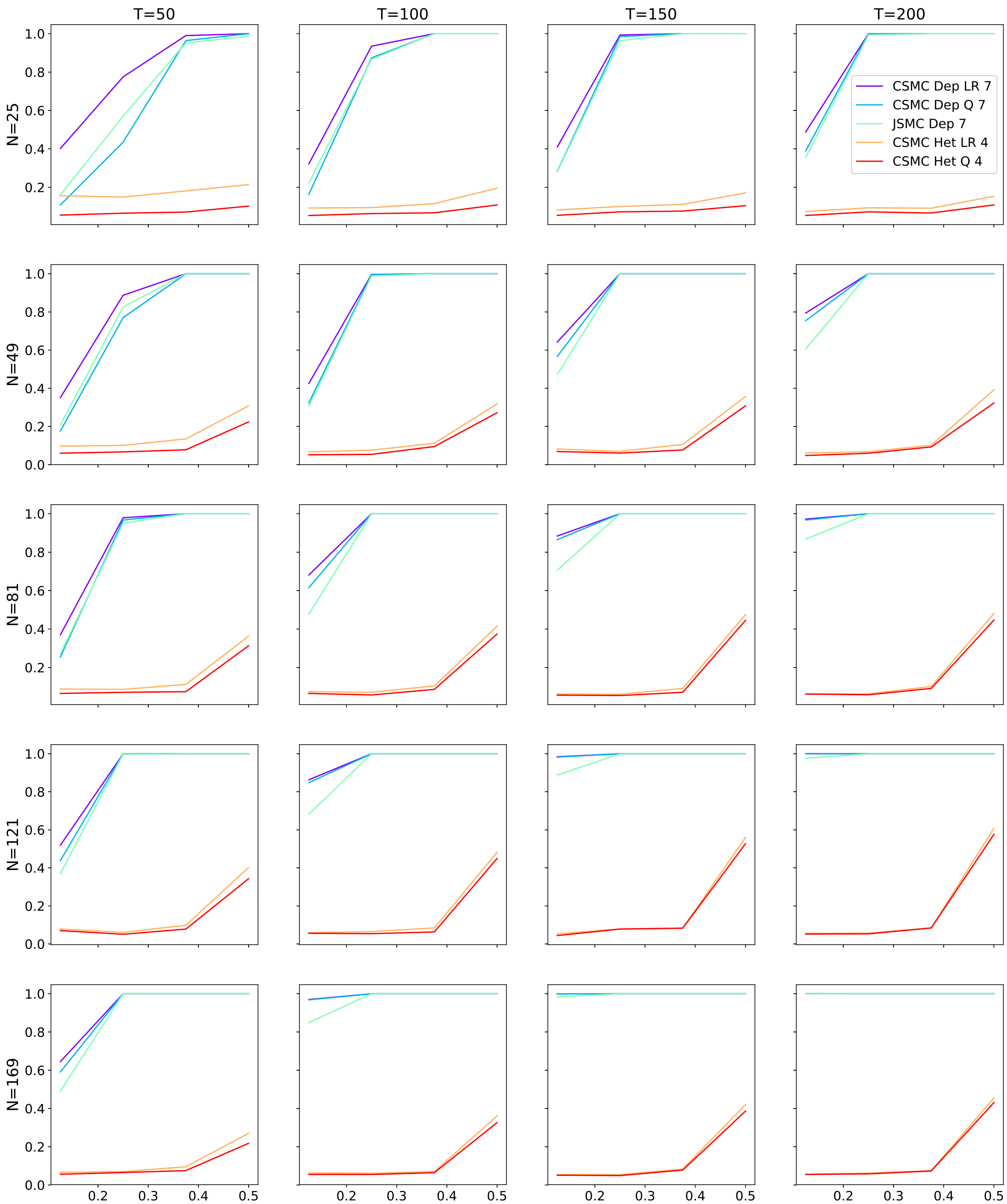

Figure 20. Rejection Frequencies of 5 Test Statistics for All Sample Sizes. The Level of Discretization Granularity is High: $m = 7$. 4 Spatial Regimes are Constructed for the Alternative of CSMC Het LR and CSMC Het Q.

3.5 Discussion and Conclusion

Application of DMC models to the study of regional income distribution dynamics and convergence are usually conducted through imposing some implicit assumptions, including spatial independence, spatial homogeneity and Markov property preservation after discretization. This chapter investigates the properties of 5 test statistics for spatial effects including 2 CSMC tests statistics for spatial heterogeneity, 2 CSMC tests statistics for temporally lagged spatial dependence and 1 JSMC tests statistic for contemporaneous spatial dependence under a number of circumstances. A vector autoregressive model is exploited to simulate time series under the null, as well as under 4 alternatives.

Results indicate that all of the 5 test statistics display good size properties. The exception is CSMC likelihood ratio test statistics. When sample size is fairly small, it tends to be biased upwards either it is used to test for temporally lagged spatial dependence or spatial heterogeneity. Thus, although it is asymptotically equivalent to CSMC χ^2 test statistic, its behavior is less satisfactory in small sample setting. In light of this, CSMC χ^2 test statistic is recommended when the available sample size is small.

All test statistics display strong power, but some of them are sensitive to the alternative form of spatial effect they are not designed for. Both CSMC Dep LR and CSMC Dep Q are not robust to the presence of mean heterogeneity, while JSMC Dep is robust if adopting a high level of discretization granularity. CSMC Het LR and CSMC Het Q are not robust to the presence of strong spatial dependence. The lack of robustness poses challenges for the application of the test statistics in empirical studies. Developing robust test (Anselin and Rey, 1991; Anselin, 1990) to aid these 5 test statistics is a promising research direction.

In addition to the non-robustness issue, since a VAR will always introduce contemporaneous spatial dependence if temporally lagged spatial dependence is specified, we could not discriminate one from the other in this setting, nor could we examine the sensitivity of JSMC (CSMC) test to the other form of spatial dependence. Future work may be focused on designing the data generating process which will only introduce one form of spatial dependence based on which an thorough investigation of the robustness of the other test could be conducted.

As increasing the level of discretization granularity lowers the sensitivity of almost all test statistics (except for CSMC heterogeneity tests when dependence is very strong) without compromising the power when sample size is large, it is recommended to adopt higher level to prevent picking up the "wrong" spatial effect in large sample setting. Otherwise, a balance should be made to preserve Markov property without impairing estimation precision. In other words, a relatively low granularity strategy should be considered to facilitate estimation, but caution should be taken in case the Markov property is lost due to discretization. Future research could be conducted in this realm to develop some procedures to select the best granularity level especially for the alternatives in small sample settings.

3.6 Appendix: Rejection Frequencies

Table 4. Size Results.

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	1.00	25	50	2	0.070	0.070	3	0.051	0.048	0.064
0.50	0	1.00	25	50	3	0.065	0.065	3	0.051	0.048	0.064
0.50	0	1.00	25	50	4	0.065	0.062	3	0.051	0.048	0.064
0.50	0	1.00	25	100	2	0.068	0.068	3	0.061	0.059	0.067
0.50	0	1.00	25	100	3	0.071	0.070	3	0.061	0.059	0.067
0.50	0	1.00	25	100	4	0.075	0.075	3	0.061	0.059	0.067
0.50	0	1.00	25	150	2	0.055	0.056	3	0.050	0.051	0.079
0.50	0	1.00	25	150	3	0.066	0.069	3	0.050	0.051	0.079
0.50	0	1.00	25	150	4	0.064	0.067	3	0.050	0.051	0.079
0.50	0	1.00	25	200	2	0.053	0.053	3	0.051	0.053	0.053
0.50	0	1.00	25	200	3	0.067	0.068	3	0.051	0.053	0.053
0.50	0	1.00	25	200	4	0.070	0.070	3	0.051	0.053	0.053
0.50	0	1.00	49	50	2	0.074	0.073	3	0.053	0.051	0.054
0.50	0	1.00	49	50	3	0.064	0.064	3	0.053	0.051	0.054
0.50	0	1.00	49	50	4	0.076	0.075	3	0.053	0.051	0.054
0.50	0	1.00	49	100	2	0.049	0.049	3	0.053	0.056	0.066
0.50	0	1.00	49	100	3	0.053	0.054	3	0.053	0.056	0.066
0.50	0	1.00	49	100	4	0.068	0.066	3	0.053	0.056	0.066
0.50	0	1.00	49	150	2	0.065	0.065	3	0.055	0.053	0.062
0.50	0	1.00	49	150	3	0.056	0.059	3	0.055	0.053	0.062
0.50	0	1.00	49	150	4	0.066	0.065	3	0.055	0.053	0.062
0.50	0	1.00	49	200	2	0.068	0.068	3	0.062	0.063	0.097
0.50	0	1.00	49	200	3	0.070	0.071	3	0.062	0.063	0.097
0.50	0	1.00	49	200	4	0.073	0.073	3	0.062	0.063	0.097
0.50	0	1.00	81	50	2	0.055	0.055	3	0.051	0.052	0.062
0.50	0	1.00	81	50	3	0.080	0.079	3	0.051	0.052	0.062
0.50	0	1.00	81	50	4	0.080	0.079	3	0.051	0.052	0.062
0.50	0	1.00	81	100	2	0.050	0.050	3	0.058	0.058	0.060
0.50	0	1.00	81	100	3	0.058	0.057	3	0.058	0.058	0.060
0.50	0	1.00	81	100	4	0.063	0.063	3	0.058	0.058	0.060
0.50	0	1.00	81	150	2	0.070	0.069	3	0.045	0.044	0.061
0.50	0	1.00	81	150	3	0.066	0.067	3	0.045	0.044	0.061
0.50	0	1.00	81	150	4	0.064	0.062	3	0.045	0.044	0.061
0.50	0	1.00	81	200	2	0.052	0.052	3	0.054	0.053	0.068
0.50	0	1.00	81	200	3	0.074	0.075	3	0.054	0.053	0.068
0.50	0	1.00	81	200	4	0.081	0.079	3	0.054	0.053	0.068
0.50	0	1.00	121	50	2	0.078	0.078	3	0.044	0.045	0.056
0.50	0	1.00	121	50	3	0.077	0.074	3	0.044	0.045	0.056
0.50	0	1.00	121	50	4	0.076	0.078	3	0.044	0.045	0.056
0.50	0	1.00	121	100	2	0.062	0.062	3	0.061	0.059	0.073
0.50	0	1.00	121	100	3	0.048	0.049	3	0.061	0.059	0.073
0.50	0	1.00	121	100	4	0.064	0.066	3	0.061	0.059	0.073

Continued on next page

Table 4 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	1.00	121	150	2	0.065	0.065	3	0.056	0.056	0.067
0.50	0	1.00	121	150	3	0.074	0.073	3	0.056	0.056	0.067
0.50	0	1.00	121	150	4	0.070	0.069	3	0.056	0.056	0.067
0.50	0	1.00	121	200	2	0.063	0.063	3	0.071	0.070	0.073
0.50	0	1.00	121	200	3	0.066	0.067	3	0.071	0.070	0.073
0.50	0	1.00	121	200	4	0.065	0.064	3	0.071	0.070	0.073
0.50	0	1.00	169	50	2	0.065	0.065	3	0.052	0.052	0.077
0.50	0	1.00	169	50	3	0.059	0.060	3	0.052	0.052	0.077
0.50	0	1.00	169	50	4	0.075	0.075	3	0.052	0.052	0.077
0.50	0	1.00	169	100	2	0.075	0.075	3	0.060	0.059	0.067
0.50	0	1.00	169	100	3	0.064	0.066	3	0.060	0.059	0.067
0.50	0	1.00	169	100	4	0.070	0.070	3	0.060	0.059	0.067
0.50	0	1.00	169	150	2	0.052	0.052	3	0.040	0.040	0.062
0.50	0	1.00	169	150	3	0.065	0.065	3	0.040	0.040	0.062
0.50	0	1.00	169	150	4	0.054	0.054	3	0.040	0.040	0.062
0.50	0	1.00	169	200	2	0.058	0.058	3	0.040	0.040	0.052
0.50	0	1.00	169	200	3	0.065	0.064	3	0.040	0.040	0.052
0.50	0	1.00	169	200	4	0.065	0.066	3	0.040	0.040	0.052
0.50	0	1.00	25	50	2	0.077	0.069	5	0.074	0.039	0.052
0.50	0	1.00	25	50	3	0.082	0.054	5	0.074	0.039	0.052
0.50	0	1.00	25	50	4	0.096	0.067	5	0.074	0.039	0.052
0.50	0	1.00	25	100	2	0.067	0.062	5	0.067	0.047	0.046
0.50	0	1.00	25	100	3	0.065	0.062	5	0.067	0.047	0.046
0.50	0	1.00	25	100	4	0.076	0.063	5	0.067	0.047	0.046
0.50	0	1.00	25	150	2	0.065	0.063	5	0.057	0.051	0.061
0.50	0	1.00	25	150	3	0.059	0.056	5	0.057	0.051	0.061
0.50	0	1.00	25	150	4	0.066	0.059	5	0.057	0.051	0.061
0.50	0	1.00	25	200	2	0.065	0.061	5	0.059	0.048	0.061
0.50	0	1.00	25	200	3	0.064	0.061	5	0.059	0.048	0.061
0.50	0	1.00	25	200	4	0.065	0.062	5	0.059	0.048	0.061
0.50	0	1.00	49	50	2	0.061	0.059	5	0.065	0.050	0.057
0.50	0	1.00	49	50	3	0.060	0.054	5	0.065	0.050	0.057
0.50	0	1.00	49	50	4	0.062	0.054	5	0.065	0.050	0.057
0.50	0	1.00	49	100	2	0.059	0.054	5	0.051	0.041	0.052
0.50	0	1.00	49	100	3	0.061	0.054	5	0.051	0.041	0.052
0.50	0	1.00	49	100	4	0.065	0.058	5	0.051	0.041	0.052
0.50	0	1.00	49	150	2	0.057	0.057	5	0.051	0.042	0.048
0.50	0	1.00	49	150	3	0.063	0.062	5	0.051	0.042	0.048
0.50	0	1.00	49	150	4	0.047	0.042	5	0.051	0.042	0.048
0.50	0	1.00	49	200	2	0.047	0.046	5	0.044	0.042	0.047
0.50	0	1.00	49	200	3	0.057	0.056	5	0.044	0.042	0.047
0.50	0	1.00	49	200	4	0.059	0.056	5	0.044	0.042	0.047
0.50	0	1.00	81	50	2	0.050	0.050	5	0.069	0.055	0.046
0.50	0	1.00	81	50	3	0.065	0.063	5	0.069	0.055	0.046
0.50	0	1.00	81	50	4	0.072	0.066	5	0.069	0.055	0.046
0.50	0	1.00	81	100	2	0.046	0.044	5	0.049	0.045	0.057
0.50	0	1.00	81	100	3	0.048	0.046	5	0.049	0.045	0.057

Continued on next page

Table 4 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	1.00	81	100	4	0.046	0.047	5	0.049	0.045	0.057
0.50	0	1.00	81	150	2	0.054	0.054	5	0.062	0.055	0.051
0.50	0	1.00	81	150	3	0.061	0.059	5	0.062	0.055	0.051
0.50	0	1.00	81	150	4	0.051	0.051	5	0.062	0.055	0.051
0.50	0	1.00	81	200	2	0.061	0.061	5	0.041	0.039	0.048
0.50	0	1.00	81	200	3	0.054	0.055	5	0.041	0.039	0.048
0.50	0	1.00	81	200	4	0.056	0.057	5	0.041	0.039	0.048
0.50	0	1.00	121	50	2	0.054	0.053	5	0.060	0.052	0.056
0.50	0	1.00	121	50	3	0.051	0.047	5	0.060	0.052	0.056
0.50	0	1.00	121	50	4	0.063	0.057	5	0.060	0.052	0.056
0.50	0	1.00	121	100	2	0.044	0.043	5	0.060	0.058	0.056
0.50	0	1.00	121	100	3	0.051	0.053	5	0.060	0.058	0.056
0.50	0	1.00	121	100	4	0.055	0.053	5	0.060	0.058	0.056
0.50	0	1.00	121	150	2	0.056	0.056	5	0.041	0.043	0.062
0.50	0	1.00	121	150	3	0.060	0.059	5	0.041	0.043	0.062
0.50	0	1.00	121	150	4	0.048	0.049	5	0.041	0.043	0.062
0.50	0	1.00	121	200	2	0.064	0.064	5	0.050	0.048	0.061
0.50	0	1.00	121	200	3	0.052	0.053	5	0.050	0.048	0.061
0.50	0	1.00	121	200	4	0.058	0.059	5	0.050	0.048	0.061
0.50	0	1.00	169	50	2	0.056	0.054	5	0.065	0.061	0.057
0.50	0	1.00	169	50	3	0.057	0.057	5	0.065	0.061	0.057
0.50	0	1.00	169	50	4	0.052	0.051	5	0.065	0.061	0.057
0.50	0	1.00	169	100	2	0.053	0.052	5	0.054	0.052	0.052
0.50	0	1.00	169	100	3	0.070	0.068	5	0.054	0.052	0.052
0.50	0	1.00	169	100	4	0.064	0.062	5	0.054	0.052	0.052
0.50	0	1.00	169	150	2	0.049	0.049	5	0.054	0.052	0.044
0.50	0	1.00	169	150	3	0.067	0.065	5	0.054	0.052	0.044
0.50	0	1.00	169	150	4	0.068	0.067	5	0.054	0.052	0.044
0.50	0	1.00	169	200	2	0.055	0.055	5	0.056	0.054	0.054
0.50	0	1.00	169	200	3	0.059	0.060	5	0.056	0.054	0.054
0.50	0	1.00	169	200	4	0.061	0.061	5	0.056	0.054	0.054
0.50	0	1.00	25	50	2	0.094	0.059	7	0.315	0.059	0.088
0.50	0	1.00	25	50	3	0.112	0.054	7	0.315	0.059	0.088
0.50	0	1.00	25	50	4	0.159	0.053	7	0.315	0.059	0.088
0.50	0	1.00	25	100	2	0.062	0.053	7	0.158	0.059	0.069
0.50	0	1.00	25	100	3	0.087	0.055	7	0.158	0.059	0.069
0.50	0	1.00	25	100	4	0.109	0.073	7	0.158	0.059	0.069
0.50	0	1.00	25	150	2	0.060	0.050	7	0.115	0.054	0.058
0.50	0	1.00	25	150	3	0.071	0.050	7	0.115	0.054	0.058
0.50	0	1.00	25	150	4	0.069	0.046	7	0.115	0.054	0.058
0.50	0	1.00	25	200	2	0.066	0.060	7	0.104	0.060	0.059
0.50	0	1.00	25	200	3	0.059	0.056	7	0.104	0.060	0.059
0.50	0	1.00	25	200	4	0.072	0.064	7	0.104	0.060	0.059
0.50	0	1.00	49	50	2	0.061	0.049	7	0.159	0.050	0.063
0.50	0	1.00	49	50	3	0.078	0.057	7	0.159	0.050	0.063
0.50	0	1.00	49	50	4	0.095	0.059	7	0.159	0.050	0.063
0.50	0	1.00	49	100	2	0.074	0.060	7	0.095	0.048	0.057

Continued on next page

Table 4 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	1.00	49	100	3	0.064	0.056	7	0.095	0.048	0.057
0.50	0	1.00	49	100	4	0.077	0.059	7	0.095	0.048	0.057
0.50	0	1.00	49	150	2	0.052	0.048	7	0.084	0.053	0.064
0.50	0	1.00	49	150	3	0.061	0.054	7	0.084	0.053	0.064
0.50	0	1.00	49	150	4	0.052	0.045	7	0.084	0.053	0.064
0.50	0	1.00	49	200	2	0.046	0.041	7	0.075	0.055	0.066
0.50	0	1.00	49	200	3	0.070	0.066	7	0.075	0.055	0.066
0.50	0	1.00	49	200	4	0.050	0.048	7	0.075	0.055	0.066
0.50	0	1.00	81	50	2	0.050	0.047	7	0.108	0.047	0.063
0.50	0	1.00	81	50	3	0.071	0.057	7	0.108	0.047	0.063
0.50	0	1.00	81	50	4	0.082	0.055	7	0.108	0.047	0.063
0.50	0	1.00	81	100	2	0.061	0.055	7	0.066	0.040	0.044
0.50	0	1.00	81	100	3	0.071	0.066	7	0.066	0.040	0.044
0.50	0	1.00	81	100	4	0.071	0.063	7	0.066	0.040	0.044
0.50	0	1.00	81	150	2	0.056	0.055	7	0.073	0.052	0.049
0.50	0	1.00	81	150	3	0.056	0.055	7	0.073	0.052	0.049
0.50	0	1.00	81	150	4	0.063	0.057	7	0.073	0.052	0.049
0.50	0	1.00	81	200	2	0.051	0.050	7	0.055	0.041	0.042
0.50	0	1.00	81	200	3	0.054	0.053	7	0.055	0.041	0.042
0.50	0	1.00	81	200	4	0.072	0.070	7	0.055	0.041	0.042
0.50	0	1.00	121	50	2	0.065	0.062	7	0.091	0.055	0.064
0.50	0	1.00	121	50	3	0.066	0.055	7	0.091	0.055	0.064
0.50	0	1.00	121	50	4	0.076	0.059	7	0.091	0.055	0.064
0.50	0	1.00	121	100	2	0.052	0.048	7	0.052	0.042	0.052
0.50	0	1.00	121	100	3	0.055	0.052	7	0.052	0.042	0.052
0.50	0	1.00	121	100	4	0.062	0.057	7	0.052	0.042	0.052
0.50	0	1.00	121	150	2	0.056	0.055	7	0.049	0.043	0.055
0.50	0	1.00	121	150	3	0.056	0.055	7	0.049	0.043	0.055
0.50	0	1.00	121	150	4	0.055	0.051	7	0.049	0.043	0.055
0.50	0	1.00	121	200	2	0.059	0.059	7	0.063	0.056	0.046
0.50	0	1.00	121	200	3	0.069	0.065	7	0.063	0.056	0.046
0.50	0	1.00	121	200	4	0.054	0.052	7	0.063	0.056	0.046
0.50	0	1.00	169	50	2	0.050	0.044	7	0.077	0.044	0.063
0.50	0	1.00	169	50	3	0.059	0.059	7	0.077	0.044	0.063
0.50	0	1.00	169	50	4	0.060	0.054	7	0.077	0.044	0.063
0.50	0	1.00	169	100	2	0.048	0.045	7	0.049	0.040	0.042
0.50	0	1.00	169	100	3	0.065	0.061	7	0.049	0.040	0.042
0.50	0	1.00	169	100	4	0.051	0.045	7	0.049	0.040	0.042
0.50	0	1.00	169	150	2	0.052	0.052	7	0.074	0.060	0.055
0.50	0	1.00	169	150	3	0.062	0.061	7	0.074	0.060	0.055
0.50	0	1.00	169	150	4	0.056	0.054	7	0.074	0.060	0.055
0.50	0	1.00	169	200	2	0.049	0.047	7	0.064	0.060	0.044
0.50	0	1.00	169	200	3	0.045	0.044	7	0.064	0.060	0.044
0.50	0	1.00	169	200	4	0.039	0.041	7	0.064	0.060	0.044

Table 5. Power & Robustness Results, Mean Heterogeneity.

