NBER WORKING PAPER SERIES

MINING CHINESE HISTORICAL SOURCES AT SCALE:
A MACHINE LEARNING-APPROACH TO QING STATE CAPACITY

Wolfgang Keller
Carol H. Shiue
Sen Yan

Working Paper 32982
http://www.nber.org/papers/w32982

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2024

We thank Rui Esteves, Morgan Kelly, and Ian Miller as well as members of the audience at the CEPR 2024 Economic History Symposium in Dublin for useful comments that have improved this paper. We are also grateful to Ling Liu for her contributions to an earlier draft. Matt Butner provided valuable research assistance. We have uploaded all the Python code and data necessary to replicate the results of this paper to our GitHub repository [https://github.com/SenYan1999/qingshilu-riot-ml/tree/master]. Interested readers should refer to the README file. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Mining Chinese Historical Sources At Scale: A Machine Learning-Approach to Qing State Capacity
Wolfgang Keller, Carol H. Shiue, and Sen Yan
NBER Working Paper No. 32982
September 2024
JEL No. C8, N45

ABSTRACT

Primary historical sources are often by-passed for secondary sources due to high human costs of accessing and extracting primary information–especially in lower-resource settings. We propose a supervised machine-learning approach to the natural language processing of Chinese historical data. An application to identifying different forms of social unrest in the Veritable Records of the Qing Dynasty shows that approach cuts dramatically down the cost of using primary source data at the same time when it is free from human bias, reproducible, and flexible enough to address particular questions. External evidence on triggers of unrest also suggests that the computer-based approach is no less successful in identifying social unrest than human researchers are.

Wolfgang Keller
Department of Economics
University of Colorado Boulder
Boulder, CO 80309-0256
and NBER
Wolfgang.Keller@colorado.edu

Carol H. Shiue
Department of Economics
University of Colorado Boulder
Boulder, CO 80309
and NBER
carol.shiue@colorado.edu

Sen Yan
Leeds School of Business
University of Colorado Boulder
Boulder, CO 80309
Sen.Yan@colorado.edu

1 Introduction

Historical data is uniquely well suited for providing new insight into key questions of economic well-being, including the role of institutional quality, trust, and migrants' speed of assimilation in the labor market (Acemoglu, Johnson, and Robinson 2001, Nunn and Wantchekon 2011, and Abramitzky, Boustan, and Eriksson 2014). And yet, in many settings the barriers to working with primary data remain prohibitively high and researchers have no choice but to approach their questions through the lens of secondary sources that provide a summary of the original documents. This is a particular concern for less researched, lower-resource settings outside of Europe and North America. To make progress, this paper proposes a computer-based approach to exploiting information in Chinese primary sources from the country's Qing dynasty (1644-1911). Employing supervised machine learning to identify social unrest events in the Veritable Records of the Qing Dynasty (*Qing Shilu*), we show that the approach cuts dramatically down the cost of using primary source data. We also show how the researcher can begin directly with the primary source and tailor the data to specific questions of interest without having to use secondary compilations. Finally, by comparing the machine learning approach with existing manually compiled data sets, we demonstrate how the data we generate—as based on external evidence—is at least as successful as manually extracted information in applications.

Our primary source are the Veritable Records of the Qing Dynasty (*Qing Shilu).* Veritable Records are historical records compiled by court historians of Chinese dynasties since the 6th century—in essence, a record of what happened in the empire as the information crossed the emperor's desk. The *Qing Shilu* is highly detailed, containing a wealth of political, economical, military, and biographical information, and is generally considered to be the most comprehensive source on government documentation of social unrest during the Qing Dynasty (Hung 2013). At the same time, one well-known edition of the *Qing Shilu* encompasses more than 1,200 volumes.[1] Given its sheer size, and the fact that the content is in text, the resources required to read and extract content from the textual information is very high and sometimes infeasible. This creates a situation where primary human-based research oftentimes cover only an incomplete subset of periods within the Qing Dynasty. When there is broader coverage, there is much less depth in detail, again due to the limits of manual collection. In either case, researchers who wish to employ unrest information from the *Qing Shilu* as, for example, a control variable in a regression,

[1] See http://www.chinaknowledge.de/Literature/Historiography/qingshilu.html.

will be forced to rely on secondary compilations, even when the objectives of the secondary compilation are at best a partial proxy for what the researcher is actually trying to control for in their regression specification.

This paper shows that a computer-based AI approach does not only address the main challenge of feasibility but also several secondary issues. One is the issue of consistency. Textual analysis by two different researchers may produce different and inconsistent assessments. A team of researchers will, in practice, have difficulties applying the same exact criteria to a given research question. A computer-based classification is not affected by "long-established results" in the literature that act as focal points that might generate bias. Last not least, once a secondary compilation is published, there is very little incentive to redo the output to check for accuracy and replicability. Our computer-based classification approach is replicable, something that is difficult to achieve in the sense of record-keeping with a human-based research approach involving numerous individuals over longer periods of time, and it allows for adjustments, improvements, and future updates as needed.

We show that the computer-based approach can be tailored to particular research questions by focusing on one issue that is important not only in Chinese history, but also in many other contexts both in the past as well as today. State capacity, which may be defined as the ability of the state to reach its objectives (Acemoglu and Robinson 2019), was a key concern of the Qing. Protests in any part of the empire was considered a potential threat to the stability and the legitimacy of the state, and the bureaucratic apparatus was attuned to deal with these threats in several ways. Officials reported information on the events, and they communicating this information to superiors. We know not only when and where the unrest took place, but also how each of the reported event transpired. The individuals that were responsible for the unrest were identified by name, and the nature of the unrest was reported. Thereafter, either they continued to cause mass disturbances in the same location or different locations, or they were were captured and dealt with.

We instruct the computer first to identify reports on any unrest in the *Qing Shilu*, before we ask for a classification into three categories: peasant unrest, militia unrest, and secret societies.[2] Reports of unrest involving peasants generally encapsulate agricultural concerns such as protest of high tax, issues surrounding the landlord and the lease, burdens of disasters, poor harvest, and population pressure.

[2]A militia is a group consisting of individuals that are part-time civilian and part-time soldiers.

Militia unrest typically involved armed uprisings, fueled by local conflicts between landlords and the poor, or by feuding warlords, which usually started locally but at times also spread to a larger scale. Unrest that grew and spread was a larger concern for the ruling dynasty, no matter who was the target. However, activities of so-called secret societies such as the Red Spears or the White Lotus Sect mounted the highest concern because the Qing state itself became the target (Kuhn 1970).[3] Unrest that was fueled by anti-Manchu (anti-Qing) sentiment or activities by groups that also assumed quasi-state functions—such as, for example, secret societies that provided mutual aid and security to members, who were considered bandits, thieves, or rebels by the Qing state— was a clear and immediate threat to the state capacity of the Qing Dynasty.

Adopting a supervised machine-learning approach based on the seminal work on Qing unrests by Chen et al. (1989) as a training sample, we employ several natural language processing approaches before settling on the GUWEN-BERT classifier as the leading approach in our context. In our relatively disaggregated analysis of unrest—a given prefecture, month, and year—it achieves an accuracy in out-of-sample analysis of above 97%. The GUWEN-BERT classifier's performance sets it apart from other, more traditional methods such as Bayesian or neural networks because it is pre-trained on a large opus of historical documents in Chinese. Even in the more challenging task of classifying a given unrest into peasant, militia, or secret society, the GUWEN-BERT classifier is accurate at above 98%. With a powerful (albeit standard) computer, the classification of events into "unrest" or "not unrest" for the entire *Qing Shilu* takes about 40 minutes.

We provide two metrics of external validation for our approach. First, we exploit the relationship between unrest and economic variables that are related to food prices. Whatever the particular situation of the people in a specific region of China during the 250+ years of the Qing dynasty was, social unrest becomes more likely when grain prices are high and when weather conditions are extreme. High grain prices are indicative of poor harvests and food pressure in an economy in which the diet of the large majority of the people consists almost entirely of some form of grain. In addition to leading to poor harvests, extreme weather can mean flooding or droughts that leave the land barren, fueling more unrest. We show that our GUWEN-BERT identified events of social unrest are generally positively correlated with these known triggers of popular unrest.

[3] See section B in the Appendix for additional information.

Second, we compare the way the GUWEN-BERT-identified unrest events relate to grain price and weather shocks with how existing measures relate to these shocks. Aggregating to the national and annual level to maximize the number of existing measures in this comparison, we find that that grain price and weather shocks are more strongly correlated with the GUWEN-BERT-identified unrest event patterns compared with other measures. Furthermore, grain price and weather shocks explain seven times, or more, as much variation in the GUWEN-BERT-identified measures of unrest than in the variation of other measures. This suggests that our computer-based approach to social unrest in the Qing performs at least as well, and certainly no worse than, existing human approaches to this question.

This paper is part of the turn to digital methods in economic history (see Abramitzky 2015, Mitchener 2015). Producing compelling novel evidence that allows researchers to study fundamental questions of history turns on our ability to utilize new sources of (big) data, and to this end, the application of machine-learning can be impactful. While there is a sizable literature on the automated linking of US census records (Abramitzky, Mill, and Perez 2020, Bailey, Lin, Mohammed, Mohnen, Murray, Zhang, and Prettyman 2023), the present paper is closer in spirit to natural language processing of big data as in Card, Chang, Becker, Mendelsohn, Voigt, Boustan, Abramitzky, and Jurafsky (2022). Our focus on Chinese language complements other work for non-English textual analysis contexts that have received less attention thus far, such as Yang, Arora, Jheng, and Dell (2023). Computer-based strategies have the potential of dramatically reducing the gap of what we know about economies in higher versus lower resource settings, just as the ability of Sub-Saharan farmers to predict price fluctuations compared to their European counterparts has dramatically caught up with the arrival of the cell phone (which makes non-existing land lines a moot point). In the context of the *Qing Shilu*, our analysis complements the topical modeling methods by Miller (2013). More generally, we believe that this is one of the first attempts to employ supervised machine-learning to the *Qing Shilu*, a source from a time as early as the 17th century, over almost three centuries, to the early 20th century.

