FOREST COVER CHANGE IN NORTHEAST CHINA DURING THE PERIOD OF 1977-2007 AND ITS DRIVING FORCES

By

Miaoying Shi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Forestry - Doctor of Philosophy

2015

ABSTRACT

FOREST COVER CHANGE IN NORTHEAST CHINA DURING THE PERIOD OF 1977-2007 AND ITS DRIVING FORCES

By

Miaoying Shi

Motivated by asking the question whether or not the large Natural Forest Protection Program (NFPP) had been effective in protecting the natural forests in northeast China. Ten adjacent counties were selected in Sanjiang Plain area of Heilongjiang, upon which region the NFPP had been heavily concentrated. The three chief hypotheses are: (1) the region had undergone severe deforestation and forest degradation before the implementation of NFPP; (2) while the decline of forest cover might have been slowed down following the initiation of NFPP, it would take a longer time to see any significant gain; (3) farmland expansion is the dominant driver of deforestation, whereas population increase, economic growth, and management policy are among the more fundamental forces. Thus the specific tasks were set to detect the regional LUCC over a period of 30 years (1977-2007) and to explore the demographic, economic, political, and other determinants of the detected changes.

Landsat images for six periods were acquired to derive the Land Use Land Cover (LUCC) information. With minor classes being merged, classification resulted in four classes—forestland, farmland, built-up land and other (wetland being a main component). Rule-based rationality evaluation and formal accuracy assessment both proved the classification results are acceptable. The detection results show that: (1) farmland and forestland are the two predominant classes of the regional land use; (2) farmland and built-up land increased persistently during the 30 years; (3) forestland suffered an extended, heavy loss before the end of last century and the decline slowed down significantly thereafter; and (4) "other" land declined continuously. Detailed examination

based on extended conversion matrixes reveals that although forestland experienced the most loss, while wetland suffered the largest proportional reduction. Moreover, the calculated landscape diversity and integrity indexes show that the distribution of land-cover types became more uneven, and land-use patches became more interspersed.

During the investigation the effects of various forces driving deforestation based on series of single equation models, it was found that directly taking farmland as regressor suffer problems, e.g. endogeneity. Thus instrument variables analysis and simultaneous equation modelling were employed to remedy the endogeneity problem and to incorporate the interaction and feedback effects between different land uses.

The outcomes of using the instrumental variable (IV) method were much improved—the coefficients of NFPP is significant, implying that the program has played a positive role in protecting local forests. In addition, the coefficient of the "Forestland-Farmland-Wetland" system are generally consistent with those derived from the IV method. The area of wetland is negatively correlated with the area of forestland, indicating a mutual substitution in farmland expansion; likewise, farmland is negatively correlated with wetland. The significantly positive coefficient of built-up area in the farmland equation suggests a strong link between farming activities and residential construction. The significant negative coefficient of irrigation confirms that wetland loss is adversely affected by the change in local cropping structure. However, due to the limitations of small sample data, estimates could possibly suffer an upward bias while inferences are not reliable.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

2SLS   Two Stage Least Square

3SLS   Three Stage Least Square

ABM   Agent-Based Modelling

AI   Aggregation Index

AIC   Akaike's Information Criterion

BIC   Bayesian Information Criterion

CM   Cellular Model

CONTAG   Contagion Index

COST   Cosine Approximation Model

DOS   Dark Object Subtraction

FE   Fixed Effects

FGLS   Feasible Generalized Least Square

GLS   Generalized Least Square

LSI   Landscape Shape Index

LUCC   Land Use Land Cover Change

MSIDI   Modified Simpson's Diversity Index

MSIEI   Modified Simpson's Evenness Index

NFPP   Natural Forest Protection Program

OLS   Ordinary Least Square

PCA   Principal Component Analysis

PLADJ   Percentage of Like Adjacencies

RE   Random Effects

SEM   Simultaneous Equation Modeling

CHAPTER 1

BACKGROUND, LITERATURE REVIEW, AND RESEARCH OBJECTIVE

1.1 Introduction

Forests in China used to play an important role in the national economy by supplying energy, lumber, and pulp and papers. Like all other sectors, the forest sector has undergone tremendous change over times (Yin 1998; Zhang 2001; Yin et al. 2003). In the northeast state forest region, the growth of logging and forest products manufacturing, and the expansion of agriculture, among other things, had depleted vast natural forests and led to worsening ecological conditions. Subsequently, the government was forced to take drastic policy measures to halt the deforestation and improve the forest condition in the region at the turn of the century (Zhang et al. 2000; Xu et al. 2004).

Nevertheless, some important questions concerning the resource dynamics and factors influencing them remain poorly addressed. These questions include: How severe the regional deforestation and forest degradation had become before the Natural Forest Protection Program (NFPP) was initiated at the end of the 1990s? Whether the forest condition has significantly improved ever since? And what are the major forces that have affected the forest dynamics over time? The goal of this study is to address these questions in a theoretically sound and practically relevant manner. Answering the above questions is not only worthwhile but also important in improving our knowledge of the resource dynamics and environmental consequences and their socioeconomic, policy, and other drivers, and in improving the effectiveness of policy making and implementation and, ultimately, the resource condition. In the following section, I will first briefly examine the major policy changes in China, with particular attention to the northeast state-owned forest region. Then, I will present a literature survey regarding the effects of the NFPP and the driving forces of the forest dynamics in the broader context of the land-use and land-cover change

in the region. Finally, I will outline the analytic tasks that I will undertake in this dissertation project and how the chapters are organized.

1.2 Overview of the Forest History and Policy

China's forest sector, just like its overall economy, witnessed several stages of development since the new republic was founded in 1949 (Wang et al. 2007). A brief overview of the history is beneficial to a clear understanding of the socioeconomic and policy evolution and the associated changes of the resource conditions over time.

After the People's Republic of China was established in 1949, large tracts of primary natural forests remained in the northeast. In 1958, the national wide 'Great Leap Forward' campaign was launched, thousands of inefficient furnaces were built to produce steel and massive forests were destroyed (Zhang 2001). Several years later, state-owned forest bureaus were gradually set up in these forests and nearly 1 million forest workers were dispatched to forested areas to produce timber (SFA 2000; Zhao & Shao 2002). Prior to 1978, under the policy of "Prioritizing Food Production", China emphasized grain production. At the same time, the Chinese Ministry of Forestry had tight control over the forests (Wang et al. 2004). Supplies from both the agricultural and forest sectors were underpriced in order to support the economic development. The state-owned forest companies in northeast China were under the government control, with little freedom related to decision making in forest management. Over-cutting became prevalent and regeneration was neglected. During the period of "Cultural Revolution" in 1966-1977, large-scale deforestation and over-harvesting gradually depleted the natural forest resources in the region (Zhang et al. 2000; Li 2004).

Started in 1978, the economic reform and opening up policy have stimulated economic growth. In the agricultural sector, the introduction of Household Responsible System (HRS) provided incentives for households and thus increased land productivity as well as per-capita incomes. During 1981-1985, the HRS found its way into the forest sector. Due to the long rotation periods and high uncertainty of forestry policies, however, incentives of planting trees were inadequate (Yin 1998; Wang et al. 2007). Despite the repeated upward adjustments of timber prices by the government, the pricing signals failed to reflect societal needs during that time. In northeast China, the rapid national economic growth increased demands for its forest products. There were heavy logging activities. After years of experimenting, the country's first Forest Law officially entered into force in 1984 (Zhang et al. 2000; Wang et al. 2004).

In 1985, the compulsory production quotas and the dual-price system for agricultural products were abandoned. The HRS success in the agricultural sectors provided incentives for a series of policy reforms. Contract Responsibility System (CRS) was developed in the non-agricultural enterprises in rural areas and Township and Village Enterprises (TVEs) emerged under contract with the local administrative authorities (Hyde et al. 2003). Disparities between household incomes increased. In the forest sector, industries producing wood products and pulp and paper grew rapidly in the TVEs. One year after the Forestry Law was enacted, the logging quota system was introduced by the Ministry of Forestry (Wang et al. 2004). In northeast China, the government relaxed its monopolistic role in most state-owned enterprises but continued to control most capital investment decisions. Prices still suffered from distortion in the forest sector, with forest rents arbitrarily captured by downstream manufacturers.

Beginning in 1991, some state-owned enterprises were privatized and some were shut down. Timber prices became mostly market determined and household incomes continued to increase (Yin et al. 2003). In 1989, the Ministry of Forestry reinforced the logging quota system and required that forest growth must exceed timber removal (Zhang et al. 2000; Yu et al. 2011). As a result of a series of reforms in the administrative hierarchy, the state-owned enterprises in the northeast China became more autonomous. Large forest industry groups emerged in the early 1990s; with reduced government control, forest companies were more flexible with responding to market signals and thus improved economic efficiency. Nonetheless, excessive cutting and deforestation continued. According to Yu et al. (2011), about 50% of the matured stands in the northeast disappeared in less than 20 years, with stocking volume falling from 1660 million m^3 in 1981 to 860 million m^3 in 1998. In Heilongjiang province, logging beyond quota limits was most severe, reaching 843,000 m^3, or 31% beyond the allowable quota (MOF 1997). Based on Jiang et al. (2011), the percentage of mature stock in timber forests in Heilongjiang dropped from 65.6% in 1984 to 3.2% in 2004. Muldavin (1997) noted that logging in Heilongjiang caused serious soil erosions, and the "forest 'reserves' consist of barren and rocky yellow subsoils. The slopes, still classified as protected forests, are barren and eroded."

The booming economy along with population expansion has put great pressure on the natural resources and ecosystems. Deforestation, wetland destruction, and farmland degradation have caused severe problems of soil erosion, water shortages, dust storms, and habitat losses over the last few decades (Liu and Diamond 2005; Xu et al. 2006). To combat these problems, the Chinese government has launched several ecological restoration programs since the late 1990s, including the Natural Forest Protection Program (NFPP) and the Sloping Land Conversion Program (SLCP) (Yamane 2001b; Yin & Yin 2010). Among those huge ecological restoration

programs, the NFPP is recognized as one of the largest in terms of geographic scope, financial investment, and number of people impacted (Zhang et al. 2000). The NFPP is also regarded as a far-reaching historic step toward protecting the natural forest resources and carrying out strategic changes in forestry management. It was initiated in the wake of the huge floods of 1998 in the Yangtze River basin and some major waterways in the northeast (Xu et al. 2005). It covers 17 provinces with an initial investment commitment of 96.4 billion (US$14.1 billion) (SFA 2000). The specific goals of the NFPP are to: (1) reduce commercial timber harvests in the natural forests from 32 million m^3 in 1997 to 12 million m^3 by 2003; (2) conserve nearly 90 million ha of natural forests; and (3) afforest and revegetate an additional 8.7 million ha by 2010 by means of mountain closure, aerial seeding, and artificial planting (Liu 2002).

Now the NFPP has entered into its second phase, under a total budget of 244.02 billion yuan (US$38.5 billion). According to the decision made by the State Council, 219.52 billion yuan would be invested by the central government and 24.5 billion by local governments. It is hoped that by 2020, the forestland, stock volume, and carbon sequestration would increase, respectively, by 780 million mu (or 52 million hectares), 1.1 billion cubic meters, and 416 million tons (NFPP Management Center 2011).

1.3 Existing Studies of the NFPP

There have been studies of the effects, as well as the effectiveness, of the NFPP. Xu et al. (2006a) summarized its preliminary economic impacts using Qinhe forest bureau (in Heilongjiang) as an example. Their descriptive statistics showed that from 1998 to 2001, logging and processing revenues together with the local tax incomes had sharply declined. Meanwhile, along with the increased government investments, the earnings of employees in the forest bureau improved while

the local farmers experienced a large decline in their income. As this study was published soon after the NFPP was initiated, the data were insufficient to support a more comprehensive analysis. Later, Zhang et al. (2011) built a panel dataset based on 35 forest farms in northeast China in 2000, 2003, and 2006. The study explored the forest condition change with respect to the new plantation area, the area under protection, and the volume of harvested timber. Their results indicate that the NFPP policy measures, like afforestation, forest protection, and forest management, all have had positive effects. A shortcoming of the study lies in that it assumes the geographic and socioeconomic characters are homogenous in northeast China.

In response, Huang et al. (2010) relaxed the homogeneity assumptions and concluded differently. They formulated three regression equations in a structural model to explore the causes of forest changes in northeast China from 1985 to 2005. They claimed that the socioeconomic factors, like total population, rural population, and GDP, play an influential role in influencing forest dynamics. Also, the geographic and meteorological indicators, like terrain slope, elevation, and climate conditions, are important factors leading to the forest changes. This study provides some interesting results, but its analytical framework is problematic. For example, the whole model is not predicated on any existing theory, and the variable selection seems ad hoc.

A more rigorous model is developed by (Mullan et al. 2009). This study employed two-period survey data from the collective forest areas to estimate the NFPP impact on local household income and labour decision. They took the NFPP as a natural experiment, using the difference-in-differences method to compare the changes between households in the NFPP and non-NFPP areas. Their results suggest that the NFPP has had a negative impact on the income of timber harvesting. And more importantly, the NFPP has stimulated more off-farm labour supply in the NFPP areas than in the non-NFPP areas and made a positive impact on overall household income. However,

data based on two points of time (1997 and 2004) would not capture the whole process of policy implementation. An inherent problem lies in the recall data for the local situations before the introduction of the NFPP.

Jiang et al. (2011) conducted a more convincing analysis, which integrated theoretical analysis and empirical estimation. They analysed the harvest and investment behaviour of the state-owned forest enterprises (SOFEs) under the utility maximization assumption and built a panel dataset based on 75 state-owned forest enterprises in northeast China during 1980-2004 to test their hypothesis. Their results demonstrate that policy measures can have positive effects on the development of forest resources through changing the SOFEs managerial behaviour. Moreover, due to the inability of making significant changes related to employee adjustment and social services provision, the SOFEs "have had relatively few effects on harvest and investment decisions, and on development of the forest resources" (Jiang et al. 2011).

Previous studies have provided useful background information and interesting case descriptions related to the NFPP implementation and impacts. Most research findings indicate that the NFPP has positive impacts on improving local environment and farmers' income, as well as infrastructure and public services. However, their analyses are hardly comprehensive; various aspects of the regional social and natural environments were not clearly examined. First, most papers were based on forest census statistics, while these statistics are generally viewed as being less comprehensive and of lower quality. Thus, rigorous statistical analyses are uncommon (Xu et al. 2005). Second, efforts of studying the NFPP from the perspective of land-use and land-cover change (LUCC) are limited, and long-term comparisons of the forest dynamics induced by policy and other forces are rare. Also, insufficient attention was paid to the dramatic forestland change during the last a few decades, and it still remains unclear whether the natural forests have indeed

been well protected and what are the factors that have led to their depletion and recovery, if any. These observations have motivated my study.

1.4 Review of LUCC in Northeast China

As forestland is part of overall land use and forest cover is part of the regional land cover, it is natural to study forest cover change through the typical LUCC lens.[1] LUCC is a complex process combining natural and social systems through the linkage of human interventions at different temporal and spatial scales (Lambin et al. 2001b; Turner et al. 2008b). Consensus exists in literature that human demand induced social driving forces play a dominant role in LUCC process, and the conversions between farmland, forestland, wetland, etc. are one of the important external display of human activities (Foley et al. 2005; Lambin & Meyfroidt 2011).

LUCC has become a global research thrust as the land surface processes affect ecosystem services and human wellbeing (Foley et al. 2005; Lambin & Geist 2008). It has greatly influenced the soil carbon storage (Post & Kwon 2000; Fargione et al. 2008) and greenhouse emissions (Searchinger et al. 2008), and has contributed to watershed degradation (Sliva & Williams 2001; Tong & Chen 2002), habitat fragmentations (Wang et al. 1997; Fischer & Lindenmayer 2007), and biodiversity losses (Jetz et al. 2007; Kleijn et al. 2009). Meanwhile, the current demographic and economic trends will possibly lead to further degradation of the environmental conditions (Millennium Ecosystem Assessment 2005). Commonly, land use is studied at the regional/local scale. LUCC studies tend to implement scenario-based analyses to identify critical land

---

[1] "Forest cover" refers to the land present with a specified density of trees, not to a land use pertaining to forestry, nor to a continuous representation canopy density (Hansen et al. 2010; Kim et al. 2014). Lund (2006) shows that more than 800 definitions of forests and wooded areas are used around the world. The Food and Agriculture Organization (2015) states that a forest has "a minimum threshold for the height of trees (5 m), at least 10 percent crown cover, and a minimum forest area size (0.5 hectares)."

conversions, and sometimes predict the short- and long-term land-use dynamics. Occasionally, they also explore the proximate and underlying causes (Verburg et al. 2002; Foley et al. 2005). Regional land use studies overview present and past land use histories, recognizing how land uses are interconnected and how they change under human interferences. Developing and implementing regional land study can help foster a vision of land use dynamics in human-dominated ecosystems and shed light on better future land use managements in a fast-developing environment (Foley et al. 2005).

China's forests are unevenly distributed and generally divided into four geographic regions: the northeast state forest region, the northern plains agroforests, the southern collective forest region, and the southwest state forest region (Harkness 1998; Zhang et al. 1999). Among the four regions, the northeast state forest region, which covers Heilongjiang, Jilin, and Liaoning provinces, and the eastern part of Inner Mongolia autonomous region, has the largest natural forests (Zhang et al. 2000; Yu et al. 2011). Within in the region, Heilongjiang, sitting on one of the world's three major black soil zones, is a resource-rich province and used to be the national base of timber and grain. It has the name of "the great northern granary" (Muldavin 1997) and owns the highest percentages of forested land area (40.7 %) (SFA 2005; Yu et al. 2011). The province has gone through extensive landscape changes during the past decades, which has in turn put great pressure on its natural resources and ecosystems. Deforestation, wetland destruction, and farmland degradation have caused severe problems of soil erosion, water shortages, and habitat losses over the last several decades (Xu et al. 2006a; Yin & Yin 2010; Jiang et al. 2011).

While there have been numerous LUCC studies of China, not many of them have been done in the northeast in general, and in Heilongjiang in particularly. Song et al. (2009b) mapped the LUCC in the Amur River basin using MODIS 250 m normalized difference vegetation index

(NDVI), land surface water index (LSWI) time series data in 2001. The study suggested this type of time series data has great potential for large-region LUCC monitoring, but the results lacked sufficient confidence as the spatial resolution was too coarse. Tang et al. (2005) used Landsat images of three periods (1990, 1996 and 2000) to capture the LUCC trajectory of Daqing in Heilongjiang. It is found that the most significant change is wetland degradation and fragmentation, whereas grassland was converted to agriculture.

The study of Huruyama et al. (2009) was based on two-period JERS-1 SAR images (1992 and 1996) in the middle reaches of the Amur River basin. Their results showed that cropland was increasing on all of the geomorphologic landforms, mainly at the expense of wetland on the alluvial plain. Wang et al. (2006) used Landsat MSS and/or TM imagery in three periods of time (1980, 1996 and 2000) to estimate the area changes and the transition of land-use types in the Sanjiang Plain area. The conclusion is similar to that of Huruyama et al. (2009) in terms of the general LUCC trend. Wang et al. (2006) also examined the impact of land-use change on variation in ecosystem services. They found that the total annual ecosystem service value in the the Sanjiang Plain declined by 40% between 1980 and 2000 and this large decline was mainly attributed to the 53.4% loss of wetland. A follow-up paper by the same team (Wang et al. 2009) estimated the impacts of land-use change on regional vegetation productivity in the area. They concluded that the considerable increase of cropland area came mainly from the reclamation of forestland, grassland, and wetland during 2000-2005. Also, they pointed out that the regional LUCC negatively impacted carbon sequestration and food supply.

Because the study areas were selected in the alluvial lowland and the main attention was paid to the wetland, these earlier works are not necessary complete or systematic. Meanwhile, there have been few studies of forestland, despite its importance in Heilongjiang. Further, the

NFPP and other ecological restoration efforts could have exerted great influence on LUCC in this region after 2000. Unfortunately, most of the existing studies dealt with changes before 2000 (Song et al. 2009, Huruyama et al. 2009, Tang et al. 2005, and Wang et al. 2006). LUCC in the region after 2000 has rarely been examined. Therefore, it is interesting to detect the induced land cover change of these initiatives in a timely manner. To capture the reduction of forestland and the expansion of cropland, I will go back as far as the late 1970s when continuously and consistently archived Landsat images became available. This will put the forest degradation, deforestation, and the recent implementation of the NFPP in an appropriate historical context.

1.5 Objectives and Organization

With a focus on forestland, the primary objectives of this study are to examine the underlying land conversion trends in the Sanjiang Plain region of Heilongjiang and to investigate the driving forces of LUCC in general and forestland dynamics in particular. Therefore, the two tasks of this dissertation project are: (1) detecting LUCC dynamics in northeast China for the past 30 years; and (2) exploring the demographic, economic, political, and other determinants of these changes. My hypotheses are: (1) the region had suffered severe deforestation and forest degradation before the Natural Forest Protection Program (NFPP) was initiated; (2) while the decline of forest cover might have been slowed down following the NFPP implementation, it would take a longer time and more effective management measures to see any significant gain in it; and (3) farmland expansion is a direct driver of deforestation, and population increase, economic growth, and management policy are among the more fundamental drivers.

The first task, detecting LUCC dynamics, will be carried out in the next chapter, based on satellite image interpretation using Eardas Imaging and GIS. The general LUCC trends and the

internal land-use interactions will be examined. While I will pay particular attention to LUCC since the late 1990's when various forest and wetland protection and restoration projects were launched, I will trace the regional LUCC back to the late 1970's. By doing so, I will be able to obtain a much longer LUCC series, allowing a more thorough analysis of causal factors and feedback effects in later chapters.

The second task, discerning the LUCC driving forces, will be based on multiple modeling efforts informed by a systematic literature review in Chapter 3, as well as a careful characterization of the regional land-use situation. Using a panel dataset that integrates LUCC information and observations of other social-ecological variables, I will explore various modeling schemes and estimation techniques. The important direct and indirect natural and human-induced causes will be investigated with theoretically sound and empirically practical approaches. Specifically, I will develop reduced-form single-equation models first in Chapter 4 and then more sophisticated strategies, such as instrumental variable method and system of simultaneous equations, in Chapter 5 to explore the LUCC driving forces in general and those of the forestland change in particular.

It is expected that this study will improve the understanding of the dynamics and driving forces of the regional LUCC process and that the integration of multivariate analysis and careful specification of the relationship between the LUCC and its primary drivers will lead to more rigorous findings. Certainly, this will shed light on the important policy question of how to improve the regional land use and environmental management.

REFERENCES

REFERENCES

Fargione, J., Hill, J., Tilman, D., Polasky, S., Hawthorne, P., 2008. Land clearing and the biofuel carbon debt. Science 319, 1235-1238

Fischer, J., Lindenmayer, D.B., 2007. Landscape modification and habitat fragmentation: a synthesis. Global Ecology and Biogeography 16, 265-280

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., 2005. Global consequences of land use. science 309, 570-574

Hansen, M.C., Stehman, S.V., Potapov, P.V., 2010. Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences 107, 8650-8655

Harkness, J., 1998. Recent trends in forestry and conservation of biodiversity in China. The China Quarterly 156, 911-934

Huang, W., Deng, X., Lin, Y., Jiang, Q., 2010. An Econometric Analysis of Causes of Forestry Area Changes in Northeast China Procedia Environmental Sciences 2

Hyde, W.F., Belcher, B.M., Xu, J., 2003. China's forests: global lessons from market reforms. Rff Press.

Jetz, W., Wilcove, D.S., Dobson, A.P., 2007. Projected impacts of climate and land-use change on the global diversity of birds. PLoS Biol 5, e157

Jiang, X., Gong, P., Bostedt, G., Xu, J., 2011. Impacts of Policy Measures on the Development of State-Owned Forests in Northeastern China: Theoretical Results and Empirical Evidence. Environment for Development

Kim, D., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S., Feng, M., Townshend, J.R., 2014. Global, Landsat-based forest-cover change from 1990 to 2000. Remote Sensing of Environment 155, 178-193

Kleijn, D., Kohler, F., Báldi, A., Batáry, P., Concepción, E., Clough, Y., Diaz, M., Gabriel, D., Holzschuh, A., Knop, E., 2009. On the relationship between farmland biodiversity and land-use intensity in Europe. Proceedings of the Royal Society of London B: Biological Sciences 276, 903-909

Lambin, E.F., Geist, H.J., 2008. Land-use and land-cover change: local processes and global impacts. Springer Science & Business Media.

Lambin, E.F., Meyfroidt, P., 2011. Global land use change, economic globalization, and the looming land scarcity. Proceedings of the National Academy of Sciences 108, 3465-3472

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skånes, H., Steffen, W., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C., Xu, J., 2001. The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11, 261-269

Li, W., 2004. Degradation and restoration of forest ecosystems in China. Forest Ecology and Management 201, 33-41

Lund, H.G., 2006. Definitions of forest, deforestation, afforestation, and reforestation. Forest Information Services.

Millennium Ecosystem Assessment, 2005. Ecosystems and human well-being. Island Press Washington, DC.

MOF, 1997. China Forestry Yearbook 1996. China Forestry Publishing House (Ministry of Forestry), Beijing (in Chinese).

Muldavin, J.S., 1997. Environmental degradation in Heilongjiang: policy reform and agrarian dynamics in China's new hybrid economy. Annals of the Association of American Geographers 87, 579-613

Mullan, K., Kontoleon, A., Swanson, T., Zhang, S., 2009. An evaluation of the impact of the Natural Forest Protection Programme on Rural Household Livelihoods. In: An Integrated Assessment of China's Ecological Restoration Programs. Springer, pp. 175-199.

NFPP Management Center, 2011. Authoritative interpretations for the second phase policies of natural forest protection project

Post, W.M., Kwon, K.C., 2000. Soil carbon sequestration and land-use change: processes and potential. Global change biology 6, 317-327

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., Yu, T.-H., 2008. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319, 1238-1240

SFA, 2000. Statistics on the national forest resources (the 5th National Forest Inventory 1994-1998). State Forestry Administration, Beijing (in Chinese).

SFA, 2005. Statistics on the national forest resources (the 6th National Forest Inventory 1999-2003). State Forestry Administration, Beijing (in Chinese).

Sliva, L., Williams, D.D., 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water research 35, 3462-3472

Song, k., Wang, Z., Liu, Q., Lu, D., Yang, G., Zeng, L., Liu, D., Zhang, B., Du, J., 2009. Land use/land cover (LULC) characterizaitoin with MODIS time series data in the Amu River

Basin. In: Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, pp. IV-310-IV-313

Tang, J., Wang, L., Zhang, S., 2005. Investigating landscape pattern and its dynamics in Daqing, China. International Journal of Remote Sensing 26, 2259-2280

Tong, S.T., Chen, W., 2002. Modeling the relationship between land use and surface water quality. Journal of environmental management 66, 377-393

Turner, B.L., Lambin, E.F., Reenberg, A., 2008. Land Change Science Special Feature: The emergence of land change science for global environmental change and sustainability (vol 104, pg 20666, 2007). Proceedings of the National Academy of Sciences of the United States of America 105, 2751-2751

Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., Mastura, S.S., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. Environmental management 30, 391-405

Wang, G., Innes, J.L., Lei, J., Dai, S., Wu, S.W., 2007. China's Forestry Reforms. Science 318, 1556-1557

Wang, L., Lyons, J., Kanehl, P., Gatti, R., 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22, 6-12

Wang, S., Cornelis van Kooten, G., Wilson, B., 2004. Mosaic of reform: forest policy in post-1978 China. Forest Policy and Economics 6, 71-83

Wang, Z., Liu, Z., Song, K., Zhang, B., Zhang, S., Liu, D., Ren, C., Yang, F., 2009. Land use changes in Northeast China driven by human activities and climatic variation. Chinese Geographical Science 19, 225-230

Wang, Z., Zhang, B., Zhang, S., Li, X., Liu, D., Song, K., Li, J., Li, F., Duan, H., 2006. Changes of land use and of ecosystem service values in Sanjiang Plain, Northeast China. Environmental Monitoring and Assessment 112, 69-91

Xu, J., Tao, R., Amacher, G.S., 2004. An empirical analysis of China's state-owned forests. Forest Policy and economics 6, 379-390

Xu, J., Yin, R., Li, Z., Liu, C., 2005. China's ecological rehabilitation: the unprecedented efforts and dramatic impacts of reforestation and slope protection in western China. Ecological Economics 57, 595-607

Xu, J., Yin, R., Li, Z., Liu, C., 2006. China's ecological rehabilitation: Unprecedented efforts, dramatic impacts, and requisite policies. Ecological Economics 57, 595-607

Yamane, M., 2001. China's recent forest-related policies: Overview and background. Policy Trend Report 1, 1-12

Yin, R., 1998. Forestry and the environment in China: the current situation and strategic choices. World Development 26, 2153-2167

Yin, R., Xu, J., Li, Z., 2003. Building institutions for markets: Experiences and lessons from China's rural forest sector. Environment, Development and Sustainability 5, 333-351

Yin, R., Yin, G., 2010. China's primary programs of terrestrial ecosystem restoration: initiation, implementation, and challenges. Environmental management 45, 429-441

Yu, D., Zhou, L., Zhou, W., Ding, H., Wang, Q., Wang, Y., Wu, X., Dai, L., 2011. Forest management in Northeast China: history, problems, and challenges. Environmental management 48, 1122-1135

Zhang, K., Hori, Y., Zhou, S., Michinaka, T., Hirano, Y., Tachibana, S., 2011. Impact of Natural Forest Protection Program policies on forests in northeastern China. Forestry Studies in China 13, 231-238

Zhang, P., Shao, G., Zhao, G., Le Master, D.C., Parker, G.R., Dunning Jr, J.B., Li, Q., 2000. China's forest policy for the 21st century. Science 288, 2135-2136

Zhang, Y., 2000. Costs of Plans vs Costs of Markets: Reforms in China's State-owned Forest Management. Development Policy Review 18, 285-306

Zhang, Y., 2001. Deforestation and forest transition: theory and evidence in China. In: Palo M & Vanhanen H (eds.) World forests from deforestation to transition? Springer, Netherlands, pp. 41-65.

Zhang, Y., Dai, G., Huang, H., Kong, F., Tian, Z., Wang, X., Zhang, L., 1999. The forest sector in China: Towards a market economy. In: World forests, society and environment. Springer, pp. 371-393.

Zhang, y., Li, z., Jiang, l., 2012. Measures on Forest Right System Reform of Local State-Owned Forest Farm in Heilongjiang Province. China Forestry Economy 112, 35-48

Zhao, G., Shao, G., 2002. Logging Restrictions in China: A Turning Point for Forest Sustainability. Journal of Forestry 100, 34-37

CHAPTER 2

LAND USE AND LAND COVER CHANGE IN HEILONGJIANG

2.1 Introduction

China's booming economy, along with its population expansion, has put great pressure on its natural resources and ecosystems. Deforestation, desertification, wetland destruction, and farmland degradation have caused severe problems such as soil erosion, water shortages, dust storms, and habitat losses over the last few decades (Liu & Diamond 2005; Xu et al. 2006b; Yin & Yin 2009). To combat these problems, the Chinese government has launched several ecological restoration programs since the late 1990s. One of these programs is the Natural Forest Protection Program (NFPP), which I have described in Chapter 1. The tremendous efforts to date notwithstanding, it remains questionable whether the existing natural forests have been effectively protected under the NFPP. To address this question, I have selected a primary area of natural forests in northeast China that experienced heavy logging and farming expansion in the three decades prior to the program as the focus of this study (Yin 1998).

While there have been many studies of China's land use/land cover change (LUCC), few of them have been done in the northeast, especially the forest ecosystems in Heilongjiang. As discussed in last chapter, a large portion of the literature has concentrated on wetland in the region, with study sites mostly located in the Sanjiang and Armu river basins (Tang et al. 2005; Wang et al. 2006; Song et al. 2009a). These studies have concluded that wetland degradation and fragmentation was widespread in the region, but they have not provided sufficient insight into changes in forestland. Also, most of the existing studies dealt with changes that occurred before the year 2000, which means that they were not able to consider the NFPP and other ecological restoration efforts that could have potentially influenced the regional LUCC in recent years (Zhang et al. 2003; Liu et al. 2004; Tang et al. 2005; Wang et al. 2006; Song et al. 2009a; Wang et al. 2009;

Gao & Liu 2011). So, it is interesting and beneficial to detect the land cover change induced by these initiatives in a timely manner.

To capture the early reduction of forestland and the expansion of farmland, I will go back as far as the late 1970s, when continuously archived Landsat images became available. Considering both relevance and feasibility, I have selected 10 adjacent counties in Heilongjiang province as my study site (see Figure 2.1). Heilongjiang lies in the corner of northeast China bordering Russia by the Amur River in the north and the Wusuli River in the east. Forests in Heilongjiang are made up of three large tracks—the Yichun Forest Region, the Mudanjiang Forest Region, and the Daxinanling Forest Region—covering nearly 36% of the province's total land area.

The 10 counties selected for this study are located in the Mudanjiang Forest Region. They are: Fangzheng, Yilan, Huachuan, Huanan, Jixian, Shuangyashan, Qitaihe, Suibin, Youyi and Boli. The whole area ranges from 128.15°-132.33°E and 45.32°-47.45°N, and covers about 29,029 square kilometers. Relatively flat and low in altitude, this area includes a large part of the Sanjiang Plain, which consists of alluvial deposits from the Amur, Songhua, and Wusuli rivers. The Songhua River has two primary tributaries—the Mudan and Mayi rivers. The study site features a temperate continental monsoon climate. Winters are long and bitter, with an average temperature of −31 to −15°C in January, and summers are short and warm with an average temperature of 18 to 23°C in July. The annual rainfall ranges from 400 to 700 millimeters, concentrated mostly in summer. The main soils are black soil (*Luvic Phaeozem*), chernozem (*Haplic Chernozem*), and meadow soil (*Eutric Vertisol*). Agricultural production is mainly based upon such crops as rice, soybeans, maize, and wheat. Cash crops include beets, flax, and sunflowers.

Figure 2.1 Study site in Heilongjiang, northeast China
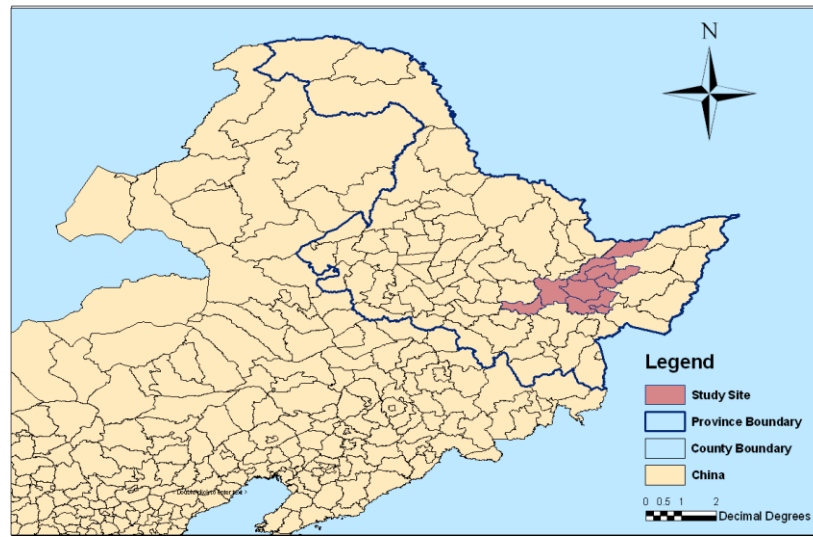
The most common tree species in the study site are Dahurian larch (*Larix gmelinii*), white birch (*Betula papyrifera)*, and Mongolian oak (*Quercus macrocarpa*). Like all other natural forests in China, forests in this region have undergone tremendous changes over time. Massive deforestation in Heilongjiang started in the 1920's following the construction of the "Mid-eastern" Railway. According to Yin (1998), large tracts of primary natural forests still remained after the People's Republic of China was established in 1949, but in order to spur the young economy, over-cutting became prevalent without enough incentive and autonomy from local forest farms to manage and utilize the resources efficiently. In the meantime, population and employment expansion in these forest regions led to more fuelwood consumption, housing construction, and land clearing. (Yamane 2001a) estimated that timber extraction from northeast China accounted for more than 40% of the total national log production in the 1970s and over 20 percent thereafter.

2.2 Data and Methodology

2.2.1 Pre-Classification Preparations and Classification Processes

For my study, Landsat images for six periods were acquired, covering the time span of the late 1970s to 2007. They include two sets of MSS images for the late 1970s (roughly 1977) and 1984; three sets of TM images for 1993, 2004, and 2007; and one set of ETM+ images for 2000. The images for the first four periods were downloaded from the United States Geological Survey website (U.S. Department of the Interior 2009). The images for 2004 were ordered from the China Remote Sensing Satellite Ground Station (Chinese Academy of Sciences 2008). For each period, there are three Landsat scenes to cover the entire study area. Notably, due to quality concerns, images for a given year may not be useable, in which case, a common practice is to assemble them around a given year as closely as possible. Also, due to the low quality of ETM+ images for 2004 and 2007, TM images are used instead.

The images were first georeferenced and rectified by the GLCF to UTM projection zone 52 and WGS84 datum. Based on the image-to-image registration method, I manually geo-encoded and matched the 1984 MSS images and the 2004 ETM+ images one by one using a second order polynomial transformation with an average root mean square error (RMSE) of less than 0.5 pixel units. Since atmospheric influences are acute to multi-temporal studies of land cover change, I employed the cosine approximation model (COST) to correct the ETM+ and TM images (Chavez 1996; Song et al. 2001), and Dark-Object-Subtraction (DOS) method to correct the MSS images (Chavez 1996). Then, the geometrically corrected and radiometrically calibrated images were cropped to the extent of the study area.

Before classification, the Principal Component Analysis (PCA) method was used to account for over 98% of the variance (Deng et al. 2008). Then, the PCA-enhanced images were first classified using unsupervised classification. Initially, a modified version of the U.S. Geological Survey Land Use/Land cover Classification System was employed (Anderson et al. 1976), which includes nine classes—farmland (dry land and paddy land), forestland (dense forest and sparse forest), grassland (dense grass and sparse grass), water body, built-up land, and unused land. Fieldwork was carried out from May to June 2010 to gain better knowledge of the study area and improve the accuracy of the LUCC maps obtained from my classification. During the classification process, ground truth knowledge can help to identify what is actually present and group the various sub-categories together into land use categories more accurately. Still, as the classification went on, I found that some classes of land use, like grassland, are quite difficult to be identified and differentiated from unused land. Finally, I decided to merge those minor categories because they are not the focus of this study. These minor categories of land use, including water bodies, wetland, and grassland add up to less than 4% of the whole region. As a result, the classes of land use examined in this study were reduced to four—farmland, forestland, built-up area, and "other."

2.2.2 Post-Classification Analysis

A traditional "conversion matrix", which is also called a "transition matrix", is commonly employed to demonstrate the land use transitions. Rows in a conversion matrix display the categories of the staring period, and the columns display the categories of the ending period. Entries on the diagonal line indicate the persistence of each category, while those off the diagonal line indicate transitions from the row category to the column category.

Pontius et al. (2004) pointed out a deficiency underlying the traditional conversion matrix. For example, if 12.9% of the total land area was transformed from forest to farmland or built-up land (or F&B), does this indicate that the most systematic process of the land-use transition is from forestland to F&B? Pontius et al. (2004) demonstrated that it is not necessarily so. To answer the question properly, they proposed to consider the size of each land use category. For a particular category of land use, the changes of land-use are mainly about "gains" and "losses". Pontius et al. (2004) calculated the expected value representing a random process of gain based on Equation 2.1 below. This equation assumes the gain of each land-use category is fixed, and this gain is then distributed across other categories according to the relative proportions of other categories at time 1 (the starting time point of land-use change in the matrix); that is, the gain in each column is distributed among the off-diagonal entries within that column. In Equation 2.1, i stands for row and j for column, so P_{i+} stands for the total percentage of in row i and P_{+j} for the total percentage in column j.

$$G_{ij} = (P_{+j} - P_{jj})(\frac{P_{i+}}{\sum_{i=1,i\neq j}^{J} P_{i+}})$$ (Eq. 2.1)

Similarly, Pontius et al. (2004) generated a table for losses of different classes of land use. The expected percentages of the loss in a category were random, as given by

$$L_{ij} = (P_{i+} - P_{ii})(\frac{P_{i+}}{\sum_{i=1,i\neq j}^{J} P_{i+}})$$ (Eq. 2.2)

where L_{ij} represents the loss on the off-diagonal cells in conversion matrix. Eq. 2.2 assumes the loss in each category of land use is fixed, and distributes the loss across other categories according to the relative proportions of the other categories at time 1. In my study, I chose to use time 1 where Pontius et al. (2004) calculated the loss based on the relative proportion of other categories

at time 2. This is because when one category is replaced by a combination of other categories through random processes, it should be based on how those categories populate the landscape *in situ*, not on the landscape structure *in future*. These extended conversion matrixes of specific gains and losses provide more detailed information than one can get from the conventional conversion matrixes.

Another difficulty with the common conversion matrixes is that it is possible for changes to occur within a class of land use while its aggregate quantity remains the same, which is not represented clearly in this type of matrixes. For example, forests could be cleared in some places while the same amount of forest could be gained elsewhere. (Pontius Jr et al. 2004) called this kind of change a "swap." Thus, swap (locational change) and net change (quantity change) together represent a composite of the total changes of LUCC transitions.

2.3 Results

Two accuracy assessment methods were employed to validate the classification results—the rule-based rationality evaluation technique (Liu & Zhou 2004) and the spatially balanced sampling method (SBS) (Foody 2009a). The assessment results demonstrate that my classification results are fairly robust and accurate. Appendix A and Appendix B reports the assessment details. Figure 2.2 shows the trajectories of the changes in the four land-use classes from 1977 to 2007. Farmland and built-up land increased dramatically, while forestland and the other lands declined sharply during the 30 years. Examining these changes in more detail, in 1977 there were only 14,301 km^2 of farmland and built-up land combined; this figure increased to 17194 km^2 by 2007. Forestland experienced a dramatic decrease during the period. In 1977, forestland amounted to 12,294 km^2, but it shrank to only 9,509 km^2 in 2007, showing a more than 20% loss. After the

introduction of the NFPP, the rate of forestland decrease became much smaller, suggesting a stabilization of forest cover.