$\bar{\beta}$	$\bar{\lambda}$	μ_h	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	–	25	50	2	1.000	1.000	3	0.076	0.069	0.184
0.50	0	–	25	50	3	1.000	1.000	3	0.768	0.795	0.955
0.50	0	–	25	50	4	1.000	1.000	3	0.271	0.253	0.385
0.50	0	–	25	100	2	1.000	1.000	3	0.079	0.079	0.244
0.50	0	–	25	100	3	1.000	1.000	3	0.993	0.996	1.000
0.50	0	–	25	100	4	1.000	1.000	3	0.567	0.533	0.926
0.50	0	–	25	150	2	1.000	1.000	3	0.096	0.094	0.409
0.50	0	–	25	150	3	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	25	150	4	1.000	1.000	3	0.783	0.749	0.995
0.50	0	–	25	200	2	1.000	1.000	3	0.130	0.132	0.522
0.50	0	–	25	200	3	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	25	200	4	1.000	1.000	3	0.923	0.912	1.000
0.50	0	–	49	50	2	1.000	1.000	3	0.099	0.099	0.097
0.50	0	–	49	50	3	1.000	1.000	3	0.644	0.668	0.891
0.50	0	–	49	50	4	1.000	1.000	3	0.207	0.186	0.290
0.50	0	–	49	100	2	1.000	1.000	3	0.119	0.119	0.149
0.50	0	–	49	100	3	1.000	1.000	3	0.972	0.979	1.000
0.50	0	–	49	100	4	1.000	1.000	3	0.400	0.350	0.767
0.50	0	–	49	150	2	1.000	1.000	3	0.172	0.175	0.230
0.50	0	–	49	150	3	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	49	150	4	1.000	1.000	3	0.566	0.508	0.951
0.50	0	–	49	200	2	1.000	1.000	3	0.199	0.199	0.273
0.50	0	–	49	200	3	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	49	200	4	1.000	1.000	3	0.747	0.714	0.992
0.50	0	–	81	50	2	1.000	1.000	3	0.123	0.117	0.225
0.50	0	–	81	50	3	1.000	1.000	3	0.305	0.280	0.431
0.50	0	–	81	50	4	1.000	1.000	3	0.819	0.794	0.802
0.50	0	–	81	100	2	1.000	1.000	3	0.230	0.229	0.422
0.50	0	–	81	100	3	1.000	1.000	3	0.631	0.600	0.957
0.50	0	–	81	100	4	1.000	1.000	3	0.988	0.988	0.996
0.50	0	–	81	150	2	1.000	1.000	3	0.317	0.316	0.643
0.50	0	–	81	150	3	1.000	1.000	3	0.827	0.809	0.999
0.50	0	–	81	150	4	1.000	1.000	3	0.998	0.998	1.000
0.50	0	–	81	200	2	1.000	1.000	3	0.428	0.429	0.784
0.50	0	–	81	200	3	1.000	1.000	3	0.955	0.952	1.000
0.50	0	–	81	200	4	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	121	50	2	1.000	1.000	3	0.107	0.107	0.159
0.50	0	–	121	50	3	1.000	1.000	3	0.259	0.233	0.669
0.50	0	–	121	50	4	1.000	1.000	3	0.564	0.514	0.448
0.50	0	–	121	100	2	1.000	1.000	3	0.132	0.132	0.243
0.50	0	–	121	100	3	1.000	1.000	3	0.595	0.541	0.998
0.50	0	–	121	100	4	1.000	1.000	3	0.920	0.918	0.908
0.50	0	–	121	150	2	1.000	1.000	3	0.202	0.203	0.356
0.50	0	–	121	150	3	1.000	1.000	3	0.835	0.818	1.000
0.50	0	–	121	150	4	1.000	1.000	3	0.986	0.987	0.994
0.50	0	–	121	200	2	1.000	1.000	3	0.241	0.245	0.443

Continued on next page

Table 5 continued

$\bar{\beta}$	$\bar{\lambda}$	μ_h	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	–	121	200	3	1.000	1.000	3	0.944	0.936	1.000
0.50	0	–	121	200	4	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	169	50	2	1.000	1.000	3	0.217	0.215	0.264
0.50	0	–	169	50	3	1.000	1.000	3	0.682	0.635	0.624
0.50	0	–	169	50	4	1.000	1.000	3	0.479	0.435	0.537
0.50	0	–	169	100	2	1.000	1.000	3	0.425	0.425	0.521
0.50	0	–	169	100	3	1.000	1.000	3	0.971	0.966	0.979
0.50	0	–	169	100	4	1.000	1.000	3	0.879	0.874	0.956
0.50	0	–	169	150	2	1.000	1.000	3	0.597	0.597	0.729
0.50	0	–	169	150	3	1.000	1.000	3	0.999	1.000	1.000
0.50	0	–	169	150	4	1.000	1.000	3	0.984	0.984	1.000
0.50	0	–	169	200	2	1.000	1.000	3	0.756	0.756	0.874
0.50	0	–	169	200	3	1.000	1.000	3	1.000	1.000	1.000
0.50	0	–	169	200	4	1.000	1.000	3	0.999	0.999	1.000
0.50	0	–	25	50	2	1.000	1.000	5	0.117	0.056	0.100
0.50	0	–	25	50	3	1.000	1.000	5	0.758	0.769	0.065
0.50	0	–	25	50	4	1.000	1.000	5	0.331	0.242	0.032
0.50	0	–	25	100	2	1.000	1.000	5	0.098	0.062	0.159
0.50	0	–	25	100	3	1.000	1.000	5	0.995	0.997	0.339
0.50	0	–	25	100	4	1.000	1.000	5	0.557	0.527	0.095
0.50	0	–	25	150	2	1.000	1.000	5	0.112	0.082	0.222
0.50	0	–	25	150	3	1.000	1.000	5	1.000	1.000	0.794
0.50	0	–	25	150	4	1.000	1.000	5	0.789	0.790	0.339
0.50	0	–	25	200	2	1.000	1.000	5	0.103	0.080	0.267
0.50	0	–	25	200	3	1.000	1.000	5	1.000	1.000	0.995
0.50	0	–	25	200	4	1.000	1.000	5	0.919	0.929	0.752
0.50	0	–	49	50	2	1.000	1.000	5	0.081	0.055	0.106
0.50	0	–	49	50	3	1.000	1.000	5	0.596	0.583	0.037
0.50	0	–	49	50	4	1.000	1.000	5	0.195	0.160	0.025
0.50	0	–	49	100	2	1.000	1.000	5	0.087	0.060	0.105
0.50	0	–	49	100	3	1.000	1.000	5	0.962	0.953	0.108
0.50	0	–	49	100	4	1.000	1.000	5	0.335	0.329	0.045
0.50	0	–	49	150	2	1.000	1.000	5	0.103	0.080	0.130
0.50	0	–	49	150	3	1.000	1.000	5	0.999	0.998	0.403
0.50	0	–	49	150	4	1.000	1.000	5	0.515	0.513	0.131
0.50	0	–	49	200	2	1.000	1.000	5	0.114	0.092	0.117
0.50	0	–	49	200	3	1.000	1.000	5	1.000	1.000	0.856
0.50	0	–	49	200	4	1.000	1.000	5	0.662	0.649	0.303
0.50	0	–	81	50	2	1.000	1.000	5	0.104	0.072	0.125
0.50	0	–	81	50	3	1.000	1.000	5	0.284	0.250	0.016
0.50	0	–	81	50	4	1.000	1.000	5	0.477	0.470	0.039
0.50	0	–	81	100	2	1.000	1.000	5	0.133	0.104	0.184
0.50	0	–	81	100	3	1.000	1.000	5	0.592	0.561	0.035
0.50	0	–	81	100	4	1.000	1.000	5	0.878	0.884	0.125
0.50	0	–	81	150	2	1.000	1.000	5	0.165	0.137	0.209
0.50	0	–	81	150	3	1.000	1.000	5	0.824	0.808	0.046
0.50	0	–	81	150	4	1.000	1.000	5	0.991	0.991	0.319

Continued on next page

Table 5 continued

$\bar{\beta}$	$\bar{\lambda}$	μ_h	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	–	81	200	2	1.000	1.000	5	0.196	0.163	0.296
0.50	0	–	81	200	3	1.000	1.000	5	0.951	0.948	0.072
0.50	0	–	81	200	4	1.000	1.000	5	1.000	0.998	0.696
0.50	0	–	121	50	2	1.000	1.000	5	0.103	0.075	0.121
0.50	0	–	121	50	3	1.000	1.000	5	0.364	0.319	0.030
0.50	0	–	121	50	4	1.000	1.000	5	0.309	0.295	0.038
0.50	0	–	121	100	2	1.000	1.000	5	0.115	0.090	0.136
0.50	0	–	121	100	3	1.000	1.000	5	0.740	0.710	0.047
0.50	0	–	121	100	4	1.000	1.000	5	0.687	0.665	0.102
0.50	0	–	121	150	2	1.000	1.000	5	0.142	0.118	0.192
0.50	0	–	121	150	3	1.000	1.000	5	0.952	0.934	0.112
0.50	0	–	121	150	4	1.000	1.000	5	0.916	0.895	0.201
0.50	0	–	121	200	2	1.000	1.000	5	0.156	0.144	0.228
0.50	0	–	121	200	3	1.000	1.000	5	0.996	0.992	0.294
0.50	0	–	121	200	4	1.000	1.000	5	0.991	0.979	0.470
0.50	0	–	169	50	2	1.000	1.000	5	0.103	0.080	0.119
0.50	0	–	169	50	3	1.000	1.000	5	0.386	0.348	0.035
0.50	0	–	169	50	4	1.000	1.000	5	0.202	0.201	0.042
0.50	0	–	169	100	2	1.000	1.000	5	0.200	0.182	0.189
0.50	0	–	169	100	3	1.000	1.000	5	0.797	0.768	0.040
0.50	0	–	169	100	4	1.000	1.000	5	0.487	0.447	0.095
0.50	0	–	169	150	2	1.000	1.000	5	0.230	0.218	0.240
0.50	0	–	169	150	3	1.000	1.000	5	0.967	0.959	0.077
0.50	0	–	169	150	4	1.000	1.000	5	0.782	0.717	0.137
0.50	0	–	169	200	2	1.000	1.000	5	0.313	0.304	0.353
0.50	0	–	169	200	3	1.000	1.000	5	0.994	0.992	0.164
0.50	0	–	169	200	4	1.000	1.000	5	0.922	0.892	0.283
0.50	0	–	25	50	2	1.000	1.000	7	0.200	0.057	0.041
0.50	0	–	25	50	3	1.000	1.000	7	0.637	0.488	0.016
0.50	0	–	25	50	4	1.000	1.000	7	0.255	0.222	0.013
0.50	0	–	25	100	2	1.000	1.000	7	0.163	0.069	0.094
0.50	0	–	25	100	3	1.000	1.000	7	0.956	0.905	0.021
0.50	0	–	25	100	4	1.000	1.000	7	0.412	0.400	0.023
0.50	0	–	25	150	2	1.000	1.000	7	0.157	0.081	0.110
0.50	0	–	25	150	3	1.000	1.000	7	0.997	0.996	0.030
0.50	0	–	25	150	4	1.000	1.000	7	0.660	0.621	0.040
0.50	0	–	25	200	2	1.000	1.000	7	0.154	0.092	0.128
0.50	0	–	25	200	3	1.000	1.000	7	1.000	1.000	0.061
0.50	0	–	25	200	4	1.000	1.000	7	0.847	0.823	0.060
0.50	0	–	49	50	2	1.000	1.000	7	0.147	0.045	0.053
0.50	0	–	49	50	3	1.000	1.000	7	0.383	0.309	0.009
0.50	0	–	49	50	4	1.000	1.000	7	0.172	0.133	0.012
0.50	0	–	49	100	2	1.000	1.000	7	0.121	0.069	0.084
0.50	0	–	49	100	3	1.000	1.000	7	0.675	0.667	0.013
0.50	0	–	49	100	4	1.000	1.000	7	0.325	0.239	0.018
0.50	0	–	49	150	2	1.000	1.000	7	0.118	0.053	0.116
0.50	0	–	49	150	3	1.000	1.000	7	0.916	0.923	0.019

Continued on next page

Table 5 continued

$\bar{\beta}$	$\bar{\lambda}$	μ_h	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0	–	49	150	4	1.000	1.000	7	0.483	0.436	0.023
0.50	0	–	49	200	2	1.000	1.000	7	0.109	0.064	0.114
0.50	0	–	49	200	3	1.000	1.000	7	0.986	0.985	0.017
0.50	0	–	49	200	4	1.000	1.000	7	0.588	0.562	0.039
0.50	0	–	81	50	2	1.000	1.000	7	0.110	0.050	0.080
0.50	0	–	81	50	3	1.000	1.000	7	0.160	0.137	0.005
0.50	0	–	81	50	4	1.000	1.000	7	0.316	0.242	0.013
0.50	0	–	81	100	2	1.000	1.000	7	0.141	0.084	0.130
0.50	0	–	81	100	3	1.000	1.000	7	0.315	0.266	0.010
0.50	0	–	81	100	4	1.000	1.000	7	0.649	0.604	0.024
0.50	0	–	81	150	2	1.000	1.000	7	0.129	0.083	0.145
0.50	0	–	81	150	3	1.000	1.000	7	0.459	0.429	0.011
0.50	0	–	81	150	4	1.000	1.000	7	0.844	0.850	0.043
0.50	0	–	81	200	2	1.000	1.000	7	0.138	0.106	0.194
0.50	0	–	81	200	3	1.000	1.000	7	0.631	0.598	0.012
0.50	0	–	81	200	4	1.000	1.000	7	0.949	0.950	0.052
0.50	0	–	121	50	2	1.000	1.000	7	0.115	0.066	0.084
0.50	0	–	121	50	3	1.000	1.000	7	0.232	0.199	0.012
0.50	0	–	121	50	4	1.000	1.000	7	0.274	0.211	0.020
0.50	0	–	121	100	2	1.000	1.000	7	0.123	0.086	0.113
0.50	0	–	121	100	3	1.000	1.000	7	0.392	0.373	0.013
0.50	0	–	121	100	4	1.000	1.000	7	0.447	0.420	0.034
0.50	0	–	121	150	2	1.000	1.000	7	0.114	0.086	0.156
0.50	0	–	121	150	3	1.000	1.000	7	0.633	0.630	0.015
0.50	0	–	121	150	4	1.000	1.000	7	0.680	0.693	0.039
0.50	0	–	121	200	2	1.000	1.000	7	0.136	0.103	0.179
0.50	0	–	121	200	3	1.000	1.000	7	0.752	0.783	0.017
0.50	0	–	121	200	4	1.000	1.000	7	0.859	0.862	0.059
0.50	0	–	169	50	2	1.000	1.000	7	0.129	0.075	0.112
0.50	0	–	169	50	3	1.000	1.000	7	0.209	0.191	0.011
0.50	0	–	169	50	4	1.000	1.000	7	0.175	0.135	0.015
0.50	0	–	169	100	2	1.000	1.000	7	0.149	0.111	0.143
0.50	0	–	169	100	3	1.000	1.000	7	0.503	0.482	0.013
0.50	0	–	169	100	4	1.000	1.000	7	0.245	0.233	0.026
0.50	0	–	169	150	2	1.000	1.000	7	0.204	0.156	0.184
0.50	0	–	169	150	3	1.000	1.000	7	0.698	0.712	0.019
0.50	0	–	169	150	4	1.000	1.000	7	0.404	0.415	0.035
0.50	0	–	169	200	2	1.000	1.000	7	0.206	0.176	0.195
0.50	0	–	169	200	3	1.000	1.000	7	0.872	0.897	0.022
0.50	0	–	169	200	4	1.000	1.000	7	0.609	0.599	0.046

Table 6. Power & Robustness Results, Lag Heterogeneity.

$\boldsymbol{B}_{hh}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
–	0	1.00	25	50	2	0.835	0.833	3	0.042	0.038	0.061
–	0	1.00	25	50	3	1.000	1.000	3	0.073	0.069	0.075
–	0	1.00	25	50	4	0.997	0.997	3	0.064	0.063	0.074
–	0	1.00	25	100	2	0.998	0.998	3	0.061	0.061	0.061
–	0	1.00	25	100	3	1.000	1.000	3	0.075	0.073	0.083
–	0	1.00	25	100	4	1.000	1.000	3	0.072	0.069	0.068
–	0	1.00	25	150	2	1.000	1.000	3	0.047	0.048	0.063
–	0	1.00	25	150	3	1.000	1.000	3	0.073	0.070	0.080
–	0	1.00	25	150	4	1.000	1.000	3	0.058	0.058	0.054
–	0	1.00	25	200	2	1.000	1.000	3	0.054	0.053	0.068
–	0	1.00	25	200	3	1.000	1.000	3	0.064	0.064	0.099
–	0	1.00	25	200	4	1.000	1.000	3	0.060	0.059	0.084
–	0	1.00	49	50	2	0.991	0.991	3	0.068	0.066	0.059
–	0	1.00	49	50	3	1.000	1.000	3	0.071	0.066	0.072
–	0	1.00	49	50	4	1.000	1.000	3	0.066	0.063	0.072
–	0	1.00	49	100	2	1.000	1.000	3	0.045	0.046	0.051
–	0	1.00	49	100	3	1.000	1.000	3	0.064	0.062	0.097
–	0	1.00	49	100	4	1.000	1.000	3	0.064	0.063	0.070
–	0	1.00	49	150	2	1.000	1.000	3	0.055	0.056	0.069
–	0	1.00	49	150	3	1.000	1.000	3	0.062	0.063	0.075
–	0	1.00	49	150	4	1.000	1.000	3	0.060	0.058	0.067
–	0	1.00	49	200	2	1.000	1.000	3	0.051	0.051	0.070
–	0	1.00	49	200	3	1.000	1.000	3	0.062	0.064	0.086
–	0	1.00	49	200	4	1.000	1.000	3	0.069	0.068	0.081
–	0	1.00	81	50	2	0.999	0.999	3	0.051	0.051	0.058
–	0	1.00	81	50	3	1.000	1.000	3	0.063	0.061	0.078
–	0	1.00	81	50	4	1.000	1.000	3	0.071	0.070	0.080
–	0	1.00	81	100	2	1.000	1.000	3	0.067	0.066	0.061
–	0	1.00	81	100	3	1.000	1.000	3	0.059	0.059	0.087
–	0	1.00	81	100	4	1.000	1.000	3	0.056	0.055	0.085
–	0	1.00	81	150	2	1.000	1.000	3	0.056	0.055	0.069
–	0	1.00	81	150	3	1.000	1.000	3	0.050	0.049	0.066
–	0	1.00	81	150	4	1.000	1.000	3	0.058	0.060	0.072
–	0	1.00	81	200	2	1.000	1.000	3	0.059	0.057	0.064
–	0	1.00	81	200	3	1.000	1.000	3	0.079	0.081	0.097
–	0	1.00	81	200	4	1.000	1.000	3	0.077	0.077	0.085
–	0	1.00	121	50	2	1.000	1.000	3	0.055	0.056	0.064
–	0	1.00	121	50	3	1.000	1.000	3	0.067	0.067	0.065
–	0	1.00	121	50	4	1.000	1.000	3	0.068	0.067	0.068
–	0	1.00	121	100	2	1.000	1.000	3	0.065	0.065	0.064
–	0	1.00	121	100	3	1.000	1.000	3	0.050	0.050	0.078
–	0	1.00	121	100	4	1.000	1.000	3	0.059	0.059	0.093
–	0	1.00	121	150	2	1.000	1.000	3	0.055	0.054	0.083
–	0	1.00	121	150	3	1.000	1.000	3	0.080	0.079	0.073
–	0	1.00	121	150	4	1.000	1.000	3	0.066	0.067	0.071
–	0	1.00	121	200	2	1.000	1.000	3	0.052	0.052	0.058

Continued on next page

Table 6 continued

$\boldsymbol{B}_{hh}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
–	0	1.00	121	200	3	1.000	1.000	3	0.061	0.062	0.082
–	0	1.00	121	200	4	1.000	1.000	3	0.055	0.057	0.065
–	0	1.00	169	50	2	1.000	1.000	3	0.043	0.043	0.059
–	0	1.00	169	50	3	1.000	1.000	3	0.060	0.059	0.073
–	0	1.00	169	50	4	1.000	1.000	3	0.068	0.068	0.082
–	0	1.00	169	100	2	1.000	1.000	3	0.053	0.053	0.071
–	0	1.00	169	100	3	1.000	1.000	3	0.060	0.059	0.082
–	0	1.00	169	100	4	1.000	1.000	3	0.070	0.071	0.079
–	0	1.00	169	150	2	1.000	1.000	3	0.059	0.058	0.061
–	0	1.00	169	150	3	1.000	1.000	3	0.066	0.066	0.076
–	0	1.00	169	150	4	1.000	1.000	3	0.057	0.057	0.073
–	0	1.00	169	200	2	1.000	1.000	3	0.054	0.053	0.059
–	0	1.00	169	200	3	1.000	1.000	3	0.059	0.059	0.070
–	0	1.00	169	200	4	1.000	1.000	3	0.052	0.051	0.062
–	0	1.00	25	50	2	0.760	0.747	5	0.086	0.056	0.054
–	0	1.00	25	50	3	0.997	0.997	5	0.116	0.067	0.100
–	0	1.00	25	50	4	0.986	0.982	5	0.100	0.061	0.087
–	0	1.00	25	100	2	0.984	0.984	5	0.083	0.074	0.055
–	0	1.00	25	100	3	1.000	1.000	5	0.080	0.066	0.082
–	0	1.00	25	100	4	1.000	1.000	5	0.067	0.050	0.065
–	0	1.00	25	150	2	1.000	1.000	5	0.044	0.039	0.046
–	0	1.00	25	150	3	1.000	1.000	5	0.067	0.058	0.067
–	0	1.00	25	150	4	1.000	1.000	5	0.064	0.054	0.064
–	0	1.00	25	200	2	1.000	1.000	5	0.047	0.043	0.055
–	0	1.00	25	200	3	1.000	1.000	5	0.076	0.069	0.084
–	0	1.00	25	200	4	1.000	1.000	5	0.061	0.055	0.060
–	0	1.00	49	50	2	0.978	0.977	5	0.068	0.055	0.060
–	0	1.00	49	50	3	1.000	1.000	5	0.085	0.063	0.077
–	0	1.00	49	50	4	1.000	1.000	5	0.077	0.057	0.064
–	0	1.00	49	100	2	1.000	1.000	5	0.053	0.050	0.050
–	0	1.00	49	100	3	1.000	1.000	5	0.064	0.053	0.066
–	0	1.00	49	100	4	1.000	1.000	5	0.060	0.050	0.074
–	0	1.00	49	150	2	1.000	1.000	5	0.051	0.048	0.057
–	0	1.00	49	150	3	1.000	1.000	5	0.064	0.064	0.070
–	0	1.00	49	150	4	1.000	1.000	5	0.062	0.055	0.078
–	0	1.00	49	200	2	1.000	1.000	5	0.045	0.042	0.051
–	0	1.00	49	200	3	1.000	1.000	5	0.059	0.057	0.072
–	0	1.00	49	200	4	1.000	1.000	5	0.069	0.066	0.067
–	0	1.00	81	50	2	0.999	0.999	5	0.052	0.053	0.054
–	0	1.00	81	50	3	1.000	1.000	5	0.068	0.058	0.068
–	0	1.00	81	50	4	1.000	1.000	5	0.066	0.059	0.065
–	0	1.00	81	100	2	1.000	1.000	5	0.055	0.053	0.054
–	0	1.00	81	100	3	1.000	1.000	5	0.073	0.066	0.064
–	0	1.00	81	100	4	1.000	1.000	5	0.064	0.058	0.062
–	0	1.00	81	150	2	1.000	1.000	5	0.064	0.061	0.064
–	0	1.00	81	150	3	1.000	1.000	5	0.064	0.056	0.050
–	0	1.00	81	150	4	1.000	1.000	5	0.061	0.056	0.064

Continued on next page

Table 6 continued

$\boldsymbol{B}_{hh}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
–	0	1.00	81	200	2	1.000	1.000	5	0.046	0.044	0.058
–	0	1.00	81	200	3	1.000	1.000	5	0.055	0.053	0.066
–	0	1.00	81	200	4	1.000	1.000	5	0.062	0.061	0.068
–	0	1.00	121	50	2	1.000	1.000	5	0.076	0.064	0.053
–	0	1.00	121	50	3	1.000	1.000	5	0.074	0.067	0.076
–	0	1.00	121	50	4	1.000	1.000	5	0.060	0.054	0.051
–	0	1.00	121	100	2	1.000	1.000	5	0.051	0.051	0.041
–	0	1.00	121	100	3	1.000	1.000	5	0.069	0.065	0.072
–	0	1.00	121	100	4	1.000	1.000	5	0.082	0.075	0.069
–	0	1.00	121	150	2	1.000	1.000	5	0.049	0.046	0.054
–	0	1.00	121	150	3	1.000	1.000	5	0.070	0.066	0.072
–	0	1.00	121	150	4	1.000	1.000	5	0.064	0.063	0.063
–	0	1.00	121	200	2	1.000	1.000	5	0.044	0.043	0.050
–	0	1.00	121	200	3	1.000	1.000	5	0.071	0.066	0.057
–	0	1.00	121	200	4	1.000	1.000	5	0.064	0.061	0.061
–	0	1.00	169	50	2	1.000	1.000	5	0.066	0.061	0.068
–	0	1.00	169	50	3	1.000	1.000	5	0.061	0.059	0.062
–	0	1.00	169	50	4	1.000	1.000	5	0.061	0.058	0.059
–	0	1.00	169	100	2	1.000	1.000	5	0.057	0.056	0.043
–	0	1.00	169	100	3	1.000	1.000	5	0.063	0.067	0.053
–	0	1.00	169	100	4	1.000	1.000	5	0.067	0.065	0.051
–	0	1.00	169	150	2	1.000	1.000	5	0.052	0.054	0.050
–	0	1.00	169	150	3	1.000	1.000	5	0.074	0.071	0.049
–	0	1.00	169	150	4	1.000	1.000	5	0.056	0.055	0.058
–	0	1.00	169	200	2	1.000	1.000	5	0.057	0.058	0.050
–	0	1.00	169	200	3	1.000	1.000	5	0.061	0.061	0.067
–	0	1.00	169	200	4	1.000	1.000	5	0.057	0.056	0.063
–	0	1.00	25	50	2	0.663	0.611	7	0.326	0.050	0.054
–	0	1.00	25	50	3	0.995	0.990	7	0.319	0.046	0.105
–	0	1.00	25	50	4	0.961	0.928	7	0.321	0.067	0.083
–	0	1.00	25	100	2	0.954	0.947	7	0.125	0.044	0.054
–	0	1.00	25	100	3	1.000	1.000	7	0.169	0.055	0.110
–	0	1.00	25	100	4	1.000	1.000	7	0.174	0.062	0.108
–	0	1.00	25	150	2	0.999	0.999	7	0.105	0.055	0.053
–	0	1.00	25	150	3	1.000	1.000	7	0.131	0.048	0.124
–	0	1.00	25	150	4	1.000	1.000	7	0.098	0.039	0.071
–	0	1.00	25	200	2	1.000	1.000	7	0.090	0.059	0.056
–	0	1.00	25	200	3	1.000	1.000	7	0.116	0.064	0.109
–	0	1.00	25	200	4	1.000	1.000	7	0.099	0.053	0.080
–	0	1.00	49	50	2	0.950	0.946	7	0.152	0.053	0.055
–	0	1.00	49	50	3	1.000	1.000	7	0.178	0.061	0.108
–	0	1.00	49	50	4	1.000	1.000	7	0.179	0.065	0.101
–	0	1.00	49	100	2	1.000	1.000	7	0.084	0.052	0.060
–	0	1.00	49	100	3	1.000	1.000	7	0.104	0.058	0.086
–	0	1.00	49	100	4	1.000	1.000	7	0.091	0.043	0.076
–	0	1.00	49	150	2	1.000	1.000	7	0.065	0.051	0.043
–	0	1.00	49	150	3	1.000	1.000	7	0.094	0.066	0.084

Continued on next page

Table 6 continued

$\boldsymbol{B}_{hh}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
–	0	1.00	49	150	4	1.000	1.000	7	0.101	0.071	0.089
–	0	1.00	49	200	2	1.000	1.000	7	0.055	0.044	0.061
–	0	1.00	49	200	3	1.000	1.000	7	0.084	0.059	0.097
–	0	1.00	49	200	4	1.000	1.000	7	0.084	0.059	0.079
–	0	1.00	81	50	2	0.997	0.997	7	0.090	0.045	0.044
–	0	1.00	81	50	3	1.000	1.000	7	0.114	0.052	0.057
–	0	1.00	81	50	4	1.000	1.000	7	0.129	0.061	0.058
–	0	1.00	81	100	2	1.000	1.000	7	0.061	0.046	0.044
–	0	1.00	81	100	3	1.000	1.000	7	0.070	0.048	0.048
–	0	1.00	81	100	4	1.000	1.000	7	0.086	0.064	0.081
–	0	1.00	81	150	2	1.000	1.000	7	0.077	0.066	0.064
–	0	1.00	81	150	3	1.000	1.000	7	0.075	0.055	0.063
–	0	1.00	81	150	4	1.000	1.000	7	0.071	0.058	0.068
–	0	1.00	81	200	2	1.000	1.000	7	0.069	0.057	0.057
–	0	1.00	81	200	3	1.000	1.000	7	0.057	0.048	0.053
–	0	1.00	81	200	4	1.000	1.000	7	0.064	0.050	0.050
–	0	1.00	121	50	2	1.000	1.000	7	0.076	0.052	0.064
–	0	1.00	121	50	3	1.000	1.000	7	0.093	0.054	0.079
–	0	1.00	121	50	4	1.000	1.000	7	0.090	0.049	0.074
–	0	1.00	121	100	2	1.000	1.000	7	0.068	0.059	0.069
–	0	1.00	121	100	3	1.000	1.000	7	0.080	0.060	0.063
–	0	1.00	121	100	4	1.000	1.000	7	0.067	0.051	0.050
–	0	1.00	121	150	2	1.000	1.000	7	0.057	0.052	0.056
–	0	1.00	121	150	3	1.000	1.000	7	0.065	0.055	0.075
–	0	1.00	121	150	4	1.000	1.000	7	0.072	0.061	0.065
–	0	1.00	121	200	2	1.000	1.000	7	0.072	0.067	0.044
–	0	1.00	121	200	3	1.000	1.000	7	0.061	0.050	0.059
–	0	1.00	121	200	4	1.000	1.000	7	0.055	0.051	0.066
–	0	1.00	169	50	2	1.000	1.000	7	0.053	0.048	0.057
–	0	1.00	169	50	3	1.000	1.000	7	0.082	0.052	0.062
–	0	1.00	169	50	4	1.000	1.000	7	0.076	0.056	0.050
–	0	1.00	169	100	2	1.000	1.000	7	0.048	0.043	0.045
–	0	1.00	169	100	3	1.000	1.000	7	0.066	0.052	0.057
–	0	1.00	169	100	4	1.000	1.000	7	0.053	0.046	0.060
–	0	1.00	169	150	2	1.000	1.000	7	0.054	0.052	0.044
–	0	1.00	169	150	3	1.000	1.000	7	0.063	0.057	0.056
–	0	1.00	169	150	4	1.000	1.000	7	0.058	0.053	0.055
–	0	1.00	169	200	2	1.000	1.000	7	0.056	0.053	0.051
–	0	1.00	169	200	3	1.000	1.000	7	0.065	0.062	0.055
–	0	1.00	169	200	4	1.000	1.000	7	0.071	0.064	0.056

Table 7. Power & Robustness Results, Spatial Dependence.