We also contribute to the analysis of state capacity, a concept of great importance not only in economics but also in other social sciences (Skocpol 1985, Besley and Persson 2011, Dincecco 2017, and Johnson and Koyama 2023). Proposing a machine-learning approach that can be flexibly tailored to specific research questions may open up new ways of measuring state capacity (Vaccaro 2023). Our analysis distinguishing different motives behind unrest also complements work highlighting alternative motives to participate in protests in modern-day China (Cantoni, Heizlsperger, Yang, Yuchtman, and Zhang 2022).

The remainder of the paper is as follows. The next section 2 introduces our approach and compares it with human classification approaches. Section 3 describes the three classifiers that we evaluate in the present context, and it describes the metrics of evaluation that are employed. New descriptive evidence on social unrest in Qing China based on our analysis is presented in section 4. Section 5 turns to the external validity of our approach, and it compares its performance using external data with existing analyses based on the *Qing Shilu*. Some conclusions are drawn in section 6, while the Appendix includes additional information as well as robustness analysis using alternative classification methods.

2 Alternative Approaches to Research based on the Qing Shilu

Research on the extent, causes, and consequences of civil unrest in late imperial China includes Chan (1983), Chen et al. (1989), Zhang et al. (2007), Hung (2011), Miller (2013), Jia (2014), and Chen (2015). Existing work on events in the *Qing Shilu* related to unrest has virtually always relied on humans reading through the source.[4] Typically, existing research on unrest based on the *Qing Shilu* would require a skilled researcher to analyze this document by explicitly going through the daily memorials, and demarcating events related to civil unrest as they are spotted. For example, Hung (2013) identifies a total of 514 popular uprising events and 450 cases of petitions to the central government against local injustices (capital appeals) for the period 1740-1839, with rebellions, class conflicts, and other conflicts excluded. Work based on different primary sources such as the Editing Committee of China's Military History has produced analogous lists based on human reading of the primary source material (Zhang 2007).

While these list of unrest-related events based on the work of highly skilled researchers are useful, they also have some limitations. For one, any research employing these lists is conditional on varying definitions of unrest that are employed by the compiler. Hung (2013), for example, counts popular uprisings but not rebellions, although the difference between the two may be small, while the Editing Committee of China's Military History will tend to emphasize events with a substantial level of military action. Given these differences in emphasis, consistent results from different measures cannot necessarily be expected in applications. Furthermore, human classifiers may be influenced by prior research in ways not known to

[4] An exception is Miller (2013) who uses an unsupervised computational algorithm to group together words to identify topics.

them, and if the classification task is performed by a larger team as would be needed for the *Qing Shilu* it raises the question of consistency across persons.

In contrast, the computer-based machine-learning approach is free from interpersonal and intertemporal consistency problems, and fatigue plays no role. Furthermore, by employing alternative methods and comparing them in external validity checks, as done below, non-robustness and major performance differences can be easily assessed. The machine-learning approach can be transparently modified to fit a particular research objective, and it is reproducible.

The only bias that is left is the intrinsic one from a particular primary source, in this case the *Qing Shilu*. Given that the *Qing Shilu* is the administrative record of the dynasty, what is considered as a "unrest" has to be viewed through the lens of the Qing rulers. For example, during the 19th century, the Qing government often relied on militiamen, typically organized by local gentry, when it would not want to involve the regular Qing army (called Banners and the Green Standard Army). Such pro-Qing militia activity would typically not be mentioned in the *Qing Shilu*. In our approach of unrest events from the Qing Shilu, therefore, we do not extract all militia activity, but only militia activity that constituted a threat to Qing rule, and similarly with other unrest events.

For our baseline approach of identifying unrest events in the *Qing Shilu*, then, we assume that the definition of what constitutes a particular form of unrest is constant over time and across prefectures. Under this assumption, our unrest figures are intertemporally and cross-regionally comparable, and we are not aware of systematic evidence that contradicts this assumption. For the external validity checks using grain price and weather shocks, we can relax this assumption because the inclusion of controls at the prefecture, year, and month level means that common national trends or persistent regional differences are eliminated. For example, one would expect that Qing rulers would be particularly sensitive to unrest in the region in which their capital (Beijing) is located, and the inclusion of prefecture fixed effects accounts for this.

We begin with a training sample on Qing unrest from the seminal work by Chen et al. (1989) that are manually identified–the 'ground truth' for the purposes of our analysis. Employing a training sample is what turns our analysis from an unsupervised into a supervised machine-learning approach. These known unrest events are further classified in Chen et al. (1989) into (i) peasant, (ii) militia, and (iii) and secret society unrest, whose classification we follow. Further discussion on the nature of specifically the secret society unrest events is given in section B of the Appendix. Thus, our approach starts out from the knowledge of skilled human researchers and leverages it by using a computational approach to protests in the *Qing Shilu*.

The *Qing Shilu* is accessed as a singular digital text document, which is broken up into separate entries. Information given in the *Qing Shilu* follows a strictly prescribed format. Figure 1 shows an example.

Figure 1: Example from *Qing Shilu, juan 212, fol. 9. Jiaqing Edition.*

‘

旗蒙古都統○兵部議覆土司田舜年疏請
以伊職令其嫡子田昞如承襲其次子田旻
如願入京備使令所給敕命免繳部俱無庸
議得旨逆賊吳三桂作亂之時田舜年一心
効力頗有勞績著依所請令伊長子田昞如
襲職原領敕命亦著免繳部其次子田旻如
著來京効力○己亥
上諭八旗都統前鋒統領護軍統領副都統叅
領佐領等曰朕爲官兵生計不時厪念前已

聖祖仁皇帝實錄　卷二百十二　九

屢施大澤今年詔款內復特沛鴻恩不惜數
百萬帑金遍行賞賜嗣後軍卒人等應人人
務立生計清償夙逋豐裕度日儻有不肖之
輩不思撙節儉約惟知縱酒酣飲鮮衣肥馬
過於費用則不數日間仍如未沛恩澤時爾
等俱有督率之責不當徒以督率爲名亦當
誘之向善使人人以孝弟爲本各知自守愛
惜產業則不特風俗可至淳樸而朕恤兵之
心亦不致徒勞矣可將此旨刋刻遍示軍卒

Notes: Source is http://www.chinaknowledge.de/Literature/Historiography/qingshilu.html

One entry is a single correspondence or memorandum related to an event. Different entries in Figure 1 are separated by circles.[5] The 775 entries of our training sample from Chen et al. (1989) are entries within the corpus that have been manually and correctly identified as being unrest-related. We combine these unrest entries with 775 half manually picked entries from the Qing Shilu that are not unrest-related (we

[5]The first entry of each month is dated, and we observe multiple entries within a single month. It is common for there to be multiple entries for a single event as the situation develops.

say half manually picked because in choosing them we have made sure than none of them is related to an unrest). Using this training data, we construct a classifier, which is a function that maps an entry from the *Qing Shilu* into a specific category. Classifiers can be based on different algorithms, and below we consider several of them. Our classifier categorizes the data as either being unrest related, or not unrest related, and we focus on entries in the *Qing Shilu* of 50 or more characters to ensure that the algorithm has sufficiently much data to robustly classify the entry. The entries of the *Qing Shilu* that most resemble unrest-related entries in the training data are classified as unrest events.

For any classifier considered in this paper, we proceed to test it. To do so, we train the classifier on 4/5 of the training data set (620 + 620 = 1,240 entries), and then use it to classify the remaining 1/5 cases of the original data set (= 155 + 155 = 310 entries). Since we know *a priori* which entries in the training data are unrest events and which are not, we can calculate false positive and false negative rates. These measures of error are the basic criteria used to evaluate alternative classifiers; results are presented below.

Given this approach, any bias in the training data is a potential concern. We address this in a number of ways. First, relying on the multi-author, multi-year research of Chen et al. (1989) we employ the definitive research on identifying unrest events in the *Qing Shilu* as training data. Second, to address issues resulting from the relatively small size of the training data set, we augment the training data by hand-reading additional entries from the *Qing Shilu* to reach 1,210 unrest-related entries. Third, to reduce the number of false positives, we consider the output of multiple classifiers. In our final data set, we only consider the entries that are identified as unrest events by multiple classifiers.

Once we have settled on a classifier with the lowest error rate, we apply this classifier to the entire *Qing Shilu*. The output is all the entries in the *Qing Shilu* which are identified as unrest events by the classifier. This is our basic machine-learning output on Qing unrests. Based on the classification of unrest in Chen et al. (1989) into (i) peasant, (ii) militia, and (iii) and secret society unrest, we go one step further and construct another classifier that allocates an identified unrest event as either peasant, militia, or secret society-related. We also attach a date and location to each identified event. The date, in terms of a month-year, comes from where the entry is listed within the chronological *Qing Shilu*. The geographic location in terms of the prefecture and province is found using the China Historical Geographic Information System (CHGIS, https://chgis.fas.harvard.edu/). Finally, we count the number of unrest-related entries for each month-prefecture and protest type to construct a panel of civil unrest during the Qing.

Overall, we find that the approach of applying natural language processing to the *Qing Shilu* is promising. Compared to existing lists of unrest based on human classification, our machine learning approach tends to identify a greater number of unrest events, and a BERT-based Classifier achieves an accuracy of above 97% in classying entries into unrest versus not unrest events.[6] The following describes several classifiers that have been considered in this research.

3 Machine-Learning Approaches and Evaluation

We consider several machine-learning approaches in Natural Language Processing (NLP) for our task of identifying unrest events in the Qing Shilu, and some details as well as benchmark results in our case for three of these methods are given below. The methods are (1) the Multinomial Naive Bayes classifier, (2) the Perceptron classifier, and a (3) BERT-based classifier. We have uploaded all the Python code and data necessary to replicate the results of this paper to our GitHub repository [https://github.com/SenYan1999/qingshilu-riot-ml/tree/master]. Interested readers should refer to the README file.