Figure 2.2 LUCC trajectories during 1977-2007
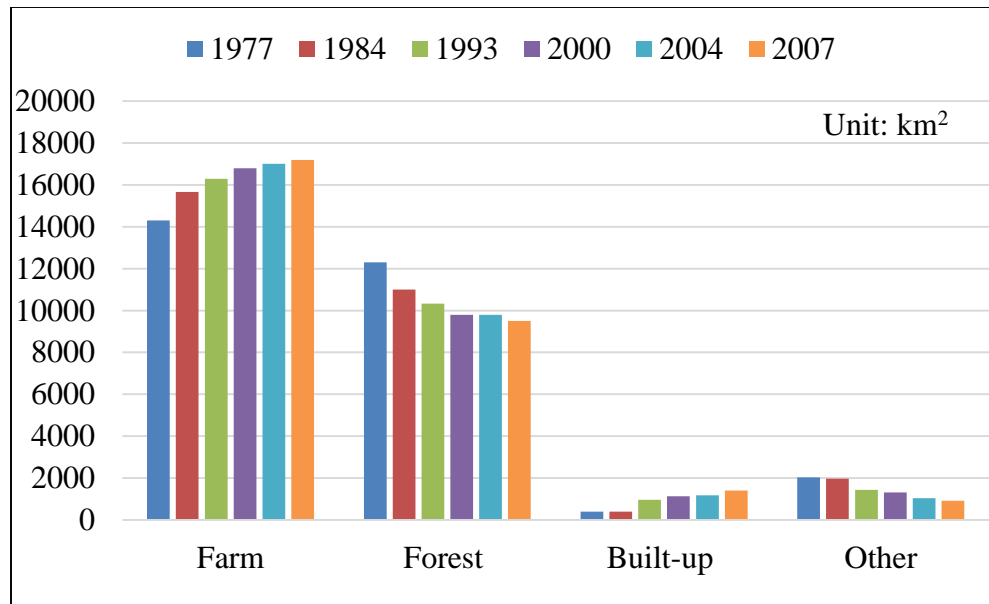
Note: The "other" category, including water bodies, wetland, and grassland, decreased persistently during the 30-year study period.

Tables 2.1 and 2.2 are the extended conversion matrixes with specific gains and losses. Due to a confusion between farmland and built-up land during the MSS data classification process, I merged these two classes for the period of 1977-1984 for a clearer presentation. Each block in these tables contains four values, listed vertically: (1) the observed value, (2) the expected value, (3) the difference between the observed and expected value, and (4) the percentage ratio of difference calculated by dividing the difference by the expected amount of land conversion and multiplied by 100 percent.

Table 2.1 Percentages of land-use changes during 1977-2007 based on Equation 2.1
(The expected gain represents a random process)

	2007			1977 Total	Losses
	F&B	**Forest**	**Other**		
1977					
F&B	**47.40**	**2.76**	**0.45**	**50.62**	**3.22**
	47.40	2.96	0.31	50.67	3.26
	0.00	*-0.20*	*0.15*	*-0.05*	*-0.05*
	0.00	*-6.62*	*48.07*	*-0.10*	*-1.51*
Forest	**12.86**	**29.39**	**0.11**	**42.35**	**12.97**
	14.31	29.39	0.26	43.95	14.56
	-1.45	*0.00*	*-0.15*	*-1.60*	*-1.60*
	-10.13	*0.00*	*-57.45*	*-3.63*	*-10.96*
Other	**3.82**	**0.61**	**2.60**	**7.03**	**4.43**
	2.37	0.41	2.60	5.38	2.79
	1.45	*0.20*	*0.00*	*1.65*	*1.65*
	61.05	*47.71*	*0.00*	*30.57*	*59.09*
2007 Total	**64.09**	**32.76**	**3.16**	**100.00**	**20.61**
	64.09	32.76	3.16	100.00	20.61
	0.00	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*
Gains	**16.68**	**3.37**	**0.56**	**20.61**	
	16.68	3.37	0.56	20.61	
	0.00	*0.00*	*0.00*	*0.00*	
	0.00	*0.00*	*0.00*	*0.00*	

Note: (1) Due to the coarse image resolution, there was confusion regarding farmland and built-up land in 1977, so these two classes were merged into "F&B" to allow for a more meaningful comparison between different time periods. (2) The bold figures are the observed percentages, and the regular figures are the expected percentages under the assumption that the loss to each category is random; the figures in both bold and italics are the difference between the observed and the expected values; and the figures in italics only are the result of the differences divided by the expected values, multiplied by 100 percent.

Table 2.2 Percentages of land-use changes during 1977-2007 based on Equation 2.2
(The expected loss represents a random process)

| | 2007 | | | 1977 | Losses |
	F&B	Forest	Other	Total	
1977					
F&B	**47.40**	**2.76**	**0.45**	**50.62**	**3.22**
	47.40	2.76	0.46	50.62	3.22
	0.00	*0.01*	*-0.01*	*0.00*	*0.00*
	0.00	*0.21*	*-1.29*	*0.00*	*0.00*
Forest	**12.86**	**29.39**	**0.11**	**42.35**	**12.97**
	11.39	29.39	1.58	42.35	12.97
	1.47	*0.00*	*-1.47*	*0.00*	*0.00*
	12.93	*0.00*	*-93.13*	*0.00*	*0.00*
Other	**3.82**	**0.61**	**2.60**	**7.03**	**4.43**
	2.41	2.02	2.60	7.03	4.43
	1.41	*-1.41*	*0.00*	*0.00*	*0.00*
	58.51	*-69.93*	*0.00*	*0.00*	*0.00*
2007 Total	**64.09**	**32.76**	**3.16**	**100.00**	**20.61**
	61.20	34.16	4.64	100.00	20.61
	2.88	*-1.41*	*-1.48*	*0.00*	*0.00*
	4.71	*-4.11*	*-31.88*	*0.00*	*0.00*
Gains	**16.68**	**3.37**	**0.56**	**20.61**	
	13.80	4.78	2.04	20.61	
	2.88	*-1.41*	*-1.48*	*0.00*	
	20.90	*-29.43*	*-72.51*	*0.00*	

Note: See the note beneath Table 2.1 for definitions of the different values.

A positive difference between expectation and observation indicates that the category in that row lost more to the category in the column than would be predicted by a truly random process of gain (or loss). Obviously, the two largest categories of land use are F&B and forest in the region. Forestland experienced the largest loss—12.97% of the total landscape—while F&B had the largest gain—16.68% of the total landscape. During 1977-2007, 20% of the study site underwent land use changes. From Tables 2.1 and 2.2, I further generated a summary of the systematic land-use transitions during 1977-2007, as shown in Table 2.3. In the region, farmland and other land converted mutually, with farmland having gained more from other. Even though a large amount of

forestland was converted into farmland, it appears that forestland was not the first option for farmers to reclaim.

Table 2.3 Land-use transitions, 1977-2007

LUCC transition	Important Transition		Diff	Diff %	Interpretation
	1977	2007			
Gains	F&B	Other	0.15	48.07	Other gains, it replaces F&B more
	Forest	F&B	-1.45	-10.13	F&B gains, it replaces forest less
	Forest	Other	-0.15	-57.45	Other gains, it replaces forest less
	Other	F&B	1.45	61.05	F&B gains, it replaces other more
	Other	Forest	0.20	47.71	Forest gains, it replaces other more
Losses	Forest	F&B	1.47	12.93	Forest loses, F&B replaces it more
	Forest	Other	-1.47	-93.13	Forest loses, other replaces it less
	Others	F&B	1.41	58.51	Other loses, F&B replaces it more
	Others	Forest	-1.41	-69.93	Other loses, forest replaces it less

Note: (1) "Diff" is the difference between observed and expected values. "Diff %" is the percentage of difference calculated in the previous column divided by expected amount of land conversion. F&B stands for the combination of farmland and built-up area. (2) In the interpretation column, "more" means "more than as expected and "less" means "less than as expected".

To understand the LUCC before and after the introduction of the NFPP, I selected two time periods—the period of 1993-2000 and the period of 2000-2007—in my assessment. Due to the better quality of TM/ETM images for these periods, built-up land can be more easily differentiated from farmland, and thus comparisons can be made more thoroughly. Also, to save space, I merged the "gain" and "loss" conversion matrixes. Table 2.4 shows the matrix for the period 1993-2000, and that for the period of 2000-2007 is displayed in Table 2.5.

Table 2.4 Percentages of land change in terms of gains and losses, 1993-2000

1993	Farm Gain	Farm Loss	Forest Gain	Forest Loss	Built-up Gain	Built-up Loss	Other Gain	Other Loss	1993 Total Gain	1993 Total Loss	Losses Gain	Losses Loss
Farm	**50.25**		**3.74**		**0.80**		**1.34**		**56.11**		**5.87**	
	50.25	50.25	3.45	4.76	0.55	0.45	0.84	0.66	55.10	56.11	4.85	5.87
	0.00	*0.00*	*0.28*	*-1.02*	*0.24*	*0.35*	*0.49*	*0.67*	*1.02*	*0.00*	*1.02*	*0.00*
	0.00	*0.00*	*8.24*	*-21.48*	*43.72*	*77.90*	*58.37*	*101.58*	*1.85*	*0.00*	*21.01*	*0.00*
Forest	**5.66**		**29.76**		**0.07**		**0.08**		**35.58**		**5.82**	
	6.17	5.07	29.76	29.76	0.35	0.30	0.54	0.45	36.81	35.58	7.05	5.82
	-0.50	*0.59*	*0.00*	*0.00*	*-0.28*	*-0.23*	*-0.45*	*-0.36*	*-1.23*	*0.00*	*-1.23*	*0.00*
	-8.14	*11.70*	*0.00*	*0.00*	*-79.29*	*-75.95*	*-84.24*	*-81.17*	*-3.34*	*0.00*	*-17.46*	*0.00*
Built-up	**0.38**		**0.02**		**2.93**		**0.01**		**3.34**		**0.42**	
	0.58	0.24	0.21	0.15	2.93	2.93	0.05	0.02	3.76	3.34	0.84	0.42
	-0.20	*0.14*	*-0.18*	*-0.13*	*0.00*	*0.00*	*-0.04*	*-0.01*	*-0.42*	*0.00*	*-0.42*	*0.00*
	-33.89	*58.99*	*-88.57*	*-84.61*	*0.00*	*0.00*	*-83.23*	*-60.39*	*-11.17*	*0.00*	*-50.32*	*0.00*
Other	**1.56**		**0.20**		**0.09**		**3.11**		**4.96**		**1.85**	
	0.86	1.09	0.31	0.69	0.05	0.06	3.11	3.11	4.33	4.96	1.21	1.85
	0.70	*0.47*	*-0.10*	*-0.49*	*0.04*	*0.02*	*0.00*	*0.00*	*0.63*	*0.00*	*0.63*	*0.00*
	81.25	*42.93*	*-33.47*	*-70.63*	*74.19*	*31.21*	*0.00*	*0.00*	*14.62*	*0.00*	*52.12*	*0.00*
2000 Total	**57.85**		**33.72**		**3.88**		**4.54**		**100.00**		**13.95**	
	57.85	56.65	33.72	35.36	3.88	3.74	4.54	4.25	100.00	100.00	13.95	13.95
	0.00	*1.20*	*0.00*	*-1.64*	*0.00*	*0.14*	*0.00*	*0.30*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*2.12*	*0.00*	*-4.64*	*0.00*	*3.72*	*0.00*	*7.00*	*0.00*	*0.00*	*0.00*	*0.00*
Gains	**7.61**		**3.96**		**0.95**		**1.43**		**13.95**			
	7.61	6.40	3.96	5.60	0.95	0.81	1.43	1.13	13.95	13.95		
	0.00	*1.20*	*0.00*	*-1.64*	*0.00*	*0.14*	*0.00*	*0.30*	*0.00*	*0.00*		
	0.00	*18.80*	*0.00*	*-29.27*	*0.00*	*17.09*	*0.00*	*26.23*	*0.00*	*0.00*		

Table 2.5 Percentages of land change in terms of gains and losses, 2000-2007

2000	2007 Farm Gain	Farm Loss	Forest Gain	Forest Loss	Build-up Gain	Build-up Loss	Other Gain	Other Loss	2000 Total Gain	2000 Total Loss	Losses Gain	Losses Loss
Farm	**52.84**		**3.79**		**0.96**		**0.26**		**57.85**		**5.01**	
	52.84	52.84	3.70	4.01	0.65	0.46	0.20	0.54	57.39	57.85	4.55	5.01
	0.00	*0.00*	*0.09*	*-0.22*	*0.31*	*0.50*	*0.06*	*-0.28*	*0.46*	*0.00*	*0.46*	*0.00*
	0.00	0.00	2.42	-5.44	48.23	107.37	30.49	-51.33	0.81	0.00	10.17	0.00
Forest	**5.09**		**28.52**		**0.05**		**0.07**		**33.72**		**5.21**	
	5.11	4.55	28.52	28.52	0.38	0.30	0.12	0.36	34.12	33.72	5.61	5.21
	-0.02	*0.54*	*0.00*	*0.00*	*-0.33*	*-0.25*	*-0.05*	*-0.29*	*-0.40*	*0.00*	*-0.40*	*0.00*
	-0.45	*11.96*	*0.00*	*0.00*	*-86.45*	*-83.29*	*-42.53*	*-81.10*	*-1.17*	*0.00*	*-7.11*	*0.00*
Built-up	**0.09**		**0.01**		**3.78**		**0.00**		**3.88**		**0.10**	
	0.59	0.06	0.25	0.03	3.78	3.78	0.01	0.00	4.63	3.88	0.85	0.10
	-0.50	*0.03*	*-0.24*	*-0.03*	*0.00*	*0.00*	*-0.01*	*0.00*	*-0.75*	*0.00*	*-0.75*	*0.00*
	-85.04	49.04	-96.76	-76.61	0.00	0.00	-84.85	-55.84	-16.23	0.00	-88.46	0.00
Other	**1.21**		**0.44**		**0.06**		**2.83**		**4.54**		**1.72**	
	0.69	1.04	0.29	0.61	0.05	0.07	2.83	2.83	3.86	4.54	1.03	1.72
	0.52	*0.17*	*0.15*	*-0.17*	*0.01*	*-0.01*	*0.00*	*0.00*	*0.69*	*0.00*	*0.69*	*0.00*
	76.01	*16.42*	*51.80*	*-27.32*	*27.56*	*-7.45*	*0.00*	*0.00*	*17.84*	*0.00*	*66.80*	*0.00*
2007 Total	**59.23**		**32.76**		**4.86**		**3.16**		**100.00**		**12.03**	
	57.85	58.49	33.72	33.17	3.88	4.62	4.54	3.73	100.00	100.00	12.03	12.03
	1.38	*0.74*	*-0.97*	*-0.41*	*0.97*	*0.24*	*-1.39*	*-0.57*	*0.00*	*0.00*	*0.00*	*0.00*
	2.38	*1.27*	*-2.87*	*-1.24*	*25.10*	*5.11*	*-30.50*	*-15.27*	*0.00*	*0.00*	*0.00*	*0.00*
Gains	**6.39**		**4.24**		**1.07**		**0.33**		**12.03**			
	7.61	5.65	3.96	4.65	0.95	0.84	1.43	0.90	13.95	12.03		
	-1.22	*0.74*	*0.28*	*-0.41*	*0.12*	*0.24*	*-1.10*	*-0.57*	*-1.92*	*0.00*		
	-16.00	*13.17*	*6.96*	*-8.83*	*12.48*	*28.24*	*-76.76*	*-63.14*	*-13.76*	*0.00*		

In the first period (1993-2000), 13.95% of the landscape was transformed. Farmland gained 7.61% and lost 5.87%, and forestland gained 3.96% and lost 5.82%. Built-up area increased by 0.53% and other decreased by 0.42%, respectively. In the second period (2000-2007), the total gain of farmland was 6.39% while its total loss was 5.01%, leading to a net gain of 1.38%. Forestland experienced a smaller net loss than in the prior period, with a total gain of 4.24% and a total loss of 5.21%.

As stated before, the LUCC statistics do not mean only quantity changes but also locational transformation. Table 2 demonstrates the percentages of losses, gains, net changes, and swaps of the three categories of land use. Among them, a major swap occurred in forestland—6.74% of the total land that was forestland in 1977 was cleared and reforested by 2007. To better understand the LUCC in Heilongjiang and the NFPP's effects, I aggregated the LUCC transitions.

Table 2.6 Percentages of gains, losses, net changes, and swaps of the land use categories, 1977-2007

	1977	2007	Gains	Losses	Total Change	Net	Swap
F&B	50.62	64.09	16.68	3.22	19.90	13.47	6.43
Forest	42.35	32.76	3.37	12.97	16.34	9.60	6.74
Other	7.03	3.16	0.56	4.43	4.99	3.87	1.12
Total	100.00	100.00	20.61	20.61	41.23	26.93	14.29

Table 2.7 Percentages of gains, losses, net changes, and swaps of the land use categories in 1977-2000 and 2000-2007

Period	Classes	Time 1	Time 2	Gains	Losses	Total Change	Net	Swap
	Farm	56.11	57.85	7.61	5.87	13.48	1.74	11.74
	Forest	35.58	33.72	3.96	5.82	9.78	1.86	7.93
1993-2000	Built-up	3.34	3.88	0.95	0.42	1.37	0.54	0.83
	Other	4.96	4.54	1.43	1.85	3.28	0.42	2.86
	Total	100	100	13.95	13.95	27.90	4.55	23.36
	Farm	57.85	59.23	6.39	5.01	11.40	1.38	10.02
	Forest	33.72	32.76	4.24	5.21	9.45	0.97	8.48
2000-2007	Built-up	3.88	4.86	1.07	0.10	1.17	0.97	0.20
	Other	4.54	3.16	0.33	1.72	2.05	1.39	0.66
	Total	100	100	12.03	12.03	24.07	4.71	19.36
	Farm	-1.74	-1.38	1.22	0.86	2.08	0.36	1.72
	Forest	1.86	0.96	-0.28	0.61	0.33	0.89	-0.55
Difference	Built-up	-0.54	-0.98	-0.12	0.32	0.20	-0.43	0.63
	Other	0.42	1.38	1.10	0.13	1.23	-0.97	2.20
	Total	0.00	0.00	1.92	1.92	3.83	-0.16	4.00

Note: Differences resulted from the values from 1993-2000 minus the values from 2000-2007.

To better examine the effects of the NFPP, I selected two time periods that bookended its introduction in 2000—the period of 1993-2000 and the period of 2000-2007. We can get quite amount information from Table 2.7. Built-up land expanded considerably during 2000-2007, with a net increase of 0.43%. There is also a small increase in other land, which means there was a small gain in wetland, or grassland, etc. In particular, there are two important messages conveyed in the "Difference" block. Firstly, we can see that forestland gained more and lost less in the period of 2000-2007 and the net change is smaller in the period of 2000-2007 compared to the period of 1993-2000. Meanwhile, larger swap change in 2000-2007 suggests local farmers reforested more than before, which could result from a large area of reforestation as well as agriforestation in most farmland-dominant counties, like Suibin and Youyi.

2.4 Conclusion

In this chapter, I have assessed the temporal dynamics of LUCC in Heilongjiang between 1977 and 2007. My LUCC classification results show that the study region underwent enormous land-use changes during 1977-2007. A large quantity of forestland was converted into farmland, and built-up land increased continuously. Moreover, by taking the relative land use sizes into consideration, the extended conversation matrixes reveal that wetland and grassland tended to be the first targets to be converted into farmland, but the loss of forestland was much larger. Based on these findings, the dominant land use conversions are depicted in Figure 2.3.

Figure 2.3 Relationship between the two major land-use classes
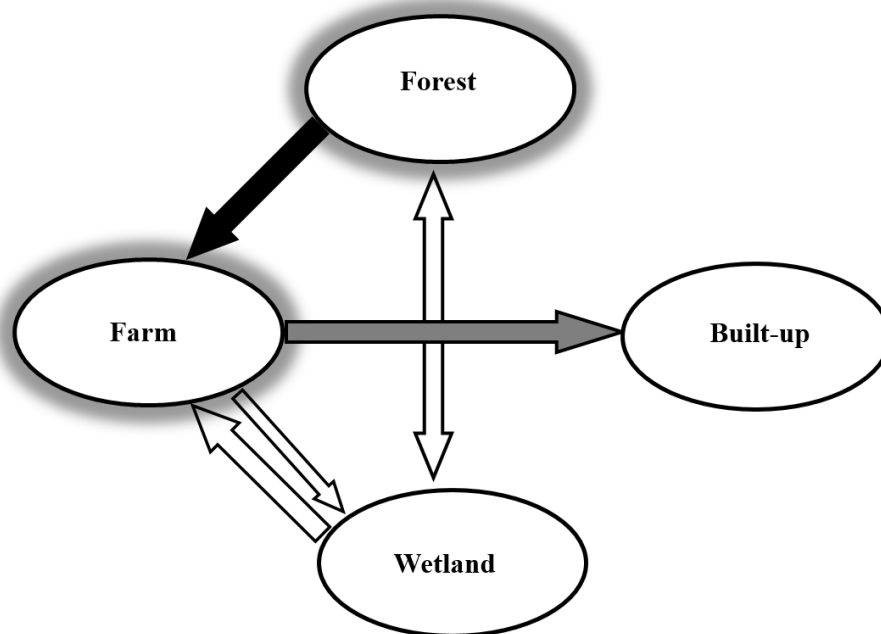
To better examine the effects of the NFPP, I also made a close comparison using images of two neighboring periods before and after 2000—1993-2000 and 2000-2007. My results show that there was no net increase in forestland following the initiation of the NFPP, but the decrease slowed down in the latter period. Also, the location change in forestland was larger after 2000, which

implies that local farmers reforested more than before the NFPP. These are good signs regarding the initial impact of the program.

In this chapter, I have tried to understand the LUCC transitions in different ways. This knowledge of the regional LUCC is of scientific and policy significance. Based on the detected LUCC outcomes, I am now in a position to take further steps to investigate the driving forces of the LUCC—what factors have affected the LUCC dynamics and to what extent. Certainly, my central focus will be on the determinants of forest cover change, including the potential impact of the NFPP. Before undertaking these tasks, though, a careful review of the literature of LUCC driving force analysis is called for so that I can better understand the various theoretical frameworks, modeling approaches, data requirements, and estimation techniques.

APPENDICES

Appendix A: Accuracy Assessment of LUCC Classification—Rule-based Classification Rationality Evaluation

Validating classified results from long-series of images is always a problem because simultaneous reference data is frequently not available. The rule-based rationality evaluation, suggested by (Liu & Zhou 2004), can be employed as an alternative accuracy assessment technique in certain cases, including this study. The advantage of the method is that it only employs a set of rules while no reference map is needed.

Given that the classified images cover six time periods (1977, 1984, 1993, 2000, 2004, and 2007), the maximum chance for land use change is five. If t denotes the number of potential changes over the six periods, then $0 \leq t \leq 5$. If t equals 0, it implies that the pixel under analysis did not change at all during the whole time under study; if t equals 5, the pixel under investigation changed classes in each period. Each pixel in each of the six periods was generalized into one of four different assessment results: "Consistent", "Fuzzy", "Uncertain" or "Misclassified." These four statuses denote that "the pixel is correctly classified," "the pixel is in a fuzzy state," "whether the pixel was fuzzy or it was misclassified, or it is actually a real change remains uncertain," and "the pixel is not correctly classified," respectively.

The images were classified into four classes: C_1="Farmland," C_2="Forestland," C_3="Other," and C_4="Built-up." If change was detected between two neighboring periods, it was denoted as $T(C_a, C_b)$. So, $T(C_2, C_4)$ describes a pixel that changed from forestland to built-up in images from two consecutive periods. As shown in Figure 2.4, six rules were employed to assess the rationality of each pixel change trajectory. For each pixel, the rules are examined in sequential order.

Figure 2.4 Rationality evaluation rules
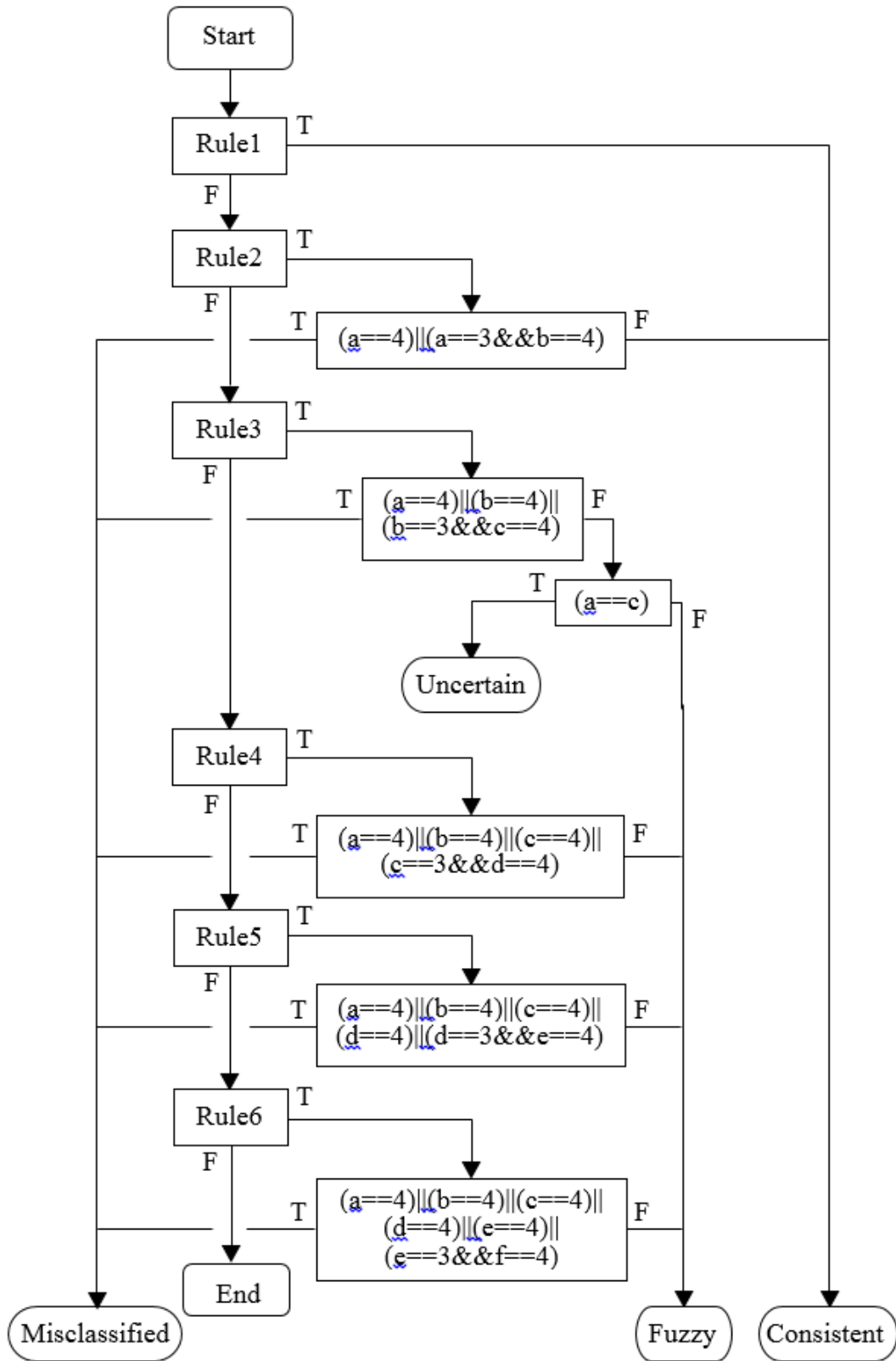
The six rules are defined and explained as follows:

Rule 1: If t=0, then accept "Consistent."

Rule 2: If t=1, i.e. $T(C_a, C_b)$, AND if (a==4)||(a==3&&b==4), THEN accept "Misclassified;"

otherwise, "Consistent."

Rule 3: If t=2, i.e. $T(C_a, C_b, C_c)$, AND if (a==4)||(b==4)||(b==3&&c==4), THEN accept

"Misclassified." Otherwise, check if (a==c). If so, "Uncertain;" otherwise, "Fuzzy."

Rule 4: If t=3, i.e. $T(C_a, C_b, C_c, C_d)$, AND if (a==4)||(b==4)||(b==3&&c==4), THEN accept

"Misclassified;" otherwise, "Fuzzy."

Rule 5: If t=4, i.e. $T(C_a, C_b, C_c, C_d, C_e)$, AND if (a==4)||(b==4)||(c==4)||(d==4)||(d==3&&e==4),

THEN accept "Misclassified;" otherwise, "Fuzzy."

Rule 6: If t=5, i.e. $T(C_a, C_b, C_c, C_d, C_e, C_f)$, AND if

(a==4)||(b==4)||(c==4)||(d==4)||(e==4)||(e==3&&f==4), THEN accept "Misclassified;"

otherwise, "Fuzzy."

There are two most important assumptions behind these six rules. First, the change to built-up from other land-use classes is irreversible, so that any pixel that is classified as built-up in a previous period and later placed into any other land use class would be regarded as a misclassification. Second, it is also uncommon to construct on wetland, therefore, conversions from wetland to built-up are all processed as misclassifications. These two underlying rules are generally applied to all cases during the six periods.

Rule 1 is quite straightforward; if a pixel is classified as the same land use class for all six periods, then the pixel is regarded as "consistent." Rule 2 concerns the situation when a once-only change is detected for a certain pixel. If the land conversion direction is true (T) with the two misclassification statements, then the change is labeled "misclassified." In other cases, I take it as a possible change, and regard it as correctly classified ("consistent"). Similar to Rule 2, Rule 3 first defines that if the reverse process (i.e. change from built-up area to another land use type) or the unlikely process (i.e. the change to built-up from other) were detected, the changes are taken as not correctly classified. This rule then deals with a one-time error of multi-temporal remote sensing image classification. If a pixel is found to have changed from one class (C_a) to another (C_b) and back to its original status (i.e. C_a), this situation could either be taken as a one-time classification error (i.e. C_b is the incorrect class), or it could be that the pixel itself is a fuzzy pixel, in which case the pixel could be classified as C_a or C_b. This one-time inconsistent situation does not affect the final result of cover detection, but it is hard to tell if it is a real classification error or not, so the pixel is regarded as "Uncertain". Finally, Rule 3 specifies the treatment of a case where the land use type changed twice to two different classes during the study period. In this case, I consider the pixel "Fuzzy" with a composite land use type.

Rules 4, 5 and 6 consider pixels that change frequently between cover types. This is most likely a consequence of mis-registration in geometric image rectification (Townshend et al. 1992, Stow 1999). Obviously, the reverse process and the unlikely process would be both improbable according to Rule 2, which indicates that the pixel may not be correctly classified. For other similar pixels, this can be considered as a "Fuzzy" case with frequent cover classes.

Since in this project, a county is the basic unit of observation and analysis, all the pixel-based results of LUCC detection are aggregated into the ten counties. The rationality evaluation results,

shown in Figure 2.5, are generally acceptable. "Consistent" classified pixels in each county are above 80%, the "Misclassified" rate is low, and the "Uncertain" and "Fuzzy" rates are both below 10%.

Figure 2.5 Rule based rationality evaluation results
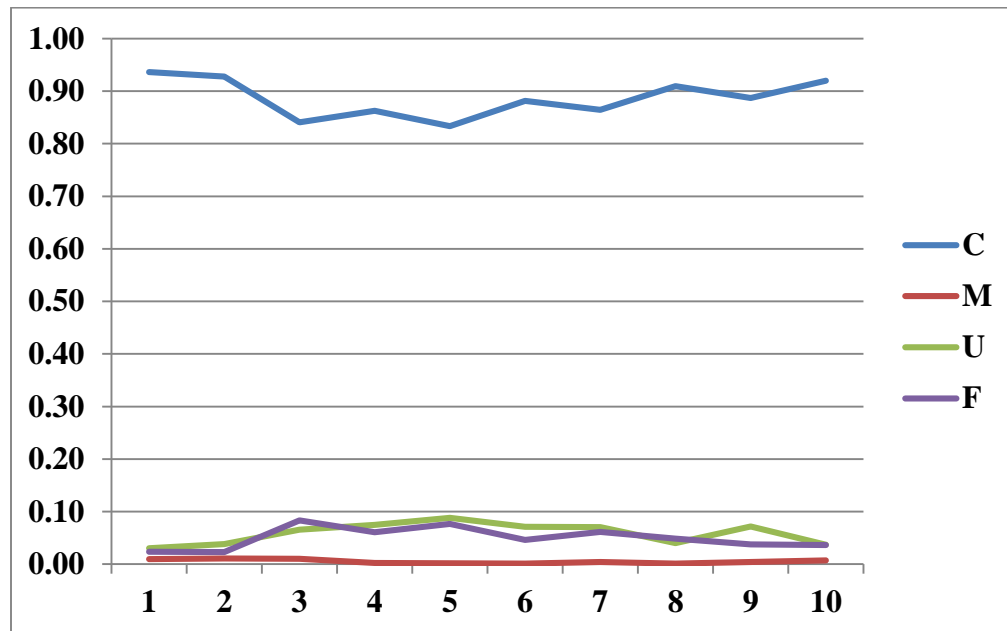
Note: C, M, U, F stand for "Consistent," "Misclassified," "Uncertain," and "Fuzzy," respectively. The ten counties are: 1, Fangzheng; 2, Yilan; 3, Huachuan; 4, Suibin; 5, Youyi; 6, Jixian; 7, Shuangyashan; 8, Huanan; 9, Qitaihe; and 10, Boli.

The "Uncertain" and "Fuzzy" classes are possibly the most active pixels where land conversion tends to take place. Since some of the once-only land use changes determined by in Rule 2 are also regarded as "Consistent," the potential LUCC change is larger than that reflected in the proportion of "Uncertain" and "Fuzzy." The rule-based rationality evaluation is beneficial especially in identifying the misclassification rate. This could be helpful for further classification correction. However, there are also some logical limits in this flow chart design. For example, it is hard to clearly differentiate once-only changes from fuzzy pixels, thus the "Uncertain" and "Fuzzy" rates are subject to dispute.

Appendix B: Accuracy Assessment of LUCC Classification—Traditional Accuracy Assessment Results

To validate the accuracy of my classified LUCC results under this method, I first adopted the simple equation used to estimate sample size in this context: $N = Z_{\alpha/2}^2 P(1-P)/CI^2$ (Foody 2009b). The overall accuracy P for each class of land use is usually assumed to be 80%. CI is the half width of the confidence interval; a value of 0.05 is often taken. And following conventional practice, $Z_{\alpha/2}$ is set at 1.96. The calculated results show that sample size for each category should be 246. Given that I have four landscape classes, about 1000 points needed to be drawn from the map of my study site.

To this end, I employed the spatially balanced sampling method (SBS), which draws sample points proportional to the presence of the area (Stevens Jr & Olsen 2004). I generated 1200 points in my study site and used images in Google Earth as the reference data for my classification results for 2000, 2004 and 2007, respectively. After the layer of randomly sampled points was created, I converted it into a KML file readable by Google Earth, and marked the categories of those points on Google Earth. Next, the extracted Google Earth map information was compared to the classification results (Boulos 2005; Du et al. 2009). So, I got two datasets for the same points, based on which Kappa indices and conversion matrixes can be derived. After I started counting whether the sampled points are correctly classified, I identified an error in ArcMap 10, which provided wrong numbers in the attributes table. This led me to estimate the density of sampling points incorrectly, with less than 40 points for the minor LUCC categories (built-up and other). To get a larger sample to alleviate this problem, I added another 400 sample points to the two minor categories. In the end, I reached a total sample size of 1550 points.

But for the land-use maps before 2000—covering 1977, 1984 and 1990, it is not feasible to directly take a reference map from Google Earth, because most images in Google Earth are post-2000. Because there was not any other kind of map available, it was extremely difficult to get a reliable reference for those earlier periods. In this case, I took the following two steps to address the problem. First, note that the four classes of land use are not easily re-convertible. For example, it is highly unlikely for forestland to be converted to farmland and then reconverted back to forestland. So, my first step was to select those consistent points from a land-use classification map from an earlier time period and the Google Earth data from 2004 in the whole sample and take those points as unchanged. My second step was to extract the inconsistent points and compare them with the original images. I realized that the geo-corrected and atmospheric adjusted images are the best available reference data. So, I manually recorded the classes of land use for those inconsistent points to distinguish points of real change from those misclassified.

Based on the above steps, the accuracy assessment results are summarized in Table 2.7. The overall accuracy rates for the six periods are around or above 85%. For 1977 to 1984, as the MSS data have coarser spatial resolution than TM and ETM+ images, I merged farmland and built-up land into one category, called F&B. The overall accuracy for 1977 and 1984 is 91.6 % and 90.5%, respectively, and the overall Kappa indexes are 86.1% and 84.2%, which are generally higher than for the rest of the images used in this study. The classifications of the maps for the other four periods include four LUCC categories: farmland, forestland, built-up land and other. The overall accuracy rates for these four periods are around 85%, and the Kappa indexes are about 80%. The accuracy of the 1993 and 2007 maps is a bit higher than that of the remaining two periods. The Kappa indexes for these two periods are around 80%, while that of 1993 is 82% and 2000 is about 77%. Due to the large sample size, the standard deviations and coefficients of variation for

both overall accuracy and kappa indexes are very small. As I use the high accuracy Google Earth maps as reference maps, it is inevitable that the classification map accuracy would be lower than expected.

I also calculated the classification accuracy for each land-use class and the results are reported in Tables 2.8 and 2.9. In both tables, the left block is the common confusion matrix (Foody 2002); the middle block contains the calculated indices of user's accuracy (UA); and the right block contains the indices of producer's accuracy (PA). There are no clear patterns of performance for the assessed user's and producer's accuracies. For example, in Table 2.8, the user's accuracy for F&B and other is generally higher than the producer's accuracy. But the producer's accuracy for forestland is higher than the user's accuracy. For a more thorough assessment of classification accuracy, the tables also included the Kappa index, which reflects the difference between the classification agreement and the agreement expected by chance (Stehman 1997). Some authors argue that this index tends to underestimate the accuracy (Rosenfield & Fitzpatrick-Lins 1986). The calculated values are generally lower than those from the user's and producer's accuracy statistics.

Table 2.8 Overall accuracy report of LUCC classification results

Year	OA%	Std(10^{-2})	CV%	Kappa%	Std(10^{-2})	CV%
1977	91.61	0.70	0.76	86.14	1.16	0.74
1984	90.52	0.74	0.82	84.17	1.24	0.68
1993	87.81	0.83	0.95	82.21	1.21	0.68
2000	84.24	0.93	1.10	77.15	1.35	0.57
2004	86.24	0.88	1.02	80.09	1.28	0.63
2007	89.08	0.79	0.89	84.44	1.13	0.75

Note: OA stands for overall accuracy, Std stands for standard deviation, and CV is short for coefficient of variation, which shows the extent of variability in relation to the overall accuracy.

Table 2.9 LUCC category-based accuracy report for 1977 and 1984

		F&B	Ft	Other	UA	Kappa	Std	PR	Kappa	Std
	F&B	705	16	18	0.95	0.91	0.02	0.89	0.78	0.02
	Ft	63	513	4	0.88	0.82	0.02	0.97	0.95	0.01
1977	Other	28	1	201	0.87	0.85	0.03	0.90	0.88	0.02
	F&B	741	12	29	0.95	0.89	0.02	0.88	0.76	0.02
	Ft	61	459	7	0.87	0.81	0.02	0.97	0.96	0.01
1984	Other	38	0	203	0.84	0.81	0.03	0.85	0.82	0.03

Note: F&B stands for farmland and built-up, Ft stands for forest, and Other mainly includes wetland, grassland and unused land. UA and PA are short for user's and producer's accuracy, respectively. Std stands for standard deviation. The number of observations in 1977 was 1549 while the number of observations in 1984 was 1550.

Table 2.10 LUCC category-based accuracy report for 1993, 2000, 2004 and 2007

		Fm	Ft	Other	Bltup	UA	Kappa	Std	PR	Kappa	Std
	Fm	585	15	65	19	0.86	0.75	0.02	0.89	0.80	0.02
	Ft	33	443	5	3	0.92	0.88	0.02	0.96	0.95	0.01
	Other	28	1	170	1	0.85	0.82	0.03	0.69	0.65	0.03
1993	Bltup	12	1	6	163	0.90	0.88	0.03	0.88	0.86	0.03
	Fm	559	38	36	12	0.87	0.76	0.02	0.81	0.67	0.02
	Ft	64	393	2	5	0.85	0.79	0.02	0.89	0.84	0.02
	Other	56	9	186	3	0.73	0.69	0.03	0.81	0.78	0.03
2000	Bltup	13	1	5	166	0.90	0.88	0.03	0.89	0.88	0.03
		Fm	Ft	Other	Bltup	UA	kappa	Std	PR	kappa	Std
	Fm	564	30	30	7	0.89	0.81	0.02	0.82	0.69	0.02
	Ft	63	406	2	7	0.85	0.79	0.02	0.92	0.89	0.02
	Other	50	4	195	2	0.78	0.74	0.03	0.85	0.82	0.03
2004	Bltup	15	1	2	170	0.90	0.89	0.02	0.91	0.90	0.02
	Fm	561	13	6	3	0.96	0.93	0.01	0.81	0.70	0.02
	Ft	43	422	3	0	0.90	0.86	0.02	0.96	0.94	0.01
	Other	71	4	216	5	0.73	0.68	0.03	0.95	0.94	0.02
2007	Bltup	17	2	2	180	0.90	0.88	0.02	0.96	0.95	0.02

Note: Fm stands forfarmland, Ft stands for forestland, Bltup is short for built-up and Other mainly includes wetland, grassland and unused land. UA and PA are short for user's and producer's accuracy, respectively. Std stands for standard deviation.

It can be seen from the above tables that the classification of farmland and forestland— the focal classes of land use—is reasonably good, despite some misclassifications between the two classes. The accuracy for built-up land is relatively low because it was hard to clearly distinguish

built-up areas from farmland in certain cases. While people can easily differentiate forestland and farmland using Google Earth, classification differences can happen in a 30-by-30-meter pixel given the possibility that an area of that size may include more than one use. Meanwhile, small positional deviations between Landsat images and images in Google Earth could also be a potential source for lower accuracy (Dai & Khorram 1998; Potere 2008).

Appendix C: Landscape Composition and Configuration Change

The composition and configuration of a landscape are fundamental aspects of landscape pattern, and studies of these patterns are useful for quantifying human impact. Development of quantitative indexes of spatial patterns (O'Neill 1988) enables the analysis and characterization of landscapes in terms of their patch composition, spatial relations, and dynamics. FRAGSTATS (McGarigal & Marks 1995; McGarigal 2012) is widely used for the description and analysis of landscape configuration. Various landscape metrics offer a wide range of measures of varying complexity and facilitate making comparisons across landscapes. Table 2.10 shows some of the most popular and frequently employed landscape metrics, which I employed to monitor landscape diversity and integrity.

Table 2.11 Landscape diversity and integrity change, 1977~2007

Year	MSIDI	MSIEI	LSI	CONTAG	PLADJ	AI
1977	0.85	0.61	76.32	61.19	97.49	97.52
1984	0.82	0.59	99.90	60.45	96.70	96.73
1993	0.81	0.58	118.02	59.17	96.10	96.12
2000	0.79	0.57	114.06	59.50	96.23	96.26
2004	0.77	0.56	128.16	59.59	95.76	95.78
2007	0.77	0.55	141.96	59.05	95.29	95.32

Note: The 8-neighbor rule was selected to capture the adjacency of neighboring land cover, under which the 8 pixels adjacent vertically, horizontally, and diagonally are included.

MSIDI and MSIEI quantify composition at the landscape level, which refers to the number and occurrence of different classes of land use. The most frequently employed measures of landscape composition include the Shannon and Simpson indexes. The Shannon index is sensitive to rare cover types and emphasizes landscape richness, whilst the Simpson index places more weight on the dominant cover types and the landscape evenness (McGarigal & Marks 1995;

Nagendra 2002). Because my focus is primarily on forestland and farmland, the Simpson Index family fits better.