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.125	1.00	25	50	2	0.065	0.065	3	0.274	0.270	0.291
0.50	0.250	1.00	25	50	2	0.081	0.080	3	0.919	0.921	0.922
0.50	0.375	1.00	25	50	2	0.115	0.113	3	1.000	1.000	1.000
0.50	0.500	1.00	25	50	2	0.242	0.240	3	1.000	1.000	1.000
0.50	0.125	1.00	25	100	2	0.071	0.071	3	0.585	0.587	0.597
0.50	0.250	1.00	25	100	2	0.079	0.078	3	0.999	0.999	0.999
0.50	0.375	1.00	25	100	2	0.119	0.119	3	1.000	1.000	1.000
0.50	0.500	1.00	25	100	2	0.261	0.262	3	1.000	1.000	1.000
0.50	0.125	1.00	25	150	2	0.059	0.059	3	0.826	0.826	0.831
0.50	0.250	1.00	25	150	2	0.067	0.067	3	1.000	1.000	1.000
0.50	0.375	1.00	25	150	2	0.110	0.110	3	1.000	1.000	1.000
0.50	0.500	1.00	25	150	2	0.288	0.288	3	1.000	1.000	1.000
0.50	0.125	1.00	25	200	2	0.071	0.071	3	0.917	0.916	0.927
0.50	0.250	1.00	25	200	2	0.065	0.065	3	1.000	1.000	1.000
0.50	0.375	1.00	25	200	2	0.090	0.090	3	1.000	1.000	1.000
0.50	0.500	1.00	25	200	2	0.261	0.262	3	1.000	1.000	1.000
0.50	0.125	1.00	49	50	2	0.070	0.069	3	0.505	0.500	0.530
0.50	0.250	1.00	49	50	2	0.075	0.074	3	0.999	0.999	1.000
0.50	0.375	1.00	49	50	2	0.074	0.073	3	1.000	1.000	1.000
0.50	0.500	1.00	49	50	2	0.122	0.121	3	1.000	1.000	1.000
0.50	0.125	1.00	49	100	2	0.054	0.054	3	0.872	0.870	0.878
0.50	0.250	1.00	49	100	2	0.056	0.056	3	1.000	1.000	1.000
0.50	0.375	1.00	49	100	2	0.075	0.075	3	1.000	1.000	1.000
0.50	0.500	1.00	49	100	2	0.127	0.127	3	1.000	1.000	1.000
0.50	0.125	1.00	49	150	2	0.052	0.052	3	0.986	0.987	0.987
0.50	0.250	1.00	49	150	2	0.069	0.069	3	1.000	1.000	1.000
0.50	0.375	1.00	49	150	2	0.084	0.084	3	1.000	1.000	1.000
0.50	0.500	1.00	49	150	2	0.132	0.128	3	1.000	1.000	1.000
0.50	0.125	1.00	49	200	2	0.079	0.079	3	0.998	0.998	0.998
0.50	0.250	1.00	49	200	2	0.075	0.075	3	1.000	1.000	1.000
0.50	0.375	1.00	49	200	2	0.078	0.078	3	1.000	1.000	1.000
0.50	0.500	1.00	49	200	2	0.133	0.133	3	1.000	1.000	1.000
0.50	0.125	1.00	81	50	2	0.062	0.061	3	0.762	0.760	0.771
0.50	0.250	1.00	81	50	2	0.061	0.061	3	1.000	1.000	1.000
0.50	0.375	1.00	81	50	2	0.072	0.072	3	1.000	1.000	1.000
0.50	0.500	1.00	81	50	2	0.196	0.197	3	1.000	1.000	1.000
0.50	0.125	1.00	81	100	2	0.051	0.051	3	0.986	0.986	0.985
0.50	0.250	1.00	81	100	2	0.059	0.059	3	1.000	1.000	1.000
0.50	0.375	1.00	81	100	2	0.087	0.086	3	1.000	1.000	1.000
0.50	0.500	1.00	81	100	2	0.226	0.228	3	1.000	1.000	1.000
0.50	0.125	1.00	81	150	2	0.072	0.072	3	1.000	1.000	1.000
0.50	0.250	1.00	81	150	2	0.073	0.073	3	1.000	1.000	1.000
0.50	0.375	1.00	81	150	2	0.098	0.098	3	1.000	1.000	1.000
0.50	0.500	1.00	81	150	2	0.280	0.280	3	1.000	1.000	1.000
0.50	0.125	1.00	81	200	2	0.060	0.060	3	1.000	1.000	1.000
0.50	0.250	1.00	81	200	2	0.078	0.078	3	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.375	1.00	81	200	2	0.088	0.088	3	1.000	1.000	1.000
0.50	0.500	1.00	81	200	2	0.288	0.288	3	1.000	1.000	1.000
0.50	0.125	1.00	121	50	2	0.076	0.076	3	0.932	0.933	0.937
0.50	0.250	1.00	121	50	2	0.080	0.080	3	1.000	1.000	1.000
0.50	0.375	1.00	121	50	2	0.114	0.114	3	1.000	1.000	1.000
0.50	0.500	1.00	121	50	2	0.340	0.340	3	1.000	1.000	1.000
0.50	0.125	1.00	121	100	2	0.065	0.065	3	1.000	1.000	1.000
0.50	0.250	1.00	121	100	2	0.071	0.071	3	1.000	1.000	1.000
0.50	0.375	1.00	121	100	2	0.107	0.107	3	1.000	1.000	1.000
0.50	0.500	1.00	121	100	2	0.405	0.405	3	1.000	1.000	1.000
0.50	0.125	1.00	121	150	2	0.073	0.073	3	1.000	1.000	1.000
0.50	0.250	1.00	121	150	2	0.082	0.082	3	1.000	1.000	1.000
0.50	0.375	1.00	121	150	2	0.114	0.113	3	1.000	1.000	1.000
0.50	0.500	1.00	121	150	2	0.486	0.487	3	1.000	1.000	1.000
0.50	0.125	1.00	121	200	2	0.065	0.065	3	1.000	1.000	1.000
0.50	0.250	1.00	121	200	2	0.069	0.069	3	1.000	1.000	1.000
0.50	0.375	1.00	121	200	2	0.094	0.094	3	1.000	1.000	1.000
0.50	0.500	1.00	121	200	2	0.484	0.483	3	1.000	1.000	1.000
0.50	0.125	1.00	169	50	2	0.058	0.058	3	0.990	0.990	0.993
0.50	0.250	1.00	169	50	2	0.061	0.061	3	1.000	1.000	1.000
0.50	0.375	1.00	169	50	2	0.087	0.087	3	1.000	1.000	1.000
0.50	0.500	1.00	169	50	2	0.202	0.202	3	1.000	1.000	1.000
0.50	0.125	1.00	169	100	2	0.068	0.068	3	1.000	1.000	1.000
0.50	0.250	1.00	169	100	2	0.074	0.074	3	1.000	1.000	1.000
0.50	0.375	1.00	169	100	2	0.094	0.094	3	1.000	1.000	1.000
0.50	0.500	1.00	169	100	2	0.291	0.291	3	1.000	1.000	1.000
0.50	0.125	1.00	169	150	2	0.042	0.042	3	1.000	1.000	1.000
0.50	0.250	1.00	169	150	2	0.052	0.052	3	1.000	1.000	1.000
0.50	0.375	1.00	169	150	2	0.074	0.074	3	1.000	1.000	1.000
0.50	0.500	1.00	169	150	2	0.370	0.370	3	1.000	1.000	1.000
0.50	0.125	1.00	169	200	2	0.056	0.056	3	1.000	1.000	1.000
0.50	0.250	1.00	169	200	2	0.072	0.072	3	1.000	1.000	1.000
0.50	0.375	1.00	169	200	2	0.098	0.098	3	1.000	1.000	1.000
0.50	0.500	1.00	169	200	2	0.391	0.390	3	1.000	1.000	1.000
0.50	0.125	1.00	25	50	2	0.067	0.056	5	0.184	0.125	0.144
0.50	0.250	1.00	25	50	2	0.081	0.071	5	0.738	0.681	0.659
0.50	0.375	1.00	25	50	2	0.122	0.097	5	0.996	0.993	0.994
0.50	0.500	1.00	25	50	2	0.164	0.147	5	1.000	1.000	1.000
0.50	0.125	1.00	25	100	2	0.060	0.059	5	0.322	0.288	0.271
0.50	0.250	1.00	25	100	2	0.078	0.075	5	0.981	0.980	0.953
0.50	0.375	1.00	25	100	2	0.106	0.101	5	1.000	1.000	1.000
0.50	0.500	1.00	25	100	2	0.198	0.182	5	1.000	1.000	1.000
0.50	0.125	1.00	25	150	2	0.063	0.058	5	0.535	0.522	0.414
0.50	0.250	1.00	25	150	2	0.065	0.063	5	1.000	1.000	0.998
0.50	0.375	1.00	25	150	2	0.084	0.082	5	1.000	1.000	1.000
0.50	0.500	1.00	25	150	2	0.218	0.201	5	1.000	1.000	1.000
0.50	0.125	1.00	25	200	2	0.049	0.048	5	0.670	0.662	0.586

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.250	1.00	25	200	2	0.062	0.061	5	1.000	1.000	1.000
0.50	0.375	1.00	25	200	2	0.070	0.067	5	1.000	1.000	1.000
0.50	0.500	1.00	25	200	2	0.221	0.188	5	1.000	1.000	1.000
0.50	0.125	1.00	49	50	2	0.056	0.055	5	0.301	0.259	0.247
0.50	0.250	1.00	49	50	2	0.062	0.057	5	0.959	0.957	0.931
0.50	0.375	1.00	49	50	2	0.077	0.073	5	1.000	1.000	1.000
0.50	0.500	1.00	49	50	2	0.104	0.082	5	1.000	1.000	1.000
0.50	0.125	1.00	49	100	2	0.062	0.061	5	0.603	0.590	0.513
0.50	0.250	1.00	49	100	2	0.059	0.057	5	1.000	1.000	1.000
0.50	0.375	1.00	49	100	2	0.071	0.067	5	1.000	1.000	1.000
0.50	0.500	1.00	49	100	2	0.107	0.092	5	1.000	1.000	1.000
0.50	0.125	1.00	49	150	2	0.058	0.055	5	0.863	0.859	0.749
0.50	0.250	1.00	49	150	2	0.063	0.061	5	1.000	1.000	1.000
0.50	0.375	1.00	49	150	2	0.061	0.061	5	1.000	1.000	1.000
0.50	0.500	1.00	49	150	2	0.109	0.098	5	1.000	1.000	1.000
0.50	0.125	1.00	49	200	2	0.060	0.060	5	0.965	0.965	0.893
0.50	0.250	1.00	49	200	2	0.066	0.062	5	1.000	1.000	1.000
0.50	0.375	1.00	49	200	2	0.052	0.052	5	1.000	1.000	1.000
0.50	0.500	1.00	49	200	2	0.108	0.098	5	1.000	1.000	1.000
0.50	0.125	1.00	81	50	2	0.058	0.058	5	0.435	0.424	0.376
0.50	0.250	1.00	81	50	2	0.068	0.063	5	0.998	0.998	0.994
0.50	0.375	1.00	81	50	2	0.056	0.053	5	1.000	1.000	1.000
0.50	0.500	1.00	81	50	2	0.141	0.125	5	1.000	1.000	1.000
0.50	0.125	1.00	81	100	2	0.052	0.051	5	0.876	0.876	0.781
0.50	0.250	1.00	81	100	2	0.051	0.050	5	1.000	1.000	1.000
0.50	0.375	1.00	81	100	2	0.069	0.069	5	1.000	1.000	1.000
0.50	0.500	1.00	81	100	2	0.157	0.147	5	1.000	1.000	1.000
0.50	0.125	1.00	81	150	2	0.055	0.055	5	0.985	0.983	0.944
0.50	0.250	1.00	81	150	2	0.064	0.063	5	1.000	1.000	1.000
0.50	0.375	1.00	81	150	2	0.075	0.075	5	1.000	1.000	1.000
0.50	0.500	1.00	81	150	2	0.183	0.175	5	1.000	1.000	1.000
0.50	0.125	1.00	81	200	2	0.061	0.059	5	0.998	0.998	0.993
0.50	0.250	1.00	81	200	2	0.057	0.052	5	1.000	1.000	1.000
0.50	0.375	1.00	81	200	2	0.076	0.076	5	1.000	1.000	1.000
0.50	0.500	1.00	81	200	2	0.212	0.200	5	1.000	1.000	1.000
0.50	0.125	1.00	121	50	2	0.059	0.057	5	0.705	0.695	0.587
0.50	0.250	1.00	121	50	2	0.071	0.068	5	1.000	1.000	1.000
0.50	0.375	1.00	121	50	2	0.074	0.071	5	1.000	1.000	1.000
0.50	0.500	1.00	121	50	2	0.232	0.216	5	1.000	1.000	1.000
0.50	0.125	1.00	121	100	2	0.058	0.058	5	0.982	0.984	0.943
0.50	0.250	1.00	121	100	2	0.061	0.058	5	1.000	1.000	1.000
0.50	0.375	1.00	121	100	2	0.092	0.092	5	1.000	1.000	1.000
0.50	0.500	1.00	121	100	2	0.316	0.310	5	1.000	1.000	1.000
0.50	0.125	1.00	121	150	2	0.053	0.053	5	0.999	0.999	0.994
0.50	0.250	1.00	121	150	2	0.058	0.058	5	1.000	1.000	1.000
0.50	0.375	1.00	121	150	2	0.088	0.088	5	1.000	1.000	1.000
0.50	0.500	1.00	121	150	2	0.362	0.353	5	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.125	1.00	121	200	2	0.064	0.064	5	1.000	1.000	1.000
0.50	0.250	1.00	121	200	2	0.065	0.065	5	1.000	1.000	1.000
0.50	0.375	1.00	121	200	2	0.090	0.089	5	1.000	1.000	1.000
0.50	0.500	1.00	121	200	2	0.367	0.360	5	1.000	1.000	1.000
0.50	0.125	1.00	169	50	2	0.061	0.060	5	0.864	0.855	0.746
0.50	0.250	1.00	169	50	2	0.055	0.053	5	1.000	1.000	1.000
0.50	0.375	1.00	169	50	2	0.063	0.063	5	1.000	1.000	1.000
0.50	0.500	1.00	169	50	2	0.143	0.130	5	1.000	1.000	1.000
0.50	0.125	1.00	169	100	2	0.045	0.045	5	0.999	0.999	0.995
0.50	0.250	1.00	169	100	2	0.063	0.061	5	1.000	1.000	1.000
0.50	0.375	1.00	169	100	2	0.062	0.062	5	1.000	1.000	1.000
0.50	0.500	1.00	169	100	2	0.207	0.199	5	1.000	1.000	1.000
0.50	0.125	1.00	169	150	2	0.055	0.054	5	1.000	1.000	1.000
0.50	0.250	1.00	169	150	2	0.073	0.073	5	1.000	1.000	1.000
0.50	0.375	1.00	169	150	2	0.070	0.069	5	1.000	1.000	1.000
0.50	0.500	1.00	169	150	2	0.238	0.233	5	1.000	1.000	1.000
0.50	0.125	1.00	169	200	2	0.049	0.049	5	1.000	1.000	1.000
0.50	0.250	1.00	169	200	2	0.055	0.055	5	1.000	1.000	1.000
0.50	0.375	1.00	169	200	2	0.075	0.074	5	1.000	1.000	1.000
0.50	0.500	1.00	169	200	2	0.263	0.259	5	1.000	1.000	1.000
0.50	0.125	1.00	25	50	2	0.083	0.053	7	0.402	0.108	0.160
0.50	0.250	1.00	25	50	2	0.097	0.054	7	0.775	0.435	0.570
0.50	0.375	1.00	25	50	2	0.130	0.091	7	0.990	0.964	0.952
0.50	0.500	1.00	25	50	2	0.178	0.100	7	1.000	1.000	0.986
0.50	0.125	1.00	25	100	2	0.068	0.052	7	0.322	0.163	0.222
0.50	0.250	1.00	25	100	2	0.069	0.054	7	0.935	0.874	0.867
0.50	0.375	1.00	25	100	2	0.096	0.073	7	1.000	1.000	1.000
0.50	0.500	1.00	25	100	2	0.179	0.131	7	1.000	1.000	1.000
0.50	0.125	1.00	25	150	2	0.056	0.049	7	0.410	0.283	0.281
0.50	0.250	1.00	25	150	2	0.080	0.065	7	0.993	0.985	0.963
0.50	0.375	1.00	25	150	2	0.081	0.067	7	1.000	1.000	1.000
0.50	0.500	1.00	25	150	2	0.190	0.149	7	1.000	1.000	1.000
0.50	0.125	1.00	25	200	2	0.070	0.065	7	0.488	0.389	0.356
0.50	0.250	1.00	25	200	2	0.071	0.065	7	1.000	1.000	0.994
0.50	0.375	1.00	25	200	2	0.075	0.068	7	1.000	1.000	1.000
0.50	0.500	1.00	25	200	2	0.181	0.140	7	1.000	1.000	1.000
0.50	0.125	1.00	49	50	2	0.066	0.048	7	0.351	0.176	0.210
0.50	0.250	1.00	49	50	2	0.078	0.067	7	0.887	0.770	0.825
0.50	0.375	1.00	49	50	2	0.092	0.072	7	1.000	0.999	1.000
0.50	0.500	1.00	49	50	2	0.115	0.077	7	1.000	1.000	1.000
0.50	0.125	1.00	49	100	2	0.050	0.046	7	0.425	0.324	0.306
0.50	0.250	1.00	49	100	2	0.055	0.046	7	0.996	0.995	0.988
0.50	0.375	1.00	49	100	2	0.067	0.054	7	1.000	1.000	1.000
0.50	0.500	1.00	49	100	2	0.107	0.073	7	1.000	1.000	1.000
0.50	0.125	1.00	49	150	2	0.068	0.060	7	0.642	0.567	0.472
0.50	0.250	1.00	49	150	2	0.053	0.050	7	1.000	1.000	1.000
0.50	0.375	1.00	49	150	2	0.070	0.067	7	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.500	1.00	49	150	2	0.103	0.085	7	1.000	1.000	1.000
0.50	0.125	1.00	49	200	2	0.050	0.047	7	0.794	0.754	0.606
0.50	0.250	1.00	49	200	2	0.056	0.055	7	1.000	1.000	1.000
0.50	0.375	1.00	49	200	2	0.061	0.057	7	1.000	1.000	1.000
0.50	0.500	1.00	49	200	2	0.103	0.080	7	1.000	1.000	1.000
0.50	0.125	1.00	81	50	2	0.064	0.062	7	0.370	0.255	0.271
0.50	0.250	1.00	81	50	2	0.065	0.058	7	0.979	0.967	0.951
0.50	0.375	1.00	81	50	2	0.076	0.061	7	1.000	1.000	1.000
0.50	0.500	1.00	81	50	2	0.151	0.113	7	1.000	1.000	1.000
0.50	0.125	1.00	81	100	2	0.050	0.046	7	0.681	0.615	0.477
0.50	0.250	1.00	81	100	2	0.044	0.040	7	1.000	1.000	1.000
0.50	0.375	1.00	81	100	2	0.070	0.063	7	1.000	1.000	1.000
0.50	0.500	1.00	81	100	2	0.142	0.118	7	1.000	1.000	1.000
0.50	0.125	1.00	81	150	2	0.052	0.051	7	0.884	0.865	0.706
0.50	0.250	1.00	81	150	2	0.059	0.059	7	1.000	1.000	1.000
0.50	0.375	1.00	81	150	2	0.061	0.057	7	1.000	1.000	1.000
0.50	0.500	1.00	81	150	2	0.150	0.128	7	1.000	1.000	1.000
0.50	0.125	1.00	81	200	2	0.050	0.049	7	0.972	0.967	0.868
0.50	0.250	1.00	81	200	2	0.063	0.061	7	1.000	1.000	1.000
0.50	0.375	1.00	81	200	2	0.085	0.085	7	1.000	1.000	1.000
0.50	0.500	1.00	81	200	2	0.173	0.155	7	1.000	1.000	1.000
0.50	0.125	1.00	121	50	2	0.061	0.055	7	0.519	0.438	0.371
0.50	0.250	1.00	121	50	2	0.060	0.055	7	0.999	0.999	0.997
0.50	0.375	1.00	121	50	2	0.081	0.074	7	1.000	1.000	1.000
0.50	0.500	1.00	121	50	2	0.212	0.174	7	1.000	1.000	1.000
0.50	0.125	1.00	121	100	2	0.062	0.062	7	0.863	0.848	0.682
0.50	0.250	1.00	121	100	2	0.073	0.070	7	1.000	1.000	1.000
0.50	0.375	1.00	121	100	2	0.066	0.062	7	1.000	1.000	1.000
0.50	0.500	1.00	121	100	2	0.265	0.232	7	1.000	1.000	1.000
0.50	0.125	1.00	121	150	2	0.053	0.051	7	0.984	0.982	0.888
0.50	0.250	1.00	121	150	2	0.058	0.056	7	1.000	1.000	1.000
0.50	0.375	1.00	121	150	2	0.071	0.068	7	1.000	1.000	1.000
0.50	0.500	1.00	121	150	2	0.286	0.267	7	1.000	1.000	1.000
0.50	0.125	1.00	121	200	2	0.065	0.063	7	1.000	1.000	0.976
0.50	0.250	1.00	121	200	2	0.058	0.058	7	1.000	1.000	1.000
0.50	0.375	1.00	121	200	2	0.066	0.063	7	1.000	1.000	1.000
0.50	0.500	1.00	121	200	2	0.320	0.297	7	1.000	1.000	1.000
0.50	0.125	1.00	169	50	2	0.045	0.045	7	0.645	0.591	0.491
0.50	0.250	1.00	169	50	2	0.058	0.057	7	1.000	1.000	1.000
0.50	0.375	1.00	169	50	2	0.065	0.059	7	1.000	1.000	1.000
0.50	0.500	1.00	169	50	2	0.114	0.093	7	1.000	1.000	1.000
0.50	0.125	1.00	169	100	2	0.046	0.044	7	0.970	0.968	0.849
0.50	0.250	1.00	169	100	2	0.053	0.052	7	1.000	1.000	1.000
0.50	0.375	1.00	169	100	2	0.070	0.068	7	1.000	1.000	1.000
0.50	0.500	1.00	169	100	2	0.158	0.144	7	1.000	1.000	1.000
0.50	0.125	1.00	169	150	2	0.060	0.058	7	0.999	0.999	0.984
0.50	0.250	1.00	169	150	2	0.042	0.042	7	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.375	1.00	169	150	2	0.073	0.072	7	1.000	1.000	1.000
0.50	0.500	1.00	169	150	2	0.206	0.183	7	1.000	1.000	1.000
0.50	0.125	1.00	169	200	2	0.042	0.040	7	1.000	1.000	0.999
0.50	0.250	1.00	169	200	2	0.060	0.060	7	1.000	1.000	1.000
0.50	0.375	1.00	169	200	2	0.059	0.056	7	1.000	1.000	1.000
0.50	0.500	1.00	169	200	2	0.220	0.201	7	1.000	1.000	1.000
0.50	0.125	1.00	25	50	3	0.067	0.069	3	0.274	0.270	0.291
0.50	0.250	1.00	25	50	3	0.089	0.090	3	0.919	0.921	0.922
0.50	0.375	1.00	25	50	3	0.141	0.142	3	1.000	1.000	1.000
0.50	0.500	1.00	25	50	3	0.343	0.344	3	1.000	1.000	1.000
0.50	0.125	1.00	25	100	3	0.073	0.071	3	0.585	0.587	0.597
0.50	0.250	1.00	25	100	3	0.100	0.099	3	0.999	0.999	0.999
0.50	0.375	1.00	25	100	3	0.144	0.145	3	1.000	1.000	1.000
0.50	0.500	1.00	25	100	3	0.393	0.395	3	1.000	1.000	1.000
0.50	0.125	1.00	25	150	3	0.072	0.074	3	0.826	0.826	0.831
0.50	0.250	1.00	25	150	3	0.092	0.094	3	1.000	1.000	1.000
0.50	0.375	1.00	25	150	3	0.175	0.174	3	1.000	1.000	1.000
0.50	0.500	1.00	25	150	3	0.421	0.416	3	1.000	1.000	1.000
0.50	0.125	1.00	25	200	3	0.072	0.071	3	0.917	0.916	0.927
0.50	0.250	1.00	25	200	3	0.088	0.089	3	1.000	1.000	1.000
0.50	0.375	1.00	25	200	3	0.134	0.135	3	1.000	1.000	1.000
0.50	0.500	1.00	25	200	3	0.399	0.400	3	1.000	1.000	1.000
0.50	0.125	1.00	49	50	3	0.074	0.075	3	0.505	0.500	0.530
0.50	0.250	1.00	49	50	3	0.085	0.083	3	0.999	0.999	1.000
0.50	0.375	1.00	49	50	3	0.136	0.139	3	1.000	1.000	1.000
0.50	0.500	1.00	49	50	3	0.465	0.463	3	1.000	1.000	1.000
0.50	0.125	1.00	49	100	3	0.075	0.072	3	0.872	0.870	0.878
0.50	0.250	1.00	49	100	3	0.088	0.088	3	1.000	1.000	1.000
0.50	0.375	1.00	49	100	3	0.128	0.127	3	1.000	1.000	1.000
0.50	0.500	1.00	49	100	3	0.520	0.518	3	1.000	1.000	1.000
0.50	0.125	1.00	49	150	3	0.058	0.059	3	0.986	0.987	0.987
0.50	0.250	1.00	49	150	3	0.097	0.096	3	1.000	1.000	1.000
0.50	0.375	1.00	49	150	3	0.174	0.175	3	1.000	1.000	1.000
0.50	0.500	1.00	49	150	3	0.564	0.566	3	1.000	1.000	1.000
0.50	0.125	1.00	49	200	3	0.085	0.081	3	0.998	0.998	0.998
0.50	0.250	1.00	49	200	3	0.100	0.101	3	1.000	1.000	1.000
0.50	0.375	1.00	49	200	3	0.144	0.143	3	1.000	1.000	1.000
0.50	0.500	1.00	49	200	3	0.577	0.580	3	1.000	1.000	1.000
0.50	0.125	1.00	81	50	3	0.070	0.070	3	0.762	0.760	0.771
0.50	0.250	1.00	81	50	3	0.080	0.082	3	1.000	1.000	1.000
0.50	0.375	1.00	81	50	3	0.144	0.144	3	1.000	1.000	1.000
0.50	0.500	1.00	81	50	3	0.596	0.603	3	1.000	1.000	1.000
0.50	0.125	1.00	81	100	3	0.049	0.050	3	0.986	0.986	0.985
0.50	0.250	1.00	81	100	3	0.077	0.075	3	1.000	1.000	1.000
0.50	0.375	1.00	81	100	3	0.145	0.144	3	1.000	1.000	1.000
0.50	0.500	1.00	81	100	3	0.687	0.692	3	1.000	1.000	1.000
0.50	0.125	1.00	81	150	3	0.074	0.075	3	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.250	1.00	81	150	3	0.093	0.094	3	1.000	1.000	1.000
0.50	0.375	1.00	81	150	3	0.163	0.162	3	1.000	1.000	1.000
0.50	0.500	1.00	81	150	3	0.752	0.753	3	1.000	1.000	1.000
0.50	0.125	1.00	81	200	3	0.071	0.073	3	1.000	1.000	1.000
0.50	0.250	1.00	81	200	3	0.090	0.089	3	1.000	1.000	1.000
0.50	0.375	1.00	81	200	3	0.145	0.144	3	1.000	1.000	1.000
0.50	0.500	1.00	81	200	3	0.782	0.780	3	1.000	1.000	1.000
0.50	0.125	1.00	121	50	3	0.062	0.062	3	0.932	0.933	0.937
0.50	0.250	1.00	121	50	3	0.082	0.083	3	1.000	1.000	1.000
0.50	0.375	1.00	121	50	3	0.120	0.119	3	1.000	1.000	1.000
0.50	0.500	1.00	121	50	3	0.384	0.387	3	1.000	1.000	1.000
0.50	0.125	1.00	121	100	3	0.064	0.064	3	1.000	1.000	1.000
0.50	0.250	1.00	121	100	3	0.092	0.090	3	1.000	1.000	1.000
0.50	0.375	1.00	121	100	3	0.131	0.131	3	1.000	1.000	1.000
0.50	0.500	1.00	121	100	3	0.510	0.509	3	1.000	1.000	1.000
0.50	0.125	1.00	121	150	3	0.075	0.074	3	1.000	1.000	1.000
0.50	0.250	1.00	121	150	3	0.074	0.075	3	1.000	1.000	1.000
0.50	0.375	1.00	121	150	3	0.149	0.148	3	1.000	1.000	1.000
0.50	0.500	1.00	121	150	3	0.552	0.550	3	1.000	1.000	1.000
0.50	0.125	1.00	121	200	3	0.069	0.070	3	1.000	1.000	1.000
0.50	0.250	1.00	121	200	3	0.092	0.092	3	1.000	1.000	1.000
0.50	0.375	1.00	121	200	3	0.111	0.112	3	1.000	1.000	1.000
0.50	0.500	1.00	121	200	3	0.586	0.588	3	1.000	1.000	1.000
0.50	0.125	1.00	169	50	3	0.062	0.062	3	0.990	0.990	0.993
0.50	0.250	1.00	169	50	3	0.069	0.067	3	1.000	1.000	1.000
0.50	0.375	1.00	169	50	3	0.091	0.091	3	1.000	1.000	1.000
0.50	0.500	1.00	169	50	3	0.397	0.395	3	1.000	1.000	1.000
0.50	0.125	1.00	169	100	3	0.058	0.058	3	1.000	1.000	1.000
0.50	0.250	1.00	169	100	3	0.071	0.072	3	1.000	1.000	1.000
0.50	0.375	1.00	169	100	3	0.095	0.094	3	1.000	1.000	1.000
0.50	0.500	1.00	169	100	3	0.552	0.551	3	1.000	1.000	1.000
0.50	0.125	1.00	169	150	3	0.072	0.071	3	1.000	1.000	1.000
0.50	0.250	1.00	169	150	3	0.076	0.075	3	1.000	1.000	1.000
0.50	0.375	1.00	169	150	3	0.095	0.093	3	1.000	1.000	1.000
0.50	0.500	1.00	169	150	3	0.635	0.635	3	1.000	1.000	1.000
0.50	0.125	1.00	169	200	3	0.065	0.064	3	1.000	1.000	1.000
0.50	0.250	1.00	169	200	3	0.076	0.075	3	1.000	1.000	1.000
0.50	0.375	1.00	169	200	3	0.110	0.110	3	1.000	1.000	1.000
0.50	0.500	1.00	169	200	3	0.670	0.671	3	1.000	1.000	1.000
0.50	0.125	1.00	25	50	3	0.080	0.057	5	0.184	0.125	0.144
0.50	0.250	1.00	25	50	3	0.097	0.068	5	0.738	0.681	0.659
0.50	0.375	1.00	25	50	3	0.136	0.113	5	0.996	0.993	0.994
0.50	0.500	1.00	25	50	3	0.267	0.228	5	1.000	1.000	1.000
0.50	0.125	1.00	25	100	3	0.075	0.066	5	0.322	0.288	0.271
0.50	0.250	1.00	25	100	3	0.092	0.080	5	0.981	0.980	0.953
0.50	0.375	1.00	25	100	3	0.121	0.103	5	1.000	1.000	1.000
0.50	0.500	1.00	25	100	3	0.305	0.271	5	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.125	1.00	25	150	3	0.071	0.067	5	0.535	0.522	0.414
0.50	0.250	1.00	25	150	3	0.069	0.065	5	1.000	1.000	0.998
0.50	0.375	1.00	25	150	3	0.113	0.108	5	1.000	1.000	1.000
0.50	0.500	1.00	25	150	3	0.298	0.270	5	1.000	1.000	1.000
0.50	0.125	1.00	25	200	3	0.061	0.060	5	0.670	0.662	0.586
0.50	0.250	1.00	25	200	3	0.063	0.057	5	1.000	1.000	1.000
0.50	0.375	1.00	25	200	3	0.117	0.113	5	1.000	1.000	1.000
0.50	0.500	1.00	25	200	3	0.301	0.280	5	1.000	1.000	1.000
0.50	0.125	1.00	49	50	3	0.071	0.069	5	0.301	0.259	0.247
0.50	0.250	1.00	49	50	3	0.071	0.066	5	0.959	0.957	0.931
0.50	0.375	1.00	49	50	3	0.090	0.085	5	1.000	1.000	1.000
0.50	0.500	1.00	49	50	3	0.334	0.307	5	1.000	1.000	1.000
0.50	0.125	1.00	49	100	3	0.063	0.060	5	0.603	0.590	0.513
0.50	0.250	1.00	49	100	3	0.070	0.063	5	1.000	1.000	1.000
0.50	0.375	1.00	49	100	3	0.090	0.086	5	1.000	1.000	1.000
0.50	0.500	1.00	49	100	3	0.413	0.380	5	1.000	1.000	1.000
0.50	0.125	1.00	49	150	3	0.059	0.061	5	0.863	0.859	0.749
0.50	0.250	1.00	49	150	3	0.078	0.074	5	1.000	1.000	1.000
0.50	0.375	1.00	49	150	3	0.109	0.106	5	1.000	1.000	1.000
0.50	0.500	1.00	49	150	3	0.426	0.410	5	1.000	1.000	1.000
0.50	0.125	1.00	49	200	3	0.067	0.065	5	0.965	0.965	0.893
0.50	0.250	1.00	49	200	3	0.072	0.068	5	1.000	1.000	1.000
0.50	0.375	1.00	49	200	3	0.099	0.097	5	1.000	1.000	1.000
0.50	0.500	1.00	49	200	3	0.447	0.438	5	1.000	1.000	1.000
0.50	0.125	1.00	81	50	3	0.059	0.053	5	0.435	0.424	0.376
0.50	0.250	1.00	81	50	3	0.076	0.074	5	0.998	0.998	0.994
0.50	0.375	1.00	81	50	3	0.094	0.091	5	1.000	1.000	1.000
0.50	0.500	1.00	81	50	3	0.420	0.401	5	1.000	1.000	1.000
0.50	0.125	1.00	81	100	3	0.058	0.058	5	0.876	0.876	0.781
0.50	0.250	1.00	81	100	3	0.066	0.065	5	1.000	1.000	1.000
0.50	0.375	1.00	81	100	3	0.100	0.099	5	1.000	1.000	1.000
0.50	0.500	1.00	81	100	3	0.537	0.532	5	1.000	1.000	1.000
0.50	0.125	1.00	81	150	3	0.056	0.055	5	0.985	0.983	0.944
0.50	0.250	1.00	81	150	3	0.063	0.058	5	1.000	1.000	1.000
0.50	0.375	1.00	81	150	3	0.094	0.095	5	1.000	1.000	1.000
0.50	0.500	1.00	81	150	3	0.600	0.592	5	1.000	1.000	1.000
0.50	0.125	1.00	81	200	3	0.054	0.055	5	0.998	0.998	0.993
0.50	0.250	1.00	81	200	3	0.065	0.062	5	1.000	1.000	1.000
0.50	0.375	1.00	81	200	3	0.085	0.083	5	1.000	1.000	1.000
0.50	0.500	1.00	81	200	3	0.651	0.643	5	1.000	1.000	1.000
0.50	0.125	1.00	121	50	3	0.058	0.053	5	0.705	0.695	0.587
0.50	0.250	1.00	121	50	3	0.066	0.062	5	1.000	1.000	1.000
0.50	0.375	1.00	121	50	3	0.090	0.083	5	1.000	1.000	1.000
0.50	0.500	1.00	121	50	3	0.269	0.250	5	1.000	1.000	1.000
0.50	0.125	1.00	121	100	3	0.054	0.055	5	0.982	0.984	0.943
0.50	0.250	1.00	121	100	3	0.054	0.055	5	1.000	1.000	1.000
0.50	0.375	1.00	121	100	3	0.088	0.087	5	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.500	1.00	121	100	3	0.367	0.345	5	1.000	1.000	1.000
0.50	0.125	1.00	121	150	3	0.054	0.053	5	0.999	0.999	0.994
0.50	0.250	1.00	121	150	3	0.068	0.068	5	1.000	1.000	1.000
0.50	0.375	1.00	121	150	3	0.097	0.096	5	1.000	1.000	1.000
0.50	0.500	1.00	121	150	3	0.408	0.398	5	1.000	1.000	1.000
0.50	0.125	1.00	121	200	3	0.061	0.059	5	1.000	1.000	1.000
0.50	0.250	1.00	121	200	3	0.059	0.057	5	1.000	1.000	1.000
0.50	0.375	1.00	121	200	3	0.090	0.088	5	1.000	1.000	1.000
0.50	0.500	1.00	121	200	3	0.459	0.449	5	1.000	1.000	1.000
0.50	0.125	1.00	169	50	3	0.050	0.050	5	0.864	0.855	0.746
0.50	0.250	1.00	169	50	3	0.060	0.060	5	1.000	1.000	1.000
0.50	0.375	1.00	169	50	3	0.074	0.067	5	1.000	1.000	1.000
0.50	0.500	1.00	169	50	3	0.262	0.246	5	1.000	1.000	1.000
0.50	0.125	1.00	169	100	3	0.061	0.061	5	0.999	0.999	0.995
0.50	0.250	1.00	169	100	3	0.065	0.066	5	1.000	1.000	1.000
0.50	0.375	1.00	169	100	3	0.062	0.060	5	1.000	1.000	1.000
0.50	0.500	1.00	169	100	3	0.408	0.392	5	1.000	1.000	1.000
0.50	0.125	1.00	169	150	3	0.047	0.047	5	1.000	1.000	1.000
0.50	0.250	1.00	169	150	3	0.044	0.045	5	1.000	1.000	1.000
0.50	0.375	1.00	169	150	3	0.081	0.078	5	1.000	1.000	1.000
0.50	0.500	1.00	169	150	3	0.447	0.433	5	1.000	1.000	1.000
0.50	0.125	1.00	169	200	3	0.067	0.065	5	1.000	1.000	1.000
0.50	0.250	1.00	169	200	3	0.061	0.063	5	1.000	1.000	1.000
0.50	0.375	1.00	169	200	3	0.083	0.084	5	1.000	1.000	1.000
0.50	0.500	1.00	169	200	3	0.530	0.527	5	1.000	1.000	1.000
0.50	0.125	1.00	25	50	3	0.105	0.051	7	0.402	0.108	0.160
0.50	0.250	1.00	25	50	3	0.118	0.060	7	0.775	0.435	0.570
0.50	0.375	1.00	25	50	3	0.181	0.104	7	0.990	0.964	0.952
0.50	0.500	1.00	25	50	3	0.262	0.180	7	1.000	1.000	0.986
0.50	0.125	1.00	25	100	3	0.082	0.055	7	0.322	0.163	0.222
0.50	0.250	1.00	25	100	3	0.090	0.065	7	0.935	0.874	0.867
0.50	0.375	1.00	25	100	3	0.133	0.079	7	1.000	1.000	1.000
0.50	0.500	1.00	25	100	3	0.276	0.209	7	1.000	1.000	1.000
0.50	0.125	1.00	25	150	3	0.071	0.060	7	0.410	0.283	0.281
0.50	0.250	1.00	25	150	3	0.092	0.070	7	0.993	0.985	0.963
0.50	0.375	1.00	25	150	3	0.113	0.084	7	1.000	1.000	1.000
0.50	0.500	1.00	25	150	3	0.283	0.211	7	1.000	1.000	1.000
0.50	0.125	1.00	25	200	3	0.067	0.059	7	0.488	0.389	0.356
0.50	0.250	1.00	25	200	3	0.085	0.065	7	1.000	1.000	0.994
0.50	0.375	1.00	25	200	3	0.095	0.070	7	1.000	1.000	1.000
0.50	0.500	1.00	25	200	3	0.277	0.223	7	1.000	1.000	1.000
0.50	0.125	1.00	49	50	3	0.092	0.056	7	0.351	0.176	0.210
0.50	0.250	1.00	49	50	3	0.081	0.057	7	0.887	0.770	0.825
0.50	0.375	1.00	49	50	3	0.116	0.082	7	1.000	0.999	1.000
0.50	0.500	1.00	49	50	3	0.282	0.221	7	1.000	1.000	1.000
0.50	0.125	1.00	49	100	3	0.049	0.043	7	0.425	0.324	0.306
0.50	0.250	1.00	49	100	3	0.074	0.062	7	0.996	0.995	0.988