3.1 Machine Learning Approaches

3.1.1 Multinominal Naive Bayes

The Naive Bayes is a classic model for document classification based on Bayes' Theorem (e.g., Xu, Li, and Wang 2017). This model operates on the principle of identifying the class with the highest conditional probability, given the presence of specific words in the document to be classified. For a document containing n words, the formula to predict its class c is as follows:

$$\underset{c \in C}{argmax}\ Pr(c|w_0, w_1, \cdots, w_{n-1}) \tag{1}$$

where C refers to the set of all possible classes, and w_i refers to the words in the document, with $i = 0, ..., n-1$.

Suppose $Pr(w_i|c)$ and $Pr(w_j|c)$ are independent for any given word i and j. Equation (1) can be rewritten as follows based on the Bayes Theorem:

[6] BERT stands for Bidirectional Encoder Representations from Transformers.

$$\begin{aligned}
\underset{c \in C}{argmax} Pr(c|w_0, w_1, \cdots, w_{n-1}) &= \frac{\underset{c \in C}{argmax}\ Pr(c)Pr(w_0, w_1, \cdots, w_{n-1}|c)}{Pr(w_0, w_1, \cdots, w_{n-1})} \qquad (2) \\
&= \underset{c \in C}{argmax}\ (Pr(c)Pr(w_0, w_1, \cdots, w_{n-1}|c)) \\
&= \underset{c \in C}{argmax}\ (Pr(c)Pr(w_0|c)Pr(w_1|c) \cdots Pr(w_{n-1}|c)) \\
&= \underset{c \in C}{argmax}\ (log(Pr(c)) + log(Pr(w_0|c)) + log(Pr(w_1|c)) + \cdots + log(Pr(w_{n-1}|c)))
\end{aligned}$$

In equation (2), we apply the logarithmic function to convert the product of multiple terms into a sum, which speeds up computation. However, utilizing equation (2) to classify entries involves estimating the probabilities in the expression, including $Pr(c), Pr(w_i|c)\ \forall\ i \in I$, where I denotes the index set containing all word indices. The estimation process uses the following:

$$Pr(c) = \frac{\#\ of\ entries\ with\ class\ c}{\#\ of\ all\ entries}, \qquad (3)$$

and

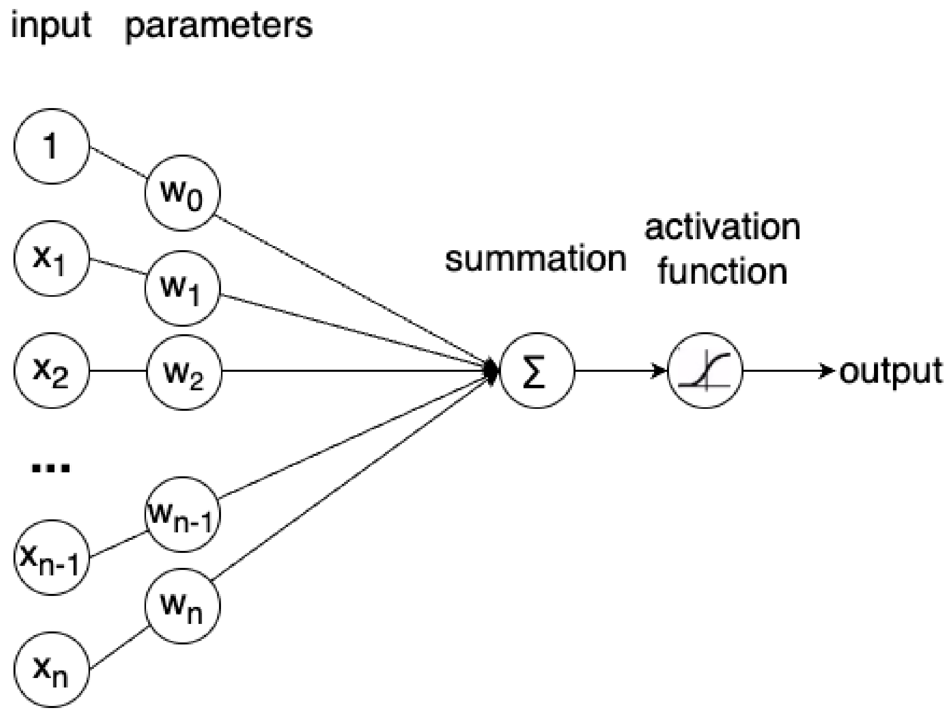
$$Pr(w_i|c) = \frac{1 + \#\ of\ times\ w_i\ appearing\ in\ all\ entries\ with\ class\ c}{\#\ of\ entries\ with\ class\ c}. \qquad (4)$$

In estimating equation (4), we use Laplace Smoothing to avoid the problem that one word appears in the inference stage but does not appear in the training corpus.

3.1.2 Perceptron

The Perceptron is a classic machine learning algorithm, see Rosenblatt (1958), and its architecture is illustrated in Figure 2.

Figure 2: Overview of the Perceptron Classifier



Notes: Synthesis of Rosenblatt (1958).

Specifically, given the input $X \in \mathbb{R}^d$ where d represents the dimension of the input vector, the formula to compute the predicted probability of the class Unrest is:

$$p = \sigma(W^T X + b) \tag{5}$$

where $W \in \mathbb{R}^d$ and b are the parameters in the Perceptron model; $\sigma(*)$ is called the sigmoid activation function in neural network analysis, and $\sigma(x) = \frac{1}{1+e^{-x}}$. The range of the output of the sigmoid activation function is $(0, 1)$ which can be use as the probability of the class Unrest directly. We adopt the gradient descent optimization method to estimate the parameters W and b during the training stage.

However, the input to the Perceptron model is a one-dimensional vector, which is inconsistent with the multi-dimensional nature of the entries to be classified. This discrepancy necessitates the encoding of the document texts into uniform vectors. Several methods exists to facilitate this encoding, and in this paper, we adopt a method known as TF-IDF (Sparck Jones 1972), where TF stands for Term Frequency and IDF stands for Inverse Document Frequency. The approach is based on word frequency.

Term Frequency indicates the number of times a word appears in a specific entry relative to its total occurrences across all entries in the training dataset. Furthermore, Inverse Document Frequency quantifies the rarity of a term, calculated as the inverse of the proportion of entries containing the term across all entries in the training dataset. To apply TF-IDF, it is necessary first to define the TF and IDF terms and then aggregate the TF-IDF values of these terms within an entry to construct its TF-IDF vector. The formula for the TF-IDF value of term i in entry j is computed as follows:

$$TF_i^j = \frac{\#\ of\ word_i\ in\ entry_j}{\#\ of\ word_i\ in\ all\ entries} \tag{6}$$

$$IDF_i = \frac{\#\ of\ all\ entries}{\#\ of\ entries\ containing\ word_i} \tag{7}$$

$$TF\text{-}IDF_i^j = TF_i^j * IDF_i \tag{8}$$

$$TF\text{-}IDF^j = [TF\text{-}IDF_0^j, TF\text{-}IDF_1^j, ..., TF\text{-}IDF_{n-1}^j]^T \tag{9}$$

where TF_i^j is the term frequency of term i in entry j, IDF_i^j is the inverse document frequency of term i, $TF\text{-}IDF_i^j$ is the TF-IDF value of term i in entry j, and $TF\text{-}IDF^j$ is the final TF-IDF vector of the entry j.

3.1.3 GUWEN-BERT Classifier

Devlin, Chang, Lee, and Toutanova (2018) propose a transformer-based model known as BERT, which has achieved state-of-the-art performance on most NLP leaderboards. Consequently, this successful model represents a significant milestone, fundamentally altering the NLP model development paradigm from training from scratch to a two-stage schema comprising the pretraining and fine-tuning stages. Specifically, this involves researchers first pretraining a large language model on a vast corpus from the Internet to learn implicit knowledge, followed by other practitioners utilizing this pretrained model and fine-tuning its parameters for specific tasks with less dependency on extensive datasets. In this paper, we adopt the same two-stage schema. Our pretrained large language model is GUWEN-BERT (see GuwenBert 2023), which is a customized version of the original BERT and has been pretrained on a broad range of historical Chinese texts encompassing various fields including history, medicine, Taoism, and others.

The GUWEN-BERT takes one entry in the *Qing Shilu* as input, and then tokenizes the sentences in the entry into historical Chinese characters. After that, a dictionary in the GUWEN-BERT can map each historical Chinese character to its embedding, a vector which was estimated during the pretraining stage. Finally, the GUWEN-BERT concatenates these embedding vectors in the order of their related Chinese characters in the sentences, generating a matrix followed by a series of matrix operation, and finally outputs a vector $H \in \mathbb{R}^h$ representing the input entry for classification, with h denoting the dimension of the output.

Equation (10) demonstrates the architecture of the GUWEN-BERT Classifier. We append a linear transformation layer to the pretrained GUWEN-BERT so that we can transform the dimension of the output from GUWEN-BERT to a vector with c dimension which can be used to do text classification. There, the value in the each dimension refers to the logits of the specific class. For example, the 2-way classifier outputs a 2-dimension vector where the two values in the vector refer to Unrest and Non-Unrest respectively. We will note the GUWEN-BERT classifier generating other numbers of classes below.

This short exposition is not the place to present the internal architecture and underlying formulas of the GUWEN-BERT as they are complex. However, given an output vector $H \in \mathbb{R}^h$ from the GUWEN-BERT, the GUWEN-BERT classifier employs the following equation:

$$L = W^T H + b \tag{10}$$

where $L \in \mathbb{R}^c$ are the logits of the c classes, $W \in \mathbb{R}^{h \times c}$ and b are the parameters of a linear transformation layer, and c denotes the number of the classes.