The value of SDI is expressed as the probability that any two cells selected at random would be different patch types. Thus, the higher the value, the greater the likelihood that any two randomly drawn cells would be different patch types. The Modified Simpson Diversity Index is adapted from the SDI. It combines evaluations of richness and evenness. It increases when the number of land-cover types (landscape richness) increases, or the land distribution balance amongst the various cover types (landscape evenness) increases (Pielou 1975; Turner 1990). As the number of land-cover types in my study is fixed at four, the richness information can be excluded from the MSDI. So the change in MSIDI in Table 2.10 reflects the decreasing trend of landscape evenness.

The MSIEI is measured as the observed level of diversity divided by the maximum possible diversity for a given patch richness (Wickham & Rhtters 1995). It facilitates evaluating evenness by normalizing comparisons of landscapes differing in the number of cover types (Hunziker & Kienast 1999). MSIEI takes a value between 0 and 1, with 0 indicating the exclusivity of one land use category, and 1 signifying an equal abundance of all the land use categories. As shown in Table 2.10, MSIEI drops considerably from 0.6102 to 0.5533 over the 30-year period, indicating that the balance of distribution of land amongst the four cover types (landscape evenness) decreases.

In assessing the biological integrity of the landscape, it is of importance to measure landscape aggregation. To measure the land aggregation, I tried to incorporate metrics with different emphases, including LSI, CONTAG, PLADJ, and AI. LSI is a normalized perimeter-to-area ratio, which is equal to 0.25 (adjustment for raster format) times the sum of the entire landscape boundary and all edge segments (m) divided by the square root of the total landscape

area (McAlpine & Eyre 2002; McGarigal 2012). In contrast to total edge or edge density, LSI provides a standardized measure that adjusts for the size of the landscape (McGarigal 2012). Thus, by measuring the geometric complexity of the landscape, LSI is usually interpreted as a measure of landscape disaggregation: the greater the value of LSI, the more dispersed the patch types are. At the landscape level, LSI equals 1 when the landscape only consists of a single patch, and it increases as levels of internal edges increase and patch shape becomes more irregular. From Table 2.11, it can be seen that LSI increased during the study period. Compared to all the other indexes, the absolute change in LSI value is largest. From 1977 to 2007, LSI approximately doubled, indicating dramatically increased levels of internal edge and corresponding decreases in the aggregation of patch types in the study area. A limitation of LSI is that it assumes that a square is the most aggregated shape in a raster data format. However, if the set of patches comprises multiple circular patches of different sizes, LSI will never equal 1. Table 2.10 shows that the LSI value in 2000 is smaller than that in 1993, which does not match my expectation. As LSI includes two aspects—edges and patch shape—I would conclude this result indicates that patches in 2000 are more compact. It is also possible that image quality in 1993 (clarity, cloud situation, seasonal effects, etc.) is better, and that in the classification process I distinguished more small patches.

CONTAG implies that pixels having the same attribute class tend to be adjacent. The Contagion Index, as defined by O'Neill et al. (1988), has been widely used in landscape ecology because it is an effective summary of overall "clumpiness" on categorical maps (Turner 1989; Graham et al. 1991). CONTAG is defined as proportion of all adjacencies that are same-class adjacencies, and it incorporates two distinct components—patch type interspersion (i.e., the intermixing of units of different patch types) and patch dispersion (i.e., the spatial distribution of a patch type) at the landscape level (Li & Reynolds 1993). The CONTAG values in Table 2.10

show a decreasing trend during the study period; the higher value in 1977 indicates that the study area had large, contiguous patches then, and these patches became more interspersed and dispersed over the study period.

Though by design CONTAG values are converted to a proportion percentage, the relative amount of value change of is much smaller than that of LSI. Also, like LSI, CONTAG values still show a small reversal in 2000 compared to those of 1993. CONTAG has its own advantage, as it is affected by both the dispersion and interspersion of patch types, and it has a complex, nonlinear formulation and multiple input components (Li & Wu 2004). PLADJ, measuring the proportion of cell adjacencies involving the same class, computes the sum of the diagonal elements of the adjacency matrix divided by the total number of adjacencies (McGarigal 2012). Due to the design of the metric, PLADJ measures patch dispersion of land use classes—a landscape containing larger patches with simple shapes will have a higher PLADJ value. It can be seen in Table 2.10 that while the PLADJ values remain high, they did decrease during the study period. Compared to CONTAG, PLADJ measures only patch-type dispersion, not interspersion. Accordingly, the relative value of PLADJ is larger. Also, as PLADJ calculation relates to the proportion of the landscape focal class P (farmland in this study), and both farmland and forestland in the study area are contagiously distributed, the PLADJ value is very high in our case.

AI is the ratio of the observed number of like adjacencies to the maximum possible number of like adjacencies given the proportion (P) of the landscape comprised of each patch type (He et al. 2000; McGarigal 2012). Like PLADJ, AI adjusts for P in different ways. At the landscape level, it is computed as an area-weighted mean class aggregation index where each class is weighted by its proportional area in the landscape. In Table 2.10, the AI values are close to the values for of

PLADJ. Also, the magnitude of decrease is similar. As AI measures land-patch dispersion—the same as PLADJ—the information I obtained tend to be consistent.

REFERENCES

REFERENCES

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use and land cover classification system for use with remote sensor data. In: Geological Survey Professional Paper. USGS, Reston, VA

Boulos, M.N., 2005. Web GIS in practice III: creating a simple interactive map of England's strategic Health Authorities using Google Maps API, Google Earth KML, and MSN Virtual Earth Map Control. International Journal of Health Geographics 4, 22

Chavez, P.S., 1996. Image-based atmospheric corrections-revisited and improved. Photogrammetric engineering and remote sensing 62, 1025-1035

Chinese Academy of Sciences, 2008. China Remote Sensing Satellite Ground Station.

Dai, X., Khorram, S., 1998. The effects of image misregistration on the accuracy of remotely sensed change detection. Geoscience and Remote Sensing, IEEE Transactions on 36, 1566-1577

Deng, J., Wang, K., Deng, Y., Qi, G., 2008. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. International Journal of Remote Sensing 29, 4823-4838

Du, Y., Yu, C., Jie, L., 2009. A study of GIS development based on KML and Google Earth. In: INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on, pp. 1581-1585. IEEE

Edström, F., Nilsson, H., Stage, J., 2012. The Natural Forest Protection Program in China: A Contingent Valuation Study in Heilongjiang Province. Journal of Environmental Science and Engineering B 1, 426-432

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote sensing of environment 80, 185-201

Foody, G.M., 2009a. Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing 30, 5273-5291

Foody, G.M., 2009b. Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing 30, F5273-5291

Gao, J., Liu, Y., 2011. Climate warming and land use change in Heilongjiang Province, Northeast China. Applied Geography 31, 476-482

Gao, J., Liu, Y., 2012. De (re) forestation and climate warming in subarctic China. Applied Geography 32, 281-290

Graham, R., Hunsaker, C., O'neill, R., Jackson, B., 1991. Ecological risk assessment at the regional scale. Ecological applications, 196-206

He, H.S., DeZonia, B.E., Mladenoff, D.J., 2000. An aggregation index (AI) to quantify spatial patterns of landscapes. Landscape Ecology 15, 591-601

Hunziker, M., Kienast, F., 1999. Potential impacts of changing agricultural activities on scenic beauty–a prototypical technique for automated rapid assessment. Landscape Ecology 14, 161-176

Li, H., Reynolds, J.F., 1993. A new contagion index to quantify spatial patterns of landscapes. Landscape Ecology 8, 155-162

Li, H., Wu, J., 2004. Use and misuse of landscape indices. Landscape Ecology 19, 389-399

Liu, H., Zhang, S., Li, Z., Lu, X., Yang, Q., 2004. Impacts on Wetlands of Large-scale Land-use Changes by Agricultural Development: The Small Sanjiang Plain, China. AMBIO: A Journal of the Human Environment 33, 306-310

Liu, H., Zhou, Q., 2004. Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison. International Journal of Remote Sensing 25, 1037-1050

Liu, J., Diamond, J., 2005. China's environment in a globalizing world. Nature 435, 1179-1186

Liu, Y., Wang, D., Gao, J., Deng, W., 2005. Land Use/Cover Changes, the Environment and Water Resources in Northeast China. Environmental Management 36, 691-701

McAlpine, C.A., Eyre, T.J., 2002. Testing landscape metrics as indicators of habitat loss and fragmentation in continuous eucalypt forests (Queensland, Australia). Landscape Ecology 17, 711-728

McGarigal, K., Marks, B.J., 1995. Spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station

McGarigal, K., SA Cushman, and E Ene, 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts. Amherst

Nagendra, H., 2002. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. Applied Geography 22, 175-186

NFPP Management Center, 2011. Authoritative interpretations for the second phase policies of natural forest protection project

O'Neill, R.V., Krummel, J.R., Gardner, R.H., Sugihara, G., Jackson, B., DeAngelis, D.L., Milne, B.T., Turner, M.G, Zygmunt, B., Christensen, S.W., Dale, V.H. and Graham, R.L., 1988. Indices of landscape pattern. Landscape Ecology. Landscape Ecology 1, 153-162

Pielou, E.C., 1975. Ecological Diversity. Wiley-Interscience, New York.

Pontius Jr, R.G., Shusas, E., McEachern, M., 2004. Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems & Environment 101, 251-268

Potere, D., 2008. Horizontal positional accuracy of Google Earth's high-resolution imagery archive. Sensors 8, 7973-7981

Rosenfield, G.H., Fitzpatrick-Lins, K., 1986. A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric engineering and remote sensing 52, 223-227

Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A., 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote sensing of Environment 75, 230-244

Song, k., Wang, Z., Liu, Q., Lu, D., Yang, G., Zeng, L., Liu, D., Zhang, B., Du, J., 2009. Land use/land cover (LULC) characterizaitoin with MODIS time series data in the Amu River Basin. In: Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, pp. IV-310 - IV-313

Stanturf, J., Madsen, P., Lamb, D., 2012. A goal-oriented approach to forest landscape restoration. Springer Science & Business Media.

Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment 62, 77-89

Stevens Jr, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99, 262-278

Tang, J., Wang, L., Zhang, S., 2005. Investigating landscape pattern and its dynamics in Daqing, China. International Journal of Remote Sensing 26, 2259-2280

Turner, M.G., 1989. Landscape ecology: the effect of pattern on process. Annual review of ecology and systematics, 171-197

Turner, M.G., 1990. Spatial and temporal analysis of landscape patterns. Landscape Ecology 4, 21-30

U.S. Department of the Interior, 2009. U.S. Geological Survey.

Wang, X., Sun, L., Zhou, X., Wang, T., Li, S., Guo, Q., 2003. Dynamic of forest landscape in Heilongjiang Province for one century. Journal of Forestry Research 14, 39-45

Wang, Z., Liu, Z., Song, K., Zhang, B., Zhang, S., Liu, D., Ren, C., Yang, F., 2009. Land use changes in Northeast China driven by human activities and climatic variation. Chinese Geographical Science 19, 225-230

Wang, Z., Zhang, B., Zhang, S., Li, X., Liu, D., Song, K., Li, J., Li, F., Duan, H., 2006. Changes of land use and of ecosystem service values in Sanjiang Plain, Northeast China. Environmental Monitoring and Assessment 112, 69-91

Wickham, J., Rhtters, K., 1995. Sensitivity of landscape metrics to pixel size. International Journal of Remote Sensing 16, 3585-3594

Xu, J., Yin, R., Li, Z., Liu, C., 2006a. China's ecological rehabilitation: Unprecedented efforts, dramatic impacts, and requisite policies. Ecological Economics 57, 595-607

Xu, J., Yin, R., Li, Z., Liu, C., 2006b. China's ecological rehabilitation: the unprecedented efforts and dramatic impacts of reforestation and slope protection in western China. Ecological Economics 57, 595-607

Yamane, M., 2001. China's Recent Forest-Related Policies: Overview and Background. Policy Trend Report 1, 1-12

Yin, R., 1998. Forestry and the environment in China: the current situation and strategic choices. World Development 26, 2153-2167

Yin, R., Yin, G., 2009. China's Ecological Restoration Programs: Initiation, Implementation, and Challenges. In: An Integrated Assessment of China's Ecological Restoration Programs. Springer Netherlands, pp. 1-19.

Yin, R., Yin, G., 2010. China's primary programs of terrestrial ecosystem restoration: initiation, implementation, and challenges. Environmental management 45, 429-441

Zhang, B., Cui, H., Yu, L., He, Y., 2003. Land reclamation process in northeast China since 1900. Chinese Geographical Science 13, 119-123

CHAPTER 3

LITERATURE REVIEW OF LUCC DRIVING FORCE ANALYSIS: MODELING

APPROACHES, RESEARCH FINDINGS AND KNOWLEDGE GAPS

3.1 Modeling LUCC Driving Forces

Much progress has been made since the 1990s in capturing the effects of socioeconomic and biophysical factors on land use and land cover change (LUCC) (Kaimowitz & Angelsen 1998; Geist & Lambin 2001; Irwin & Geoghegan 2001b; Lambin et al. 2001a; Geist & Lambin 2002a; Verburg et al. 2002; Parker et al. 2003; Verburg et al. 2004a; Railsback et al. 2006; Turner et al. 2008a). There are many modeling methods that vary across disciplines and exhibit different strengths and weaknesses. As synthesized by Kaimowitz and Angelsen (1998), the three main LUCC modeling approaches are: analytical, regression, and simulation. The decision on which model to use depends largely on the specialty and preference of the researcher. A review by Irwin (2010), based on modeling works of land-use changes published from 2003 to 2010, suggests that 94 percent of the papers in geography were simulation models, either cellular automata or agent-based; more than 60 percent of the papers in economics journals were statistical; and 30 percent were analytical with a simulation extension. In this chapter, I will review these approaches and the empirical results generated from their applications, as part of preparation for my own modeling work. In addition, because my empirical work will be largely econometric by nature, I will discuss the different strategies and techniques employed in estimating single-equation and system of equations models.

3.1.1 Analytical Models

Analytical models, which study individual behavior at the micro-level, are deeply rooted in economic theories and presented in a rigorous framework. The most representative analytical models are the household or firm-level models, in which agents are assumed to allocate their inputs (e.g., land, labor, and capital) to maximize the expected utility by consuming goods—home-

produced or purchased—and leisure under labor, time, market, preference, and property constraints (Chomitz & Gray 1996; Angelsen 1999). Usually, standard mathematical techniques, such as Lagrange optimization (with equality constraints) and linear programming (with inequality constraints), are employed to solve the objective functions.

These types of models have sound theoretical underpinnings that allow researchers to focus on some key aspects of human behavior associated with land use changes. But as indicated by Taylor and Adelman (2003), a major limitation of the household- or firm-level models is that they "risk missing an array of indirect influences shaped by fundamental features of rural economics." It is true that these models take endogenous variables into consideration, but it is unlikely for them to cover all the endogenous variables involved in the behavioral process. Along with the model abstraction and simplification process, the underlying assumptions (household's preferences, market constraints/mechanisms, and property regimes) often carry strong implications, and, to some extent, the conclusions depend greatly on the initial assumptions and sometimes they even produce ambiguous results (Kaimowitz & Angelsen 1998; Parker et al. 2003) . At the same time, since most analytical models mimic human behavior and work at the micro-level, difficulties arise from scaling these models up (Verburg et al. 2004a; Verburg et al. 2004b). Consequently, inferences drawn from micro-level findings for aggregate level outcomes should be avoided.

3.1.2 Regression Models

Empirical studies of LUCC driving forces tend to employ discrete-choice models. These models prescribe that the share or quantity of a particular land use is determined by land rent, land and landowner characteristics, among other factors (Chomitz & Gray 1996; Turner et al. 1996; Nelson & Geoghegan 2002; Walker et al. 2002; Vojáček & Pecáková 2010). In general, scholars

use multinomial logit or probit models since the dependent variable is typically a discrete category of land use. These regression models using relatively reliable data and large samples tend to give modelers a higher degree of freedom. Thus, discrete-choice models have received a considerable amount of attention in published studies. But a key problem for using discrete-choice models in LUCC lies in the *i.i.d.* (independent and identically distributed) assumption—that is, the errors should be independent of each other and with the same probability distribution. In a regional study, spatial dependence, such as neighborhood effect, spatial spillover, and network effect, is common, and the unobserved factors related to different land uses might be similar. For example, certain land use is determined perhaps by similar soil types in neighboring locations; hence, the random component might have different covariance structures (Anselin 2002). By the same token, the dependent variables can possibly be related. For instance, farmland at one location tends to form clusters with other farmland in the neighborhood. In both cases, the assumption of independence of these variable is inappropriate (Anselin 2002; Vojáček & Pecáková 2010). Meanwhile, the discrete-choice models usually fail to account for the endogenous effects of certain variables. For example, in the household- or firm-level models, prices are exogenous, while at more aggregate levels they could become endogenous to land-use change. Further, other important variables, like road development and population change, should be considered more carefully (Irwin & Geoghegan 2001b). As such, using the discrete-choice models while not taking the spatial dependence and endogeneity issues into consideration could lead to inconsistent parameter estimates and inappropriate policy conclusions (Berry 1994; Staiger & Stock 1994; Fleming 2004; Wooldridge 2005).

Geographers were pioneers in estimating spatially explicit empirical models by incorporating remote sensing data on land use/cover change (Chomitz & Gray 1996; Mertens &

Lambin 1997; Anselin & Bera 1998). These models do a good job of analyzing the correlation between land use and location-specific variables, such as distance from markets, topography, soil quality, and forest fragmentation. One shortcoming of these models is that they fail to incorporate the human behavior underlying spatial pattern/process/outcome of land use change because the models are often specified without a theoretical underpinning (Irwin & Geoghegan 2001b). Another deficiency of these models is that the study units were often based on individual pixels or some aggregate particles. From the behavioral perspective, none of these units were the true decision-making units. Finally, while many models incorporate spatial interactions, the spatial correlation still remains poorly reflected in their specifications. So, the models cannot contribute much to understanding how or why these interactions occur (Anselin 2010).

3.1.3 Simulation Models

Simulation methods are rooted in natural sciences. Cellular models and agent-based models are the most frequently used simulation systems. Tobler (1979) was one of the first to use a cellular model (CM) to simulate geographical processes. CMs define the interaction between land use at a certain location, the conditions in the surrounding pixels, and the transition rules, with all cells updated simultaneously according to those rules (Hogeweg 1988; Clarke 1997; Alonso & Sole 2000). Because CMs provide a good representation of the spatial dynamics of land use, they have been useful for modeling the ecological aspects of LUCC. However, they face challenges when human decision-making is incorporated (White & Engelen 2000; Parker et al. 2003). Thus, CMs have recently become hybrids with agent-based models (ABM).

An ABM couples social and environmental models and focuses primarily on human actions. It consists of units of "agents" that interact both with one another and with their environment, and

can make decisions and change their actions as a result of this interaction (Ferber 1999). In studying LUCC, ABMs incorporate the influence of micro-level human decision-making on land uses so that the linkages between human behavior and biophysical processes occurring in the landscape and the possible future land use situations can be clearly represented (Matthews et al. 2007). Compared to the traditional analytical and empirical methods, ABMs are superior in handling spatial interactions, socioeconomic processes, and decision feedbacks under multiple spatial scales. Because of the advent of powerful and flexible ABMs, various agent-based simulation platforms such as Swarm, Repast, MASON, and NetLogo, have evolved over the past decade (Railsback et al. 2006). Criticism of ABMs has surfaced mostly from concerns about model verification and validation of the potential outcome, since the traditional hypothesis testing methodology is not possible in this context, and more appropriate strategies and standards are yet to be developed (White & Engelen 2000).

3.1.4 Structural Equation Modeling

Structural equation modeling (SEM) originated in the sciences of psychology and education. SEM is also referred to as "causal modeling," which takes a hypothesis-testing approach to analyzing the structural relationships of interested variables. This involves integrating a series of statistical tools such as simultaneous equation modeling, path analysis, and confirmatory factor analysis (Anderson & Gerbing 1988; MacCallum & Austin 2000; Ullman & Bentler 2001; Byrne 2010).

The SEM system allows a network of relationships between independent variables and dependent variables. It integrates advantages from psychometrics, econometrics, sociometrics, and multivariate statistics (Bound et al. 1995). Thus, it is always presented diagrammatically to give a

clear conceptualization of the issues under study. Both the independent and dependent variables can be either measured variables or factors (Anderson & Gerbing 1988). Measured variables also are called manifest variables, indicators, or observed variables. Factors are hypothetical constructs that cannot be directly measured, often referred to as latent variables, composite variables, constructs, or unobserved variables (Ullman & Bentler 2001). Latent variables usually have multiple indicators and are defined, in effect, by whatever their indicators have in common; thus, they provide multidimensional representations of ideas of interest and overcome the imperfect validity and reliability of a single measured variable (MacCallum & Austin 2000). In a SEM, the manner in which the measured variables are linked to the latent variables is called the "measurement model," and the hypothesized linkages between the latent variables are called the "structural model" (Ullman & Bentler 2001). As a whole, the measurement model combined with the structural model provides a comprehensive, confirmatory assessment of interdependences between constructs (Anderson & Gerbing 1988).

Existing literature describing the possible determinants of LUCC and their systematical structures provides the initial framework for SEM model development, so relevant variables are related in a theoretically sound way. In addition, both direct and indirect effects on LUCC can be incorporated into an SEM, resulting in more complex linkages between the different variables that are being hypothesized (Grace 2006; Byrne 2010). Further, because "SEM is a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions" (Pearl 2000), this empirical testing of causality will help advance our understanding of the complex LUCC relationships and simulate future LUCC scenarios.

Due to its confirmatory nature, SEM aims to assess and modify theoretical models, so the model-building task should be based on existing comprehensive studies. Before specifying the

structural equation model however, it is necessary to look into previous studies and understand the potential causes and effects they revealed. Because the interactions between land and humans are very complicated and vary from place to place, the structural equation model is not as popular as the three previously detailed modeling techniques in land use research.

In addition, due to stringent confirmatory demands, SEM is very sample-size demanding to achieve stable results. In psychology and education publications, it is quite common to see thousands of observations for the targeted analysis (Bound et al. 1995; Baltagi & Liu 2009). In land use analysis, it is easy for us to get thousands of pixels; but, as previously noted, these abundant pixels are not decision-making units. Collecting and matching relevant socioeconomic and biological data for large samples normally is not feasible for many researchers.

In summary, each of the modeling approaches discussed above has its own advantage(s) and encounters its own specific difficulties in explaining LUCC. The popularity of the analytical models lies in their clear logical and theoretical basis. But the simplification of model representations and their underlying assumptions limit their policy implications in the real world. The regression-based empirical models can handle relatively large numbers of independent variables, but these models commonly have limited ability to explain human behavior, and it is not easy to separate correlation from causality. The simulation models are versatile, but the model fitness and validation are greatly dependent on the modeler's knowledge and experience. SEM provides a framework for analyzing the interdependencies among a set of variables with multiple equations that are logically linked. However, it is a confirmatory model, not suitable for explorative studies. Overall, econometric and simulation models provide evidence confirming certain basic conclusions from the analytical models, but they have not contributed much so far to resolving issues where analytical models provide inconclusive results (Kaimowitz 1998).

3.2 Main Results of LUCC Driving Force Analysis

My dissertation will explore the causes of LUCC, with a focus on deforestation in northeast China. It is thus necessary to examine the variables employed and the interactions and feedbacks between them in the deforestation and LUCC literature to gain a comprehensive understanding of the driving forces affecting forest cover changes.

Deforestation is induced by various proximate causes, which in turn are mediated by certain underlying determinants. A large number of published studies have tried to explore the causes of deforestation (Nelson & Hellerstein 1997; Kaimowitz & Angelsen 1998; Angelsen & Kaimowitz 1999; Pfaff 1999b; Angelsen et al. 2001; Zhang 2001; Geist & Lambin 2002a). According to these studies, proximate causes of deforestation include wood extraction, transport costs, and agricultural expansion. Included in the underlying causes are population growth, input and output prices, wage rates, agricultural productivities, and off-farm employment. Overall, deforestation is a complex process stemming from the multifaceted interactions among many socioeconomic and biophysical factors. In the following section, I will synthesize the potential relationships of those variables relevant to my study region, rather than providing a general review of the causes of deforestation.

3.2.1 The Direct Causes of Deforestation

Studies completed by Geist and Lambin (2002) revealed: 102 out of 152 cases of deforestation related to wood extraction, 146 cases from agricultural expansion, and 110 cases due to transport extension and settlement/market expansion. As such, the authors came to the conclusion that agricultural expansions, wood extraction/logging, and infrastructure development are the three main direct causes for deforestation.

Wood Extraction/Logging

In certain times and/or phases of development, wood extraction does improve the level of social welfare. Swanson (1994) characterized deforestation as an economic activity that is "stock disinvestment" or "resource mining." Commonly, selective wood extraction alone would not necessarily lead to deforestation because it does not necessarily result in a dramatic loss of canopy cover (Rudel & Roper 1997; Mainardi 1998). However, the impact of wood extraction is likely to become more significant over time, and studies found that wood production and deforestation are positively correlated (Burgess 1993; Asner et al. 2005; Bekker & Ploeg 2005; Asner et al. 2006). A study of deforestation in the Amazon by Asner et al. (2005) showed that logging annually impacts a forest area of between 12,000 and 19,000 square kilometers. Subsequent analysis by Asner et al. (2006) revealed that 76% of selective logging resulted in high levels of forest canopy damage. The study predicted the logged forests would be cleared within four years.

Agricultural Expansion

Agriculture expansion has been cited as another major cause of deforestation (Chichilnisky 1994; Barbier 2004). A sizable number of analyses start with the hypothesis that forest loss is the result of competing land use between agriculture and forestry (Barbier & Burgess 1997; Angelsen et al. 1999; Walker et al. 2002). Competing land-use models occasionally measure the cost of farmland by figuring lost net revenue from timber production plus the evaluated environmental benefits if the forest stands remain (Hausman et al. 2007). When exploring the underlying determinants of land conversion to agriculture, studies tend to focus on the decisions of agricultural households. The classic general equilibrium model helps integrate linkages between the agricultural and forestry sectors. In such models, the equilibrium level of deforestation is frequently hypothesized to be determined by output and input prices and other factors affecting

the famers' incomes (Rudel & Horowitz 1993; Bawa & Dayanandan 1997; Angelsen et al. 1999; Van Soest et al. 2002; Hausman et al. 2007).

Infrastructural Development

Infrastructural development is another proximate cause that promotes the conversion of forest to other land uses. The von Thünen theory, which posits that the agricultural frontier will expand until the net profit or land rent becomes zero, is still widely used in empirical studies (Angelsen et al. 2001); Chomitz and Gray (1996). Integrating the spatial dimension into an economic model of land use in Belize, the study found that road access would expose the forest to various forms of degradation, and that market access and distance to roads are key determinants of the type of land use. Pfaff (1999b) developed a deforestation equation from an economic land-use model and tested a number of factors influencing forest clearing at the county level. The results suggest that factors affecting transportation costs, road density and distance to major markets are significant. Mertens et al. (2002) examined the relationship between roads and deforestation by further classifying the roads into main and secondary road networks, and concluded that the improved road network along with other factors has made the remote forests more likely to be converted into pasture. All this empirical evidence suggests that lower access costs fuel deforestation. But Angelsen and Kaimowitz (1999) present a caveat, pointing out that studies tend to overstate the causality between road construction and deforestation because, in reality, roads are commonly built on cleared land rather than forested land that needs to be cleared.

3.2.2 The Underlying Causes of Deforestation

Demographic Factors

Population growth is widely recognized as a trigger of LUCC (Cropper et al. 1997; Angelsen 1999; Carr et al. 2005). For instance, limited farmland per capita can lead farmers to clear forests. Studies in the Neo-Malthusian tradition often view population expansion as an underlying cause of deforestation (Sandler 1993; Vanclay 1993). But Mather and Needle (2000) pointed out that attempts to link deforestation with population growth usually neglect to take into account that children require years to be considered a factor. Mertens et al. (2000) considered a five-year lag in the influence of population on deforestation. Meanwhile, studies in the Neo-Boserupian tradition argued that increasing population could also induce technological and other changes without overexploiting the natural resources (Goldman 1993; Drechsel et al. 2001). The two cases in West Africa reported by Leach and Fairhead (2000) suggest that an increase in the number of people can even lead to the development of more forests in the forest-savanna transition area. Overall, higher population density is associated with more deforestation in most cases; while in certain context, population increase could correlate with forest land expansion.

Technological Change

Local farmers face a production constraint, or technology, that depicts the relationship between inputs and outputs. In the agricultural sector, technology takes various forms—some are embodied in inputs, such as improved plant seeds, and some are disembodied, like the use of new machines (Lambin et al. 2003). The employment of new technologies in agricultural production requires labor and/or capital investments; for instance, the use of fertilizers requires cash for purchasing them and labor expenditure for applying them. Technological progress can change the relative scarcities of inputs, exerting contradictory effects on productivity. Findings of the effects

of agricultural technology on forests are ambiguous, depending on the production constraints and the forms of the technological progress: On one hand, technological progress may increase the marginalized return for labor, making households willing to supply more labor, which may lead to a greater demand for clearing land. On the other hand, technology may also raise a household's income, resulting in more spending on goods and leisure activities, which may reduce the pressure placed on land-based production activities. So, the overall effect of agricultural technology on forests depends on which scenario dominates in the local area (Van Soest et al. 2002; Pacheco 2006; Varian 2009).

Market and Price

The case study by Geist and Lambin (2002b) revealed that the growing prices of cash crops constitute a robust driver for deforestation. Timber price increase would lead to more logging in the short run but possibly to more forestation in the long run (Vincent 1990). Meanwhile, low timber prices make profit-orientated farmers less motivated to institute logging and prone to more crop production. Barbier (1994) pointed out that policies designed to ''improve'' the overall economy in the presence of market failures, such as the lack of prices for converted forests, may result in incentives that worsen forest loss. According to Zhang (2001), from the late 1970s to the mid-1990s, the timber prices in China went up sharply due to scarcity, but the prices increase became subsided as timber imports and plantation forests grew.

Studies also confirm that agricultural conversion is positively related to agricultural output prices but negatively correlated with rural wage rates (Barbier & Burgess 1996; Lopez 1997). Rent-seeking behavior in the agricultural sector will lead to farming intensification as well as farmland expansion. According to the study by (Deininger & Minten 1999), biased price

policies also could increase resource consumption and become a motivation for agricultural expansion.

Economic Growth (GDP)

Poverty is one of the frequently used drivers of deforestation (Dradjad H. Wibowo 1999). Deininger and Minten (1999) pointed out that higher levels of poverty significantly contribute to increased deforestation, and poverty- or capital-driven deforestation is often seen in developing counties (Rudel & Roper 1997). The environmental Kuznets curve (EKC) postulates that during the early stage of economic development in a country with substantial natural forests, deforestation will get worsened. As per-capita income increases, though, deforestation will slow down along with the emergence of reforestation and even afforestation (Zhang 2001). Studies by Grainger (1995) and Mather et al. (1999) confirmed the existence of Kuznets-type trends in forestry. They also found out that forests expanded more in emerging market economies. Rudel and Roper (1997) concluded that "Rates of deforestation are high in impoverished places; they increase with an initial surge of economic growth, and they decline when additional wealth creates other economic opportunities."

Policies

Given that the social costs of deforestation are usually not taken into account under the market mechanism, government policy becomes an important tool for internalizing various social costs. Angelsen et al. (1999) argued that many policies, including adopting improved technologies that are good for agricultural development, frequently promote deforestation. A panel-data analysis for all Mexican states confirmed that the potential impact of agricultural policy reform on the expansion of agricultural area is the direct effect of changes in pricing on the incentives for frontier expansion and forest conversion by rural households (Barbier & Burgess 1996).

71

3.3 Data Structure and Strength

The availability of annual observations on socioeconomic conditions for each sample county is an advantage of my research. In order to optimize the utilization of my data and better understand the linkages between various social-ecological factors and forest dynamics, I will interpolate the LUCC information into annual observations to enable the attainment of a panel dataset that integrates the LUCC information with information for other variables. This type of panel data, or cross-sectional time series data, involve two dimensions—a cross-sectional dimension (county) denoted by subscript i, and a time dimension (year) denoted by subscript t (Beck 2001; Hsiao 2003; Frees 2004). As county i is observed in each year t, it is a balanced panel. In an unbalanced panel, there are missing data on some units in some years (Baltagi & Song 2006).

According to the relative magnitude of *N* and *T (i=1, 2,...N; t=1, 2,...T)*, a panel dataset can be called a macro panel, in which *N* is moderate (typically less than 100) and *T* is substantial (usually larger than 20), or a micro panel, in which *N* is large (hundreds or even thousands) and *T* is small (usually less than 10 and most commonly less than 5 (Judson & Owen 1999; Baltagi 2008). The two-dimensional panel data set generally has a large number of data points, so more detailed and sophisticated econometric questions can be addressed that may not be handled using conventional cross-sectional or time-series datasets. (Baltagi & Giles 1998; Hsiao 2003) illustrated several major advantages in panel data applications. The enlarged dataset can lead to more variability among the variables. Also, it allows us to make different transformations, and we can get more reliable estimates and test more sophisticated assumptions and hypotheses (Hsiao 2014). For instance, as typical in cross-sectional data, the unobserved individual-specific effects usually

leads to biased estimates, while under the panel data setting, the advantages of controlling the effects of individual heterogeneity or omitted (mis-measured or unobserved) variables are widely recognized. Also, it is often difficult to make inferences about the dynamics based on cross-sectional evidence, while panel datasets are better able to identify the before-and-after effects and even the effects of dynamic behavior. Another important advantage occurs in the case of a non-stationary time-series where the data no longer follow normal distribution and the least-squares estimators and the maximum likelihood estimators would be biased. But when observations of cross-sectional units are available, under the independently distributed assumption, the central limit theorem based on cross-sectional units points out that the limiting distributions of estimators remain asymptotically normal (Hsiao 2007).

3.4 Basic Econometric Methods Using Panel Data

Because my econometric estimation of the LUCC driving forces will be primarily using the two main approaches of regression analysis under panel data setting—fixed effects (FE) model and random effects (RE) model—it is worthwhile to review these approaches here as well. A clear illustration of these methods is necessary to understanding my empirical analysis later.

3.4.1 Fixed Effects Model

The fixed effects (FE) estimator is known as the within estimator because only variations within a unit over time are used in the regression. Sometimes, it is also called the least-squares dummy-variable (LSDV) estimator (Cameron & Trivedi 2009). Without loss of generality, the fixed effect model can be illustrated as the following:

$$Y_{it} = \alpha_i + \beta' X_{it} + \varepsilon_{it}, \qquad i = 1,...N, \quad t = 1,...T \tag{3.1}$$

where β' is a $1 \times K$ vector of constants and α_i is a 1×1 scalar constant representing the unobserved heterogeneity peculiar to the ith individual over time. The FE model treats α_i to be fixed, and allows possible correlation between individual unobserved effect α_i and any regressor of interest, so regressor x_{it} may be endogenous (with respect to α_i but not ε_{it}). The error term, ε_{it}, represents the effects of the omitted variables that are peculiar to both the individual units and time periods. It is assumed that ε_{it} is uncorrelated with ($x_{i1}, ..., x_{iT}$) and can be characterized by an independently identically distributed random variable with mean zero and variance σ_ε^2.

The idea of using the FE model to obtain a consistent estimator is to remove α_i from the estimated equation. After calculating the means of time-series observations separately for each cross-sectional unit, the FE model transforms the observed variables by subtracting out the corresponding time-series means, and then apply the least squares method to the transformed data. That is, the individual-demeaned y is regressed against individual demeaned x.

$$y_{it} - \overline{y}_i = \beta_0 + \beta_1(x_{it1} - \overline{x}_{i1}) + \beta_k(x_{itk} - \overline{x}_{ik}) + \mu_{it} - \overline{\mu}_i, \quad t = 1,2,...T \tag{3.2}$$

With such a transformation, variations between individuals are not used in the estimation, so we cannot obtain the coefficients of the regressors that are time-invariant.

In the panel data case, the individual unit is sampled more than once. Repeated observations for the same unit are often referred as a "group," or more officially a "cluster." Cluster analysis is very popular now, and various econometric studies have used clusters in their modeling procedure (Kaufman & Rousseeuw 2009; Anderberg 2014).

The cluster-specific FE model is an extension if the original fixed effect model (Cameron et al. 2011). It includes a separate intercept for each cluster, $y_{ig} = x_{ig}'\beta + \sum_{h=1}^{G} \alpha_g dh_{ig} + u_{ig}$ where

dh_{ig} is the hth of G dummy variables, equals one if the ig^{th} observation is in cluster h and zero otherwise (Wooldridge 2003). There are two main approaches to obtain the cluster-specific FE estimators: The least squares dummy variable employs OLS with regression of y_{ig} on x_{ig} together with G dummies, and the FE estimator also uses OLS but with the mean-difference model $(y_{ig} - \bar{y}_g) = (x_{ig} - \bar{x}_g)'\beta + (u_{ig} - \bar{u}_g)$. Mainstream empirical researchers tend to use the FE estimator as it controls for a certain form of endogeneity of regressors when the regressors are correlated with the cluster invariant component α_g, in which case, the traditional OLS and Feasible Generalized Least Square (FGLS) estimators would be inconsistent while the FE estimator eliminates α_g by the design and is consistent if either $G \to \infty$ or $N_g \to \infty$ (Cameron & Miller 2015).

The major attraction of an FE estimator is that it suites well for non-experiment research fields. It controls for unobserved and stable characteristics of the unit in the study, and it allows unobserved variables are correlated with observed variables (Hsiao 1985; Lau et al. 1998; Allison 2009). In a regression equation the unobserved effects can either be directly estimated or parceled out. Thus, it is a huge advantage when omitted variable bias is an issue. On the other hand, it has some crucial limitations that should not be ignored. First, if a researcher wants to estimate the individual effects, the dummy variable approach is costly in terms of degrees of freedom (Allison 2009). Second, as stated, a classic FE model will not produce any estimates of the effects of variables that don't change over time. Third, when the between variation is larger than the within variations in the predictor variables, the fixed effect estimates will be imprecise, leading to larger standard errors and wider confidence intervals (Hedges & Vevea 1998; Allison 2009). This is because in estimating an FE model, the differences between individuals are essentially discarded

during the process of subtracting the mean differences across the units of observation, leaving only the within-individual differences in the estimated equation.

3.4.2 Random Effects Model

An RE model can be written as

$$y_{it} = \mu + x_{it}{}'\beta + \alpha_i + \varepsilon_{it} = \mu + x_{it}{}'\beta + \mu_{it} \tag{3.3}$$

The error term μ_{it} contains two components, that is, $\mu_{it} = \alpha_i + \varepsilon_{it}$, where α_i is referred to as individual random effects. In the RE model, there are two fundamental assumptions. First, the unobserved individual effects α_i are random draws from a common population. Second, there is no correlation between the observed explanatory variables and the unobserved effect, or α_i is assumed to be uncorrelated with x_{it}. Thus, $cov(x_{itj}, \alpha_i) = 0$ with $t = 1,2,...T; j = 1,2,...,k$ (Laird & Ware 1982; Hedges & Vevea 1998).

The RE model is a weighted average of the within (or fixed effects) estimator (variation within units over time) and the between estimator (variation between units at the cross-sectional level) (Hedges & Vevea 1998; Wooldridge 2012). It can be estimated by Generalized Least Square (GLS), which id obtained using a least squares regression of

$$y_{it} - \hat{\theta}_i \bar{y}_i = \beta_0(1 - \hat{\theta}_i) + \beta_1(x_{it} - \hat{\theta}_i \bar{x}_i) + \{(1 - \hat{\theta}_i)\alpha_i + (1 - \hat{\theta}_i)\bar{\varepsilon}_i\}, \quad t = 1,2,...T \tag{3.4}$$

In the above equation, regressor x_{it} is exogenous. All the feasible GLS estimators are efficient asymptotically as N and T goes to infinity. The constant θ measures the weight given to the between-group variation, the equation for weight is as following:

$$\theta = 1 - \sqrt{\frac{\sigma_\mu^2}{\sigma_\mu^2 + T\sigma_\alpha^2}} \tag{3.5}$$

As the quantity under the square root sign approaches zero, θ is close to 1, then the model would become the fixed effect model. It is likely when the idiosyncratic variation σ_μ^2 is small relative to $T\sigma_\alpha^2$, that is, more of the variation is from fixed effect. Also, when the time span is long (T is large), there would be greater variation across time for each individual, or the FE σ_α^2 is big, θ approaches to 1, and the FE dominants. Vice versa, when σ_μ^2 is relative lager in magnitude, the pooled OLS suites (Laird & Ware 1982; Wooldridge 2012).

The RE estimator offers distinct advantages over the FE estimator in terms of efficiency because the former uses more of the variation in X (specifically, the cross sectional/between variation), which leads to smaller standard errors (Robinson 1991). Meanwhile, with random effects, we can estimate the effects of stable covariates such as race and gender. The most serious drawback of the RE approach is that it doesn't control for unmeasured, stable characteristics of the individuals (Semykina & Wooldridge 2010; Wooldridge 2012). Suppose that there is a variable z omitted from the model specification when predicting y in the RE model, any correlation between x and α_i can imply an omitted variable z that produces bias in estimates of β (Baltagi 2008).

3.4.3 Choice between FE and RE

When deciding whether to employ a FE or RE estimator, there are a number of practical and technical issues to be taken into account. First, an important misunderstanding of the frequently used terminology needs to be noted here. In FE models, the α_i term is treated as a set of fixed parameters which may either be estimated directly or conditionally on the estimation process. In RE models, however, the α_i term is treated as a random variable with a specified

probability distribution (usually normal, homoscedastic, and independent of all measured variables). So the term "FE model" is usually contrasted with "RE model."

Unfortunately, this terminology is the cause of much confusion. As suggested by (Mundlak 1978), the key issue involving α_i is whether or not it is uncorrelated with the observed explanatory variables x_{it}, for $t = 1, ..., T$. In a more advanced framework (Wooldridge 2002), the authors avoid referring to α_i as RE or FE. Instead, they suggest referring to α_i as unobserved effect, or unobserved heterogeneity; and what truly distinguishes the two approaches is the structure of the correlations between the observed variables and the unobserved variables. So, as pointed out by Mundlak (1978), the "FE" specification can be viewed as a case in which α_i is a random parameter with $\text{cov}(x_{itj}, \alpha_i) \neq 0$, whereas the RE model correspond to the situation in which $\text{cov}(x_{itj}, \alpha_i) = 0$

Theoretically, the decision to treat the between-unit variation as fixed or random is a trade-off choice between the problem of high variance and that of bias. As stated earlier, the FE model is making inferences conditional on the effects that are in the sample; it will produce unbiased estimates of β, but those estimates can be subject to high sample-to-sample variability (Hedges & Vevea 1998; Clark & Linzer 2012). The RE model makes unconditional or marginal inferences with respect to the population of all effects; so, it often introduces bias in the estimates of β, but it can greatly constrain the variance, leading to estimates that are closer (on average) to the true value.