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.375	1.00	49	100	3	0.101	0.085	7	1.000	1.000	1.000
0.50	0.500	1.00	49	100	3	0.318	0.273	7	1.000	1.000	1.000
0.50	0.125	1.00	49	150	3	0.073	0.061	7	0.642	0.567	0.472
0.50	0.250	1.00	49	150	3	0.079	0.069	7	1.000	1.000	1.000
0.50	0.375	1.00	49	150	3	0.087	0.070	7	1.000	1.000	1.000
0.50	0.500	1.00	49	150	3	0.335	0.295	7	1.000	1.000	1.000
0.50	0.125	1.00	49	200	3	0.059	0.048	7	0.794	0.754	0.606
0.50	0.250	1.00	49	200	3	0.079	0.069	7	1.000	1.000	1.000
0.50	0.375	1.00	49	200	3	0.084	0.077	7	1.000	1.000	1.000
0.50	0.500	1.00	49	200	3	0.374	0.337	7	1.000	1.000	1.000
0.50	0.125	1.00	81	50	3	0.076	0.065	7	0.370	0.255	0.271
0.50	0.250	1.00	81	50	3	0.078	0.057	7	0.979	0.967	0.951
0.50	0.375	1.00	81	50	3	0.094	0.075	7	1.000	1.000	1.000
0.50	0.500	1.00	81	50	3	0.354	0.302	7	1.000	1.000	1.000
0.50	0.125	1.00	81	100	3	0.060	0.053	7	0.681	0.615	0.477
0.50	0.250	1.00	81	100	3	0.079	0.075	7	1.000	1.000	1.000
0.50	0.375	1.00	81	100	3	0.103	0.092	7	1.000	1.000	1.000
0.50	0.500	1.00	81	100	3	0.438	0.395	7	1.000	1.000	1.000
0.50	0.125	1.00	81	150	3	0.048	0.043	7	0.884	0.865	0.706
0.50	0.250	1.00	81	150	3	0.047	0.040	7	1.000	1.000	1.000
0.50	0.375	1.00	81	150	3	0.082	0.079	7	1.000	1.000	1.000
0.50	0.500	1.00	81	150	3	0.488	0.454	7	1.000	1.000	1.000
0.50	0.125	1.00	81	200	3	0.052	0.050	7	0.972	0.967	0.868
0.50	0.250	1.00	81	200	3	0.057	0.055	7	1.000	1.000	1.000
0.50	0.375	1.00	81	200	3	0.093	0.086	7	1.000	1.000	1.000
0.50	0.500	1.00	81	200	3	0.531	0.507	7	1.000	1.000	1.000
0.50	0.125	1.00	121	50	3	0.074	0.064	7	0.519	0.438	0.371
0.50	0.250	1.00	121	50	3	0.069	0.061	7	0.999	0.999	0.997
0.50	0.375	1.00	121	50	3	0.081	0.069	7	1.000	1.000	1.000
0.50	0.500	1.00	121	50	3	0.225	0.183	7	1.000	1.000	1.000
0.50	0.125	1.00	121	100	3	0.055	0.053	7	0.863	0.848	0.682
0.50	0.250	1.00	121	100	3	0.065	0.060	7	1.000	1.000	1.000
0.50	0.375	1.00	121	100	3	0.065	0.059	7	1.000	1.000	1.000
0.50	0.500	1.00	121	100	3	0.280	0.236	7	1.000	1.000	1.000
0.50	0.125	1.00	121	150	3	0.064	0.063	7	0.984	0.982	0.888
0.50	0.250	1.00	121	150	3	0.064	0.061	7	1.000	1.000	1.000
0.50	0.375	1.00	121	150	3	0.085	0.079	7	1.000	1.000	1.000
0.50	0.500	1.00	121	150	3	0.321	0.294	7	1.000	1.000	1.000
0.50	0.125	1.00	121	200	3	0.065	0.059	7	1.000	1.000	0.976
0.50	0.250	1.00	121	200	3	0.052	0.052	7	1.000	1.000	1.000
0.50	0.375	1.00	121	200	3	0.067	0.063	7	1.000	1.000	1.000
0.50	0.500	1.00	121	200	3	0.354	0.319	7	1.000	1.000	1.000
0.50	0.125	1.00	169	50	3	0.069	0.062	7	0.645	0.591	0.491
0.50	0.250	1.00	169	50	3	0.051	0.044	7	1.000	1.000	1.000
0.50	0.375	1.00	169	50	3	0.077	0.068	7	1.000	1.000	1.000
0.50	0.500	1.00	169	50	3	0.192	0.150	7	1.000	1.000	1.000
0.50	0.125	1.00	169	100	3	0.056	0.051	7	0.970	0.968	0.849

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.250	1.00	169	100	3	0.054	0.054	7	1.000	1.000	1.000
0.50	0.375	1.00	169	100	3	0.065	0.066	7	1.000	1.000	1.000
0.50	0.500	1.00	169	100	3	0.313	0.286	7	1.000	1.000	1.000
0.50	0.125	1.00	169	150	3	0.055	0.052	7	0.999	0.999	0.984
0.50	0.250	1.00	169	150	3	0.044	0.044	7	1.000	1.000	1.000
0.50	0.375	1.00	169	150	3	0.066	0.063	7	1.000	1.000	1.000
0.50	0.500	1.00	169	150	3	0.357	0.331	7	1.000	1.000	1.000
0.50	0.125	1.00	169	200	3	0.052	0.052	7	1.000	1.000	0.999
0.50	0.250	1.00	169	200	3	0.061	0.058	7	1.000	1.000	1.000
0.50	0.375	1.00	169	200	3	0.067	0.066	7	1.000	1.000	1.000
0.50	0.500	1.00	169	200	3	0.394	0.374	7	1.000	1.000	1.000
0.50	0.125	1.00	25	50	4	0.070	0.066	3	0.274	0.270	0.291
0.50	0.250	1.00	25	50	4	0.077	0.073	3	0.919	0.921	0.922
0.50	0.375	1.00	25	50	4	0.089	0.087	3	1.000	1.000	1.000
0.50	0.500	1.00	25	50	4	0.219	0.209	3	1.000	1.000	1.000
0.50	0.125	1.00	25	100	4	0.073	0.074	3	0.585	0.587	0.597
0.50	0.250	1.00	25	100	4	0.075	0.070	3	0.999	0.999	0.999
0.50	0.375	1.00	25	100	4	0.103	0.096	3	1.000	1.000	1.000
0.50	0.500	1.00	25	100	4	0.202	0.204	3	1.000	1.000	1.000
0.50	0.125	1.00	25	150	4	0.061	0.059	3	0.826	0.826	0.831
0.50	0.250	1.00	25	150	4	0.070	0.067	3	1.000	1.000	1.000
0.50	0.375	1.00	25	150	4	0.105	0.103	3	1.000	1.000	1.000
0.50	0.500	1.00	25	150	4	0.233	0.237	3	1.000	1.000	1.000
0.50	0.125	1.00	25	200	4	0.064	0.067	3	0.917	0.916	0.927
0.50	0.250	1.00	25	200	4	0.071	0.070	3	1.000	1.000	1.000
0.50	0.375	1.00	25	200	4	0.083	0.084	3	1.000	1.000	1.000
0.50	0.500	1.00	25	200	4	0.210	0.204	3	1.000	1.000	1.000
0.50	0.125	1.00	49	50	4	0.083	0.084	3	0.505	0.500	0.530
0.50	0.250	1.00	49	50	4	0.100	0.096	3	0.999	0.999	1.000
0.50	0.375	1.00	49	50	4	0.130	0.129	3	1.000	1.000	1.000
0.50	0.500	1.00	49	50	4	0.477	0.484	3	1.000	1.000	1.000
0.50	0.125	1.00	49	100	4	0.069	0.072	3	0.872	0.870	0.878
0.50	0.250	1.00	49	100	4	0.074	0.071	3	1.000	1.000	1.000
0.50	0.375	1.00	49	100	4	0.133	0.131	3	1.000	1.000	1.000
0.50	0.500	1.00	49	100	4	0.504	0.506	3	1.000	1.000	1.000
0.50	0.125	1.00	49	150	4	0.074	0.073	3	0.986	0.987	0.987
0.50	0.250	1.00	49	150	4	0.084	0.084	3	1.000	1.000	1.000
0.50	0.375	1.00	49	150	4	0.148	0.146	3	1.000	1.000	1.000
0.50	0.500	1.00	49	150	4	0.566	0.569	3	1.000	1.000	1.000
0.50	0.125	1.00	49	200	4	0.079	0.074	3	0.998	0.998	0.998
0.50	0.250	1.00	49	200	4	0.085	0.086	3	1.000	1.000	1.000
0.50	0.375	1.00	49	200	4	0.134	0.135	3	1.000	1.000	1.000
0.50	0.500	1.00	49	200	4	0.560	0.558	3	1.000	1.000	1.000
0.50	0.125	1.00	81	50	4	0.065	0.067	3	0.762	0.760	0.771
0.50	0.250	1.00	81	50	4	0.087	0.085	3	1.000	1.000	1.000
0.50	0.375	1.00	81	50	4	0.148	0.149	3	1.000	1.000	1.000
0.50	0.500	1.00	81	50	4	0.575	0.576	3	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.125	1.00	81	100	4	0.075	0.072	3	0.986	0.986	0.985
0.50	0.250	1.00	81	100	4	0.092	0.092	3	1.000	1.000	1.000
0.50	0.375	1.00	81	100	4	0.164	0.163	3	1.000	1.000	1.000
0.50	0.500	1.00	81	100	4	0.658	0.660	3	1.000	1.000	1.000
0.50	0.125	1.00	81	150	4	0.085	0.085	3	1.000	1.000	1.000
0.50	0.250	1.00	81	150	4	0.091	0.091	3	1.000	1.000	1.000
0.50	0.375	1.00	81	150	4	0.159	0.159	3	1.000	1.000	1.000
0.50	0.500	1.00	81	150	4	0.735	0.733	3	1.000	1.000	1.000
0.50	0.125	1.00	81	200	4	0.069	0.073	3	1.000	1.000	1.000
0.50	0.250	1.00	81	200	4	0.090	0.090	3	1.000	1.000	1.000
0.50	0.375	1.00	81	200	4	0.153	0.154	3	1.000	1.000	1.000
0.50	0.500	1.00	81	200	4	0.750	0.749	3	1.000	1.000	1.000
0.50	0.125	1.00	121	50	4	0.088	0.087	3	0.932	0.933	0.937
0.50	0.250	1.00	121	50	4	0.087	0.085	3	1.000	1.000	1.000
0.50	0.375	1.00	121	50	4	0.131	0.130	3	1.000	1.000	1.000
0.50	0.500	1.00	121	50	4	0.623	0.623	3	1.000	1.000	1.000
0.50	0.125	1.00	121	100	4	0.077	0.077	3	1.000	1.000	1.000
0.50	0.250	1.00	121	100	4	0.090	0.089	3	1.000	1.000	1.000
0.50	0.375	1.00	121	100	4	0.144	0.145	3	1.000	1.000	1.000
0.50	0.500	1.00	121	100	4	0.748	0.749	3	1.000	1.000	1.000
0.50	0.125	1.00	121	150	4	0.070	0.068	3	1.000	1.000	1.000
0.50	0.250	1.00	121	150	4	0.083	0.083	3	1.000	1.000	1.000
0.50	0.375	1.00	121	150	4	0.167	0.167	3	1.000	1.000	1.000
0.50	0.500	1.00	121	150	4	0.793	0.795	3	1.000	1.000	1.000
0.50	0.125	1.00	121	200	4	0.069	0.068	3	1.000	1.000	1.000
0.50	0.250	1.00	121	200	4	0.070	0.069	3	1.000	1.000	1.000
0.50	0.375	1.00	121	200	4	0.129	0.131	3	1.000	1.000	1.000
0.50	0.500	1.00	121	200	4	0.823	0.820	3	1.000	1.000	1.000
0.50	0.125	1.00	169	50	4	0.070	0.071	3	0.990	0.990	0.993
0.50	0.250	1.00	169	50	4	0.088	0.087	3	1.000	1.000	1.000
0.50	0.375	1.00	169	50	4	0.123	0.122	3	1.000	1.000	1.000
0.50	0.500	1.00	169	50	4	0.443	0.444	3	1.000	1.000	1.000
0.50	0.125	1.00	169	100	4	0.065	0.065	3	1.000	1.000	1.000
0.50	0.250	1.00	169	100	4	0.080	0.080	3	1.000	1.000	1.000
0.50	0.375	1.00	169	100	4	0.114	0.116	3	1.000	1.000	1.000
0.50	0.500	1.00	169	100	4	0.605	0.605	3	1.000	1.000	1.000
0.50	0.125	1.00	169	150	4	0.058	0.057	3	1.000	1.000	1.000
0.50	0.250	1.00	169	150	4	0.060	0.060	3	1.000	1.000	1.000
0.50	0.375	1.00	169	150	4	0.109	0.111	3	1.000	1.000	1.000
0.50	0.500	1.00	169	150	4	0.694	0.694	3	1.000	1.000	1.000
0.50	0.125	1.00	169	200	4	0.076	0.076	3	1.000	1.000	1.000
0.50	0.250	1.00	169	200	4	0.107	0.107	3	1.000	1.000	1.000
0.50	0.375	1.00	169	200	4	0.128	0.129	3	1.000	1.000	1.000
0.50	0.500	1.00	169	200	4	0.716	0.717	3	1.000	1.000	1.000
0.50	0.125	1.00	25	50	4	0.084	0.067	5	0.184	0.125	0.144
0.50	0.250	1.00	25	50	4	0.091	0.065	5	0.738	0.681	0.659
0.50	0.375	1.00	25	50	4	0.112	0.082	5	0.996	0.993	0.994

