Litigation in China: Development, Drivers, and Forecasts

By

Douglas J. Bujakowski

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Risk and Insurance)

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

Date of final oral examination: 07/06/2017

The dissertation is approved by the following members of the Final Oral Committee:

Joan Schmit, Professor, Risk and Insurance
Ty Leverty, Associate Professor, Risk and Insurance
Peng Shi, Associate Professor, Risk and Insurance
Justin Sydnor, Associate Professor, Risk and Insurance
John Ohnesorge, Professor, Law

Dissertation Overview

The dissertation consists of three individual papers, which share a common theme; they enhance our understanding of litigation in China. Each paper focuses on a different element of this understanding. Paper 1 provides a review of the development and current state of Chinese legal institutions. The utilization of those institutions to resolve legal disputes is examined in paper 2. In paper 3, methods to predict future utilization patterns are investigated. A short summary of each paper and its connection to the other pieces can be found below.

Paper 1: Legal Liability in China: A Brief History and Current Practice Page 1

Paper 1 is a historical piece in which the development of civil law in China is examined, and current issues with the implementation of legal rules are discussed. We consider the connection between economic development and legislative change in various jurisdictions, including China. We also discuss China's enduring legal traditions and their potential implications for legislation and litigation. Our findings help to motivate the specific economic, social, and legal factors considered in paper 2.

Paper 2: Litigation in China: Development and Drivers Page 15

Paper 2 is an empirical piece in which the utilization of China's legal institutions is investigated. We observe changes in litigation rates across provinces and time, and examine the relationship between those changes and a variety of economic, social, and legal factors. We then investigate how these relationships vary according to a province's level of development. This analysis leads us to inquire whether statistically meaningful variables are also useful predictors of litigation rates, the focus of paper 3.

Paper 3: Forecasting Litigation Rates in China Page 64

We develop a predictive model for litigation rates in paper 3. We employ traditional techniques to forecast province-level litigation rates and then investigate whether forecast accuracy can be improved by incorporating variables shown to be correlated with litigation rates. This analysis provides further insight into variables identified in papers 1 and 2.

Legal Liability in China: A Brief History and Current Practice

Abstract

We examine the historical development of civil law in China and discuss current issues with the implementation of legal rules. We observe that economic development has been closely linked with legislative change, but that at times, long-standing legal traditions undermine written law, dampening the impact of such legislation. As a result, legal rights and responsibilities for individuals and organizations remain uncertain.

I. Introduction

China, with its enormous population and rapid economic growth, has been of great interest to academics, business people, and public policy makers for decades. One interesting avenue of consideration is the development of legal institutions alongside China's economic expansion. China's history differs somewhat dramatically from that of other nations, setting its legal and political institutions on a unique foundation. Furthermore, China's legal institutions are developing in a global environment with established successful institutions, which serve as examples to Chinese policymakers of what has worked and what has not worked elsewhere. For all of these reasons, much can be gained from understanding the conditions and evolution of China's legal environment.

This paper sets out to contribute to that understanding. Specifically, the purpose of this article is to identify China's legal foundations, to explore China's legal development, and to discuss obstacles to the utilization of China's current judicial system. The results are important for any enterprise having direct and indirect business ties with Chinese institutions.

To achieve our goals, we first highlight key legal traditions in China which lay the foundation for today's practices, including the presence of a strong community-based system of mediation and a high degree of government influence in the legal sphere. We then discuss the importance of economic growth as a driver of legal development and the adoption of legal rules from other jurisdictions. Next, we present various obstacles faced today by parties wishing to use legal processes for dispute resolution, including specifics detailed in two high-profile injury cases. These experiences demonstrate the extent to which Chinese legislation on the books differs from law in practice.

II. Legal Traditions

Chinese culture, including legal traditions, are among the oldest known in history. Two traditions in particular are relevant as we discuss China's current legal systems: (1) the presence of a strong community-based system of mediation and (2) a high degree of government influence in the legal sphere. These traditions, despite their ancient history, have influenced China's recent legal development and remain important element's of China's current systems.

Community mediation in China dates back to China's imperialist era (221 BC – 1911 AD). During this period, emperors over many centuries were vested with power in all arenas, including the legal sector. The law itself focused on the "law of punishment," or today's criminal law. Civil harms were dealt with primarily through social norms developed through Confucianism. Emperors attempted to develop general rules for civil behavior by issuing edicts associated with certain types of disputes. In practice however, these rules were often too specific in their application and failed to provide general principles concerning civil liability. As a result, emperors resorted to adjudicating individual cases as they arose. Unlike judicial rulings in modern common law jurisdictions, rulings in China's imperial era did not have a binding effect upon the adjudication of later cases (Lin et. al., 1989).

Faced with the large burden of handling individual cases, emperors encouraged the local resolution of disputes. Minor disagreements, such as those dealing with family matters or land, comprised the majority of disputes (Zhang, J., 2014). These disputes were generally mediated by respected community leaders or village elders, who applied customary rules and concepts of morality to reach harmonious resolutions, as per the teachings of Confucius. More serious disputes were decided by a hierarchy of ministers and officials, all ultimately accountable to the emperor. The existence of a strong, community-based mediation system arising out of emperor rule is a tradition somewhat unique to China that is still widely employed as a modern mechanism of dispute resolution.

Like emperors in imperialist China, more recent leaders of the Communist Party of China (CPC) have retained power over the country's legal sphere. From 1966 to 1976, the CPC pursued political goals leading to China's Cultural Revolution. During this time, China's Communist leader Mao Zedong mobilized the nation's youth to reassert his authority over the Chinese government, which he believed was being eroded by other party leaders. As a result of these efforts, citizens lost many of their rights, and laws passed just years prior were no longer enforced (Rou & Ocko, 1989). Though individual freedoms have increased since the rule of Mao Zedong ended, modern Chinese legal processes still

experience a relatively high degree of government influence. As a result, law is still commonly viewed as an instrument of the ruling party (Li, 2012).

III. Legal Development

An overview of China's legal development reveals a link between economic growth and subsequent legislative change. Furthermore, we observe that when legislative changes occur, China, like other jurisdictions, often borrows from rules created elsewhere.

As the rule of emperors was coming to an end in the final years of the Qing Dynasty during the late 1800s and early 1900s, China entered into an unsuccessful war with Japan over China's control of the Korean peninsula. Previously, China had considered Japan a subordinate part of the Chinese cultural sphere, but the war demonstrated that China had lost ground economically to its Asian neighbors. Perhaps in partial response, government reform advocates encouraged the adoption of various elements of the Japanese legal system, which were in turn based on German judicial precedent.

Similar legal transplantation occurred in 1949 when the CPC and Mao Zedong came to power. Beginning with the banking sector, and then moving into all other domains, China transitioned quickly to a centrally planned command economy modeled on that of the Soviet Union. Under this new economy, the state allocated industrial inputs and outputs, including allocation of workers to jobs, setting wages, and rationing consumer goods (Perkins, 2013). Alongside economic reforms, the government promulgated a host of new tort regulations based on Soviet precedent, which outlined compensation for damages in a number of specific industries (Rou & Ocko, 1989).

Since the rule of Mao, China has transitioned from a planned economy to one with a robust private sector. In the early years of reform, law was seen as a mechanism for regulating state owned enterprises (SOEs). Transition to a market economy and a reduction in the relative number of SOEs means that the government today directly controls relatively fewer entities than in the recent past. In the past, the government could resolve disputes across the entities it controlled. Today, with a smaller percentage of entities under government control, the need for an alternative dispute resolution mechanism beyond the ruling party is apparent (Clark et. al., 2008).

China's changing economy also spurred demand for legislative change among individuals. According to Lin (1989), an expanding market increased the frequency of certain types of tortious

conduct, such as product liability, torts arising from dismantling old houses, the use of land, irrigation systems, forests, and orchards, and for damage to crops caused by inadequate supervision of livestock. Additionally, economic changes brought about new types of tortious conduct, such as trademark infringement. These conditions necessitated revision of the nation's legal framework.

In additional to domestic demand for legal change, similar pressure from foreign entities began to mount following the opening of the Chinese economy to foreign trade and investment in the late 1970s and early 1980s (Clark et. al., 2008). In an effort to facilitate market access and stability, trading partners sought stronger laws that would detail rights and protect their interests, improved enforcement of laws, and greater transparency throughout legislative processes (Zimmerman, 2005). As an initial response, the Chinese government attempted to maintain separate legal systems for domestic and foreign entities but recognized that a segregated legal system did not make sense in the long term. The ultimate unified system is most similar to what had been developed for foreign investors.

Demand for legal change from all of these sources as well as from foreign trading partners culminated in the passage of the General Principles of Civil Law (GPCL), promulgated in 1986. The GPCL was the first piece of Chinese legislation to apply broadly to all realms of civil relations. Prior legislation had focused on certain industries and/or areas of society. Importantly, the GPCL defines the rights of citizens and legal persons, lays out fundamental principles to which civil acts must adhere and conditions which constitute a civil act, and asserts general rules for commodities exchange (Rou, 1989).

In the years since the passage of the GPCL, China's economy has continued to grow rapidly. The occurrence of a number of high-profile liability events has also grown in recent years, especially with regard to food safety. At least 39 high-profile liability events have occurred since 2003. W. Zhang (2014a) contends that these events have expedited the passage of subsequent legislation. According to Johnson (2011), the prior tort system was not equipped to handle increased exposure to product-related injuries brought about by rapid economic growth, and thus required change.

Notable laws include the Civil Procedural Law (1991; 2007; 2012), Consumer Protection Law (1993; 2013), Product Quality Law (1993; 2000), Litigation Cost Payment Act (2006), Property Law (2007), Patent Law (2008), Food Safety Law (2009; 2015), Copyright Law (2010), and Law of the Application of Law for Foreign-Related Civil Relations (2010).[1] These laws have increased protections for

[1] The first date listed refers to the time of original promulgation and subsequent dates refer to times of amendments.

potential plaintiffs in China in a number of ways: by expanding the scope of rights under protection, expanding the scope of liable parties, tightening liability rules favorable to potential defendants, and augmenting damages. W. Zhang (2014b) provides a detailed discussion of these items, offering several instances in which the litigation of specific torts would change following the enactment of various pieces of legislation.

This trend toward increased protections eventually lead to passage of the the Tort Law in 2009. Unlike many of the other laws passed in the late twentieth and early twenty-first century, which apply only to specific industries or classifications of torts, the Tort Law, like the GPCL, outlines provisions for torts generally. Importantly, the Tort Law extends provisions initially found in many of the specifically-focused laws preceding it to all types of disputes (Johnson, 2011).

China's experience of legislative reform following economic and social change is mirrored in other developing regions, such as Hong Kong and Brazil among others. In the mid-19th century, Hong Kong experienced rapid industrialization and urbanization. In this environment even if a defendant behaved carefully they ran the risk of injuring others. Litigation laws at the time (known as forms of action) only required that the defendant be the cause of harm for litigation to proceed. Thus, the defendant might be found liable despite having been careful. This situation was seen as unfair and the law was changed. The Common Law Procedure Act of 1852 and the Judicature Act of 1873 provided that carelessness (negligence) be a requirement for establishing liability. These new laws favored those who engaged in activities that could result in accidents and injury to others such as industrial organizations and operators of motorized and mass transportation and was seen as more conducive for a developing economy (Glofcheski, 2012).

Similar to the experience of China and Hong Kong, industrialization and urbanization in early 20th century Brazil was met with mounting public pressure on lawmakers to reform tort legislation to keep pace with the rapidly changing society. A widespread legal modernization occurred in the 1940s resulting in reforms to contract and tort legislation. The modernization continued through the 1950s and 60s and in 1969 the Brazilian government set out to "to render civil law coherent with the societal realities of Brazil in the ongoing century" (Bruggemeier, 2011).

IV. Current Legal Obstacles

Though the codification of civil procedural laws provides newfound opportunities to utilize the legal system as a mechanism to resolve disputes, legal and political obstacles remain. We consider these obstacles in two categories: those stemming from codified law that remains unfavorable to potential plaintiffs; and those stemming from the practice of law that at times undermines codified law.

Even as legal provisions continue to develop, certain aspects of codified law discourage potential plaintiffs from utilizing China's judicial system. For example, "discovery" is a pre-trial procedure in which each party can obtain evidence from the other party or parties. In China, discovery is generally not available and thus, plaintiffs must find their own evidence to meet their burden of proof. Furthermore, the period for evidence submission is short, typically between 30 and 60 days. Given that no new evidence may be entered after this window has closed, plaintiffs must gather the majority of their evidence before filing their case. As a result, plaintiffs spend a significant amount of resources before knowing if their case will be heard (Bennett, 2008).

Upon filing a lawsuit, plaintiffs are subject to extensive litigation costs, including filing costs, court acceptance costs, and possibly costs of transportation, accommodation, living expenses, and compensation for lost work hours for a witness, a notary, an interpreter, or an adjuster requested by the court to appear at a court hearing (Litigation Cost Payment Act, 2006). Among these costs, court acceptance costs are uniquely applied in China, in that they depend on the amount or value of damages sought in the lawsuit.[2] As a result, court acceptance costs may shape damages sought (Liufang, 2007).

During and after trial, legal provisions remain unfavorable for plaintiffs. Witness testimony is generally not admissible as evidence; rather, evidence consists of written documents and occasionally, court appointed experts. Even after a verdict has been reached, enforcement can be difficult. In the case of domestic judgments, the losing party may move assets to a region outside the jurisdiction of local law enforcement. Given that asset tracking systems in China are still primitive, the assets may never be found or seized and the plaintiff may never receive payment. If the plaintiff still hopes to collect from the defendant, they must try the case again in a higher court, where the decision will have a broader reach (Bennett, 2008).

[2] For a table of court acceptance costs see http://chineseip.jmls.edu/sites/en/court-acceptance-cost-chinese-civil-litigation.

In addition to codified law which remains unfavorable to plaintiffs, current legal practices in China at times undermine codified law altogether, adding uncertainty to legal outcomes and discouraging the pursuit of certain types of claims. In China, a large number of administrative bodies have the authority to pass laws including the National People's Congress and its standing committee, the state council, People's Congresses at the province level as well as in certain large cities, and ministries of the state council and local governments. In many cases, laws passed by these bodies conflict with one another. Though an order of precedence exists in theory, in practice, judicial decisions regularly reverse this order (Clark et. al., 2008).

Even when laws do not conflict and seem to confer private rights of action, the state has stepped in to limit these rights. For example, the state sometimes requires that all private suits be preceded by an adverse finding against the defendant in administrative or criminal proceedings (Clark et. al., 2008).

Government officials have also demonstrated a willingness to intervene in certain high-profile cases, often through required mediation. Like governments in other jurisdictions, the Chinese government maintains that intervention within legal processes is intended to maintain a "harmonious society." Government intervention is arguably a more harmonious solution to disputes than use of the judicial system because while the former involves the coordination of efforts among parties, the latter involves an adversarial process in which one party tends to "win" and one party tends to "lose."

Others dispute the government's stated motivation for intervening, citing at least three other possible motives. First, judicial litigation may result in the depletion of state assets and/or public officials may be held responsible. Given that many corporations are owned or controlled by the government, litigation against these corporations may equate to litigation against the state (Fu, 2006). Second, judicial litigation may reduce the resources available to the local government. For many local governments, up to 90% of their revenues come from taxes on local corporate profits. Thus, inflicting financial penalties on these corporations will lower government revenue. Third, judicial litigation may jeopardize career advancement opportunities of local officials who tend to be promoted based on their municipality's contribution to GDP. As a result, legal impediments to business ventures may reduce the probability of promotion (Fu & Nicoll, 2011).

Regardless of their actual motivation, experience from several high-profile cases suggest that the government has pressured judges to reject claim applications and has pressured lawyers to reject

clients with particular types of disputes. To understand why it is difficult, if not impossible, for judges to go against the wishes of the government, consider the nature of the relationship between these entities. The People's Congress has the power to appoint and dismiss court officials. In practice, this power lays in the hands of the local communist party organizational department. Furthermore, local governments control court finances, material supplies, and other welfare benefits for court officials and their families. Finally, local authorities have a natural power over the courts in that they have the ability to enforce or not enforce court decisions.[3]

The government's ability to influence lawyers stems from the fact that in in many cases, the government has the ability to revoke a lawyer's license to practice law.[4] Additionally, in order to become and remain licensed, most lawyers are required to join provincial judicial bureaus or lawyers' associations, which are controlled by The Ministry of Justice.

Specific high-profile events illustrate how the Chinese government has exercised this influence over judges and lawyers. In addition, such discussion provides an understanding of government dealings with other parties, such as potential plaintiffs and the media, and may shed light onto possible government motives for intervention within the legal system generally. Two such events are discussed here: the sale and distribution of tainted milk in 2008 and the Wenzhou high-speed train crash in 2011. Our account here comes from various western news stories. We acknowledge that we cannot know for certain what occurred, and that Western outlets may represent conditions differently from domestic outlets, yet the extent of reporting and the ultimate response of the government suggest that the news reports are reasonably accurate.

In 2008, an estimated 300,000 people in China became sick after consuming milk and infant formula adulterated with melamine (a chemical added to dairy products to make them appear to have a higher protein content). Government inspections following the incident revealed that products from 22 companies were contaminated. The greatest offender among these companies was Sanlu, a Shijiazhuang based company that led the market in dairy sales at the time.

[3] This point is true in many Western nations as well.

[4] In order to become and remain licensed, most lawyers are required to join provincial judicial bureaus or lawyers' associations, which are controlled by The Ministry of Justice.

When local officials became aware of the contamination, they initially did nothing. It took the Shijiazhuang government 38 days to report the incident to the Hebei provincial government, even though regulation specifies that they should have done so within two hours. Other non-official state media reports indicate that Fonterra, a New Zealand diary producer that owned 43% of Sanlu's shares, was the first to break the story. These reports claim that after Fonterra was told about the incident, they met with Sanlu and Shijiazhuang government officials three times to try to persuade them to disclose the problem. Fontera was unsuccessful in their attempts and eventually reported the incident to the New Zealand Prime Minister, who then contacted the Chinese state government.

When local officials were asked why they had waited to disclose the report, allowing the contamination to escalate, a spokesman of the Shijiazhuang government stated "support for local businesses" was the primary factor. Further investigation uncovered a letter from Sanlu to the Shijiazhuang officials that pleaded with the government to "increase control and coordination of the media, to create a good environment for the recall of the company's problem products.... to avoid whipping up the issue and creating a negative influence in society" (Fu & Nicoll, 2011).

After the incident became publicly known, officials attempted to limit litigation against Sanlu in a number of ways. Initially, officials pressured potential plaintiffs not to file lawsuits and instead to seek out-of-court compensation payments. A journalist who spoke out against this pressure was jailed for "disrupting social harmony" after encouraging parents of affected children to sue. This type of government response is known to have occurred in at least one prior incident. Parents of children who had died in a school collapse during an earthquake were also instructed not to file lawsuits regarding poor school construction and instead to rely on out-of-court methods of compensation (Johnson, 2011).

Officials apparently also instructed lawyers not to provide legal aid. Just after the tainted milk incident was uncovered, more than 100 lawyers across the country volunteered to give legal advice to those affected. On October 7, 2008, the volunteer group issued a brief indicating that officials had instructed the group not to provide any legal aid. Following these instructions, over two-dozen lawyers removed their names from the volunteer list.

Commentators further believe that officials used their influence to discourage judges from taking cases related to the tainted milk incident. As a result, courts rejected many individuals who attempted to file claims against Sanlu without explaining why the case would not be heard (Fu & Nicoll, 2011).

As a result of these conditions, few plaintiffs successfully brought claims against Sanlu. Instead, most potential plaintiffs accepted nominal compensation offered by dairy companies and backed by the government. In doing so, these potential plaintiffs forfeited their right to bring legal claims.

As noted above, debate exists as to why the Chinese government took measures to limit litigation against Sanlu. Some contend that the primary motive was to protect Chinese business and government officials associated with the situation. Others argue that the government's intention was to maintain harmony and to deal with issues more systematically and officially. Supporters of government intervention following the incident point to the fact that two company executives were executed, while others were given long prison sentences, as evidence of the government's interest in protecting the public.

A second high-profile incident that illustrates the government's influence in the claiming process is the Wenzhou high-speed train crash. In July 2011, two high-speed trains collided in China's Wenzhou province, derailing each other. They were traveling on China's newest high-speed railway, the Harmony Express, which is the largest and fastest high-speed railway in the world. Government officials claim that a lightning strike caused part of the rail's signal system to malfunction, leaving drivers and dispatchers unable to locate trains and freezing the signal on the color green. As a result, two trains were able to enter the same stretch of track and dispatchers were not alerted by the signal system until it was too late. In total, 40 people were killed and 192 were injured.

The rapid pace of railroad construction may have been to blame for the incident. The government cut a year from the rail's construction schedule in an effort to finish a prominent line from Beijing to Shanghai by the 19th anniversary of the CPC. High-speed rail staff were warned that delays would affect the size of their bonuses. Zhao Jian, a professor at Beijing Transportation University, publicly objected to the hasty development of the high-speed rail system but was told by high-ranking officials and the president of the university not to voice these opinions (Osnos, 2012).

Following the collision, the government tried to avoid media attention by hastily dealing with the incident and threatening editors. Within 24 hours of the collision, the search for missing persons had been called off, derailed train cars were torn apart by backhoes and partially buried, and the rail line was reopened. The Department of Propaganda initially told editors not to link the story with high-speed rail development or to give the incident much attention, stating, "Do not question, do not elaborate."

Shortly after the government banned all coverage of the disaster 'except positive news or that issued by the authorities.'

On the first anniversary of the train crash, the government again blocked any public discussion of the collision. All media coverage of the disaster's first anniversary was blocked, Internet searches for "Wenzhou" or "high-speed train" were (and remain) censored, and police physically prevented people from visiting the crash site (Tovrov, 2012).

In the event's aftermath, government issued directives to restrict media coverage were met with limited compliance, even on state-owned networks, and the story quickly took hold as the latest in a series of incidents caused by government corruption. After the partial train burial, reporters accused officials of trying to hinder an investigation. Many people, especially habitual internet users, expressed skepticism about the government's explanation for the cause of the disaster. Wang Yongping, a spokesman for the Ministry of Railways replied, "Whether or not you believe [the government's explanation], I believe it." This phrase went viral on the Internet and became the mark of the government's diminishing credibility.

Although the story achieved widespread media attention, a legal response was non-existent. Three days after the collision, an "urgent statement" was sent to law firms in Wenzhou on behalf of the city's Judicial Bureau and the Lawyers' Association. According to Xinhua, a news agency covering the matter, the statement described the accident as a "major sensitive issue concerning social stability" and instructed lawyers to report anyone seeking legal help to the Judicial Bureau and Lawyers Association. The government paid compensation to the families of those involved in the amount of 915,000 Yuan ($143,000), roughly 15.5 times the average annual household income of Zhejiang province at the time. No known lawsuits have been filed (Associated Press, 2011).