Then, the decision about whether α_i should be treated as random variables or as parameters sometimes is dependent on the researcher—different researchers in different disciplines have different preferences. For example, economists tend to use fixed effect models because, in most cases, the data are not randomly drawn from experiments and they are more likely

to focus on estimating the effects of stable covariates, such as personal and family characteristics (Todd & Wolpin 2003). Similarly, the choice of different models also are predicated on answers to such questions as whether it's important to control for unmeasured characteristics of individuals and whether the loss of information from discarding the between-individual variation is acceptable (Clarke et al. 2010).

Another consideration relates to sample size. If the situation were one of analyzing a few numbers of units, say five or six, and the only interest lay in just these units, then α_i would more appropriately be fixed, not random. However, if the observed units are a sample from a larger population, and inferences will be made about the effects of a population, then the effects should be considered random. Also, as pointed out by Wooldridge (2003), with a large number of random draws from the cross-section, it almost always makes sense to treat the unobserved effects α_i as random draws from the population, along with y_{it} and x_{it}. However, random and FE models yield vastly different estimates, especially if T is small and N is large. While T is large, whether to treat the individual effects as fixed or random makes no differences. (Clark & Linzer 2012) summarized their advice for selecting the best approach based on the sample size. When both N and T are very small (say, N is smaller than 10 and T is smaller than 5), they suggest using the random effect model; when N is abundant while T is smaller than 5, the final decision lies in the value of $\text{cov}(x_{itj}, \alpha_i)$—choose random effect when the correlation is low and fixed effect otherwise. In the case that both N and T are large, they generally encourage using the fixed effect model; and if N fewer than 10 while T is large, the choice is correlation-dependent—large correlation leading to fixed effect while small correlation leading to random effect.

A common technique of choosing between FE and RE estimators is to employ the *Durbin–Wu–Hausman* tool, or *Hausman's* test (Hausman 1978), which is intended to tell the researcher

how significantly parameter estimates differ between the two approaches. The null hypothesis of the *Hausman's* test (1978) is that the unobserved heterogeneities are not correlated with the x_{it} ($\mathrm{cov}(x_{itj}, \alpha_i) = 0$) and the test is generally presented as a test of specification (fixed or random) of the unobserved effects. The basic rationale of this test is that the FE estimator is consistent whether the effects are or are not correlated with x_{it}. If the null hypothesis is true, the FE estimator is not efficient, because it relies only on the within variation in the data. On the other hand, when the effects are correlated with the x_{it}, the RE estimator is efficient under the null hypothesis but is biased and inconsistent (Baltagi & Giles 1998). So a statistically significant difference is interpreted as evidence against the random effect assumption. More specifically, if $\mathrm{cov}(x_{it}, \alpha_i) = 0$, both $\hat{\beta}_{re}$ and $\hat{\beta}_{fe}$ are consistent, but the RE model is more efficient than the FE model, or $se(\hat{\beta}_{re}) < se(\hat{\beta}_{fe})$. If $\mathrm{cov}(x_{it}, \alpha_i) \neq 0$, only $\hat{\beta}_{fe}$ is consistent, and with null hypothesis $\mathrm{cov}(x_{it}, \alpha_i) = 0$, $w = (\hat{\beta}_{fe}^* - \hat{\beta}_{re}^*)^2 / (Var(\hat{\beta}_{fe}) - Var(\hat{\beta}_{re}))$ is distributed with Chi-squared of x_1^2 (Wooldridge 2002). When the null hypothesis is true, the numerator of w would be small while the denominator would be large. If the null hypothesis is false, the difference between coefficients estimated by FE and RE is large, so the numerator would be large; because of the large numerator, w is large and we would choose the FE model. The above decision rules are summarized in Table 3.1.

Table 3.1 Model rules based on *Durbin–Wu–Hausman* test

	H₀ is true	**H₁ is true**
$\hat{\beta}_{re}$ (RE estimator)	Consistent and Efficient (choose RE)	Inconsistent
$\hat{\beta}_{fe}$ (FE estimator)	Consistent but Inefficient	Consistent (choose FE)

The *Hausman* test has been quite popular in helping to decide between the FE or RE models. However, it is not without problems since the null hypothesis of the *Hausman* test requires the random effect estimator to be efficient and thus requires the α_i and ε_{it} are *i.i.d* , which violates the assumption of cluster robust standard error for the random effect estimator. A simpler version of the test is

$$y_{it} - \theta \bar{y}_i = (x_{it} - \theta \bar{x}_i)\beta + (\chi_{it} - \bar{\chi}_i)\gamma + \upsilon_{it}, \quad t = 1,2,...T \tag{3.6}$$

This is simply the RE equation augmented with the additional variables. This equation consists of the time-demeaned original regressors. Here, y and x are defined as previously and χ includes the $1 \times M$ subset of time varying variables included in x (dummy variables are excluded). A test of $\gamma = 0$ can be implemented after the pooled OLS estimator. The *F* statistic is computed when $M > 1$. When the homoscedasticity assumption is violated, the robust version of test is needed (Wooldridge 2002, pp. 290-91). When heteroskedasticity as well as serial correlation are present, it is advisable to use cluster-robust standard errors (Baltagi & Giles 1998; Schmidheiny & Basel 2011).

In STATA, the model estimation procedure can be implemented manually. One could also take advantage of the user-written module "*xtoverid*," which is used to test over-identification restrictions after *xtreg, xtivreg, xtivreg2* or *xthtaylor*. STATA will report this test after standard panel data estimation with *xtreg, re*. The rationale of using an over-identification restrictions test to decide the FE or RE estimator is that the additional orthogonality conditions the RE estimator uses, i.e., $\text{cov}(x_{it}, \alpha_i) = 0$, are used to compare to the FE assumption. Unlike the *Hausman* test, the test executed by *xtoverid* guarantees to generate a nonnegative test statistic. Further, it extends straightforwardly to heteroskedastic- and cluster-robust test versions.

3.5 Summary

Selecting the appropriate framework, model, and estimation technique is crucial for adequately elucidating the driving forces of LUCC in general and deforestation in particular. Therefore, this chapter has reviewed the modeling approaches used and empirical results generated by previous studies of the driving forces of deforestation, and discussed the advantages and limitations of various models as well as the FE and RE estimation strategies associated with single-equation models. It has also articulated why we need and how we build more advanced modeling systems. These steps have prepared me well for my own modeling work in the next two chapters.

While there can be analytic, simulation, and regression models, I have decided to develop regression models, including single-equation and system of equations ones, in my empirical research. The variables to be considered in these models will be consistent with the proximate and underlying drivers of deforestation identified in the literature and the specific context of my study site. Because FE and RE estimations have their own advantages and shortcomings in estimating single-equation models, it is necessary for me to explore the ramifications of these different strategies and obtain more robust empirical findings, based on the data that I have collected. This will be the task of Chapter 4. Moreover, because single equations cannot deal with the endogeneity of certain right-hand side variable(s) or capture the interaction between different classes of land use, it is essential to develop and estimate better models to represent the complex linkages between the LUCC dynamics and their drivers. This will be the task of Chapter 5, where the Instrumental Variable Method and a system of equations will be used.

I know that completing these empirical tasks will require a skillful and careful application of economic principles and econometric tools. I am confident that I can complete get them done

successfully. Certainly, I hope that my work will contribute to an improved understanding of both the multi-faceted linkages between deforestation and its determinants in northeast China and the economic and policy implications to more effective forest protection and management as well as land use there and elsewhere.

REFERENCES

REFERENCES

Allison, P.D., 2009. Fixed effects regression models. SAGE publications, Thousand Oaks.

Alonso, D., Sole, R.V., 2000. The DivGame simulator: a stochastic cellular automata model of rainforest dynamics. Ecological Modelling 133, 131-141

Anderberg, M.R., 2014. Cluster Analysis for Applications: Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic press.

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: A review and recommended two-step approach. Psychological bulletin 103, 411-423

Angelsen, A., 1999. Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights. Journal of Development Economics 58, 185-218

Angelsen, A., Kaimowitz, D., 1999. Rethinking the Causes of Deforestation: Lessons from Economic Models. The World Bank Research Observer 14, 73-98

Angelsen, A., Shitindi, E.F.K., Aarrestad, J., 1999. Why do farmers expand their land into forests? Theories and evidence from Tanzania. Environment and Development Economics 4, 313-331

Angelsen, A., van Soest, D., Kaimowitz, D., Bulte, E., 2001. Technological change and deforestation: A theoretical overview. Agricultural technologies and tropical deforestation, 19-34

Anselin, L., 2002. Under the hood issues in the specification and interpretation of spatial regression models. Agricultural Economics 27, 247-267

Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89, 3-25

Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. Statistics Textbooks and Monographs 155, 237-290

Asner, G.P., Broadbent, E.N., Oliveira, P.J., Keller, M., Knapp, D.E., Silva, J.N., 2006. Condition and fate of logged forests in the Brazilian Amazon. Proceedings of the National Academy of Sciences 103, 12947-12950

Asner, G.P., Knapp, D.E., Broadbent, E.N., Oliveira, P.J., Keller, M., Silva, J.N., 2005. Selective logging in the Brazilian Amazon. Science 310, 480-482

Baltagi, B., 2008. Econometric analysis of panel data. John Wiley & Sons.

Baltagi, B.H., Giles, M.D., 1998. Panel data methods. Statistics Textbooks and Monographs 155, 291-324

Baltagi, B.H., Liu, L., 2009. A note on the application of EC2SLS and EC3SLS estimators in panel data models. Statistics & Probability Letters 79, 2189-2192

Baltagi, B.H., Song, S.H., 2006. Unbalanced panel data: a survey. Statistical Papers 47, 493-523

Barbier, E., 1994. The economics of the tropical timber trade. CRC Press.

Barbier, E.B., 2004. Agricultural Expansion, Resource Booms and Growth in Latin America: Implications for Long-run Economic Development. World Development 32, 137-157

Barbier, E.B., Burgess, J.C., 1996. Economic analysis of deforestation in Mexico 31. Environment and Development Economics 1, 203-239

Barbier, E.B., Burgess, J.C., 1997. The economics of tropical forest land use options. Land Economics 73, 174-195

Bawa, K.S., Dayanandan, S., 1997. Socioeconomic factors and tropical deforestation. Nature (London) 386, 562-563

Beck, N., 2001. Time-series-cross-section data: What have we learned in the past few years? Annual review of political science 4, 271-293

Bekker, P.A., Ploeg, J., 2005. Instrumental variable estimation based on grouped data. Statistica Neerlandica 59, 239-267

Berry, S.T., 1994. Estimating discrete-choice models of product differentiation. The RAND Journal of Economics 25, 242-262

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American statistical association 90, 443-450

Burgess, J.C., 1993. Timber production, timber trade and tropical deforestation. Ambio 22, 136-143

Byrne, B.M., 2010. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Psychology Press.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. Journal of Business & Economic Statistics 29, 238-249

Cameron, A.C., Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. Journal of Human Resources 50, 317-372

Cameron, A.C., Trivedi, P.K., 2009. Microeconometrics using stata. Stata Press College Station, TX.

Carr, D., Suter, L., Barbieri, A., 2005. Population Dynamics and Tropical Deforestation: State of the Debate and Conceptual Challenges. Population & Environment 27, 89-113

Chichilnisky, G., 1994. North-south trade and the global environment. American Economic Review 84, 851-874

Chomitz, K.M., Gray, D.A., 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. The World Bank Economic Review 10, 487-512

Clark, T.S., Linzer, D.A., 2012. Should I use fixed or random effects. Unpublished paper

Clarke, K., 1997. A self-modifying cellular automaton model of historical. Environment and planning B: planning and design 24, 247-261

Clarke, P., Crawford, C., Steele, F., Vignoles, A.F., 2010. The choice between fixed and random effects models: some considerations for educational research. Social Science Research Network

Cropper, M., Griffiths, C., Mani, M., 1997. Roads, population pressures, and deforestation in Thailand, 1976-89. World Bank Policy Research Working Paper

Deininger, K.W., Minten, B., 1999. Poverty, policies, and deforestation: the case of Mexico. Economic Development and Cultural Change 47, 313-344

Dradjad H. Wibowo, R.N.B., 1999. Deforestation mechanisms: a survey. International Journal of Social Economics 26, 455 - 474

Drechsel, P., Kunze, D., De Vries, F.P., 2001. Soil nutrient depletion and population growth in sub-Saharan Africa: a Malthusian nexus? Population and Environment 22, 411-423

Ferber, J., 1999. Multi-agent systems: an introduction to distributed artificial intelligence. Addison-Wesley Reading.

Fleming, M.M., 2004. Techniques for estimating spatially dependent discrete choice models. In: Advances in spatial econometrics. Springer, pp. 145-168.

Frees, E.W., 2004. Longitudinal and panel data: analysis and applications in the social sciences. Cambridge University Press.

Geist, H.J., Lambin, E.F., 2001. What drives tropical deforestation? A meta-analysis of proximate and underlying causes of defores-tation based on subnational scale case study evidence. In: LUCC Report Series No. 4., University of Louvain, Louvain-la-Neuve

Geist, H.J., Lambin, E.F., 2002a. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. BioScience 52, 143-150

Geist, H.J., Lambin, E.F., 2002b. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. BioScience 52, 143-150

Goldman, A., 1993. Agricultural Innovation in Three Areas of Kenya: Neo-Boserupian Theories and Regional Characterization. Economic Geography 69, 44-71

Grace, J.B., 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge.

Grainger, A., 1995. The Forest Transition: An Alternative Approach. Area 27, 242-251

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica: Journal of the Econometric Society 46, 1251-1271

Hausman, J.A., Newey, W.K., Woutersen, T.M., 2007. IV Estimation with Heteroskedasticity and Many Instruments. Centre for microdata methods and practice

Hedges, L.V., Vevea, J.L., 1998. Fixed-and random-effects models in meta-analysis. Psychological methods 3, 486-504

Hogeweg, P., 1988. Cellular automata as a paradigm for ecological modeling. Applied mathematics and computation 27, 81-100

Hsiao, C., 1985. Benefits and limitations of panel data. Econometric Reviews 4, 121-174

Hsiao, C., 2003. Analysis of panel data. Cambridge university press.

Hsiao, C., 2007. Panel data analysis—advantages and challenges. Test 16, 1-22

Hsiao, C., 2014. Analysis of panel data. Cambridge university press, Cambridge.

Irwin, E.G., 2010. New directions for urban economic models of land use change: incorporating spatial dynamics and heterogeneity. Journal of Regional Science 50, 65-91

Irwin, E.G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment 85, 7-24

Judson, R.A., Owen, A.L., 1999. Estimating dynamic panel data models: a guide for macroeconomists. Economics letters 65, 9-15

Kaimowitz, D., Angelsen, A., 1998. Economic models of tropical deforestation: a review. Centre for International Forestry Research, Jakarta.

Kaimowitz, D., Angelsen, A, 1998. Economic Models of Tropical Deforestation. A Review. Centre for International Forestry Research, Jakarta.

Kaufman, L., Rousseeuw, P.J., 2009. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, Hoboken.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963-974

Lambin, E.F., Geist, H.J., Lepers, E., 2003. Dynamics of land-use and land-cover change in tropical regions. Annual review of environment and resources 28, 205-241

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., 2001. The causes of land-use and land-cover change: moving beyond the myths. Global environmental change 11, 261-269

Lau, J., Ioannidis, J.P., Schmid, C.H., 1998. Summing up evidence: one answer is not always enough. The lancet 351, 123-127

Leach, M., Fairhead, J., 2000. Challenging Neo-Malthusian Deforestation Analyses in West Africa's Dynamic Forest Landscapes. Population and Development Review 26, 17-43

Lopez, R., 1997. Environmental externalities in traditional agriculture and the impact of trade liberalization: the case of Ghana. Journal of Development Economics 53, 17-39

MacCallum, R.C., Austin, J.T., 2000. Applications of structural equation modeling in psychological research. Annual review of psychology 51, 201-226

Mainardi, S., 1998. An economitric analysis of factors affecting tropical and subtropical deforestation. Agrekon 37, 23-65

Mather, A.S., Needle, C.L., 2000. The relationships of population and forest trends. Geographical Journal 166, 2-13

Mather, A.S., Needle, C.L., Fairbairn, J., 1999. Environmental Kuznets Curves and Forest Trends. Geography 84, 55-65

Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G., Gotts, N.M., 2007. Agent-based land-use models: a review of applications. Landscape Ecology 22, 1447-1459

Mertens, B., Lambin, E.F., 1997. Spatial modelling of deforestation in southern Cameroon: Spatial disaggregation of diverse deforestation processes. Applied Geography 17, 143-162

Mertens, B., Poccard-Chapuis, R., Piketty, M.G., Lacques, A.E., Venturieri, A., 2002. Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: the case of São Félix do Xingú in South Pará. Agricultural Economics 27, 269-294

Mertens, B., Sunderlin, W.D., Ndoye, O., Lambin, E.F., 2000. Impact of macroeconomic change on deforestation in South Cameroon: Integration of household survey and remotely-sensed data. World Development 28, 983-999

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica: Journal of the Econometric Society 46, 69-85

Nelson, G.C., Geoghegan, J., 2002. Deforestation and land use change: sparse data environments. Agricultural Economics 27, 201-216

Nelson, G.C., Hellerstein, D., 1997. Do roads cause deforestation? Using satellite images in econometric analysis of land use. American Journal of Agricultural Economics 79, 80-88

Pacheco, P., 2006. Agricultural expansion and deforestation in lowland Bolivia: the import substitution versus the structural adjustment model. Land Use Policy 23, 205-225

Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J., Deadman, P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. Annals of the Association of American Geographers 93, 314-337

Pearl, J., 2000. Causality: models, reasoning and inference. Cambridge University Press, Cambridge.

Pfaff, A.S., 1999. What drives deforestation in the Brazilian Amazon?: evidence from satellite and socioeconomic data. Journal of Environmental Economics and Management 37, 26-43

Railsback, S.F., Lytinen, S.L., Jackson, S.K., 2006. Agent-based simulation platforms: Review and development recommendations. Simulation 82, 609-623

Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. Statistical science, 15-32

Rudel, T., Roper, J., 1997. The paths to rain forest destruction: Crossnational patterns of tropical deforestation, 1975–1990. World Development 25, 53-65

Rudel, T.K., Horowitz, B., 1993. Tropical deforestation: Small farmers and land clearing in the Ecuadorian Amazon. Columbia University Press.

Sandler, T., 1993. Tropical Deforestation: Markets and Market Failures. Land Economics 69, 225-233

Schmidheiny, K., Basel, U., 2011. Panel Data: Fixed and Random Effects. URL http://www.schmidheiny.name/teaching/panel2up.pdf

Semykina, A., Wooldridge, J.M., 2010. Estimating panel data models in the presence of endogeneity and selection. Journal of Econometrics 157, 375-380

Staiger, D.O., Stock, J.H., 1994. Instrumental variables regression with weak instruments. Econometrica 65, 557-586

Taylor, J.E., Adelman, I., 2003. Agricultural household models: Genesis, evolution, and extensions. Review of Economics of the Household 1, 33-58

Tobler, W., 1979. Cellular geography. In: Philosophy in geography. Springer, pp. 379-386.

Todd, P.E., Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement. The Economic Journal 113, F3-F33

Turner, B.L., Lambin, E.F., Reenberg, A., 2008. Land Change Science Special Feature: The emergence of land change science for global environmental change and sustainability. Proceedings of the National Academy of Sciences of the United States of America 105, 2751-2751

Turner, M.G., Wear, D.N., Flamm, R.O., 1996. Land ownership and land-cover change in the southern Appalachian highlands and the Olympic peninsula. Ecological applications 6, 1150-1172

Ullman, J.B., Bentler, P.M., 2001. Structural equation modeling. John Wiley & Sons, Hoboken.

Van Soest, Daan P., Bulte, Erwin H., Angelsen, A., Van Kooten, G.C., 2002. Technological change and tropical deforestation: a perspective at the household level. Environment and Development Economics 7, 269-280

Vanclay, J.K., 1993. Saving the tropical forest : needs and prognosis. Ambio 22, 225-231

Varian, H.R., 2009. Intermediate Microeconomics: A Modern Approach. W. W. Norton & Company, New York City.

Verburg, P., Schot, P., Dijst, M., Veldkamp, A., 2004a. Land use change modelling: current practice and research priorities. GeoJournal 61, 309-324

Verburg, P.H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004b. Land use change modelling: current practice and research priorities. GeoJournal 61, 309-324

Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., Mastura, S.S., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. Environmental management 30, 391-405

Vincent, J.R., 1990. Don't boycott tropical timber. Journal of Forestry 88, 56

Vojáček, O., Pecáková, I., 2010. Comparison of Discrete Choice Models for Economic Environmental Research. Prague Economic Papers 19, 35-53

Walker, R., Perz, S., Caldas, M., Silva, L.G.T., 2002. Land use and land cover change in forest frontiers: The role of household life cycles. International Regional Science Review 25, 169-199

White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban Systems 24, 383-400

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge.

Wooldridge, J.M., 2003. Cluster-sample methods in applied econometrics. American Economic Review 93, 133-138

Wooldridge, J.M., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of applied econometrics 20, 39-54

Wooldridge, J.M., 2012. Introductory econometrics: A modern approach. Cengage Learning, Boston.

Zhang, Y., 2001. Deforestation and forest transition: theory and evidence in China. In: Palo M & Vanhanen H (eds.) World forests from deforestation to transition? Springer, Netherlands, pp. 41-65.

CHAPTER 4

AN ANALYSIS OF THE FORCES DRIVING FOREST COVER CHANGE

4.1 Introduction

As stated in Chapter 1, the vast forests in Heilongjiang have a paramount status in China. The province has more natural forests than any other one in China, and it is also home to much of the country's timber industry (Jiang et al., 2011). Meanwhile, due to its high quality black soil, Heilongjiang has been long recognized as the "Northeastern Granary." Its forests are regarded as playing an important role in stabilizing the local ecological system and helping secure the country's grain supply by protecting soils, preserving water supply, sheltering farmland, and moderating strong winds (Wang et al., 2006). Moreover, the Natural Forest Protection Program (NFPP), initiated in 2000, signified a major shift from traditional forest utilization to a new era of forest conservation (Xu et al., 2006; Yin & Yin, 2010). For all of these reasons, it is essential for me to focus on the determinants of Heilongjiang's forestland change.

Also, as depicted in Chapter 2, forestland and farmland are the two dominant classes of land use in the study region. In combination, they occupy around 80% of the total land area; and the predominant type of land transition has been the conversion of forestland to farmland. Therefore, the relationship between forestland and farmland requires close scrutiny. In this chapter, I will derive a theoretically consistent empirical model for analyzing the driving forces of forest cover change in Heilongjiang. Chapter 3 has reviewed a variety of approaches to deforestation analysis, some of which are theoretically motivated while others are empirical investigations. The economic and human behavior-based analytical models illuminate the causal relations in farmers' land allocation decision, such as how farmers react to price change and technology development under different market and/or land constraints. Understanding the findings based on this theoretical reasoning and other empirical studies as well as the intrinsic relationships between different indicators/variables will lay a solid foundation for me to specify my own empirical models.

Meanwhile, I will also emphasize the application of different estimating methods. Different regression tools could produce different empirical results, and variation in empirical results could be largely dependent on modeling specifications (Hegre & Sambanis, 2006). To validate the robustness of my results, a number of well-established and commonly used methods will be used with the same dataset. Comprehensive, though not exhaustive, exploration of the performance of different estimators can help me avoid poor empirical results and thus enhance their robustness.

To both ends, my strategy is to begin with simple regressions by specifying only the primary driving forces of deforestation in the empirical model, namely, the proximate factors in land use conversions—farmland expansion and wetland loss. As a second step, I will move on to augmented specifications where I will capture the effects of additional factors of deforestation identified in the literature review, such as socioeconomic development, political transformation, and demographic change.

The first part of this chapter is based on the land conversion data I have derived in Chapter 2. I will use all 48 observations (eight counties in six periods) in this analysis. To better organize the material of this chapter and present the analytic results, I will summarize the key findings in sub-sections 4.1.1 and 4.1.2, with the detailed modeling steps and between-model comparisons being covered in the Appendix (sub-section 4.4.1). The second section of this chapter begins with a discussion of the selected variables. Regression results are then presented in sub-sections 4.2.2 and 4.2.3, where I employ the most frequently used Fixed Effects (FE) and Random Effects (RE) estimators in the single-equation model with panel data. Here, Land Use and Land Cover Change (LUCC) data from the six periods (1977, 1984, 1993, 2000, 2004, and 2007) are linearly interpolated to derive annual observations, so that these land-use data can be more effectively integrated with social economic data in the driving force analysis. With a time span of from 1977

to 2007 for 8 counties: Suibin, Boli, Yilan, Fangzheng, Huanan, Huachuan, Qitaihe, and Jixian (Youyi and the municipality of Shuangyashan were dropped due to limited forest cover in their jurisdictions). The analysis in the second section will thus be based on 248 observations. By taking advantage of these long time-series, the analysis in sub-section 4.2.4 is intended to complement the earlier regressions. Finally, the implications of my modeling of the deforestation driving forces are discussed in section 4.3.

4.1.1 Initial Analysis Based on Land Use Categories: Fixed-Effects Estimation

For an initial analysis of the main driving forces of forestland changes, I have decided to include both farmland expansion and wetland loss in my models. As shown in Chapter 2, forestland and wetland (wetland and grassland constitute the most part of the "Other" class of my LUCC classification) are the two main sources for farmland expansion. Thus, the two types of land use can be mutually substitutable. Including the "Other" category in the regressions will help identify the underlying relationships in the LUCC dynamics. The general form of the regression models is:

$$Ft_{it} = f(Fm_{it}, Wt_{it}) + \alpha_i + \varepsilon_{it} \qquad \text{(Eq. 4.1)}$$

In Eq.4.1, *i* denotes observation units (counties), and *t* indexes time (year). The variables are the total areas (km^2) of different land uses, respectively; α_i is the fixed county effect, and ε_{it} is the random error. Table 4.1 reports the FE estimates of the driving forces of the forest cover changes based on six alternative modeling schemes. Mathematically, all the models in Table 4.1 are equivalent to the within-groups method and therefore estimated results are very similar.

Table 4.1 Initial results of the drivers of forestland change with unobserved heterogeneities being assumed as fixed

Forestland	I reg_lsdv	II Xtreg	III xtivreg2	IV xtreg_clbs	V areg_clbs	VI Fese
Farm	-1.14***	-1.14***	-1.14***	-1.14***	-1.14***	-1.14***
	(0.04)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)
Others	-0.82***	-0.82***	-0.82***	-0.82***	-0.82***	-0.82***
	(0.14)	(0.11)	(0.13)	(0.37)	(0.37)	(0.11)
Fangzheng	-160,842***					
	(5,723)					
Huachuan	-176,619***					
	(2,536)					
Huanan	910.0					
	(983.9)					
Jixian	-209,300***					
	(766.8)					
Qitaihe	-386,032***					
	(6620)					
Suibin	-107,539***					
	(7,165)					
Yilan	38,755***					
	(8,673)					
Constant	460,474***	335,390***		335,390***	335,390***	335,390***
	(2,182)	(8,450)		(47,850)	(47,850)	(8,450)
R^2	0.99	0.95	0.95	0.95	0.99	0.99

Note: Standard errors are in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

Column I reports results derived from a regression with dummy variables for each county estimated with Ordinary Least Squares (OLS) and clustered variances. Column II presents results derived from the most frequently employed FE modeling routine using the STATA "xtreg, fe" command. Results in column III come from the user-written "xtivreg2" command, which is robust to heteroscedasticity Standard Errors (SE). Column IV presents the "xtreg, fe" results derived from 400 bootstrap replications with the cluster-robust SE. Results in column V are derived from the "areg" command with 400 bootstrap replications clustering on counties. Results in column VI are estimated by the user-written codes "fese."

All the estimation strategies give consistent point estimates with varying SE. The consistency is due to the six models all being based on the same rationale. A small portion of the varying SE is due to the programming design behind different estimation routines, and a more significant portion lies in the degree of freedom adjustments (see Appendix A for detail). However, the dominant difference of the SE is due to the variance-covariance structures specified for exploring the data's potentials (also see Appendix A for details). Of course, the limited sample size is another reason for the unstable SE when bootstrapping is employed.

The coefficients of farmland and wetland estimated by the six alternative strategies match very well—1.14 units of farmland expansion is associated with one unit of forestland loss; meanwhile, 0.82 unit of wetland loss prevents one unit of forestland from loss. Therefore, the evidence supports the inclusion of wetland change in the regressions.

4.1.2 Initial Analysis Based on Land Use Categories: Random Effects Estimation

The general specification of a RE model is similar to the FE counterpart, with the fixed effect α_i being absorbed. In the following equation, τ_{it} stands as observation-specific random errors.

$$Ft_{it} = f(Fm_{it}, Wt_{it}) + \tau_{it} \qquad \text{(Eq. 4.2)}$$

I will employ four commonly used estimators in my analysis, all of which assume the unobserved heterogeneities are uncorrelated with the independent variables. They are the between-model estimator (Model I), the generalized least square (GLS) random-effects estimator (Model II, IV and V), the maximum likelihood estimator (MLE) (Model III), and the generalized estimation equation (GEE) with population-averaged estimator (Model VI). As shown in Table 4.2, the four different estimators have produced different results.

Table 4.2 Preliminary results of the drivers of forestland change assuming that the unobserved heterogeneities are random

VARIABLES	I xtreg_be	II Mdlk	III xtreg_mle	IV xtreg_re	V xtreg_rebs	VI xtreg_paexbs
Farm	0.37	-1.14***	-0.71*	-1.12***	-1.12***	-1.13***
	(0.81)	(0.04)	(0.41)	(0.05)	(0.26)	(0.21)
Others	-2.31	-0.82***	-0.44	-0.80***	-0.80	-0.81***
	(5.03)	(0.11)	(1.42)	(0.12)	(1.08)	(0.25)
mean_Farm		1.51***				
		(0.46)				
mean_Others		-1.50				
		(1.70)				
Constant	89,424	89,424	251,614*	332,581***	332,581***	334,040***
	(167,906)	(82,084)	(130,276)	(37,012)	(66,133)	(37,532)

Note: Standard errors in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

The models in Table 4.2 offer different perspectives of the data structure (see Appendix B). For example, contrary to fixed effects estimates, which discarded the differences between counties through the process of subtracting the mean differences across unit of observation, Model I treats the cross sectional/between-county variations as its focus. As the between variations have little explanatory power, this relationship is weak and proves that the FE model (i.e., the within estimator) did not lose much useful information during the demeaning process and is valid in explaining the general forestland transitions. Also, in Model II, the significant correlations of the averaged farmland and other land call into question the validity of the RE assumption—the observed variables are uncorrelated with the unobserved heterogeneities. This indicates that when the coefficients differ a lot between the FE and RE models, the FE estimates are probably more appropriate. Moreover, the poor performance of Model III cautions me that the small dataset may not fit the normal distribution assumption related to the classical MLE.

In addition to the important ramifications discussed above, these models have also verified the key findings shown in Table 4.1. Under the assumption that the unobserved heterogeneities are

random, the correlations between deforestation and farmland expansion fall in the range of -1.12

and -1.13. These are close to the fixed effect estimates -1.14. Also, the correlation coefficient of

wetland change with farmland expansion between is -0.80 and -0.82, and the corresponding

coefficient from the FE models is -0.82. These results confirm the dominant role of agricultural

expansion in forestland loss as well as the importance of considering substitution between

forestland and wetland in analyzing the driving forces behind the LUCC in general and

deforestation in particular.

In short, this section not only serves as the analytic basis but also offers guidelines for

model selection in the following section. Based on the LUCC data extracted from satellite images,

deforestation is mainly correlated with farmland expansion and wetland change; the estimated

coefficients are descriptive of the average land conversion ratios. As such, these coefficients could

also be a gauge for evaluating the appropriateness for the following models.

4.2 Augmented Analysis of Deforestation Drivers

4.2.1 Model Specification

Coupled with a clear understanding of the advancement in land change science (Angelsen

& Kaimowitz, 1999; Geist & Lambin, 2002; Kaimowitz, 1998; Lambin et al., 2001; Turner, et al.,

2008) and the history of forest transition in northeast China (Xu et al., 2006; Yu et al., 2011; Zhang

et al., 2011; Zhang et al., 2000), the initial results in section 4.1 have presented a solid starting

point to specify my own model of the forces driving deforestation, in which I will include

agricultural expansion and wood extraction as the two main direct causes for deforestation.

Farmland (*Fm*) and *forestland* (*Ft*) are variables derived from the LUCC detection. Wood

extraction includes government-sanctioned timber harvests and local farmers' consumption of

fuelwood as well as construction timber. As there are no direct and accurate measures of wood extraction, I will use the *gross output value of forestry (O)* as a proxy. The data for this variable came from the Heilongjiang Statistical Yearbook, and the nominal output values were deflated with the GDP deflator (1976 as the base year).

During the study period (1977-2007), the regional forest sector witnessed heavy logging and thus resource degradation in the 1980s and 1990s; by the turn of the century, however, the *Natural Forest Protection Program (NFPP*, shortened as *N* in Eq.4.3), one of the largest ecological restoration programs in China (Xu et al., 2006), had been initiated. So, the year 2000 could be a turning point of the overall management policy affecting forestland use (Yin & Yin, 2009). A dummy variable is created to reflect the implementation of the NFPP.

Timber price (Tp) change is another important factor that influences the behavior of forest enterprises and farmers and thus the forest condition. Low prices could make profit-orientated farmers switch their production efforts from logging to cropping (Yin & Newman, 1996; Yin et al., 2003) and cause the forest entities to neglect their management duties (Yin, 1998). Thus, timber price change could affect the aggregate timber supply as well as local timber inventories, and, coupled with excessive logging, could even lead to the deterioration of forest resources and subsequently impact the LUCC (Lambin et al., 2001). Timber price data were gathered from the Forest Industry Bureau of Heilongjiang Province with a unit of yuan/m^3 and they were deflated with the provincial-level Consumer Price Index (or CPI, with a base year of 1976) to obtain the real price series.

I also assume that a shorter distance and thus lower transportation cost facilitate wood extraction and annual-crop cultivation by local farmers, and even make it possible to convert land being used for other purposes into farmland. More specifically, I will take *distance (D)* from the

forest farms to the nearby timber markets, as well as the seats of the counties where the farms are located, as a proxy measure of transport costs. The process of data generation on this variable is the following. First, I extracted the centroids of each forest farm polygons with a total of 171 points. The number is larger than the total number of forest farms in the study area, because sometimes one forest farm has jurisdiction over several patches of forestland. Then, I extracted the centers of the county seats and included the largest timber markets located close to the study region. These markets are in the cities of Harbin, Suifenghe, Jiamusi, and Mudanjiang. Utilizing the "spatial join" tool in ArcMap, I got attributes of the 171 points from the county polygon layer. Then I employed the "near" tool to calculate the distances between the forest farm centroids to the 14 cities with a distance ranging up to 1000 km. After that, I calculated the mean distance (Km) from a forest farm to each city for each sample county.

As stated-owned enterprises, forest farms follow specific regulations imposed by the central government, such as the logging and reforestation quotas (Xu et al., 2004). I include the *numbers of government-owned forest farms* (*Nf*) in my model based on the assumption that the more clustered forest farms are in a county, the larger their aggregate effect is in protecting forests from farming encroachment. Such effects could be reflected geographically and institutionally—the locally clustered forest farms reduced the possibility of disturbance of human activities and thus avoiding fragmentation and further forestland loss; also, with more organizational presence, there would be more supervisory power that could lead to less excessive deforestation and better policy implementation (Key & Runsten, 1999).

Further, *population (P)* and *Gross Domestic Product* (or *GDP*), are two most frequently used indicators in land use change analyses. The widely acknowledged effects of population dynamics on LUCC mainly occur through the direct actions of clearing land for shelter and

meeting increasing demand for forest products (Carr et al., 2005; Geist & Lambin, 2002). As local population grows and spreads, more farmland is converted into built-up areas; clearing patches of forest for farming is inevitable in order to meet local farmers' growing production needs. Also, population growth is closely linked to increases in wood products consumption and fuelwood demand. GDP is an indirect indicator, predicated on the theoretical reasoning embedded in the environmental Kuznets curve, which hypothesizes that as an economy develops, deforestation rates tend to first increase and then decrease (Bhattarai & Hammig, 2001; Koop & Tole, 1999).

Based on the above discussion, the general model of deforestation determinants can be expressed as:

$$Ft_{it} = f(Fm_{it}, O_{it}, N_t, Tp_t, D_i, Nf_i, P_{it}, GDP_{it},) + u_{it} \qquad \text{(Eq. 4.3)}$$

In Eq. 4.3, the subscript i denotes county; if i is not present, it means that county level data are not available and provincial data are used instead. Similarly, t denotes time; if a variable, such as distance to markets, does not vary with time, t is absent from the variable's subscript. The error term, u_{it} represents the effects of the omitted variables that are peculiar to both the individual units and time periods. Under the fixed-effect assumption, u_{it} is the combination of an independently identically distributed (*i.i.d.*) random error ε_{it} and an unobserved heterogeneity α_i peculiar to county i over time (Hausman & Taylor, 1981; Nickell, 1981). Under the assumption that u_{it} is random, then it is just an *i.i.d.* random variable with zero mean and variance σ_τ^2. For detailed statistical information of the variables in Eq.4.3, see Table 4.3 below. The above model will be estimated with the panel dataset of 248 observations—31 years (from 1977 to 2007) in 8 counties.

Table 4.3 Variables for the single equation analysis of deforestation

Var	Definition	Unit	Mean	Std. Dev.	Min	Max
Ft	Forest Area	km^2	1194.52	901.92	5.13	2622.70
Fm	Farm Area	km^2	1773.47	799.59	206.25	2876.01
Tp	Price Index of Timber	1976=100	88.90	23.46	54.50	161.60
O	Gross Output Value of Forestry	1000 ¥	4538.87	5165.00	164.99	33424.47
D	Mean Distance to Large Markets	Km	26.10	9.57	15.96	46.56
Nf	No. of Forest Farm in County	None	6.38	4.04	1.00	13.00
N	0 before 2000; otherwise 1	None	0.30	0.46	0.00	1.00
P	Total Population	1000	305.76	99.79	104.00	527.50

Note: Var means variable and ¥ is a unit of Chinese currency.

4.2.2 Fixed-Effects Estimation

As before, six different estimating methods were adopted in the augmented model in correspondence to the different variance-covariance structures. Results are presented in Table 4.4. Here, I will first focus on illustrating the alterative estimators and their implications.

Table 4.4 Estimation results of the drivers of forestland change with the unobserved heterogeneities being fixed

Forestland	I reg_lsdv_cl	II xtreg_cl	III areg_cl	IV xtivreg2_hac	V fese_hc	VI xtreg_clbs
Farm *(Fm)*	-1.04***	-1.04***	-1.04***	-1.04***	-1.04***	-1.04***
	(0.06)	(0.06)	(0.06)	(0.04)	(0.03)	(0.13)
ForOpt *(O)*	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
NFPP *(N)*	16.96	16.96	16.96	16.96	16.96	16.96
	(12.52)	(12.34)	(12.52)	(13.56)	(11.13)	(14.59)
TimberPrice *(TP)*	0.16	0.16	0.16	0.16	0.16	0.16
	(0.44)	(0.43)	(0.44)	(0.26)	(0.22)	(0.38)
Meandist *(D)*	98.71***					
	(5.63)					
NForFarm *(Nf)*	389.02***					
	(10.51)					
TotalPop *(P)*	-0.55***	-0.55***	-0.55***	-0.55***	-0.55***	-0.55
	(0.14)	(0.14)	(0.14)	(0.08)	(0.10)	(0.44)
Fangzheng	246.09***				2915.48***	
	(48.44)				(24.34)	
Huachuan	1,191.06***				2568.26**	
	(58.12)				(59.70)	
Huanan	289.41***				4549.63**	
	(23.86)				(72.12)	
Jixian	1,322.85***				2454.66**	
	(82.70)				(56.25)	
Qitaihe					896.57**	
					(20.74)	
Suibin					2750.12**	
					(82.89)	
Yilan	-158.19***				4781.62**	
	(22.16)				(80.91)	
Boli					4548.90**	
					(62.12)	
Constant	-2,235***	3,183***	3,183***		3,183***	3,183***
	(114.6)	(128.0)	(129.8)		(55.55)	(532.8)
Observations	248	248	248	248	248	248
R^2	0.996	0.884	0.996	0.884	0.996	0.884

Note: Standard errors are in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

The first estimator in Table 4.4, marked as "reg_lsdv," used the least square dummy variable and the corresponding standard error was estimated using the clustered-robust variance-

covariance matrix. The second estimator used the widely used fixed effect analysis routine of "xtreg, fe." The third estimator employed a similar STATA routine "areg" with the cluster-robust standard errors (CRSE), the same as Estimator V in Table 4.1. The fourth estimator utilizes a user-written package "xtivreg2" as a wrapper for *ivreg2*. So, it is close to Estimator III in Table 4.1. A major alteration I made for the Estimator IV was that, rather than using the bootstrapping cluster-robust standard errors, I specified the heteroscedasticity and autocorrelation consistent (HAC) standard errors with the Bartlett kernel; the bandwidth I chose here was 2. Estimator V implemented the "fese" module with heteroscedasticity-robust (Hr) errors. The last estimator used the *xtreg* routine with 400 bootstrap replications clustering on counties.

The first three estimators were based on the clustered-robust variance covariance matrixes, but some subtle differences between them can still be seen. The SE from estimators *LSDV* and *areg* are relatively larger than those from estimator *xtreg*, which can be attributed to the different degrees of freedom adjustments: *areg* subtracts the degree of freedom by the number of unit effects that were swept away in the within-group transformation in FE estimation, while *LSDV* and *xtreg* do not make such degrees of freedom adjustments. When observations for any group are classified exactly within the same cluster, *xtreg*'s output is considered to be more appropriate (Gould, 1996; Gould, 2013).

I considered three different standard errors: *Newey-West* standard errors (or HAC) (Hoechle 2011) in Estimator IV (see Appendix 4.4.3 for more detail), Hr in Estimator V, and CRSE in Estimator VI. Compared to Estimator V, which considers autocorrelations in the time dimension, SE in Estimator IV are larger. Estimator VI reports the largest SE; my interpretation of the difference is that the SE estimated by OLS are biased downward when a large proportion of variability is due to fixed effects. The HAC are also biased but with relatively small magnitude.

Of the three estimators, the clustered standard errors should be closer to the true errors (Petersen, 2009).

Fixed unit effects are reported only for the LSDV and *fese* estimators. These unit effects were generated by different mechanics. Dummy variables were created in the LSDV estimation. In order to avoid the multicollinearity, STATA automatically excluded one unit (Boli in my sample). All the other unit effects reported are the disparities from the unreported fixed effect of Boli. In the *fese* estimation, the intercept is the average value of the fixed effects while the specific unit effects were the differences to the mean fixed effects. So, STATA drops dummy variables in LSDV due to multicollinearity, but this does not happen to the *fese* estimator.

All the six regressions report identical coefficient estimates. First, one unit of forestland loss is associated with 1.04 units of farmland expansion. Second, the policy dummy NFPP has a positive but insignificant effect on forestland. Similarly, deforestation is correlated with slowly rising timber prices, but the relationship is not significant. Further, the gross output value of forestry is little correlated with deforestation. The coefficient of mean distance suggests that forests closer to the timber markets have a greater likelihood to be depleted. Finally, the significant positive coefficient of number of forest farms indicates that counties with more forest agencies tend to have less deforestation.