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.500	1.00	25	50	4	0.158	0.115	5	1.000	1.000	1.000
0.50	0.125	1.00	25	100	4	0.068	0.058	5	0.322	0.288	0.271
0.50	0.250	1.00	25	100	4	0.073	0.061	5	0.981	0.980	0.953
0.50	0.375	1.00	25	100	4	0.086	0.069	5	1.000	1.000	1.000
0.50	0.500	1.00	25	100	4	0.173	0.138	5	1.000	1.000	1.000
0.50	0.125	1.00	25	150	4	0.058	0.055	5	0.535	0.522	0.414
0.50	0.250	1.00	25	150	4	0.060	0.051	5	1.000	1.000	0.998
0.50	0.375	1.00	25	150	4	0.089	0.077	5	1.000	1.000	1.000
0.50	0.500	1.00	25	150	4	0.166	0.135	5	1.000	1.000	1.000
0.50	0.125	1.00	25	200	4	0.061	0.057	5	0.670	0.662	0.586
0.50	0.250	1.00	25	200	4	0.072	0.067	5	1.000	1.000	1.000
0.50	0.375	1.00	25	200	4	0.075	0.075	5	1.000	1.000	1.000
0.50	0.500	1.00	25	200	4	0.163	0.140	5	1.000	1.000	1.000
0.50	0.125	1.00	49	50	4	0.071	0.063	5	0.301	0.259	0.247
0.50	0.250	1.00	49	50	4	0.071	0.064	5	0.959	0.957	0.931
0.50	0.375	1.00	49	50	4	0.119	0.098	5	1.000	1.000	1.000
0.50	0.500	1.00	49	50	4	0.323	0.298	5	1.000	1.000	1.000
0.50	0.125	1.00	49	100	4	0.062	0.059	5	0.603	0.590	0.513
0.50	0.250	1.00	49	100	4	0.069	0.064	5	1.000	1.000	1.000
0.50	0.375	1.00	49	100	4	0.107	0.088	5	1.000	1.000	1.000
0.50	0.500	1.00	49	100	4	0.388	0.370	5	1.000	1.000	1.000
0.50	0.125	1.00	49	150	4	0.070	0.066	5	0.863	0.859	0.749
0.50	0.250	1.00	49	150	4	0.068	0.063	5	1.000	1.000	1.000
0.50	0.375	1.00	49	150	4	0.091	0.084	5	1.000	1.000	1.000
0.50	0.500	1.00	49	150	4	0.450	0.427	5	1.000	1.000	1.000
0.50	0.125	1.00	49	200	4	0.051	0.048	5	0.965	0.965	0.893
0.50	0.250	1.00	49	200	4	0.063	0.060	5	1.000	1.000	1.000
0.50	0.375	1.00	49	200	4	0.103	0.100	5	1.000	1.000	1.000
0.50	0.500	1.00	49	200	4	0.429	0.413	5	1.000	1.000	1.000
0.50	0.125	1.00	81	50	4	0.072	0.065	5	0.435	0.424	0.376
0.50	0.250	1.00	81	50	4	0.078	0.068	5	0.998	0.998	0.994
0.50	0.375	1.00	81	50	4	0.105	0.100	5	1.000	1.000	1.000
0.50	0.500	1.00	81	50	4	0.421	0.401	5	1.000	1.000	1.000
0.50	0.125	1.00	81	100	4	0.053	0.052	5	0.876	0.876	0.781
0.50	0.250	1.00	81	100	4	0.066	0.064	5	1.000	1.000	1.000
0.50	0.375	1.00	81	100	4	0.102	0.094	5	1.000	1.000	1.000
0.50	0.500	1.00	81	100	4	0.510	0.484	5	1.000	1.000	1.000
0.50	0.125	1.00	81	150	4	0.061	0.057	5	0.985	0.983	0.944
0.50	0.250	1.00	81	150	4	0.059	0.056	5	1.000	1.000	1.000
0.50	0.375	1.00	81	150	4	0.101	0.097	5	1.000	1.000	1.000
0.50	0.500	1.00	81	150	4	0.576	0.557	5	1.000	1.000	1.000
0.50	0.125	1.00	81	200	4	0.060	0.063	5	0.998	0.998	0.993
0.50	0.250	1.00	81	200	4	0.068	0.065	5	1.000	1.000	1.000
0.50	0.375	1.00	81	200	4	0.113	0.108	5	1.000	1.000	1.000
0.50	0.500	1.00	81	200	4	0.600	0.596	5	1.000	1.000	1.000
0.50	0.125	1.00	121	50	4	0.062	0.061	5	0.705	0.695	0.587
0.50	0.250	1.00	121	50	4	0.076	0.071	5	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.375	1.00	121	50	4	0.092	0.087	5	1.000	1.000	1.000
0.50	0.500	1.00	121	50	4	0.491	0.447	5	1.000	1.000	1.000
0.50	0.125	1.00	121	100	4	0.054	0.054	5	0.982	0.984	0.943
0.50	0.250	1.00	121	100	4	0.070	0.069	5	1.000	1.000	1.000
0.50	0.375	1.00	121	100	4	0.101	0.100	5	1.000	1.000	1.000
0.50	0.500	1.00	121	100	4	0.615	0.595	5	1.000	1.000	1.000
0.50	0.125	1.00	121	150	4	0.056	0.057	5	0.999	0.999	0.994
0.50	0.250	1.00	121	150	4	0.059	0.057	5	1.000	1.000	1.000
0.50	0.375	1.00	121	150	4	0.106	0.103	5	1.000	1.000	1.000
0.50	0.500	1.00	121	150	4	0.660	0.648	5	1.000	1.000	1.000
0.50	0.125	1.00	121	200	4	0.067	0.067	5	1.000	1.000	1.000
0.50	0.250	1.00	121	200	4	0.073	0.073	5	1.000	1.000	1.000
0.50	0.375	1.00	121	200	4	0.111	0.110	5	1.000	1.000	1.000
0.50	0.500	1.00	121	200	4	0.727	0.717	5	1.000	1.000	1.000
0.50	0.125	1.00	169	50	4	0.054	0.053	5	0.864	0.855	0.746
0.50	0.250	1.00	169	50	4	0.072	0.071	5	1.000	1.000	1.000
0.50	0.375	1.00	169	50	4	0.095	0.092	5	1.000	1.000	1.000
0.50	0.500	1.00	169	50	4	0.308	0.277	5	1.000	1.000	1.000
0.50	0.125	1.00	169	100	4	0.052	0.055	5	0.999	0.999	0.995
0.50	0.250	1.00	169	100	4	0.062	0.060	5	1.000	1.000	1.000
0.50	0.375	1.00	169	100	4	0.073	0.071	5	1.000	1.000	1.000
0.50	0.500	1.00	169	100	4	0.464	0.447	5	1.000	1.000	1.000
0.50	0.125	1.00	169	150	4	0.047	0.047	5	1.000	1.000	1.000
0.50	0.250	1.00	169	150	4	0.054	0.055	5	1.000	1.000	1.000
0.50	0.375	1.00	169	150	4	0.091	0.086	5	1.000	1.000	1.000
0.50	0.500	1.00	169	150	4	0.519	0.503	5	1.000	1.000	1.000
0.50	0.125	1.00	169	200	4	0.062	0.059	5	1.000	1.000	1.000
0.50	0.250	1.00	169	200	4	0.074	0.073	5	1.000	1.000	1.000
0.50	0.375	1.00	169	200	4	0.099	0.098	5	1.000	1.000	1.000
0.50	0.500	1.00	169	200	4	0.551	0.554	5	1.000	1.000	1.000
0.50	0.125	1.00	25	50	4	0.156	0.055	7	0.402	0.108	0.160
0.50	0.250	1.00	25	50	4	0.149	0.065	7	0.775	0.435	0.570
0.50	0.375	1.00	25	50	4	0.181	0.071	7	0.990	0.964	0.952
0.50	0.500	1.00	25	50	4	0.214	0.102	7	1.000	1.000	0.986
0.50	0.125	1.00	25	100	4	0.092	0.053	7	0.322	0.163	0.222
0.50	0.250	1.00	25	100	4	0.094	0.063	7	0.935	0.874	0.867
0.50	0.375	1.00	25	100	4	0.115	0.067	7	1.000	1.000	1.000
0.50	0.500	1.00	25	100	4	0.195	0.108	7	1.000	1.000	1.000
0.50	0.125	1.00	25	150	4	0.082	0.054	7	0.410	0.283	0.281
0.50	0.250	1.00	25	150	4	0.100	0.072	7	0.993	0.985	0.963
0.50	0.375	1.00	25	150	4	0.111	0.076	7	1.000	1.000	1.000
0.50	0.500	1.00	25	150	4	0.171	0.104	7	1.000	1.000	1.000
0.50	0.125	1.00	25	200	4	0.074	0.053	7	0.488	0.389	0.356
0.50	0.250	1.00	25	200	4	0.093	0.072	7	1.000	1.000	0.994
0.50	0.375	1.00	25	200	4	0.091	0.066	7	1.000	1.000	1.000
0.50	0.500	1.00	25	200	4	0.153	0.108	7	1.000	1.000	1.000
0.50	0.125	1.00	49	50	4	0.097	0.060	7	0.351	0.176	0.210

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.250	1.00	49	50	4	0.101	0.067	7	0.887	0.770	0.825
0.50	0.375	1.00	49	50	4	0.135	0.078	7	1.000	0.999	1.000
0.50	0.500	1.00	49	50	4	0.308	0.224	7	1.000	1.000	1.000
0.50	0.125	1.00	49	100	4	0.067	0.052	7	0.425	0.324	0.306
0.50	0.250	1.00	49	100	4	0.076	0.054	7	0.996	0.995	0.988
0.50	0.375	1.00	49	100	4	0.112	0.095	7	1.000	1.000	1.000
0.50	0.500	1.00	49	100	4	0.318	0.272	7	1.000	1.000	1.000
0.50	0.125	1.00	49	150	4	0.082	0.069	7	0.642	0.567	0.472
0.50	0.250	1.00	49	150	4	0.070	0.061	7	1.000	1.000	1.000
0.50	0.375	1.00	49	150	4	0.106	0.077	7	1.000	1.000	1.000
0.50	0.500	1.00	49	150	4	0.357	0.308	7	1.000	1.000	1.000
0.50	0.125	1.00	49	200	4	0.061	0.048	7	0.794	0.754	0.606
0.50	0.250	1.00	49	200	4	0.068	0.060	7	1.000	1.000	1.000
0.50	0.375	1.00	49	200	4	0.102	0.093	7	1.000	1.000	1.000
0.50	0.500	1.00	49	200	4	0.392	0.323	7	1.000	1.000	1.000
0.50	0.125	1.00	81	50	4	0.088	0.065	7	0.370	0.255	0.271
0.50	0.250	1.00	81	50	4	0.086	0.071	7	0.979	0.967	0.951
0.50	0.375	1.00	81	50	4	0.112	0.074	7	1.000	1.000	1.000
0.50	0.500	1.00	81	50	4	0.363	0.313	7	1.000	1.000	1.000
0.50	0.125	1.00	81	100	4	0.074	0.065	7	0.681	0.615	0.477
0.50	0.250	1.00	81	100	4	0.071	0.057	7	1.000	1.000	1.000
0.50	0.375	1.00	81	100	4	0.104	0.086	7	1.000	1.000	1.000
0.50	0.500	1.00	81	100	4	0.415	0.374	7	1.000	1.000	1.000
0.50	0.125	1.00	81	150	4	0.062	0.056	7	0.884	0.865	0.706
0.50	0.250	1.00	81	150	4	0.061	0.054	7	1.000	1.000	1.000
0.50	0.375	1.00	81	150	4	0.091	0.071	7	1.000	1.000	1.000
0.50	0.500	1.00	81	150	4	0.475	0.445	7	1.000	1.000	1.000
0.50	0.125	1.00	81	200	4	0.062	0.062	7	0.972	0.967	0.868
0.50	0.250	1.00	81	200	4	0.063	0.058	7	1.000	1.000	1.000
0.50	0.375	1.00	81	200	4	0.101	0.091	7	1.000	1.000	1.000
0.50	0.500	1.00	81	200	4	0.481	0.447	7	1.000	1.000	1.000
0.50	0.125	1.00	121	50	4	0.078	0.070	7	0.519	0.438	0.371
0.50	0.250	1.00	121	50	4	0.061	0.051	7	0.999	0.999	0.997
0.50	0.375	1.00	121	50	4	0.098	0.078	7	1.000	1.000	1.000
0.50	0.500	1.00	121	50	4	0.402	0.343	7	1.000	1.000	1.000
0.50	0.125	1.00	121	100	4	0.059	0.056	7	0.863	0.848	0.682
0.50	0.250	1.00	121	100	4	0.065	0.054	7	1.000	1.000	1.000
0.50	0.375	1.00	121	100	4	0.084	0.063	7	1.000	1.000	1.000
0.50	0.500	1.00	121	100	4	0.482	0.449	7	1.000	1.000	1.000
0.50	0.125	1.00	121	150	4	0.053	0.044	7	0.984	0.982	0.888
0.50	0.250	1.00	121	150	4	0.079	0.078	7	1.000	1.000	1.000
0.50	0.375	1.00	121	150	4	0.083	0.083	7	1.000	1.000	1.000
0.50	0.500	1.00	121	150	4	0.560	0.527	7	1.000	1.000	1.000
0.50	0.125	1.00	121	200	4	0.056	0.052	7	1.000	1.000	0.976
0.50	0.250	1.00	121	200	4	0.057	0.053	7	1.000	1.000	1.000
0.50	0.375	1.00	121	200	4	0.084	0.084	7	1.000	1.000	1.000
0.50	0.500	1.00	121	200	4	0.609	0.576	7	1.000	1.000	1.000

Continued on next page

Table 7 continued

$\bar{\beta}$	$\bar{\lambda}$	$\bar{\mu}$	N	T	Het Regime	CSMC Het LR	CSMC Het Q	m	CSMC Dep LR	CSMC Dep Q	JSMC Dep
0.50	0.125	1.00	169	50	4	0.066	0.056	7	0.645	0.591	0.491
0.50	0.250	1.00	169	50	4	0.070	0.065	7	1.000	1.000	1.000
0.50	0.375	1.00	169	50	4	0.095	0.075	7	1.000	1.000	1.000
0.50	0.500	1.00	169	50	4	0.270	0.218	7	1.000	1.000	1.000
0.50	0.125	1.00	169	100	4	0.063	0.055	7	0.970	0.968	0.849
0.50	0.250	1.00	169	100	4	0.061	0.055	7	1.000	1.000	1.000
0.50	0.375	1.00	169	100	4	0.070	0.064	7	1.000	1.000	1.000
0.50	0.500	1.00	169	100	4	0.361	0.326	7	1.000	1.000	1.000
0.50	0.125	1.00	169	150	4	0.054	0.051	7	0.999	0.999	0.984
0.50	0.250	1.00	169	150	4	0.054	0.049	7	1.000	1.000	1.000
0.50	0.375	1.00	169	150	4	0.083	0.078	7	1.000	1.000	1.000
0.50	0.500	1.00	169	150	4	0.420	0.386	7	1.000	1.000	1.000
0.50	0.125	1.00	169	200	4	0.056	0.055	7	1.000	1.000	0.999
0.50	0.250	1.00	169	200	4	0.062	0.058	7	1.000	1.000	1.000
0.50	0.375	1.00	169	200	4	0.075	0.073	7	1.000	1.000	1.000
0.50	0.500	1.00	169	200	4	0.455	0.431	7	1.000	1.000	1.000

Chapter 4

SMOOTHED ESTIMATORS FOR MARKOV CHAINS WITH SPARSE SPATIAL OBSERVATIONS

Markov modeling has been applied to a wide array of domains including land use and land cover change, crime patterns and economic convergence, to obtain insights in the dynamics nature of the process under study and to investigate the equilibrium under the current dynamics (McMillen and McDonald, 1991; Quah, 1993a; Rey et al., 2014). A specific class of Markov models, discrete Markov chain, specify a k state classification of the phenomena under consideration. The dynamics of the transitions across these k states are summarized in a (k, k) transition probability matrix which is comprised of k^2 probabilities of transitioning from one state to another across two consecutive periods. Interest centers not only on the individual probability estimates of transitioning between a pair of states, but also a variety of parameters defined on these probabilities such as ergodic distributions, mobility indice, and hitting time. Therefore, estimators with sound statistical properties for the transition probabilities are desirable.

Maximum likelihood estimators (MLEs) are commonly used for estimating the transition probabilities in discrete Markov chain models in the spatial sciences. Although they have very nice *asymptotic* statistical properties, their usage in small sample settings could be problematic. One prominent issue is that MLEs could easily lead to a sparse transition probability matrix where 0 probability estimates are quite a few. On one hand, the interpretation of 0 probability estimates claims that transitioning between two states across consecutive periods is impossible, which is very different from that for nonzero ones, even though they might be quite small. On the other hand, excessive 0 probability estimates constitute a sparse transition probability matrix whose properties are quite different from a non-sparse one. This is especially true for some parameters derived from the matrix such as the steady state distribution and hitting time. A sparse transition probability matrix also raises issues for the bootstrap inference about Markov chains (Teodorescu, 2009; Polansky, 2009).

The sparsity issue becomes more relevant for extensions of the classic first-order Markov chain model which requires estimating a larger number of parameters, including the higher-order Markov chain model (Bickenbach and Bode, 2003), multivariate Markov chain model (Ching et al., 2002), spatial Markov chain model (Rey, 2001) and LISA Markov chain model. Focusing on the spatial Markov chain model which requires estimating k^3 transition probabilities, I find most empirical studies of regional income distribution dynamics employing this model produced a large portion of 0 transition probabilities, such as the U.S. (Rey, 2001), China (Pu et al., 2005) and Europe (Le Gallo, 2004; Maza et al., 2012). The high sparsity of the transition probability matrix could easily lead to false interpretation about the underlying dynamics, as well as the inference about the spatial Markov. Specifically, when sample size is small, the spatial Markov test which tests for spatial dependence in the Markov chain model would give rise to inflated Type I error rate because of the high sparsity of k spatially dependent (k, k) transition probability matrices. An ad-hoc approach to addressing the issue is to ignore the rows full of zero probability estimates and reduce the degree of freedom accordingly (Bickenbach and Bode, 2003; Rey et al., 2016; Kang and Rey, 2018). Though the issue seems to be alleviated, it is with the loss of potentially important information hidden in the abandoned rows. Further, the tests could still suffer from sparsity in the remaining rows. Therefore, seeking estimators which avoid producing too many 0 probability estimates when the sample size is small is important.

Teodorescu (2009) and Polansky (2009) applied multinomial smoothing techniques to address the Markov chain model's sparsity issue. Smoothing a multinomial distribution could be accomplished by "borrowing" information from other cells (Simonoff, 1995, 1998), which is similar to kernel density estimation for continuous data. The discrete kernel for smoothing the multinomial distribution can take various forms, and can also incorporate the ordinal structure of the categories if there is any. By considering each row of the transition probability matrix as a multinomial distribution, these two papers applied the discrete kernel estimator to each of them, thus producing a smoothed Markov transition probability matrix which is exempt from zero estimates. Both of these two papers demonstrated the superiority of the smoothed estimators to the conventional MLEs in small sample settings based on Monte Carlo simulations.

One shortcoming of these papers is that they did not come up with an effective way to select the smoothing parameter or kernel bandwidth which controls the degree of smoothness of the distribution. Similar to kernel density estimation for continuous data whose performance hinges on the choice of a bandwidth (Sheather and Jones, 1991; Henderson and Parmeter, 2015), smoothing parameter/bandwidth selection is also crucial to discrete kernel estimation and is data-dependent (Bowman et al., 1984; Ouyang et al., 2006; Chu et al., 2015). While smoothing tends to lower the variance of the probability estimator, it introduces bias. To balance the tradeoff between variance and bias, an appropriate smoothing parameter is vital. The other shortcoming is that smoothing is applied along the rows only although smoothing along the columns is just as feasible and could potentially improve the performance. I attempt to overcome these two shortcomings in this chapter. I utilize the cross-validation technique for smoothing parameter selection. I follow Kullback et al. (1962) and view the (k, k) transition matrix for classic Markov chain and the (k, k, k) transition matrix for spatial Markov chain as two-way and three-way *conditional* contingency tables in the sense that all the cells are filled with conditional rather than joint probabilities. Then I modify the smoothing techniques for high-order contingency tables and the relevant cross-validation technique for smoothing parameter selection to suit the *conditional* contingency tables for Markov and spatial Markov chain models.

Based on a series of Monte Carlo experiments with classic Markov chains and spatial Markov chains, I find that discrete kernel estimators with cross validation-based smoothing parameters have nice small sample properties and converge to MLEs when more spatial observations become available. They are quite effective in alleviating issues caused by the sparsity of the MLE-based transition probability matrix estimate. What's more, they seem also to be superior to MLEs in terms of minimizing mean squared error for individual and total transition probability as well as the irreducibility property, thus giving rise to a better recovery of the true underlying dynamics.

For the rest of the chapter, I first introduce the kernel smoothing techniques for high-order contingency tables and the least square cross validation technique for smoothing parameters selection. Then I explain how I adjust these techniques and apply them to the transition probability matrix for Markov and spatial Markov chain models. Next I introduce the Monte Carlo experiments I conducted to evaluate the properties of the smoothed estimators which were further compared to MLEs. I conclude the chapter with a summary of key findings and a discussion of future research directions.

4.1 Smoothing Estimators for Discrete Data

Smoothing techniques serve as an important approach to dealing with the sparsity of contingency tables when the number of categories is large compared with the sample size (Burman, 1987, 2004). Various strategies of smoothing have been proposed, including Bayes methods (Fienberg and Holland, 1973; Agresti and Hitchcock, 2005) and discrete kernels (Kokonendji and Kiessé, 2011). Take smoothing a multinomial distribution as an example. The Bayes method requires the assumption of a prior distribution with specified values of hyperparameters. Usually, a symmetric Dirichlet distribution is adopted, giving rise to the so-called *additive smoothing*. Specifying the hyperparameter α as 1 is basically increasing the sample size n to $n + k$ by adding 1 to each cell and then use MLEs for probability estimation. Other prior distributions such as logistic-normal and poisson-normal distributions were suggested to incorporate ordinal structure of categories if there exists one (Titterington and Bowman, 1985).

Compare to the Bayes smoothing methods, the logic of discrete kernels is similar to the continuous kernels which are more intuitive and have been widely applied to continuous density estimation. Moreover, its extension to joint probability estimation for several discrete variables (high-order contingency tables) is more tractable and so is the smoothing parameter selection algorithm. Therefore, I consider two discrete kernels for high-order contingency tables in this chapter and develop the smoothing notion towards estimating the transition probability matrix for Markov chains and spatial Markov chains.

4.1.1 Maximum Likelihood Estimators (MLEs)

Before introducing discrete kernels for smoothing, I first present the MLEs conventionally used for estimating joint probabilities for high-order contingency tables. Suppose we have n observations and each can be categorized based on d criteria of classification. Thus, each observation consists of d parts and could be represented by a row vector of length d. If we stack these vectors into a (n, d) matrix where n is the number of observations, then each column represents a multinomial random variable taking values in a sample space, e.g. $\mathbb{S}^s = \{1, 2, \ldots, k^s\}$ for the sth column. We name this matrix $\boldsymbol{X}$ and denote each **unique** category vector as $\boldsymbol{x}$. For the sth dimension/column $\boldsymbol{X}^s$, the MLE for the marginal probability of its jth category x_j^s is shown in Equation (4.1):

$$\hat{p}^{mle}(x_j^s) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i^s = x_j^s),\ x_j^s \in \mathbb{S}^s = \{1, 2, \ldots, k^s\}, \tag{4.1}$$

where $\mathbb{1}(X_i^s = x_j^s)$ is the indicator function which takes 1 if $X_i^s = x_j^s$ and 0 otherwise. Clearly, the marginal probability is the observed frequency of each category. Thus, the MLEs are also called frequency estimators (Ouyang et al., 2006).