3.2 Evaluation

3.2.1 Evaluation Metrics

We evaluate the three classifiers on two distinct tasks: (1) the binary classification into Unrest vs. Not Unrest, and the (2) tri-category classification into (i) peasant unrest, (ii) militia unrest, and (iii) secrety society unrest, given that the event is classified as unrest. Performance metrics employed in this paper are well-known in the literature, and are called accuracy A, precision Pr, recall R, and the F_1 score. These performance metrics depend on four key variables: true negatives TN, true positive TP, false negative FN and false positive FP. True negatives refer to the number of negative events correctly classified as negative by the model. True positives denote the count of positive events accurately classified as positive. False negatives are positive events incorrectly classified as negative. Finally, false positives represent negative events mistakenly classified as positive. With these definitions in hand, the first metric, average accuracy, A, is defined as:

$$A = \frac{TP + TN}{TP + TN + FP + FN}. \tag{11}$$

Average accuracy measures the percentage of events which are correctly predicted.

The second measure is average precision, Pr, which is equal to:

$$Pr = \frac{TP}{TP + FP}. \tag{12}$$

Precision measures the percentage of actual positive events among all predicted positive events, i.e. the tendency that positively classified events are indeed positive.

Our third measure is average recall, R, which is defined as

$$R = \frac{TP}{TP + FN}. \tag{13}$$

Recall measures the percentage of predicted positive events among all actual positive events.

Finally, the last measure is the F_1, defined as

$$F_1 = 2 \times \frac{Pr \times R}{Pr + R} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}. \tag{14}$$

The F_1 score represents the harmonic mean of precision and recall, thereby providing a symmetric summary that captures encapsulates both recall and precision in a single metric.

3.2.2 Evaluation Results

Table 1 shows the results of the three models on binary classification into Unrest versus Not Unrest. Based on 4/5 of the initial training data, our tuned GUWEN-BERT Classifier performs best in this task and returns values of $A = 0.974$ on average accuracy, $Pr = 0.976$ for average precision, $R = 0.974$ for average recall, and an F_1 value of $F_1 = 0.974$.

Table 1: Unrest vs. Not Unrest Classification: Evaluation Results

	Accuracy	Precision	Recall	F_1
Multinominal Naive Bayes	0.896	0.899	0.896	0.896
TF-IDF + Perceptron	0.925	0.926	0.925	0.925
GUWEN-BERT	**0.974**	**0.976**	**0.974**	**0.974**

Notes:Authors' computations.

In the next step, we ask our algorithms to classify the identified unrest events into one of the three categories, namely peasant, militia, or secret society unrest. Employing the same classifiers as above, their performance is summarized in Table 2. One difference in the definition of the metrics in the present, tri-category classification task is that precision, recall and F_1 score have to be re-defined in the present multi-class classification task. We compute these metrics for the three classes-peasant, militia, or secret society unrest–separately, and then average them to get the so-called macro-average precision, recall and F_1.[7]

Consist with the binary classification results above, we achieve the best results with GUWEN-BERT classifier. It obtains an average accuracy of 98.7% in out-of-sample tests for this three-way classification into peasant, militia, or secret society unrest. Furthermore, the measures of Precision, Recall, and F_1 score are also above 98%.

[7] To compute the metric for one specific class, e.g. peasant unrest, we treat peasant unrest entries as positive samples and non-peasant unrest entries (secret society and militia) as negative samples.

Table 2: Protest Type Classification Evaluation Results

	Accuracy	Precision	Recall	F_1
Multinominal Naive Bayes	0.852	0.793	0.786	0.796
TF-IDF + Perceptron	0.903	0.902	0.898	0.899
GUWEN-BERT Classifier	**0.987**	**0.991**	**0.986**	**0.988**

Notes: Authors' computations.

We believe that the advantage of the GUWEN-BERT classifier over other approaches is rooted in the fact that the classifier is pre-trained on a large opus of historical Chinese documents. Thus, in the remaining analysis we show results based on applying the GUWEN-BERT classifier to the entire *Qing Shilu*.

The robustness of these findings is analyzed in the Appendix, which presents results both for employing the underlying probabilities of the tri-category classification into peasant, militia, and secret society unrest and for utilizing a four-way classifier (no unrest/peasant unrest/militia unrest/secret society unrest), see section B.

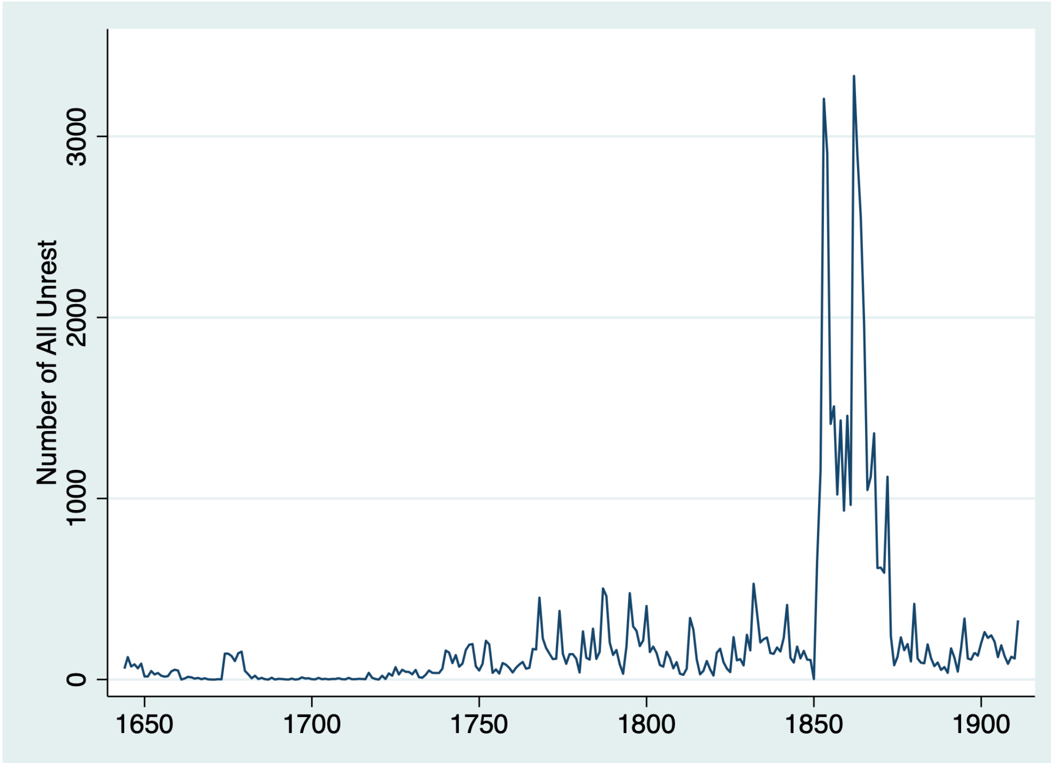
4 Social Unrest in Qing China: New Evidence from the Qing Shilu

4.1 National Patterns

In this section, we analyze the unrest distribution over time. We leverage the *Qing Shilu*'s structure to extract the date when the specific event occurred. In the *Qing Shilu*, the entries are recorded chronologically. Each new month is indicated by a date in the format "<Chinese Era Name Calendar>, <Sexagenary Cycle Year>, <Season>, <Month>": 嘉庆元年， 丙辰，春，正月. Following this pattern, a regular expression can be used to identify paragraphs by the relevant date for the entire corpus. In this paper, each unrest is assigned a date with the Chinese Era Name Calendar, e.g. Jiaqing year 1, but in addition we use a Python script to convert the Chinese Era Name Calendar into the Gregorian Calendar. If an unrest event in a given month spreads over multiple prefectures, we treat it as one unrest event for this month in each of the prefectures.

Applying this classifier to the entire *Qing Shilu* yields the pattern of Figure 3 at the national level.

Figure 3: Social Unrest in Qing China: The National Time Series



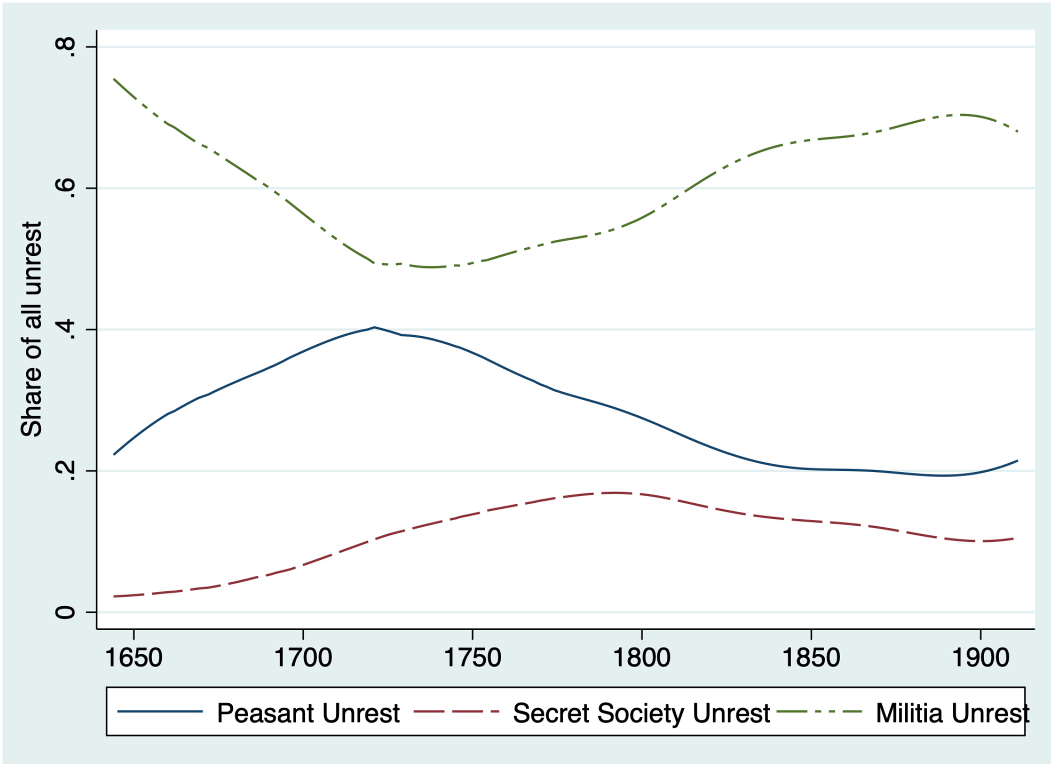
Notes: Figure shows the total number of unrest events identified by the baseline GUWEN-BERT classifier.