4.2.3 Random Effects Modeling Results

The key estimation options for random effect models are the between-effects estimator (BE) (I in Table 4.5 below), the Mundlak estimator (II), the random effect estimator (or RE and MLE) (III, IV, and VI), and the population-averaged estimator (or PA). Except estimator II, consistency estimation requires that the error term be uncorrelated with the regressors.

Estimator I used only the cross-sectional information in the data, the information reflected in the changes between counties. Estimator II was developed to relax the assumption that the observed variables are uncorrelated with the unobserved heterogeneities, providing additional details on the within and between variation of the independent variables. Here, the coefficients of the original regressors were calculated based on the within estimator, so these values are the same as those of the fixed effects model in Table 3. Meanwhile, the coefficients related to the mean of time-varying variables are tabulated based on the difference of between and within estimators. Estimator VI was based on 400 bootstrap samples; as the error term is likely to be correlated over time for a given county, it is essential that OLS SE be corrected for clustering on the counties. Estimator IV assumes the observed heterogeneities and the idiosyncratic errors are normally distributed. Through maximizing the log of the likelihood function, the MLE coefficients are consistent when T is large (Laird & Ware, 1982; Raudenbush et al., 2000). Estimator V is also called the generalized least square estimator in the literature. As the observed heterogeneities are assumed to be random and averaged out, this estimator is consistent.

Table 4.5 Single equation models assuming that the unobserved heterogeneities are random

Forestland	I xtreg_be	II Mdlk	III xtreg_re	IV xtreg_mle	V xtreg_paex	VI xtreg_rebs
Farmland*(Fm)*	-0.20	-1.04***	-1.00***	-0.99***	-1.03***	-1.00**
	(0.12)	(0.03)	(0.03)	(0.07)	(0.06)	(0.41)
ForOpt*(O)*	0.12*	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
NFPP*(N)*		16.96	15.01	14.86	16.66	15.01
		(11.13)	(12.17)	(28.90)	(12.29)	(36.65)
TimberPrice*(Tp)*		0.16	0.08	0.07	0.15	0.08
		(0.22)	(0.24)	(0.56)	(0.43)	(1.24)
Meandist*(D)*	11.60	11.60	71.21***	71.02***	73.39***	71.21
	(10.99)	(10.99)	(7.35)	(16.90)	(21.56)	(104.65)
NForFarm*(Nf)*	187.89**	187.89***	304.82***	304.58***	307.66***	304.82**
	(33.74)	(33.74)	(17.06)	(39.07)	(44.43)	(147.22)
TotalPop*(P)*	-2.57	-0.55***	-0.55***	-0.55**	-0.55***	-0.55
	(0.91)	(0.10)	(0.11)	(0.25)	(0.14)	(1.89)
M(Farm)		0.83***				
		(0.12)				
M(ForOpt)		0.12***				
		(0.04)				
M(TotalPop)		-2.02**				
		(0.91)				
Constant	292.26	273.71	-679.67***	-677.57	-703.30	-679.67
	(374.92)	(375.35)	(255.33)	(584.20)	(874.21)	(1,516.02)

Note: Standard errors in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

Results derived from Estimator I indicate that there is not much between-county variation with respect to the driving forces. Compared to the results of *xtreg, fe* in Table 4.4, it can be inferred that a large part of the changes in forest cover came from the time changing effects within counties. Estimator II incorporates both between- and within-county variations. The significances in coefficients of farmland, forest output and total population indicate that the random effect assumption may be too strict, i.e., these variables are probably correlated with some of the unobserved heterogeneities.

Estimator IV assumes that the cross-sectional effects are normally distributed. This normal distribution assumption was rejected in an earlier analysis of the 48 original observations (see Table 4.2 and Appendix B for further information). But when the 248 annual observations are used and the model is augmented, the coefficients of the current estimator become closer to those derived from other estimators.

There are several other covariance matrix options, like "independent (absence of covariance and correlations)" and "unstructured" (unconstrained pairwise correlation) for the population-averaged estimator V, but many of them are not realistic due to the small sample size. Thus, relatively, it seems to fit better under the default "exchangeable" error correlation assumption (uniform correlations across time). The difference between Estimators III and VI is that Estimator VI is based on 400 bootstrap samples. From the results, it can be seen that the standard errors changed considerably.

Overall, the differences between the estimated RE results and the FE ones are relatively small. The RE coefficients of farmland are around -1, close to those derived from the FE estimators. Also, the coefficient magnitudes of other variables, like NFPP, forestry output, and timber price, as well as population, are similar. Further, the coefficient significances of all the variables are identical between the two approaches. In Table 4.5, the coefficients of time-invariant variables— number of forest farms in a county and average distance from the forest farms to nearby county seats and markets—are not dropped. Thus, the effect of administrative arrangements and the geographic influence can be quantified by the RE model, which is complimentary.

4.2.4 Long Panel Data Analysis

As the dataset covers 31 years, exploring the information of the panel dataset with annual observations could offer more insight into how I might improve my results. A key difference in model specification between the repeated cross-sectional and panel data is that with the former, it is impossible, and perhaps unnecessary, to deal with serial correlation, while with the latter, it is necessary and feasible to consider serial correlation. Thus, serial correlation is generally assumed for the error term when panel data are used (see Tables 4.6 and 4.7).

Forestland	I pw_iid	II pw_car1	III pw_ar2	IV pw_psar1	V pw_psar1dw	VI fgls_psar1	VII fgls_cpsar1	VIII regar_fear1	IX regar_rear1
Farm*(Fm)*	-0.41***	-0.59***	-0.41***	-0.71***	-0.67***	-0.73***	-0.76***	-0.67***	-0.75***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)
NFPP*(N)*	57.60	10.11	57.60*	12.56**	11.02*	3.07	11.45***	15.64***	12.09**
	(45.04)	(6.76)	(30.22)	(6.16)	(6.14)	(4.90)	(2.04)	(5.32)	(5.80)
TimberPrice*(Tp)*	1.11	0.09	1.11**	0.07	-0.01	-0.06	-0.18***	0.05	-0.06
	(0.88)	(0.15)	(0.40)	(0.14)	(0.12)	(0.10)	(0.04)	(0.11)	(0.11)
ForOpt*(O)*	0.01***	0.00	0.01***	0.00	0.00	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Meandist*(D)*	26.10***	45.56***	26.10***	49.75***	39.19***	43.57***	63.87***		57.13***
	(2.10)	(3.26)	(3.00)	(3.71)	(4.34)	(3.21)	(1.18)		(10.61)
NForFarm*(Nf)*	252.07***	264.51***	252.07***	290.01***	256.08***	271.66***	276.20***		277.83***
	(5.21)	(5.84)	(5.15)	(7.01)	(10.69)	(6.72)	(2.63)		(24.96)
TotalPop*(P)*	-1.15***	-0.23*	-1.15**	-0.24*	-0.11	-0.06	-0.10***	-0.02	-0.14
	(0.33)	(0.13)	(0.42)	(0.13)	(0.11)	(0.08)	(0.03)	(0.10)	(0.10)
GDP	-0.00***	-0.00***	-0.00**	-0.00***	-0.00**	-0.00***	-0.00***	-0.00	-0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-28.70	- 512.85***	-28.70	- 603.22***	-106.71	-224.29**	-733.86***	2,201.17***	-672.45*
	(110.73)	(70.81)	(150.32)	(84.40)	(73.94)	(94.95)	(37.76)	(2.30)	(374.48)
R^2	0.94	0.93	0.94	0.97	0.92				

Note: (1) The model specification details are listed in Table 4.7. (2) Standard errors in parentheses. (3) *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively. (4) Model VI and Model VII do not report R-squared but with *Wald* chi^2 (8) equals to 2313.96 and 30351.54 respectively. The within R-squared for Model VIII is 0.71 and the between R-squared for Model 9 is around 0.87.

Table 4.7 Different autocorrelation and panel correlation specifications

	Panels	Autocorrelation	Estimator
Model 1	Heteroskedastic	No	Pooled OLS
Model 2	Correlated	AR(1)	Prais-Winsten
Model 3	Heteroskedastic with CS correlation	AR(2)	Pooled OLS
Model 4	Correlated	panel-specific AR(1)	Prais-Winsten
Model 5	Correlated	panel-specific AR(1)	Prais-Winsten
Model 6	Heteroskedastic	panel-specific AR(1)	Two-step FGLS
Model 7	Heteroskedastic with CS correlation	panel-specific AR(1)	Two-step FGLS
Model 8	Independent	AR(1)	FE
Model 9	Independent	AR(1)	RE GLS

Note: Table 4.7 is an explanation of the models used in Table 4.6, and CS stands as cross-sectional.

In order to improve modeling efficiency, I have employed several different techniques of coefficient estimation. Estimators I and III are pooled OLS ones; Estimators II, IV and V are Prais-Winsten ones; Estimators VI and VII use the FGLS and Estimators VIII and IX apply the within estimator and the GLS to obtain the FE and RE results. Moreover, in order to account possible correlations over time and between counties and insure the reliability of estimation results. I included different estimation packages (*xtpcse, xtgls, xtscc,* and *xtregar*) to adjust the SEs of the coefficient estimates for possible dependence in the residuals. Brief introduction of these packages and their specialties are generalized in the sub-section of 4.4.3, and the specific estimation procedures and interpretation of the corresponding result will follow.

Compared to results reported in the previous sections, it is obvious to see that the overall correlation between farmland and forestland is smaller in magnitude in the panel-data regressions. For example, the estimated minimum coefficient is -0.76, while the coefficients are around -1 in the FE and RE versions of the model. A straightforward way to decide the appropriateness of different estimators is to check the estimated results against the proportion of land change. From the conversion matrixes in Chapter 2, we know that the farmland gain is always a little larger than forestland loss. Thus, it is easy to tell that the FE and RE versions of my model in the previous

sections better fit the data. The under performance of the panel-data estimation has to do with the data generation mechanism. That is, the deficiencies of interpolated data make the estimated results less reliable when capturing the autocorrelation or differences from the means.

Nonetheless, the panel-data analysis provides some useful information. In the case of a small N, it seems that specifying the contemporaneous correlation between cross-sections is not suitable; but exploring the autocorrelation of panel data becomes beneficial. For instance, with all the disturbances being cross-sectionally correlated, the results of Estimators II-V vary a lot; however, once the panel-specific $AR(1)$ is considered by Estimators IV and V, the coefficient of farmland sees an immediate increase and is much closer to its counterpart found in sub-section of 4.1.1. Also, Estimator VII gives the most expected coefficient signs to the results, under the assumptions that the data are heteroskedastic with cross-sectional correlation and that each cross-section is auto-correlated. Estimator IX is less optimal because while it assumes data auto-correlated with one lag, it does not consider the cross-sectional correlation.

The panel-data analysis is also helpful for choosing a more appropriate estimation method. Different estimators are rooted in different methods of parameterization. With the same model specification, it seems that the FGLS estimators present relatively more consistent and efficient parameters (see estimators VI and VII). FGLS enables me to account for dependence over time for each county; and more importantly, the asymptotic properties of FGLS with a small sample size make it out-perform other estimators (Altonji & Segal, 1996).

4.2.5 Model Validations

Estimation Model Selection

Models listed in the previous sections explore the potentials of how the data would be utilized under different specifications and error structures. Thus the first question I am going to address for this small section is which model reflects the data and covariance structure best. It is straightforward to see that the data interpolation has caused the estimates of long panel analysis in section 4.2.4 to be biased. Thus, comparisons here will be only between the FE and RE models.

The between-effects estimator (Model I) in Table 4.5 utilizes the variations that are discarded from the within estimators, i.e. the fixed effects estimators. The poor estimation results of Estimator I in turn suggest that the FE model actually captured the dominant variations of forestland change. This implies that the FE models are more reliable. Meanwhile, the Mundlak model (Model II) in Table 4.5 also proves that the FE analysis fit the data better. The significances in coefficients of the mean values of farmland, output value of forestry and total population imply that the random effect assumption are relatively too strict; that is, some explanatory variables are potentially correlated with the unobserved heterogeneities. So, the within estimators instead could do better by taking into account the cross-sectional heterogeneities.

Thus, both Estimator I and II in Table 4.5 confirm the validity of FE models. To be cautious, though, other tests are also considered here. Among them, the *Hausman* test is the most widely employed one. However, a weakness of the *Hausman* test is that it assumes the RE model is efficient by default, which violates the assumption of cluster-robust standard errors in several of the estimators listed in Table 4.4 and Table 4.5. To overcome this weakness, I constructed the Sargan-Hansen test suggested by Arellano (1993) and Wooldridge (2002, pp. 290-91). As an RE model requires that the independent variables are uncorrelated with the county-based unobserved

heterogeneities. This additional orthogonality condition features the over-identification restrictions. The P-value of Sargan-Hansen test is less than 1%, which rejected the null that these additional orthogonality restrictions are valid. Thus, it is safe to conclude that the FE model is more appropriate.

Variable Selection

The drivers in the augmented models are predicated on insights found in the literature, and they are thus expected to be relevant causes to the deforestation in northeast China. Some of the variables didn't turn out as expected, like the insignificant coefficient of the NFPP. For some reason, these unexpected results could be possibly be attribute to the specific local context as well as overall model speciation problem (see further analysis in Chapter 5). In order to seek a model that is more concise in capturing the deforestation mechanisms, I employed the Akaike's information criterion (AIC) and Bayesian information criterion (BIC) as two indicators for better balancing between models fit and complexity. A model is considered to be closer to the truth as the AIC and BIC values are the smallest. I started with the whole set of variables in FE models and recorded the corresponding AIC and BIC values. The formal stepwise selection method doesn't support panel data analysis. As I gained knowledge of the data, I could manually try out different variable combinations. Table 4.8 below listed the all the AIC and BIC values with respect to each model.

Table 4.8 Variable selection process and corresponding AIC and BIC values

Forestland	(I) All	(II) ForOpt	(III) TimbPrice	(IV) NFPP	(V) TotalPop	(VI) Wetland
Farmland	-1.10***	-1.10***	-1.12***	-1.13***	-1.15***	-1.03***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.07)
Wetland	-0.96***	-0.96***	-0.92***	-0.88***	-0.85***	
	(0.08)	(0.08)	(0.09)	(0.10)	(0.10)	
TotalPop	-0.40**	-0.40**	-0.47***	-0.54***		
	(0.12)	(0.13)	(0.10)	(0.06)		
NFPP	-12.19	-11.42	-21.09			
	(11.09)	(9.61)	(13.80)			
TimbPrice	-0.48**	-0.47**				
	(0.19)	(0.19)				
ForOpt	0.00					
	(0.00)					
Constant	3,485.54***	3,483.98***	3,487.54***	3,523.29***	3,378.55***	3,016.37***
	(47.79)	(53.08)	(55.45)	(58.02)	(53.80)	(127.32)
AIC	2383.55	2381.80	2396.74	2410.08	2510.29	2723.67
BIC	2404.63	2399.37	2410.80	2420.62	2517.32	2727.18
R^2	0.97	0.97	0.97	0.96	0.94	0.87

Note: (1) All the models in Table 4.8 were estimated using the most frequently used xtreg, fe routine with heteroskedastic-robust standard errors. (2) Robust standard errors in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

From Table 4.8, it is easy to recognize that Model II giving smallest AIC and BIC values over the set of models considered. Thus, model II meets the requirement with the annual output value of forestry being dropped out. The coefficient of output value of forestry is approximately 0; so, from now on this variable will not be included in the following analysis.

4.3 Discussion and Conclusions

In this chapter, I have employed a series of empirical methods to investigate the effects of various forces driving deforestation in northeast China. Although variations resulted from different estimators, the coefficients tended to be in general agreement. First, the rate of deforestation is highly associated with farmland expansion—a one-unit loss of forestland is tied to more than one

unit of farmland expansion. Also, forests located closer to a county seat and/or a large timber market tend to have a higher probability of deforestation; counties with more forest farms and thus a greater presence of forestry administration in their jurisdictions seem to have a lower risk of forestland loss. In addition, population growth is also strongly associated with a higher rate of deforestation. As for the effect of implementing the NFPP, all the models corroborated the finding that it is positive, though insignificant. Finally, the influences of forestry output and GDP on forestland reduction are weak and thus negligible.

Some of the estimated coefficients seem counter-intuitive. For instance, timber price is positively, and insignificantly, correlated with forestland changes. It is generally thought that timber price increases would lead to more logging and thus deforestation at least in the short run, so that the impact should be significantly negative. My conjecture is that under government market control, timber prices were depressed and thus did not play much of a role in the study region. Thus, my analysis reflects that timber price has little correlation with forestland change. Moreover, it is conceivable that the long-run price effect may be positive if the incentive structure for reforestation and forest management can be improved persistently.

Because there are no direct and accurate measurements of annual wood extraction, the gross output value of forestry was used as a proxy. It can be seen from Table 4.4 and Table 4.5 that forestry output is negatively associated with forestland change, as expected. But the coefficient is insignificant, too. This could partly be attributed to the imperfect approximation using the gross output value of forestry, but it could also indicate that local farmers as well as forest-based industries tend to under-report the actual quantities of wood extraction.

The results of the augmented single-equation reveal a strong linkage between population growth and deforestation, which is consistent with a majority of the reported evidence in the

118

literature (Angelsen & Kaimowitz 1999; Geist & Lambin 2001; Carr et al. 2005). As other income opportunities for local farmers are limited, families living on the edges of forests continue clearing land to expand farming and increase their revenues. Even in the days of more developed agricultural technologies and labor shifts away from agriculture, it remains a common practice for local farmers to reclaim forestland for cultivation.

The effects of "*Meandist*" (distance from the forest farms to the nearby timber markets and county seats) and "*NForFarm*" (numbers of forest farms within a county) are detected mainly through the RE versions of my model. As expected, the evidence indicates that forests closer to the large markets and cities have a larger probability of being cleared. Similarly, because forest farms are the grassroots units of forest organization, I presumed that counties with more forest farms tend to have less deforestation. The estimated effects in the RE analysis and the LSDV version of the FE analysis give clear support to my hypothesis.

My results also suggest that there is considerable variation across counties. Both from the initial and augmented single-equation analyses, the county dummy variables are statistically different from zero at a 95% or higher significance level. This implies that even if I have tried to incorporate the potentially important causes of deforestation, it appears that the data I gathered may not allow me to capture the heterogeneity in my model due to either the missing variable problem or the limited size of my observations. With a small sample, of course, empirical results are sensitive to the model specification and related assumptions.

In this chapter, I have explored both FE and RE approaches to econometric estimation of a single-equation model. The differences between the estimated results o0f the FE and RE methods are fairly small. Still, a close comparison between these results has led to some interesting implications. First, results from different FE estimators are consistent, with a major difference

lying in the specifications of error structures and degrees of freedom adjustments embedded in the estimators. When the unobserved heterogeneities are assumed to be random, the weak explanatory power of the between estimator lends further confidence to the FE method. Also, the significant coefficients of the time-varying variables confirm the validity of the FE assumption.

Thus far, my empirical work has assumed that the explanatory variables of deforestation analysis are exogenous. Within the RE modeling framework, the assumption that the error term and the regressors are uncorrelated has been crucial. In comparison, the FE methods can moderately mitigate the threat of endogenous bias as they can deal with the dependence between the disturbances and the regressors. However, when the unobservable effects are time-varying, an FE estimator cannot fully rule out the endogeneity bias. Additionally, a key limitation of FE methods is that they are not able to determine the effect of a variable that has little within-group variation. Therefore, in the next chapter I will try to address the potential endogeneity problem by developing and estimating alternative models based on the instrumental variable method and a system of structural equations. I hope that combining my efforts here and in the next chapters will enable me to derive robust findings.

APPENDICES

Appendix A: Description of the Initial Fixed-Effects Regressions

In this sub-section, I will present the detailed estimation procedures and outcomes of the initial FE regressions. A number of estimators have been used to explore the stability of the regression outcomes. These estimators make different assumptions about the variance-covariance structure of the empirical model. Specifically, Estimator I, called the least-squares dummy variable estimator (LSDV), combines the traditional OLS procedure with dummy variables. It captures the unobserved heterogeneity (or unobserved effect) with the coefficients of the individual-specific dummy variables (Andrews et al., 2006; Stimson, 1985). A dummy variable is a binary variable that is coded either 1 or 0, and it is commonly used to examine individual (or group) and time effects in a regression model. In my case, dummy variables represent different counties, or cross-sections in the sample. In STATA, a dummy variable is created by prefixing the notation *xi* with the *regress* command and specifying the sample unit. To avoid the dummy variable trap (perfect multicollinearity), STATA arbitrarily chooses one unit to be the reference (without coding this county as a dummy). Given the need for dummy variables and computational feasibility, the LSDV estimator is not very practical when there are a large number of individuals in the panel data (Andrews et al., 2006).

In Estimator II, *xtreg* is used for the purpose of estimation in panel-data settings—fixed-, between-, random-effects, and population-averaged linear models. In a fixed-effects (FE) model, *xtreg* captures within-group variation by computing the differences between observed values and their means. But the output of *xtreg* is less informative than what is derived from an LSDV estimator with explicit dummy variables. On the other hand, when creating a dummy for each unit leads to too many explanatory variables, *xtreg* becomes more efficient (Hamilton, 2012). The STATA software estimates an FE model $y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$ with grand means of y_{it},

X_{it}, μ_i and ε_{it}. That is, it estimates $y_{it} - \bar{y}_i + \tilde{y} = \alpha + (X_{it} - \bar{X}_i + \tilde{X})\beta + (\varepsilon_{it} - \bar{\varepsilon}_i + \tilde{\mu}) + \tilde{\varepsilon}$ under the constraint $\tilde{\mu} = \sum \mu_i = 0$. So, adding grand means to both sides of the equation has no effect on the estimated coefficients β (Gould, 2013).

In comparison, Estimator V (*areg*) handles a model by absorbing its categorical factors (unit effect or unobserved heterogeneity). Note that *areg* was designed for identifying linear regression with many groups, but not groups that increase with the sample size (that is, the number of parameters remains unchanged while the sample size increases). On the other hand, *xtreg, fe* handles cases where as sample size increases, the dimension of unit effects also increases (Andrews et al., 2006; Guimaraes & Portugal, 2010). Both *xtreg, fe* and *areg* present the intercept calculated at the means of the independent variables as equal to the mean of the dependent variable, or $\bar{y} = \bar{x}\hat{\beta}$; the reported intercept is therefore the average value of the fixed effects. But the calculation of R^2 is different with these two procedures. In *xtreg, fe,* the unit effects for different groups are subtracted, whereas in *areg*, R^2 is based on the part explained by X plus each dummy variable for the unit effect (Gould, 1996). The standard errors also differ when cluster-robust variance–covariance matrix is used. That is, *areg* reports larger cluster-robust standard errors because it subtracts the degree of freedom from the number of unit effects swept away in the within-group transformation, but *xtreg, fe* does not use such degree of freedom adjustments. When observations for any group are classified in the same cluster, *xtreg* is considered to be more appropriate (Wooldridge, 2010).

The code of Estimator III, *xtivreg2*, is user-written. It is an upgraded version of STATA program *ivreg2*, which mainly implements IV/GMM estimations. By omitting the IV options, *xtivreg2* also supports a FE model with no endogenous variables, and this is not allowed in the official STATA program of *xtivreg* (Schaffer, 2012). So, *xtivreg2* offers a variety of choices

between HAC standard errors and cluster-robust options, and thus the standard errors given by

xtivreg2 can be made consistent to various violations of *i.i.d.* error assumption (Baum et al., 2007).

The R^2 reported by *xtivreg2* for the FE estimation is the "within R^2" obtained by the mean-

differenced regression. Standard errors displayed by *xtivreg2* with clusters are by default without

degrees-of-freedom adjustments for the number of fixed effects. While for FE estimation without

cluster, the standard errors are adjusted for the number of fixed effects. In a small sample setting

and with no endogenous variable, the "small" option corrects standard errors by the degree of

adjustment (N-N_g-K), where N_g is the number of groups (clusters) and K is the number of regressors.

And the small-sample adjusted standard error matches those from *areg* and *xtreg*.

Estimator VI (*fese*) is also a user-written package built on the *areg* procedure. More than

what *xtreg* and *areg* do, *fese* also estimates FE and their standard errors, which are saved into the

dataset by default (Mihaly et al., 2010). This estimator produces the standard errors not usually

generated in other programs of FE estimation. Like *xtreg* and *areg, fese* can incorporate the

ordinary, heteroskedasticity-robust, and cluster-robust SE as well. But Nichols (2008) cautions

that when implementing the cluster-robust SE, the usual asymptotic justification does not apply,

so it is better to avoid using cluster-robust SE for application purposes. Also, note that the FE

standard errors generated by *fese* only vary across panels, not by individuals.

The coefficients derived with the six estimators are the same, while the estimated standard

errors differ. Estimators II and VI report the FE results with no extra or special data structure

assumptions. The post-estimation heteroskedasticity test is based on the null hypothesis that the

errors are homoskedastic across units ($P=0$ while the null hypothesis is $\sigma(i)^2 = \sigma^2$, where here i

refers to county). With Estimator III, I choose the conventional sandwich variance-covariance

estimator, and statistics reported are robust to heteroskedasticity. Further, a correction of small

sample size bias is made, so the results report the small-sample statistics (F and t-statistics) instead of large-sample statistics (χ^2 and z statistics). Estimators II, III, and VI relax the within-panel serial correlation in the idiosyncratic error term, which is reasonable as the dataset used is not continuous in the time dimension. It includes 6 periods covering a time span of 31 years with irregular intervals. Estimator III employs the heteroskedasticity-robust standard errors as well as a degree of freedom adjustment; thus, among these three estimators, it provides more reliable standard errors.

Now, let me discuss how to incorporate the autocorrelation patterns in the residuals and create a pseudo-sample to relax the constraint of a limited sample size. With Estimator I, I specify the *vce (robust)* option in the model specification by clustering on the unit (county) in order to produce estimates that are robust to cross-sectional heteroskedasticity and within-panel (serial) correlation (Arellano, 1987). It is worth noting that Estimator I in Table 4.1 is a least square dummy variable estimator, while the rest are all within estimators. LSDV and within estimation result in identical coefficient estimates but different standard errors, due to different degrees of freedom corrections. LSDV correctly counts the parameters as $G+K$ rather than the within estimator views as K. LSDV also automatically generate the FE output when dummy variables are included. Estimator IV and V employ the bootstrapping cluster-robust errors. They share almost same estimation procedures; so, their outputs are the same, except for the R^2 values. A closer look at the standard errors in Table 4.1 suggests that the bootstrapping results produced slightly larger standard errors than the others. This is counter-intuitive, as bootstrapping cluster-robust errors are usually downward-biased. Petersen (2009) showed that when fixed effects exist in both the independent variable and the residual, the standard errors estimated by OLS are biased downward. They also conclude that the *Newey-West* standard errors are also biased, but the magnitude of bias

is relatively small. Of the most frequently used approaches, the clustered standard errors are very close to the true errors.

Under different modeling routines, there exist two different R^2 values in Table 4.1. R^2 reported by *xtreg* and *xtivreg2* procedures are 0.951 and R^2 reported by LSDV, *areg*, and *fese* are 0.998. Generally, R^2 reported by the *xtreg* and *xtivreg2* models are lower than the rest. This is because *xtreg* and *xtivreg2* report the within R^2, and the method of calculation for these is different from the usual method. Specifically, R^2 is equal to 1 minus the Residual Sum of Squares (RSS) divided by the Total Sum of Squares (TSS). In my considered cases, the RSSs are all the same, however, the TSSs differ: Conventionally, $\text{TSS} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (y_{it} - \bar{y})^2$; in the *xtreg, fe* routine, it does not report the TSS, but the within sum of squares (or model sum of squares) is calculated by $\sum_{i=1}^{N} \sum_{t=1}^{T_i} (y_{it} - \bar{y}_i)^2$. Based on the different uses of grand mean $\bar{y}$ and unit mean $\bar{y}_i$ during the computation, LSDV, *areg* and *fese* estimators include the variance explained by the absorbed dummies (McCaffrey et al., 2010; Nichols, 2008), whereas *xtreg, fe*, and *xtivreg2* do not

Appendix B: Description of the Initial Random-Effects Estimation

Estimator I employed the between estimator that only utilizes cross-section variation of the data. The between estimator is the OLS estimator of $\bar{y}_i = \alpha + \bar{x}_i'\beta + (\alpha_i - \alpha + \bar{\varepsilon}_i)$. Here, consistency requires that the error term $\alpha_i - \alpha + \bar{\varepsilon}_i$ be uncorrelated with x_{it}. Thus, the between estimator is inconsistent under the FE assumption. In STATA, the between estimator is obtained by specifying the *be* option of the *xtreg* command (Cameron & Trivedi, 2009). From the results in Table 4.2 derived by this estimator, we can see that the coefficients of farmland and other land changes are insignificant, indicating only using the between variations of the predictors cannot effectively explain overall forest land transitions.

Estimator II relaxed the assumption that the unobserved heterogeneities are uncorrelated with the independent variables in the traditional RE estimators by integrating the group-means of $\bar{X}_{i.}'$ in the overall model: $y_{it} = X_{it}'\beta + \bar{X}_{i.}'\pi + \varepsilon_i + v_{it}$ $\quad i=1,...G; \quad t=1,...T$ (Mundlak, 1978), and showed that the generalized least squares estimation yields $\hat{\beta}_{GLS} = \tilde{\beta}_{within} = (X'QX)^{-1}X'Qy$ and $\hat{\pi}_{GLS} = \hat{\beta}_{Between} - \tilde{\beta}_{within} = (X'PX)^{-1}X'Py - (X'QX)^{-1}X'QY$, where P is a matrix that averages the observations across time for each individual and $Q = I_{GT} - P$ is a matrix that obtains the deviations from individual means (Baltagi, 2006; Debarsy, 2012; Mundlak, 1978). With this estimation method, the coefficients on farmland and other land are just the fixed effects estimates in Table 4.1. The averaged values based on county-specific farmland and other land were automatically generated by the estimation techniques. The importance of these mean values in the model proposed by Mundlak (1978) is to test whether the assumption that the observed variables are uncorrelated with the unobserved heterogeneities. Statistical significance of the estimated

coefficients on the group mean of farmland indicates that such an assumption may not hold (Wooldridge, 2010).

Estimator III employed the MLE model (*xtreg, mle*). More than assuming that the unobserved heterogeneities are uncorrelated with *X*, this model also requires that they follow the normal distribution. The coefficients are smaller than those from both the FE and other RE estimators. For instance, a one-unit forestland decrease is associated with a 0.71-unit farmland expansion, which is small compared to the result derived from the conversion matrix. This could be due to the MLE method, which is sensitive to small sample size when distributional assumption for the unobserved heterogeneities is inappropriate (Breusch, 1987; De Janvry et al., 1991; Zellner & Theil, 1992).

The GLS RE estimation *xtreg, re* is widely used in the literature. As stated before, it takes a weighted average of the fixed and between estimates by assuming there is no correlation between the unobserved heterogeneities and *X*. Compared to the coefficients (-1.14 and -0.82) estimated in Table 4.1, the coefficients under the RE assumptions in Table 4.2 are very close to those under the fixed effects assumption (-1.12 and -0.80). The difference of the standard errors originates from the error specification that Estimator V employed in bootstrapping. As the same situation happened in the FE analysis, cluster-robust bootstrapping results produced slightly larger standard errors. This is also due to the within-county correlation between the two predictors.

Estimator VI is a pooled estimator, which simply regresses y_{it} on an intercept and x_{it}, using both cross-sectional and within variation in the data, that is, $y_{it} = \alpha + x_{it}^{'}\beta + (\alpha_i - \alpha + \varepsilon_{it})$. The individual effects $\alpha_i - \alpha$ are now centered on zero. Consistency of OLS requires that the error term $\alpha_i - \alpha + \varepsilon_{it}$ be uncorrelated with x_{it}. Under the assumption that the unobserved heterogeneities are averaged out, the pooled OLS is consistent if the RE assumption is appropriate

128

but inconsistent if the FE one is appropriate. Standard errors need to adjust for any error correlation and, given that, more-efficient FGLS estimation is possible. In STATA, "*pa*" is specified, as the individual effects are assumed to be random and are averaged out. A deficiency of this estimator is the assumption of constant correlation ($\rho_{ts} = $ c) by using the "*exchangeable*" option, which may not be good given that the time intervals of repeated cross-sections in my data are not even. The other options of error correlation forms that available in STATA (e.g., "*independent*" "*AR (n)*") are not appropriate, and results from "*unstructured*" error correlation cannot achieve convergence, so I did not include them here. Compared to the FE coefficients in Table 4.1 and the RE ones in Table 4.2, the coefficients from the pooled estimators are close to those of *xtreg,* as *pa* with "*exchangeable*" is asymptotically equivalent to *xtreg, re* (Cameron & Trivedi, 2009)

Appendix C: Description of Long Panel Estimation

The *xtpcse* command in STATA is specifically designed for estimating panel-corrected standard errors in long panel models (Hoechle, 2007). The standard error estimates are robust to heteroskedasticity, contemporaneously cross-sectionally correlated, and autocorrelated to type *AR(1)* disturbances. *AR(1)* denotes that $\mu_{it} = \rho_i \mu_{i,t-1} + \varepsilon_{it}$, where ε_{it} are serially uncorrelated but are correlated over i with $Cor(\varepsilon_{it}, \varepsilon_{is}) = \sigma_{ts}$.) Beck and Katz (1995) demonstrate that the large *T*-based standard error performs well in correcting for contemporaneous correlation in small panels (the ratio of *T/N* is not small).

Just as is seen with *xtpcse*, the *xtgls* command also allows the presence of *AR(1)* autocorrelation within panels and cross-sectional correlation and heteroskedasticity across panels (Chen et al., 2010; StataCorp, 2005). This estimator fits panel-data linear models by using FGLS. It is commonly more efficient asymptotically than *xtpcse* (Reed & Ye, 2011; StataCorp, 2005).

The *xtregar* command in STATA estimates panel data regression when the disturbance term is *AR(1)*. It is a within estimator under the FE assumption and a GLS estimator under the RE assumption (StataCorp, 2005). Its advantage lies in its ability to fit to an unbalanced longitudinal dataset with observations unequally spaced over time (Baltagi & Wu, 1999). A limitation of *xtregar* is that it does not incorporate the White correction for heteroskedasticity.

Rather than restricting errors to be *AR(1)* in *xtpcse* and *xtgls*, the user-written *xtscc* command (Hoechle, 2011) applies the method proposed by Driscoll and Kraay (1998). It obtains *Newey-West* type standard errors that allow auto-correlated errors of a general form, which allows the error to be serially correlated for m lags.

In Table 4.6, Estimator I assumes that μ_{it} is heteroskedastic, meaning that each county has a different variance of $E(\mu_{it}^2) = \sigma_i^2$. With no correlation between or within panels, this estimator

provides a base scenario. Compared to the results derived from other estimators, the effect of farmland expansion is relatively small. Estimators II, *xtpcse*, performs a *Prais-Winsten* regression (StataCorp, 2005), which assumes *AR(1)* with the same ρ across the panel $\mu_{it} = \rho\mu_{i,t-1} + \varepsilon_{it}$. The estimates reveal a stronger association between farmland expansion and forestland loss.

Estimator III is a pooled OLS estimator with *Driscoll-Kraay* standard errors (Hoechle, 2011). The initial intention here was to see how the results vary with different autocorrelation lags. The calculated default maximum lag period is *3(m(T)=floor[4(T/100)^(2/9)])*. Because results changed little under the *AR(1), AR(2)* and *AR(3)*, I included the *AR(2)* case in the table by specifying the disturbance as heteroskedastic with cross-sectional correlation. Still, the results are not much improved from those derived by Estimator I. The problem is possibly attributable to the inappropriate use of pooled OLS estimation. Coefficients derived by Estimator IV are slightly better than those of Estimator II—the coefficient of farmland is larger and the NFPP turns out to be significant at the 95% level. Then, results derived with Estimator V show that different ρ_i computation methods affect both the parameter and standard error estimation, but the effects are not large here. Results derived by estimators VI and VII seem more realistic in terms of the estimated effect of farmland. Also, both the coefficients of NFPP and timber price become significant at the level of 1%. But a double check of the literature suggests that results from *xtgls* tend to produce smaller standard error estimates (Beck & Katz, 1995). So, it is good to be cautious with interpreting the standard error in the two regressions. Estimators VIII and IX perform FE and RE regressions with overall panel *AR(1)*. As the FE regression cancelled the county-specific FEs, the only two variables with significant coefficients are farmland and NFPP. Results of RE regression are similar to those derived by Estimator VII.

REFERENCES

REFERENCES

Altonji, J. G., & Segal, L. M. (1996). Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics, 14*(3), 353-366.

Andrews, M., Schank, T., & Upward, R. (2006). Practical fixed-effects estimation methods for the three-way error-components model. *Stata journal, 6*(4), 461-481.

Angelsen, A., & Kaimowitz, D. (1999). Rethinking the Causes of Deforestation: Lessons from Economic Models. *The World Bank Research Observer, 14*(1), 73-98.

Arellano, M. (1987). Practitioners' Corner: Computing Robust Standard Errors for Within-groups Estimators. *Oxford bulletin of Economics and Statistics, 49*(4), 431-434.

Baltagi, B. H. (2006). An Alternative Derivation of Mundlak's Fixed Effects Results Using System Estimation. *Econometric Theory, 22*(6), 1191-1194.

Baltagi, B. H., & Wu, P. X. (1999). Unequally spaced panel data regressions with AR (1) disturbances. *Econometric Theory, 15*(6), 814-823.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression.

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with Time-Series Cross-Section Data. *The American Political Science Review, 89*(3), 634-647.

Bhattarai, M., & Hammig, M. (2001). Institutions and the environmental Kuznets curve for deforestation: a crosscountry analysis for Latin America, Africa and Asia. *World Development, 29*(6), 995-1010.

Breusch, T. S. (1987). Maximum likelihood estimation of random effects models. *Journal of Econometrics, 36*(3), 383-389.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using stata* (Vol. 5): Stata Press College Station, TX.

Carr, D. L., Suter, L., & Barbieri, A. (2005). Population dynamics and tropical deforestation: State of the debate and conceptual challenges. *Population and environment, 27*(1), 89-113.

Chen, X., Lin, S., & Reed, W. R. (2010). A Monte Carlo evaluation of the efficiency of the PCSE estimator. *Applied Economics Letters, 17*(1), 7-10.

De Janvry, A., Fafchamps, M., & Sadoulet, E. (1991). Peasant household behaviour with missing markets: some paradoxes explained. *The Economic Journal*, 1400-1417.

Debarsy, N. (2012). The Mundlak approach in the spatial Durbin panel data model. *Spatial Economic Analysis, 7*(1), 109-131.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics, 80*(4), 549-560.

Geist, H.J., Lambin, E.F., (2001). What drives tropical deforestation? A meta-analysis of proximate and underlying causes of defores-tation based on subnational scale case study evidence. In: LUCC Report Series No. 4., University of Louvain, Louvain-la-Neuve

Geist, H. J., & Lambin, E. F. (2002). Proximate Causes and Underlying Driving Forces of Tropical Deforestation Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. *BioScience, 52*(2), 143-150.

Gould, W. (1996). Why isn't the calculation of R2 the same for areg and xtreg, fe? . from http://www.stata.com/support/faqs/statistics/areg-versus-xtreg-fe/

Gould, W. (2013). How can there be an intercept in the fixed-effects model estimated by xtreg, fe? . from http://www.stata.com/support/faqs/statistics/intercept-in-fixed-effects-model/

Guimaraes, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata journal, 10*(4), 628.

Hamilton, L. (2012). *Statistics with STATA: Version 12*. Boston: Cengage Learning.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric Society*, 1377-1398.

Hegre, H., & Sambanis, N. (2006). Sensitivity analysis of empirical results on civil war onset. *Journal of conflict resolution, 50*(4), 508-535.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata journal, 7*(3), 281.

Hoechle, D. (2011). XTSCC: Stata module to calculate robust standard errors for panels with cross-sectional dependence. https://ideas.repec.org/c/boc/bocode/s456787.html#cites

Jiang, X., Gong, P., Bostedt, G., & Xu, J. (2011). Impacts of Policy Measures on the Development of State-Owned Forests in Northeastern China: Theoretical Results and Empirical Evidence. *Environment for Development*(Discussion Paper Series).

Kaimowitz, D., Angelsen, A. (1998). *Economic Models of Tropical Deforestation. A Review*. Jakarta: Centre for International Forestry Research.

Key, N., & Runsten, D. (1999). Contract farming, smallholders, and rural development in Latin America: the organization of agroprocessing firms and the scale of outgrower production. *World Development, 27*(2), 381-401.

Koop, G., & Tole, L. (1999). Is there an environmental Kuznets curve for deforestation? *Journal of Development Economics, 58*(1), 231-244.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., . . . Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change, 11*(4), 261-269. doi: http://dx.doi.org/10.1016/S0959-3780(01)00007-3

McCaffrey, D. F., Lockwood, J., Mihaly, K., & Sass, T. R. (2010). A review of Stata routines for fixed effects estimation in normal linear models. *Unpublished manuscript*.

Mihaly, K., McCaffrey, D. F., Lockwood, J., & Sass, T. R. (2010). Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg. *Stata journal, 10*(1), 82.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, 69-85.

Nichols, A. (2008). FESE: Stata module to calculate standard errors for fixed effects. *Statistical Software Components*.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417-1426.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of financial studies, 22*(1), 435-480.

Raudenbush, S. W., Yang, M., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics, 9*(1), 141-157.

Reed, W. R., & Ye, H. (2011). Which panel data estimator should I use? *Applied Economics, 43*(8), 985-1000.

Schaffer, M. E. (2012). xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. *Statistical Software Components*.

StataCorp, L. (2005). *Stata base reference manual* (Vol. 2): Citeseer.

Stimson, J. A. (1985). Regression in space and time: A statistical essay. *American Journal of Political Science, 29*(4), 914-947.

Turner, B. L., Lambin, E. F., & Reenberg, A. (2008). Land Change Science Special Feature: The emergence of land change science for global environmental change and sustainability (vol 104, pg 20666, 2007). *Proceedings of the National Academy of Sciences of the United States of America, 105*(7), 2751-2751.

Wang, Z., Zhang, B., Zhang, S., Li, X., Liu, D., Song, K., . . . Duan, H. (2006). Changes of land use and of ecosystem service values in Sanjiang Plain, Northeast China. *Environmental Monitoring and Assessment, 112*(1-3), 69-91.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT press.

Xu, J., Tao, R., & Amacher, G. S. (2004). An empirical analysis of China's state-owned forests. *Forest Policy and Economics, 6*(3), 379-390.

Xu, J., Yin, R., Li, Z., & Liu, C. (2005). China's ecological rehabilitation: the unprecedented efforts and dramatic impacts of reforestation and slope protection in western China. *Ecological Economics, 57*(4), 595-607.

Xu, J., Yin, R., Li, Z., & Liu, C. (2006). China's ecological rehabilitation: Unprecedented efforts, dramatic impacts, and requisite policies. *Ecological Economics, 57*(4), 595-607.

Yin, R. (1998). Forestry and the environment in China: the current situation and strategic choices. *World Development, 26*(12), 2153-2167.