In both the tainted milk scandal and the Wenzhou high-speed train crash, officials went to great lengths to avoid the publicity associated with lawsuits. Given that the government compensated those affected by these through non-legal means, the motive for these interventions appears to be the maintenance of social stability and control. Whether similar responses will continue into the future remains to be seen.

V. Concluding Remarks

Over the past half century, China's economy has undergone remarkable change. The country has opened itself to foreign trade, investment, and travel and has shifted to a more capitalistic economic model. These changes have produced rapid and dramatic increases in GDP and the general standard of living as well as general knowledge among Chinese people of how systems operate outside of China. All of these changes have been accompanied by legal system modifications.

The Chinese government has demonstrated a willingness to support economic development though modifications and additions to codified law. Indeed, many such laws have been enacted in recent history. Additionally, China has the great advantage of being able to observe decades of practice elsewhere and implement systems that appear most effective.

Yet, the government has been less inclined to alter long-standing legal traditions, including the emphasis on mediation (instead of litigation) to resolve disputes and government control over legal system. At times, the latter undermines written law, dampening the impact of such legislation. Thus, while legislative development may continue, utilization of judicial processes may increase at a slower rate than it would in the absence of legal obstacles.

The changing forces within China's legal system make it a fascinating area for future legal research. Indeed, a watchful eye on the government's ability to maintain control as the economy and society continue to develop in a more open and entrepreneurial spirit will be needed to anticipate future legal changes.

Bibliography

Associated Press. (2011, July 31). Lawyers in China Told to Turn Down Train Crash Cases. *Jakarta Globe*. Retrieved from http://jakartaglobe.beritasatu.com/archive/lawyers-in-china-told-to-turn-down-train-crash-cases/

Bennett, S. C. (2008). Litigation in China: Ten Things You Must Know. *The Metropolitan Corporate Counsel*. Retrieved from http://www.metrocorpcounsel.com/pdf/2008/October/19.pdf

Brüggemeier, G. (2011). Modernising civil liability law in Europe, China, Brazil and Russia: texts and commentaries. *Cambridge University Press.*

Clarke, D., Murrell, P., & Whiting, S. (2008, June). The Role of Law in China's Economic Development. *Within "China's Great Economic Transformation" by Loren Brandt and Thomas Rawski.*

Epstein, E. J. (1986). The Evolution of China's General Principles of Civil Law. *The American Journal of Comparative Law*, 705-713.

Liufang, F. (2007, June 24). A Study on Civil Litigation Fee. Retrieved from http://blogs.harvard.edu/guorui/files/2007/06/Litigationfee.pdf

Fu, J., & Nicoll, G. (2011). Milk Scandal and Corporate Governance in China, The. *Canberra L. Rev.*, *10*, 103.

Fu, J. (2006). Corporate Governance in China: A Research Agenda for a Corporate Group and Shareholder Control Perspective. *Journal of Business Systems, Governance and Ethics*, *1*(1), 1.

Garber, S. (2013). Economic effects of product liability and other litigation involving the safety and effectiveness of pharmaceuticals. *Rand Corporation.*

General Principles of the Civil Law of the People's Republic of China (1986).

Glofcheski, R. (2002). Tort law in Hong Kong. Hong Kong: Sweet & Maxwell Asia.

Johnson, V. R. (2011). The Rule of Law and Enforcement of Chinese Tort Law.

Li, J. (2013). Suing the Leviathan—An Empirical Analysis of the Changing Rate of Administrative Litigation in China. *Journal of Empirical Legal Studies*,*10*(4), 815-846.

Li, Y., & Yan, W. (2011, April 1). PRC People's Mediation Law. Retrieved from http://www.faegrebd.com/prc-peoples-mediation-law

Lin, Y., Xiaoming, L., & Pitney, H. (1989). The tort system in China. *Law and Contemporary Problems*, *52*(3), 143-167.

Osnos, E. (2012). Boss rail: the disaster that exposed the underside of the boom. *The New Yorker*, *22*. Retrieved from http://www.newyorker.com/magazine/2012/10/22/boss-rail

People's Mediation Law of the People's Republic of China (2010).

Rou, T., & Ocko, J. K. (1989). The" General Principles of Civil Law of the PRC": Its Birth, Characteristics, and Role. *Law and Contemporary Problems*,*52*(2), 151-175.

Tort Liability Law of the People's Republic of China (2010).

Tovrov, D. (12, July 23). China Propaganda Ministry Censors News Of Wenzhou Train Crash Anniversary. *International Business Times*. Retrieved from http://www.ibtimes.com/china-propaganda-ministry-censors-news-wenzhou-train-crash-anniversary-730189

Wong, E. (2008, October 16). Courts Compound Pain of China's Tainted Milk. *The New York Times*. Retrieved from http://www.nytimes.com/2008/10/17/world/asia/17milk.html

Zhang, J. (2014). The Tradition and Modern Transition of Chinese Law. *Springer*.

Zhang, M. (2011). Tort liabilities and torts law: the new frontier of Chinese legal horizon. *Richmond Journal of Global Law and* Business, 10(4).

Zhang, W. (2014a). Understanding the Law of Torts in China: A Political Economy Perspective. *Singapore Management University School of Law Research Paper*, (40).

Zhang, W. (2014b) The Evolution of the Law of Torts in China: A Bird's Eye View. *Singapore Management University School of Law Research Paper*.

Litigation in China: Development and Drivers

Abstract

We explore litigation in China using three decades of province-level legal data. Specifically, we observe changes in litigation rates across provinces and time and examine the relationship between those changes and a host of economic, social, and legal factors. We find that gross regional product, education, Internet use, tourism, access to lawyers, and privatization appear to influence litigation rates. A closer inspection of these factors reveals that factors are important to litigation rates in highly developed areas but not in less developed areas, as defined by the China National Human Development Report. Additionally, we find that a requisite level of development must be attained before these variables become relevant and that this level roughly corresponds with a human development index value of 0.57.

I. Introduction

The enforcement or defense of legal rights (litigation) is an important social process. It provides a means of allocating costs when disputes occur and, in turn, shapes incentives for safety. As such, understanding which factors influence litigation is beneficial both in setting public policy and in preparing for such claims through compensation mechanisms such as insurance.

Determinants of litigation have been studied extensively in developed economies, yielding insight into the factors influencing litigation in these areas. Such analyses have been less common in emerging economies, likely because legal and other data are often unavailable for these jurisdictions. Our intent is to test the relationship between litigation and economic, social, and legal development in an emerging economy by gathering and examining litigation data in China.

China's rapid economic rise and enormous size make the country an ideal environment in which to investigate the relationship between development and litigation. An understanding of the forces shaping China's legal system may prove useful in understanding the evolution of civil disputes in other emerging economies. Furthermore, China is a nation of unique and well-established legal traditions, including the presence of a strong community-based system of mediation and a high degree of government influence in the legal sphere. An important element of this study, therefore, is to test the influence of these traditions alongside factors shown to influence litigation in other jurisdictions.

To achieve our objectives, we employ three decades of data from China across 24 provinces.[5] Over that time period, litigated cases per 10,000 people (which we refer to as litigation rates) have risen over 1400 percent across the country, including nearly 2700 percent in Zhejiang, nearly 1800 percent in Fujian, and over 1600 percent in Guangdong (the three provinces in China with the greatest litigation growth over this time period). Along with data on litigation rates, we collect and analyze economic, social, and legal data to identify the most salient factors influencing litigation rates in China.

The paper is organized as follows. In section two, we describe a general economic theory of litigation and review empirical evidence regarding that theory. The result is a set of testable hypotheses. In section three, we present information about the data, methods, and analyses used to test our hypotheses. In section four, we conclude with a discussion of our findings and suggestions for future research.

II. Model Development and Literature Review

The advancement of the field of "law and economics" in the 1960s saw the application of classic economic theory to the study of legal participant behavior. By the 1990s, a variety of scholars had developed general models of the litigation process, including the well-known model by Cooter and Rubinfeld (1989). Within this framework, parties decide (1) how much to spend on safety; (2) whether to bring a claim; (3) whether to pursue the claim further if initially rebuffed; and (4) whether to settle prior to a trial outcome.[6] Decisions at each stage are considered within the context of a multistage game. Ultimately, decisions rely upon the possible actions of each party, estimates of costs and benefits given those actions, and risk attitudes, among other conditions.

Following Ramseyer's (2012) approach, we simplify the model to consider lawsuits as the product of two pieces: the existence of disputes and willingness to make claims for damages from those disputes. A "dispute," as we define it, is an event in which one party (the defendant) allegedly harms another (the plaintiff) and the parties disagree about who should pay for the harms. These events may be accidents or crimes and generally produce losses for the plaintiff. Given the existence of a dispute,

---

[5] China consists of 31 provinces, municipalities, and autonomous regions. We observe data for 24 of these regions and refer to them collectively as "provinces."

[6] Sociologists have developed a very similar structure of the litigation process, generally referred to as a "dispute pyramid." See for example, Felstiner et. al. (1980) and Miller and Sarat (1980).

"willingness to claim" represents the likelihood that the dispute results in litigation. The following equation illustrates the relationship between lawsuits, disputes, and willingness to claim.

$$\text{Lawsuits} = \text{Disputes} \times \text{Willingness to Claim}$$

This simple model is valuable, in part, because it calls attention to disputes, the pool from which lawsuits originate. As Ramseyer indicates, commentators often equate litigation with litigiousness, while occurrences of underlying disputes are not considered. It may be that that disputes are sometimes overlooked because researchers rarely have information about the prevalence of such events. Indeed, only a handful of litigation studies attempt to measure underlying disputes. These studies estimate the number of disputes in a few different ways: by asking survey respondents about such experiences (e.g. American Bar Association, 1994; Hensler et. al., 1991; Miller and Sarat, 1980), by examining medical conditions in which the underlying harms are reported along with disputes such as medical malpractice (e.g. Harvard Medical Practice Study, 1990; Mills, 1977; Pocincki et. al., 1973), or by considering types of litigation for which all disputes are reported to authorities, such as divorces and auto accidents (e.g. Ramseyer, 2012). In recognizing both the relevance of underlying disputes and the difficulty in obtaining this information, Eisenberg (2010) calls for the creation of a "National Civil Justice Survey" to track the occurrence of disputes and their evolution.

The vast majority of disputes do not result in litigation. For example, in a telephone survey of approximately 1,000 households in South Carolina, Eastern Pennsylvania, Eastern Wisconsin, New Mexico, and Central California, Miller and Sarat (1980) find that only about 10% of disputes come to the attention of lawyers and that only about 5% of disputes become lawsuits. Some authors suggest that a reluctance to litigate stems from the fact that many potential plaintiffs do not realize that they have a viable legal claim or do not know the procedure for hiring a lawyer and filing a lawsuit (Campbell & Talarico, 1982; Carlin et. al., 1966; Cheatham, 1963; Morrill et. al. 2010; Pleasence et. al., 2011; Sandefur, 2007). Other authors argue that litigation may appear unfavorable from a cost-benefit perspective and that this fact, rather than a lack of understanding about claim viability or procedure, is the primary explanation for a reluctance to litigate (Bezanson, 1987; Miller & Sarat, 1980). Regardless of the reason, it is clear that lawsuits comprise only a small fraction of all disputes. Still, when a dispute does result in litigation, this transition is captured within society's willingness to claim.

Like other studies, this study lacks data on underlying disputes and thus does not attempt to disentangle disputes and willingness to claim. Here, we distinguish between disputes and willingness to

claim simply to facilitate discussion about the factors affecting each piece. If a factor affects disputes, willingness to claim, or both, we would expect that factor to affect litigation rates as well.

To identify factors influencing disputes, we begin with the observation that barring intentionally-caused harms, such events are random. As such, the total number of disputes in a jurisdiction derives from a mostly random process. Economic circumstances may affect this process by changing the frequency of interactions among individuals, organizations, and the government and by altering the level of safety surrounding those interactions. As the frequency of interactions increases, so too should the number of disputes, all else equal. Conversely, as safety measures increase, the number of disputes should fall, all else equal.

To identify factors influencing willingness to claim, we begin by recognizing that the process underlying willingness to claim derives from numerous individual decisions about whether or not to litigate. The litigation process itself is costly and the outcome is uncertain. If a harmed individual chooses not to pursue a legal claim, they forego the gamble and retain their current financial state.

A decision involving risk, such as the one just depicted, is highly suitable to well-established economic models such as expected utility theory and prospect theory. Thus, it comes as little surprise that the field of law and economics has taken up the study of legal participant behavior. We look to two foundational papers from this field which consider the decision to litigate within the context of expected utility theory (Cooter and Rubinfeld, 1989; Shavell, 1982). If the expected utility of litigating exceeds the expected utility of not litigating, litigation is expected to occur.[7]

Several macro-level factors may affect the expected utility of litigation, including economic circumstances, social characteristics, and a jurisdiction's legal environment. These factors capture in part financial resources available to litigants, educational attainment, social interactions and attitudes, access to legal services, the presence of alternative dispute resolution systems, and expectations about legal proceedings. All of these conditions shape the calculus surrounding the decision to litigate.

In summary, three general conditions are expected to influence litigation: economic circumstances, social characteristics, and a jurisdiction's legal environment. Economic circumstances are unique in that they may influence both the number of disputes and willingness to claim. Conversely, social characteristics and a jurisdiction's legal environment are expected to influence litigation only

[7] Alternatives include seeking compensation through non-legal channels, such as China's system of community mediation, and taking no action.

through their influence on willingness to claim.

a. Economic Circumstances

The relationship between economic circumstances and litigation has been studied in a variety of jurisdictions, including China (Michelson, 2007; Mingming & Yuhua, 2009), Hong Kong (Glofcheski, 2012), India (Eisenberg, 2012), Italy (Carmignani & Giacomelli, 2010), Japan (Ginsburg & Hoetker, 2006; Ramseyer, 2012), the United Kingdom (Genn & Beinart, 1999; Patel et. al., 2008; Sandefur, 2009), and the United States (Atkins & Glick, 1976; Daniels, 1982; Grossman & Sarat, 1974; Jacobi, 2009; McIntosh, 1983; Yates et. al., 2001 & 2010). Circumstances most commonly studied on a macro scale include economic activity and urbanization.

Economic activity may influence litigation by altering the number of disputes in a jurisdiction and/or by changing the willingness of individuals to file lawsuits given the occurrence of disputes. Economic activities, such as the production and sale of goods and services, the development of new technologies, and the formation of business relationships, carry the potential for accidents and harms but also the potential for safer products and better working conditions. As such, the net effect of these opposing forces will determine whether economic activities increase or decrease the overall number of disputes in a jurisdiction.

In addition to its effect on disputes, economic activity may influence willingness to claim. Economic activity is often associated with rising incomes. Given that financial resources are required to file a lawsuit, higher incomes afford legal access to a greater share of potential plaintiffs (Ramseyer, 2012). However, these gains may be partially offset by the fact that augmenting plaintiff income raises the opportunity cost of filing a lawsuit (Mingming & Yuhua, 2009). Defendant income may also be important, as plaintiffs may not find litigation worthwhile when defendants have little or no wealth and thus, little or no ability to pay damages resulting from a lawsuit. Conversely, when defendants are exceptionally wealthy, plaintiffs might fear being “outspent” during litigation proceedings, which could diminish expected awards (He & Su, 2013).

Given the the array of effects that economic activity may produce, it is unsurprising that studies of economic activity and litigation have uncovered mixed results. Although a majority of empirical studies find a positive relationship between economic activity and litigation (e.g. Eisenberg, 2012; Glofcheski, 2012; Jacobi, 2009), some studies observe the reverse (e.g. Ginsburg & Hoetker, 2006),

others identify non-linear patterns (e.g. McIntosh, 1983; Mingming & Yuhua, 2009), and still others find no significant correlation whatsoever (e.g. Carmignani & Giacomelli, 2010; Michelson, 2007).

Like economic activity, urbanization may influence both disputes and willingness to claim. As people move to urban areas, high incidence of interactions arise from the close proximity among individuals, increasing the potential for disputes. Given the occurrence of disputes, individuals living in urban areas may exhibit a greater willingness to sue. In rural settings, parties to a dispute may share ongoing social interactions or close-knit community bonds, whereas in urban settings, parties to a dispute are more likely to be strangers. As a result, litigation costs for urban residents need not incorporate the discounting of long-term payoffs from established relationships (Berstein, 1992; Ellickson, 1991).

Unlike economic activity, the hypothesis surrounding urbanization is clear: urbanization is expected to exert a positive influence on litigation. This theoretical conclusion is supported in a number of empirical studies, including Carmignani & Giacomelli (2010), Daniels (1982), Eisenberg et al. (2012), Patel et. al. (2008, and Yates et. al. (2001 & 2010).

Given the theoretical and empirical evidence suggesting the importance of economic activity and urbanization, we include in our analysis real gross regional product per capita ***(GRP per capita)*** and ***urbanization rate***, defined as the proportion of the population residing in urban areas. In accordance with our theory, we anticipate a positive relationship between urbanization rates and litigation rates. Given the varied and opposing influences of economic activity on litigation, we do not hypothesize a particular relationship between GRP per capita and litigation rates. We test for a possible non-linear relationship between these variables by including ***GRP per capita^2*** in our analysis.

b. Social Characteristics

In addition to economic circumstances, we are interested in how social forces shape litigation. Prior research in this area reveals that legal knowledge is an important predictor of litigation rates. Legal knowledge refers to one's understanding of legal rights and responsibilities as well as one's ability to navigate judicial systems. Legal knowledge is measured in a number of ways throughout the literature, with several studies employing more than one proxy. The most commonly used proxies are education, prior court experience, and the score one receives on a quiz of legal information.

The literature generally indicates a positive relationship between legal knowledge and litigation (Eisenberg et al., 2012; Genn & Beinart, 1999; Mingming & Yuhua, 2009; Ramseyer, 2012; Vidmar & Schuller, 1987; Yates et. al., 2001 & 2010). Individuals who are knowledgeable about their legal rights may be better able to assess whether a particular dispute constitutes a viable legal claim. Upon filing a lawsuit, legally informed individuals may require less legal assistance and argue their cases more effectively, yielding more favorable settlement and trial outcomes.

Although legal knowledge may lead to more litigation by improving legal outcomes, some evidence suggests that education is associated with greater risk aversion (Eling et. al., 2013; Halek & Eisenhauer, 2001; Hersch, 1996). Given that litigation is an inherently risky proposition, risk averse individuals may be less inclined to utilize this process. As a result, the relationship between education and willingness to claim remains unclear.

In addition to legal knowledge, a number of theoretical articles suggest (but do not test) the idea that attitudes toward claiming may be important to litigation (Friedman, 1985; Engel, 2012; Galanter, 1974). When social norms discourage litigation, plaintiffs may experience reputational consequences for utilizing legal channels.[8] Reputational consequences represent additional costs to litigation, which may suppress willingness to claim. We hypothesize that as Chinese citizens become increasingly exposed to the legal norms of other jurisdictions, litigation will become an increasingly acceptable remedy for disputes and litigation rates will rise.

To account for the potential importance of social characteristics, we include the following three variables in our analysis: (1) ***high school graduation rate***, defined as the number of new graduates from secondary schools (high schools) per 10,000 people, (2) ***Internet penetration rate***, defined as the proportion of the population accessing the Internet for at least one hour per week, and (3) ***tourism***, defined as the number of international tourist visits per 10,000 people. High school graduation rate is employed as a proxy for legal knowledge, while Internet penetration rate and tourism proxy exposure to the rules and norms of other jurisdictions, which may influence attitudes toward claiming. We do not hypothesize a particular relationship between high school graduation rates and litigation rates, but anticipate a positive relationship between the other two variables and litigation rates.

[8] The Chinese government has repeatedly signaled a preference for non-litigious methods of dispute resolution (see Bujakowski, 2017). It may be that these signals have influenced social norms surrounding litigation.

c. Legal Environment

Like economic circumstances and social characteristics, a jurisdiction's legal environment may shape relative costs and benefits associated with litigation. Aspects of a jurisdiction's legal environment include, but are not limited to, access to lawyers, the presence of alternative dispute resolution systems, and expectations about litigation proceedings.

Prior research suggests that access to lawyers is an important determinant of litigation rates. Ginsburg and Hoetker (2006) find that in addition to a slowing economy, an expansion of the Japanese bar in the 1990s, which resulted in greater numbers of lawyers and judges, contributed to a rapid increase in Japanese litigation rates during that time. Carmignani and Giacomelli (2010) support this result in a study of litigation rates in Italian provinces from 2000 to 2005. In their analysis, Carmignani and Giacomelli use proximity of a law school in 1975 as an exogenous instrumental variable for lawyer density and conclude that the presence of lawyers is highly predictive of future claims. Similarly, Ramseyer (2012) finds that conditional on the occurrence of automobile accidents, individuals with greater access to lawyers are more likely to file a legal claim.

In addition to lawyer access, less formal (and less expensive) alternatives to litigation may impact the extent of litigation in a jurisdiction. In China, People's Mediation Committees are the embodiment of the informal system. These committees operate entirely at the village level and can be used free of charge to resolve any manner of dispute. The Chinese government has encouraged the use of community mediation as an alternative to litigation and has strengthened the system in recent years with the 2010 passage of the People's Mediation Law (Mealey-Lohmann, 2010).

During mediation, mediators offer guidance and encourage voluntary resolutions between parties. If a mediated resolution is acceptable to all parties, a legally-binding settlement agreement stating the terms of the resolution may be drawn up and signed. If however, a mediated resolution cannot be reached, other channels of dispute resolution such as the court system remain available to the parties (People's Mediation Law[9]). Given that mediation has the ability to resolve disputes before they reach the court system, it may serve as a substitute to litigation.

[9] See articles 2, 3, 5, 8, 31, and 32.

A third aspect of China's legal environment that we consider involves expectations about litigation proceedings. Though a host of legislative changes have been enacted to clarify civil procedural rules in China over the past few decades, government involvement in the legal sphere influences the implementation of such legislation. As a result, government influence remains a powerful determinant of legal outcomes.

The Chinese government appears most apt to become involved in specific legal cases when the defendant is a publicly-owned enterprise. In these situations, litigation against a corporation may equate to litigation against the state. As such, litigation may result in the depletion of state assets and/or loss of reputation to public officials, two outcomes which the government might prefer to prevent (Fu, 2006). We anticipate that as enterprises become privatized and the incentive for the government to intervene diminishes, uncertainty surrounding Chinese judicial should lessen, making litigation a more attractive proposition.

To test our hypotheses regarding the influence of China's legal environment, we include three variables in our analysis: ***lawyer density***, ***mediation rate***, and ***privatization rate***. Lawyer density is defined as the number of full time lawyers per 10,000 people and is intended to capture one's ability to access a lawyer. Mediation rate is defined as the number of mediation claims per 10,000 people and measures the frequency with which alternative dispute resolution is utilized. Privatization rate is defined as the proportion of assets owned by the private sector and is employed as a proxy for the diminishing incentive of the government to become involved in specific legal cases. In accordance with our theory, we anticipate that lawyer density and privatization rate will be positively associated with litigation rates and that mediation rate will be negatively associated with litigation rates.