Similarly, the MLE for the joint probability for the category vector $\boldsymbol{x}_j$ is the observed frequency of $\boldsymbol{x}_j$ which is shown in Equation (4.2):

$$\hat{p}^{mle}(\boldsymbol{x}_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\boldsymbol{X}_i = \boldsymbol{x}_j),\ \boldsymbol{x}_j \in \mathbb{S}^1 \times \mathbb{S}^2 \times \cdots \times \mathbb{S}^d. \tag{4.2}$$

4.1.2 Smoothing Joint Probabilities for High-Order Contingency Tables

4.1.2.1 Discrete Kernel Smoothers

I focus on two discrete kernels, one for pure categorical data (e.g. land use and land cover types) and the other for discrete data with an inherent ordinal structure (e.g. income levels). I start with utilizing them for smoothing one-way contingency tables (a multinomial distribution), and then extend each for smoothing d-way contingency tables.

Following the notation above, I now define the discrete kernel estimators for the marginal probabilities. I consider the sth dimension/column $\boldsymbol{X}^s$ as before. The discrete kernel estimator for falling in the x_j^s category is defined as shown in Equation (4.3) where l is the kernel function and λ^s is the smoothing parameter controlling for the smoothing level for the sth dimension.

$$\tilde{p}(x_j^s) = \frac{1}{n}\sum_{i=1}^{n} l(X_i^s, x_j^s, \lambda^s). \tag{4.3}$$

I consider the kernel function for pure categorical data proposed by Aitchison and Aitken (1976) and label the resulted kernel estimator $\tilde{p}^c$. The kernel function is shown in Equation (4.4) where $\lambda^s \in [0, \frac{k^s-1}{k^s}]$. When $\lambda^s = 0$, the kernel estimator reduces to the MLE estimator in Equation (4.1) which does not borrow information from other cells. When $\lambda^s = \frac{k^s-1}{k^s}$, the weights on all categories are identical, that is, $\frac{1}{k^s}$, giving rise to equal probabilities. As is shown in Ouyang et al. (2006), it could be written in a similar fashion to a Bayes estimator $\tilde{p}^c(x_j^s) = \frac{\lambda^s k^s}{k^s-1}\frac{1}{k^s} + (1 - \frac{\lambda^s k^s}{k^s-1})\hat{p}^{mle}(x_j^s)$. Therefore, $\tilde{p}^c(x^s)$ is a weighted average of equal probabilities and observed frequencies. The larger λ^s is, the more smoothed the estimated probabilities and the closer they are to the equal probability $\frac{1}{k^s}$.

$$l^c(X_i^s, x_j^s, \lambda^s) = \begin{cases} 1 - \lambda^s & if\ X_i^s = x_j^s, \\ \frac{\lambda^s}{k^s-1} & if\ X_i^s \neq x_j^s. \end{cases} \tag{4.4}$$

When the categories are ordered, it could be very beneficial to exploit the ordinal structure. Intuitively, we would consider the ordering as implying that the true probabilities on **nearby** categories in a multinomial distribution are more similar (Titterington and Bowman, 1985). Here, the term **nearby** relates to attribute space and refers to the closeness of consecutive classes which could be an artifact of discretizing continuous data (e.g. incomes in the study of regional income distribution dynamics). Here, I consider the discrete kernel function l^o as defined in Equation (4.5) where $\lambda^s \in [0, 1]$. The interpretation of the smoothing parameter λ^s is similar to that for l^c: when $\lambda^s = 0$, the kernel estimators $\tilde{p}^o$ reduce to MLEs; when $\lambda^s = 1$, all the estimated probabilities are identical. One problem with this kernel function is that the kernel weights do not always sum to 1 and thus the smoothed estimators $\tilde{p}^o$ are not appropriate when $\lambda^s \neq 0$. To address this issue, I normalize them so that they always sum to 1.

$$l^o(X_i^s, x_j^s, \lambda^s) = \begin{cases} 1 & \text{if } X_i^s = x_j^s \\ \lambda^{s|X_i^s - x_j^s|} & \text{if } X_i^s \neq x_j^s \end{cases} \tag{4.5}$$

The smoothed estimator for a d-way contingency table could be constructed based on a geometric combination of d smoothed estimators for multinomial data with d different smoothing parameters (Dong and Simonoff, 1995). Here, I follow the product kernel convention (Li and Racine, 2003; Ouyang et al., 2006; Li and Racine, 2007) which requires the multiplication of d discrete kernels l as shown in Equation (4.6) where $\boldsymbol{\lambda} = (\lambda^1, \lambda^2, \cdots, \lambda^d)$ represents the varying smoothing parameters for different classifications.

$$L(\boldsymbol{X}_i, \boldsymbol{x}_j, \boldsymbol{\lambda}) = \prod_{s=1}^{d} l(X_i^s, x_j^s, \lambda^s) \tag{4.6}$$

More specifically, for the discrete kernel function l^c for pure categorical data and the kernel function l^o taking account of the ordinal structure, the product kernel L^c and L^o are defined in Equation (4.7).

$$\begin{aligned} L^c(\boldsymbol{X}_i, \boldsymbol{x}_j, \boldsymbol{\lambda}) &= \prod_{s=1}^{d} (1-\lambda^s)^{\mathbb{1}(X_i^s = x_j^s)} \left(\frac{\lambda^s}{k^s - 1}\right)^{\mathbb{1}(X_i^s \neq x_j^s)}, \\ L^o(\boldsymbol{X}_i, \boldsymbol{x}_j, \boldsymbol{\lambda}) &= \prod_{s=1}^{d} 1^{\mathbb{1}(X_i^s = x_j^s)} (\lambda^{s|X_i^s - x_j^s|})^{\mathbb{1}(X_i^s \neq x_j^s)} = \prod_{s=1}^{d} \lambda^{s|X_i^s - x_j^s|} \end{aligned} \tag{4.7}$$

For a given smoothing parameter vector $\boldsymbol{\lambda}$, the smoothed estimator for the joint probability of $\boldsymbol{x}_j$ is shown in Equation(4.8). I can acquire the specific formula for $\tilde{p}^c(\boldsymbol{x}_j)$ and $\tilde{p}^o(\boldsymbol{x}_j)$ by plugging in product kernel functions in Equation (4.7).

$$\tilde{p}(\boldsymbol{x}_j) = \frac{1}{n} \sum_{i=1}^{n} L(\boldsymbol{X}_i, \boldsymbol{x}_j, \boldsymbol{\lambda}) \tag{4.8}$$

4.1.2.2 Smoothing Parameters Selection

Selecting an appropriate smoothing parameter(s) is of paramount importance to the performance of the discrete kernel estimators. Approaches to doing so could be similar to those for continuous kernel estimation which has received more attention, including the plug-in method (Chu et al., 2015), the cross-validation method (Henderson and Parmeter, 2015), and the Bayesian method (Agresti and Hitchcock, 2005; Belaid et al., 2016). The plug-in bandwidth for a one-way contingency table is derived by minimizing mean squared error (MSE) summed over the sample space. It is analytically challenging and only has a closed solution for the unordered kernel estimator l^c (Chu et al., 2015). Both of the cross-validation method and Bayesian method are data-driven and hence computationally intensive. The former often refers to the least square cross-validation (LSCV) method which involves the minimization of the LSCV function defined in Equation (4.9) where $\tilde{p}_{-i}(\boldsymbol{X}_i) = \frac{1}{n-1}\sum_{j,j\neq i}^{n-1} L(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{\lambda})$ is the leave-one-out estimator for the joint probability of $\boldsymbol{X}_i$. Ouyang et al. (2006) demonstrated that the discrete kernel estimators with the cross-validated smoothing parameters perform better than MLEs in terms of the summed MSE based on some Monte Carlo simulations. Belaid et al. (2016) proposed the Bayesian Markov chain Monte Carlo (MCMC) method through the likelihood cross-validation criterion for selecting the optimal smoothing parameters and compared it with LSCV. They did find better performance of the Bayesian MCMC method, but obviously it is much more computationally intensive. In this chapter, I focus on the LSCV method for selecting optimal smoothing parameters.

$$LSCV(\boldsymbol{\lambda}) = \sum_{\boldsymbol{x}_j} \tilde{p}(\boldsymbol{x}_j)^2 - \frac{2}{n}\sum_{i=1}^{n} \tilde{p}_{-i}(\boldsymbol{X}_i) \tag{4.9}$$

4.1.3 Smoothing Transition Probabilities for (Spatial) Markov Chains

4.1.3.1 Classic Markov Chains

In this section, I demonstrate how the transition probability matrix, which is core of a Markov chains model, could be viewed as a *conditional* contingency table (Kullback et al., 1962) and how the discrete kernel estimators for contingency tables could be utilized to address the sparsity issue commonly encountered in empirical studies.

$$\boldsymbol{P} = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,k) \\ p(2,1) & p(2,2) & \cdots & p(2,k) \\ \vdots & \vdots & \ddots & \vdots \\ p(k,1) & p(k,2) & \cdots & p(k,k) \end{bmatrix}, \tag{4.10}$$

$$0 \leq p(i,j) \leq 1, \sum_{j=1}^{k} p(i,j) = 1 \; \forall \; i, j \in \mathbb{S} = \{1, 2, \cdots, k\}.$$

Here, without loss of generality, let us consider a first-order time-homogenous Markov chains which transitions across k states and whose state at t is solely determined by its immediate preceding state (at $t-1$). The dynamics could be organized in a (k, k) transition probability matrix $\boldsymbol{P}$ shown in Equation (4.10). $\boldsymbol{P}$ is comprised of transition probabilities between categories across two consecutive periods. For example, $p(i, j)$ represents the probability of transitioning from category i at $t-1$ to category j at t. If there only exist positive entries for some power of $\boldsymbol{P}$, meaning that every two states could communicate, the Markov chain is said to be **irreducible**. For those irreducible Markov chains, an unique steady-state distribution $\boldsymbol{\pi}$ which is solely determined by $\boldsymbol{P}$ exists as shown in Equation (4.11). However, for Markov chains with sparse transition probability matrix, it is highly possible that not all states are able to communicate with each other. We call them reducible

Markov chains and more than one steady-state distributions exist each for a communicating class.

$$\boldsymbol{\pi P} = \boldsymbol{\pi} \tag{4.11}$$

By viewing **transitions** as the observed entities classified by two classification criteria (1 and 2) which determine the categories at $t-1$ and t, we could consider $p(i, j)$ here as a conditional probability, that is, the probability of falling in category j at t (classification 2) given that it fell in category i at $t-1$ (classification 1). Formally, $p(i, j)$ is equivalent to $Prob(x^2 = j|x^1 = i) = \frac{Prob((x^2=j)\cap(x^1=i))}{Prob(x^1=i)}$. The nominator is a joint probability which is the element constituting a two-way contingency table while the denominator is the marginal probability based on the "classification criterion 1" - the category at $t-1$.

The logic leads to the MLEs conventionally used for estimating transition probabilities in empirical studies (shown as $\hat{p}^{mle}(i, j)$ in Equation (4.12)). Similarly, since the discrete kernel estimators for joint probabilities for a two-way contingency table have been given in Equation (4.7) and the discrete kernel estimators for marginal probabilities were given in Equation (4.3), we can easily obtain smoothed estimators of conditional probabilities for the transition probability matrix ($\tilde{p}(i, j)$ in Equation (4.12)).

One exception that deserves additional attention especially when faced with sparse spatial observations is that there could be cases where there is no observation whose first classification is i. As far as the transition probability matrix is concerned, this means that there is no transitions from i over consecutive periods and the MLE estimates for all the transition probabilities in row i will be 0, failing to suffice the condition that $\boldsymbol{P}$ is a stochastic matrix. I deal with the exception by following the common practice which basically adds an observation whose first and second classifications are both i to the original sample. The resulted MLE estimate would be 1 for $p(i, i)$ and 0 for the other entries in the ith row.

$$\begin{aligned}\hat{p}^{mle}(i,j) &= \frac{\hat{p}(\boldsymbol{x}=(i,j))}{\hat{p}(x^1=i)}\\ \tilde{p}(i,j) &= \frac{\tilde{p}(\boldsymbol{x}=(i,j))}{\tilde{p}(x^1=i)}\end{aligned} \tag{4.12}$$

The LSCV method for smoothing parameter selection follows Equation (4.9) and is shown in Equation (4.13) where $\tilde{p}_{-m}(\boldsymbol{x} = (X_m^1, X_m^2))$ is the leave the mth observation out estimator for the joint probability of (X_m^1, X_m^2).

$$LSCV(\lambda^1,\lambda^2) = \sum_{i=1}^{k}\sum_{j=1}^{k}\tilde{p}(\boldsymbol{x}=(i,j))^2 - \frac{2}{n}\sum_{m=1}^{n}\tilde{p}_{-m}(\boldsymbol{x}=(X_m^1,X_m^2)) \tag{4.13}$$

4.1.3.2 Spatial Markov Chains

The spatial Markov chain model was proposed by Rey (2001) to interrogate space in a classic Markov chain model for studying the evolution of regional income distributions. It is formulated by decomposing the unique transition probability matrix into k transition probability matrices, based on which k steady state distributions $\boldsymbol{\pi}_1$,$\boldsymbol{\pi}_2$,. . . , $\boldsymbol{\pi}_k$ could potentially be derived. These matrices are estimated from mutually exclusive and exhaustive subsamples of **transitions** based on MLEs in the same way as that for the classic Markov chain model. Determining which subsample each transition falls into adds to the two-way contingency table a third classification criterion. For the spatial Markov chain model, the criterion would be spatial lag category $h \in \mathbb{S} = \{1, 2, \cdots, k\}$ at $t-1$. Spatial lag refers to the average level of neighbors for continuous data (e.g. incomes) which needs further discretization to acquire k categories, and can also refer to the most common category among neighbors for discrete data (e.g. land use types). Thus, following the logic of the two-way *conditional* contingency table for classic Markov chains, we could derive a three-way *conditional* contingency table here. The extra classification criterion h is that applied to the spatial lags. Given the entity itself fell into i and its spatial lag was classified as h at period $t-1$, the probability of transitioning to j at t is $Prob(x^2 = j|(x^1 = i \cap x^3 = h)) = \frac{Prob((x^2=j)\cap(x^1=i)\cap(x^3=h))}{Prob((x^1=i)\cap(x^3=h))}$. Both of the nominator and the denominator are joint probabilities and can be estimated based on the MLEs or discrete kernel estimators for contingency tables introduced in the preceding section (Equation (4.14)). The LSCV method for simultaneously selecting three optimal parameters $(\lambda^1, \lambda^2, \lambda^3)$ for the three-way *conditional* contingency table can be easily derived as shown in Equation (4.15).

$$\begin{aligned} \hat{p}^{mle}(i,j,h) &= \frac{\hat{p}(\boldsymbol{x} = (i,j,h))}{\hat{p}(\boldsymbol{x} = (i,,h)} \\ \tilde{p}(i,j,h) &= \frac{\tilde{p}(\boldsymbol{x} = (i,j,h))}{\tilde{p}(\boldsymbol{x} = (i,,h))} \end{aligned} \tag{4.14}$$

$$LSCV(\lambda^1, \lambda^2, \lambda^3) = \sum_{i=1}^{k}\sum_{j=1}^{k} \tilde{p}(\boldsymbol{x} = (i,j,h))^2 - \frac{2}{n}\sum_{m=1}^{n} \tilde{p}_{-m}(\boldsymbol{x} = (X_m^1, X_m^2, X_m^3)) \tag{4.15}$$

4.2 Monte Carlo Experiments

In this section, I introduce a series of Monte Carlo experiments which have been conducted to evaluate the performance of two discrete kernel estimators for smoothing the transition probability matrix for the classic Markov chain model and the spatial Markov chain model.

4.2.1 Evaluation Criteria - Mean Squared Error (MSE)

As has been discussed before, the purpose of proposing smoothed estimators for the Markov chain model or the spatial Markov chain model is to deal with the unsatisfactory properties of MLEs in the presence of sparse spatial observations - the most prominent is the tendency of producing many zero transition probability estimates. This is done by evening up the probabilities, which introduces bias but should lower the variance. The tradeoff between bias and variance for the whole transition probability matrix could be balanced via the LSCV method that selects the optimal global smoothing parameters. Therefore, we expect a decrease of mean squared error (MSE) summed over all transition probabilities compared to the nonsmoothed case. However, this might not be true for individual transition probabilities as well as other parameters derived from the matrix such as the steady state probabilities. By comparing MLEs and smoothed estimators in terms of the MSEs for these parameters of interest, we could gain a comprehensive understanding of the performance of the smoothed estimators.

Suppose we generate M samples for each Monte Carlo experiment, for the (k, k) transition probability matrix of a classic Markov chain model, and for the (k, k, k) transition probability matrix of a spatial Markov chain model, the MSE summed over all transition probabilities are defined as shown in Equation (4.16). Aside from the MSE for the whole transition matrix, I also look at the MSEs for individual probabilities and steady-state probabilities. Suppose the parameter of interest is y, we could calculate its MLE as defined in Equation (4.17).

$$MSE(\hat{\boldsymbol{P}}) = \begin{cases} \frac{1}{M} \sum_{r=1}^{M} \sum_{i=1}^{k} \sum_{j=1}^{k} (\hat{p}(i,j) - p(i,j))^2 & \boldsymbol{P} \text{ is } (k,k), \\ \frac{1}{M} \sum_{r=1}^{M} \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{h=1}^{k} (\hat{p}(i,j,h) - p(i,j,h))^2 & \boldsymbol{P} \text{ is } (k,k,k). \end{cases} \tag{4.16}$$

$$MSE(\hat{y}) = \frac{1}{M} \sum_{r=1}^{M} (\hat{y} - y)^2. \tag{4.17}$$

4.2.2 Experiments for Smoothing Markov Chains

I followed the conventional approach to generate a classic Markov chain which requires a **True** transition probability matrix, an initial state, and a time length. I experimented with several different dimensions including three transition probability matrices of different structures and varying time lengths which determines the number of transitions.

4.2.2.1 True Transition Probability Matrices

The point of departure of the Monte Carlo experiments is the 5 by 5 transition probability matrix (Equation (4.18)) estimated from quintile discretized U.S. state relative per capita income time series 1929-2010 based on the MLE. This transition probability matrix is strongly diagonally dominant[8] and irreducible, although it is rather sparse in that: 5 out of 25 cells are 0 meaning that it is impossible for transitioning between corresponding classes over consecutive periods; 12 cells are smaller than 0.01 indicating very unlikely transitions. The sparseness could be a result of small sample size or it is just reflecting the extremely low possibility of transitioning between some categories. Either way, we treat it as the **true** transition probability matrix used for simulating Markov chains of various lengths which could then be used to investigate the performance of smoothed estimators.

[8] Diagonal domination of a matrix refers to the fact that the diagonal cells in a row is no smaller than the sum of all the other cells in the same row.

$$
\boldsymbol{P5} = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0.915 & 0.075 & 0.009 & 0.001 & 0.000 \\ 0.066 & 0.827 & 0.105 & 0.001 & 0.001 \\ 0.005 & 0.103 & 0.794 & 0.095 & 0.003 \\ 0.000 & 0.009 & 0.094 & 0.849 & 0.048 \\ 0.000 & 0.000 & 0.000 & 0.062 & 0.938 \end{bmatrix} \end{array}. \tag{4.18}
$$

A type of perturbation called **diagonalising shifting** (Dardanoni, 1995) which shifts probability mass away from the main diagonal was applied to $\boldsymbol{P5}$. The amount of shifted mass is controlled by a portion parameter β, meaning that β of the diagonal elements and is equally assigned to the nondiagonal elements in each row. By setting $\beta = \beta_1 = 0.3$ and $\beta = \beta_2 = 0.7$, I obtained two less sparse transition probability matrices, $\boldsymbol{P5}_{\beta_1}$ and $\boldsymbol{P5}_{\beta_2}$. Both are now exempt from the sparsity issue, while the former is still diagonally dominant and the latter is not.

$$
\boldsymbol{P5}_{\beta_1} = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0.640 & 0.144 & 0.078 & 0.070 & 0.068 \\ 0.128 & 0.579 & 0.167 & 0.063 & 0.063 \\ 0.064 & 0.162 & 0.556 & 0.155 & 0.063 \\ 0.064 & 0.073 & 0.157 & 0.594 & 0.112 \\ 0.070 & 0.071 & 0.070 & 0.132 & 0.657 \end{bmatrix} \end{array}, \quad \boldsymbol{P5}_{\beta_2} = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0.275 & 0.235 & 0.169 & 0.161 & 0.160 \\ 0.211 & 0.248 & 0.250 & 0.146 & 0.145 \\ 0.144 & 0.242 & 0.238 & 0.234 & 0.142 \\ 0.149 & 0.157 & 0.243 & 0.255 & 0.196 \\ 0.164 & 0.164 & 0.164 & 0.226 & 0.282 \end{bmatrix} \end{array} \tag{4.19}
$$

4.2.2.2 Other Experimental Dimensions

Based on each of the **true** transition probability matrices $\boldsymbol{P5}$, $\boldsymbol{P5}_{\beta_1}$, and $\boldsymbol{P5}_{\beta_2}$, I simulate a time series of length $T = 25, 49, 81$ 1000 times. Each time series gives rise to a total number of transitions $n = 24, 48, 80$. The initial state is randomly assigned for each simulation. It should be noted that although the initial state does not impact the steady state distribution, it could be a key influence in estimating the transition probability matrix when the time length is short. This is because it is highly possible that the system will be stuck in its initial state for a long time if the underlying transition probability matrix is strongly diagonally dominant, e.g. $\boldsymbol{P}\mathbf{5}$ ($\beta = 0$). Therefore, simulating 2 time series of of length $T = 13$ which also results in 24 transitions may not be equivalent to simulating 1 time series of $T = 25$. Moreover, the former could contain much more information if the 2 initial states are different. In other words, its effective sample size could be larger.

4.2.3 Experiments for Smoothing Spatial Markov Chains

4.2.3.1 Data Generating Process (DGP) for a Spatial Markov Chain Model

Simulating a spatial Markov chain model is done differently from the approach used for a classic Markov chain model. The differences are not only in the transition probability matrix which is (k, k, k) for the spatial case compared to (k, k) for the classic case, but also lie in the fact that spatiotemporal interactions are incorporated in the former case. Therefore, a map displaying the geographical locations of the N spatial units and a reasonable perspective from which we conceptualize neighboring relationships are needed. For example, spatial units sharing an edge/node can be defined neighbors and the neighboring relationships are usually formalized in a (N, N) spatial weight matrix (Anselin, 1988).

Instead of simulating one time series as a sample for the classic Markov, here I simultaneously simulate N time series each of which represents the evolution of the variable of interest in a spatial unit (e.g. per capita income time series of US states or land use time series of US tracts). Given N initial states and the k (k, k) transition probability matrices, I first obtain N categorical spatial lags which are the most common category among the neighbors for each spatial unit[9]. The spatial lags at the initial period determine which of the k transition probability matrices to use for generating each of the N categories at period 2. After obtaining N categories at period 2, I calculate their categorical spatial lags, determine which transition probability matrix should drive the dynamics and then generate N categories at period 3. I repeat the procedures until N categories at period T are obtained.

4.2.3.2 True Transition Probability Matrices

Similar to the classic Markov case, I experimented with three **true** transition probability matrices for the spatial Markov experiments. The point of departure is the 5 spatial lag-conditional $(5, 5)$ transition probability matrices estimated from an empirical dataset based on MLEs. The dataset consists of 31 China provincial real per capita income series from 1978 to 2016[10]. I first obtained relative per capita incomes by dividing the real per capita incomes by annual national averages for correcting for business cycle and trends in the Chinese average income. Then these relative per capita incomes as well as their contemporaneous continuous spatial lags were further discretized into 5 categories based on the global quintiles. Afterwards, the 5 transition probability matrices were estimated for each spatial lag-dependent subsamples based on MLEs. As shown in Equation (4.20), some

[9] I have also experimented with ordinal spatial lags which can be defined for categories with an inherent ordinal structure. This type of spatial lag is defined as the category closest to the average of neighbors' category values. I do not report the results here because they are similar to what I have obtained for categorical spatial lags.

[10] Nominal average per capita income series of 31 Chinese provinces 1978-2016 were downloaded from China Data Center and they were converted to temporally comparable real incomes by using the deflator, Consumer Price Index (CPI) which was collected from National Data National Bureau of Statistics of China.

serious sparsity issues are encountered here. Take the fifth row of the first $(5,5)$ matrix as an example. Since we did not observe any provinces with poor neighbors at $t-1$ (spatial lag was in category 1) in the rich (category 5) across period $t-1$ and t, this row should be filled with 0 probabilities based on MLEs. To satisfy the requirement of row sum equal to 1, I followed the conventional ad-hoc way to fill the diagonal element with 1. All the four bold numbers indicate such cases. For the three matrices $(1,2,5)$ suffering from the zero row issue, they are linked to reducible Markov chains even after such adjustments. Moreover, these five matrices are still very sparse and strongly diagonally dominant.

$$
\begin{array}{l}
1\;\begin{array}{cc} & \begin{array}{ccccc}\;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; & \;\;5\;\;\end{array}\\ \begin{array}{c}1\\2\\3\\4\\5\end{array} & \begin{bmatrix}0.940 & 0.060 & 0.000 & 0.000 & 0.000\\ 0.040 & 0.853 & 0.107 & 0.000 & 0.000\\ 0.000 & 0.097 & 0.855 & 0.048 & 0.000\\ 0.000 & 0.000 & 0.158 & 0.842 & 0.000\\ 0.000 & 0.000 & 0.000 & 0.000 & \mathbf{1.000}\end{bmatrix}\end{array}\\
2\;\begin{array}{cc} & \begin{array}{ccccc}\;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; & \;\;5\;\;\end{array}\\ \begin{array}{c}1\\2\\3\\4\\5\end{array} & \begin{bmatrix}0.945 & 0.055 & 0.000 & 0.000 & 0.000\\ 0.042 & 0.931 & 0.028 & 0.000 & 0.000\\ 0.000 & 0.069 & 0.759 & 0.172 & 0.000\\ 0.000 & 0.000 & 0.357 & 0.607 & 0.036\\ 0.000 & 0.000 & 0.000 & 0.000 & \mathbf{1.000}\end{bmatrix}\end{array}\\
3\;\begin{array}{cc} & \begin{array}{ccccc}\;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; & \;\;5\;\;\end{array}\\ \begin{array}{c}1\\2\\3\\4\\5\end{array} & \begin{bmatrix}0.842 & 0.158 & 0.000 & 0.000 & 0.000\\ 0.086 & 0.793 & 0.121 & 0.000 & 0.000\\ 0.000 & 0.000 & 0.872 & 0.128 & 0.000\\ 0.000 & 0.000 & 0.000 & 0.964 & 0.036\\ 0.000 & 0.000 & 0.000 & 0.045 & 0.955\end{bmatrix}\end{array}\\
4\;\begin{array}{cc} & \begin{array}{ccccc}\;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; & \;\;5\;\;\end{array}\\ \begin{array}{c}1\\2\\3\\4\\5\end{array} & \begin{bmatrix}0.909 & 0.091 & 0.000 & 0.000 & 0.000\\ 0.094 & 0.844 & 0.062 & 0.000 & 0.000\\ 0.000 & 0.041 & 0.857 & 0.102 & 0.000\\ 0.000 & 0.000 & 0.051 & 0.897 & 0.051\\ 0.000 & 0.000 & 0.000 & 0.115 & 0.885\end{bmatrix}\end{array}\\
5\;\begin{array}{cc} & \begin{array}{ccccc}\;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; & \;\;5\;\;\end{array}\\ \begin{array}{c}1\\2\\3\\4\\5\end{array} & \begin{bmatrix}\mathbf{1.000} & 0.000 & 0.000 & 0.000 & 0.000\\ 0.000 & \mathbf{1.000} & 0.000 & 0.000 & 0.000\\ 0.000 & 0.000 & 0.875 & 0.125 & 0.000\\ 0.000 & 0.000 & 0.036 & 0.909 & 0.055\\ 0.000 & 0.000 & 0.000 & 0.006 & 0.994\end{bmatrix}\end{array}
\end{array} \tag{4.20}
$$

I produced two non-sparse (k, k, k) transition probability matrices by applying the **diagonalising shifting** technique to each (k, k) matrix. The proportion of shifted mass is still $\beta = \beta_1 = 0.3$ and $\beta = \beta_2 = 0.7$.