Yin, R., & Newman, D. H. (1996). The effect of catastrophic risk on forest investment decisions. *Journal of Environmental Economics and Management, 31*(2), 186-197.

Yin, R., Xu, J., & Li, Z. (2003). Building institutions for markets: Experiences and lessons from China's rural forest sector. *Environment, Development and Sustainability, 5*(3-4), 333-351.

Yin, R., & Yin, G. (2009). China's Ecological Restoration Programs: Initiation, Implementation, and Challenges *An Integrated Assessment of China's Ecological Restoration Programs* (pp. 1-19): Springer Netherlands.

Yin, R., & Yin, G. (2010). China's primary programs of terrestrial ecosystem restoration: initiation, implementation, and challenges. *Environmental Management, 45*(3), 429-441.

Yu, D., Zhou, L., Zhou, W., Ding, H., Wang, Q., Wang, Y., . . . Dai, L. (2011). Forest management in Northeast China: history, problems, and challenges. *Environmental Management, 48*(6), 1122-1135.

Zellner, A., & Theil, H. (1992). Three-stage least squares: Simultaneous estimation of simultaneous equations *Henri Theil's Contributions to Economics and Econometrics* (pp. 147-178): Springer.

Zhang, K., Hori, Y., Zhou, S., Michinaka, T., Hirano, Y., & Tachibana, S. (2011). Impact of Natural Forest Protection Program policies on forests in northeastern China. *Forestry Studies in China, 13*(3), 231-238.

Zhang, P., Shao, G., Zhao, G., Le Master, D. C., Parker, G. R., Dunning Jr, J. B., & Li, Q. (2000). China's forest policy for the 21st century. *Science, 288*(5474), 2135-2136.

CHAPTER 5

A SYSTEMATIC ANALYSIS OF LAND USE CHANGE DRIVERS

5.1 Introduction

Building upon what I have done in Chapter 4, this chapter attempts to achieve more rigorous results through systematic analysis of the driving forces of LUCC in northeast China. The emphasis of Chapter 4 was to explore the drivers of deforestation using conventional single-equation regression models and typical estimation techniques. However, my extensive work indicated that the single-equation models have some weaknesses. First, while it is reasonable to focus on the determinants of deforestation within a single-equation model, these determinants of deforestation are assumed to be exogenous (Mertens et al. 2000; Geoghegan et al. 2001; Schneider & Pontius 2001; Deininger & Minten 2002; Munroeaic et al. 2002; Pan et al. 2004; Franzese & Hays 2007; Song et al. 2008). However, the Mundlak model I have estimated shows that the mean value of farmland is correlated with the error term. Therefore, ignoring the potential issue that farmland expansion might not truly be exogenous and thus taking it as independent variables could cause biased estimation, which I will address here.

Endogeneity usually refers to situations where nonzero correlation exists between the error terms and observed explanatory variables in a model (Louviere et al. 2005; Chenhall & Moers 2007). This can lead to biased and inconsistent parameter estimates, making reliable inference impossible (Semykina & Wooldridge 2010). Endogeneity comes from various sources; the most common ones are omitted variables, measurement error, and simultaneity (Brownstone et al. 2002; Semykina & Wooldridge 2010). So, characterizing the endogenous land use changes is both necessary and desirable (Jöreskog & Sörbom 1986; Baltagi 2006; Fingleton & Gallo 2007). In my study region, the LUCC dynamics indicate that potential endogeneity could arise from: (1) simultaneity that is intrinsic in the land-use conversions; (2) spatial dependences of LUCC between

different classes of land use; and/or (3) indirect or spillover effects induced by other land-use changes.

Simultaneity arises when one or more of the explanatory variables are jointly determined with the dependent variable, usually through an equilibrium mechanism (Baltagi 1981; Zellner & Theil 1992). A frequently used example is the "supply and demand" relationships in determining the price and quantity of "beef or pork," in which the consumer's demand for beef is not just affected by the price of beef itself, but also by the price of a substitutive good, such as pork (Epple 1987; Angrist & Krueger 2001). Models of this sort are known as simultaneous-equations models (SEMs), which are an important class of empirical models in economics (Wooldridge 2010, 2012). For an equation system to be viewed as an SEM, at least one of the right-hand-side variables in one of the equations should be endogenous and thus correlated with the error term.

Simultaneity is also embedded in LUCC conversion. In my study region, farmland expansion comes at the expense of loss of forestland as well as wetland. Numerous studies have documented the encroachment of agriculture on wetland (Liu et al. 2004; Wang et al. 2006; Zhang et al. 2010; Wang et al. 2011). As an important food basket of China, Heilongjiang has experienced a rapid expansion of rice growth due to the higher yield and better quality of rice there (Jiang et al. 2006; Sun et al. 2010). Meanwhile, the acreage of other crops has declined substantially. For instance, the statistics Zhou et al. (2009) calculated, based on 15 farms surrounding the Honghe Natural Reserve in the Sanjiang Plain, suggest that the rice fields there increased from about 200 km^2 in 1993 to more than 2000 km^2 in 1998. By 2002, the overall area of crop fields had reached 3,781 km^2, of which rice accounted for 2,024 km^2. So, when characterizing the relationship of farmland demand and supply, agricultural growth is a primary factor on the demand side, whereas forestland and wetland ("Others" in my classification scheme) are two of the main variables on the

supply side. Also, since wetland is an alternative source of for farmland expansion, it could be a substitute for forestland. Therefore, it is worthwhile and beneficial to adopt a more integrated framework to identify the indirect linkages between wetland and forestland, as well as the direct linkages between farmland and the other two classes of land use.

Moreover, there can be spatial dependence, which tends to be multi-directional: each location serves as a "neighbor" for its nearby locations, and according to the Tobler's law, things that are located closer to each other tend to be and behave more alike (Tobler 1970). Thus, land use conversion between two neighboring classes is two-directional; each type of land use can be an intruder on the other. In this hedonic case, the "strict exogeneity" assumption fails. Instead, it is more appropriate to hypothesize that changes in farmland and forestland are endogenous (Anselin 2003). From the land conversion matrixes derived in Chapter 2, however, we see that forestland generally did not take over farmland. So, it is natural to hypothesize that farmland is endogenous to forestland.

Endogeneity and the potentially biased estimation when it is ignored are well accounted for in econometrics, despite the slow progress of adopting the idea and procedure of endogeneity testing and correction in analyzing the forces driving LUCC. Examples of endogeneity testing of driving forces in land-use studies are particularly limited before 2000s (Irwin & Geoghegan 2001a; Lambin et al. 2001b; Verburg et al. 2004a). Lambin et al. (2001) reviewed some of the recent models of spatial land-use changes and affirmed the contribution of structural economic models in addressing spatial dependency and endogeneity. Verburg et al. (2004) conducted a thorough review of land-use models and related concepts regarding the forces driving changes in land use, and pointed out that road development, population change and production prices could be endogenous under certain circumstances. Following a discussion of advances in understanding the

causes and consequences of land conversion, Irwin and Geoghegan (2001) built a system of interactive equations for population migration and government expenditures and revenues. Then, they illustrated a decision framework for land use conversion, showing how to estimate the implicit residential land value with a spatially explicit hedonic pricing model.

Studies linking LUCC to socioeconomic factors with recognition and careful handling of endogenous variables are still rare in literature (Chomitz & Gray 1996; Pfaff 1999a; Mertens & Lambin 2000; Herbert & Arild 2009; Yin & Xiang 2010). Chomitz & Gray (1996), Pfaff (1999) and Mertens & Lambin (2000) developed land-use models by starting with land allocation according to the rule of maximizing expected profits. They perceived potential endogeneity problems when selecting variables that are included in the land-use conversion model. Chomitz & Gray (1996) found that road development suffers endogeneity as the siting of roads is affected by agricultural production. Pfaff (1999) examined the possible endogeneity problem associated between population change and forest conversion. He argued that population may be endogenous, or it may be collinear with government policies that encourage development of targeted areas.

Per a suggestion by Chomitz and Gray (1996), Mertens & Lambin (2000) developed their modeling approach by introducing a variable to measure the suitability of land for agriculture to reduce the endogenous bias. Herbert & Arild (2009) further suspected that indicators, like plot area, land under bush fallow, farm-related assets, and number of livestock are endogenous variables. They applied the three-stages least squares method to control for potential unobserved heterogeneity and simultaneity. Yin and Xiang (2010) developed a structural model with four equations featuring the multiple dimensions of agriculture (cropland use, grain production, soil erosion and technical change); by solving this system of equations, the interactions and feedback

142

of cropland change dynamics were clearly validated within the complex human and natural connections.

In sum, the complex land-use system being examined in this study calls for a more sophisticated modelling strategy. A fundamental problem of the single-equation regression models lies in their failure to capture the underlying interactions between drivers of different classes and processes of land use. Meanwhile, when we consider the complex relationships of a land-use system, the assumption of consistent OLS estimation—where the error term is unrelated to any of the regressors—may become no longer valid because of potential endogeneity (Semykina and Wooldridge 2010).

5.2 Model Specification

There are two ways to estimate a model consistently with the endogeneity issues—single-equation estimation with instrumental variables (IV) and system of equations estimation. Single-equation estimation, by definition, involves one equation of main interest, while it considers an endogenous variable to be determined by a set of exogenous variables in a "side" equation (Angrist et al. 1996). In other words, these exogenous variables are used to identify the effects of an endogenous variable in the main equation. The exogenous variables in the side equation are called the instrumental variables for the identification. In the first stage regression, thus, all the exogenous variables in main equation and side equation(s) are taken as explanatory regressors for the endogenous variable. To distinguish the exogenous variables in the main equation from those in the side equation, the instruments in the side equation are called excluded instrument variables while the instruments in the main equation are called included instrument variables. So, a single-

equation estimation, when endogeneity appears, is oftentimes viewed as a simultaneous system with jointly determined dependent variable *Y* and endogenous variable *X(s)* (Wooldridge 2002).

Compared to single-equation estimation with one endogenous variable, system of equations estimation involves estimating a set of equations in which one or more explanatory variables are jointly determined with the dependent variables. So, the conventional regressors that appear only on the right hand side of an equation can also have their own equation(s). Equations in the system that contains endogenous variables are usually referred as structural equations. Structural equations cannot be directly estimated. Using algebra, the endogenous variables could be expressed as functions of only exogenous regressors on the right hand side, leading to an equation in reduced form. As the error term in one equation is likely to be contemporaneously correlated with the error terms in other equations of the system, estimating the system of equations jointly captures the interactions of underlying causes and improves the estimation efficiency from cross-equation coefficient restrictions and correlations (Zellner & Theil 1992; Wooldridge 1996).

In the following two subsections, I will first define a single equation with instrumental variables to examine linkages between the two dominate land-use classes—forestland and farmland. Then, I will specify a system of equations to depict the LUCC relationships when wetland is considered as well. For both models, the detailed steps of estimation will be elaborated.

5.2.1 Analysis of the Two Dominant Land-Use Classes: An Instrumental Variable Method

The single-equation models in Chapter 4 have already included variables for the most relevant forces driving changes in forest cover that are frequently used in the literature: timber price (*Tp*), gross output value of forestry (*O*), dummy variables capturing the effect of implementing the NFPP (*N*), distance between a forest farm and its closest timber market and county seat (*D*), number of forest farms in each county (*Nf*), and the local population (*P*).

From a land-use perspective, agricultural expansion is the extension of cultivation into previously uncultivated areas. This process may require increased inputs, including 1) increased labor use for land conversion (e.g. construction of swamp drainage and irrigation channels) and cultivation, 2) increased spending on purchasing production materials, and 3) capital investment in technical capacity that can raise land productivity (De Janvry et al. 1991; Grossman & Helpman 1993; Färe et al. 1994; Kalirajan et al. 1996). In practice, the relative feasibility of these factors is likely to vary in different places. Meanwhile, farmland expansion is often driven by an increased demand for food products, which is partly reflected in the prices of agricultural products.

The above-mentioned inputs seems to be relevant candidate instruments for the potential endogenous variable "farmland": number of agricultural laborers (L), per capita annual net income (C) (as the potential expenditure of farming), and total agricultural machinery power (T) (a proxy for technological development). I will also incorporate the price index of agricultural products (AP) to reflect market demand relative to supply. Built-up area (B) is included as a determinant of farmland growth based on the assumption that more settlement leads to greater agricultural expansion.

With farmland expansion encroaching upon forestland, the equation of farmland use is linked to the equation of forestland as follows:

$$Ft_{it} = f(Fm_{it}, Tp_t, O_{it}, N_t, D_i, Nf_i, P_{it}) + u_{it} \qquad \text{(Eq.5.1)}$$

$$Fm_{it} \sim f(L_{it}, C_{it}, T_{it}, AP_t, B_{it}) + v_{it} \qquad \text{(Eq.5.2)}$$

In both equations, i denotes county; if i is not present in a variable, it means that county-level data are not available and provincial data are used instead. Similarly, t denotes time; if a variable, such as distance to markets, does not vary with time, t is absent from the variable's subscript. In Eq.5.1, forestland is a function of the right-hand-side variables that are independent, except for farmland.

Farmland, on the other, is assumed to be endogenous and instrumented with a set of selected variables on the right side of Eq.5.2.

Figure 5.1 The relationship between the two major land-use classes
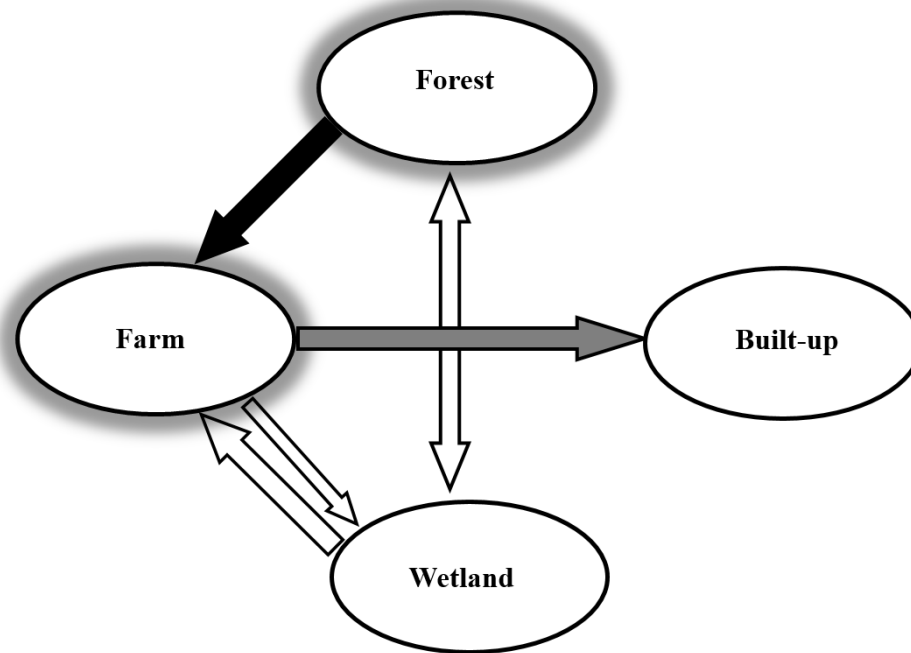
Note: The dominant conversion is from forestland to farmland. Built-up land doesn't interact with forestland directly, so it is taken as an instrument candidate for the expansion of farmland.

Figure 5.1 above depicts this relationship. In addition to this major linkage of the LUCC dynamics, considerable conversion of farmland to built-up area is also involved. With built-up area being an exogenous variable, the strong correlation between farmland and built-up area makes built-up area an important instrument candidate in Eq.5.2. The error term, u_{it} represents the effects of the omitted variables that are peculiar to both the individual units and time periods. Under the fixed-effect assumption, u_{it} is a combination of an independently identically distributed (*i.i.d.*) random error ε_{it} and an unobserved heterogeneity α_i peculiar to county i over time (Hausman & Taylor 1981; Nickell 1981).

The instrumental variables method (IV) is used as follows. The potentially endogenous variable (farmland in this case) is first regressed on the excluded instrumental variables in Eq.5.2 as well as all the exogenous variables in Eq.5.1. Given the least squares regime, this first-stage regression produces an optimal linear combination of exogenous variables. Then, the predicted values of farmland are used as the independent variable in Eq.5.1 in the second stage regression (Wooldridge 2002; Murray 2006). Therefore, this procedure is also called the two-stage least squares, or 2SLS (Wooldridge 2002). The 2SLS regression, coupled with a fixed-effect estimator, controls for not only the endogeneity in farmland but also unobserved heterogeneity. However, this procedure does not account for the potential simultaneity among different classes of land use.

5.2.2 A More Integrated System of Land Use: Simultaneous Equations Modelling

To disentangle the direct and indirect effects of LUCC and eliminate the potential endogeneity, I will further analyze the LUCC processes by developing and estimating a simultaneous equations model. For the three closely interrelated categories of land use—forestland, farmland, and wetland, I can specify a system of three equations to describe their behavior and reflect their interaction. For simplicity, I have decided to name them the deforestation equation, the farmland expansion equation, and the wetland loss equation, respectively. Meanwhile, built-up land comes from converting farmland, but after it is built up it will no longer be converted into any other type of land use. Built-up area can thus be viewed as an external factor that affects the "forestland-farmland-wetland" system one way or another. This theoretical consideration is indeed confirmed by my empirical evidence from the identification tests (see the section of 5.3.1).

Similar to the analytic system of the two dominant classes of land use specified above, the deforestation equation in the SEM is defined on the basis of the existing literature investigating its driving forces. In the farmland expansion equation, I will deliberately include the full set of

147

explanatory variables in Eq.5.2. As noted earlier, wetland is one of the targets of agricultural expansion, and it also serves as a substitute for forestland in farmland demand. Thus, the status of wetland is connected to the dynamics of farmland and forestland.

Agricultural production in the region used to be comprised mostly of water-saving crops such as wheat, corn, and soybeans, but it has gradually shifted to paddy rice (Yun et al. 2005). The rapid increase in paddy rice fields has greatly propelled water demand in the Sanjiang Plain—pumping groundwater for irrigation; this has in turn led to a continual decline of groundwater level (Zhang et al. 2009). Local farmers' establishment of extensive irrigation networks has thus accelerated the wetland loss: reservoir construction disturbs the local natural waterways, and the corresponding expansion of dams, canals, and dikes also cut off the wetlands' water supply from nearby rivers or lakes (Zhou et al. 2009). As such, I will use the effective irrigation area to approximate the aggregate water use for irrigation. Natural factors, such as climate change, may also affect the status of wetland. For example, a warming climate and decreasing precipitation could possibly result in wetland reduction in the long run (Yan et al. 2001; Yan et al. 2002; Song et al. 2008; Zhang et al. 2010).

Based on the above discussion, I can define wetland loss (*Wt*) as being associated with farmland expansion (*Fm*), forest-cover change (*Ft*), human water withdrawal and reservoir construction (*I*), and climate change as reflected in decreased precipitation (*Pr*) and increased temperature (*T*). This leads to Eq.5.5 below.

$$Ft_{it} = f(Fm_{it}, Wt_{it}, Tp_t, O_{it}, N_t, D_i, Nf_i, P_{it}) + u_{it} \tag{Eq.5.3}$$

$$Fm_{it} = f(L_{it}, C_{it}, T_{it}, AP_t, B_{it}) + \varphi_{it} \tag{Eq.5.4}$$

$$Wt_{it} = f(Fm_{it}, Ft_{it}, I_t, Pr_t, Te_t) + \omega_{it} \tag{Eq.5.5}$$

The land conversion dynamics underlying the above specification are illustrated in Figure 5.2, with the dark arrows indicating the linkages among the three classes of land use embodied in Eq. 5.3-5.5. Eq.5.3 and 5.4 are similar to Eq.5.1 and 5.2 for the two dominant classes of land use, but an important distinction is that farmland change is instrumented with a set of candidate variables in Eq. 5.2, whereas those variables are now treated as regular regressors in Eq.5.4.

Compared to a single-equation model, a system of equations estimated with panel data has an even shorter intellectual history (Biørn 2004). A general strategy in adopting the three-equation system is to combine the features of simultaneous equations while allowing for possible interaction between some of the dependent variables. The three-stage least squares procedure (3SLS) exactly fulfils these two important objectives. It combines insights from instrumental variable and GLS methods to achieve consistency and efficiency through appropriate weighting in the variance-covariance matrix (Wooldridge 1996; Baltagi & Liu 2009).

Figure 5.2 A driving force analysis of the "Forest-Farm-Wetland" system
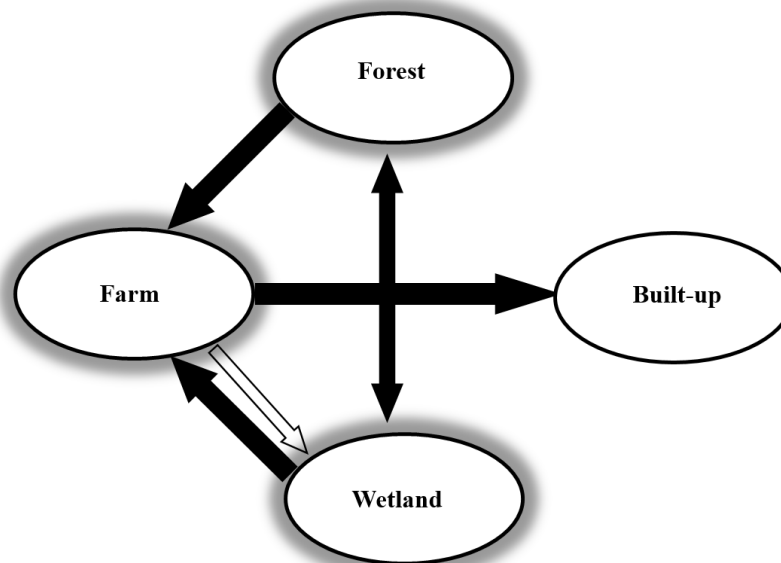
The 3SLS procedure consists of the following steps. First, convert the structural equations containing endogenous explanatory variables into reduced form equations, in which only exogenous variables appear on the right-hand side, and then estimate the reduced-form equations by OLS to obtain fitted values for the endogenous variables. Second, estimate the structural equation through 2SLS by replacing the endogenous regressors with their fitted values derived in step one and retrieve the covariance matrix of the equations disturbances. Finally, perform a GLS-type estimation on the stacked system using the covariance matrix from the first step (Cornwell et al. 1992; Wooldridge 1996).

Before proceeding, it is necessary to verify whether the order condition for identification is satisfied. That condition for an equation requires that the number of excluded exogenous variables (See the model specification part for "excluded instrument variables") is at least as many as the number of included right-hand-side endogenous variables (Baumol & Hall 1977; Engle & Kroner 1995). It is easy to see that each equation in the "Forest-Farm-Wetland" system contains more than three exogenous variables—6 in Eq.5.3, 5 in Eq.5.4 and 3 in Eq.5.5. On the other hand, the maximum number of endogenous variables is 2 in Eq. 5.3 and Eq. 5.5. Therefore, the order condition is satisfied.

5.3 Data and Variables

Table 5.1 below presents a general description of all the variables. The variables in bold are the three land-use classes (forestland, farmland, and wetland), which are taken as endogenous, and thus have their own explanatory variables. My panel data in this study span 31 years and 8 counties. Recall that the original LUCC data were derived from six periods of time (1976, 1984, 1993, 2000, 2004, and 2007) and they were then interpolated to obtain annual observations. In

Table 5.1, column 1 lists the variables with their corresponding name abbreviations; the full name of each variable is given in column 2 and their units in column 3; and columns 4-7 summarize their basic statistic values. Details regarding the data sources of the variables and potential concerns about them are discussed below.

Table 5.1 Summary data description

Variable Definition	Abbreviation	Unit	Mean	Std	Min	Max
Forest Area	Forest (*Ft*)	Km^2	1194.52	901.92	5.13	2622.70
Price Index of Timber	TimberPrice (*Tp*)	1976=100	88.90	23.46	54.50	161.60
Gross Output Value of Forestry	ForOpt (*O*)	1,000 ¥	4538.87	5165.00	164.99	33424.47
Mean Distance to Nearby Large Markets	Meandist (*D*)	Km	26.10	9.57	15.96	46.56
Number of Forest Farm in County	NForFarm (*Nf*)	None	6.38	4.04	1.00	13.00
0 before 2000; otherwise 1	NFPP (*N*)	None	0.30	0.46	0.00	1.00
Total Population	TotalPop (*P*)	1,000 P	305.76	99.79	104.00	527.50
Farm Area	Farm (*Fm*)	Km^2	1773.47	799.59	206.25	2876.01
Built-up Area	Builtup (*B*)	Km^2	92.63	55.50	12.38	243.04
Number of Agricultural Laborers	Aglabor (*L*)	1,000 L	52.15	29.23	11.40	146.04
Per Capita Annual Net Income of Rural Population	IncmRurPop (*C*)	Yuan	312.06	192.39	36.04	920.31
Agricultural Machinery Power	AgMachPowr (*T*)	1000 kWh	137.73	68.08	27.21	417.80
Price Index of Agricultural Products	AgPrice (*Ap*)	1976=100	344.00	170.34	100.00	578.04
Wetland	Wetland (*Wt*)	Km^2	173.59	231.38	2.04	1033.88
Farm Area	Farm (*Fa*)	Km^2	1773.47	799.59	206.25	2876.01
Forest Area	Forest (*Fo*)	Km^2	1194.52	901.92	5.13	2622.70
Irrigation Area in Heilongjiang	IrrigatArea (*I*)	Km^2	131.26	70.33	60.50	295.00
Average Annual Total Precipitation	Precip (*Pr*)	Mm	524.01	70.85	383.49	657.59
Average Annual Temperature	AveTemp (*Te*)	0.1 °C	30.34	7.44	17.06	46.50

Note: "¥"is the unit of Chinese currency (Yuan). The unit for "Total Population" is 1000 persons and the unit of "Number of Agricultural Laborers" is 1000. The unit of "kWh" in the "Agricultural Machinery Power" stands for kilowatt hour, and "mm" in "Average Annual Total Precipitation" stands as millimeter.

Variables Used in the Deforestation Equation

Again, *NFPP (N)* is a discrete dummy variable which takes value 0 before 2000 and 1 otherwise, indicating that the forest protection program is "on" after 2000. *Timber price (TP)* data came from Forest Industry Bureau of Heilongjiang with a unit of yuan/m^3. The real price series were obtained by deflating the nominal prices with the provincial-level Consumer Price Index (or CPI, with a base year of 1976) (Heilongjiang Statistical Bureau 1986-2008). The *number of forest farms (Nf)* in each county is included to explore the institutional effect based on the assumption that with more government owned forests being located in a county, there would be less illegal logging and thus less deforestation. As local *population growth (P)* increased and spread, more farmland was converted into built-up areas and clearing forests for farming became inevitable in order to increase local farm production and meet the demands of a larger population. Also, population growth is closely linked to rising consumption of wood products and fuelwood. *Mean distance (D)* measures the average distance from a forest farm to nearby capitals and timber markets.

Agricultural-Expansion-Related Variables

Agricultural labor (L) is a proxy for labor use in farmland. Data on agricultural laborers came from the *Heilongjiang Statistical Yearbook* (Heilongjiang Statistical Bureau 1986-2008) and the area of farmland is derived from my land-use classification results. *Per capita annual income (C)* of a rural population connects agricultural production to the local economy. As rural people gradually began participating in non-agricultural activities, a question was whether the local farmers would invest their income in increased agricultural production by purchasing commercial inputs. If they did so, the relationship between their income and farmland area should be positive; however, if local farmers had enough access to other business activities, such as commerce and

services, there would be less desire for agricultural expansion, in which case the relationship between rural income and farmland expansion would be negative.

Agricultural machinery power (T) is a main indication of the technological sophistication of agricultural production. The agricultural machinery power of each county is documented in its statistical yearbook. A concern is whether this variable is representative of local agricultural technology adoption, because technological improvement could be embedded in various inputs, such as better seeds, more fertilizer and pesticide use, and adoption of more effective methods of cultivation. Unfortunately, I could not find any statistics to capture these phenomena. Of course, even if machinery is an appropriate indicator for farming technology, a large machinery use does not guarantee a high technological efficiency.

Data on *price index of agricultural products (Ap)* were collected from the Heilongjiang Price Annals (volume 42) for the period of 1976-1985 (Compilation Committee of Heilongjiang Annals 1993) and Heilongjiang Statistical Yearbook for the period of 1986-2007 (Heilongjiang Statistical Bureau 1986-2008). After the dual-track pricing system was introduced in 1985 (Qian 2000), agricultural product prices gradually went up. Prices reached their peak in 1996 and 1997, partly caused by the high levels of countrywide inflation in 1994 (Wang 2008).

Wetland-Loss-Related Variables

Irrigation area (I) data were taken from the publication "Sixty Years of Heilongjiang" (Heilongjiang Statistics Bureau 2009). This variable is an important indicator for agricultural water consumption; along with increasing local rice production, the effective irrigation area increased rapidly. *Precipitation (Pr)* and *temperature (Te)* were the annual averages over the 13 meteorological stations in Heilongjiang, which were acquired from the website of the China Meteorological Data Sharing Service System（National Meteorological Information Center，

154

2009). Yan et al. (2002) pointed out that in the Sanjiang Plain, the annual average temperature rose from 1.2°C to 2.3°C from 1955 to 1999. The average temperature during the period of 1976-2007 trended upwards from 1.71 °C in 1977 to 4.65 °C in 2007. Zhou et al. (2009) also confirmed the decreasing precipitation trend with data from the Jiansanjiang Weather Station during 1957 to 2000. Therefore, I assume that in addition to the human drivers, natural factors like decreased precipitation and warming temperatures have also contributed to wetland loss.

5.4 Estimated Results

5.4.1 Two Dominant Classes of Land Use

Model Validation

As a preliminary step, it is necessary to validate the selected instruments and the goodness of fit of first-stage regression. Table 5.2 reports my testing results in terms of under-identification, weak identification, and weak-instrument-robust inference. Four diagnostic tests are conducted in the second-stage: endogeneity test, under-identification test, weak identification test, and over-identification test. The statistics for the under-identification and weak identification tests are the same as those in the first stage, while the endogeneity and over-identification tests are specific to the second stage (see Appendix A for more detail).

Table 5.2 1st and 2nd stage test results of instrumental variable analysis

Tests	Statistics	All IV	No B	No AP	No C	No L	No T	No T or AP	Only B
Under-Identification	SW χ^2	86.56	79.53	82.04	78.62	83.36	60.54	57.88	37.92
	P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KP χ^2	42.65	41.63	39.29	40.88	42.61	38.27	35.96	31.60
	P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Weak-Identification	CD F	22.60	21.70	26.86	21.79	27.92	22.41	29.64	50.89
	KP F	16.67	19.22	19.83	19.00	20.15	14.63	18.73	37.13
Weak-instrument-robust inference	AR F	30.63	23.01	38.08	31.70	35.56	28.61	37.65	82.36
	P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AR χ^2	159.11	95.21	157.59	131.18	147.14	118.39	116.34	84.12
	P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	SW χ^2	61.76	50.02	61.27	57.95	61.26	59.87	59.05	56.51
	P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Endogeneity	Ed-test	6.78	6.15	33.95	7.56	7.28	12.16	35.57	46.27
	P-value	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00
Over-Identification	Hsen J	33.95	13.43	8.88	22.58	22.73	17.84	7.41	
	P-value	0.00	0.00	0.03	0.00	0.00	0.00	0.02	

Note: (1) B represents Built-up Area, AP Agricultural Price Index, C Per Capita Net Income, L Average Laborers per Unit Farmland, and T Agricultural Machinery per Unit Farmland. (2) SW χ^2 indicates *Sanderson-Windmeijer* χ^2 statistic; KP χ^2 *Kleibergen-Paap rk LM* χ^2 statistics; CD F *Cragg-Donald (CD) Wald F* statistic; KP F *Kleibergen-Paap Wald F* statistic; AR F *Anderson-Rubin (AR) Wald F* statistics; AR χ^2 *Anderson-Rubin (AR) Wald* test; Ed-test endogeneity test of endogenous regressor; Hsen J: *Hansen J* statistic.

When all instruments were included, they passed all the tests except the Hansen J test (Pitt 2011), which rejected the assumption that "all the excluded instruments are valid" (see Appendix A), casting doubt over the validity of this instrument combination. To make sure that only the exogenous instrumental variables are included, I took a further step to try different instrument combinations and recorded the corresponding test statistics (see Table 5.2). However, all of the over-identification test results still could not eliminate of the doubt over the validity of these instruments. This model validation process continued until I found that the variable built-up area fits as an instrument.

It is known that built-up area includes the areas that have been most intensely changed by human activities, such as cities, towns, villages, and road networks. My classification results suggest that both built-up area and farmland experienced an expansion, but the built-up area does not necessarily encroach onto forestland. These relationships perfectly satisfy the requirements of a suitable instrument variable. Moreover, the existing literature confirms a strong correlation between settlement and road development on the one hand and agricultural land expansion on the other. Thus, built-up area can serve as a good instrument for farmland change. Subsequently, my statistical testing results nicely validated this assertion.

The endogenous test strongly rejected the null hypothesis of exogeneity while the under-identification power did not lose much strength by only keeping one instrument in the model. The weak-identification statistics also outperformed the previous tests based on various combinations of instruments. These results consistently point to the choice of built-up area as an instrument for farmland and, therefore, I dropped all the other instrument candidates.

Modelling Results from the System of Two Dominant Classes

Results reported in Table 5.2 are based on the system of two dominant land-use classes, with the endogenous variable farmland being replaced by built-up area. Models I-VI were estimated using different FE estimators. The 2SLS is the most widely used IV estimator (Model I), but it is also known to likely cause substantial bias in over-identified models, and especially when the first stage partial R^2 is low (Bound et al. 1995). The Limited Information Maximum Likelihood (LIML) estimator naturally comes as a remedy for this problem (Staiger & Stock 1994) (Model II), and is believed to outperform both the 2SLS or the GMM estimators in finite samples (Murray 2006; Cameron & Trivedi 2009). However, Morimune (1983) pointed out that the LIML has the potential problem of considerable large dispersion in the estimates.

Subsequently, Bekker and Ploeg (2005) and Hausman et al. (2007) argued that the LIML is inconsistent with the presence of heteroskedasticity when the number of instruments is large. The continuous updating estimator (Model III) which is GMM-like generalization of the LIML, could tackle possible heteroskedastic and auto-correlated disturbances but still has the moment problem and exhibits wide dispersion (Hausman et al. 2007). On the other hand, the widely applied GMM estimation methods have the virtue of avoiding unnecessary structure assumptions in the data generating process, and thus the specification of a particular distribution of the error terms (Model IV and Model V). Compared to the one-step GMM estimators which use weight matrices that are independent of estimated parameters, the two-step GMM constructs a weighting matrix with a consistent estimate of the parameters in its first step (Windmeijer 2005). The two-step efficient GMM estimator in Model IV is robust to arbitrary heteroskedasticity while Model V implemented the kernel-based heteroskedasticity and autocorrelation consistent (HAC) covariance matrix. Still, like the 2SLS, the GMM procedures have a finite sample bias. Thus in Model VI, I

bootstrapped 400 replications by clustering on the unit of "County" with the HAC covariance

matrix, Model VI is robust to arbitrary heteroskedasticity and intra-group correlations.

Forestland	(I) IV	(II) IV_limlr	(III) IV_cuer	(IV) IV_gmm2sr	(V) IV_hacr	(VI) IV_bscr	(VII) IV_re	(VIII) IV_ec2sls	(IX) IV_nosa	(X) IV_be
Farm(Fm)	-1.47***	-1.47***	-1.47***	-1.47***	-1.47***	-1.47*	-1.46***	-1.34***	-1.43***	-0.09
	(0.10)	(0.09)	(0.09)	(0.09)	(0.13)	(0.78)	(0.11)	(0.09)	(0.14)	(0.27)
TbPrice(Tp)	1.15***	1.15***	1.15***	1.15***	1.15***	1.15	1.13***	0.85**	1.08**	
	(0.36)	(0.31)	(0.31)	(0.31)	(0.43)	(0.81)	(0.41)	(0.35)	(0.52)	
ForOpt(O)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.05)
NFPP(N)	40.15**	40.15***	40.15***	40.15***	40.15**	40.15	39.60**	33.04**	38.05	
	(16.45)	(12.81)	(12.81)	(12.81)	(19.80)	(41.87)	(18.65)	(16.29)	(23.93)	
TotalPop(P)	-0.73***	-0.73***	-0.73***	-0.73***	-0.73***	-0.73	-0.74***	-0.69***	-0.75***	-2.69
	(0.14)	(0.09)	(0.09)	(0.09)	(0.13)	(0.87)	(0.16)	(0.14)	(0.20)	(1.10)
Meandist(D)							101.66***	93.51***	99.55***	4.05
							(11.17)	(9.55)	(10.53)	(20.52)
NFtFarm(Nf)							347.02***	335.76***	344.44***	175.36*
							(22.83)	(19.82)	(18.22)	(47.78)
Constant	3,913.86***						-968.55***	-890.81***	-942.73***	381.94
	(164.47)						(320.12)	(280.61)	(227.20)	(481.21)
R^2		0.77	0.77	0.77	0.77	0.77				

Note: TbPrice = TimberPrice, and NFrFarm = NForFarm; standard errors are in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively. As most 2SLS modelling tests were based on the degrees of freedom, with the same variable and data set, the modelling testing results are close, please refer to the final column of Table 5.2 for testing information.

Models VII-IX apply RE estimators in conjunction with the IV method. Model VII was estimated by default with the G2SLS RE estimator, Model VIII was based on Baltagi's EC2SLS RE estimator, and Model IX used the Baltagi-Chang estimators for the variance components. The G2SLS and EC2SLS estimators differ in how they construct the GLS instruments. The traditional G2SLS estimator passes each exogenous variable in X through the feasible GLS transformation (See Eq.3.4 and 3.5 in Chapter 3), while Baltagi's EC2SLS spans the set of instruments used by including the group means of each variable X_{it}. Baltagi and Liu (2009) argued that the extra instruments in EC2SLS can lead to efficiency gains in small samples. Model VII and Model VIII used the default adapted Swamy-Arora estimators (Swamy & Arora 1972) when computing the variance components, while Model IX employed the Baltagi-Chang estimators. The difference between these two methods is that the Swamy-Arora estimator considers degree-of-freedom corrections which are supposed to improve the model performance for small samples. Given the two different model and variance estimators, we can see in Table 5.3 that the magnitude of coefficients and standard errors in Model VIII, based on the EC2SLS estimator and default Swamy-Arora variance estimator, are all smaller relative to the default G2SLS estimator in Model VII. The coefficients of Model IX generally lie between those of Model VII and Model VIII, but the standard errors fall outside of the corresponding range of Model VII and Model VIII due to no degree adjustment in its variance estimator.

The ultimate goal of including so many estimators in the fixed-effects IV analysis was to get the most robust estimation. In case all the four candidate instruments are included for one endogenous regressor, these estimators report results with more variations. In the just-identified fixed-effect analysis, with the endogenous regressor being instrumented by one variable, the 2SLS is equivalent to the IV method. Meanwhile, all these models were required to report results that

are at least robust to heteroskedasticities, making the estimation differences under different estimators small and negligible.

The RE models in Table 5.3 are meant to offer insights complimentary to the system of two classes of land-use. The correlations between forestland change and the two time-invariant variables—the mean distance to nearby cities and timber markets and the number of forest farms located within the same county—are dropped in FE analysis. These two drivers apparently play important roles in influencing forestland change. The highly significant coefficients of "Meandist" suggest that forest farms located farther away from timber markets and large cities tend to suffer less deforestation. Also, with more forest farms clustered in same county, the forestland tends to be better protected. These additional findings generated by the random-effect analysis are useful for understanding the driving forces of deforestation and their interaction. Further, the seemingly non-significant between-effect derived from Model 10, where the regressors explain little of the variance in the dependent variable, actually confirms that changes of regressors between counties are small, validating the appropriateness of choosing FE (or within-effects) estimators in this analysis of two land-use classes.

Generally speaking, the signs and magnitudes of the 2SLS coefficients outperform considerably those from the single-equation regressions in Chapter 4. Specifically, farmland use is strongly correlated with forestland change, with a coefficient of -1.47—larger than that derived from the FE OLS estimators. The dummy variable for the NFPP is now significant, suggesting a positive effect on forestland protection. Also, the effect of population change is consistent with the general finding that deforestation occurs under human pressure in developing countries. Meanwhile, the coefficient of timber price is positive, which seems counterintuitive.

5.4.2 A More Systematic Analysis of Land Use Driving Forces

Various model validation routines are presented in the Appendix B. Table 5.4 in the next page presents the 3SLS estimates for the "forest-farm-wetland" system. The first column is the variables and their notations specified in section 5.2.2. The second column lists the to-be-checked hypothesis (sign of the coefficient). The estimated results are listed in the last three columns of the table.

The coefficient estimates of the deforestation equation are generally consistent with those of the two land-use classes. The statistically significant coefficient of "Farm" indicates that farmland expansion has a strong and negative correlation with forestland (-1.40). The area of wetland is also negatively correlated (-0.39) to the area of forestland, attributable to their mutual substitution in farmland expansion. The negative coefficient of population change shows that the increasing population could have put pressure on forest resource extraction, leading to more forestland losses. On the other hand, timber price and the NFPP are positively correlated with forestland change. It is easy to interpret the positive policy effect—the NFPP has played a role in protecting local forests. While the positive effect of timber price seems counterintuitive, it is possible that the forest cover will expand, partially in response to higher timber prices over the long run.

Table 5.4 Results of 3SLS analysis of the "Farmland-Forestland-Wetland" system

VARIABLES	Expected Sign	(1) Forestland	(2) Farmland	(3) Wetland
Farmland	-	-1.40***		-0.24***
		(0.03)		(0.07)
Wetland	-	-0.39***		
		(0.12)		
Price Index of Timber	-	0.40**		
		(0.17)		
Population	-	-0.25***		
		(0.07)		
NFPP	+	25.19***		
		(7.33)		
Forestland	-			-0.26***
				(0.05)
Irrigation Area	-			-4.05***
				(0.45)
Average Annual Total Precipitation	+			-0.04
				(0.03)
Average Annual Temperature	-			-0.96***
				(0.31)
Built-up Land	+		2.18***	
			(0.21)	
Net Income of Rural Population	+		0.12**	
			(0.05)	
Number of Agricultural Laborers	+		1.05***	
			(0.39)	
Agricultural Machinery Power	+		0.11	
			(0.14)	
Price Index of Agricultural Products	+		-0.25***	
			(0.08)	
Constant		3,775.21***	1,552.02***	1,003.48***
		(64.73)	(19.68)	(189.34)
Number of Observations		248	248	248
R^2		0.86	0.33	0.58

Note: (1) The signs indicate that the dependent variable is expected to be associated with the independent variables positively or negatively. (2) Standard errors are in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

The fitted equation of farmland expansion demonstrates that the increases of built-up area and farmland expansion are strongly correlated. I employed per capita annual net income of rural population, number of agricultural laborers, and the total agricultural machinery power to capture the effects of changed inputs and outputs on farmland. The significantly positive coefficient (0.12) of farmers' net income indicates that it is highly correlated with agricultural expansion. Likewise, rural laborer is positively correlated with agricultural expansion; but the coefficient of agricultural machinery power is not statistically significant. Finally, the coefficient of price index for agricultural products is negatively correlated with farmland expansion, revealing that price increase may not necessarily result in farmland expansion at the extensive margin.