Table 1 provides a summary of our hypotheses and the associated proxies used to measure relevant factors. Other factors, including the evolving nature of China's civil law doctrine and the occurrence of large-scale liability events, likely influence litigation as well, yet we do not possess data to account for them.[10] To combat the possibility of omitted variable bias and account for residual variation across provinces and time, we include in our analyses linear yearly time trends for each province as well

[10] A jurisdiction's legal doctrine specifies types of harms which constitute viable legal claims and sets costs to file suit, evidence standards, burdens of proof, and damage amounts, among other things.

as province and year effects. Province specific time trends are intended to control for legislative shifts among other things, while province and year effects are expected to pick up the effects of large-scale liability events that may exert a sizable impact on litigation rates in a given province or year.

Table 1: Summary of Potentially Influential Factors

Category	Variable	Definition	Hypothesized Sign
Economic	GRP per Capita	Real gross regional product per capita in 10,000s	+/-
	Urbanization Rate	Proportion of the population living in urban areas	+
Social	High School Graduation Rate	Number of new graduates from secondary schools (high schools) per 10,000 people	+/-
	Internet Penetration Rate	Proportion of the population accessing the Internet for at least one hour per week	+
	Tourism	Number of international tourist visits per 10,000 people	+
Legal	Lawyer Density	Full-time lawyers per 10,000 people	+
	Mediation Rate	Mediation claims per 10,000 people	-
	Privatization Rate	Proportion of assets owned by the private sector	+

The table lists eight variables (column two) which we hypothesize to be associated with litigation rates, defined as the number of litigated cases in a given province-year per 10,000 people and measured at the end of each year. These variables are grouped into three broad categories (column one) in an effort to highlight the mechanism by which they relate to litigation rates. Specifically, variables classified as “economic” are thought to influence both the number of disputes in a jurisdiction and likelihood that disputes become lawsuits, while variables classified as “social” and “legal” are thought to influence only the likelihood that disputes become lawsuits. Column three provides definitions for each variable. All variables except for litigation rates are measured at the beginning of each year. Column four indicates the hypothesized sign on variable coefficients in a regression of litigation rates on the eight variables.

III. Data, Methods, and Analysis

To examine litigation patterns in China and test hypotheses developed in the previous section, we employ 29 years of province-level data from China (1986-2014) across 24 provinces. These data are taken from the China Statistical Yearbook Database (CSYD), which includes provincial statistical yearbooks published annually by the National Bureau of Statistics of China. The CSYD is a compilation of data from a number of enterprises and governmental institutions. As a result, this database contains a host of economic, social, and legal information.

Available litigation data reported in the CSYD are derived from two sources, courts and law firms. Court data are preferred because they account for all litigated cases, as opposed to law firm data which represent only litigated cases that involve a plaintiff attorney. Unfortunately, court data are much less complete than law firm data. For example, court data are available for 11 Chinese provinces and are typically missing before year 2006, while law firm data are available for 24 Chinese provinces and become available for many provinces beginning in 1985. Given the data limitations for court data, we opt to use law firm data in conducting this study. An important result of this choice is that we do not observe when plaintiffs litigate without the assistance of a lawyer.

Figure 1 is a depiction of litigation rates by province across time. Provinces are organized in decreasing order of litigation rate in the most recent year of reporting. We observe that litigation rates have increased for each province over the timespan, but that the magnitude of this growth varies considerably by province. Zhejiang displays the greatest absolute gain of almost 43 cases per 10,000 people and the greatest relative gain of almost 2700 percent, while Jilin displays the lowest absolute gain of less than 0.7 cases per 10,000 people and the lowest relative gain of less than 20 percent.

Figure 2 is a map of China which illustrates variation in average litigation rates in the most recent 10 years of our study. From this map we observe that coastal regions, including Beijing, Zhejiang, Shanghai, Jiangsu, and Fujian tend to have the highest average litigation rates. This finding is unsurprising given that these regions are also the most economically developed (China National Human Development Report, 2016). In our analysis, we will consider the differential importance of factors in highly developed coastal provinces and less developed in-land provinces.

Given that we are working with a panel dataset, we can specify panel data regressions which exploit both cross sectional and time series variation to test the hypotheses outlined in Table 1. Equation 1 provides this general regression framework.

Equation 1

Litigation Rate = *f* (GRP per Capita; Urbanization Rate; High School Graduation Rate; Internet Penetration Rate; Tourism; Lawyer Density; Mediation Rate; Privatization Rate)

Where:

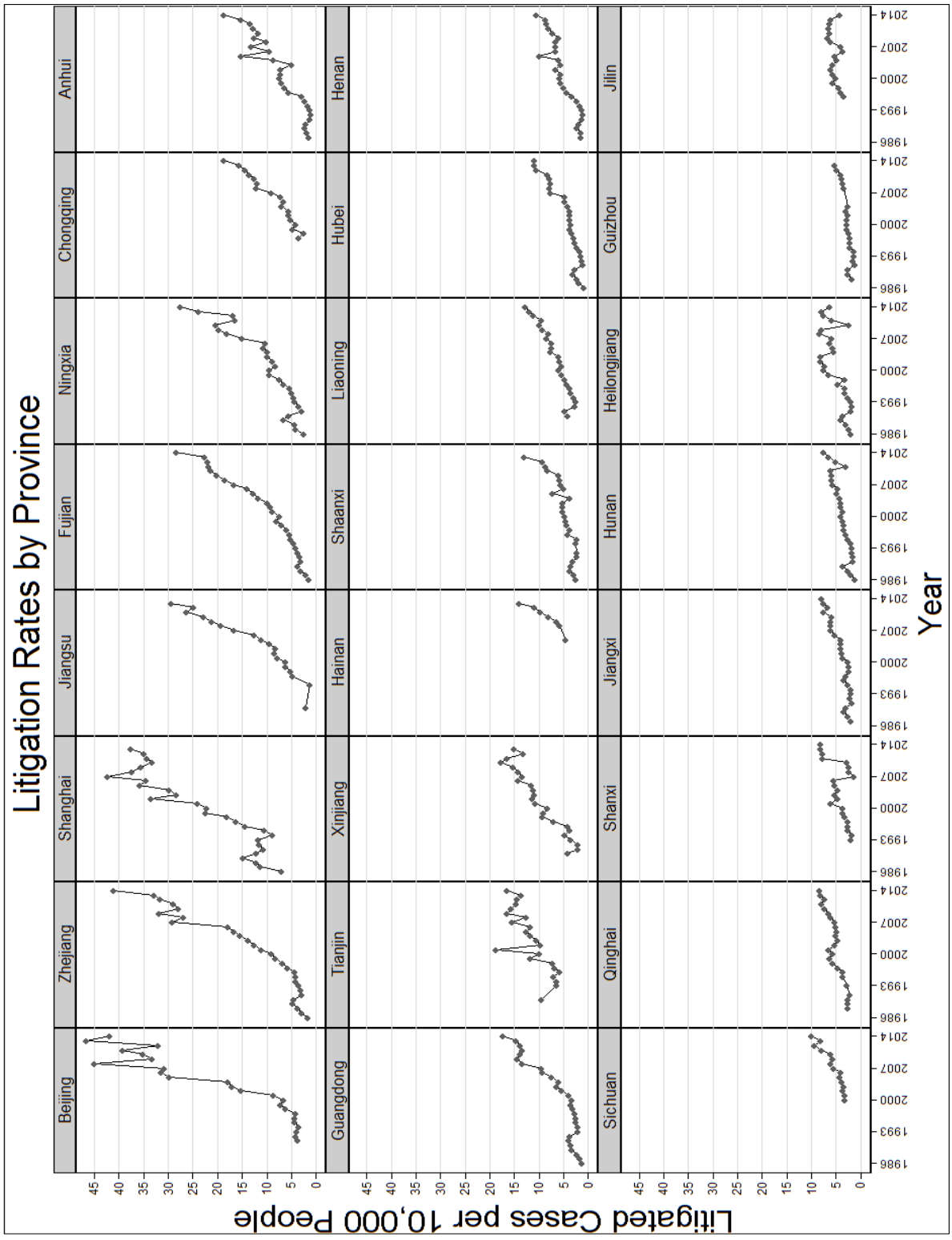
- Litigation Rate refers to the total number of litigated cases involving a lawyer per 10,000 people
- Other variables are defined as specified in Table 1.

Litigation rates are measured at the end of each year, while regressor values are measured at the beginning of each year. The difference in the time is intended to reduce the potential for endogeneity. Certain regressors, such as lawyer density and mediation rates, might both influence litigation rates and be influenced by litigation rates. For example, one can imagine that more lawyers might choose to work in a region or time period where/when litigation rates are high. By ensuring that regressor measurements are taken before those of litigation rates, we can reduce the potential for reverse causality.

We employ fixed effects to account for heterogeneity among provinces. We believe that fixed effects offers several advantages over random effects. Recall that the inclusion of any effects likely absorb, at least partially, omitted variables such as legislative changes and the occurrence of large-scale liability events. If, as we believe, these omitted variables are correlated with other regressors, only fixed effects are consistent. Furthermore, we utilize the Hausman specification test, which indicates that fixed effects are preferred. Despite the preference for fixed effects, we redo our analyses using random effects and find that results are generally robust to this specification. Results using random effects are in Appendix A.

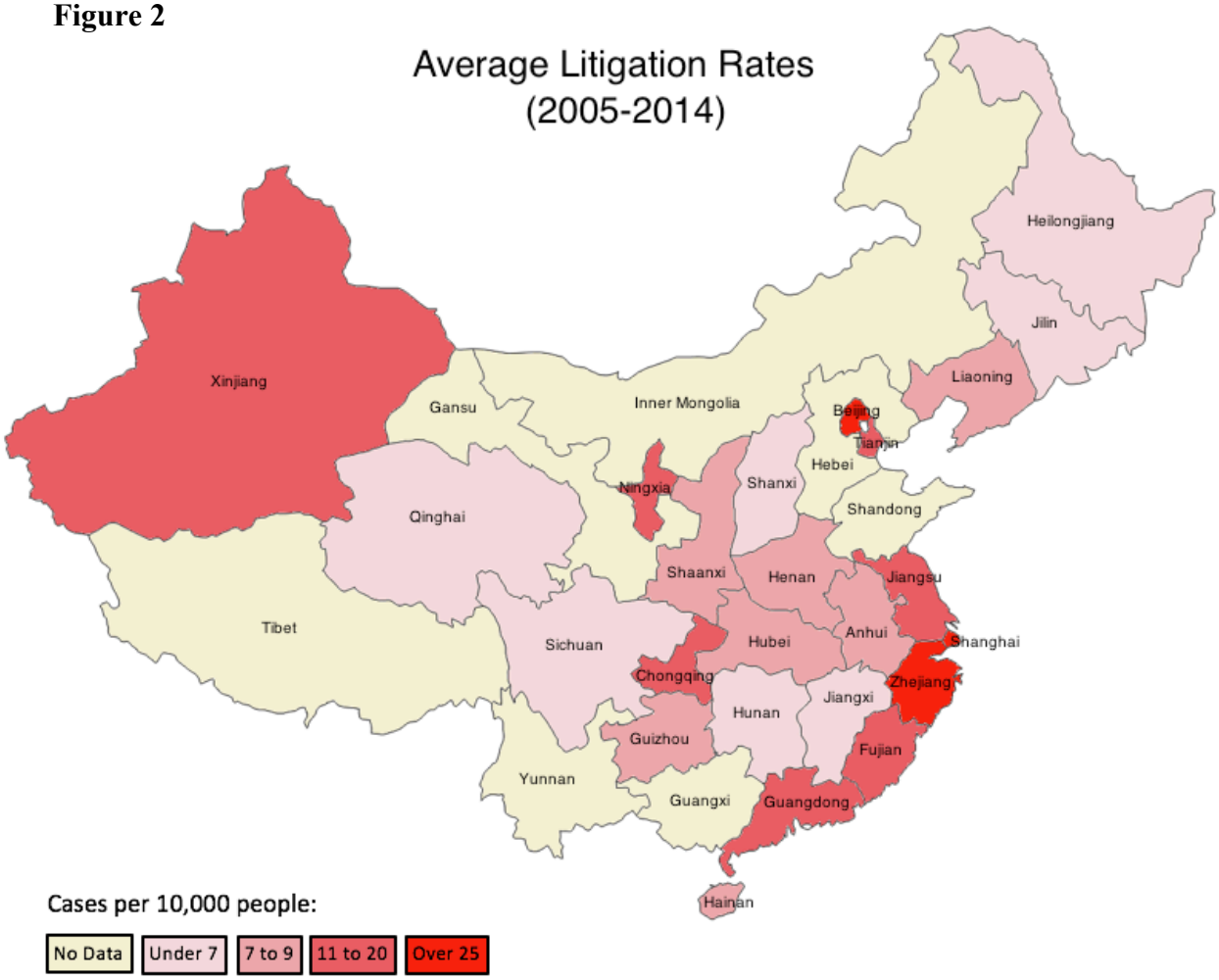
Summary statistics are reported in Table 2. It is interesting to note that for all variables except urbanization rate, variation between provinces is at least as great as variation within provinces. This result illustrates the vast disparities that exist within Chinese society. Consider that our analysis spans almost three decades in which China underwent significant economic, social, and legal change, and yet, variation between provinces is still greater than that across time. Further investigation into these disparities is likely warranted and will be untaken in our analyses.

Figure



The graph shows litigation rates across 29 years (1986-2014) for 24 of China's 31 provinces, municipalities, and autonomous regions (collectively referred to as provinces in this paper). Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. We do not possess litigation rate data for seven provinces excluded from this table. Similarly, missing data points indicate that we do not possess litigation rate data for a given province-year. Panels are organized by that province's litigation rate in the most recent year of reporting, with the higher litigation rate provinces appearing before lower ones.

Figure 2



The map shows average litigation rates across 10 years (2005-2014) for 24 of China's 31 provinces, municipalities, and autonomous regions (collectively referred to as provinces in this paper). Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. We do not use the full 29-year timespan of this study (1986-2014) to calculate average litigation rates because litigation rates are more frequently missing in early years. Shades of red are used to group provinces by their average litigation rate, with darker shades of red indicating higher average litigation rates. Groups were chosen by natural breaks in the distribution of average litigation rates. We do not possess litigation rate data for the seven provinces in beige.

Data in Table 2 also reveal that although we have 586 province-year litigation rate observations, we are sometimes missing corresponding economic, social, or legal data. The most severe case of missing data relates to urbanization rate, which has data for only 483 of the 586 province-years. In such situations, analyses can proceed in at least three ways. First, analyses can be conducted using only those province-years for which all data are available. Second, variables with high rates of missing values can be excluded from the analyses. Third, missing values can be imputed and analyses can proceed using both known and imputed values. Each of these methods has advantages and disadvantages and none is clearly superior to the others. We utilize method one throughout the paper and note that findings are generally robust to the other two specifications. Results under methods two and three are shown in Appendix A.

A matrix of pairwise correlations among variables is presented in Table 3. A high degree of correlation exists between many of the regressors. Provinces with high GRP per capita, for instance, also tend to have high urbanization rates, Internet penetration rates, rates of tourism, and lawyer densities, while those with low values for one of these variables tend to have low values across the board. Despite these high correlations, near multicollinearity does not seem to be an issue, as standard errors do not appear susceptible to the inclusion or exclusion of regressors.

High correlations also exist between regressors and litigation rates. One possibility is that these correlations suggest preliminary support for our hypotheses. Another possibility is that these correlations are simply the result of an underlying temporal process in the variables.[11] To determine the true relationships between regressors and litigation rates, we must control for the influence of time in our analysis. The influence of time is partially addressed through the inclusion of province specific time trends. Given that litigation rate growth varies considerably across provinces, province specific time trends are likely preferable to a single time variable. The former allows the influence of time to vary by province, while the latter does not. As a further measure, we redo our analyses imposing a one-year autoregressive (AR1) structure among the error terms. An AR1 structure accounts for previous-year

[11] A temporal process is a time trend. If over time, both litigation rates and regressors are changing monotonically, we will observe a correlation between these variables even if no causal relationship exists.

litigation rates in the estimation of current-year litigation rates. After accounting for this influence, we can determine if other regressors provide additional insight into the development of litigation rates.[12]

[12] We also re-estimate the model without imposing an AR1 structure but with robust standard errors. The results of this analysis can found in Appendix A.

Table 2: Summary Statistics

Variable	Obs.	Mean	Standard Deviation			Min.	Max.
			Overall	Within	Between		
Litigation Rate	586	8.40	7.92	5.05	6.05	0.93	46.63
GRP per Capita (10,000s)	571	2.03	1.84	1.03	1.53	0.03	9.71
Urbanization Rate	483	0.45	0.17	0.15	0.09	0.15	0.89
High School Graduation Rate	564	38.21	18.90	6.50	17.91	9.43	88.89
Internet Penetration Rate	585	0.12	0.17	0.06	0.16	0.00	0.74
Tourism	578	272.75	453.56	386.16	240.85	1.48	2574.90
Lawyer Density	562	1.00	1.36	0.94	0.98	0.05	10.92
Mediation Rate	560	52.36	24.80	16.11	19.91	5.02	209.18
Privatization Rate	585	0.43	0.20	0.09	0.19	0.03	0.83

The table shows summary statistics for all variables in our analysis (column one) across 24 provinces and 2 9 years (1986-2016). All variables are measured at the province-year level. Litigation rates are measured at the end of each year, while all other variables are measured at the beginning of each year. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector. Columns two through eight provide observation counts, means, standard deviations, minimums, and maximums across all province-years. Within standard deviation refers to the average standard deviation within provinces across time. Between standard deviation refers to the average standard deviation within years across provinces.

Table 3: Pairwise Correlations

	Litigation Rate	GRP per Capita	Urban. Rate	HS Grad. Rate	Internet Pen. Rate	Tourism	Lawyer Density	Mediation Rate	Private. Rate
Litigation Rate	1.00								
GRP per Capita	0.77	1.00							
Urbanization Rate	0.68	0.75	1.00						
HS Grad Rate	0.33	0.45	0.29	1.00					
Internet Pen. Rate	0.69	0.87	0.60	0.60	1.00				
Tourism	0.72	0.70	0.77	0.08	0.54	1.00			
Lawyer Density	0.79	0.74	0.65	0.14	0.63	0.80	1.00		
Mediation Rate	0.19	0.22	0.21	-0.05	0.25	0.30	0.28	1.00	
Privatization Rate	0.50	0.65	0.33	0.73	0.75	0.25	0.38	-0.01	1.00

The table shows pairwise correlations for all variables in our analysis across 24 provinces and 29 years (1986-2016). All variables are measured at the province-year level. Litigation rates are measured at the end of each year, while all other variables are measured at the beginning of each year. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Results of our regressions are shown in Table 4 and suggest that over 64 percent of all variation in litigation rates is explained by the regressors (including province specific time trends and year fixed effects not reported). In examining the coefficients and p-values of individual regressors, we find support for most, but not all, of our hypotheses. Specifically, GRP per Capita, High School Graduation Rate, Internet Penetration Rate, Tourism, Lawyer Density, and Privatization Rate appear consistent with our hypotheses and statistically significant in all regressions, indicating that economic, social, and legal factors are each related to the development of legal claims. On the other hand, Urbanization Rate and Mediation Rate are not statistically different from zero in any of the regressions.

We also test for the presence of a non-linear relationship between economic growth and litigation. Some studies suggest that economic growth may exhibit a positive but diminishing influence on the presence of litigation. We do not find support for this claim. If a positive but diminishing relationship does exist, we should expect a positive coefficient on GRP per Capita and a negative coefficient on GRP per Capita^2; however, in both model specifications which include GRP per Capita^2, the negative coefficient on GRP per Capita^2 is not statistically significant. One possibility is that China has not yet reached a stage of economic development in which litigation claims begin to level off.

A closer inspection of variable coefficients helps to draw additional information which can be used to distinguish their relative magnitudes. Looking across the four regression specifications, we find that:

- A 10,000 Yuan (about US$1,500) increase in GRP per capita is associated with 1.09 to 2.79 additional lawsuits per 10,000 people.[13]
- One new high school graduate per 10,000 people is associated with 0.039 to 0.065 additional lawsuits per 10,000 people.
- One additional tourist visit per 10,000 people is associated with 0.0016 to 0.0019 additional lawsuits per 10,000 people.
- A one percent increase in the proportion of people accessing the Internet for at least one hour per week is associated with 8.21 to 15.18 additional lawsuits per 10,000 people.
- One additional lawyer per 10,000 people is associated with 2.37 to 3.19 additional lawsuits per 10,000 people.
- A one percent increase in the proportion of assets owned by the private sector is associated with 4.87 to 5.56 additional lawsuits per 10,000 people.

[13] A 10,000 Yuan increase represents a roughly 20% increase in average GRP per capita.

Table 4: Regression Results – All Regions

Dependent Variable = Litigation Rate	*All Provinces*			
	Fixed Effects	*Fixed Effects*	*FE & AR1*	*FE & AR1*
GRP per Capita (10,000s)	1.09**	1.83*	1.63**	2.79**
	(0.52)	(0.97)	(0.76)	(1.35)
GRP per Capita^2 (10,000s)		-0.0000057		-0.0000086
		(0.0000063)		(0.0000083)
Urbanization Rate	1.21	0.77	0.48	0.055
	(3.07)	(3.11)	(3.71)	(3.73)
High School Graduation Rate	0.065***	0.052**	0.058**	0.039
	(0.021)	(0.025)	(0.027)	(0.033)
Internet Penetration Rate	15.18***	14.84***	9.04*	8.21
	(4.57)	(4.59)	(5.37)	(5.43)
Tourism	0.0019**	0.0016*	0.0018*	0.0016
	(0.00090)	(0.00094)	(0.0010)	(0.0011)
Lawyer Density	3.15***	3.19***	2.37***	2.39***
	(0.40)	(0.40)	(0.56)	(0.56)
Mediation Rate	-0.013	-0.012	-0.0028	-0.0012
	(0.010)	(0.0094)	(0.011)	(0.010)
Privatization Rate	5.47**	4.92**	5.56**	4.87*
	(2.19)	(2.28)	(2.60)	(2.68)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	441	441	417	417
R^2	0.65	0.66	0.64	0.67

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. In the regressions titled "fixed effects" (columns two and three), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns four and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

From this set of results, we consider the relative impact of GRP per Capita, High School Graduation Rate, Internet Penetration Rate, Tourism, Lawyer Density, and Privatization Rate on the development of litigation rates. One way to compare the influence of these variables is by determining how rapidly they increase in value over time and then normalizing their effect on litigation rates by a common time standard. Specifically, we determine the average annual change for a given variable over the 29-year time horizon of this study and multiply it by the variable's regression coefficient. In doing so, we estimate the average annual change in litigation associated with a particular factor after controlling for the influence of other conditions. The results of this method are displayed in Table 5.