4.2.3.3 Other Experimental Dimensions

As has been discussed in the last subsection, simulating 1 time series of length T could be very different from simulating 2 time series of length $\frac{T-1}{2} + 1$ though they seem to give the same number of transition observations $T - 1$. The latter could give rise to a larger effective sample size because of a larger number of initial states and possibly different values of these initial states. Here, to control for the potential impacts of initial states in short-run dynamics of the spatial Markov chains, I assigned the $N = 31$ per capita income classes of Chinese provinces in 1978 to the initial states and keep them fixed across all simulations.

I adopted the queen contiguity matrix for formalizing the neighboring relationships of 31 Chinese provinces across all simulations. I also varied the time length $T = 5, 20, 39$ where $T = 39$ replicates the time length of the empirical dataset. For each combination of experimental parameters, including the (k, k, k) **true** transition probability matrix and the time length T, I simulated N time series of length T $1,000$ times, and estimated the transition probability matrices based on MLEs, and the two discrete kernel estimators.

Table 8. (Average) Optimal Smoothing Parameters via LSCV for $1,000$ Markov Simulations.

			$\tilde{p}^c$		$\tilde{p}^o$	
N	T	β	λ^1	λ^2	λ^1	λ^2
		0	0.045	0.046	0.030	0.031
1	25	0.3	0.205	0.206	0.171	0.170
		0.7	0.653	0.648	0.733	0.732
		0	0.027	0.027	0.017	0.017
1	49	0.3	0.113	0113	0.089	0.089
		0.7	0.627	0.627	0.684	0.694
		0	0.018	0.018	0.012	0.012
1	81	0.3	0.073	0.073	0.055	0.055
		0.7	0.591	0.589	0.623	0.618

4.3 Results

4.3.1 Smoothing Markov Chains

The average LSCV-based optimal smoothing parameters of the two smoothed estimators for each set of $1,000$ Monte Carlo simulations are shown in Table 8. The optimal parameters λ^1 and λ^2 for both smoothed estimators are similar in magnitude indicating that the amount of information borrowed from categories at period t is similar to that borrowed from categories at period $t-1$. For the same sample size, as the **true** probability matrix becomes less diagonally dominant (increasing β), both of λ^1 and λ^2 increase, meaning that the conditional probability estimates are increasingly smoothed. For the same **true** transition probability matrix, both of λ^1 and λ^2 decrease as more observations become available.

The MSEs for the whole transition probability matrix with the LSCV-based optimal smoothing parameters are displayed in Table 9. Bold numbers indicate the smallest MSE among the MLE and the two smoothed estimators. The performance of both smoothed estimators is superior to the MLEs almost under all circumstances. Focusing on cases

Table 9. MSEs for the Whole Transition Probability Matrix for $1,000$ Markov Simulations.

N	T	β	$MSE(\hat{\boldsymbol{P}})$		
			$\hat{p}^{mle}$	$\tilde{p}^{c}$	$\tilde{p}^{o}$
		0	0.493	0.523	**0.484**
1	25	0.3	1.245	0.774	**0.758**
		0.7	1.078	**0.095**	0.101
		0	**0.438**	0.458	0.442
1	49	0.3	0.591	0.481	**0.479**
		0.7	0.479	0.072	**0.071**
		0	**0.316**	0.339	0.328
1	81	0.3	0.278	0.270	**0.269**
		0.7	0.273	0.062	**0.059**

when sample size is very small ($N = 1, T = 25$), all three estimators are similar in terms of minimizing the MSE for the whole matrix when the **true** transition probability matrix is strongly diagonally dominant. As β increases (the **true** transition probability matrix becomes less diagonally dominant), both smoothed estimators produce much smaller MSEs than the MLE. The superiority is more obvious as β gets larger. With β fixed, the MLE for each estimator is smaller as the sample size increases, and still, both smoothed estimators seem to be a better choice than the MLE.

I further decompose the MSE into MSEs for individual transition probability estimators. They are visualized for three **true** transition probability matrices in three parallel coordinates plots for cases where the length of the Markov chain is 25[11] (Figure 21). The red curves denote MSEs for MLEs ($\hat{p}^{mle}$), the blue curves denote MSEs for smoothed estimators ($\tilde{p}^{c}$) which do not consider potential ordinal structures of categories and the green curves denote MSEs for smoothed estimators ($\tilde{p}^{o}$) which account for ordinal structures of categories. Looking at the top plot when $\beta = 0$, the MLEs for individual entries are quite similar across

[11]The MSEs for individual transition probability estimators for $T = 49, 81$ are similar to the case for $T = 25$ and are not displayed here for space saving, but are available upon request.

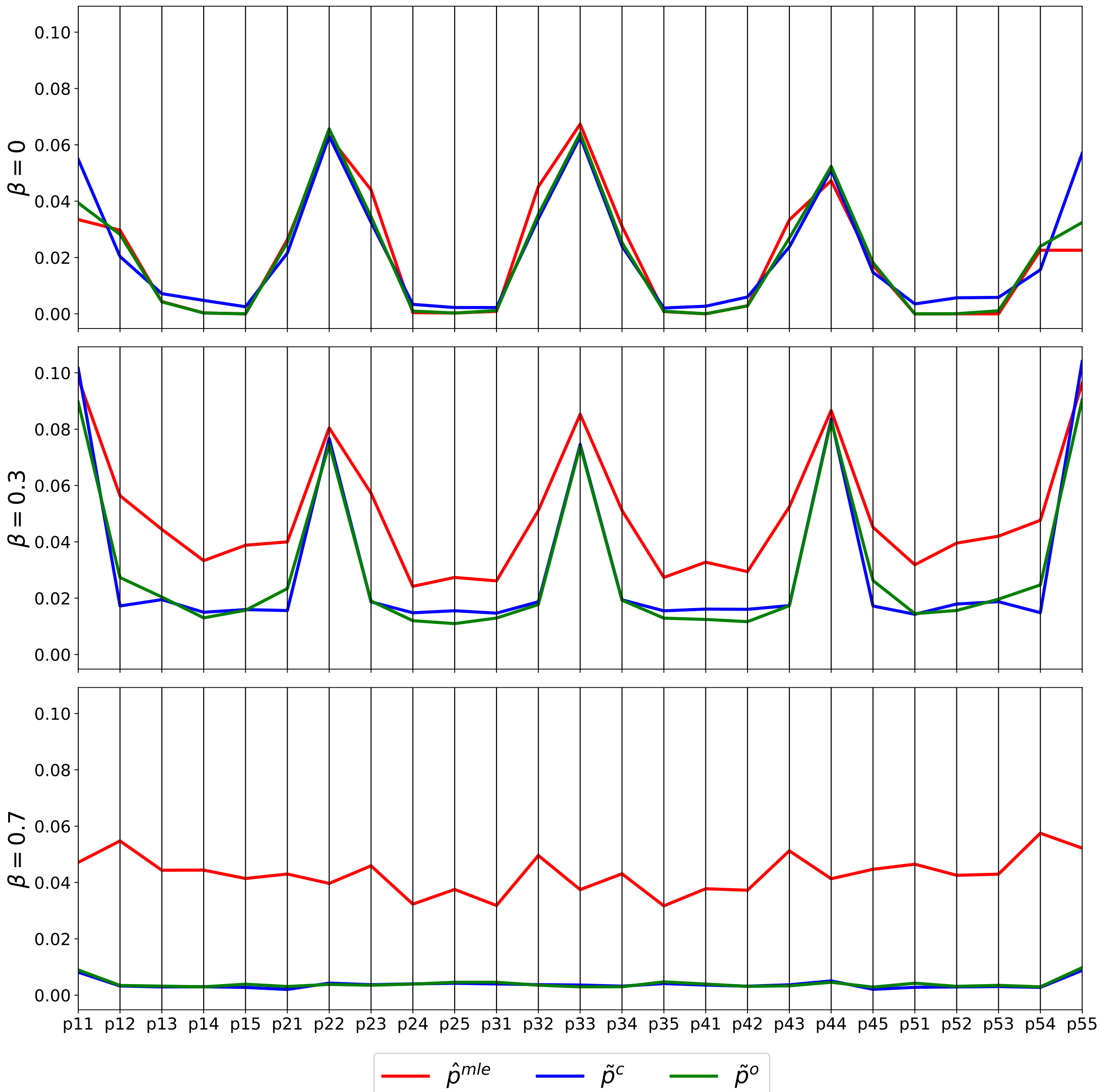

Figure 21. MSEs for Individual Probability Estimators for a Short Markov Chain ($T = 25$).

three estimators. As we move to the second plot when $\beta = 0.3$, the MSEs for MLEs are almost always larger than those for the two smoothed estimators. Moving to the bottom plot when $\beta = 0.7$, we find that both smoothed estimators produce a much smaller MSE for every individual transition probability than MLEs.

Moreover, if we look at other properties of the estimated transition probability matrix

Table 10. Proportion of Estimated Transition Probability Matrices Giving Rise to Irreducible Markov Chains in $1,000$ Markov Simulations.

			Irreducible proportion		
N	T	β	$\hat{p}^{mle}$	$\tilde{p}^{c}$	$\tilde{p}^{o}$
		0	0.001	1.000	1.000
1	25	0.3	0.513	1.000	1.000
		0.7	0.949	1.000	1.000
		0	0.041	1.000	1.000
1	49	0.3	0.937	1.000	1.000
		0.7	0.999	1.000	1.000
		0	0.167	1.000	1.000
1	81	0.3	0.994	1.000	1.000
		0.7	1.000	1.000	1.000

such as the irreducibility, which is an essential property concerning whether any two states are able to communicate with each other, as well as the steady state distribution, which is usually considered as prediction into the future assuming the current dynamics last, we can similarly find superior performance of the smoothed estimators than MLEs. Since we know that all the experimental **true** transition probability matrices link to irreducible Markov chains, we prefer the estimated matrices to have the same property. Table 10 displays the proportion of estimated transition probability matrices linking to irreducible Markov chains in the $1,000$ simulated Markov chain samples for each experiment. In all cases, smoothed estimators ensure the irreducible property of the estimates while the performance of MLE is not quite satisfactory. Particularly, for the sparse **true** transition probability matrix ($\beta = 0$), the MLE only produces 1 transition probability matrix estimate owing the irreducible property when sample size is very small ($N = 1, T = 25$). Its performance is slightly better as the sample size becomes larger, but is still less than satisfactory.

Looking at the MSEs for individual entries in the steady state distribution $\boldsymbol{\pi}$ in Figure 22, we could find a more obvious superior performance of the smoothed estimators. This is very true for cases where the **true** transition probability matrix is either moderately

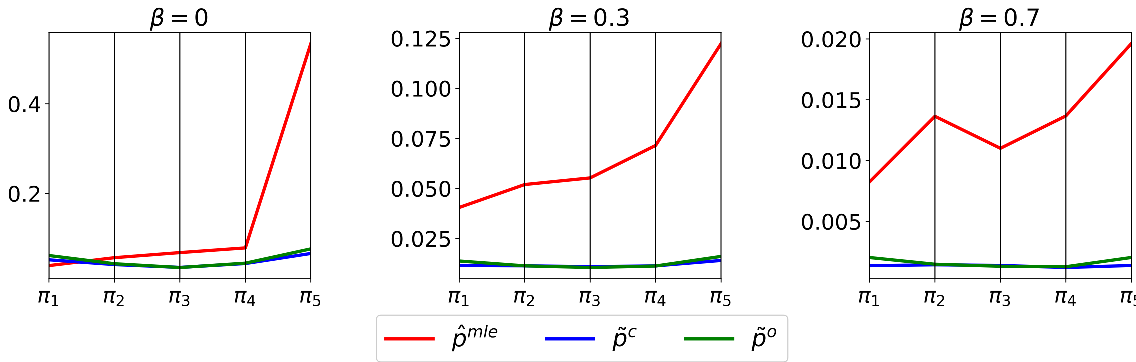

Figure 22. MSEs for Individual Steady-State Probability Estimators for a Short Markov Chain ($T = 25$).

diagonally dominant ($\beta = 0.3$) or not diagonally dominant ($\beta = 0.7$). Even for the strongly diagonally dominant $\boldsymbol{P}$ ($\beta = 0$), the MSEs of individual steady state probabilities for smoothed estimators seem to be consistently lower.

4.3.2 Smoothing Spatial Markov Chains

In general, the performance of the two discrete kernel estimators for estimating the k (k, k) transition probability matrices for spatial Markov chains is similar to that for classic Markov chains.

The LSCV-based optimal smoothing parameters averaged for $1,000$ spatial Markov simulations are displayed in Table 11. For cases which are driven by k diagonally dominant transition probability matrices ($\beta = 0, 0.3$), the optimal smoothing parameter λ_3 is larger than λ_1 and λ_2 indicating that the estimates borrow more information from categories of spatial lags than from categories at period $t-1$ and t. In contrast, when **true** transition probability matrices are not diagonally dominant ($\beta = 0.5$), the smoothing process borrow

Table 11. (Average) Optimal Smoothing Parameters via LSCV for $1,000$ Spatial Markov Simulations.

			$\tilde{p}^c$			$\tilde{p}^o$		
N	T	β	λ^1	λ^2	λ^3	λ^1	λ^2	λ^3
		0	0.016	0.018	0.156	0.013	0.016	0.146
31	5	0.3	0.019	0.095	0.334	0.037	0.054	0.294
		0.7	0.316	0.634	0.211	0.305	0.665	0.195
		0	0.003	0.004	0.034	0.002	0.004	0.030
31	20	0.3	0.00004	0.032	0.152	0.001	0.023	0.095
		0.7	0.014	0.497	0.144	0.037	0.480	0.071
		0	0.001	0.002	0.017	0.001	0.002	0.013
31	39	0.3	0.00002	0.018	0.094	0.00001	0.013	0.053
		0.7	0.00009	0.377	0.104	0.006	0.340	0.052

Table 12. MSEs for the Whole Transition Probability Matrix for $1,000$ Spatial Markov Simulations.

				$MSE(\hat{\boldsymbol{P}})$	
N	T	β	$\hat{p}^{mle}$	$\tilde{p}^c$	$\tilde{p}^o$
		0	1.767	0.946	**0.938**
1	25	0.3	4.994	**1.378**	1.523
		0.7	6.545	**0.441**	0.448
		0	0.678	**0.457**	0.499
1	49	0.3	0.839	**0.519**	0.562
		0.7	1.031	**0.295**	0.302
		0	1.000	**0.257**	0.294
1	81	0.3	0.381	**0.298**	0.312
		0.7	0.496	**0.226**	0.229

more information from categories at period t than from the others. Similar to before, a smaller value is selected for each smoothing parameter as more observations are available.

The MSEs of the whole (k, k, k) transition probability matrix for all the estimators are shown in Table 12. Clearly, the MLEs are superior in terms of minimizing MSE even when each of the (k, k) spatial lag-conditional transition probability matrix is strongly diagonally dominant. When these matrices are not diagonally dominant, we observe a much better performance of smoothed estimators.

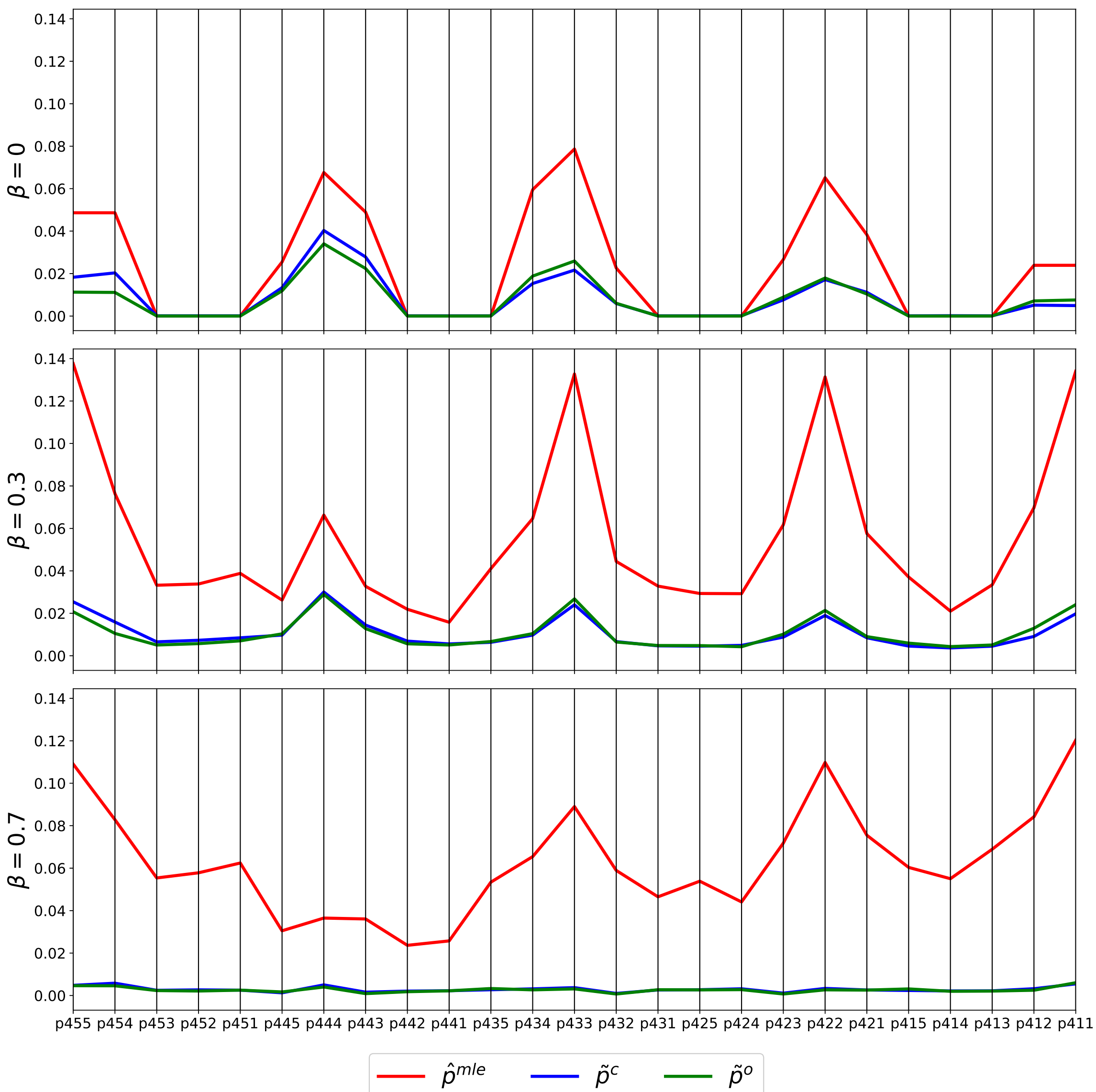

Figure 23. MSEs for Individual Probability Estimators for the Fourth (k, k) Matrix for Short Spatial Markov Chains $(N = 31, T = 5)$.

Table 13. Proportion of Fourth (k, k) Conditional Transition Probability Matrix Estimates Giving Rise to Irreducible Markov Chains in $1,000$ Spatial Markov Simulations.

			Irreducible proportion		
N	T	β	$\hat{p}^{mle}$	$\tilde{p}^{c}$	$\tilde{p}^{o}$
		0	0	1.000	0.999
1	25	0.3	0.062	1.000	1.000
		0.7	0.515	1.000	1.000
		0	0.06	1.000	1.000
1	49	0.3	0.99	1.000	1.000
		0.7	1.000	1.000	1.000
		0	0.426	1.000	1.000
1	81	0.3	1.000	1.000	1.000
		0.7	1.000	1.000	1.000

Decomposing the $MSE(\hat{\boldsymbol{P}})$ into MSEs of individual probability estimators, we could discern a similar pattern as we have observed for classic Markov case. In Figure 23, I visualize the MSEs individual probability estimators for the fourth matrix for the smallest sample case. Clearly, the performance of smoothed estimators is contingent on the structure of the **true** transition probability matrix. The weaker diagonally dominant the matrix, the more superior the smoothed estimators in reducing the individual MSEs.

Now we turn to the irreducibility and the steady state probabilities for estimators for each of k conditional transition probability matrix. Since the fourth (k, k) transition probability matrix in the **true** k matrices when $\beta = 0$ (Equation 4.20) links to irreducible Markov chains, we expect the transition probability matrix estimate should render the same property. As displayed in the first row of Table **??**, smoothed estimators ensure the irreducibility property while the MLE serves as a negative example which ensures the reducibility property. Comparatively, the MLEs seem to better recover the reducibility property linked to the first (k, k) transition probability matrix ($\beta = 0$) as shown in the first row of Table **??**.

Further, let us focus on the spatially conditional steady state probabilities implied by the

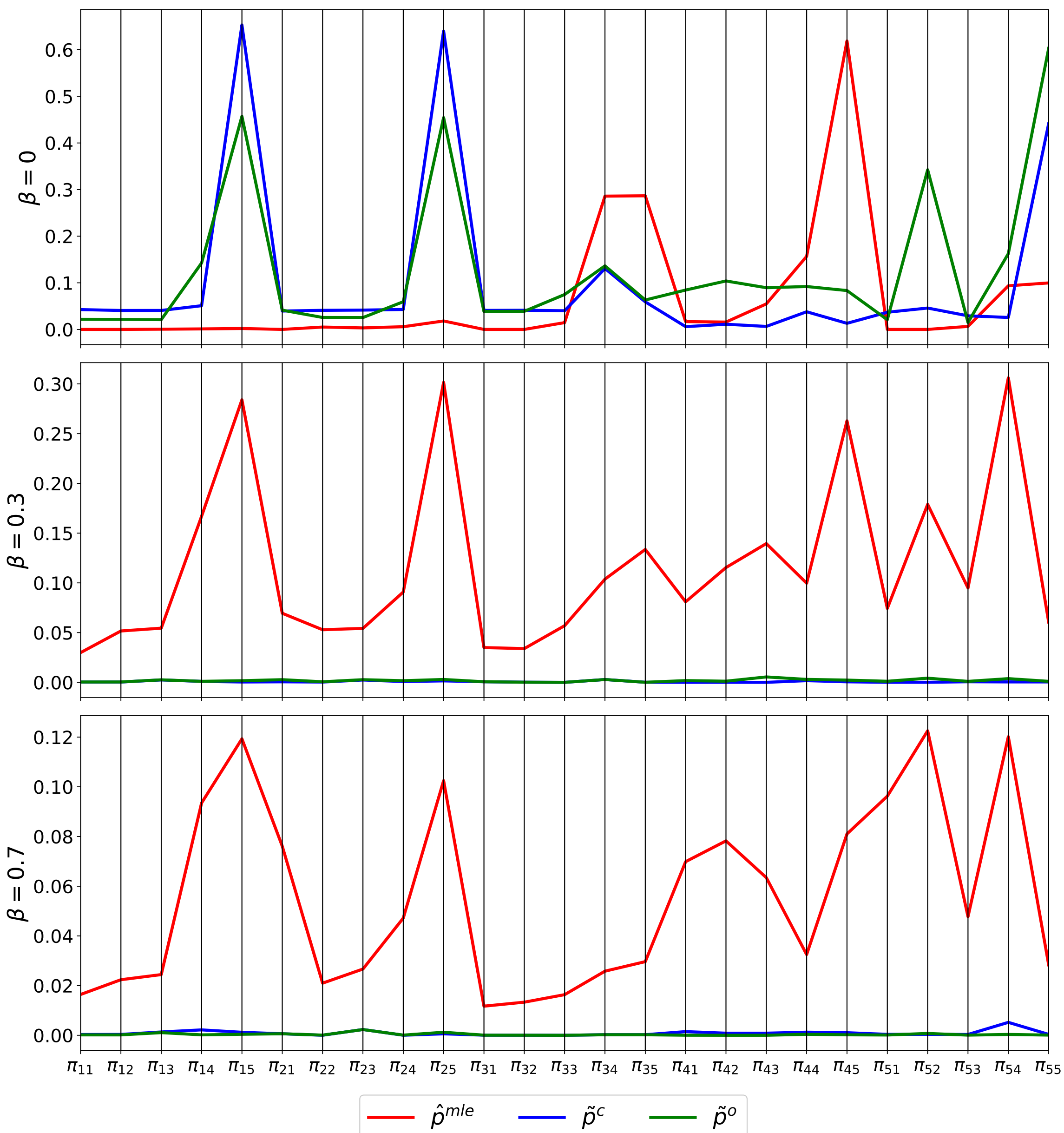

Figure 24. MSEs for Individual Steady-State Probability Estimators for Short Spatial Markov Chains ($N = 31, T = 5$).

Table 14. Proportion of First (k, k) Conditional Transition Probability Matrix Estimates Giving Rise to Irreducible Markov Chains in $1,000$ Spatial Markov Simulations.