In wetland loss equation, as expected, farmland expansion is strongly negatively correlated with wetland loss, with a coefficient of -0.24. The relationship between wetland and forestland is substitutional. The significant negative coefficient of irrigation area confirms the view that wetland loss is strongly related to the change in local cropping pattern (from dryland crops to irrigated crops). In this region, pumping water greatly disturbs the local natural water system; at the same time, the irrigation network also cuts off the hydraulic relationships of the local natural water system. All these practices have limited water supplies from rivers to wetland, exerting a strong negative correlation (-4.05) between irrigation area increase and wetland loss. In addition, as the warming climate (-0.96) also contributed to wetland loss over the past 30 years.

Various other model validation techniques are listed in Appendix B below. Here, I took a sensitivity analysis by dropping out variables one by one for each step. The first variable I omitted from the system is the built-up land which is assumed to be exogenous of the "Forestland-Farmland-Wetland" system, then price index of timber, price index of agricultural products,

agricultural machinery power, and average annual total precipitation were dropped out step by step. The results are listed in Table 5.9 below.

From the Table 5.5 below, by omitting the built-up land from the explanatory variable set, model performance actually improved. With a coefficient of the farmland being less than 1.30, the value is more trustworthy according to the extended land conversion matrixes. And the model progress a little with the price index of timber excluded. In this model, the wetland are negatively correlated to forestland, and the coefficient magnitude were verified by following on regression.

In sum, Table 5.5 demonstrates that there are model improvement spaces by omitting variables from the explanatory variable set. With the exogenous variable built-up land and the price index of timber casted out, coefficients in the "Forestland-Farmland-Wetland" model are more close to the magnitude as expected.

Table 5.5 Sensitivity analysis of "Forestland-Farmland-Wetland" model

VAR	Forestland	Farmland	Wetland	Forestland	Farmland	Wetland	Forestland	Farmland	Wetland
	Built-up Land			Price Index of Timber			Price Index of Agricultural Products		
Farmland	-1.28***		-0.54***	-1.24***		-0.47***	-1.25***		-0.38***
	(0.04)		(0.07)	(0.03)		(0.07)	(0.03)		(0.08)
Wetland	-0.16			-0.31***			-0.38***		
	(0.13)			(0.11)			(0.11)		
TimberPrice	0.30								
	(0.20)								
TotalPop	-0.48***			-0.43***			-0.41***		
	(0.08)			(0.06)			(0.06)		
NFPP	36.79***			32.17***			28.51***		
	(9.09)			(8.47)			(8.35)		
AgPrice		-0.04			-0.05				
		(0.07)			(0.07)				
IncmRurPop		0.24***			0.25***			0.24***	
		(0.06)			(0.05)			(0.05)	
Aglabor		2.72***			2.81***			2.60***	
		(0.45)			(0.45)			(0.38)	
AgMachPowr		0.21			0.18			0.15	
		(0.17)			(0.16)			(0.15)	
Forestland			-0.55***			-0.51***			-0.47***
			(0.06)			(0.06)			(0.06)
IrrigatArea			-4.19***			-4.62***			-4.97***
			(0.36)			(0.37)			(0.38)
Precip			-0.05**			-0.05**			-0.05**
			(0.02)			(0.02)			(0.02)
AveTemp			-1.07***			-1.06***			-1.09***
			(0.26)			(0.27)			(0.27)
Constant	3,601.78	1,541.25	1,899.62	3,577.86	1,539.79	1,733.10	3,592.24	1,541.19	1,534.10
	(77.20)	(20.31)	(196.64)	(69.91)	(20.34)	(200.64)	(68.89)	(20.28)	(208.29)
R^2	0.88	0.39	0.79	0.90	0.39	0.75	0.91	0.39	0.69

Table 5.5 (cont'd)

VAR	Forestland	Farmland	Wetland	Forestland	Farmland	Wetland
	Agricultural Machinery Power			Average Annual Total Precipitation		
Farmland	-1.25***		-0.36***	-1.26***		-0.32***
	(0.03)		(0.08)	(0.03)		(0.08)
Wetland	-0.32***			-0.30***		
	(0.10)			(0.10)		
TotalPop	-0.41***			-0.41***		
	(0.06)			(0.06)		
NFPP	31.97***			33.90***		
	(8.41)			(8.47)		
IncmRurPop		0.26***			0.26***	
		(0.04)			(0.04)	
Aglabor		2.74***			2.77***	
		(0.35)			(0.35)	
Forestland			-0.44***			-0.41***
			(0.06)			(0.07)
IrrigatArea			-4.98***			-4.80***
			(0.39)			(0.38)
Precip			-0.05**			
			(0.02)			
AveTemp			-1.04***			-1.02***
			(0.27)			(0.27)
Constant	3,591.27	1,548.19	1,456.41	3,591.79	1,547.89	1,330.27
	(69.65)	(19.53)	(210.97)	(70.48)	(19.55)	(220.49)
R^2	0.90	0.38	0.67	0.90	0.38	0.64

Note (1) All the constant are significant at the level of $p<0.01$, the "***" mark were deleted for the purpose of saving space. (2) Numbers of Observations are 248. (3) Standard errors in parentheses, *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

5.5 Discussion and Conclusions

The basic purpose of this chapter is to explore the underlying driving forces in more systematic frameworks. Based on the single-equation OLS analysis results in Chapter 4, I first constructed an interactive system of two classes of land use, forestland and farmland, assuming that farmland could be endogenous in explaining the deforestation process. Then, a series of formal

statistical tests was conducted to select appropriate instruments among multiple combinations of candidate variables that were thought to be relevant to agricultural development. It was found that built-up land, which increased along with farmland expansion but did not have a direct relationship with forestland, was the only satisfactory instrument. Meanwhile, tests also demonstrated that the finite sample bias of IV analysis is smaller than that of OLS. The IV results provided strong evidence of endogeneity of land use; thus, I went one step further by including another class of land use, wetland, in a system of three classes of land use, with forestland-farmland-wetland being jointly determined. The three interrelated classes of land use—deforestation, farmland expansion, and wetland loss—were investigated together through three equations. The interactive relationships of the three classes of land use rendered this system of three equations to be a simultaneous equations model.

Clearly, results derived from the forestland-and-farmland and forest-farm-wetland systems are more encouraging and robust. All of the included variables, except for price indices for timber price and agricultural products, have the correct signs. This study was partly motivated to investigate the effect of implementing the NFPP, which was positive but insignificant in the OLS analysis of Chapter 4. Now, it is confirmed that the program has played a significantly positive role in protecting local forests in both systems. Meanwhile, deforestation is more strongly correlated with farmland expansion, and wetland change has a strong substitutive effect with forestland—loss of wetland tends to save forestland from loss, and vice versa. Additionally, with an IV method and an SEM, exploring the underlying driving forces became more likely to answer such questions as how the population growth and urbanization, irrigation system construction and climate condition changes have affected wetland change, and how local farmers' revenue increases,

amount of available agricultural labor, and machinery purchases have influenced their land allocation decisions.

Moreover, different estimation strategies have allowed comparisons of the performance of regression methods as well as estimated results. From the single-equation model used in Chapter 4 to the IV method and SEM analysis in this chapter, I have employed a number of typically-used modeling approaches: fixed-, random-, and between-effects models; and ML, LIML, GMM, 2SLS and 3SLS estimation techniques. The between-effects models have little power in explaining forestland change in a single-equation model, lending confidence to the validity of choosing to use fixed-effects estimators. Indeed, the Mundlak model in Chapter 4 shed light on the existence of endogeneity; in this chapter, endogeneity has been formally tested and addressed. What is even more important is that the alternative models have generally corroborated the consistency of my empirical results, making them more robust and reliable.

Because the coefficients of prices for timber and farm products are insignificant, however, a closer examination of the price indices is necessary. Data show that the timber price index went up sharply after the year 2000, exactly when the NFPP was initiated; before that, it fluctuated within a relatively small range, but did not demonstrate any trend over time as deforestation did. This implies that timber price had little effect on local farmers' decisions over timber harvesting or forestland clearing. As we traced the data back to the earlier years, we realized that prices for both forest and farm products in this region were under strict government control for quite a long time. It appears that this had depressed prices and caused some abnormal association between the dynamics of farmland and forestland and output prices. Similarly, machinery power grew much faster after 2000, but with the NFPP and wetland protection programs having been put in place

further farmland expansion was halted, making the relationship between machinery power and agricultural land not as strong as expected.

There are some other limitations in the current study. The small sample size has made the estimated results sometimes sensitive to the modeling framework used and assumptions made. Also, the small sample size did not allow me to take into consideration the spatial autocorrelation. Because the original LUCC data covered six periods, I had to linearly interpolate these periodic data to obtain annual observations to match the existing socioeconomic data. This has made it a challenge to apply panel-data and other estimation methods. It is hoped that future research will be able to overcome these problems.

APPENDICES

Appendix A: A Description of Various Tests When Instrument Variables Are Used

In the case of a weak instrument variable problem, several tests are needed during the first and second stages of estimation: the under-identification tests, weak identification tests, and weak-instrument-robust inference tests during the first-stage regression; and the endogeneity test and under-, weak- and over-identification tests during the second stage regression.

First Stage Test Result

The under-identification tests detect whether the equation is "identified;" in other words, whether the instrument variables are "relevant." An instrument is relevant if it correlates with the endogenous regressors Fa and thus accounts for significant variation in Fa (Baum et al. 2007b; Schaffer 2012). The *Sanderson-Windmeijer (SW) chi-squared* statistic (Sanderson & Windmeijer 2013) and *Kleibergen-Paap (KP) rk LM chi-squared* statistics are used for testing under-identification. The *KP* statistic is robust to various forms of heteroskedasticity, autocorrelation, and clustering (Kleibergen & Paap 2006). The null hypothesis is that the endogenous regressor Fa in regression is unidentified. The large statistics and corresponding small P-values in Table 5.2 suggest that the null hypothesis is rejected, and the model is identified.

Based on the under-identification tests, weak-identification tests discern whether the excluded instruments are "weakly" correlated with the endogenous regressors. Table 5.2 contains two diagnostic statistic values for weak identification: the *Cragg-Donald (CD) Wald statistic* (Cragg & Donald 1993) and the *Kleibergen-Paap Wald* statistic (Baum et al. 2007a). Commonly, it is required that the maximal bias in IV be no more than 10% of the bias of OLS. Thus, according to a rule of thumb proposed by Staiger and Stock (1994), F values larger than 10 are required, and in my results, the values of the F statistics all exceed 10. Compared to the critical values tabulated by Stock and Yogo (2005) for a single endogenous regressor with 5 excluded instruments, the

threshold value of 10% maximal LIML size is 4.84. So, we can infer that the instruments are not weak as all the first stage *F* statistics are larger than the critical values.

Table 5.2 also presents results of weak-instrument-robust inference tests. The null hypothesis is that the joint significance of endogenous regressors in the structural equation equals zero. This is equivalent to testing that the coefficients for the excluded instrument variables equal zero in the reduced form (Andrews & Stock 2005; Chernozhukov & Hansen 2008). The *Anderson-Rubin (AR) Wald* test and its *F* statistics (Anderson & Rubin 1949) and the *Stock-Wright (SW) S* statistic, all these tests are robust to weak instruments, that is, no information about the correlation between the endogenous variable farmland and the exogenous variables is required (Stock & Wright 2000; Stock et al. 2002; Moreira 2003). The corresponding p-values in Table 5.2 reject the null hypothesis, indicating the coefficient of the endogenous variable "farmland" is non-zero.

Second Stage Test Results

The null hypothesis of the endogeneity test is that the specified endogenous regressors can be treated as exogenous. It is the difference of the two *Hansen* (or *Sargan*) statistics—one for the model where the suspected variable is treated as endogenous and the other for the equation with the suspect variable treated as exogenous (Schaffer 2012). So the endogeneity test resembles the *Hausman* test under the homoskedasticity assumption, but the test statistics reported in Table 5.2 are robust to heteroskedastisity of various forms (Hayashi 2000). From the Chi-squared and corresponding p-values, even with different model specifications, the assumption that farmland area change is exogenous with forestland change is easily rejected.

The *Hansen's J* statistic tests the over-identification restrictions of all instruments. Similar to the *Sargan*'s statistic, the null hypothesis of a Hansen's J test is that all the excluded instruments are valid. Under the assumption of homoskedastic errors, the *Sargan*'s statistic is reported; otherwise, the *Hansen's J* statistic is reported instead. In the case where all instruments were

included, the test statistic rejected the null assumption, casting doubts on the validity of these instruments. As the excluded variables are strongly correlated with the suspect endogenous variable farmland, this satisfies the first requirement of a good candidate for an instrument variable. Thus the potential problem of these instrument variables would lie in the non-zero correlations between the excluded instruments with the error terms.

Appendix B: Validation of the "Farmland-Forestland-Wetland" System

Variable Selection

As nested regression models do not support the criteria of AIC and BIC (StataCorp. 2013),

the variables, though based on theoretical rationale and evidence in the literature, should still

subject to close scrutiny. So, I did a pre-estimation validation based on separate equations. Also,

because stepwise variable selection method doesn't support panel data regression, I tried different

variable combinations manually. Recall that the deforestation equation was already calibrated in

Chapter 4, I have estimated the agricultural land expansion and wetland loss equations here, with

results being listed in Table 5.6 and Table 5.7 below.

Table 5.6 Farmland expansion model variable selection

Farmland	(I) All	(II) Builtup	(III) AgMachPowr	(IV) AgPrice	(V) Aglabor
Aglabor	2.01	2.25*	2.43**	2.74***	3.83***
	(1.23)	(1.06)	(1.00)	(0.60)	(0.74)
IncmRurPop	0.19	0.19	0.22*	0.26	
	(0.12)	(0.11)	(0.11)	(0.15)	
AgPrice	-0.10	0.02	0.08		
	(0.10)	(0.15)	(0.17)		
AgMachPowr	0.37	0.37			
	(0.43)	(0.40)			
Builtup	0.65				
	(0.94)				
Constant	1,532.20***	1,537.01***	1,549.66***	1,550.32***	1,573.69***
	(58.45)	(61.29)	(55.27)	(52.96)	(38.70)
AIC	3053.48	3056.75	3058.84	3058.06	3082.71
BIC	3071.04	3070.80	3069.38	3065.09	3086.22
R^2	0.41	0.40	0.39	0.38	0.32

Note: Robust standard errors in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

Model I has the smallest AIC, in which number of agricultural laborers, annual net income

of rural population, price index of agricultural products, aggregate agricultural machinery power,

and built-up land are included. At the same time, Model IV has the lowest BIC, in which only

number of agricultural laborers and annual net income of rural population are included. Also, most variables have the expected signs though some of their coefficients are not statistically significant. Based on the AIC, BIC, and estimated coefficients, thus, there is no strong reason to differentiate Model I, II, III and IV.

Table 5.7 Wetland loss model variable selection

Wetland	(1) All	(2) Precip	(3) AveTemp	(4) IrrigatArea	(5) Forest
Farmland	-0.78***	-0.78***	-0.78***	-0.84**	-0.14*
	(0.16)	(0.16)	(0.18)	(0.28)	(0.06)
Forestland	-0.73***	-0.73***	-0.72***	-0.68**	
	(0.15)	(0.15)	(0.17)	(0.27)	
IrrigatArea	-3.63***	-3.42***	-4.12***		
	(0.54)	(0.49)	(0.67)		
AveTemp	-1.20**	-1.20**			
	(0.36)	(0.36)			
Precip	-0.05**				
	(0.02)				
Constant	2,554.47***	2,518.32***	2,475.36***	2,488.80**	426.61***
	(453.32)	(466.06)	(516.56)	(828.76)	(107.69)
AIC	2245.60	2249.63	2270.72	2456.59	2669.97
BIC	2263.16	2263.69	2281.26	2463.61	2673.48
R^2	0.85	0.85	0.83	0.64	0.14

Note: Robust standard errors in parentheses. *, **, and *** indicate the significance levels of 90%, 95%, and 99%, respectively.

As the linkages in the equation of wetland loss are more straightforward, the included variables are all strongly correlated with it. Consequently, the AIC and BIC criteria point to an agreement to include all the variables, among which irrigation area increase played a dominant role in wetland decrease and climate change, as reflected in average temperature increase and precipitation increase, also had a significant effect.

Model Validation

Here, I first manually verified the correlation between equations. The correlation coefficient between the error terms of forestland equation and farmland equation is 0.74; the same coefficient between forestland and wetland equations is -0.37, and that between farmland and wetland equations is -0.17. The Breusch-Pagan LM Diagonal Covariance Matrix Test is a formal test which hypothesizes that the OLS estimate is appropriate. Test outcomes rejected the null with a P-Value close to zero (Lagrange Multiplier Test = 176.61), in favor of the alternative 3SLS.

Table 5.8 Breusch-Pagan LM diagonal covariance matrix

	Forestland	Farmland	Wetland
Forestland	2208.29		
Farmland	3944.68	12819.11	
Wetland	-520.90	-575.87	915.47

Additionally, I tried to compare the forecasted and observed values of the relevant variables as part of my model validation efforts. As the year of imagery classified land use data are 1977, 1984, 1993, 2000, 2004 and 2007, I dropped data for the last three years, the estimation results are very close to the results produced with the full data set (see Table 5.9 below).

Table 5.9 Results of 3SLS analysis of the "Farmland-Forestland-Wetland" system based on data from 1977 to 2004

VARIABLES	Expected Sign	(1) Forestland	(2) Farmland	(3) Wetland
Farmland	-	-1.39***		-0.27***
		(0.04)		(0.07)
Wetland	-	-0.60***		
		(0.13)		
Price Index of Timber	-	0.36		
		(0.23)		
Population	-	-0.26***		
		(0.07)		
NFPP	+	18.35**		
		(7.61)		
Forestland	-			-0.28***
				(0.05)
Irrigation Area	-			-3.89***
				(0.52)
Average Annual Total Precipitation	+			-0.03
				(0.03)
Average Annual Temperature	-			-1.15***
				(0.33)
Built-up Land	+		1.99***	
			(0.21)	
Net Income of Rural Population	+		-0.21**	
			(0.09)	
Number of Agricultural Laborers	+		0.13**	
			(0.06)	
Agricultural Machinery Power	+		0.93**	
			(0.38)	
Price Index of Agricultural Products	+		0.31*	
			(0.18)	
Constant		3,809.43***	1,535.17***	1,090.76***
		(75.62)	(22.24)	(176.76)
Number of Observations		248	248	248
R^2		0.88	0.32	0.59

Note: The signs indicate that the dependent variable is expected to be associated with the independent variables positively or negatively.

Forecasting of simultaneous equation based on panel data doesn't work in STATA. As a compromise, I made land-use predications based on the estimated coefficients of the "Forestland-farmland-wetland" system and compared the predicated values to the observations across the study period. Results for forestland are shown in Figure 5.3.

Figure 5.3 Predicted and observed values of forestland based on the "Forestland-Farmland-Wetland" model

Unit: Square kilometers
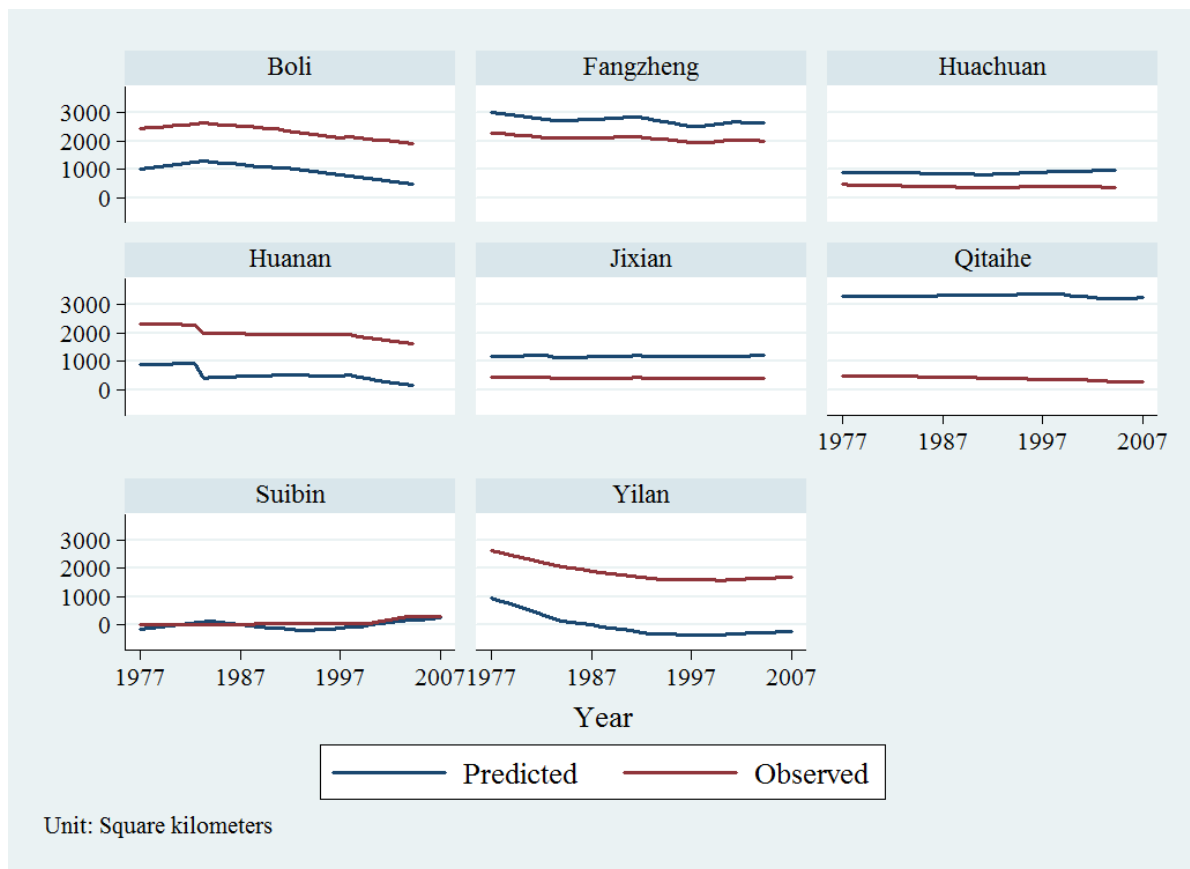
Overall, the predicted areas of forestland capture the general patterns of observed forestland dynamics. Meanwhile, gaps exist between predicted and observed changes of forestland, due to the heterogeneities of the initial forestland areas. Since I am more interested in the land dynamics in the whole study region, the 8 counties are studied as an integrated landscape. The disparities across counties are not so much a concern to me. Further, within a small sample, it

would cost too many degrees of freedom to create dummies for each county. Thus, I leave the prediction gaps for certain counties as such. It is easy to find out in Figure 5.3 that Qitaihe has the largest prediction gap. Qitaihe is a prefecture-level city with large area of built-up land in its jurisdiction. During the process of urbanization, farmers flocked into the city; and the disproportionately increased number of laborers, income and non-agricultural used machinery could have made the predicted amount of forestland deviate from its observed values.

Figure 5.4 Predicted and observed values of farmland based on the "Forestland-Farmland-Wetland" model

Unit: Square kilometers
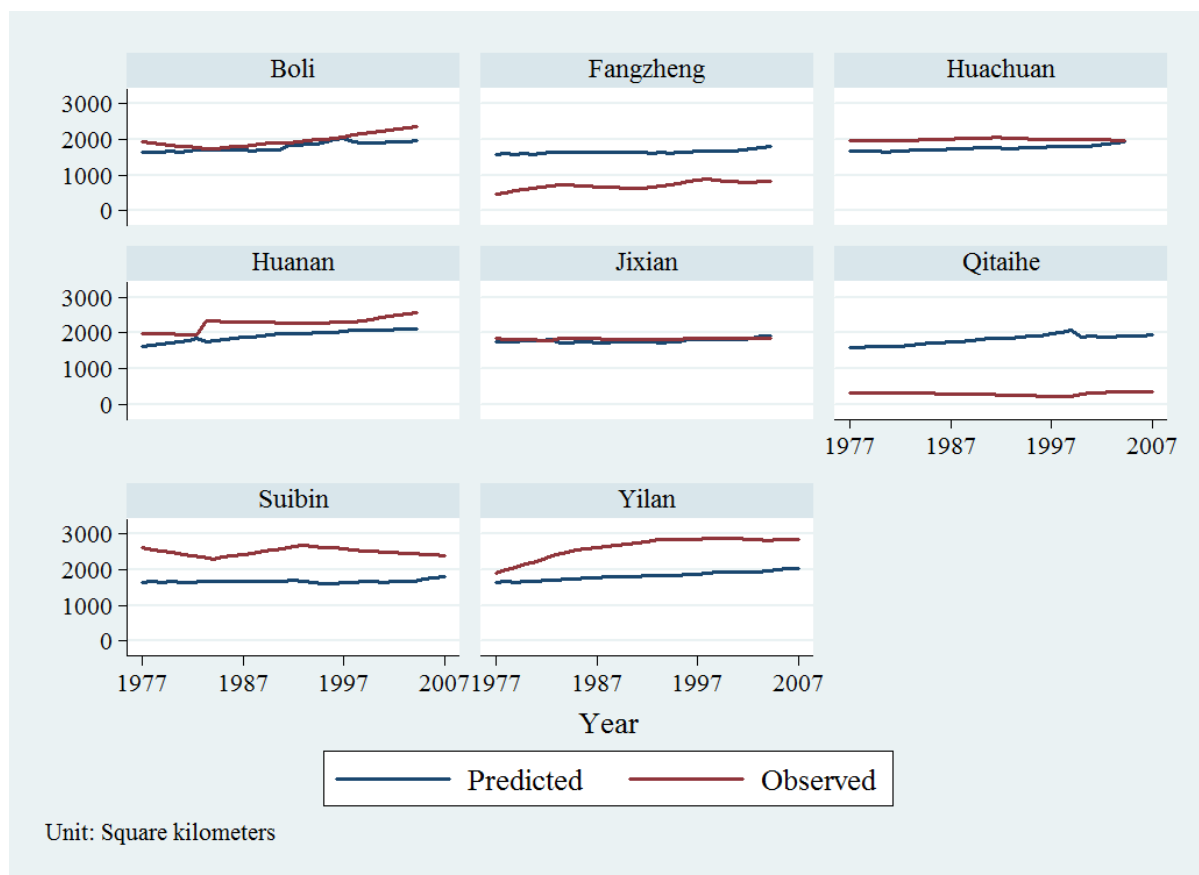
Farmland prediction pattern are not as fit compared to that of forestland while it is still adequate. The county which matches best is Jixian, and the predictions of Boli, Huachuan and Huanan are all very close. The prediction of municipal district of Qitaihe, as expected, differs

181

most from its true values. As the city area of Qitaihe shifts it production from agricultural industry into other activities, the prediction of farmland are higher than all the rest counties. Meanwhile, the Suibin county and Yilan county are agricultural dominate, the real amount of farmland are larger than as predicted.

Figure 5.5 Predicted and observed values of wetland based on the "Forestland-Farmland-Wetland" model

Unit: Square kilometers
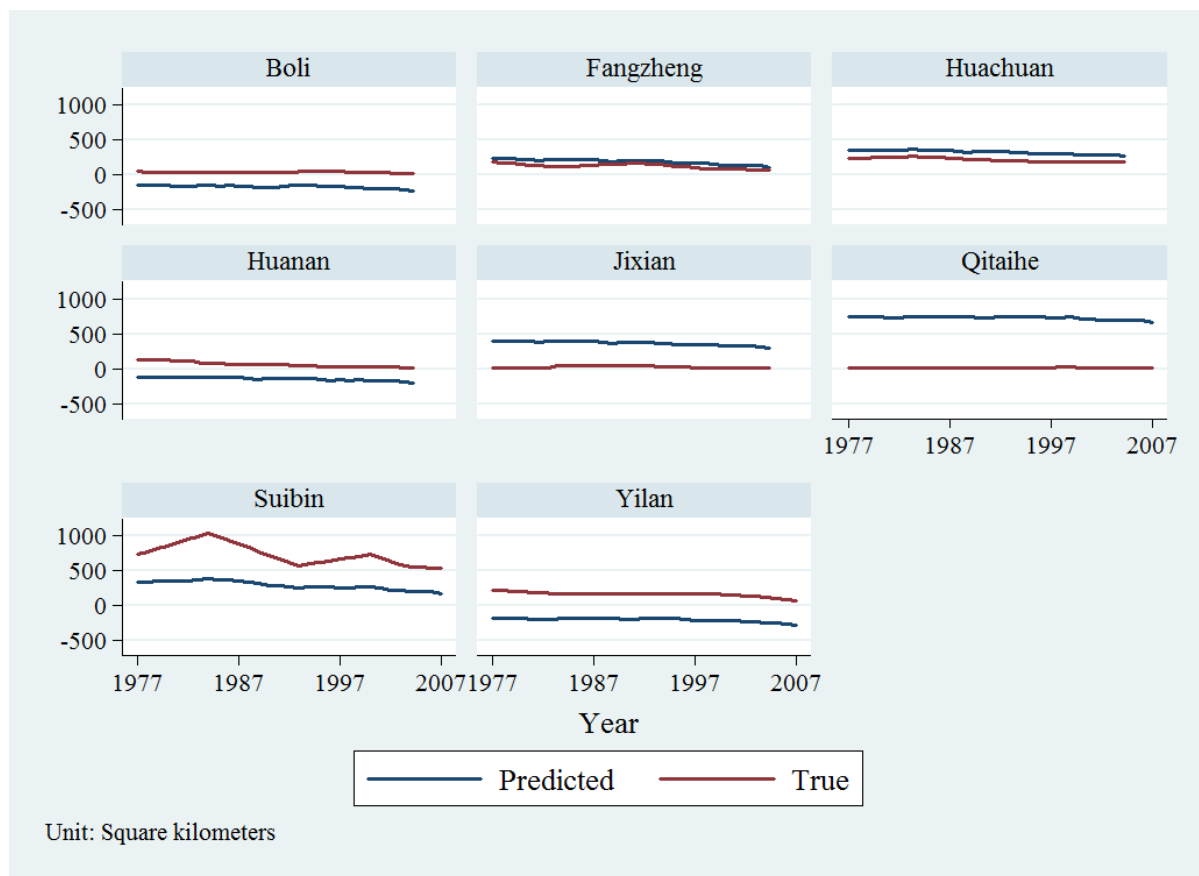
Comparisons of predicted and observed values of farmland and wetland tell a similar story—while overall patterns of change over time are largely consistent, there exist gaps between them. Wetland area in all the counties demonstrates a decreasing trend. As wetland is a minor land use category in the study region and varies according to meteorology changes. Counties like Suibin, bordering Songhua Rive and Amur River (Heilongjiang), wetland area fluctuate due to the floodplain changes according to different precipitation situations.

REFERENCES

REFERENCES

Al-Tuwaijri, S.A., Christensen, T.E., Hughes, K., 2004. The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach. Accounting, organizations and society 29, 447-471

Allison, P.D., 2009. Fixed effects regression models. SAGE publications, Thousand Oaks.

Alonso, D., Sole, R.V., 2000. The DivGame simulator: a stochastic cellular automata model of rainforest dynamics. Ecological Modelling 133, 131-141

Anderberg, M.R., 2014. Cluster Analysis for Applications: Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic press.

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: A review and recommended two-step approach. Psychological bulletin 103, 411-423

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use and land cover classification system for use with remote sensor data. In: Geological Survey Professional Paper. USGS, Reston, VA

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics, 46-63

Andrews, D., Stock, J.H., 2005. Inference with weak instruments. National Bureau of Economic Research Cambridge, Mass., USA

Angelsen, A., 1999. Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights. Journal of Development Economics 58, 185-218

Angelsen, A., Kaimowitz, D., 1999. Rethinking the Causes of Deforestation: Lessons from Economic Models. The World Bank Research Observer 14, 73-98

Angelsen, A., Shitindi, E.F.K., Aarrestad, J., 1999. Why do farmers expand their land into forests? Theories and evidence from Tanzania. Environment and Development Economics 4, 313-331

Angelsen, A., van Soest, D., Kaimowitz, D., Bulte, E., 2001. Technological change and deforestation: A theoretical overview. Agricultural technologies and tropical deforestation, 19-34

Angrist, J., Krueger, A.B., 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. National Bureau of Economic Research

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. Journal of the American statistical Association 91, 444-455

Anselin, L., 2002. Under the hood issues in the specification and interpretation of spatial regression models. Agricultural Economics 27, 247-267

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International regional science review 26, 153-166

Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89, 3-25

Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. Statistics Textbooks and Monographs 155, 237-290

Asner, G.P., Broadbent, E.N., Oliveira, P.J., Keller, M., Knapp, D.E., Silva, J.N., 2006. Condition and fate of logged forests in the Brazilian Amazon. Proceedings of the National Academy of Sciences 103, 12947-12950

Asner, G.P., Knapp, D.E., Broadbent, E.N., Oliveira, P.J., Keller, M., Silva, J.N., 2005. Selective logging in the Brazilian Amazon. Science 310, 480-482

Baltagi, B., 2008. Econometric analysis of panel data. John Wiley & Sons.

Baltagi, B.H., 1981. Simultaneous equations with error components. Journal of Econometrics 17, 189-200

Baltagi, B.H., 2006. An Alternative Derivation of Mundlak's Fixed Effects Results Using System Estimation. Econometric Theory 22, 1191-1194

Baltagi, B.H., Giles, M.D., 1998. Panel data methods. Statistics Textbooks and Monographs 155, 291-324

Baltagi, B.H., Liu, L., 2009. A note on the application of EC2SLS and EC3SLS estimators in panel data models. Statistics & Probability Letters 79, 2189-2192

Baltagi, B.H., Song, S.H., 2006. Unbalanced panel data: a survey. Statistical Papers 47, 493-523

Barbier, E., 1994. The economics of the tropical timber trade. CRC Press.

Barbier, E.B., 2004. Agricultural Expansion, Resource Booms and Growth in Latin America: Implications for Long-run Economic Development. World Development 32, 137-157

Barbier, E.B., Burgess, J.C., 1996. Economic analysis of deforestation in Mexico 31. Environment and Development Economics 1, 203-239

Barbier, E.B., Burgess, J.C., 1997. The economics of tropical forest land use options. Land Economics 73, 174-195

Baum, C.F., Schaffer, M.E., Stillman, S., 2007a. Enhanced routines for instrumental variables/GMM estimation and testing. Stata journal 7, 465-506

Baum, C.F., Schaffer, M.E., Stillman, S., 2007b. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression.

Baumol, W.J., Hall, P., 1977. Economic theory and operations analysis.

Bawa, K.S., Dayanandan, S., 1997. Socioeconomic factors and tropical deforestation. Nature (London) 386, 562-563

Beck, N., 2001. Time-series-cross-section data: What have we learned in the past few years? Annual review of political science 4, 271-293

Bekker, P.A., Ploeg, J., 2005. Instrumental variable estimation based on grouped data. Statistica Neerlandica 59, 239-267

Berry, S.T., 1994. Estimating discrete-choice models of product differentiation. The RAND Journal of Economics 25, 242-262

Biørn, E., 2004. Regression systems for unbalanced panel data: a stepwise maximum likelihood procedure. Journal of Econometrics 122, 281-291

Boulos, M.N., 2005. Web GIS in practice III: creating a simple interactive map of England's strategic Health Authorities using Google Maps API, Google Earth KML, and MSN Virtual Earth Map Control. International Journal of Health Geographics 4, 22

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American statistical association 90, 443-450

Brownstone, D., Golob, T.F., Kazimi, C., 2002. Modelling non-ignorable attrition and measurement error in panel surveys: an application to travel demand modeling. In: Earlier Faculty Research. University of California Transportation Center

Burgess, J.C., 1993. Timber production, timber trade and tropical deforestation. Ambio 22, 136-143

Byrne, B.M., 2010. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Psychology Press.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. Journal of Business & Economic Statistics 29, 238-249

Cameron, A.C., Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. Journal of Human Resources 50, 317-372

Cameron, A.C., Trivedi, P.K., 2009. Microeconometrics using stata. Stata Press College Station, TX.

Carr, D., Suter, L., Barbieri, A., 2005. Population Dynamics and Tropical Deforestation: State of the Debate and Conceptual Challenges. Population & Environment 27, 89-113

Center，, N.M.I., 2009. China Meteorological Data Sharing Service System. Beijing

Chavez, P.S., 1996. Image-based atmospheric corrections-revisited and improved. Photogrammetric engineering and remote sensing 62, 1025-1035

Chenhall, R.H., Moers, F., 2007. The issue of endogeneity within theory-based, quantitative management accounting research. European Accounting Review 16, 173-196

Chernozhukov, V., Hansen, C., 2008. The reduced form: A simple approach to inference with weak instruments. Economics Letters 100, 68-71

Chichilnisky, G., 1994. North-south trade and the global environment. American Economic Review 84, 851-874

Chinese Academy of Sciences, 2008. China Remote Sensing Satellite Ground Station.

Chomitz, K.M., Gray, D.A., 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. The World Bank Economic Review 10, 487-512

Clark, T.S., Linzer, D.A., 2012. Should I use fixed or random effects. Unpublished paper

Clarke, K., 1997. A self-modifying cellular automaton model of historical. Environment and planning B: planning and design 24, 247-261

Clarke, P., Crawford, C., Steele, F., Vignoles, A.F., 2010. The choice between fixed and random effects models: some considerations for educational research. Social Science Research Network

Compilation Committee of Heilongjiang Annals, 1993. Heilongjiang Price Annals. Helongjiang People's Press, Harbin.

Cornwell, C., Schmidt, P., Wyhowski, D., 1992. Simultaneous equations and panel data. Journal of Econometrics 51, 151-181

Cragg, J.G., Donald, S.G., 1993. Testing identifiability and specification in instrumental variable models. Econometric Theory 9, 222-240

Cropper, M., Griffiths, C., Mani, M., 1997. Roads, population pressures, and deforestation in Thailand, 1976-89. World Bank Policy Research Working Paper

Dai, X., Khorram, S., 1998. The effects of image misregistration on the accuracy of remotely sensed change detection. Geoscience and Remote Sensing, IEEE Transactions on 36, 1566-1577

De Janvry, A., Fafchamps, M., Sadoulet, E., 1991. Peasant household behaviour with missing markets: some paradoxes explained. The Economic Journal 101, 1400-1417

Deininger, K., Minten, B., 2002. Determinants of deforestation and the economics of protection: An application to Mexico. American Journal of Agricultural Economics 84, 943-960

Deininger, K.W., Minten, B., 1999. Poverty, policies, and deforestation: the case of Mexico. Economic Development and Cultural Change 47, 313-344

Deng, J., Wang, K., Deng, Y., Qi, G., 2008. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. International Journal of Remote Sensing 29, 4823-4838

Dezhbakhsh, H., Levy, D., 1994. Periodic properties of interpolated time series. Economics Letters 44, 221-228

Dradjad H. Wibowo, R.N.B., 1999. Deforestation mechanisms: a survey. International Journal of Social Economics 26, 455 - 474

Drechsel, P., Kunze, D., De Vries, F.P., 2001. Soil nutrient depletion and population growth in sub-Saharan Africa: a Malthusian nexus? Population and Environment 22, 411-423

Du, Y., Yu, C., Jie, L., 2009. A study of GIS development based on KML and Google Earth. In: INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on, pp. 1581-1585. IEEE

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric theory 11, 122-150

Epple, D., 1987. Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products. The Journal of Political Economy 95, 59-80

Färe, R., Grosskopf, S., Norris, M., Zhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. The American economic review 84, 66-83

Fargione, J., Hill, J., Tilman, D., Polasky, S., Hawthorne, P., 2008. Land clearing and the biofuel carbon debt. Science 319, 1235-1238

Ferber, J., 1999. Multi-agent systems: an introduction to distributed artificial intelligence. Addison-Wesley Reading.

Fingleton, B., Gallo, J.L., 2007. Finite Sample Properties of Estimators of Spatial Models with Autoregressive, or Moving Average, Disturbances and System Feedback. Annals of Economics and Statistics / Annales d'Économie et de Statistique, 39-62

Fischer, J., Lindenmayer, D.B., 2007. Landscape modification and habitat fragmentation: a synthesis. Global Ecology and Biogeography 16, 265-280

Fleming, M.M., 2004. Techniques for estimating spatially dependent discrete choice models. In: Advances in spatial econometrics. Springer, pp. 145-168.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., 2005. Global consequences of land use. science 309, 570-574

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote sensing of environment 80, 185-201

Foody, G.M., 2009a. Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing 30, F5273-5291

Foody, G.M., 2009b. Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing 30, 5273-5291

Franzese, R.J., Hays, J.C., 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. Political Analysis 15, 140-164

Frees, E.W., 2004. Longitudinal and panel data: analysis and applications in the social sciences. Cambridge University Press.

Gao, J., Liu, Y., 2011. Climate warming and land use change in Heilongjiang Province, Northeast China. Applied Geography 31, 476-482

Geist, H.J., Lambin, E.F., 2001. What drives tropical deforestation? A meta-analysis of proximate and underlying causes of defores-tation based on subnational scale case study evidence. In: LUCC Report Series No. 4., University of Louvain, Louvain-la-Neuve

Geist, H.J., Lambin, E.F., 2002a. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. BioScience 52, 143-150

Geist, H.J., Lambin, E.F., 2002b. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. BioScience 52, 143-150

Geoghegan, J., Villar, S.C., Klepeis, P., Mendoza, P.M., Ogneva-Himmelberger, Y., Chowdhury, R.R., Turner, B., Vance, C., 2001. Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. Agriculture, Ecosystems & Environment 85, 25-46

Goldman, A., 1993. Agricultural Innovation in Three Areas of Kenya: Neo-Boserupian Theories and Regional Characterization. Economic Geography 69, 44-71

Grace, J.B., 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge.

Graham, R., Hunsaker, C., O'neill, R., Jackson, B., 1991. Ecological risk assessment at the regional scale. Ecological applications, 196-206

Grainger, A., 1995. The Forest Transition: An Alternative Approach. Area 27, 242-251

Grossman, G.M., Helpman, E., 1993. Endogenous innovation in the theory of growth. Journal of Economic Perspectives 8, 23-44

Hansen, M.C., Stehman, S.V., Potapov, P.V., 2010. Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences 107, 8650-8655

Harkness, J., 1998. Recent trends in forestry and conservation of biodiversity in China. The China Quarterly 156, 911-934

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica: Journal of the Econometric Society 46, 1251-1271

Hausman, J.A., Newey, W.K., Woutersen, T.M., 2007. IV Estimation with Heteroskedasticity and Many Instruments. Centre for microdata methods and practice

Hausman, J.A., Taylor, W.E., 1981. Panel Data and Unobservable Individual Effects. Econometrica 49, 1377-1398

Hayashi, F., 2000. Econometrics Princeton University Press. Princeton

He, H.S., DeZonia, B.E., Mladenoff, D.J., 2000. An aggregation index (AI) to quantify spatial patterns of landscapes. Landscape Ecology 15, 591-601

Hedges, L.V., Vevea, J.L., 1998. Fixed-and random-effects models in meta-analysis. Psychological methods 3, 486-504

Heilongjiang Statistical Bureau, 1986-2008. Heilongjiang Statistical Yearbook (1986-2008). China Statistics Press, Beijing

Heilongjiang Statistics Bureau, 2009. Sixty Years of Heilongjiang. China Statistics Press, Beijing.