The analysis in Table 5 shows the relative importance of statistically significant regressors on the development of litigation rates. Under three of the four model specifications (fixed effects with GRP^2, fixed effects with AR1, fixed effects with both GRP^2 and AR1), we observe that lawsuits most strongly relate to increases in GRP per Capita, followed by Lawyer Density, Internet Penetration Rates, High School Graduation Rates, Privatization Rates and finally Tourism. Under the fourth specification (fixed effects without GRP^2 and AR1), we observe an alternate ordering of the top three variables, with Internet Penetration Rates most strongly relating to increases in lawsuits, followed by Lawyer Density and GRP per Capita. In most cases, differences between ranks are small and no single factor appears to greatly outweigh the importance of the others.

The previous analysis relies on estimates of regression coefficients. If regression coefficients are unstable due to collinearity among regressors, the results listed in Table 5 may be incorrect. Given the high correlations between certain pairs of variables (e.g. GRP per capita and Urbanization Rate), it is especially important to investigate this possibility. We do so by re-estimating each regression model eight times, excluding one of the eight independent variable each time. If collinearity is an issue, the omission of certain variables should produce dramatic shifts in the estimated coefficients of other variables. In the absence of collinearity, fluctuations will be smaller but non-zero due to omitted variable bias introduced through the exclusion of specific regressors. We then examine the range of estimated coefficients for each variable and calculate the average annual change in litigation rates for the minimums and maximums of those ranges.

Table 5: Relative Influence of Regressors on Litigation Rates

	Average Annual Change	Regression Coefficient				Average Annual Change in Litigation Rates			
		FE	FE & GRP^2	FE & AR1	FE, GRP^2 & AR1	FE	FE & GRP^2	FE & AR1	FE, GRP^2 & AR1
GRP per Capita (10,000s)	0.16	1.09	1.83	1.63	2.79	**0.174**	**0.293**	**0.261**	**0.446**
Lawyer Density	0.071	3.15	3.19	2.37	2.39	**0.224**	**0.226**	**0.168**	**0.170**
Internet Penetration Rate	0.015	15.18	14.84	9.04	8.21	**0.228**	**0.223**	**0.136**	**0.123**
Privatization Rate	0.022	5.47	4.92	5.56	4.87	**0.120**	**0.108**	**0.122**	**0.107**
High School Graduation Rate	1.79	0.065	0.052	0.058	0.039	**0.116**	**0.093**	**0.104**	**0.070**
Tourism	15.88	0.0019	0.0016	0.0018	0.0016	**0.030**	**0.025**	**0.029**	**0.025**

The table shows average annual changes in litigation rates associated with statistically significant regressors (columns seven through ten). The average annual change in litigation rates associated with a particular variable is calculated as the product of that variable's average annual change (column two) and its regression coefficient (columns three through six). Average annual changes are found by regressing each variable on a linear time trend. Litigation rate refers to the number of litigated cases per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. GRP per capita refers to real gross regional product per capita in 10,000s. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Tourism refers to the number of international tourist visits per 10,000 people.

The results of these test can be found in Appendix B. Across the eight variable-excluded regressions, the largest coefficients on High School Graduation Rate and Tourism are just over four times as large as the smallest, while the largest coefficients on other variables are less than double the smallest. When minimum and maximum coefficients are used to calculate the average annual change in litigation rates associated with each variable, the rank ordering shown in Table 5 is largely maintained.

We turn next to the variations which are observed across China's provinces. Summary statistics in Table 2 indicate that variation across provinces is high and in many cases, exceeds that within provinces across time. It may be that the conditions relevant to litigation growth vary based on a province's level of economic, social, and legal development. To determine if this is the case, we analyze separately highly developed provinces and less developed provinces. To identify a province's level of development, we make use of a metric known as the Human Development Index (HDI). The HDI was developed by the United Nations Development Programme (UNDP) in 1990 to measure the level of social and economic development in a region. The 2016 China National Human Development Report indicates that HDI is a function of per capita income, educational attainment, and health, though the the actual formula for HDI is not provided.

The most recently released HDI data for China's provinces are from 2014 and can be found in the 2016 China National Human Development Report. A graph of province HDI values can be found in Appendix C. For the purposes of our analysis, we group provinces by their 2014 HDI values. Provinces with HDI values that exceed the 2014 national average HDI of 0.754 are classified as "high HDI provinces," while those with HDI values that trail the national average are classified as "low HDI provinces." Two alternative analyses can be found in Appendix C. The first redefines high and low HDI provinces as those on the top and bottom one-third of the HDI scale, while the second uses HDI values from 1999, the earliest year in which this variable is available for China's provinces. These analyses indicate that results are mostly robust to changes in the definitions of high and low HDI provinces.

Table 6 shows the results of an analysis of high and low HDI provinces.[14] Results suggest a clear divide between the factors influencing litigation rates in these areas. In high HDI provinces, GRP per Capita, High School Graduation Rate, and Lawyer Density are all positive and statistically significant, whereas in low HDI provinces, Internet Penetration Rate and Tourism are positive and statistically significant in one of our two specifications and Lawyer Density and Mediation Rate are negative and statistically significant. These findings indicate that our proxies for attitudes toward claiming and alternative methods of dispute resolution are more important in less developed areas, while economic, intellectual, and legal resources are more important in highly developed areas.

It is interesting to note that the coefficient on lawyer density is negative and statistically significant in low HDI provinces. One possibility is that lawyers in these areas are disproportionately employed by powerful defendants such as corporate or government entities and thus, reduce the likelihood of plaintiff success. Such a situation is expected when defendants are relatively large entities with more resources to hire lawyers.

We again consider the relative importance of statistically significant regressors on the development of litigation rates, this time for high and low HDI provinces. Table 7 shows the results of this analysis. In high HDI provinces, we observe that litigation rates most strongly relate to increases in GRP per Capita, followed by Lawyer Density, and High School Graduation Rates. The difference between GRP per Capita and the other two ranks is large, indicating a dominant influence of this variable in high HDI provinces. In low HDI provinces, we observe that changes in Internet penetration rates yield the greatest gains in litigation rates, followed by tourism, and then mediation rates. The average annual change in litigation rates associated with lawyer density is negative, indicating that litigation declines as the number of lawyers increases in low HDI provinces.

[14] To be concise, we omit regressions which include GRP per Capita^2 from Table 6. These regressions can be found in Appendix A and do not suggest a non-linear relationship between GRP per Capita and litigation rates.

Table 6: Regression Results – High and Low HDI Regions

Dependent Variable = Litigation Rate	*High HDI Provinces*		*Low HDI Provinces*	
	Fixed Effects	*FE & AR1*	*Fixed Effects*	*FE & AR1*
GRP per Capita (10,000s)	7.03***	7.69***	-0.43	-0.69
	(1.76)	(2.14)	(0.42)	(0.66)
Urbanization Rate	-2.19	0.62	-3.13	-3.22
	(12.28)	(14.20)	(2.31)	(2.73)
High School Graduation Rate	0.14***	0.14***	-0.020	-0.038
	(0.042)	(0.051)	(0.027)	(0.035)
Internet Penetration Rate	0.016	-5.52	12.54*	10.29
	(8.19)	(9.63)	(7.27)	(7.83)
Tourism	0.00053	0.00030	0.0060**	0.0030
	(0.0013)	(0.0015)	(0.0028)	(0.0030)
Lawyer Density	2.72***	2.35***	-2.82**	-3.18**
	(0.57)	(0.71)	(1.14)	(1.34)
Mediation Rate	0.017	0.021	-0.026**	-0.023*
	(0.018)	(0.019)	(0.010)	(0.012)
Privatization Rate	5.21	4.70	-0.43	-0.15
	(4.71)	(5.37)	(2.26)	(2.59)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	182	172	259	245
R^2	0.60	0.55	0.47	0.41

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. The first two regressions (columns two and three) include only observations from provinces with 2014 human development index values above the national average for that year. The second two regressions (columns four and five) include only observations from provinces with 2014 human development index values below the national average for that year. In the regressions titled "fixed effects" (columns two and four), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns three and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Table 7: Relative Influence of Regressors on Litigation Rates – High and Low HDI Regions

High HDI Provinces

	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		FE	FE & AR1	FE	FE & AR1
GRP per Capita (10,000s)	0.21	7.03	7.69	**1.476**	**1.615**
Lawyer Density	0.12	2.72	2.35	**0.326**	**0.282**
High School Graduation Rate	1.49	0.14	0.14	**0.209**	**0.209**

Low HDI Provinces

	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		FE	FE & AR1	FE	FE & AR1
Internet Penetration Rate	0.12	12.54	10.29	**1.505**	**1.235**
Tourism	8.10	0.0060	0.0030	**0.049**	**0.024**
Mediation Rate	-0.44	-0.026	-0.023	**0.011**	**0.010**
Lawyer Density	0.039	-2.82	-3.18	**-0.110**	**-0.124**

The table shows average annual changes in litigation rates associated with statistically significant regressors in regressions of high and low HDI provinces (column four). The average annual change in litigation rates associated with a particular variable is calculated as the product of that variable's average annual change (column two) and its regression coefficient (column three). Average annual changes are found by regressing each variable on a linear time trend. Litigation rate refers to the number of litigated cases per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. GRP per capita refers to real gross regional product per capita in 10,000s. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Tourism refers to the number of international tourist visits per 10,000 people.

The fact that economic, educational, and legal factors are relevant to litigation rates in high HDI provinces but not in low HDI provinces raises an interesting possibility – that some minimal level of development is required before resources necessary for litigating become important. To investigate this possibility, we begin by examining the relationship between HDI and litigation. Although HDI values are only available at the province level for five years of our study, 1999, 2006, 2008, 2010, and 2014, we can use our existing dataset along with information about the HDI metric to predict province-level HDI values in other years.

Recall that HDI is a function of per capita income, educational attainment, and health. Our dataset contains information for each of these measures. As such, it is possible to construct a predictive model for HDI, as shown in Equation 2. We estimate this model using province-level HDI data for the five years of observable HDI values. We then use estimated variable coefficients along with independent variable data to predict HDI values for other years in our study period.

Equation 2: Predictive Model for HDI

$$\text{HDI} = \alpha + \text{GRP per Capita} + \gamma\ \text{High School Graduation Rate} + \delta\ \text{Medical Staff per Capita} + \varepsilon$$

Where:

- α is a constant term.
- GRP per Capita and High School Graduation Rate are defined as specified in Table 1.
- Medical Staff per Capita is defined as the number of licensed doctors, nurses, pharmacists, laboratory technicians, imaging staff, health care supervisors, and technical personnel per 10,000 people.

Regression results from an estimation of Equation 2 can be found in Appendix C. These results suggest that our predictive model accounts for the majority of variation in HDI values, with an R^2 value of 0.86. As a result, predicted values for HDI are quite close to observed values. On average, predicted HDI values differ from actual HDI values by less than three percent.

Figure 3 shows a scatterplot of predicted and actual HDI values versus litigation rates. From this graph, we observe that litigation rates do not begin to increase until HDI is nearly 0.6. Beyond this point, variation in litigation rates increases dramatically. Thus, it appears that some requisite level of development is necessary for litigation growth.

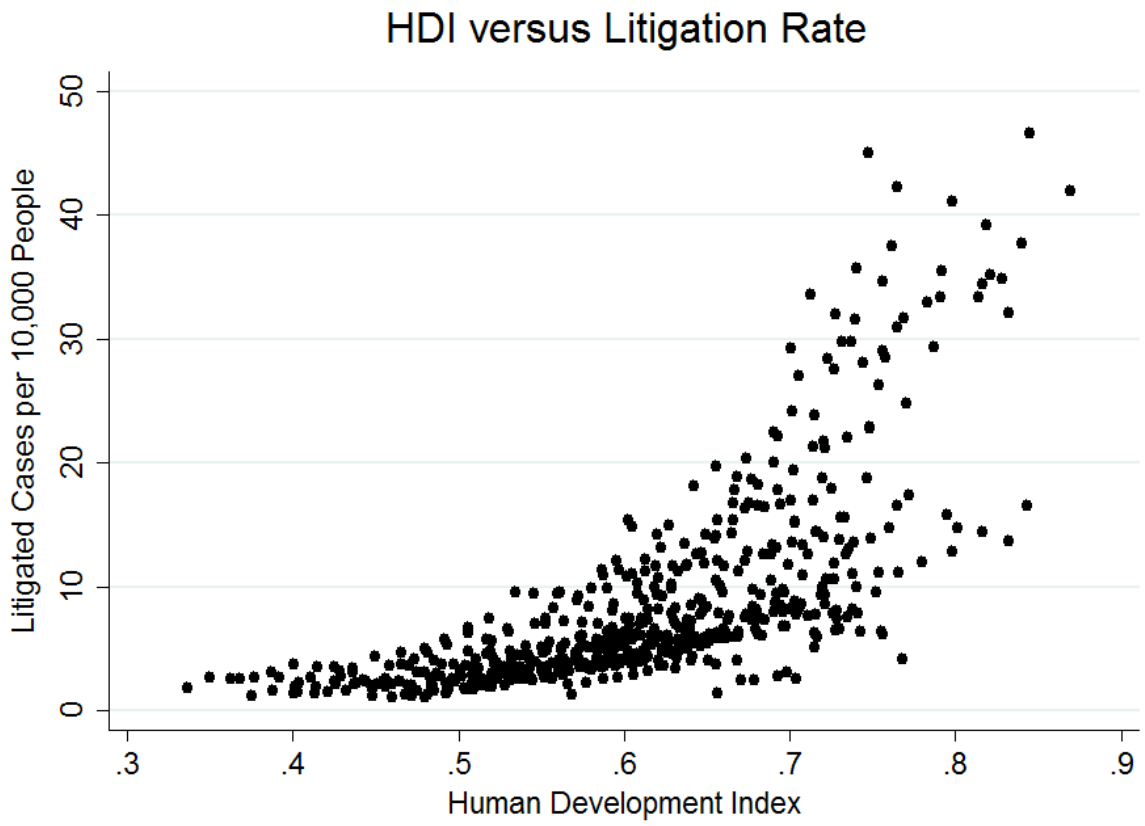
To estimate the HDI value associated with this level of development, we regress litigation rate on HDI and "High HDI," where High HDI is equal to HDI when HDI is greater than some point P and 0 otherwise. We can think of this regression as a procedure for fitting a segmented best fit line to Figure 3, where the slope of the line changes at the point P. By altering P, we can determine the value of P that produces the "sharpest" kink in the line. We define "sharpest" to mean the greatest difference in Z-score between the high and low portions of the best fit line. By examining the change in Z-score, as opposed to the change in slope, we account for both the change in slope and the uncertainty surrounding that estimate.

The results of this empirical procedure can be found in Appendix C. We observe that the greatest change in Z-score occurs at an HDI level of 0.57. It is our hypothesis that resources necessary for litigation growth will be important once a province has achieved this threshold, but not before. To test this hypothesis, we split province-year observations into two groups: those with HDI values greater than 0.57 and those with HDI values less than 0.57. We then re-estimate our model for each group.

Table 8 shows the results of this analysis. We see that when HDI is less than 0.57, economic, social, and legal variables do not exhibit a positive relationship with litigation rates. Conversely, when HDI exceeds 0.57, GRP per capita, High School Graduation Rate, Tourism, Lawyer Density and Privatization Rate each display a positive and statically significant relationship with litigation rates.[15] These results support our hypothesis that a minimal level of development is required before resources necessary for litigating become important.

[15] An analysis of the relative influence of these variables on litigation rates can be found in Appendix C. Results are similar to those associated with high HDI provinces in Table 2 except that privatization is determined to be the second most influential factor after GRP.

Figure 3



The table shows predicted and actual human development index values versus litigation rates. Predicted human development index values are generated from a regression equation in which actual human development index values are regress on gross regional product per capita, high school graduation rates, and medical staff per capita. Litigation rate refers to the number of litigated cases per 10,000 people.

Table 8: Regression Results – HDI Above and Below 0.57

Dependent Variable = Litigation Rate	*HDI > 0.57*		*HDI < 0.57*	
	Fixed Effects	*FE & AR1*	*Fixed Effects*	*FE & AR1*
GRP per Capita (10,000s)	1.66**	2.52***	-3.41	-4.31
	(0.82)	(1.06)	(2.77)	(4.23)
Urbanization Rate	3.55	-3.12	1.95	-1.15
	(4.71)	(7.85)	(4.12)	(4.62)
High School Graduation Rate	0.087***	0.11***	0.016	0.022
	(0.027)	(0.038)	(0.060)	(0.089)
Internet Penetration Rate	4.34	1.23	-114.40	-30.87
	(5.74)	(6.30)	(173.06)	(185.69)
Tourism	0.0027**	0.0027**	-0.00053	-0.0014
	(0.0012)	(0.0013)	(0.0010)	(0.0020)
Lawyer Density	2.95***	1.97***	-2.88	-2.70
	(0.46)	(0.62)	(1.94)	(2.40)
Mediation Rate	-0.016	-0.0017	-0.0025	0.0020
	(0.012)	(0.013)	(0.014)	(0.015)
Privatization Rate	10.03***	8.52**	2.71	0.99
	(3.11)	(4.01)	(2.66)	(2.98)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	342	318	99	85
R^2	0.65	0.58	0.19	0.23

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. The first two regressions (columns two and three) include only province-year observations for which actual or predicted human development index values are above 0.57. The second two regressions (columns four and five) include only province-year observations for which actual or predicted human development index values are below 0.57. In the regressions titled "fixed effects" (columns two and four), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns three and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

IV. Concluding Remarks

As China's economy has grown and expanded over the past thirty years, so too has its legal system. All provinces in our study display some amount of litigation growth over this timespan, but the magnitude of this growth varies considerably across provinces. In many cases, developed regions have witnessed the greatest gains, while less developed inland areas have seen smaller increases in litigation rates.

A variety of economic, social, and legal factors appear to be associated with rising litigation rates. Among these factors, economic activity, access to lawyers, and Internet use seem to have the greatest influence. It is interesting to note that tourism and privatization also appear relevant, as these variables capture trends that are somewhat unique to China, namely, the opening of China to foreign travel and a push toward greater private ownership.

A closer inspection of factors reveals that their importance is tied to a province's overall level of development. Prior to some requisite level of development, no correlation seems to exist between factors and litigation rates. Upon reaching the requisite level, positive correlations emerge. Our analysis indicates that this necessary level of development roughly corresponds with a human development index value of 0.57. Testing for a similar, non-linear relationship between factors and litigation rates in other emerging economies would be beneficial.

The findings of this study may be useful in setting public policy in China and in preparing for legal claims. Additionally, the study contributes to a more general understanding of litigation in developing economies – an area which has received relatively little attention in light of data limitations for these jurisdictions.

The current analysis is not immune to such data limitations. Ideally, one would incorporate more detailed information regarding the experience of potential plaintiffs, including costs to file a lawsuit, lawyer fees, damages awarded, and the probability of plaintiff success via settlement or trial. In addition, complete information for each variable in each province-year would be preferred. Gathering such data would allow for a richer understanding of the specific mechanisms which shape the decision to litigate in China.

Bibliography

American Bar Association. (1994). Comprehensive legal needs study. *Chicago, American Bar Association*.

Atkins, B. M., & Glick, H. R. (1976). Environmental and structural variables as determinants of issues in state courts of last resort. *American Journal of Political Science*, 97-115.

Bernstein, L. (1992). Opting out of the legal system: Extralegal contractual relations in the diamond industry. *The Journal of Legal Studies*, 21(1), 115-157.

Bezanson, R. P., Cranberg, G., & Soloski, J. (1987). *Libel law and the press: Myth and reality*. Free Pr.

Bujakowski, D. (2016). Legal Liability in China: A Brief History and Current Practice. *Working Paper*.

Campbell, B., & Talarico, S. M. (1982). Access to Legal Services: Examining Common Assumptions. *Judicature*, 66, 313.

Carlin, J. E., Howard, J., & Messinger, S. L. (1966). Civil justice and the poor: issues for sociological research. *Law and Society Review*, 9-89.

Carmignani, A., & Giacomelli, S. (2010). Too many lawyers? Litigation in Italian civil courts.

Cheatham, E. E. (1963). A Lawyer When Needed: Legal Services for the Middle Classes. *Columbia Law Review*, 63(6), 973-986.

Cooter, R. D., & Rubinfeld, D. L. (1989). Economic analysis of legal disputes and their resolution. *Journal of Economic Literature*, *27*(3), 1067-1097.

Daniels, S. (1982). Civil Litigation in Illinois Trial Courts-An Exploration of Rural-Urban Differences. *Law & Pol'y Q., 4*, 190.

Eisenberg, T. (2010). The need for a national civil justice survey of incidence and claiming behavior. *Fordham Urb. LJ*, 37, 17.

Eisenberg, T., Kalantry, S., & Robinson, N. (2012). Litigation as a measure of well-being. *DePaul L. Rev*., 62, 247.

Eling, M., Pradhan, S., & Schmit, J. T. (2014). The determinants of microinsurance demand. *The Geneva Papers on Risk and Insurance Issues and Practice, 39*(2), 224-263.

Ellickson, R. C. (1994). The aim of order without law. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 150(1), 97-100.

Engel, D. M. (2012). Perception and Decision at the Threshold of Tort Law: Explaining the Infrequency of Claims. *DePaul L. Rev., 62*, 293.

Felstiner, W. L., Abel, R. L., & Sarat, A. (1980). The Emergence and Transformation of Disputes: Naming, Blaming, Claiming... *Law and society review*, 631-654.

Friedman, L. M. (1985). *Total justice*. Russell Sage Foundation.

Fu, J. (2006). Corporate Governance in China: A Research Agenda for a Corporate Group and Shareholder Control Perspective*. Journal of Business Systems, Governance and Ethics, 1*(1), 1.

Galanter, M. (1974). Why the" haves" come out ahead: Speculations on the limits of legal change. *Law & society review*, 9(1), 95-160.

Genn, H. G., & Beinart, S. (1999). *Paths to justice: what people do and think about going to law*. Hart Publishing.

Ginsburg, T., & Hoetker, G. (2006). The unreluctant litigant? An empirical analysis of Japan's turn to litigation. *The Journal of Legal Studies*, 35(1), 31-59.

Glofcheski, R. A. (2012). *Tort Law in Hong Kong*. Sweet & Maxwell/Thomson Reuters.

Grossman, J. B., & Sarat, A. (1974). Litigation in the federal courts: A comparative perspective. *Law & Soc'y Rev.*, 9, 321.

Halek, M., & Eisenhauer, J. G. (2001). Demography of risk aversion. *Journal of Risk and Insurance*, 1-24.

Harvard Medical Practice Study. (1990). *Patients, Doctors, and Lawyers: Medical Injury, Malpractice Litigation, and Patient Compensation in New York: a Report*. President and Fellows of Harvard College.

He, X., & Su, Y. (2013). Do the "haves" come out ahead in Shanghai courts?. *Journal of Empirical Legal Studies*, 10(1), 120-145.