The annual distribution of machine-learning identified unrests during the Qing is in line with some of the broad outlines emphasized in earlier work. In particular, there is a relatively high number of events during the mid-19th century, associated with the Taiping Rebellion, as well as a trend-wise increase over the course of the Qing Dynasty. At the same time, the computer-based approach increases the count of identified unrest events based on the *Qing Shilu*, relative to human classifier-based approaches. As Figure 3 indicates, the number of unrest events in a single year can be above 3,000, which is higher than the total number identified in the work by Chen et al. (1989).[8] Overall, about 10 percent of all *Qing Shilu* entries are related to unrest.

How important are the different forms of unrest over the Qing dynasty? Figure 4 presents the composition of all unrest broken down into the three different forms, peasant, militia, and secret society unrest.

[8]Note, however, that Chen et al.'s (1989) analysis focuses on the period 1600 to 1800, that is, it does not include China's riotous 19th century.

Figure 4: Relative Importance of Peasant, Militia, and Secret Society Unrest



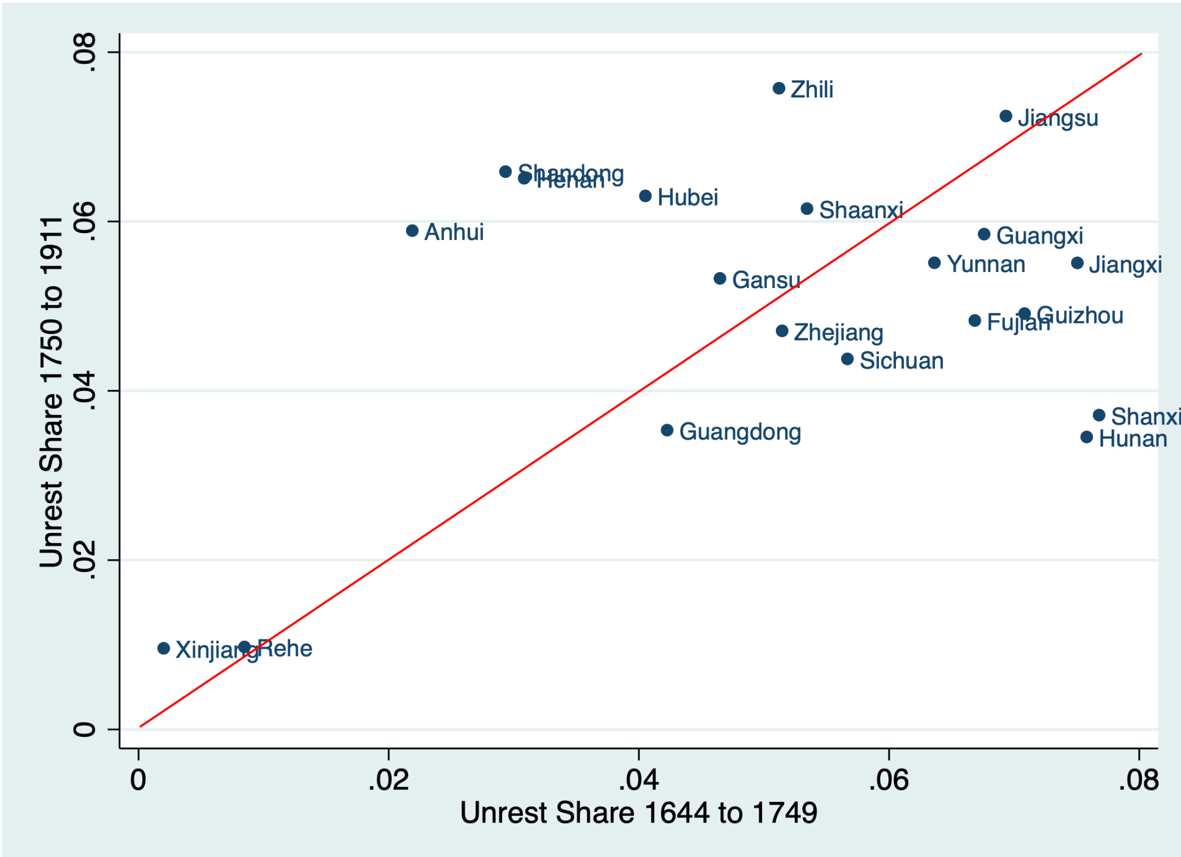
Notes: Figure shows lowess-smoothed share of peasant, militia, and secret society unrest in all unrest events identified by the baseline GUWEN-BERT classifier.

In terms of overall frequency, most unrest events are associated with militia activity, followed by peasant unrest. Secret society unrest accounts typically for less than 20% of all unrest. In relative terms, peasant unrest was most important around the year 1725, with a share of around 40%, while the relative frequency of secret society unrest peaked around the year 1800 with just under 20%.

4.2 Provincial Patterns

Broad regional patterns of unrest can be summarized at the province level. Figure 5 shows the share of each province in all unrest events for two time periods, the early and the later Qing dynasty (1644-1750 and 1750 to 1911, respectively).

Figure 5: Provincial-Level Variation in Unrest by Subperiod

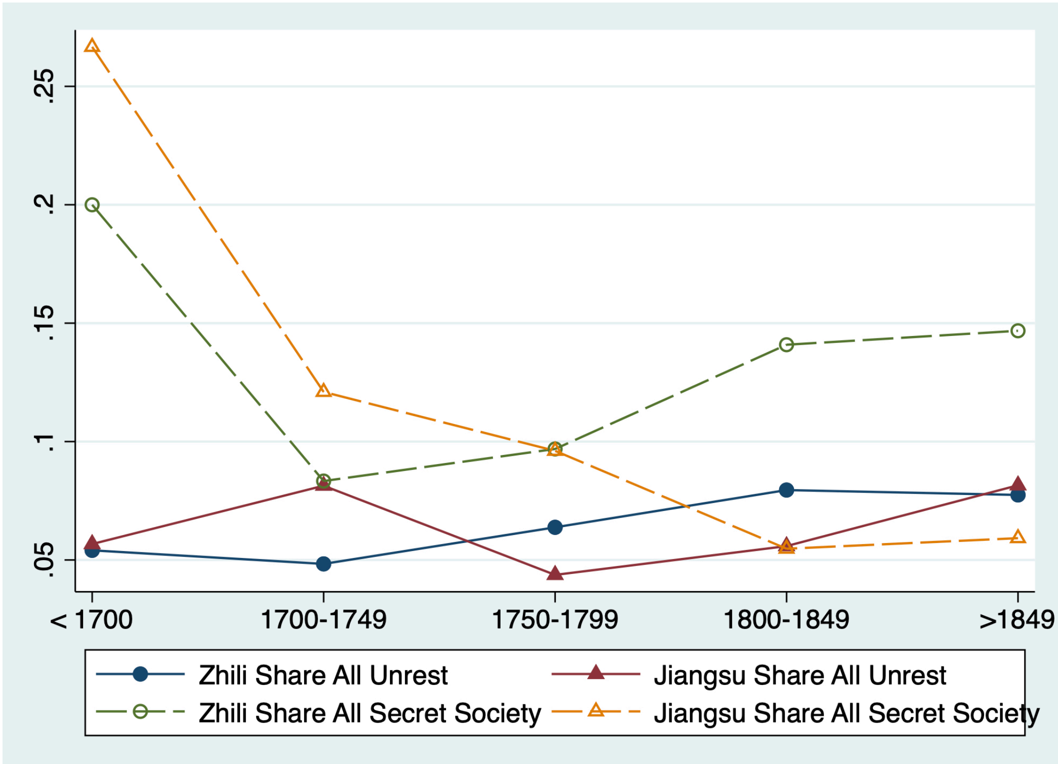


Notes: Authors' computation based on all unrest events identified by the baseline GUWEN-BERT classifier.

Figure 5 indicates that there is considerable variation in the role that a given province plays for overall unrest based on our analysis of the *Qing Shilu*, with Shanxi, e.g., accounting for almost twice the share of unrest of Guangdong during the early Qing period. Furthermore, while in some provinces unrest activity does not change much over time, for other provinces there are substantial changes. An example of the former is Jiangsu province (close to the main diagonal), and Anhui province is an example of the latter (far from the main diagonal). This is interesting because it suggests that unrest activity as identified here may drastically change over relatively short geographic distances (Jiangsu and Anhui are neighboring provinces).

Figure 6 presents a closer look at changes over time in the importance of two specific provinces, namely Zhili and Jiangsu. These provinces include the cities of Beijing and Shanghai, respectively.

Figure 6: Unrest in Zhili and Jiangsu Provinces



Notes: Authors' computation based on unrest events identified by the baseline GUWEN-BERT classifier.

Figure 6 indicates that while the share of both Zhili and Jiangsu in all China unrest increases over time, this is a moderate increase from about 5.5% before the year 1700 to 7.5% in the late 19th and early 20th century. In contrast, Zhili and Jiangsu are both experiencing secret society unrest events at a relatively high rate during the early Qing; both provinces account for well above the roughly 5% that would be their share if these unrests were distributed uniformly across provinces (there are 21 provinces in the analysis).

Furthermore, after a substantial drop in the share of secret society unrest in the early 18th century for both provinces, Zhili's share rises while Jiangsu's share falls, and by the end of the 19th century Zhili province accounts for more than double the share of Jiangsu in terms of secret society unrest. This is consistent with the hypothesis that the computer-based classification approach is able to detect changes in spatial patterns over time at the sub-class (secret society/militia/peasant) level.

The following increases the level of regional disaggregation by turning to the changing patterns of unrest activity across prefectures.