			Irreducible proportion		
N	T	β	$\hat{p}^{mle}$	$\tilde{p}^{c}$	$\tilde{p}^{o}$
		0	0	1.000	0.999
1	25	0.3	0.26	1.000	1.000
		0.7	0.658	1.000	1.000
		0	0.001	1.000	1.000
1	49	0.3	0.975	1.000	1.000
		0.7	1.000	1.000	1.000
		0	0.012	1.000	1.000
1	81	0.3	1.000	1.000	1.000
		0.7	1.000	1.000	1.000

spatial Markov transition probability matrices. As discussed earlier, steady state probabilities serve as important indicators of the long run distribution assuming the current dynamics persist. The steady state probabilities conditional on spatial context here shed light on the role of space in shaping the long run distribution. As shown in Figure 24, the smoothed estimators are better at reducing MSEs in all cases except when the **true** conditional transition probability matrix is linked to reducibility.

4.4 Discussion and Conclusion

Empirical applications of the Markov chain model and spatial Markov chain model can suffer from issues induced by the sparse transition probability matrix which is usually estimated by adopting maximum likelihood estimation (MLE) techniques. The sparsity arises from the generally short length of time series employed in empirical work using spatial data. I propose two discrete kernel estimators with cross validation-based smoothing parameters selection, which are a modification of the smoothing techniques for high-order contingency tables, to address the sparsity issue.

Based on a series of Monte Carlo experiments, it is found that the performance of discrete kernel estimators offers an improvement over traditional MLE approaches when the sample size is small compared to the number of categories in the classic and spatial Markov chain models. More specifically, the smoothed estimators produce nonzero transition probability estimates with smaller MSEs and thus a smaller MSE summed over the whole matrix; they are much better at recovering the irreducible property of a Markov chain if it is inherent in the underlying dynamics; they are also more effective at predicting the steady state distribution assuming the current dynamics last. In addition, I find that the smoothed estimators for the transition probabilities converge to the MLE as the sample size gets larger, which is similar to what have been found for smoothed estimators for contingency tables (Ouyang et al., 2006; Li and Racine, 2007). Therefore, the application of the proposed smoothed estimators would be quite straightforward in that we do not have to determine whether there is a sample size threshold beyond which the MLE should be preferred. Rather, the smoothed estimators could be adopted in both small and large sample settings.

Next steps could be directed to exploring ways to improve the performance of the proposed smoothed estimators. Monte Carlo results indicate that the superiority of the two discrete kernel estimators to MLEs is contingent on the structure of the **true** transition probability matrix such as the number of communicating classes and the extent to which the matrix is diagonally dominant. More specifically, as the true transition probability matrix becomes less diagonally dominant, the two proposed smoothed estimators produce probability estimates with much smaller MSEs, while when the true transition probability matrix is strongly diagonally dominant, their performance is merely similar to (no better than) the MLE. This is probably due to the independence assumption between the two classifications (category at $t-1$, category at t) for the (k, k) transition probability matrix, which could be invalid in the context of Markov chains which are designed to model temporal dependence. This could also be an issue for the spatial Markov chains model. Therefore, the dependence structure between d categorical variables needs to be incorporated in the smoothing process. It could be achieved either by extending the current productive multivariate discrete kernel smoothers or by expanding the d-dimensional smoothing vector to a (d, d) smoothing matrix whose non-diagonal elements control the form of orientation of the kernel (Belaid et al., 2016). For the classic Markov chain model, the dependence structure refers to the temporal dependence between category at $t-1$ and category at t, while for the spatial Markov chain model, the temporal dependence, the cross-sectional dependence (category at $t-1$ and spatial lag category at $t-1$) as well as the spatiotemporal dependence (spatial lag category at $t-1$ and category at t) need to be incorporated.

Chapter 5

CONCLUSION

The dissertation attempts to address issues caused by spatial effects and small sample settings to the *distribution dynamics approach* in studying regional economic growth and convergence in three publishable papers (Chapter 2, 3 and 4). Specifically, I focused on the discrete version of the *distribution dynamics approach*, the discrete Markov chain model, which has been widely adopted in the empirical work.

5.1 Main Findings

Chapter 2 **Inference of Income mobility measures in the presence of spatial dependence** looks at whether the conventional regional income mobility estimators are still suitable if cross-sectional spatial dependence is present or if the sample size is small. The former inflates the asymptotic variance while the latter biases the estimator. For the two-sample test about the mobility difference between two regional economic system, the size tends to become increasingly upward biased with stronger spatial dependence in either income systems, which indicates that conclusions about differences in mobility between two different regional systems need to drawn with caution as the presence of spatial dependence can lead to false positives. In light of this, critical values are suggested to be adjusted for relevant statistical tests.

Next, diagnostic tests for spatial effects in the discrete Markov chain framework are investigated in various settings in Chapter 3 **Conditional and joint tests for spatial effects in discrete Markov chain models of regional income distribution dynamics**. The classic question of differentiating spatial dependence from spatial heterogeneity is explored and the result is not optimistic. Tests for spatial heterogeneity are not robust to that for spatial dependence while the pattern is mixed for tests for spatial dependence to the presence of spatial heterogeneity. When the spatial regimes are comprised of contiguous regions, tests for spatial dependence are not robust as well. The small sample issue is intertwined with the discretization granularity. That is, if the latter is large, meaning we need to estimate a large number of parameters, the small sample issue is more relevant and pressing. Based on a set of Monte Carlo experiments which simulate a spatially explicit vector autoregressive model, we find that all of the test statistics under study display good size properties except for the CSMC likelihood ratio test statistic in small sample settings - it tends to be biased upwards when it is used to test for temporally lagged spatial dependence or spatial heterogeneity. Thus, although it is asymptotically equivalent to the CSMC χ^2 test statistic, its behavior is less satisfactory in small sample settings. When the sample size is large, because increasing the level of discretization granularity lowers the sensitivity of almost all test statistics (except for CSMC heterogeneity tests when dependence is very strong) without compromising the power, it is recommended to adopt a higher level to prevent picking up the "wrong" spatial effect. Otherwise, a balance should be made to preserve the Markov property without impairing estimation precision. In other words, a relatively low granularity strategy should be considered to facilitate estimation, but caution should be taken in case the Markov property is lost due to discretization.

Chapter 4 **Smoothed estimators for Markov Chains with sparse spatial observations** specifically deals with the small sample issue induced by the maximum likelihood estimator (MLE) which is the standard method when it comes to estimating a Markov transition probability matrix. In light of the poor behavior of MLE indicated in the former chapters, as well as the sparsity of the estimated matrix in small sample settings especially for the spatial

Markov chain model, I propose discrete kernel estimators with cross-validated bandwidths as an alternative to MLEs in small sample settings. Based on a series of Monte Carlo experiments, it is demonstrated that the performance of discrete kernel estimators offers improvement over MLEs when sample size is small, giving rise to a better recovery of the true underlying dynamics. These smoothed estimators also tend to converge to MLEs when more spatial observations become available. Therefore, the application of the proposed smoothed estimators would be quite straightforward in that we do not have to determine whether there is a sample size threshold beyond which the MLE should be preferred. Rather, the smoothed estimators could be adopted in both small and large sample settings.

5.2 Limitations and Future Directions

5.2.1 Correcting mobility estimators or test statistics to the presence of spatial dependence

Current correction to regional income mobility estimators focuses on the adjustments of critical values based on the results from a limited number of Monte Carlo simulations. Further research could be directed to the generalization of the adjustments of critical values to incorporate a wider range of cases. Empirical applications of the adjusted one-sample and two-sample tests are of great potential once a general formula is readily available.

Other approaches to accounting for spatial dependence could also be promising. Among them, parametric and nonparametric spatial filtering methods (Anselin, 1988; Getis and Griffith, 2002; Griffith and Chun, 2014) are tractable and commonly used. They treat the spatial dependence as nuisance and attempt to filter out spatially correlated components while leaving the independent components as the input for classic inference. We could also resort to the spatial bootstrap technique (Nordman et al., 2007; Cavaliere et al., 2015) which extends the conventional bootstrap to take account of dependence structure in the resampling process. Besides, we could also follow the Clifford and Richardson "effective sample size" solution for bivariate correlation coefficient (Clifford et al., 1989; Haining, 1991) to seek the *effective sample size* for a decrease in the inflated variance.

5.2.2 Robust Tests for Spatial Effects

As shown in Chapter 3, all test statistics display strong power, but most of them are sensitive to the alternative form of spatial effect they are not designed for. Both CSMC Dep LR and CSMC Dep Q are not robust to the presence of mean heterogeneity, while JSMC Dep is robust if adopting a high level of discretization granularity. CSMC Het LR and CSMC Het Q are not robust to the presence of strong spatial dependence. The lack of robustness poses challenges for the application of the test statistics in empirical studies. Developing robust test (Anselin and Rey, 1991; Anselin, 1990) to aid these 5 test statistics is a promising research direction.

In addition to the non-robustness issue, since a VAR will always introduce contemporaneous spatial dependence if temporally lagged spatial dependence is specified, we could not discriminate one from the other in this setting, nor could we examine the sensitivity of JSMC (CSMC) test to the other form of spatial dependence. Future work may be focused on designing the data generating process which will only introduce one form of spatial dependence based on which an thorough investigation of the robustness of the other test could be conducted.

5.2.3 Improving Smoothed Estimators for Diagonally Dominant Transition Probability Matrix

We note that the superiority of the two discrete kernel estimators to MLEs could be contingent on the structure of the **true** transition probability matrix, such as the number of communicating classes and the extent to which the matrix is diagonally dominant. More specifically, as the true transition probability matrix becomes less diagonally dominant, the two proposed smoothed estimators produce probability estimates with much smaller MSEs, while when the true transition probability matrix is strongly diagonally dominant, their performance is merely similar to (no better than) the MLE. This is probably due to the independence assumption between the two classifications (category at $t-1$, category at t) for the (k, k) transition probability matrix, which could be invalid in the context of Markov chains which are designed to model temporal dependence. This could also be an issue for the spatial Markov chains model. Therefore, the dependence structure between d categorical variables needs to be incorporated in the smoothing process. It could be achieved either by extending the current productive multivariate discrete kernel smoothers or by expanding the d-dimensional smoothing vector to a (d, d) smoothing matrix whose non-diagonal elements control the form of orientation of the kernel (Belaid et al., 2016). For the classic Markov chain model, the dependence structure refers to the temporal dependence between category at $t-1$ and category at t, while for the spatial Markov chain model, the temporal dependence, the cross-sectional dependence (category at $t-1$ and spatial lag category at $t-1$) as well as the spatiotemporal dependence (spatial lag category at $t-1$ and category at t) need to be incorporated. Further research could be conducted in this regard.

5.2.4 Incorporating Continuous Spatial Heterogeneity

Exploring continuous spatial heterogeneity in the discrete Markov chain model could be a promising research area. Currently, empirical studies either estimate one transition probability matrix from the pooled regional income time series, or estimate several matrices for specifically constructed subsamples to account for spatial dependence, discrete spatial heterogeneity or temporal heterogeneity. For incorporating discrete spatial heterogeneity, spatial regimes which are comprised of regions are predefined and assumed to be governed by different transitional dynamics. It is quite possible that the transitional dynamics vary across space giving rise to continuous spatial heterogeneity. Obviously, the small sample issue would be more severe here as N transition probability matrices need to be estimated for N individual regional time series. A promising solution would be to follow the spirit of geographically weighted regression (GWR) (Fotheringham et al., 2002, 2017) and assume the transitional dynamics are more similar for nearby regions, thus facilitating *data-borrowing* from nearby regions. The discrete kernel smoothers with cross-validated bandwidths proposed in Chapter 4 could be properly adjusted to fulfil the purpose.

REFERENCES

Abreu, M., De Groot, H. L., and Florax, R. J. (2004). Space and growth: a survey of empirical evidence and methods.

Abreu, M., De Groot, H. L. F., and Florax, R. J. (2005). A Meta-Analysis of β-Convergence: the legendary 2%. *Journal of Economic Surveys*, 19(3):389–420.

Agresti, A. and Hitchcock, D. B. (2005). Bayesian inference for categorical data analysis. *Statistical Methods & Applications*, 14(3):297–330.

Aitchison, J. and Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3):413–420.

Allison, P. (1978). Measures of inequality. *American Sociological Review*, 43:865–880.

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 28(1):89–110.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.

Anselin, L. (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics*, 20(2):141–163.

Anselin, L. and Rey, S. J. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*, 23(2):112–131.

Arbia, G. (2006). *Spatial econometrics: statistical foundations and applications to regional convergence*. Springer Science & Business Media.

Arbia, G., Basile, R., and Piras, G. (2006). Analyzing intra-distribution dynamics: A reappraisal. Ersa conference papers, European Regional Science Association.

Barro, R. J. (2015). Convergence and modernisation. *The Economic Journal*, 125(585):911–942.

Barro, R. J. (2016). Economic growth and convergence, applied to China. *China World Econ.*, 24(5):5–19.

Barro, R. J. and Sala-i Martin, X. (2003). *Economic growth, 2nd*. MIT Press.

Baumol, W. (1986). Productivity growth, convergence and welfare: What the long run data show. *American Economic Review*, 76(5):1072–1085.

Belaid, N., Adjabi, S., Zougab, N., and Kokonendji, C. C. (2016). Bayesian bandwidth selection in discrete multivariate associated kernel estimators for probability mass functions. *Journal of the Korean Statistical Society*, 45(4):557 – 567.

Bickenbach, F. and Bode, E. (2003). Evaluating the Markov property in studies of economic convergence. *International Regional Science Review*, 26(3):363–392.

Bode, E. and Rey, S. J. (2006). The spatial dimensions of economic growth and convergence. *Papers in Regional Science*, 85:171–176.

Bowman, A. W., Hall, P., and Titterington, D. M. (1984). Cross-validation in nonparametric estimation of probabilities and probability densities. *Biometrika*, 71(2):341–351.

Bulli, S. (2001). Distribution dynamics and cross-country convergence: a new approach. *Scottish Journal of Politcal Economy*, 48(2):226–243.

Burman, P. (1987). Smoothing sparse contingency tables. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 49(1):24–36.

Burman, P. (2004). On some testing problems for sparse contingency tables. *Journal of Multivariate Analysis*, 88(1):1 – 18.

Cavaliere, G., Politis, D. N., and Rahbek, A. (2015). Recent developments in bootstrap methods for dependent data. *Journal of Time Series Analysis*, 36(3):269–271. 10.1111/jtsa.12128.

Chakravarty, S. R. (1995). A note on the measurement of mobility. *Economics Letters*, 48(1):33 – 36.

Ching, W., Fung, E., and Ng, M. (2002). A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. *IMA Journal of Management Mathematics*, 13(3):187–199.

Chu, C.-Y., Henderson, D. J., and Parmeter, C. F. (2015). Plug-in bandwidth selection for kernel density estimation with discrete data. *Econometrics*, 3(2):199–214.

Clifford, P., Richardson, S., and Hemon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45(1):123–134.

Conlisk, J. (1990). Monotone mobility matrices. *Journal of Mathematical Sociology*, 15(3-4):173–191.

Creedy, J. and Wilhelm, M. (2002). Income mobility, inequality and social welfare. *Australian Economic Papers*, 41(2):140–150.

D'Agostino, R. and Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, 60(3):613–622.

Dardanoni, V. (1995). Income distribution dynamics: monotone Markov chains make light work. *Social Choice and Welfare*, 12(2):181–192.

Debarsy, N., Ertur, C., and LeSage, J. P. (2012). Interpreting dynamic space–time panel data models. *Statistical Methodology*, 9(1–2):158 – 171.

Dong, J. and Simonoff, J. S. (1995). A geometric combination estimator for d-dimensional ordinal sparse contingency tables. *The Annals of Statistics*, 23(4):1143–1159.

Durlauf, S. N. (2001). Manifesto for growth econometrics. *Journal of Econometrics*, 100:65–69.

Durlauf, S. N., Johnson, P. A., and Temple, J. R. (2005). Chapter 8 growth econometrics. volume 1, Part A of *Handbook of Economic Growth*, pages 555 – 677. Elsevier.

Egger, P. and Pfaffermayr, M. (2006). Spatial convergence*. *Papers in Regional Science*, 85(2):199–215.

Elhorst, J. P. (2001). Dynamic models in space and time. *Geographical Analysis*, 33(2):119–140.

Ertur, C., Gallo, J. L., and Baumont, C. (2006). The European regional convergence process, 1980-1995: Do spatial regimes and spatial dependence matter? *International Regional Science Review*, 29(1):3–34.

Ertur, C. and Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22(6):1033–1062.

Ertur, C., Le Gallo, J., and LeSage, J. P. (2007). Local versus global convergence in Europe: a Bayesian spatial econometric approach. *The Review of Regional Studies*, 37(1):82.

Fields, G. S. and Ok, E. A. (1996). The meaning and measurement of income mobility. *Journal of Economic Theory*, 71(2):349 – 377.

Fields, G. S. and Ok, E. A. (1999). The measurement of income mobility: an introduction to the literature. In *Handbook of income inequality measurement*, pages 557–598. Springer.

Fienberg, S. E. and Holland, P. W. (1973). Simultaneous estimation of multinomial cell probabilities. *Journal of the American Statistical Association*, 68(343):683–691.

Fischer, M. M. (2011). A spatial Mankiw–Romer–Weil model: theory and evidence. *The Annals of Regional Science*, 47(2):419–436.

Fischer, M. M. (2016). Spatial externalities and growth in a Mankiw-Romer-Weil world: Theory and evidence. *International Regional Science Review*, pages 1–17.

Fischer, M. M. and Stirböck, C. (2006). Pan-european regional income growth and club-convergence. *The Annals of Regional Science*, 40(4):693–721.

Florax, R. J. G. M., Folmer, H., and Rey, S. J. (2003). Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33:557–579.

Formby, J., Smith, W., and Zheng, B. (2004). Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics*, 120(1):181–205.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Fotheringham, A. S., Yang, W., and Kang, W. (2017). Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6):1247–1265.

Getis, A. and Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical Analysis*, 34(2):130–140.

Griffith, D. and Chun, Y. (2014). *Spatial Autocorrelation and Spatial Filtering*, pages 1477–1507. Springer Berlin Heidelberg, Berlin, Heidelberg.

Guihenneuc-Jouyaux, C. and Robert, C. P. (1998). Discretization of continuous Markov chains and Markov chain Monte Carlo convergence assessment. *Journal of the American Statistical Association*, 93(443):1055–1067.

Haining, R. (1991). Bivariate correlation with spatial data. *Geographical Analysis*, 23(3):210–227.

Hammond, G. W. (2004). Metropolitan/non-metropolitan divergence: a spatial Markov chains approach. *Papers in Regional Science*, 83(3):543–563.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied nonparametric econometrics*. Cambridge University Press.

Islam, N. (2003). What have we learnt from the convergence debate? *Journal of Economic Surveys*, 17(3):309–362.

Kang, W. and Rey, S. J. (2018). Conditional and joint tests for spatial effects in discrete Markov chain models of regional income distribution dynamics. *The Annals of Regional Science*.

Khor, N. and Pencavel, J. (2008). Measuring income mobility, income inequality, and social welfare for households of the people's republic of china. Technical report, Asian Development Bank.

Kokonendji, C. C. and Kiessé, T. S. (2011). Discrete associated kernels method and extensions. *Statistical Methodology*, 8(6):497 – 516.

Kullback, S., Kupperman, M., and Ku, H. (1962). Tests for contingency tables and Markov chains. *Technometrics*, 4(4):573–608.

Le Gallo, J. (2004). Space-time analysis of GDP disparities among European regions: A Markov chains approach. *International Regional Science Review*, 27(2):138–163.

Le Gallo, J. and Chasco, C. (2008). Spatial analysis of urban growth in Spain, 1900–2001. *Empirical Economics*, 34(1):59–80.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

LeSage, J. P. and Cashell, B. A. (2015). A comparison of vector autoregressive forecasting performance: spatial versus non-spatial Bayesian priors. *The Annals of Regional Science*, 54(2):533–560.

Lesage, J. P. and Fischer, M. M. (2008). Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis*, 3(3):275–304.

LeSage, J. P. and Krivelyova, A. (1999). A spatial prior for Bayesian vector autoregressive models. *Journal of Regional Science*, 39(2):297–317.

Levine, R. and Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82:942–963.

Li, Q. and Racine, J. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*, 86(2):266 – 292.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Liao, F. H. and Wei, Y. D. (2012). Dynamics, space, and regional inequality in provincial China: A case study of Guangdong province. *Applied Geography*, 35(1–2):71 – 83.

López-Bazo, E., Vayá, E., and Artís, M. (2004). Regional externalities and growth: Evidence from european regions*. *Journal of regional science*, 44(1):43–73.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Maasoumi, E. (1998). On mobility. *Statistics Textbooks and Monographs*, 155:119–176.

Magrini, S. (1999). The evolution of income disparities among the regions of the European Union. *Regional Science and Urban Economics*, 29(2):257–281.

Magrini, S. (2004). Regional (Di)Convergence. In Henderson, V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*. Elsevier, New York.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2):407–437.

Maza, A., Hierro, M., and Villaverde, J. (2012). Income distribution dynamics across European regions: Re-examining the role of space. *Economic Modelling*, 29(6):2632–2640.

McMillen, D. P. and McDonald, J. F. (1991). A Markov chain model of zoning change. *Journal of Urban Economics*, 30(2):257–270.

Nordman, D. J., Lahiri, S. N., and Fridley, B. L. (2007). Optimal block size for variance estimation by a spatial block bootstrap method. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 69(3):468–493.

Ouyang, D., Li, Q., and Racine, J. (2006). Cross-validation and the estimation of probability distributions with categorical data. *Journal of Nonparametric Statistics*, 18(1):69–100.

Parent, O. and LeSage, J. P. (2012). Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics*, 42(4):727 – 738.

Piribauer, P. (2016). Heterogeneity in spatial growth clusters. *Empirical Economics*, 51(2):659–680.

Polansky, A. M. (2009). Using the smoothed bootstrap for statistical inference for markov chains. *Pakistan Journal of Statistics*, 25(4).

Prais, S. J. (1955). Measuring social mobility. *Journal of the Royal Statistical Society Series A*, 118:56–66.

Pu, Y., Ma, R., Ge, Y., and Huang, X. (2005). Spatial-temporal dynamics of regional convergence at county level in Jiangsu. *Chinese Geographical Science*, 15(2):113–119.

Qin, C., Liu, Y., and Li, C. (2013). Spatial spillovers and the convergence of regional economic growth: A case study based on the Yangtze River Delta*. *Social Sciences in China*, 34(3):159–173.

Quah, D. T. (1993a). Empirical cross-section dynamics in economic growth. *European Economic Review*, 37(2-3):426–434.

Quah, D. T. (1993b). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95:427–443.

Quah, D. T. (1996a). Empirics for economic growth and convergence. *European Economic Review*, 40(6):1353 – 1375.

Quah, D. T. (1996b). Twin peaks: Growth and convergence in models of distribution dynamics. *The Economic Journal*, 106:1045–1055.

Quah, D. T. (1997). Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of economic growth*, 2(1):27–59.

Rey, S. J. (2001). Spatial empirics for economic growth and convergence. *Geographical Analysis*, 33(3):195–214.

Rey, S. J. (2015). Discrete regional distribution dynamics revisited. *Journal of Regional and Urban Economics*, 1/2:83–103.

Rey, S. J. and Dev, B. (2006). σ-convergence in the presence of spatial effects. *Papers in Regional Science*, 85(2):217–234.

Rey, S. J. and Gutiérrez, M. L. S. (2015). Comparative spatial inequality dynamics: The case of Mexico and the United States. *Applied Geography*, 61:70–80.

Rey, S. J. and Janikas, M. V. (2005). Regional convergence, inequality and space. *Journal of Economic Geography*, 5(2):155–176.

Rey, S. J., Kang, W., and Wolf, L. (2016). The properties of tests for spatial effects in discrete Markov chain models of regional income distribution dynamics. *Journal of Geographical Systems*, 18(4):377–398.

Rey, S. J., Mack, E. A., and Koschinsky, J. (2012). Exploratory space–time analysis of burglary patterns. *Journal of Quantitative Criminology*, 28:509–531.

Rey, S. J. and Montouri, B. D. (1999). U.S. regional income convergence: A spatial econometric perspective. *Regional Studies*, 33(2):145–156.

Rey, S. J., Murray, A. T., Grubesic, T. H., Mack, E., Wei, R., Anselin, L., and Griffin, M. (2014). Sex offender residential movement patterns: A Markov chain analysis. *The Professional Geographer*, 66(1):102–11.

Rey, S. J. and Ye, X. (2010). Comparative spatial dynamics of regional systems. In Páez, A., Le Gallo, J., and Buluing, R. N., editors, *Progress in Spatial Analysis: Methods and Applications*, pages 441–463. Springer, Berlin.

Ruiz-Castillo, J. (2004). The measurement of structural and exchange income mobility. *The Journal of Economic Inequality*, 2(3):219–228.

Schluter, C. (1998). Statistical inference with mobility indices. *Economics letters*, 59(2):157–162.

Schorrocks, A. (1978). Income inequality and income mobility. *Journal of Economic Theory*, 19:376–393.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):683–690.

Shorrocks, A. (1978). The measurement of mobility. *Econometrica*, 46:1013–1024.

Shorrocks, A. F. (1976). Income mobility and the markov assumption. *The Economic Journal*, 86(343):566–578.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3):613–625.

Silos, P. (2006). Assessing Markov chain approximations: A minimal econometric approach. *Journal of Economic Dynamics and Control*, 30(6):1063–1079.

Simonoff, J. S. (1995). Smoothing categorical data. *Journal of Statistical Planning and Inference*, 47(1):41 – 69. Statistical Modelling.

Simonoff, J. S. (1998). Three sides of smoothing: Categorical data smoothing, nonparametric regression, and density estimation. *International Statistical Review*, 66(2):137–156.

Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70:65–94.

Sommers, P. M. and Conlisk, J. (1979). Eigenvalue immobility measures for Markov chains. *The Journal of Mathematical Sociology*, 6(2):253–276.

Tauchen, G. (1986). Finite state Markov-chain approximations to univariate and vector autoregressions. *Economics letters*, 20(2):177–181.

Teodorescu, I. (2009). Maximum Likelihood Estimation for Markov Chains. *ArXiv e-prints*.

Titterington, D. and Bowman, A. (1985). A comparative study of smoothing procedures for ordered categorical data. *Journal of Statistical Computation and Simulation*, 21(3-4):291–312.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.*, 46(2):234–240.

Trede, M. (1999). Statistical inference for measures of income mobility / statistische inferenz zur messung der einkommensmobilität. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics*, 218(3/4):473–490.

Villaverde, J. and Maza, A. (2012). Chinese per capita income distribution, 1992–2007: A regional perspective. *Asian Economic Journal*, 26(4):313–331.

Wolf, L. J. and Rey, S. J. (2015). On the lumpability of regional income convergence. *Letters in Spatial and Resource Sciences*, pages 1–11.