Hensler, D. R., Marquis, M. S., Abrahamse, A., Berry, S. H., Ebener, P. A., Lewis, E., Lind, E., MacCoun, R.J., Manning, W.G., Rogowski, J. and Vaiana, M.E. (1991). *Compensation for accidental injuries in the United States*. Rand Corporation.

Hersch, J. (1996). Smoking, seat belts, and other risky consumer decisions: Differences by gender and race. *Managerial and decision economics*, 471-481.

Jacobi, T. (2009). The role of politics and economics in explaining variation in litigation rates in the US States. *The Journal of Legal Studies*, 38(1), 205-233.

McIntosh, W. (1983). Private use of a public forum: A long range view of the dispute processing role of courts. *The American Political Science Review*, 991-1010.

Mealey-Lohmann, L. (2010). Using mediation to resolve disputes – Differences between China and the United States. Retrieved from https://www.chinainsight.info/culture/chinese-3/526-using-mediation-to-resolve-disputes--differences-between-china-and-the-united-states-.html

Michelson, E. (2007). Climbing the dispute pagoda: Grievances and Appeals to the Official Justice System in Rural China. *American Sociological Review*,72(3), 459-485.

Miller, R. E., & Sarat, A. (1980). Grievances, claims, and disputes: Assessing the adversary culture. *Law and Society Review*, 525-566.

Mills, D. H. (1977). Report on the medical insurance feasibility study.

Mingming, S., & Yuhua, W. (2009). Litigating economic disputes in rural China. *China Review*, 97-121.

Morrill, C., Tyson, K., Edelman, L. B., & Arum, R. (2010). Legal mobilization in schools: the paradox of rights and race among youth. *Law & Society Review*, 44(3-4), 651-694.

Patel, A., Balmer, N. J., & Pleasence, P. (2008). Geography of advice seeking. *Geoforum*, 39(6), 2084-2096.

People's Mediation Law of the People's Republic of China. (2010)

Pleasence, P., Balmer, N. J., & Reimers, S. (2011). What really drives advice seeking behaviour? Looking beyond the subject of legal disputes.

Pocincki, L. S., Dogger, S. J., & Schwartz, B. P. (1973). The incidence of iatrogenic injuries. *Appendix, Report of the Secretary's Commission on Medical Malpractice (DHEW Publication No.[OS] 73-89). Washington, DC, Government Printing Office*, 50-70.

Ramseyer, J. M. (2014). Litigation and Social Capital: Divorces and Traffic Accidents in Japan. *Journal of Empirical Legal Studies*, *11*(1), 39-73.

Sandefur, R. L. (2009). The Fulcrum Point of Equal Access to Justice: Legal and Non-Legal Institutions of Remedy.

Sandefur, R. L. (2007). The importance of doing nothing: everyday problems and responses of inaction.

Shavell, S. (1982). Suit, settlement, and trial: A theoretical analysis under alternative methods for the allocation of legal costs. *The Journal of Legal Studies*, 11(1), 55-81.

UNDP China, & Development Research Center of the State Council of China. (2016). China National Human Development Report 2016: Social Innovation for Inclusive Human Development. *China Publishing Group Corporation, China Translation & Publishing House.*

Vidmar, N., & Schuller, R. A. (1987). Individual differences and the pursuit of legal rights: A preliminary inquiry. *Law and Human Behavior*, *11*(4), 299.

Yates, J., Davis, B. C., & Glick, H. R. (2001). The politics of torts: Explaining litigation rates in the American states. *State Politics & Policy Quarterly, 1*(2), 127-143.

Yates, J., Tankersley, H., & Brace, P. (2010). Assessing the impact of state judicial structures on citizen litigiousness. *Political Research Quarterly, 63*(4), 796-810.

Appendix A

Alternative regression analyses

Regression Results – All Regions (with Random Effects)

Dependent Variable = Litigation Rate	*All Provinces*			
	RE	*RE*	*RE & AR1*	*RE & AR1*
GRP per Capita (10,000s)	0.40 (0.41)	2.08** (0.85)	0.70 (0.51)	2.22** (1.02)
GRP per Capita^2 (10,000s)		-0.000013** (0.0000057)		-0.000011* (0.0000067)
Urbanization Rate	10.41*** (1.80)	9.33*** (1.85)	8.56*** (1.61)	7.55*** (2.27)
High School Graduation Rate	0.061*** (0.018)	0.030 (0.022)	0.054** (0.022)	0.028 (0.026)
Internet Penetration Rate	23.83*** (3.79)	23.65*** (3.77)	19.23*** (4.39)	18.98*** (4.38)
Tourism	0.00080 (0.00075)	0.00037 (0.00078)	0.00084 (0.00082)	0.00043 (0.00085)
Lawyer Density	3.21*** (0.30)	3.51*** (0.32)	3.05*** (0.37)	3.32*** (0.40)
Mediation Rate	-0.010 (0.077)	-0.0075 (0.0077)	-0.0057 (0.0086)	-0.0037 (0.0087)
Privatization Rate	8.25*** (1.94)	7.20*** (1.99)	7.35*** (2.24)	6.56*** (2.28)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	441	441	441	441
R^2	0.94	0.94	0.94	0.94

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province random effects, year fixed effects, and province-specific time trends. In the regressions titled "RE" (columns two and three), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "RE & AR1" (columns four and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Regression Results – All Regions (Urbanization Rate Omitted)

Dependent Variable = Litigation Rate	*All Provinces*			
	Fixed Effects	*Fixed Effects*	*FE & AR1*	*FE & AR1*
GRP per Capita (10,000s)	0.88*	1.51*	1.36**	1.94*
	(0.47)	(0.86)	(0.69)	(1.18)
GRP per Capita^2 (10,000s)		-0.0000048		-0.0000043
		(0.0000054)		(0.0000072)
High School Graduation Rate	0.038**	0.026	0.042*	0.033
	(0.018)	(0.023)	(0.025)	(0.030)
Internet Penetration Rate	24.00***	23.82***	14.95***	14.61***
	(3.64)	(3.65)	(4.60)	(4.64)
Tourism	0.0015**	0.0013*	0.0019**	0.0018*
	(0.00077)	(0.00080)	(0.00093)	(0.00096)
Lawyer Density	2.79***	2.82***	1.91***	1.92***
	(0.38)	(0.38)	(0.53)	(0.53)
Mediation Rate	-0.0043	-0.0033	0.0018	0.0025
	(0.0081)	(0.0082)	(0.0092)	(0.0093)
Privatization Rate	7.31***	7.00***	5.88***	5.69***
	(1.71)	(1.75)	(2.14)	(2.16)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	528	528	504	504
R^2	0.61	0.62	0.67	0.68

The table shows estimates from four regressions of litigation rates on the other variables in column one. Urbanization rate is not omitted from this model – one possible remedy for missing data associated with this variable. All regressions include province and year fixed effects and province-specific time trends. In the regressions titled "fixed effects" (columns two and three), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns four and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Regression Results – All Regions (Urbanization Rate Imputed)

Dependent Variable = Litigation Rate	*All Provinces*			
	Fixed Effects	*Fixed Effects*	*FE & AR1*	*FE & AR1*
GRP per Capita (10,000s)	0.93**	1.34	1.38**	1.84
	(0.47)	(0.86)	(0.68)	(1.18)
GRP per Capita^2 (10,000s)		-0.0000032		-0.0000034
		(0.0000055)		(0.0000072)
Urbanization Rate (Imputed)	4.84*	4.56*	3.70	3.50
	(2.66)	(2.71)	(3.42)	(3.45)
High School Graduation Rate	0.041**	0.033	0.042*	0.035
	(0.018)	(0.023)	(0.025)	(0.030)
Internet Penetration Rate	24.15***	24.02***	15.23***	14.95***
	(3.64)	(3.65)	(4.58)	(4.63)
Tourism	0.0017**	0.0015*	0.0020**	0.0019**
	(0.00077)	(0.00081)	(0.00093)	(0.00096)
Lawyer Density	2.80***	2.82***	1.91***	1.92***
	(0.38)	(0.38)	(0.53)	(0.53)
Mediation Rate	-0.0049	-0.0042	0.0014	0.0021
	(0.0081)	(0.0082)	(0.0092)	(0.0093)
Privatization Rate	7.30***	7.10***	6.04***	5.87***
	(1.71)	(1.74)	(2.13)	(2.16)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	528	528	504	504
R^2	0.65	0.66	0.70	0.70

The table shows estimates from four regressions of litigation rates on the other variables in column one. Missing urbanization rates values are imputed prior to estimation – one possible remedy for missing data associated with this variable. Values are imputed in the following manner:

1. For each province regress: $\text{Urbanization Rate}_{\text{Province}} = \alpha + \beta \times \text{Urbanization Rate}_{\text{National}} + \gamma \times \text{Year} + \delta \times \text{Year}^2 + \varepsilon$
2. For each province, obtain estimates of $\alpha, \beta, \gamma,$ and δ and the variance of ε.
3. Set missing values of $\text{Urbanization Rate}_{\text{Province}}$ equal to:
   $\hat{\alpha} + \hat{\beta} \times \text{Urbanization Rate}_{\text{National}} + \hat{\gamma} \times \text{Year} + \hat{\delta} \times \text{Year}^2 + \text{N}(0, \text{var}(\varepsilon))$

All regressions include province and year fixed effects and province-specific time trends. In the regressions titled "fixed effects" (columns two and three), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns four and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Regression Results (with Robust Standard Errors)

Dependent Variable = Litigation Rate	*All Provinces*		*High HDI Provinces*	*Low HDI Provinces*
GRP per Capita (10,000s)	1.09 (1.15)	1.83 (1.80)	7.03*** (1.77)	-0.43 (0.38)
GRP per Capita^2 (10,000s)		-0.0000057 (0.000011)		
Urbanization Rate	1.21 (2.98)	0.77 (2.88)	-2.19 (17.60)	-3.13 (3.06)
High School Graduation Rate	0.065 (0.045)	0.052 (0.045)	0.14* (0.065)	-0.020 (0.038)
Internet Penetration Rate	15.18* (8.94)	14.84* (8.53)	0.016 (12.49)	12.54 (9.89)
Tourism	0.0019* (0.00094)	0.0016* (0.0010)	0.00053 (0.0013)	0.0060 (0.0043)
Lawyer Density	3.15*** (0.64)	3.19*** (0.54)	2.72*** (0.55)	-2.82* (1.48)
Mediation Rate	-0.013 (0.011)	-0.012 (0.011)	0.017 (0.013)	-0.026** (0.0094)
Privatization Rate	5.47* (2.99)	4.92* (2.84)	5.21* (2.60)	-0.43 (3.12)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	441	441	182	259
R^2	0.65	0.66	0.60	0.47

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. The first two regressions (columns two and three) include observations from all provinces. The second two regressions include only observations with 2014 human development index values above and below the national average for that year, respectively. Coefficients and robust standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Appendix B

Excluded variable analyses

Regression Coefficients with Excluded Variables

Fixed Effects without GRP^2 and AR1

Excluded Var. →	None	GRP	Urban.	H.S.	Internet	Tourism	Lawyer	Med.	Private.
GRP	1.09**		1.09**	0.82	1.36***	1.14**	1.62***	1.14**	1.30**
GRP^2									
Urbanization	1.21	1.15		0.67	1.27	0.83	0.83	1.29	0.87
High School	0.065***	0.057***	0.064***		0.074***	0.058***	0.018	0.069***	0.076***
Internet	15.18***	16.66***	15.19***	17.00***		17.53***	20.30***	15.61***	12.93***
Tourism	0.0019**	0.0020**	0.0018**	0.0014	0.0026***		0.00063	0.0018**	0.0018**
Lawyer	3.15***	3.26***	3.14***	2.79***	3.34***	3.00***		3.18***	3.04***
Mediation	-0.013	-0.015	-0.013	-0.017*	-0.015*	-0.012	-0.017*		-0.018*
Privatization	5.47**	6.20***	5.43***	6.94***	4.04*	5.41**	3.63	6.07***	
Obs.	441	441	441	441	441	441	441	441	441
R2	0.65	0.65	0.64	0.60	0.66	0.63	0.64	0.64	0.66

Fixed Effects with GRP^2

Excluded Var. →	None	GRP	Urban.	H.S.	Internet	Tourism	Lawyer	Med.	Private.
GRP	1.83*		1.86*	2.64***	2.31**	2.27**	1.50	2.01**	2.48***
GRP^2	-.0000057		-.0000059	-.000013**	-.0000074	-.0000087	.00000089	-.0000068	-.0000094
Urbanization	0.77	1.15		-0.10**	0.70	0.22	0.90	0.76	0.20
High School	0.052**	0.057***	0.051**		0.057**	0.040*	0.020	0.053**	0.053**
Internet	14.84***	16.66***	14.83***	15.38***		16.56***	20.34***	15.16***	12.74***
Tourism	0.0016*	0.0020**	0.0016*	0.0011	0.0023**		0.00067	0.0015*	0.0015
Lawyer	3.19***	3.26***	3.20***	3.06***	3.39***	3.10***		3.23***	3.14***
Mediation	-0.012	-0.015	-0.012	-0.013	-0.014	-0.011	-0.017*		-0.015*
Privatization	4.92**	6.20***	4.87**	5.00**	3.37	4.57**	3.72	5.36**	
Obs.	441	441	441	441	441	441	441	441	441
R2	0.66	0.65	0.66	0.66	0.68	0.66	0.64	0.66	0.68

Fixed Effects with AR1

Excluded Var. →	None	GRP	Urban.	H.S.	Internet	Tourism	Lawyer	Med.	Private.
GRP	1.63**		1.63**	1.47*	1.90**	1.74**	2.27***	1.66**	1.83**
GRP^2									
Urbanization	0.48	0.20		0.98	0.71	0.19	1.16	0.48	-0.32
High School	0.058**	0.050*	0.058**		0.061**	0.054**	0.038	0.058**	0.067**
Internet	9.04*	11.14**	9.06*	9.64*		10.03*	7.75	8.95*	7.41
Tourism	0.0018*	0.0020**	0.0018*	0.0017*	0.0020**		0.0020*	0.0018*	0.0018*
Lawyer	2.37***	2.53***	2.37***	2.08***	2.35***	3.36***		2.35***	2.31***
Mediation	-0.0028	-0.0045	-0.0028	-0.0048	-0.0019	-0.00063	0.0016		-0.0053
Privatization	5.56**	6.21**	5.53**	6.38**	5.00*	5.55**	5.27*	5.63**	
Obs.	417	417	417	417	417	417	417	417	417
R2	0.65	0.67	0.64	0.61	0.63	0.61	0.63	0.64	0.63

Regression Coefficients with Excluded Variables (continued)

Fixed Effects with GRP^2 and AR1

Excluded Var. →	None	GRP	Urban.	H.S.	Internet	Tourism	Lawyer	Med.	Private.
GRP	2.79**		2.79**	3.44***	3.30**	3.25**	3.61**	2.82**	3.44***
GRP^2	-.0000086		-.0000086	-.000014**	-.000011	-.000011	.0000099	-.0000087	-.000012
Urbanization	0.055	0.20		0.019**	0.17	-0.32	0.75	0.052	-0.79
High School	0.039	0.050*	0.039		0.038	0.031	0.018	0.039	0.039
Internet	8.21	11.14**	8.21	7.99		8.78	6.78	8.11	6.53
Tourism	0.0016	0.0020**	0.0016	0.0014	0.0017		0.0017	0.0016	0.0015
Lawyer	2.39***	2.53***	2.39***	2.26***	2.37***	2.39***		2.38***	2.35***
Mediation	-0.0012	-0.0045	-0.0012	-0.0011	-0.000017	0.0011	-0.0034		-0.0026
Privatization	4.87**	6.21**	4.87*	4.85*	4.23	4.63*	4.56*	4.89*	
Obs.	441	441	441	441	441	441	441	441	441
R2	0.65	0.67	0.67	0.67	0.66	0.65	0.65	0.67	0.66

The tables show regression coefficients when specific variables are excluded from each analysis. The sample is held constant across these specifications. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Relative Influence of Regressors on Litigation Rates using Min. and Max. Regression Coefficients

Fixed Effects without GRP^2 and AR1

Variable	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		Min.	Max.	Min.	Max.
GRP per Capita (10,000s)	0.16	0.82	1.62	**0.131**	**0.259**
Lawyer Density	0.071	2.79	3.34	**0.198**	**0.237**
Internet Penetration Rate	0.015	12.93	20.30	**0.194**	**0.305**
Privatization Rate	0.022	3.63	6.94	**0.080**	**0.153**
High School Graduation Rate	1.79	0.018	0.076	**0.032**	**0.136**
Tourism	15.88	0.00063	0.0026	**0.010**	**0.041**

Fixed Effects with GRP^2

Variable	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		Min.	Max.	Min.	Max.
GRP per Capita (10,000s)	0.16	1.50	2.64	**0.240**	**0.422**
Lawyer Density	0.071	3.06	3.39	**0.217**	**0.241**
Internet Penetration Rate	0.015	12.74	20.34	**0.191**	**0.305**
Privatization Rate	0.022	3.37	6.20	**0.074**	**0.136**
High School Graduation Rate	1.79	0.020	0.057	**0.036**	**0.102**
Tourism	15.88	0.00067	0.0023	**0.011**	**0.037**

Fixed Effects with AR1

Variable	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		Min.	Max.	Min.	Max.
GRP per Capita (10,000s)	0.16	1.47	2.27	**0.235**	**0.363**
Lawyer Density	0.071	2.08	3.36	**0.148**	**0.239**
Internet Penetration Rate	0.015	7.41	11.14	**0.111**	**0.167**
Privatization Rate	0.022	5.00	6.38	**0.110**	**0.140**
High School Graduation Rate	1.79	0.038	0.067	**0.068**	**0.120**
Tourism	15.88	0.0017	0.0020	**0.027**	**0.032**

Relative Influence of Regressors on Litigation Rates using Min. and Max. Regression Coefficients (continued)

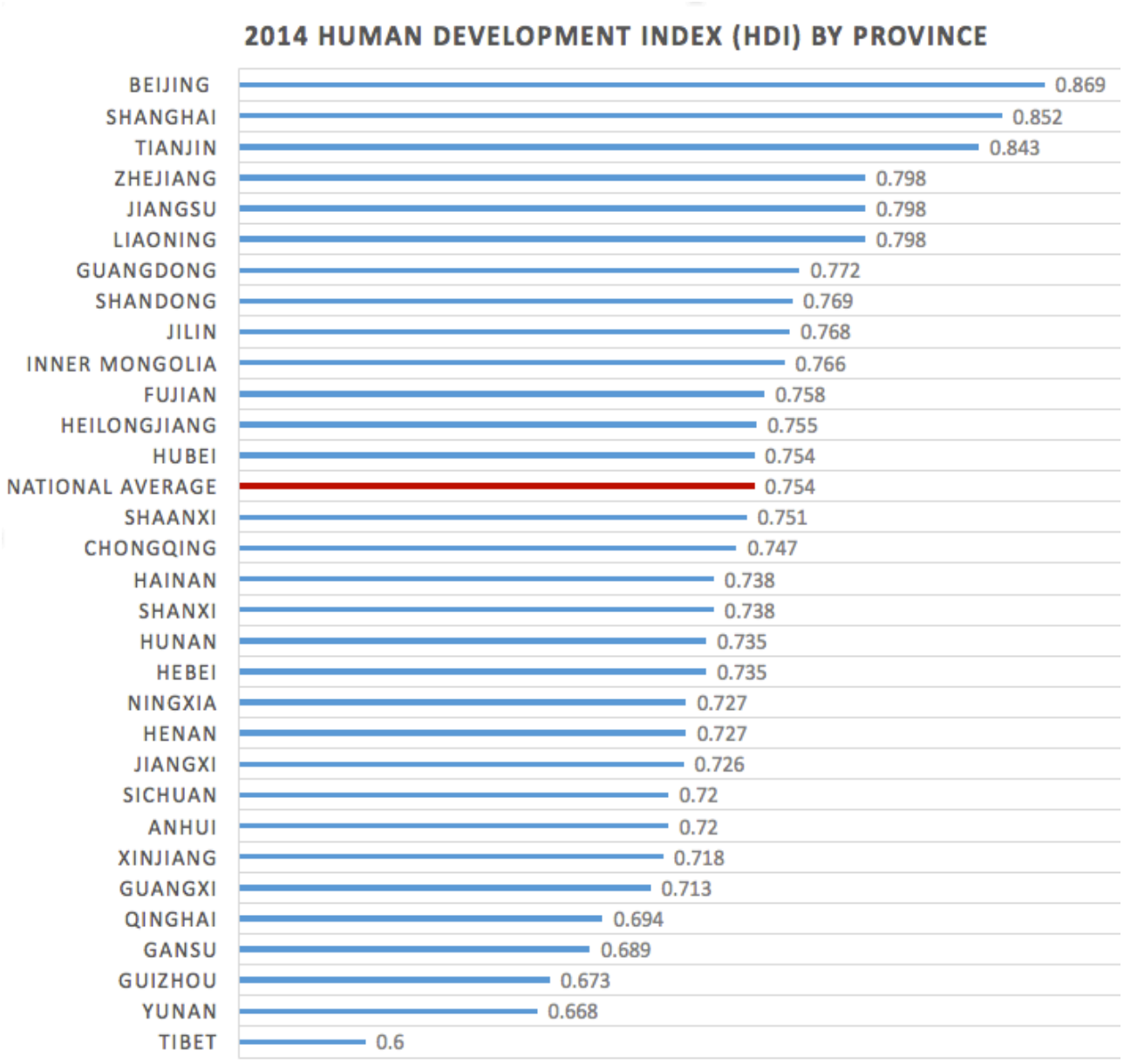
Fixed Effects with GRP^2 and AR1

Variable	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		Min.	Max.	Min.	Max.
GRP per Capita (10,000s)	0.16	2.79	3.61	**0.446**	**0.578**
Lawyer Density	0.071	2.26	2.53	**0.160**	**0.180**
Internet Penetration Rate	0.015	6.53	11.14	**0.098**	**0.167**
Privatization Rate	0.022	4.23	6.21	**0.093**	**0.137**
High School Graduation Rate	1.79	0.018	0.050	**0.032**	**0.090**
Tourism	15.88	0.0014	0.0020	**0.022**	**0.032**

The tables show average annual changes in litigation rates associated with statistically significant regressors (columns five and six). The average annual change in litigation rates associated with a particular variable is calculated as the product of that variable's average annual change (column two) and its regression coefficient (column three). Average annual changes are found by regressing each variable on a linear time trend. Minimum and maximum regression coefficients are determined through a series of regressions in which each regressor is individually excluded from the model. Litigation rate refers to the number of litigated cases per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. GRP per capita refers to real gross regional product per capita in 10,000s. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Tourism refers to the number of international tourist visits per 10,000 people.

Appendix C

Human development index metrics and analyses



The graph shows China's 2014 Human Development Index (HDI) by province. The HDI was developed by the United Nations Development Programme (UNDP) in 1990 to measure the level of social and economic development in a region. Provinces with HDI values that exceed the 2014 national average HDI of 0.754 are classified as "high HDI provinces," while those with HDI values that trail the national average are classified as "low HDI provinces."