4.3 Prefecture Distribution

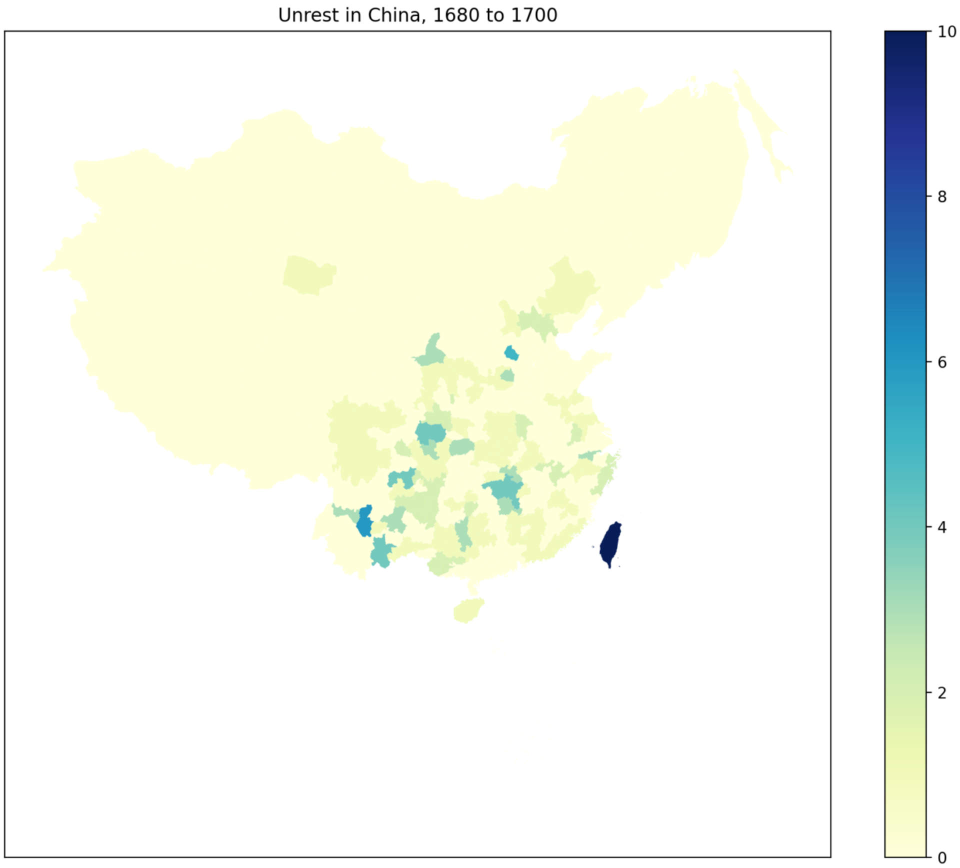
In this section, we examine protest data at the prefecture level to discern patterns related to geographical dimensions. Our Harvard Historical GIS data aligns towns and prefectures from the Qing Dynasty manually.[9] To determine the prefectures where unrest occurred in each entry, we iteratively checked for the presence of town or prefecture names, assigning any mentioned locations as the protest sites. This study focuses solely on prefecture-level information, thus towns are reclassified under their respective prefectures.

We utilize the CHGIS data to create heat maps illustrating the geographical spread of unrest in the Qing Dynasty's Chinese regions. Figures 7 and Figure 8 depict the distribution for two distinct periods, 1680-1700 and 1840-1860, at the prefectural level. We have also developed a website using Streamlit (see https://streamlit.io/) to visualize the results using heatmaps, as in Figures Figures 7 and Figure 8. The interested reader can access the website with the following link: http://5.78.89.73:8501. In addition, this website allows users to query actual unrest entries from the *Qing Shilu* shown in the heatmaps.

The comparison reveals significant differences in both the intensity and geographical extent of protests in China during these two twenty-year spans. Figure 7 shows the unrest geographical distribution from 1680 to 1700, and most severe unrest during this period was the Zheng Keshuang unrest in Taiwan prefecture. Figure 8 illustrates the unrest distribution from 1840 to 1860, when then most severe incident was the Taiping Rebelling, mainly occurring in the southeastern part of China.

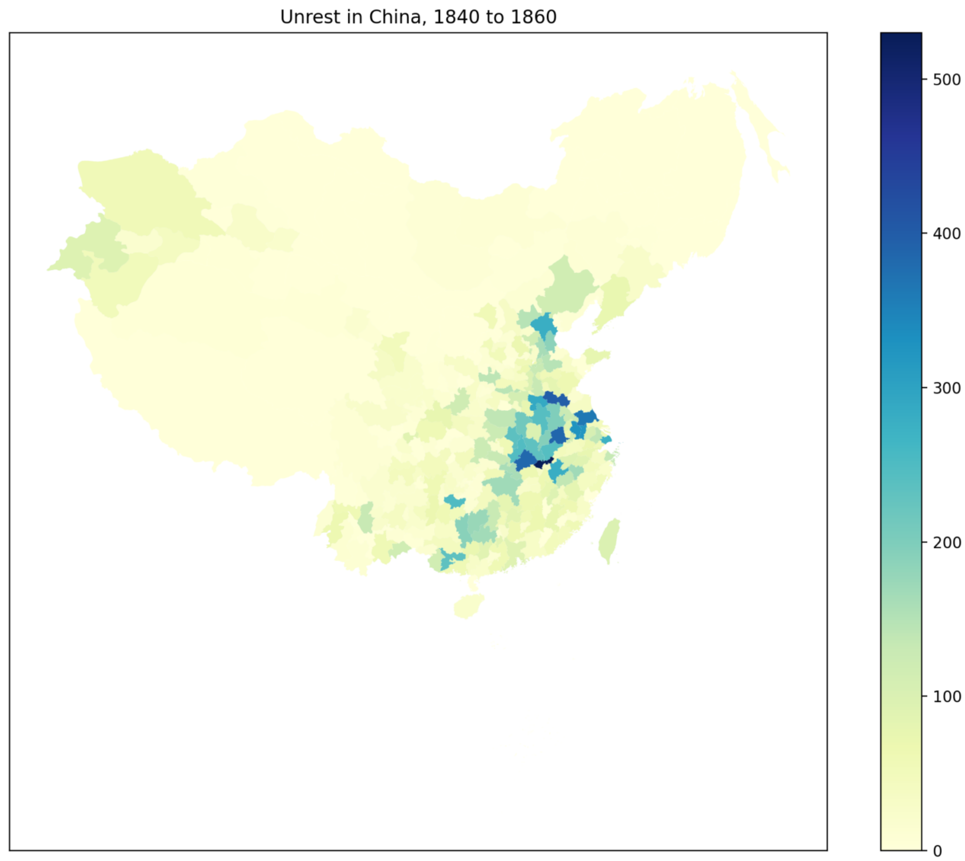
[9]CHGIS, see https://gis.harvard.edu/china-historical-gis

Figure 7: Geographical Distribution of Unrest in 1680-1700



Notes: Authors' computations based on GUWEN-BERT classifier.

Figure 8: Geographical Distribution of Unrest in 1840-1860



Notes: Authors' computations based on GUWEN-BERT classifier.

5 External Validation

This section discusses the validity of our computer-based measure in the light of external evidence. We begin by comparing the incidence of identified events from the *Qing Shilu* with China-wide information on plausible triggers of unrest and protests. They come in form of harvest shocks as reflected by high grain prices, as well as weather shocks in terms of very wet or very dry climatic conditions. Given the importance of agriculture in China during the sample period, and the fact that many persons lived close to subsistence, high grain prices and extreme weather raise the chance of various forms of social unrest. An advantage of employing grain prices analysis is that it is feasible at the same, relatively disaggregated level as our measure of *Qing Shilu* events, namely the year-month by prefecture level.

In the second part of this section, we compare our measure to existing measures aggregated to the annual-national level.

5.1 Using External Information on Weather and Grain Price Shocks

We ask whether factors that historically have triggered unrest in China and elsewhere are indeed correlated with unrest incidents that we identify from the *Qing Shilu*. We consider extreme weather and high grain prices as shocks that frequently cause local food crises and unrest. The analysis employs the highest grain price in a given month (denoted by m) and year (t) in a prefecture (p). The price data comes from the Qing grain price reports as employed in Keller, Shiue, and Wang (2021) and Shiue and Keller (2007). It is an unbalanced sample with a range from year 1736 to 1911. Information on historical weather comes originally from State Meteorological Society (1981) which presents annual information for 120 weather stations across China. They are matched to the closest prefecture following Keller, Shiue and Wang (2021). Weather is reported with five categorical variables, from a value of 1 (very wet) to 5 (very dry), where the value 3 is normal climatic conditions in this location. We employ simple OLS regressions of the following form:

$$prot^{q}_{mtp} = \beta_0 + \beta_1 price_{mtp} + \beta_2 weather_{tp} + \beta' X + \varepsilon_{mtp}, \tag{15}$$

where $prot^{q}_{mtp}$ is our unrest measure of type q, where q is militia, peasant, or secret society, for a given month m, year t, and prefecture p (entered as $\log(prot^{q}_{mtp}{+}1)$). The variable $price_{mtp}$ is the highest price of grain in that month (-year) and prefecture (in logs, with one added). We employ logs to address potentially non-linearities but results are broadly similar without taking logs. The variable $weather_{tp}$ is an indicator for extreme weather in this prefecture and year, defined as either extreme drought or extreme wetness (values 1 or 5); extreme drought and wetness are typically associated with harvest failure leading to food crisis.

The vector X includes a flexible, sixth-order polynomial trend at the national level, as well as prefecture and month fixed effects. Inclusion of prefecture fixed effects means that we allow for arbitrary differences in the propensity for unrest across prefectures, while inclusion of month fixed effects accounts for the fact that grain prices vary over the harvest cycle. In addition, the onset of unrest could also vary by month of the year.[10] Conditional on these variables, we assume that ε_{mtp} is a well-behaved mean-zero error term,

[10]Given the size of China, regions typically grow different types of grain. In order to maximize sample size we employ prices on three different types of grain, namely second-quality rice (mostly grown in the center and south of China), millet

although we allow for heteroskedasticity. Our interest lies primarily in the coefficients β_1 and β_2, namely, whether high grain prices and extreme weather are correlated with the incidence of social unrest in China at the prefecture level. Table 3 shows the results.