Herbert, A.J., Arild, A., 2009. The paradox of household resource endowment and land productivity in Uganda. In: Agricultural Economists Conference, Beijing

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric environment 42, 7561-7578

Hogeweg, P., 1988. Cellular automata as a paradigm for ecological modeling. Applied mathematics and computation 27, 81-100

Hsiao, C., 1985. Benefits and limitations of panel data. Econometric Reviews 4, 121-174

Hsiao, C., 2003. Analysis of panel data. Cambridge university press.

Hsiao, C., 2007. Panel data analysis—advantages and challenges. Test 16, 1-22

Hsiao, C., 2014. Analysis of panel data. Cambridge university press, Cambridge.

Huang, W., Deng, X., Lin, Y., Jiang, Q., 2010. An Econometric Analysis of Causes of Forestry Area Changes in Northeast China Procedia Environmental Sciences 2

Hunziker, M., Kienast, F., 1999. Potential impacts of changing agricultural activities on scenic beauty–a prototypical technique for automated rapid assessment. Landscape Ecology 14, 161-176

Hyde, W.F., Belcher, B.M., Xu, J., 2003. China's forests: global lessons from market reforms. Rff Press.

Irwin, E.G., 2010. New directions for urban economic models of land use change: incorporating spatial dynamics and heterogeneity. Journal of Regional Science 50, 65-91

Irwin, E.G., Geoghegan, J., 2001a. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment 85, 7-24

Irwin, E.G., Geoghegan, J., 2001b. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment 85, 7-24

Jaeger, A., 1990. Shock persistence and the measurement of prewar output series. Economics Letters 34, 333-337

Jenerette, G.D., Wu, J., 2001. Analysis and simulation of land-use change in the central Arizona – Phoenix region, USA. Landscape Ecology 16, 611-626

Jetz, W., Wilcove, D.S., Dobson, A.P., 2007. Projected impacts of climate and land-use change on the global diversity of birds. PLoS Biol 5, e157

Jiang, L., Yan, P., Wang, P., Shi, J., Yang, X., Dong, J., Han, J., Nan, R., 2006. Influence of climatic factors on safety of rice production in Heilongjiang Province. Journal of Natural Disasters 15, 46-51

Jiang, X., Gong, P., Bostedt, G., Xu, J., 2011. Impacts of Policy Measures on the Development of State-Owned Forests in Northeastern China: Theoretical Results and Empirical Evidence. Environment for Development

Jöreskog, K.G., Sörbom, D., 1986. LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods. Scientific Software, Ann Arbor.

Judson, R.A., Owen, A.L., 1999. Estimating dynamic panel data models: a guide for macroeconomists. Economics letters 65, 9-15

Kaimowitz, D., Angelsen, A., 1998. Economic models of tropical deforestation: a review. Centre for International Forestry Research, Jakarta.

Kaimowitz, D., Angelsen, A, 1998. Economic Models of Tropical Deforestation. A Review. Centre for International Forestry Research, Jakarta.

Kalirajan, K.P., Obwona, M.B., Zhao, S., 1996. A decomposition of total factor productivity growth: the case of Chinese agricultural growth before and after reforms. American Journal of Agricultural Economics 78, 331-338

Kaufman, L., Rousseeuw, P.J., 2009. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, Hoboken.

Kim, D., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S., Feng, M., Townshend, J.R., 2014. Global, Landsat-based forest-cover change from 1990 to 2000. Remote Sensing of Environment 155, 178-193

Kleibergen, F., Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics 133, 97-126

Kleijn, D., Kohler, F., Báldi, A., Batáry, P., Concepción, E., Clough, Y., Diaz, M., Gabriel, D., Holzschuh, A., Knop, E., 2009. On the relationship between farmland biodiversity and land-use intensity in Europe. Proceedings of the Royal Society of London B: Biological Sciences 276, 903-909

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963-974

Lambin, E.F., Geist, H.J., 2008. Land-use and land-cover change: local processes and global impacts. Springer Science & Business Media.

Lambin, E.F., Geist, H.J., Lepers, E., 2003. Dynamics of land-use and land-cover change in tropical regions. Annual review of environment and resources 28, 205-241

Lambin, E.F., Meyfroidt, P., 2011. Global land use change, economic globalization, and the looming land scarcity. Proceedings of the National Academy of Sciences 108, 3465-3472

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., 2001a. The causes of land-use and land-cover change: moving beyond the myths. Global environmental change 11, 261-269

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skånes, H., Steffen, W., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C., Xu, J., 2001b. The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11, 261-269

Lau, J., Ioannidis, J.P., Schmid, C.H., 1998. Summing up evidence: one answer is not always enough. The lancet 351, 123-127

Leach, M., Fairhead, J., 2000. Challenging Neo-Malthusian Deforestation Analyses in West Africa's Dynamic Forest Landscapes. Population and Development Review 26, 17-43

Li, H., Reynolds, J.F., 1993. A new contagion index to quantify spatial patterns of landscapes. Landscape Ecology 8, 155-162

Li, H., Wu, J., 2004. Use and misuse of landscape indices. Landscape Ecology 19, 389-399

Li, W., 2004. Degradation and restoration of forest ecosystems in China. Forest Ecology and Management 201, 33-41

Liu, H., Zhang, S., Li, Z., Lu, X., Yang, Q., 2004. Impacts on Wetlands of Large-scale Land-use Changes by Agricultural Development: The Small Sanjiang Plain, China. AMBIO: A Journal of the Human Environment 33, 306-310

Liu, H., Zhou, Q., 2004. Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison. International Journal of Remote Sensing 25, 1037-1050

Liu, J., Diamond, J., 2005. China's environment in a globalizing world. Nature 435, 1179-1186

Lopez, R., 1997. Environmental externalities in traditional agriculture and the impact of trade liberalization: the case of Ghana. Journal of Development Economics 53, 17-39

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., Deshazo, J., Fiebig, D., Greene, W., 2005. Recent progress on endogeneity in choice modeling. Marketing Letters 16, 255-265

Lund, H.G., 2006. Definitions of forest, deforestation, afforestation, and reforestation. Forest Information Services.

MacCallum, R.C., Austin, J.T., 2000. Applications of structural equation modeling in psychological research. Annual review of psychology 51, 201-226

Mainardi, S., 1998. An economitric analysis of factors affecting tropical and subtropical deforestation. Agrekon 37, 23-65

Mather, A.S., Needle, C.L., 2000. The relationships of population and forest trends. Geographical Journal 166, 2-13

Mather, A.S., Needle, C.L., Fairbairn, J., 1999. Environmental Kuznets Curves and Forest Trends. Geography 84, 55-65

Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G., Gotts, N.M., 2007. Agent-based land-use models: a review of applications. Landscape Ecology 22, 1447-1459

McAlpine, C.A., Eyre, T.J., 2002. Testing landscape metrics as indicators of habitat loss and fragmentation in continuous eucalypt forests (Queensland, Australia). Landscape Ecology 17, 711-728

McGarigal, K., Marks, B.J., 1995. Spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station

McGarigal, K., SA Cushman, and E Ene, 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts. Amherst

Mertens, B., Lambin, E.F., 1997. Spatial modelling of deforestation in southern Cameroon: Spatial disaggregation of diverse deforestation processes. Applied Geography 17, 143-162

Mertens, B., Lambin, E.F., 2000. Land-cover-change trajectories in southern Cameroon. Annals of the Association of American Geographers 90, 467-494

Mertens, B., Poccard-Chapuis, R., Piketty, M.G., Lacques, A.E., Venturieri, A., 2002. Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: the case of São Félix do Xingú in South Pará. Agricultural Economics 27, 269-294

Mertens, B., Sunderlin, W.D., Ndoye, O., Lambin, E.F., 2000. Impact of macroeconomic change on deforestation in South Cameroon: Integration of household survey and remotely-sensed data. World Development 28, 983-999

Millennium Ecosystem Assessment, 2005. Ecosystems and human well-being. Island Press Washington, DC.

MOF, 1997. China Forestry Yearbook 1996. China Forestry Publishing House (Ministry of Forestry), Beijing (in Chinese).

Moody, E.G., King, M.D., Platnick, S., Schaaf, C.B., Gao, F., 2005. Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products. Geoscience and Remote Sensing, IEEE Transactions on 43, 144-158

Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71, 1027-1048

Morimune, K., 1983. Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. Econometrica: Journal of the Econometric Society 51, 821-841

Muldavin, J.S., 1997. Environmental degradation in Heilongjiang: policy reform and agrarian dynamics in China's new hybrid economy. Annals of the Association of American Geographers 87, 579-613

Mullan, K., Kontoleon, A., Swanson, T., Zhang, S., 2009. An evaluation of the impact of the Natural Forest Protection Programme on Rural Household Livelihoods. In: An Integrated Assessment of China's Ecological Restoration Programs. Springer, pp. 175-199.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica: Journal of the Econometric Society 46, 69-85

Munroeaic, D.K., Southworth, J., Tucker, C.M., 2002. The dynamics of land-cover change in western Honduras: exploring spatial and temporal complexity. Agricultural Economics 27, 355-369

Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. The journal of economic perspectives 20, 111-132

Nagendra, H., 2002. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. Applied Geography 22, 175-186

Nelson, G.C., Geoghegan, J., 2002. Deforestation and land use change: sparse data environments. Agricultural Economics 27, 201-216

Nelson, G.C., Hellerstein, D., 1997. Do roads cause deforestation? Using satellite images in econometric analysis of land use. American Journal of Agricultural Economics 79, 80-88

NFPP Management Center, 2011. Authoritative interpretations for the second phase policies of natural forest protection project

Nickell, S., 1981. Biases in Dynamic Models with Fixed Effects. Econometrica 49, 1417-1426

O'Neill, R.V., Krummel, J.R., Gardner, R.H., Sugihara, G., Jackson, B., DeAngelis, D.L., Milne, B.T., Turner, M.G, Zygmunt, B., Christensen, S.W., Dale, V.H. and Graham, R.L., 1988. Indices of landscape pattern. Landscape Ecology. Landscape Ecology 1, 153-162

Pacheco, P., 2006. Agricultural expansion and deforestation in lowland Bolivia: the import substitution versus the structural adjustment model. Land Use Policy 23, 205-225

Pan, W.K., Walsh, S.J., Bilsborrow, R.E., Frizzelle, B.G., Erlien, C.M., Baquero, F., 2004. Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. Agriculture, Ecosystems & Environment 101, 117-134

Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J., Deadman, P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. Annals of the Association of American Geographers 93, 314-337

Pearl, J., 2000. Causality: models, reasoning and inference. Cambridge University Press, Cambridge.

Pfaff, A.S., 1999a. What Drives Deforestation in the Brazilian Amazon? Journal of Environmental Economics and Management 37, 2643

Pfaff, A.S., 1999b. What drives deforestation in the Brazilian Amazon?: evidence from satellite and socioeconomic data. Journal of Environmental Economics and Management 37, 26-43

Pielou, E.C., 1975. Ecological Diversity. Wiley-Interscience, New York.

Pitt, M.M., 2011. Overidentification tests and causality: a second response to Roodman and Morduch. Brown University, http://www. pstc. brown. edu/~ mp/papers/Overidentification. pdf

Pontius Jr, R.G., Shusas, E., McEachern, M., 2004. Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems & Environment 101, 251-268

Post, W.M., Kwon, K.C., 2000. Soil carbon sequestration and land-use change: processes and potential. Global change biology 6, 317-327

Potere, D., 2008. Horizontal positional accuracy of Google Earth's high-resolution imagery archive. Sensors 8, 7973-7981

Qian, Y., 2000. The process of China's market transition (1978-1998): The evolutionary, historical, and comparative perspectives. Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft 156, 151-171

Railsback, S.F., Lytinen, S.L., Jackson, S.K., 2006. Agent-based simulation platforms: Review and development recommendations. Simulation 82, 609-623

Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. Statistical science, 15-32

Rosenfield, G.H., Fitzpatrick-Lins, K., 1986. A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric engineering and remote sensing 52, 223-227

Rudel, T., Roper, J., 1997. The paths to rain forest destruction: Crossnational patterns of tropical deforestation, 1975–1990. World Development 25, 53-65

Rudel, T.K., Horowitz, B., 1993. Tropical deforestation: Small farmers and land clearing in the Ecuadorian Amazon. Columbia University Press.

Sanderson, E., Windmeijer, F., 2013. A weak instrument F-test in linear IV models with multiple endogenous variables. CEMMAP working paper, Centre for Microdata Methods and Practice

Sandler, T., 1993. Tropical Deforestation: Markets and Market Failures. Land Economics 69, 225-233

Schaffer, M.E., 2012. xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. Statistical Software Components

Schmidheiny, K., Basel, U., 2011. Panel Data: Fixed and Random Effects. URL http://www.schmidheiny.name/teaching/panel2up.pdf

Schneider, L.C., Pontius, R.G., 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment 85, 83-94

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., Yu, T.-H., 2008. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319, 1238-1240

Semykina, A., Wooldridge, J.M., 2010. Estimating panel data models in the presence of endogeneity and selection. Journal of Econometrics 157, 375-380

SFA, 2000. Statistics on the national forest resources (the 5th National Forest Inventory 1994-1998). State Forestry Administration, Beijing (in Chinese).

SFA, 2005. Statistics on the national forest resources (the 6th National Forest Inventory 1999-2003). State Forestry Administration, Beijing (in Chinese).

Sliva, L., Williams, D.D., 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water research 35, 3462-3472

Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A., 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote sensing of Environment 75, 230-244

Song, K., Liu, D., Wang, Z., Zhang, B., Jin, C., Li, F., Liu, H., 2008. Land use change in Sanjiang Plain and its driving forces analysis since 1954. Acta Geographica Sinica (Chinese Edition) 63, 81-93

Song, k., Wang, Z., Liu, Q., Lu, D., Yang, G., Zeng, L., Liu, D., Zhang, B., Du, J., 2009a. Land use/land cover (LULC) characterizaitoin with MODIS time series data in the Amu River Basin. In: Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, pp. IV-310 - IV-313

Song, k., Wang, Z., Liu, Q., Lu, D., Yang, G., Zeng, L., Liu, D., Zhang, B., Du, J., 2009b. Land use/land cover (LULC) characterizaitoin with MODIS time series data in the Amu River Basin. In: Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, pp. IV-310-IV-313

Staiger, D.O., Stock, J.H., 1994. Instrumental variables regression with weak instruments. Econometrica 65, 557-586

StataCorp., 2013. reg3 postestimation — Postestimation tools for reg3. Stata Press, College Station, Texas.

Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment 62, 77-89

Stevens Jr, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99, 262-278

Stock, J.H., Wright, J.H., 2000. GMM with weak identification. Econometrica 68, 1055-1096

Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics 20, 518-529

Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. Identification and inference for econometric models: Essays in honor of Thomas Rothenberg

Strasser, U., Mauser, W., 2001. Modelling the spatial and temporal variations of the water balance for the Weser catchment 1965–1994. Journal of Hydrology 254, 199-214

Sun, Q., Zhang, S., Zhang, J., Yang, C., 2010. Current Situation of Rice Production in Northeast of China and Countermeasures. North Rice 2, 2-32

Swamy, P., Arora, S.S., 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. Econometrica: Journal of the Econometric Society 40, 261-275

Tang, J., Wang, L., Zhang, S., 2005. Investigating landscape pattern and its dynamics in Daqing, China. International Journal of Remote Sensing 26, 2259-2280

Taylor, J.E., Adelman, I., 2003. Agricultural household models: Genesis, evolution, and extensions. Review of Economics of the Household 1, 33-58

Tobler, W., 1979. Cellular geography. In: Philosophy in geography. Springer, pp. 379-386.

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46, 234-240

Todd, P.E., Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement. The Economic Journal 113, F3-F33

Tong, S.T., Chen, W., 2002. Modeling the relationship between land use and surface water quality. Journal of environmental management 66, 377-393

Turner, B.L., Lambin, E.F., Reenberg, A., 2008a. Land Change Science Special Feature: The emergence of land change science for global environmental change and sustainability. Proceedings of the National Academy of Sciences of the United States of America 105, 2751-2751

Turner, B.L., Lambin, E.F., Reenberg, A., 2008b. Land Change Science Special Feature: The emergence of land change science for global environmental change and sustainability (vol 104, pg 20666, 2007). Proceedings of the National Academy of Sciences of the United States of America 105, 2751-2751

Turner, M.G., 1989. Landscape ecology: the effect of pattern on process. Annual review of ecology and systematics, 171-197

Turner, M.G., 1990. Spatial and temporal analysis of landscape patterns. Landscape Ecology 4, 21-30

Turner, M.G., Wear, D.N., Flamm, R.O., 1996. Land ownership and land-cover change in the southern Appalachian highlands and the Olympic peninsula. Ecological applications 6, 1150-1172

U.S. Department of the Interior, 2009. U.S. Geological Survey.

Ullman, J.B., Bentler, P.M., 2001. Structural equation modeling. John Wiley & Sons, Hoboken.

Vachaud, G., Passerat de Silans, A., Balabanis, P., Vauclin, M., 1985. Temporal stability of spatially measured soil water probability density function. Soil Science Society of America Journal 49, 822-828

Van Soest, Daan P., Bulte, Erwin H., Angelsen, A., Van Kooten, G.C., 2002. Technological change and tropical deforestation: a perspective at the household level. Environment and Development Economics 7, 269-280

Vanclay, J.K., 1993. Saving the tropical forest : needs and prognosis. Ambio 22, 225-231

Varian, H.R., 2009. Intermediate Microeconomics: A Modern Approach. W. W. Norton & Company, New York City.

Verburg, P., Schot, P., Dijst, M., Veldkamp, A., 2004a. Land use change modelling: current practice and research priorities. GeoJournal 61, 309-324

Verburg, P.H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004b. Land use change modelling: current practice and research priorities. GeoJournal 61, 309-324

Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., Mastura, S.S., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. Environmental management 30, 391-405

Vincent, J.R., 1990. Don't boycott tropical timber. Journal of Forestry 88, 56

Vojáček, O., Pecáková, I., 2010. Comparison of Discrete Choice Models for Economic Environmental Research. Prague Economic Papers 19, 35-53

Walker, R., Perz, S., Caldas, M., Silva, L.G.T., 2002. Land use and land cover change in forest frontiers: The role of household life cycles. International Regional Science Review 25, 169-199

Wang, G., Innes, J.L., Lei, J., Dai, S., Wu, S.W., 2007. China's Forestry Reforms. Science 318, 1556-1557

Wang, L., Lyons, J., Kanehl, P., Gatti, R., 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22, 6-12

Wang, S., Cornelis van Kooten, G., Wilson, B., 2004. Mosaic of reform: forest policy in post-1978 China. Forest Policy and Economics 6, 71-83

Wang, T., 2008. Effective measurements to tackle the problem of rising price. URL http://paper.people.com.cn/rmlt/html/2008-05/16/content_48573240.htm

Wang, Z., Liu, Z., Song, K., Zhang, B., Zhang, S., Liu, D., Ren, C., Yang, F., 2009. Land use changes in Northeast China driven by human activities and climatic variation. Chinese Geographical Science 19, 225-230

Wang, Z., Song, K., Ma, W., Ren, C., Zhang, B., Liu, D., Chen, J.M., Song, C., 2011. Loss and fragmentation of marshes in the Sanjiang Plain, Northeast China, 1954–2005. Wetlands 31, 945-954

Wang, Z., Zhang, B., Zhang, S., Li, X., Liu, D., Song, K., Li, J., Li, F., Duan, H., 2006. Changes of land use and of ecosystem service values in Sanjiang Plain, Northeast China. Environmental Monitoring and Assessment 112, 69-91

White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban Systems 24, 383-400

Wickham, J., Rhtters, K., 1995. Sensitivity of landscape metrics to pixel size. International Journal of Remote Sensing 16, 3585-3594

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics 126, 25-51

Wooldridge, J.M., 1996. Estimating systems of equations with different instruments for different equations. Journal of Econometrics 74, 387-405

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge.

Wooldridge, J.M., 2003. Cluster-sample methods in applied econometrics. American Economic Review 93, 133-138

Wooldridge, J.M., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of applied econometrics 20, 39-54

Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data. The MIT press, Cambridge.

Wooldridge, J.M., 2012. Introductory econometrics: A modern approach. Cengage Learning, Boston.

Xu, J., Tao, R., Amacher, G.S., 2004. An empirical analysis of China's state-owned forests. Forest Policy and economics 6, 379-390

Xu, J., Yin, R., Li, Z., Liu, C., 2005. China's ecological rehabilitation: the unprecedented efforts and dramatic impacts of reforestation and slope protection in western China. Ecological Economics 57, 595-607

Xu, J., Yin, R., Li, Z., Liu, C., 2006a. China's ecological rehabilitation: Unprecedented efforts, dramatic impacts, and requisite policies. Ecological Economics 57, 595-607

Xu, J., Yin, R., Li, Z., Liu, C., 2006b. China's ecological rehabilitation: the unprecedented efforts and dramatic impacts of reforestation and slope protection in western China. Ecological Economics 57, 595-607

Yamane, M., 2001a. China's Recent Forest-Related Policies: Overview and Background. Policy Trend Report 1, 1-12

Yamane, M., 2001b. China's recent forest-related policies: Overview and background. Policy Trend Report 1, 1-12

Yan, M., Deng, W., Chen, P., 2001. Climate variation in the Sanjiang Plain disturbed by large scale reclamation during the last 45 years. ACTA GEOGRAPHICA SINICA-CHINESE EDITION- 56, 170-179

Yan, M., Deng, W., Chen, P., 2002. Climate change in the Sanjiang Plain disturbed by large-scale reclamation. Journal of Geographical Sciences 12, 405-412

Yin, R., 1998. Forestry and the environment in China: the current situation and strategic choices. World Development 26, 2153-2167

Yin, R., Xiang, Q., 2010. An integrative approach to modeling land-use changes: multiple facets of agriculture in the Upper Yangtze basin. Sustainability Science 5, 9-18

Yin, R., Xu, J., Li, Z., 2003. Building institutions for markets: Experiences and lessons from China's rural forest sector. Environment, Development and Sustainability 5, 333-351

Yin, R., Yin, G., 2009. China's Ecological Restoration Programs: Initiation, Implementation, and Challenges. In: An Integrated Assessment of China's Ecological Restoration Programs. Springer Netherlands, pp. 1-19.

Yin, R., Yin, G., 2010. China's primary programs of terrestrial ecosystem restoration: initiation, implementation, and challenges. Environmental management 45, 429-441

Yu, D., Zhou, L., Zhou, W., Ding, H., Wang, Q., Wang, Y., Wu, X., Dai, L., 2011. Forest management in Northeast China: history, problems, and challenges. Environmental management 48, 1122-1135

Yun, Y., Fang, X., Wang, Y., Tao, J., Qiao, D., 2005. Main grain crops structural change and its climate background in Heilongjiang province during the past two decades. Journal of Natural Resources 20, 697-704

Zellner, A., Theil, H., 1992. Three-stage least squares: Simultaneous estimation of simultaneous equations. In: Henri Theil's Contributions to Economics and Econometrics. Springer, pp. 147-178.

Zhang, B., Cui, H., Yu, L., He, Y., 2003. Land reclamation process in northeast China since 1900. Chinese Geographical Science 13, 119-123

Zhang, J., Ma, K., Fu, B., 2010. Wetland loss under the impact of agricultural development in the Sanjiang Plain, NE China. Environmental monitoring and assessment 166, 139-148

Zhang, K., Hori, Y., Zhou, S., Michinaka, T., Hirano, Y., Tachibana, S., 2011. Impact of Natural Forest Protection Program policies on forests in northeastern China. Forestry Studies in China 13, 231-238

Zhang, P., Shao, G., Zhao, G., Le Master, D.C., Parker, G.R., Dunning Jr, J.B., Li, Q., 2000. China's forest policy for the 21st century. Science 288, 2135-2136

Zhang, S., Na, X., Kong, B., Wang, Z., Jiang, H., Yu, H., Zhao, Z., Li, X., Liu, C., Dale, P., 2009. Identifying wetland change in China's Sanjiang Plain using remote sensing. Wetlands 29, 302-313

Zhang, Y., 2000. Costs of Plans vs Costs of Markets: Reforms in China's State-owned Forest Management. Development Policy Review 18, 285-306

Zhang, Y., 2001. Deforestation and forest transition: theory and evidence in China. In: Palo M & Vanhanen H (eds.) World forests from deforestation to transition? Springer, Netherlands, pp. 41-65.

Zhang, Y., Dai, G., Huang, H., Kong, F., Tian, Z., Wang, X., Zhang, L., 1999. The forest sector in China: Towards a market economy. In: World forests, society and environment. Springer, pp. 371-393.

Zhang, y., Li, z., Jiang, l., 2012. Measures on Forest Right System Reform of Local State-Owned Forest Farm in Heilongjiang Province. China Forestry Economy 112, 35-48

Zhao, G., Shao, G., 2002. Logging Restrictions in China: A Turning Point for Forest Sustainability. Journal of Forestry 100, 34-37

Zhou, D., Gong, H., Wang, Y., Khan, S., Zhao, K., 2009. Driving forces for the marsh wetland degradation in the Honghe National Nature Reserve in Sanjiang Plain, Northeast China. Environmental Modeling & Assessment 14, 101-111

CHAPTER 6

SUMMARY, LIMITATIONS, AND FUTURE WORK

6.1 Motivations, Tasks, and Hypotheses

The initial question of my dissertation research was whether or not the large Natural Forest Protection Program (NFPP) had been effective in protecting the natural forests in northeast China, upon which region the program had been heavily concentrated. Quickly, my literature review revealed that studies of the effectiveness and impact of the NFPP were not many, and they tended to focus on the short-term outcomes. Few had put it in an adequate historical context and examined the regional depletion and possible recovery of natural forests, and even fewer had investigated the major factors affecting the resource dynamics from a more holistic view. Further, previous studies of the deforestation determinants were mainly based on forest census statistics, despite the common concern that these statistics would be less comprehensive and of poor quality. These observations motivated me to consider the NFPP from the perspective of land-use and land-cover change over a longer time span.

Therefore, I decided to investigate the land conversions in the Sanjiang Plain area of Heilongjiang and their driving forces, with a focus on the forestland dynamics. Accordingly, the two tasks of my dissertation were set to detect the regional LUCC over a period of 30 years (1977-2007) and to explore the demographic, economic, political, and other determinants of the detected changes. I had three chief hypotheses to test. First, the region had suffered severe deforestation and forest degradation before the NFPP was initiated. Second, while the decline of forest cover might have been slowed down following the NFPP implementation, it would take a longer time and more effective efforts to see any significant gain. Third, farmland expansion is a primary direct driver of deforestation, whereas population increase, economic growth, and management policy are among the more fundamental drivers.

I will report the main findings of my LUCC detection in the next section. Then, I will summarize my modeling approaches, data treatment, and empirical results in section 6.3. Finally, limitations of my research and future directions will be discussed in section 6.4.

6.2 Main Findings of Land-Use Change Detection

I selected 10 adjacent counties as my study site, based on considerations of relevance and feasibility. This area used to have large tracks of natural forests and wetland, but it experienced heavy logging and farming expansion during the second half of last century. Then, Landsat images for six periods were gathered to derive the LUCC information. Before interpretation, the images were corrected and enhanced. Next, unsupervised classification was conducted according to the USGS Classification System. In light of the overall regional land-use structure and my fieldwork knowledge, minor classes of land use were merged, resulting in four classes—forestland, farmland, built-up land and other (wetland being a main component).

Later, accuracy assessment was performed before my arrival at the final detection results. Because validating long-term image classification is always problematic due to the unavailability of simultaneous reference data, a rule-based rationality evaluation was taken as a preliminary step. Subsequently, a formal accuracy assessment was performed with the spatially balanced sampling method. Using a sample of 1550 points for each period of time, the accuracy rates for the six periods are all around 85% and thus acceptable.

As reported in Chapter 2, it is found that: (1) farmland and forestland are the two predominant classes of the regional land use; (2) farmland and built-up land increased persistently during the 30 years; (3) forestland suffered an extended, heavy loss before the end of last century and the decline slowed down significantly thereafter; and (4) "other" land declined continuously.

Meanwhile, by taking the relative land use sizes into consideration, I calculated the extended conversation matrixes, which present a clear picture of the LUCC. As large as the forestland loss may be, it was not the first to be encroached by farmland expansion; instead, local farmers targeted "other" land, especially wetland, first, for farming. Additionally, my work discovered that reforestation as well as agroforestry in the farmland-dominant counties became prevalent after the NFPP was initiated. Moreover, the calculated landscape diversity and integrity indexes show that the distribution of land-cover types became more uneven, and land-use patches became more interspersed.

In short, these findings are interesting and important in and of themselves. They also make it likely and feasible for me to undertake the other task of my research—analyzing the deriving forces of the regional LUCC in general and deforestation in particular.

6.3 Analysis of the LUCC Driving Forces

Modeling Approaches

With a satisfactory generation of the regional LUCC data for my study site, I was excited to embark on studying the determinants of the LUCC, especially those of the deforestation. I started with an extensive review of the relevant literature, which has been rapidly growing since the 1990s. As documented in Chapter 3, LUCC driving force analysis can be done with an analytic approach, a simulation approach, and/or a regression approach. Given the advantages and disadvantages of these approaches, as well as my academic background of and interest in applied economics, I decided to take the regression approach. There can be single-equation regression models or system of equations regression models reveals, and these models have their own

strengths and weaknesses, in addition to their particular data requirements and estimation techniques.

Taking all these factors into account, I decided to develop and estimate both kinds of regression models in my empirical analysis. Furthermore, my literature review indicates that deforestation is largely driven by a combination of three proximate factors—wood extraction, farming expansion, and infrastructure development. These proximate factors are in turn mediated by a whole host of more fundamental forces, including demographic change, economic growth, and institutional, policy and market factors.

Data Treatment

I had three options in compiling the dataset needed for analyzing the regional LUCC driving forces. The first option was to do a pixel-level analysis, which could give rise to a large number of observations, allowing the adoption of various econometric strategies and estimation methods. However, the fundamental problem with that option is that LUCC is a social-economic phenomenon, which is not organized at the pixel level. The unit of my observation and analysis should thus be some socioeconomic organization, be it household, community, township, county, municipality, or province. That's why I chose to do my LUCC detection and driving force determination at the county level from the beginning.

Another straightforward option would be to combine the repeated cross-sectional LUCC data that I had obtained from my first task and the corresponding social-ecological data that I had gathered from existing sources. While this dataset consists of original observations at the appropriate level, the sample size is small—only 48 observations (8 counties and 6 intermittent points of time). Given the limited degree of freedom, relying solely on this small dataset would make me severely handicapped in addressing issues like spatial and temporal correlations and to

obtain stable and reliable results. Certainly, it would not permit me to take advantage of the more advanced modeling frameworks or estimation techniques in dealing with potential endogeneity and simultaneity.

The other option was to interpolate the LUCC data for the missing years between nearby two points of time in the 31 years and then integrate the annualized LUCC information with the existing annual social-ecological data to form a panel dataset of 248 observations. With the available LUCC data in about every five years, an interpolation would be easy and reasonable. Of course, someone may wonder why I did not do my LUCC detection for more cross-sections and/or more points of time over the whole period of study. But that would be a huge amount of work, which is unfortunately beyond the reach of my dissertation project. On the other hand, the interpolated and integrated dataset could open up some substantial analytic opportunities as what I have alluded to above. So, I decided to pursue it as part of my analysis of the LUCC determinants. Below, I will synthesize my modeling efforts and findings first; then, I will discuss the effects of this data treatment.

Empirical Findings

I undertook multiple single-equation regressions in Chapter 4, whereby fixed effects and random effects estimators were considered. Indeed, I began with simple specifications of single-equation models to explore the possibilities and pitfalls of the two datasets (one with the original 48 observation and the other with the 248 observations derived through interpolation). Several useful messages emerged from this preliminary exploration. First, the results of fixed-effects analysis are more reliable than those of random-effects analysis. Second, it seems problematic to directly incorporate farmland expansion as a repressor in explaining deforestation, for example, potential endogeneity. Endogeneity could result in biased coefficient estimates. Third, the counties

under study varied a lot in their land resource endowment, leading to the inapplicability of traditional homoscedastic standard error in this study. As such, adopting the heteroskedastic robust standard errors is a basic regression requirement.

The results of estimated single-equation models demonstrated that farmland expansion and population growth are significantly correlated with deforestation. The coefficients of distance to market and number of forest farms are significantly positive. Meanwhile, the NFPP effect, while having the correct sign, is insignificant. Also, the coefficient of timber price is insignificant. It should be further noted that given the small cross sections (8 counties only), spatial correlation was impractical to capture the potential spatial correlation. And when the temporal correlation was considered, the outcomes were mixed; some of the coefficients got improved (e.g., NFPP) while others (e.g., farmland) became not as strong. Therefore, caution is called for in interpreting the estimated results.

Then, in Chapter 5, I adopted the instrumental variable method and a system of simultaneous equations model to incorporate the interaction and feedback effects between different land uses in an attempt to improve my empirical results. The outcomes of using the instrumental variable method to deal with the potential endogeneity embedded in farmland were much improved—the coefficients of NFPP and timber price are significant, implying that the program has played a positive role in protecting local forests. The bias associated with instrument variable analysis is smaller than those with the OLS estimation. In addition, the coefficient estimates of the 3SLS estimation of the system are generally consistent with those derived from the IV method. The area of wetland is negatively correlated with the area of forestland—a mutual substitution in farmland expansion; likewise, farmland is negatively correlated with wetland. The significantly positive coefficient of built-up area in the farmland equation suggests a strong tie between farming

activities and residential construction. The significant negative coefficient of irrigation confirms that wetland loss is adversely affected by the change in local cropping structure. There and other findings carry some interesting policy implications.

6.4 Limitations and Future Work

Overall, different estimation strategies have allowed me to compare the performances of alternative regression models of the LUCC driving forces, and these alternative regression models have corroborated the consistency of my empirical results. These are encouraging outcomes and they should help mitigate the concerns with my data interpolation as well as the limited number of observations in my sample. At the same time, I must admit that the two datasets I have put together do have limitations. First, as noted, I was unable to capture any of the potential spatial correlation, and I was unable to adequately capture the temporal correlation. Second, while I was able to develop more sophisticated models and use more advance estimation techniques based on the long panel dataset with interpolated observations, the small sample size made the estimated results sometimes sensitive to the modeling framework used and assumptions made. Further, I had to ignore potential time lags between dependent and independent variables due to the limited degree of freedom. So, caution is needed in interpreting the estimated results.

It is hoped that future research will be able to overcome these problems. Accumulating longer time-series and larger cross-sectional data will be a fundamental undertaking in order to accommodate more advance econometric tools and frameworks to derive more robust empirical results. Also, the quality of LUCC and other social-ecological data should be carefully scrutinized and, if possible, data with higher quality and reliability should be incorporated into the datasets. Moreover, data for other relevant variables, such as changes in the ecological conditions induced

by implementing the NFPP, should be collected or updated. To pursue these activities, it becomes

essential to develop strong collaboration with other scholars. I am confident that these steps will

go a long way in advancing research agenda along the direction that I have embarked on.

APPENDIX

APPENDIX

Model Validation and Model Limitations

Model Validation

Model validation is an important step in the model building. I employed methods like different formal hypothesis tests, descriptive statistics and graphic checks to validate the different model sets.

Variable filtering is an important step in the model building sequence. In order not to include extra and unnecessary terms, and to minimize the effects of the potential high prevalence of correlated predictors in ecological and socioeconomic dataset, even though I did primitive correlation related analysis, different examining approaches were further carried out in order to reach a concise but still powerful model. In the single equation model, tests of individual parameters and the information criteria of AIC and BIC help exclude the variable of annual output value of forestry sector and provide foundations for all the other predictors are included in the model. In the instrumental variable based two system model, various statistical tests were used to check whether there is overfitting, or over-identification situation. The statistical tests effectively ruled out all the other instrument candidates while only keeping built-up land as the one and only effective instrument. Due to the nested nature in the "Forestland-Farmland-Wetland" model, typical information criteria are not applicable. As a compromise, I examined equations one by one, thus, the final model started with relatively simple and have a few terms and most of them were turned out to be significant in the final estimation results.

Meanwhile, In order to test whether there is omitted variable problem or model misspecifications in the functional part of deforestation model, series of different models which

has different emphasis are considered. Cases happened that even some models had little explanatory power, they provided evidences and hints for modelling specification from different perspectives. For example, the failure of between-effects model layer a strong foundation for fixed-effects analysis, and the significance of the coefficient of the mean value of farmland in the Mundlak model lead me to test of the hypothesis that the single equation model is insufficient, and possibly the variable of farmland subjects to further exploration, e.g. endogeneity. Thus, exploring the applicable models provide the rich feedbacks for appropriate selecting a rigorous analysis as well as for identifying potential limits in the functional part of the model.

Model limitations

As forestland were largely replaced by farmland, though some turned into eroded and barren land (Muldavin 1997). Thus how to incorporate farmland expansion as an important causal impact need further consideration. And this may point to the single equation models in Chapter 4 suffering from problems in directly employing farmland as regressor for explaining the deforestation causes. The following on Chapter 5 partly remedied this problem by incorporating instrument variables analysis and simultaneous equation modelling. For both estimation procedures, fitted values for the variable farmland (and wetland) were estimated through reduced-form equations which is explained by the instrument variables (built-up land) and/or the underlying driving forces (Wooldridge 1996). Therefore, the instruments and driving forces in the farmland expansion and wetland loss equations as well as in the forestland loss are the true variables which have effects on deforestation. So these two methods not only just addressed the endogeneity issue, they also played an important role in mediating the problem of explaining deforestation by farmland expansion, as well as examining the indirect or spillover effects on deforestation that were induced by farmland and wetland changes.

The exogeneity of "built-up land" was based on the perspective of land conversion trend that built-up land converted from farmland and both land uses increased a lot with a high correlation and the few land interactions between forestland and built-up land. The validity of the instrument is well grounded based on land use studies, and also under strict inspection by a comprehensive statistical tests during different estimation stages, like endogeneity test, under-identification test, weak identification test, and over-identification test. The instrument built-up land passed all the examination tools. However, the model's statistical validated power in mitigating the biases that ordinary least squares estimation suffers when a troublesome explanatory farmland is correlated with the disturbances. It is suggested that the two stage least square estimator is not sufficiently robust when testing candidate instrument that potentially is not strong enough (Murray 2006). The author suggests using conditional likelihood ratio techniques and Fuller's estimators in over-identified models due to the good properties which is regard would not be diluted by weak instruments. As variable inspection procedure excluded all the other candidates, the model is an exact-identified case. In this dissertation I have only compared the efficiency given the set of instruments that in the framework and within my research sight, and I still keep the suspects for the instrument validity as it is well known that how hard it is to find an appropriate instrument.

The intrinsic interactions between the land use classes "Forestland-Farmland-Wetland" lead to the systematic analysis, and model specification further strength the nature of related equation. It was found that formal validation based on such nested models were limited and it has received little attention despite now gradually being applied to different research areas (Al-Tuwaijri et al. 2004; Herbert & Arild 2009; Yin & Xiang 2010). Mathematical computation based

215

on the errors in the post-estimation stage supports the legitimacy of using 3SLS analysis, and carrying out the Breusch-Pagan LM diagonal covariance matrix test confirm the existence of the correlations and feedback effects between different models.

Due to the small sample size, at the post-estimation process, I took a further step to simplify the big model by carry out sensitivity analysis. Dropping out regressors with relatively weaker in capturing the explanatory variance, this procedure has led to a more concise model while still keeping the explanatory power and model integrity. Meanwhile, there exists practical obstacles for examining the model fit in the "Forestland-Farmland-Wetland" system through partitioning dataset into two parts and comparing the forecasted and observed differences in forestland. The graphical checks of comparisons based on the predicted and observed values are quantitative and informative, which support the conclusion that the explained variations effectively captures the land dynamic trends for most counties.

An important issue needs to be raised is the time series problem in the "Forestland-Farmland-Wetland" model. As cautioned by Wooldridge (1996), when the instruments (included and excluded) are specified for each equation, dependence in the data exacerbates three stage least square estimation as the assumption that no temporal correlations is violated in the possible situation that instrument correlate with the errors. So, for future study, the explanatory variables should be further examined.

Many land use studies utilize such periodic sampling frequency and different interpolation methods were employed, while the effects of interpolation on the time series properties and statistical inferences were not much examined (Vachaud et al. 1985; Jenerette & Wu 2001; Strasser & Mauser 2001; Moody et al. 2005; Hoek et al. 2008; Song et al. 2008). Jaeger (1990) suggests

that segmented linear trend interpolation for constructing U.S. prewar output series may cause ambiguous findings. Subsequently, Dezhbakhsh and Levy (1994) linearly interpolated trend stationary series, data exhibits significant periodic variation. Though the land use data differs a lot from economic data, their research implications are helpful that estimates form the conventional time series methods would biased upward and corresponding inferences are not reliable.

REFERENCES

REFERENCES

Al-Tuwaijri, S.A., Christensen, T.E., Hughes, K., 2004. The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach. Accounting, organizations and society 29, 447-471

Dezhbakhsh, H., Levy, D., 1994. Periodic properties of interpolated time series. Economics Letters 44, 221-228

Herbert, A.J., Arild, A., 2009. The paradox of household resource endowment and land productivity in Uganda. In: Agricultural Economists Conference, Beijing

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric environment 42, 7561-7578

Jaeger, A., 1990. Shock persistence and the measurement of prewar output series. Economics Letters 34, 333-337

Jenerette, G.D., Wu, J., 2001. Analysis and simulation of land-use change in the central Arizona – Phoenix region, USA. Landscape Ecology 16, 611-626

Moody, E.G., King, M.D., Platnick, S., Schaaf, C.B., Gao, F., 2005. Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products. Geoscience and Remote Sensing, IEEE Transactions on 43, 144-158

Muldavin, J.S., 1997. Environmental degradation in Heilongjiang: policy reform and agrarian dynamics in China's new hybrid economy. Annals of the Association of American Geographers 87, 579-613

Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. The journal of economic perspectives 20, 111-132

Song, K., Liu, D., Wang, Z., Zhang, B., Jin, C., Li, F., Liu, H., 2008. Land use change in Sanjiang Plain and its driving forces analysis since 1954. Acta Geographica Sinica (Chinese Edition) 63, 81-93

Strasser, U., Mauser, W., 2001. Modelling the spatial and temporal variations of the water balance for the Weser catchment 1965–1994. Journal of Hydrology 254, 199-214

Vachaud, G., Passerat de Silans, A., Balabanis, P., Vauclin, M., 1985. Temporal stability of spatially measured soil water probability density function. Soil Science Society of America Journal 49, 822-828

Wooldridge, J.M., 1996. Estimating systems of equations with different instruments for different equations. Journal of Econometrics 74, 387-405

Yin, R., Xiang, Q., 2010. An integrative approach to modeling land-use changes: multiple facets of agriculture in the Upper Yangtze basin. Sustainability Science 5, 9-18