Regression Results – High and Low HDI Regions (with GRP per Capita^2)

Dependent Variable = Litigation Rate	*High HDI Provinces*		*Low HDI Provinces*	
	Fixed Effects	*FE & AR1*	*Fixed Effects*	*FE & AR1*
GRP per Capita (10,000s)	3.04	3.22	0.83	-0.040
	(1.92)	(2.21)	(1.46)	(2.03)
GRP per Capita^2 (10,000s)	0.000053***	0.000057**	0.0000081	-0.0000039
	(0.000013)	(0.000015)	(0.0000089)	(0.000012)
Urbanization Rate	5.68	8.04	-3.10	-3.23
	(11.72)	(13.04)	(2.31)	(2.73)
High School Graduation Rate	0.30***	0.31***	-0.025	-0.041
	(0.054)	(0.062)	(0.028)	(0.036)
Internet Penetration Rate	-5.98	-8.46	12.64*	10.10
	(7.85)	(8.76)	(7.28)	(7.87)
Tourism	0.0018	0.0015	0.0052*	0.0029
	(0.0013)	(0.0015)	(0.0029)	(0.0031)
Lawyer Density	1.77***	1.62**	-2.98**	-3.22**
	(0.58)	(0.66)	(1.15)	(1.35)
Mediation Rate	0.0054	0.0064	-0.026***	-0.023*
	(0.017)	(0.018)	(0.010)	(0.012)
Privatization Rate	9.11**	7.97	-0.61	-0.16
	(4.54)	(4.93)	(2.27)	(2.60)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	182	172	259	245
R^2	0.36	0.31	0.48	0.41

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. The first two regressions (columns two and three) include only observations from provinces with 2014 human development index values above the national average for that year. The second two regressions (columns four and five) include only observations from provinces with 2014 human development index values below the national average for that year. In the regressions titled "fixed effects" (columns two and four), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns three and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Regression Results – High and Low HDI Regions (using Top and Bottom One-Third of HDI Spectrum)

Dependent Variable = Litigation Rate	*High HDI Provinces*		*Low HDI Provinces*	
	Fixed Effects	*FE & AR1*	*Fixed Effects*	*FE & AR1*
GRP per Capita (10,000s)	7.64***	8.33***	-2.44	-3.26
	(2.04)	(2.44)	(4.64)	(5.68)
Urbanization Rate	-17.91	-18.81	-37.89*	-24.86
	(15.42)	(17.73)	(20.66)	(21.77)
High School Graduation Rate	0.17***	0.17**	0.017	0.012
	(0.056)	(0.068)	(0.044)	(0.059)
Internet Penetration Rate	-12.37	-15.02	31.46**	31.05***
	(10.69)	(11.97)	(9.71)	(10.70)
Tourism	-0.00051	-0.00029	-0.0017	-0.0033
	(0.0016)	(0.0018)	(0.0022)	(0.0022)
Lawyer Density	2.03***	1.86**	-5.54***	-4.22**
	(0.67)	(0.71)	(1.62)	(2.02)
Mediation Rate	0.019	0.020	-0.032*	-0.0089
	(0.021)	(0.023)	(0.019)	(0.022)
Privatization Rate	7.19	6.93	3.25	5.09
	(5.86)	(6.58)	(3.74)	(4.24)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	141	133	150	142
R^2	0.52	0.49	0.69	0.77

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. The first two regressions (columns two and three) include only observations from provinces with 2014 human development index values in the top one-third of China's HDI distribution. The second two regressions (columns four and five) include only observations from provinces with 2014 human development index values in the bottom one-third of China's HDI distribution. In the regressions titled "fixed effects" (columns two and four), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns three and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Regression Results – High and Low HDI Regions (using HDI in 1999 instead of 2014)

Dependent Variable = Litigation Rate	*High HDI Provinces*		*Low HDI Provinces*	
	Fixed Effects	*FE & AR1*	*Fixed Effects*	*FE & AR1*
GRP per Capita (10,000s)	4.47***	5.94***	-0.56	-0.73
	(1.42)	(1.88)	(0.43)	(0.65)
Urbanization Rate	7.80	1.96	-2.04	-1.72
	(7.82)	(9.30)	(2.35)	(2.81)
High School Graduation Rate	0.11***	0.11***	-0.033	-0.041
	0.034	(0.046)	(0.027)	(0.035)
Internet Penetration Rate	3.35	-1.59	30.03***	23.85***
	(7.16)	(8.46)	(8.84)	(9.41)
Tourism	0.0015	0.0011	0.0021	-0.00020
	(0.0012)	(0.0014)	(0.0030)	(0.0034)
Lawyer Density	3.10***	2.45***	-2.80**	-2.16
	(0.533)	(0.71)	(1.14)	(1.36)
Mediation Rate	-0.013	-0.0030	-0.0032	0.0023
	(0.017)	(0.019)	(0.0089)	(0.0096)
Privatization Rate	7.34	5.96	-0.032	0.43
	(3.76)	(4.43)	(2.82)	(3.09)
Province Specific Time Trends	Yes	Yes	Yes	Yes
Observations	227	214	214	203
R^2	0.60	0.47	0.54	0.59

The table shows estimates from four regressions of litigation rates on the other variables in column one. All regressions include province and year fixed effects and province-specific time trends. The first two regressions (columns two and three) include only observations from provinces with 1999 human development index values above the national average for that year. The second two regressions (columns four and five) include only observations from provinces with 1999 human development index values below the national average for that year. In the regressions titled "fixed effects" (columns two and four), error terms are assumed to be independent and identically distributed, whereas in the regressions titled "FE & AR1" (columns three and five), error terms are assumed to follow an AR1 process. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Predictive Model for HDI

Dependent Variable = HDI	*Fixed Effects*
GRP per Capita (10,000s)	0.025***
	(0.0015)
High School Graduation Rate	0.0011***
	(0.00014)
Medical Staff per Capita	0.0010***
	(0.00023)

Observations	120
R^2	0.88

The table shows estimates from a regression of HDI on GRP per Capita, High School Graduation Rate, and Medical Staff per Capita. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The constant term is not reported. HDI refers to human development index. GRP per capita refers to real gross regional product per capita in 10,000s. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Medical Staff per Capita refers to the number of licensed doctors, nurses, pharmacists, laboratory technicians, imaging staff, health care supervisors, and technical personnel per 10,000 people. Province-level HDI data are are available for five years in our study: 1999, 2006, 2008, 2010, 2014. Estimated parameters are used to predict HDI values for the other years of our study.

Tipping Point Identification

$$\text{Litigation Rate} = \alpha + \beta\ \text{HDI} + \gamma\ \text{High HDI} + \varepsilon$$

Tipping Point	HDI Z-Score	High HDI Z-Score	Difference
0.54	1.38	1.43	0.05
0.55	2.09	4.12	2.03
0.56	3.83	7.46	3.63
0.57	**6.18**	**10.64**	**4.46**
0.58	7.77	11.99	4.22
0.59	9.65	12.46	2.81
0.60	10.81	13.19	2.38
0.61	12.21	13.83	1.62
0.62	13.17	14.59	1.42
0.63	14.56	15.25	0.69

The table shows Z-scores for HDI and High HDI for various definitions of High HDI. Z-scores are obtained from a regression of litigation rate on HDI and High HDI, shown above the table. Litigation rate refers to the total number of litigated cases involving a lawyer per 10,000 people. HDI refers to refers to human development index.
High HDI is equal to HDI when HDI is greater than some tipping point P and 0 otherwise.

Relative Influence of Regressors on Litigation Rates

Provinces-year observations with HDI above 0.57

	Average Annual Change	Regression Coefficient		Average Annual Change in Litigation Rates	
		FE	FE & AR1	FE	FE & AR1
GRP per Capita (10,000s)	0.16	1.66	2.52	**0.266**	**0.403**
Privatization Rate	0.023	10.03	8.52	**0.231**	**0.196**
Lawyer Density	0.070	2.95	1.97	**0.207**	**0.138**
High School Graduation Rate	1.58	0.087	0.11	**0.137**	**0.174**
Tourism	13.62	0.0027	0.0027	**0.037**	**0.037**

The table shows average annual changes in litigation rates associated with statistically significant regressors in regressions of provinces-year observations with HDI above 0.57 (column one). The average annual change in litigation rates associated with a particular variable is calculated as the product of that variable's average annual change (column two) and its regression coefficient (column three). Average annual changes are found by regressing each variable on a linear time trend. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Privatization rate refers to the proportion of assets owned by the private sector. Lawyer density refers to the number of full-time lawyers per 10,000 people. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Tourism refers to the number of international tourist visits per 10,000 people.

Forecasting Litigation Rates in China

Abstract

We employ traditional techniques to forecast province-level litigation rates in China and then investigate whether forecast accuracy can be improved by incorporating variables shown to be correlated with litigation rates. We find that small improvements are possible but that these improvements come at the cost of model simplicity and ease of implementation. Additionally, we find that predictions are improved by first forecasting factors that are used to predict litigation rates rather than relying on historical data directly. Our work makes two contributions. First, it provides insight into the future of litigation rates in China. Second, it implements one method of forecasting with variables whose future values are presently unknown and compares the performance, parsimony, and implementation of that technique with traditional forecasting methods. In this particular situation, the predictions are improved.

I. Introduction

The use of historical information to anticipate future outcomes (forecasting) has developed considerably over the years. Methods have become increasingly refined and best practices have emerged. The application of forecasting methods to specific circumstances often brings with it unique questions and challenges. Occasionally, these situations present opportunities to expand upon established methods. Such is the case for our particular problem – the future of litigation rates in China.

Litigation rates are a common measure of the frequency of lawsuits in a jurisdiction and are defined as the number of lawsuits per 10,000 people. Anticipating future demands on a jurisdiction's legal system is relevant to the development of those systems. It offers policymakers the opportunity to shape incentives to litigate through legal rules and to ensure that the supply of legal services can meet demand. Insight into future litigation rates is also relevant to entities that may be susceptible to legal liability as a result of their activities, associations, and financial positions. By anticipating costs associated with litigation, insurers may be able to assist in managing these costs for exposed entities.

In this paper, we employ traditional techniques to forecast province-level litigation rates in China. We then investigate whether forecast accuracy can be improved by incorporating variables shown to be correlated with litigation rates. Future values for variables related to litigation rates are presently unknown. As such, we must first develop forecasts for these variables and then use those forecasts to predict future litigation rates. This technique is suggested by Frees and Miller (2004). To our knowledge, we are the first to implement this method in a panel data context.

The paper is organized as follows. In section two, we discuss the development of litigation rates in China and employ methods of forecasting future litigation rates. We compare the accuracy of these methods using out-of-sample prediction, a common practice in which a forecaster uses data from a portion of their survey timespan to produce forecasts for the remainder of the timespan. In section three, we consider whether forecast accuracy can be improved by incorporating variables shown to be correlated with litigation rates whose future values are presently unknown. In section four, we conclude with a discussion of our findings and thoughts about future efforts.

II. Model Selection

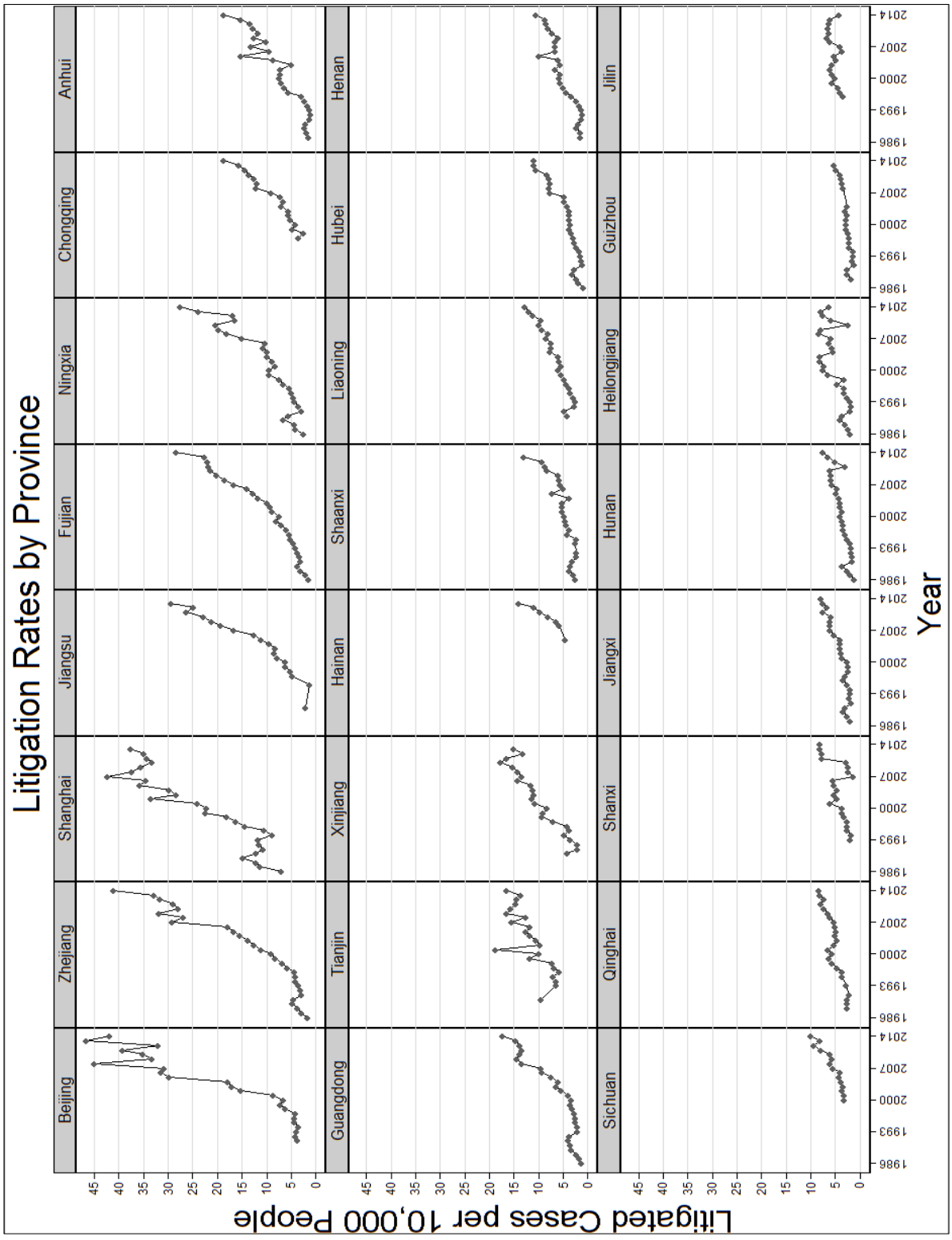
To forecast litigation rates, we employ 29 years of province-level data from China (1986-2014) across 24 provinces. These data are taken from the China Statistical Yearbook Database (CSYD), which includes provincial statistical yearbooks published annually by the National Bureau of Statistics of China. Data include litigated case counts as well as a host of economic, social, and legal information. Litigation data are derived from law firms and thus represent only those litigated cases which make use of a lawyer.

Figure 1 shows litigation rates across time by province. Provinces are organized in decreasing order of litigation rate in the most recent year of reporting. In looking at this graph, it is immediately apparent that time series for Chinese province-level litigation rates are not stationary; rather, most of the time series display a noticeable increasing trend.[16] Stationarity is an important assumption underlying certain forecasting techniques. In order to utilize these techniques, the litigation rate time series must be transformed into stationary patterns.

One common technique used to create stationary time series is first-differencing. First-differencing is a variable transformation in which adjacent values from the original series are subtracted from one another to produce the transformed series. To determine whether the transformed series is stationary, we can use the Dickey-Fuller test for stationarity. Using this test, we find that first-differencing does indeed make the litigation rate time series one that is stationary. A time series plot of first-differenced litigation rates can be found in the appendix.

[16] A stationarity time series is one whose joint probability distribution is constant over time. As a result, statistical properties such as mean and variance also remain constant over time.

Figure 1



The graph shows litigation rates across 29 years (1986-2014) for 24 of China's 31 provinces, municipalities, and autonomous regions (collectively referred to as provinces in this paper). Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. We do not possess litigation rate data for seven provinces excluded from this table. Similarly, missing data points indicate that we do not possess litigation rate data for a given province-year. Panels are organized by that province's litigation rate in the most recent year of reporting, with the higher litigation rate provinces appearing before lower ones.

Having generated a stationary version of our dependent variable, our next step is to select an appropriate forecasting model. Before discussing specific models, it is valuable to understand the spectrum of options available to anyone attempting to forecast panel data. By its very nature, panel data are a collection of time series – one for each subject in the analysis. When anticipating future values for a particular subject, a forecaster may choose to rely solely on historical data from that subject, to rely on historical data from the group (all subjects), or to rely on some combination of the two. We investigate models of each type and compare their forecasting accuracy.

The first model that we consider involves the estimation of separate time series for each province. Specifically, we regress first-differenced litigation rates on a single lag of first differenced litigation rates for each province and then use parameter estimates to generate forecasts.[17] The two stages in this process are depicted in the two lines of Equation 1. This approach rests on an assumption of stationarity and therefore, we employ the first differenced version of the litigation rate variable.

Equation 1: Forecasting with Separate Time Series

$$\text{Model 1:}\quad \Delta Lit_t = \alpha_i + \beta_i \, \Delta Lit_{t-1} + \varepsilon$$

$$\Delta \widehat{Lit}_{T+f} = \hat{\alpha}_i + \hat{\beta}_i \, \Delta Lit_{T+f-1}$$

The first line of Equation 1 establishes a relation between first-differenced litigation rates, ΔLit_t, a single lag of first-difference litigation rates, ΔLit_{t-1}, a constant term, α_i, and an error term, ε. The subscript i on α and β indicates that these parameters vary by province. This variation arises from the fact that separate regressions are estimated for each province. On one hand, allowing model parameters to vary provides flexibility. On the other hand, estimating separate regressions means that litigation rate forecasts for a particular province rely exclusively upon that provinces litigation rate history. Potentially useful information contained in the litigation experience of other provinces is not considered.[18]

[17] Error terms are assumed to equal zero in the forecast window. We investigate the possibility of including zero or two lags as well as the possibility of including a temporal process in the error term but find that these changes do not improve forecast accuracy.

[18] A vector autoregressive approach which incorporates historical litigation rates from other provinces is infeasible given missing values in our dataset.

The parameters α_i and β_i can be estimated within a regression framework and then used to forecast litigation rates. The second line of Equation 1 demonstrates how to generate forecasts using these estimates. Here, $\Delta\widehat{Lit}_{T+f}$ represents the first-differenced litigation rate forecast at time T+f, where T is the number of periods for which litigation rates are observed and f is the forecast period. It should be noted that at time T+1, forecasts are generated using litigation rates in the final year of our sample, while in years beyond T+1, forecasts are generated using the forecast from the previous period. As a result, forecasts must be generated sequentially and become less reliable as they are further removed from observed values.

The second model that we consider is equivalent to Model 1, except that observations are pooled across provinces. Pooling addresses a key shortcoming of our first model – an inability to utilize group litigation experience in generating forecasts. In this case, a single regression is estimated using observations from all provinces, as shown in Equation 2.

Equation 2: Forecasting with Pooled Data

$$\text{Model 2:} \quad \Delta Lit_t = \alpha + \beta\, \Delta Lit_{t-1} + \varepsilon$$

$$\Delta\widehat{Lit}_{T+f} = \hat{\alpha} + \hat{\beta}\, \Delta Lit_{T+f-1}$$

The empirical difference between Equation 1 and Equation 2 is that in Equation 2, neither the intercept nor the coefficient on lagged first-differenced litigation rates vary across provinces. Here, only the litigation experience of the group is relevant to the estimation of model parameters. Recall that Model 1 uses prior information in a very different way, relying only upon the litigation experience of a particular province.

Thus far, we have considered two models: one which relies entirely on subject-specific information and another which relies entirely on group information. We now consider a combined approach, in which a fixed or random effects framework is used to estimate model parameters. This combined approach is common in the panel data forecasting literature and has been used to forecast a vast array of metrics, including carbon dioxide emissions (Schmalensee et. al., 1998), company investment (Hsiao and Tahmiscioglu, 1997), cigarette consumption (Baltagi et. al., 2000), demand for electricity and natural gas (Maddala et. al.; 1997), exchange rates (Rapach and Wohar, 2004), gasoline prices (Baltagi et. al.; 2003), and lottery ticket sales (Frees and Miller, 2004), among others.

Equation 3 shows several different forecasting models that utilize a fixed or random effects framework. Each of these models includes province-specific intercepts and a single lag of the dependent variable, the coefficient on which is estimated across all provinces. Models 3 and 4 continue to employ the first-differenced litigation rate variable and thus, directly combine elements of modes 1 and 2.

Models 5-8 employ the original litigation rate variable, taking advantage of the fact that time series need not be stationary when observations are pooled across provinces. Recall that before taking first-differences, litigation rates displayed noticeable temporal patterns which appeared to vary across provinces. When forecasting with the original litigation rate variable, it may be useful to account for these temporal patterns in some other way. We investigate this possibility by estimating two models, models 7 and 8, which include linear time trends for each province in our study.

Equation 3: Forecasting with Fixed and Random Effects

Model 3: $\Delta Lit_t = \alpha_{i,FE} + \beta\ \Delta Lit_{t-1} + \varepsilon$

$$\Delta\widehat{Lit}_{T+f} = \hat{\alpha}_{i,FE} + \hat{\beta}\ \Delta Lit_{T+f-1}$$

Model 4: $\Delta Lit_t = \alpha_{i,FE} + \beta\ \Delta Lit_{t-1} + \varepsilon$

$$\Delta\widehat{Lit}_{T+f} = \hat{\alpha}_{i,RE} + \hat{\beta}\ \Delta Lit_{T+f-1}$$

Model 5: $Lit_t = \alpha_{i,FE} + \beta\ Lit_{t-1} + \varepsilon$

$$\widehat{Lit}_{T+f} = \hat{\alpha}_{i,FE} + \hat{\beta}\ Lit_{T+f-1}$$

Model 6: $Lit_t = \alpha_{i,RE} + \beta\ Lit_{t-1} + \varepsilon$

$$\widehat{Lit}_{T+f} = \hat{\alpha}_{i,RE} + \hat{\beta}\ Lit_{T+f-1}$$

Model 7: $Lit = \alpha_{i,FE} + \beta\ \Delta Lit_{t-1} + \gamma_i\ Year + \varepsilon$

$$\widehat{Lit}_{T+f} = \hat{\alpha}_{i,FE} + \hat{\beta}\ Lit_{T+f-1} + \hat{\gamma}_i\ Year$$

Model 8: $Lit = \alpha_{i,RE} + \beta\ \Delta Lit_{t-1} + \gamma_i\ Year + \varepsilon$

$$\widehat{Lit}_{T+f} = \hat{\alpha}_{i,RE} + \hat{\beta}\ Lit_{T+f-1} + \hat{\gamma}_i\ Year$$

To select among the eight potential forecasting models, we compare the accuracy of forecasts produced under each model. Evaluating forecast accuracy is a challenge faced by all forecasters. The only way to truly measure forecast accuracy is to wait until future values can be observed and then to compare these values with earlier predictions. Luckily, a second best approach is available, known as "out-of-sample prediction." Under this approach, a forecaster uses data from a portion of their survey timespan (the estimation period) to produce forecasts for the remainder of the timespan (the validation period). Accuracy metrics can then be constructed by comparing forecasts with actual observations.