Table 3: Grain Price Shocks, Extreme Weather, and the Incidence of Unrest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Militia	Secret Society	Peasant	All	Militia	Secret Society	Peasant
High Price	0.033**	0.022**	0.002**	0.010**	0.034**	0.022**	0.002*	0.010**
	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)
Extreme Weather					0.005**	0.002	0.001**	0.002**
					(0.001)	(0.001)	(0.000)	(0.000)
Month FE	Y	Y	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y	Y	Y
Rsq	0.066	0.066	0.009	0.012	0.072	0.073	0.009	0.013
N	323,281	323,281	323,281	323,281	249,813	249,813	249,813	249,813

Notes: Dependent variable is GUWEN-BERT-based measure of unrest. Estimation of equation (15) by OLS. All specifications also include a trend in form of a sixth-order polynomial in year-month at the national level, as well as an indicator for each grain (rice, wheat, and millet). FE stands for fixed effects. Robust standard errors in parentheses; **/*/+ indicates significance at the 1%/5%/10% level.

We begin by isolating the role of high grain prices because the sample is somewhat larger, see the left side of Table 3. Employing all incidents of unrest, we see that a ten percent higher grain price is associated with 3.3% more unrest events (column (1)). Furthermore, each individual form of unrest is positively associated with higher grain prices (columns (2) to (4)). Quantitatively, militia unrest is most closely related to grain price shocks, while the grain price point estimate in the secret society specification is smallest.

Next, we add the measure of extreme weather to the analysis (right side of Table 3). Generally, extreme weather is positively related to social unrest in China. Moreover, among the different forms of unrest extreme weather is most strongly related to peasant unrest, see column (8). Because farmers are highly exposed to climatic condition, this is what one might have expected. Militia events are not correlated to extreme weather at standard levels of statistical significance (column (6)).

Overall, this analysis indicates that well-known triggers of local unrest are indeed correlated with the

(mostly in the north), and wheat (in many parts of China). Differences in the level of grain prices are accounted for through an indicator for each grain that is part of the vector X.

measures that our machine-learning approach has identified. In addition, Table 3 provides evidence that alternative forms of unrest that we have identified are related to local triggers in different and often plausible ways. This is confirmed by a series of robustness checks, see TableA2 in the Appendix.

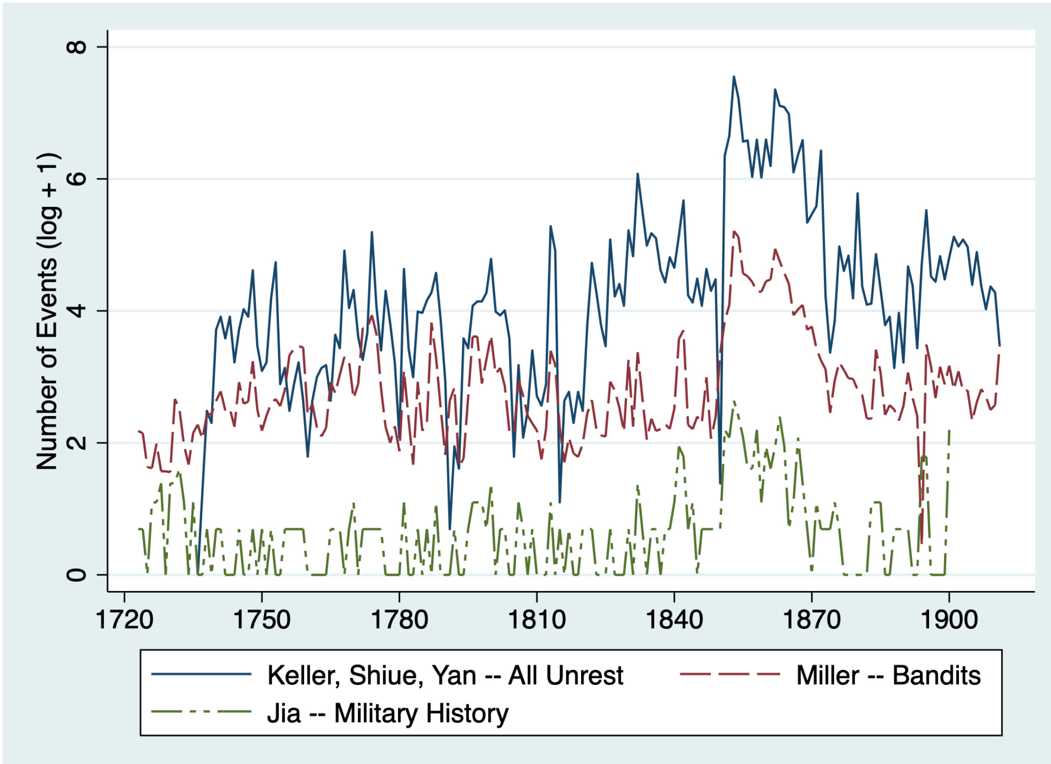
5.2 Comparison with Existing Measures

This section compares our machine-learning unrest measures to existing work on social unrest during Qing China. The existing work is summarized in Table A.1 in the Appendix. Comparisons are complicated by the fact that each study has a somewhat different focus, leading the authors to define social unrest differently (multiple measures are provided in some cases). Furthermore, the studies also differ in terms of sample period, level of geographic aggregation (national, province, or prefecture), and frequency (annual or monthly). Specifically, Chan (1983) and the source material for Zhang et al. (2007) and Jia (2014) are the only studies providing sub-national analysis. As far as we know, this study presents the only analysis of social unrest in Qing China at the local and high-frequency level (prefecture-by-year-and-month).

One way of comparing measures from different studies, albeit an imperfect one, is to aggregate them to the annual and national level.[11] Several measures are available between the year 1722 and the end of the Qing dynasty (1911), and Figure 9 shows a comparison.

[11] It is imperfect because the national average may gloss over important regional concentrations of social unrest that threaten the government, see e.g. Figures 7 and 8.

Figure 9: National Unrest Measures Compared



Notes: Authors' computations based on the sources given; see also Appendix A.

First, Figure 9 indicates that measures from these studies differ in their scale. Our supervised machine-learning approach yields typically a higher number of events than recorded in other sources. For example, there are 258 national entries in China's Military History as employed by Jia (2014), in contrast to more than 58,000 unrest events identified by the GUWEN-BERT approach. Second, there is broad agreement among all measures that events were more common during some subperiods –such as the Taiping Rebellion in the 1850s and 1860s– than other subperiods. The bilateral correlation between the series in Figure 9 is 0.62 or higher. In particular, our GUWEN-BERT-based measure has a correlation of about 0.76 with the measure of Miller (2013). Third, year-to-year variation for the GUWEN-BERT-based measure is higher than for the other measures; Miller's bandit series, e.g., exhibits a coefficient of variation of 1.2, compared to 1.8 for the GUWEN-BERT-based measure.

We also relate these national measures of unrest to the grain price and weather shock triggers considered earlier, recognizing that a national measure of high grain price is necessarily a noisy measure given the size of China.[12] Table 4 shows results from this analysis.

Table 4: Social Unrest, Grain Prices, and Weather Shocks at the National Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Militia	Secret Society	Peasant	Bandits	Mass Distur-bance	Military History
Source	Keller, Shiue, and Yan				Miller	Chan	Jia
High Price	3.397**	4.143**	2.175**	2.458**	0.521	0.188	0.800+
	(0.668)	(0.778)	(0.596)	(0.462)	(0.343)	(0.523)	(0.464)
Extreme Weather	1.427	0.355	2.158**	2.243**	−0.798	−0.531	−0.187
	(0.959)	(1.201)	(0.807)	(0.711)	(0.573)	(0.883)	(0.566)
R^2	0.141	0.131	0.134	0.179	0.016	0.005	0.018
N	175	175	175	175	174	116	164

Notes: Results by OLS. Unrest measures are the sum over all of China in a given year (log, one added). High price is average of the highest monthly prefectural grain price aggregated to the national-annual level. Extreme weather is the average of the annual extreme weather indicator across all prefectures. Number of observations is the number of years a particular measure is available. Robust standard errors in parentheses; **/*/+ indicates significance at the 1%/5%/10% level.

For all social unrest measures, we find a positive correlation with both high grain prices and extreme weather (Table 4, column (1); weather correlation not significant at standard levels). For peasant and secret society unrest events, there is not only a strong positive correlation with high grain prices but also with extreme weather (see columns (3) and (4)). In contrast, militia unrest is primarily related to high grain prices (column (2)).

How does this compare with existing measures? Results are shown in columns (5), (6), and (7). Typically, the correlation of social unrest events with high grain prices and extreme weather is weaker than according to our GUWEN-BERT measure. Only half of the coefficients are estimated to be positive, and only one is significant at standard levels (see columns (5), (6), and (7)). To the extent that extreme weather and high grain prices are triggers of unrest, the GUWEN-BERT-based measure performs better than existing measures when employed at the national level. Irrespective of coefficient sign and significance, unrest triggers such as high grain prices and extreme weather account for at least seven times of the variation in social unrest with our supervised machine learning measure, compared to existing measures (R^2 of 0.131

[12]Specifically, because the total number of social unrest events in China in a given year is not likely related to the highest grain price in a single prefecture of China, we define High Price in the present analysis as the average of the maximum grain price by prefecture.

versus R^2 of 0.018 in columns (2) and (7), respectively).