We evaluate the accuracy of forecasts produced under each model using out-of-sample prediction. In doing so, we partition our dataset into two pieces: an estimation period ranging from 1986

to 2008 from which forecasts will be constructed and a validation period ranging from 2009 to 2014 for which predictions will be made and accuracy metrics determined. The six-year length of the validation period represents approximately 20 percent of our total survey timespan, a proportion comparable to that of other studies. In a subsequent analysis, we shorten the validation period to two years, 2009-2010, to investigate whether certain models are more accurate under a shorter forecast horizon. Two provinces, Guizhou and Hainan, are omitted from all out-of-sample analyses because data limitations for these provinces make prediction infeasible.

We construct four metrics to evaluate out-of-sample prediction accuracy: (1) mean absolute error, (2) mean squared error, (3) sum of squared error, and (4) mean absolute percentage error. All four of these metrics compare out-of-sample predictions with actual observations. Definitions for each metric are shown in Equation 4. In these definitions, $Lit_{i,2008+f}$ refers to the actual litigation rate in province i in year 2008+f and $\widehat{Lit}_{i,2008+f}$ refers to the litigation rate prediction for that province and year. f refers to the forecast year and spans from 1 to 6 in the 2009-2014 validation period and from 1 to 2 in the 2009-2010 validation period.

Equation 4: Definitions of Out-of-Sample Metrics

Mean absolute error: $\frac{1}{22\times F}\sum_{i=1}^{22}\sum_{f=1}^{F}\left|\widehat{Lit}_{i,2008+f} - Lit_{i,2008+f}\right|$

Mean squared error: $\frac{1}{22\times F}\sum_{i=1}^{22}\sum_{f=1}^{F}(\widehat{Lit}_{i,2008+f} - Lit_{i,2008+f})^2$

Sum of squared error: $\sum_{i=1}^{22}\sum_{f=1}^{F}(\widehat{Lit}_{i,2008+f} - Lit_{i,2008+f})^2$

Mean absolute percentage error: $\frac{100}{22\times F}\sum_{i=1}^{22}\sum_{f=1}^{F}\left|\frac{\widehat{Lit}_{i,2008+f} - Lit_{i,2008+f}}{Lit_{i,2008+f}}\right|$

Table 1 summarizes the eight forecasting models and shows out-of-sample metrics for each. Note that 385 observations are used to develop forecasting models employing first-differenced litigation rates, while 409 observations are used to develop forecasting models employing the original litigation rate variable. This difference stems from the fact that the earliest year of observations cannot be used in the first-differenced model given that prior year's data are missing.

We see that in both the six-year and two-year validation periods, Model 2 displays the smallest forecasting error according to three of our four accuracy metrics. This model generates forecasts using a

common intercept and coefficient on lagged first-differenced litigation rates across provinces. An important implication of a common trajectory is that it represents a "leveling-off" of litigation rates for provinces which have had historically high growth rates and a "swelling" of litigation rates for provinces which have had historically low growth rates. On average, predictions under this model differ from actual litigation rates by about 2.1 cases per 10,000 people. For context, the average litigation rate across provinces during this time is about 14.9 case per 10,000 people.

Figure 2 shows out-of-sample predication results for Model 2 alongside actual litigation rate observations. Vertical lines between years 2008 and 2009 mark an end to the estimation period and the beginning of the validation period. We see that predictions appear accurate for provinces with moderate litigation rates but are especially poor for provinces with very high and very low litigation rates.[19] Despite recognizing where predictions perform poorly, it remains unclear how these estimates can be improved. For example, in high litigation provinces, predictions are not systematically higher or lower than actual litigation rate observations. As such, estimating a separate model for high litigation provinces in unlikely to improve outcomes. One possibility is that our model fails to account for information beyond historical litigation rates, which might provide insight into litigation rates changes. We investigate this possibility in the next section by incorporating variables shown to be correlated with litigation rates.

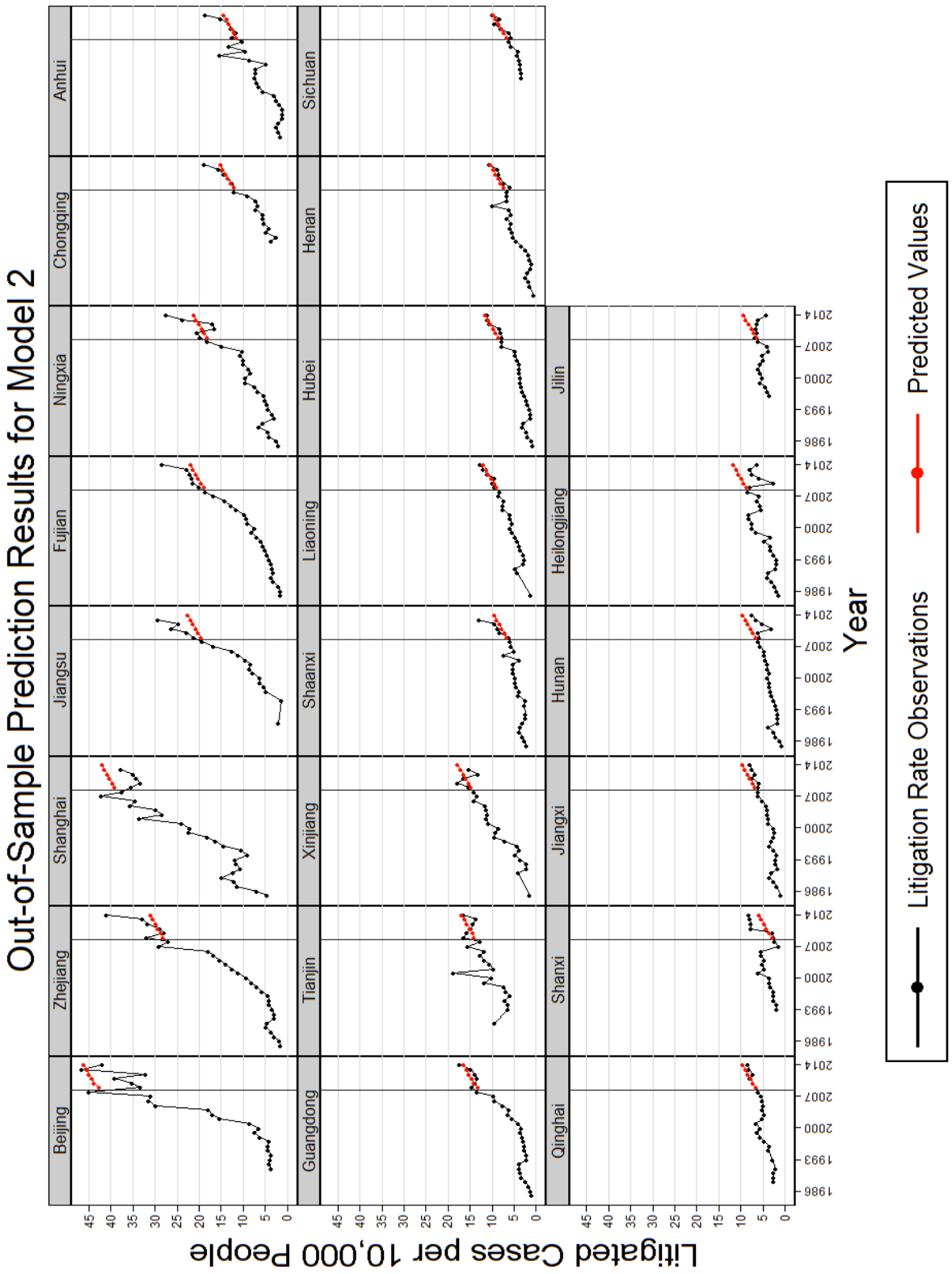
[19] This result stems from at least three things: (1) the particular model employed, which will generally (but not always) result more accurate predictions for provinces with litigation rates closer to the average, (2) the fact that moderate litigation provinces appear to follow more linear litigation rate trajectories when compared with very high and very low litigation rate provinces, and (3) the high degree of litigation rate variation across provinces, which will generally (but not always) result in less accurate prediction for very high and very low litigation rate provinces.

Table 1: Model Summaries and Out-of-Sample Metrics

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Forecast unit	Δ Lit.	Δ Lit.	Δ Lit.	Δ Lit.	Lit.	Lit.	Lit.	Lit.
Intercept	Separate	Pooled	Fixed E.	Rand. E.	Fixed E.	Rand. E.	Fixed E.	Rand. E.
Coefficient on lagged forecast unit	Separate	Pooled	Pooled	Pooled	Pooled	Pooled	Pooled	Pooled
Province-specific time trends	No	No	No	No	No	No	Yes	Yes
Observations in estimation period	385	385	385	385	409	409	409	409
Validation period: 2009-2014								
Mean absolute prediction error	2.43	**2.12**	2.37	2.38	2.71	3.27	2.52	2.32
Mean squared prediction error	17.56	**9.56**	15.08	15.43	24.94	40.94	17.50	13.52
Sum of squared prediction error	2247	**1224**	1930	1975	3192	5240	2240	1730
Mean absolute percent error	18.09%	18.53%	**17.66%**	17.71%	19.57%	22.39%	17.76%	18.53%
Validation period: 2009-2010								
Mean absolute prediction error	1.79	**1.71**	1.76	1.77	1.93	2.10	1.88	1.76
Mean squared prediction error	11.8	**7.41**	9.87	10.17	15.29	19.14	11.58	10.32
Sum of squared prediction error	519	**326**	434	448	673	842	509	454
Mean absolute percent error	**14.41%**	16.80%	15.07%	15.04%	16.11%	17.31%	15.48%	16.21%

The table summarizes eight litigation rate forecasting models and shows out-of-sample metrics for each. Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. "Estimation period" refers to years 1986-2008. Province-level litigation rate data from these years are used to generate forecasts for the period 2009-2014. "Validation period" refers to years 2009-2014. Data from these years are used to assess the accuracy of these forecasts. Accuracy is judged based on four metrics: mean absolute prediction error, mean squared prediction error, sum of squared prediction error, and mean absolute percent error. Definitions for these metrics are shown in Equation 4. Bolded numbers indicate the minimum value for a particular metric.

Figure 2



The graph shows actual litigation rates and out-of-sample forecasts for 22 of China's 31 provinces, municipalities, and autonomous regions (collectively referred to as provinces in this paper). Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people. We do not possess litigation rate data sufficient for out-of-sample tests for the nine provinces excluded from this table. Vertical lines between years 2008 and 2009 indicate the beginning of out-of-sample predictions. Panels are organized by that province's litigation rate in the most recent year of reporting, with the higher litigation rate provinces appearing before lower ones.

III. Inclusion of Relevant Variables

The forecasting models developed thus far rely only upon historical litigation rate data. We now investigate whether forecasts can be improved by incorporating variables shown to be correlated with litigation rates. Bujakowski and Schmit (2017) find that a number of economic, social, and legal variables are relevant to the development of litigation rates, including gross regional product, education, Internet use, tourism, access to lawyers, and privatization. Their analyses also include two other variables, urbanization rate and mediation rate, which are not found to have a statistically significant relationship with litigation rates.

Table 2 shows regression results from Bujakowski and Schmit (2017) as well as results of an original regression titled "First-Differenced," the form of which mirrors that of Model 2. Recall that our forecasting model uses first-differencing to make time series stationary and a single lag of the litigation rate variable to forecast. Analogously, the first-differenced regression employs first-differenced variables and one lag of the dependent variable. Results from the first-differenced regression are similar to those of the original two models. All variables that were statistically significant in the original models remain so in the first-differenced model with the exception of Internet Penetration Rate. Coefficients on these variables also appear similar, with the largest changes occurring for Internet Penetration Rate and GRP per Capita.

All three regressions in Table 2 use observations from the entire dataset (1986-2014). Regressions using only observations from the estimation period (1986-2008) can be found in the appendix. Results of these regressions indicate that GRP per capita, Internet penetration, lawyer density, and privatization rate continue to display positive and statistically significant associations with litigation rates. This is not the case for high school graduation rate and tourism. The coefficient on high school graduation rate becomes negative when only observations from the estimation period are considered, while the coefficient on tourism loses statistical significance in all but the first-differenced model.

Table 2: Regression Results

Dependent Variable = Litigation Rate	*All Provinces*		
	Fixed Effects	*FE & AR1*	*First-Differenced*
Litigation Rate$_{t-1}$			-0.38***
			0.049
GRP per Capita (10,000s)	1.09**	1.63**	2.63***
	(0.52)	(0.76)	(0.74)
Urbanization Rate	1.21	0.48	-1.80
	(3.07)	(3.71)	(8.48)
High School Graduation Rate	0.065***	0.058**	0.069**
	(0.021)	(0.027)	(0.028)
Internet Penetration Rate	15.18***	9.04*	-0.38
	(4.57)	(5.37)	(4.19)
Tourism	0.0019**	0.0018*	0.0021**
	(0.00090)	(0.0010)	(0.00099)
Lawyer Density	3.15***	2.37***	1.82***
	(0.40)	(0.56)	(0.635)
Mediation Rate	-0.013	-0.0028	-0.00035
	(0.010)	(0.011)	(0.010)
Privatization Rate	5.47**	5.56**	7.23***
	(2.19)	(2.60)	(2.33)
Province Specific Time Trends	Yes	Yes	No
Observations	441	417	391
R^2	0.65	0.64	0.21

The table shows estimates from three regressions of litigation rates on the other variables in column one. The first two regressions include province and year fixed effects and province-specific time trends. In the regression titled "fixed effects" (columns two), error terms are assumed to be independent and identically distributed, whereas in the regression titled "FE & AR1" (column three), error terms are assumed to follow an AR1 process. The regression titled "First-Differenced" uses first-differenced versions of all variables. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Given support for the notion that economic, social, and legal variables (which we refer to as "factors") are associated with litigation rates, it is natural to inquire whether these variables are also useful predictors of litigation rates. Forecasting with factors is particularly challenging because future values for these variables are presently unknown.[20] We can address this issue by first forecasting factors and then use those values to produce litigation rate forecasts. In this case, two sources of uncertainty exist within the model: (1) uncertainty associated with factor forecasts and (2) uncertainty associated with litigation rate forecasts. The extra uncertainty associated with point number one should be considered when constructing forecast confidence intervals.

This technique is suggested by Frees and Miller (2004) in their study of Wisconsin state lottery sales. Frees and Miller find that lottery sales increase dramatically when the jackpot prize for one online game, PowerBall, grows to an amount in excess of $100 million. They indicate that information about the size of future PowerBall jackpots likely would be useful in forecasting lottery sales, yet acknowledge that this information cannot be known in advance. They suggest the development of a separate forecasting model for PowerBall jackpots, the results of which could be used to forecast lottery sales. We expand upon the work of Frees and Miller by implementing their suggested technique in the context of Chinese provincial litigation rates.

We begin by forecasting factors associated with litigation rates. Using the Dickey-Fuller test for stationarity, we find that first-differencing is sufficient to make each factor stationary. As a result, the same method used to forecast litigation rates in the previous section can be used to forecast factors. This method is outlined in Equation 2. First, a regression framework is used to estimate a common intercept and coefficient for each variable. Second, parameter estimates are used alongside historical values to develop forecasts.

Once factor forecasts have been generated, these values can be incorporated into our forecasting model for litigation rates. Equation 4 shows how to incorporate these values. The first line of Equation 4 is equivalent to the first line of Equation 2, except that the term $\theta\ \Delta Factors_t$ has been added to the right side. $\Delta Factors_t$ is a vector which includes the first-difference of each factor at time t. Note that when $t < T$, $\Delta Factors_t$ is known, whereas when $T > t$, $\Delta Factors_t$ is predicted. Thus, in the second line of Equation 4, predicted values of factors ($\Delta\widehat{Factors}_{T+f}$) are used.

[20] In most forecasting applications, forecasters make use of variables with known future values, such as location, season of the year, unit of analysis, etc.

Equation 4: Forecasting with Relevant Factors

$$\Delta Lit_t = \alpha + \beta\ \Delta Lit_{t-1} + \theta\ \Delta Factors_t + \varepsilon$$

$$\Delta\widehat{Lit}_{T+f} = \hat{\alpha} + \hat{\beta}\ \Delta Lit_{T+f-1} + \hat{\theta}\ \Delta\widehat{Factors}_{T+f}$$

Column three of Table 3 shows out-of-sample prediction results after factors have been added to the model. We see that in both the two-year and six-year validation periods, adding factors makes predictions less accurate. We investigate three possible explanations for this result. First, adding factors reduces the number of observations used to construct the forecasting model from 385 to 257, a reduction of roughly 33 percent. This decrease in sample size is due to missing values of factors in the estimation period. Given that information is lost when adding factors, litigation rate forecasts might become less accurate. Second, factor forecasts might be inaccurate. Given that these values are used to forecast litigation rates, inaccurate factor forecasts might yield inaccurate litigation rate forecasts. Third, adding eight factors to the forecast model might lead to overfitting. An overfitted model is unlikely to produce accurate out-of-sample predictions.

We investigate each of these potential explanations. To determine whether a reduction in sample size is the primary reason for a decrease in forecast accuracy, we consider two additional models. The first model makes use of all factors except urbanization rate. Out-of-sample metrics for this model are listed in column four of Table 3. Given that many missing values exist for urbanization rate, dropping it from the model allows us to utilize a greater portion of the dataset. We see that when urbanization rate is excluded, observations used to generate litigation rate forecasts increase from 257 to 349. Despite the increase in sample size, out-of-sample metrics remain worse than those associated with the model without factors.

The second model that we consider is identical to the without factors model, except that observations are restricted to those in which all factors are observed. Out-of-sample metrics for this model are listed in column five of Table 3. By restricting observations in this way, models with and without factors are forced to rely on the same set of observations. Unsurprisingly, forecasts generally become less accurate when less information is used to develop the forecasting model. Even so, the model without factors and restricted observations still outperforms the model with factors.

Table 3: Out-of-Sample Metrics with Multiple Factors

	Without Factors	With Factors	With Factors (No Urban. Rate)	Without Factors (Restricted)	With Factors (Actual Values)
Observations in estimation period	385	257	349	257	257
Validation period: 2009-2014					
Mean absolute prediction error	**2.12**	3.06	2.93	2.16	3.64
Mean squared prediction error	**9.56**	17.36	15.71	9.98	25.24
Sum of squared prediction error	**1224**	2222	2010	1277	3104
Mean absolute percent error	**18.53%**	29.57%	27.90%	19.63%	35.56%
Validation period: 2009-2010					
Mean absolute prediction error	**1.71**	2.07	1.99	1.74	2.08
Mean squared prediction error	**7.41**	10.63	9.70	7.82	10.99
Sum of squared prediction error	**326**	467	427	344	484
Mean absolute percent error	**16.80%**	21.98%	21.26%	17.48%	21.82%

The table summarizes five litigation rate forecasting models and shows out-of-sample metrics for each. Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. The model "Without Factors" (column 2) is equivalent to Model 2 in Table 2. The model "With Factors" (column 3) includes eight factors: GRP per capita, urbanization rate, high school graduation rate, Internet penetration rate, tourism, lawyer density, mediation rate, and privatization rate. The model "With Factors (No Urban. Rate)" (column 4) includes all of these factors except for urbanization rate. The model "Without Factors (Restricted)" (column 5) is equivalent to the model "Without Factors" except that litigation rate observations are used to develop forecasts only when all values of factors are observed. The model "With Factors (Actual Values)" (column 6) is equivalent to the model "With Factors" except that actual values of factors, instead of forecasts, are used to forecast litigation rates. "Estimation period" refers to years 1986-2008. Province-level litigation rate data from these years are used to generate forecasts for the period 2009-2014. "Validation period" refers to years 2009-2014. Data from these years are used to assess the accuracy of these forecasts. Accuracy is judged based on four metrics: mean absolute prediction error, mean squared prediction error, sum of squared prediction error, and mean absolute percent error. Definitions for these metrics are shown in Equation 4. Bolded numbers indicate the minimum value for a particular metric.

The model with factors except urbanization rate and the model without factors and restricted observations allow us to evaluate the inclusion of factors while maintaining the underlying sample. Out-of-sample metrics for these models indicate that adding factors makes forecasts less accurate, even when the sample used to construct the forecasting model is held nearly constant. As such, we investigate other possible explanations for the decrease in forecast accuracy.

The second possible explanation that we consider regards the accuracy of factor forecasts. It may be the case that better forecasts for these variables are possible and would, in turn, improve litigation rate forecasts. We investigate this possibility by using actual values of factors in the validation period, instead of forecasts. Out-of-sample metrics for this model are listed in column six of Table 3. Surprisingly, when actual factor values are employed, forecasts appear to become less accurate. This result leads us to conclude that poorly predicted factors are not the primary reason for the decrease in accuracy when factors are included.

A third explanation that we consider involves the possibility of overfitting. An overfitted model is one that describes random error instead of an underlying relationship and often results in poor out-of-sample estimates. Overfitting typically occurs when a model contains many parameters. To investigate the possibility of overfitting, we forecast litigation rates using only one factor, instead of eight.

Table 4 shows out-of-sample metrics for models using only one of the eight factors, one at a time. We see that forecast accuracy improves when high school graduation rate, tourism, or privatization rate are added to the model without factors. Among these variables, tourism is associated with the largest improvement, yet the magnitude of this gain still appears relatively small. Other variables appear to reduce forecast accuracy when they are added to the model without factors.

Out-of-sample metrics for models that use actual factor values in the validation period, instead of forecasts, can be found in the appendix. We see that for many variables, forecasts become less accurate when actual values are used. This is not the case for tourism and high school graduation rate, the two variables associated with the greatest improvement in litigation rate forecast accuracy. As forecasts for these variables improve, so too do litigation rate forecasts.

Table 4: Out-of-Sample Metrics with a Single Factor

	Without Factors	With All Factors	GRP Only	Urban. Only	H.S. Only	Int. Only	Tour. Only	Lawyer Only	Med. Only	Pvt. Only
Obs. in estimation period	385	257	371	281	367	385	381	379	383	385
Validation period: 2009-2014										
Mean absolute prediction error	2.12	3.06	2.88	2.16	2.09	4.50	**2.07**	2.28	2.12	2.12
Mean squared prediction error	9.56	17.36	15.49	10.03	9.06	33.34	**8.68**	12.11	9.55	9.38
Sum of squared prediction error	1224	2222	1983	1283	1159	4267	**1110**	1550	1223	1200
Mean absolute percent error	18.53%	29.57%	29.98%	19.41%	18.63%	40.44%	18.40%	18.60%	18.55%	**18.09%**
Validation period: 2009-2010										
Mean absolute prediction error	1.71	2.07	1.95	1.75	1.72	2.70	**1.60**	1.83	1.71	1.70
Mean squared prediction error	7.41	10.63	8.91	8.00	6.98	14.44	**6.07**	9.94	7.41	7.06
Sum of squared prediction error	326	467	392	352	307	635	**267**	437	326	311
Mean absolute percent error	16.80%	21.98%	22.28%	17.60%	17.63%	29.77%	16.51%	16.46%	16.86%	**15.89%**

The table summarizes 10 litigation rate forecasting models and shows out-of-sample metrics for each. Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. The model "Without Factors" (column 2) is equivalent to Model 2 in Table 2. The model "With Factors" (column 3) includes eight factors: GRP per capita, urbanization rate, high school graduation rate, Internet penetration rate, tourism, lawyer density, mediation rate, and privatization rate. The other models (columns 4-11) include only one of these factors, respectively. "Estimation period" refers to years 1986-2008. Province-level litigation rate data from these years are used to generate forecasts for the period 2009-2014. "Validation period" refers to years 2009-2014. Data from these years are used to assess the accuracy of these forecasts. Accuracy is judged based on four metrics: mean absolute prediction error, mean squared prediction error, sum of squared prediction error, and mean absolute percent error. Definitions for these metrics are shown in Equation 4. Bolded numbers indicate the minimum value for a particular metric.

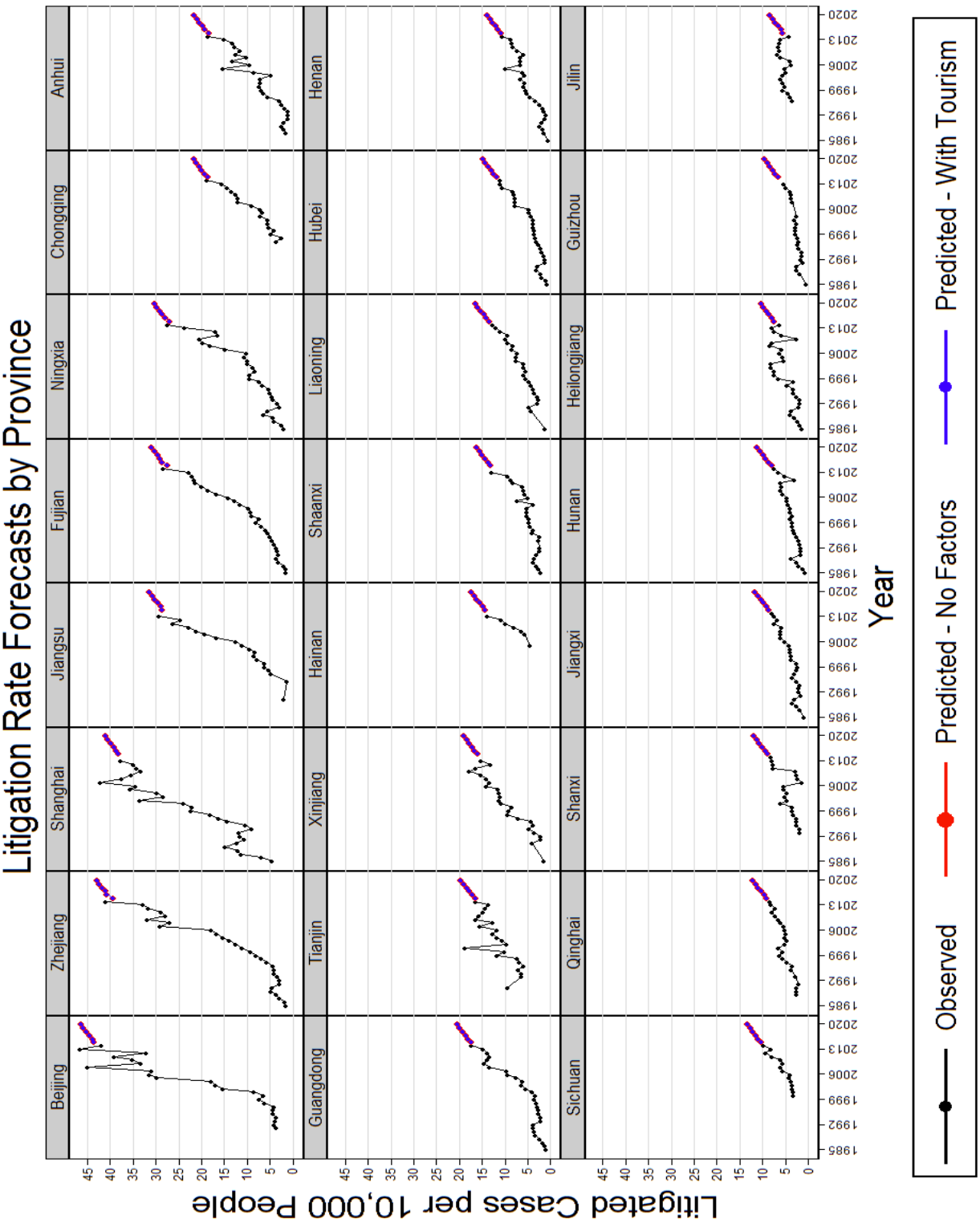
It remains unclear whether the poor forecast accuracy of the model with all eight factors is a direct consequence of overfitting. On one hand, the fact that litigation rate forecast accuracy improves when individual variables are added to the forecasting model, but not when many variables are added to the model, indicates that overfitting may be a concern. On the other hand, poor forecast accuracy might simply result from the inclusion of certain variables which are poor predictors, even when employed in isolation. Regardless of the true explanation, it is clear that among our factors, only tourism, high school graduation rate, and privatization rate appear beneficial for forecasting litigation rates.

We proceed by forecasting litigation rates beyond 2014 using two models – one without factors and one with tourism as the sole factor. The model with tourism is used because it produces the most accurate out-of-sample predictions, as determined by the metrics in Table 4. The model without factors is also employed because it is the most parsimonious of all models tested, yet produces out-of-sample predictions comparable to those of more complex models.

Figure 3 shows litigation rate forecasts generated under each model. We see that forecasts under the two models are almost identical. Indeed, the average difference between forecasts is less than 0.16 cases per 10,000 people, or approximately 0.76 percent. We also see that litigation rates are projected to increase at the same rate for all provinces. This result stems from the fact that common coefficients are estimated for all variables.

Whether or not to include Tourism or any other factor in a litigation rate forecasting model will depend on the objectives of the forecaster. We have seen that small improvements in accuracy are possible with the inclusion of specific variables; yet, these improvements come at the cost of model simplicity and ease of implementation. For example, it remains unclear how to construct forecast confidence intervals for models that include one or more factors, given that those intervals must account for uncertainty surrounding factor forecasts and litigation rate forecasts. A forecaster should balance these considerations when deciding whether to forecast with variables whose future values are presently unknown.

Figure 3



The graph shows actual litigation rates and litigation rate forecasts for 24 of China's 31 provinces, municipalities, and autonomous regions (collectively referred to as provinces in this paper). Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people. We do not possess litigation rate data for the seven provinces excluded from this table. Panels are organized by that province's litigation rate in the most recent year of reporting, with the higher litigation rate provinces appearing before lower ones.

We have seen that in our particular problem, including factors only slightly improves litigation rate forecasting accuracy. In other situations, variables with unknown future values may have greater predictive power. Here we demonstrate one method for utilizing those variables. Recall that we employ a method proposed by Frees and Miller (2004) in which forecasts for predictors are generated and used to forecast the variable of interest. One alternative to this approach is to rely on historical factor data directly. To do so, we construct a forecasting model as specified in Equation 5.

Equation 5: Forecasting with Historical Factor Data

$$\Delta Lit_t = \alpha_1 + \beta_1 \, \Delta Lit_{t-1} + \theta_1 \, \Delta Factors_{t-1} + \varepsilon \qquad \widehat{\Delta Lit}_{T+1} = \hat{\alpha}_1 + \hat{\beta}_1 \, \Delta Lit_T + \hat{\theta}_1 \, \Delta Factors_{T-1}$$

$$\Delta Lit_t = \alpha_2 + \beta_2 \, \Delta Lit_{t-2} + \theta_2 \, \Delta Factors_{t-2} + \varepsilon \qquad \widehat{\Delta Lit}_{T+2} = \hat{\alpha}_2 + \hat{\beta}_2 \, \Delta Lit_T + \hat{\theta}_2 \, \Delta Factors_{T-2}$$

… …

$$\Delta Lit_t = \alpha_f + \beta_f \, \Delta Lit_{t-f} + \theta_f \, \Delta Factors_{t-f} + \varepsilon \qquad \widehat{\Delta Lit}_{T+f} = \hat{\alpha}_f + \hat{\beta}_f \, \Delta Lit_T + \hat{\theta}_f \; \Delta Factors_{T-f}$$

Equation 5 is really a combination of several forecasting models, one for each forecast period. In the first line, a relationship is established between litigation rates and prior year predictors. That relationship is used to develop forecasts one period into the future (time T+1). In the second line, a separate relationship is established between litigation rates and predictors from two years ago. That relationship is used to develop forecasts two periods into the future (time T+2). This process is repeated until forecasts have been generated for the desired number of periods.

Table 5 shows out-of-sample metrics when litigation rate forecasts are generated from factor forecasts and from historical factor data. All accuracy metrics indicate that forecasting factors produces more accurate litigation rate predictions than relying on historical forecast data directly. This result reinforces the suggestion by Frees and Miller (2004), that separate forecasting models be developed for variables with unknown future values.

Table 5: Out-of-Sample Metrics Using Factor Forecasts and Historical Factor Data

	Factor Forecasts		Historical Factor Data	
	With All Factors	Tourism Only	With All Factors	Tourism Only
Obs. in estimation period	257	381	161 to 257	270 to 381
Validation period: 2009-2014				
Mean absolute prediction error	3.06	**2.07**	5.51	2.14
Mean squared prediction error	17.36	**8.68**	64.23	10.02
Sum of squared prediction error	2222	**1110**	8222	1282
Mean absolute percent error	29.57%	**18.40%**	49.89%	19.65%
Validation period: 2009-2010				
Mean absolute prediction error	2.07	**1.60**	2.26	1.62
Mean squared prediction error	10.63	**6.07**	11.87	6.23
Sum of squared prediction error	467	**267**	522	277
Mean absolute percent error	21.98%	**16.51%**	23.25%	16.64%

The table shows out-of-sample metrics for models which utilize "factor forecasts" and for models which rely on "historical factor data." Factor forecasts refer to the development of separate forecasting models for each factor. Forecasts from these models are then used to forecast litigation rates. Historical factor data refer to the use of factor values directly (see Equation 5). The models "With All Factors" include eight factors: GRP per capita, urbanization rate, high school graduation rate, Internet penetration rate, tourism, lawyer density, mediation rate, and privatization rate. Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. "Estimation period" refers to years 1986-2008. Province-level litigation rate data from these years are used to generate forecasts for the period 2009-2014. "Validation period" refers to years 2009-2014. Data from these years are used to assess the accuracy of these forecasts. Accuracy is judged based on four metrics: mean absolute prediction error, mean squared prediction error, sum of squared prediction error, and mean absolute percent error. Definitions for these metrics are shown in Equation 4. Bolded numbers indicate the minimum value for a particular metric.

IV. Concluding Remarks

The future of litigation in China is relevant to policymakers and to entities which may be susceptible to litigation. To shed light on that future, we consider several possible methods of forecasting province-level litigation rates. Using out-of-sample comparison techniques, we find that forecasts are most accurate when historical data from all provinces are used to develop model parameters. Under this forecasting model, litigation rates are expected to level-off in provinces that have witnessed historically high litigation growth rates and are expected to swell in provinces that have experienced historically low growth rates.

We also investigate whether forecasts can be improved by incorporating variables shown to be correlated with litigation rates whose future values are presently unknown. We find that small improvements in accuracy are possible with the inclusion of specific variables; yet, these improvements come at the cost of model simplicity and ease of implementation. Additionally, we find that forecasting factors produces more accurate litigation rate predictions than relying on historical forecast data directly. This result reinforces the suggestion by Frees and Miller (2004), that separate forecasting models be developed for variables with unknown future values.

The current analysis is limited somewhat by the availability of Chinese data. Ideally, we would incorporate litigation rate information from all of China's 31 provinces, municipalities, and autonomous regions for all 29 years of our study. Furthermore, the data would be available for the entire study period in each province. Additionally, more detailed information regarding the experience of potential plaintiffs, including costs to file a lawsuit, lawyer fees, damages awarded, and the probability of plaintiff success via settlement or trial, might be more predictive of future litigation rates than the variables which we currently consider. Gathering such data would allow for a richer understanding of the specific mechanisms which predict litigation in China.

Bibliography

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of Forecasting*, *27*(2), 153-173.

Baltagi, B. H., Bresson, G., Griffin, J. M., & Pirotte, A. (2003). Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption. *Empirical Economics*, 28(4), 795-811.

Baltagi, B. H., Griffin, J. M., & Xiong, W. (2000). To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics*, 82(1), 117-126.

Elliott, G., & Timmermann, A. (Eds.). (2013). *Handbook of economic forecasting*. Elsevier.

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

Frees, E. W., & Miller, T. W. (2004). Sales forecasting using longitudinal data models. *International Journal of Forecasting*, 20(1), 99-114.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298), 369-375.

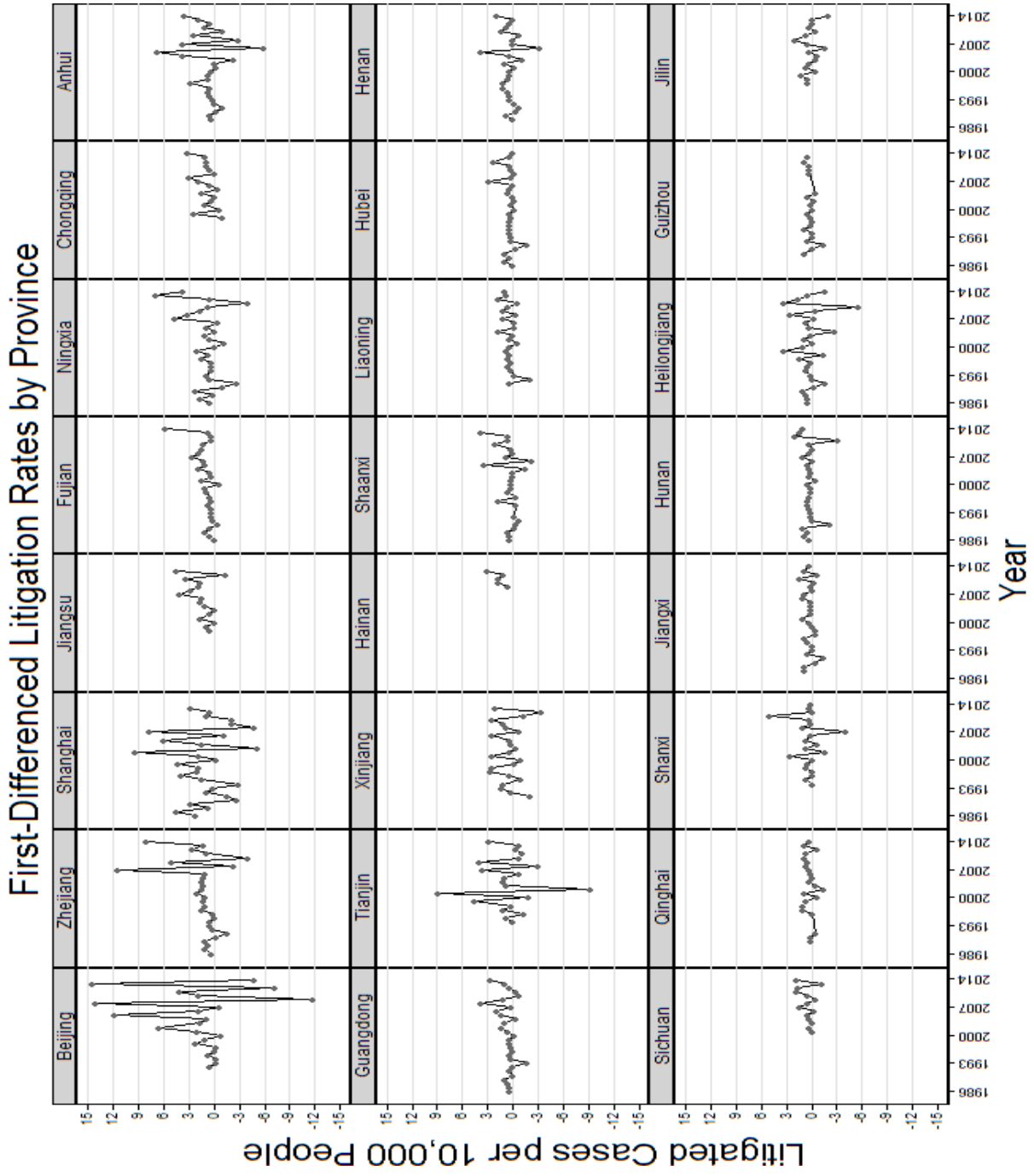
Hsiao, C., & Tahmiscioglu, A. K. (1997). A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association*, 92(438), 455-465.

Maddala, G. S., Trost, R. P., Li, H., & Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics*, 15(1), 90-100.

Rapach, D. E., & Wohar, M. E. (2004). Testing the monetary model of exchange rate determination: a closer look at panels. *Journal of international Money and Finance*, 23(6), 867-895.

Schmalensee, R., Stoker, T. M., & Judson, R. A. (1998). World carbon dioxide emissions: 1950–2050. *Review of Economics and Statistics*, 80(1), 15-27.

Appendix



The graph shows first-differenced litigation rates across 29 years (1986-2014) for 24 of China's 31 provinces, municipalities, and autonomous regions (collectively referred to as provinces in this paper). Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. We do not possess litigation rate data for seven provinces excluded from this table. Similarly, missing data points indicate that we do not possess litigation rate data for a given province-year. Panels are organized by that province's litigation rate in the most recent year of reporting, with the higher litigation rate provinces appearing before lower ones.

Regression Results – Using Only Observations from Estimation Period (1986-2008)

Dependent Variable = Litigation Rate	*All Provinces*		
	Fixed Effects	*FE & AR1*	*First-Differenced*
Litigation Rate$_{t-1}$			-0.42***
			0.068
GRP per Capita (10,000s)	2.85**	2.42	3.85***
	(1.23)	(1.59)	(1.25)
Urbanization Rate	0.59	-1.05	-1.04
	(2.70)	(3.27)	(8.30)
High School Graduation Rate	-0.056*	-0.080**	-0.012
	(0.030)	(0.038)	(0.034)
Internet Penetration Rate	19.62***	17.86***	6.06
	(5.49)	(6.65)	(5.97)
Tourism	0.00042	-0.000018	0.0030**
	(0.0011)	(0.0012)	(0.0014)
Lawyer Density	4.39***	4.36***	2.28***
	(0.50)	(0.64)	(0.72)
Mediation Rate	-0.013	-0.012	-0.0014
	(0.014)	(0.016)	(0.018)
Privatization Rate	5.96**	6.06**	7.10***
	(2.24)	(2.66)	(2.41)
Province Specific Time Trends	Yes	Yes	No
Observations	307	284	257
R^2	0.53	0.21	0.27

The table shows estimates from three regressions of litigation rates on the other variables in column one. Observations are restricted to those from years 1986-2008. The first two regressions include province and year fixed effects and province-specific time trends. In the regression titled “fixed effects” (columns two), error terms are assumed to be independent and identically distributed, whereas in the regression titled “FE & AR1” (column three), error terms are assumed to follow an AR1 process. The regression titled “First-Differenced” uses first-differenced versions of all variables. Coefficients and standard errors are reported for each variable in each regression. One, two, and three stars next to estimated variable coefficients indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Constant terms, province specific time trends, and province and year fixed effects are not reported. Litigation rate refers to the number of litigated cases per 10,000 people. GRP per capita refers to real gross regional product per capita in 10,000s. Urbanization rate refers to the proportion of the population living in urban areas. High school graduation rate refers to the number of new graduates from secondary schools (high schools) per 10,000 people. Internet penetration rate refers to the proportion of the population accessing the Internet for at least one hour per week. Tourism refers to the number of international tourist visits per 10,000 people. Lawyer density refers to the number of full-time lawyers per 10,000 people. Mediation rate refers to the number of mediation claims per 10,000 people. Privatization rate refers to the proportion of assets owned by the private sector.

Out-of-Sample Metrics with a Single Factor – Actual Values for Factors

	Without Factors	With All Factors	GRP Only	Urban. Only	H.S. Only	Int. Only	Tour. Only	Lawyer Only	Med. Only	Pvt. Only
Obs. in estimation period	385	257	371	281	367	385	381	379	383	385
Validation period: 2009-2014										
Mean absolute prediction error	2.12	3.64	3.85	2.15	2.03	3.52	2.06	2.29	2.18	2.18
Mean squared prediction error	9.56	25.24	27.88	10.00	8.79	19.13	9.02	12.56	9.98	9.86
Sum of squared prediction error	1224	3104	3569	1279	1126	2448	1155	1608	1228	1262
Mean absolute percent error	18.53%	35.56%	39.79%	19.39%	17.17%	35.42%	18.09%	18.81%	19.04%	17.95%
Validation period: 2009-2010										
Mean absolute prediction error	1.71	2.08	2.06	1.75	1.67	2.5	1.57	1.9	1.72	1.71
Mean squared prediction error	7.41	10.99	9.71	8.00	6.7	12.31	5.66	11.02	7.45	6.87
Sum of squared prediction error	326	484	427	352	295	542	249	485	328	302
Mean absolute percent error	16.80%	21.82%	23.23%	17.59%	16.81%	29.83%	15.82%	16.85%	16.67%	15.63%

The table summarizes 10 litigation rate forecasting models and shows out-of-sample metrics for each. Litigation rates are defined as the number of litigated cases in a given province-year per 10,000 people and are measured at the end of each year. The model "Without Factors" (column 2) is equivalent to Model 2 in Table 2. The model "With Factors" (column 3) includes eight factors: GRP per capita, urbanization rate, high school graduation rate, Internet penetration rate, tourism, lawyer density, mediation rate, and privatization rate. The other models (columns 4-11) include only one of these factors, respectively. Litigation rate forecasts which use factors rely on actual values of those factors, not forecasts. "Estimation period" refers to years 1986-2008. Province-level litigation rate data from these years are used to generate forecasts for the period 2009-2014. "Validation period" refers to years 2009-2014. Data from these years are used to assess the accuracy of these forecasts. Accuracy is judged based on four metrics: mean absolute prediction error, mean squared prediction error, sum of squared prediction error, and mean absolute percent error. Definitions for these metrics are shown in Equation 4. Bolded numbers indicate the minimum value for a particular metric.