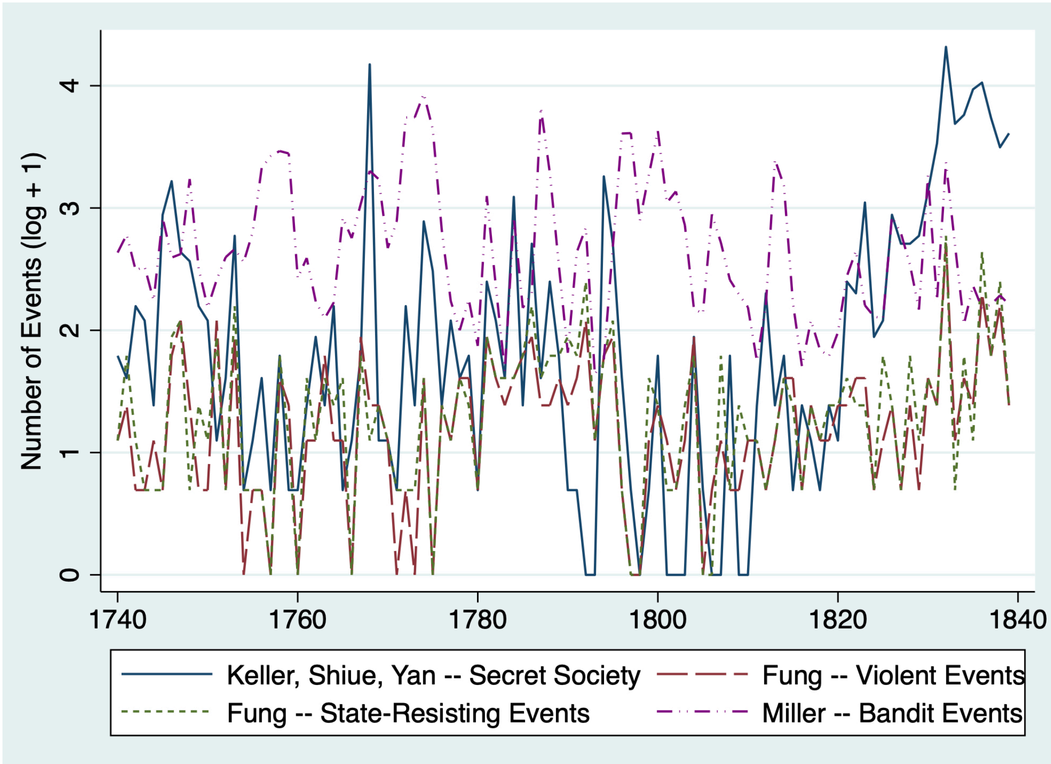
It is also interesting that the GUWEN-BERT approach yields similar patterns of correlation with grain price and weather shocks whether employed at a disaggregated or aggregated level, compare columns (5) to (8) of Table 3 and columns (1) to (4) of Table 4. In particular, peasant and secret society unrest is more strongly related to extreme weather than militia unrest.

5.3 Threats to State Capacity

A subset of unrest events during the Qing dynasty is directed squarely against the state, presenting an immediate threat to state capacity. These events are often associated with the activity of so-called secret societies (*jieshe shejiao*).[13] Among prior work studying unrest that might pose a risk to state capacity is Hung (2011), who identifies events that are described as state-resisting and violent, respectively. Figure 10 compares these series with the secret society unrest series based on the GUWEN-BERT classifier for the period 1740 to 1840. We also show Miller's (2013) Bandit series for comparison.

[13] Section B provides additional information on secret societies in China during the Qing dynasty.

Figure 10: Unrest During the Mid-Qing in Comparison



Notes: Autors' computations based on GUWEN-BERT classifier as well as Miller (2013) and Hung (2011).

Notice that the degree of correlation between these series in Figure 10 is lower than for the overall unrest series in Figure 9 above. One reason for this is that the definition of what classifies as an event across studies has a relatively large effect at the subcomponent level, whereas there is more agreement on the total number of events no matter how they are sub-classified. Another reason is that the range of years shown in Figure 10 does not include times of highest-level of unrest activity such as the Taiping Rebellion, during which it is relatively easy to agree on unrest trends.

Our GUWEN-BERT-identified secrety society series turns out to be most highly correlated to Hung's (2011) violent and state-resisting series (correlation of 0.43 and 0.36, respectively, in Figure 10). It would seem to be plausible that secret society unrest can be violent and state-resisting, thereby presenting a threat to Qing state capacity. In contrast, the correlation between secret society unrest events and Miller's (2013) bandit series is close to zero. The difference between these correlations provides additional evidence that supervised machine-learning picks up relevant variation in what motivates a particular unrest.

The following compares the correlation of Hung's state-resisting and violent event series with grain price and weather shocks to that of the GUWEN-BERT-identified secret society series. Secret society activity turns out to be related to both high grain prices and extreme weather (Table 5, column (1)). In contrast, violent events as identified by Hung (2011) are only correlated to high grain prices, while state-resisting activity is neither correlated with grain price nor with weather shocks. Furthermore, grain price shocks and extreme weather also explain more variation in secret society unrest than in violent or state resisting events, as measured by the R^2.

Table 5: Threats to the State, Grain Prices, and Weather Shocks

	(1)	(2)	(3)
Events	Secret Society	Violent	State Resisting
Source	Keller, Shiue, Yan	Hung	
High Price	2.353^{*}	1.311^{*}	0.892
	(1.166)	(0.580)	(0.650)
Extreme Weather	2.196^{+}	0.740	0.278
	(1.139)	(0.588)	(0.610)
R^2	0.095	0.070	0.024
N	100	100	100

Notes: Results by OLS. Unrest measures are the sum over all of China in a given year (log, one added). High price is average of the highest monthly prefectural grain price aggregated to the national-annual level. Extreme weather is the average of the annual extreme weather indicator across all weather stations. Sample is years 1740 to 1839. Source of dependent variables in columns (2) and (3) is Hung (2011).Number of observations is the number of years a particular measure is available. Robust standard errors in parentheses; **/*/+ indicates significance at the 1%/5%/10% level.

To the extent that grain price shocks and extreme weather matters for activity that is a threat to state capacity, the results of Table 5 indicate that the GUWEN-BERT-based approach identifies unrest activity that threatens the Qing state at least as well as earlier approaches.

6 Conclusions

To understand the ways in which social unrest was reported by Qing officials, we train different language models based on nearly one-thousand existing samples describing different kinds of protests in the *Qing Shilu*, the Veritable Records, which is based on court documents of the Qing Dynasty. We then apply the model to the full text of the *Qing shilu* to extract all the records about unrest in the empire. Unlike a human researcher, who may search for evidence on known conflicts, the machine starts with a clean slate, free from preconceptions about which time periods were more or less important. One immediate outcome is the number of records that deal with unrest and the regions that were affected by unrest far surpasses any existing compilations, which have tended to focus on the largest and most well-known events.

We employ supervised machine learning methods to recognize unrest-related entries and their types in the *Qing Shilu*. Among the three models tested, the GUWEN-BERT Classifier demonstrates superior performance in both the binary classification (unrest vs. not unrest) and the tri-category classification into peasant, militia, or secret society unrest. Employing the GUWEN-BERT Classifier to identify all unrest-related entries and their types in the *Qing Shilu*, we present new evidence on the temporal and regional patterns of social unrest in Qing China.

Because the computer-based approach distinguishes different types of unrest we can show that temporal variations are not uniform across categories of unrest. Distinguishing different types of unrest allows us to have a more precise understanding of what the Qing rulers saw as a threat to the legitimacy of Qing rule, and provides new evidence on elements of state capacity.

Confronting our computer-identified measures of unrest with well-known triggers of unrest helps to externally validate our approach. At both the regionally disaggregated and the national level, we typically find that social unrest is more likely at times of high grain prices and extreme weather. Grain price shocks and extreme weather turn also out to be more strongly correlated with our machine-learning identified measures of unrest than with measures of unrest that are available in the literature.

Our method increases the number and the scope of social unrest that was present during the Qing. While we are able to match the timing and location of some of the most famous events, most notably the Taiping Rebellion, our analysis also reveals that many social unrest episodes have not been studied yet. Thus, a general lesson is that AI output may be useful in supplementing human-based research, showing us that unrest was more frequent and regionally pervasive than previously understood. This difference may be attributable to the possibility that when documents need to be tabulated manually by humans, smaller unrest events, or events in regions that are deemed less central, receive less attention. Additional insights can be expected from the application of increasingly more powerful large language models that are currently being developed.

References

[1] Abramitzky, Ran (2015), "Economics and the Modern Economic Historian", Journal of Economic History, published online Dec 16, 2015.

[2] Abramitzky, R., L. Boustan, and K. Eriksson (2014), "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration", Journal of Political Economy Vol. 122, No. 3.

[3] Abramitzky, Ran, Roy Mill, and Santiago Perez (2020), "Linking individuals across historical sources: A fully automated approach", Historical Methods 53:2, 94-111.

[4] Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation", American Economic Review Vol. 91.

[5] Acemoglu, Daron, and James A. Robinson (2019), The Narrow Corridor: States, Societies, and the Fate of Liberty, Penguin Press, New York.

[6] Bailey, Martha, Peter Z. Lin, A. R. Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman (2023), "The Creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database Project", *Historical Methods*, published online Aug 17, 2023.

[7] Besley, Timothy, and Torsten Persson (2011), Pillars of Prosperity: The Political Economics of Development Clusters, Princeton: Princeton University Press.

[8] Cantoni, D., L. J. Heizlsperger, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2022), "The fundamental determinants of protest participation: Evidence from Hong Kong's antiauthoritarian movement", Journal of Public Economics 211, 104667.

[9] Chan, J. H.-k. (1983) *Mass Disturbances and Protest Movements in Late Imperial China, 1791-1911: A Time-series Study of Collective Actions*, Ph.D. thesis, University of Pittsburgh.

[10] Card, Dallas, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky (2022), "Computational Analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration", Proceedings of the National Academy of Sciences, July.

[11] Chen, Q. (2015) “Climate shocks, state capacity and peasant uprisings in north china during 25-1911 ce”, Economica, 82, 295_318.

[12] Chen, Zhenhan, Xiong Zhengwen, Li Chen, and Yin Hanzhang (1989), Qing shilujingiishi ziliao [Economic history materials from the Veritable Records], 4 volumes. Beijing: Beijing Daxue, 1989.

[13] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[14] Dincecco, Marco (2017), “The Rise of Effective States in Europe”, Journal of Economic History 75(3), pp. 901-918.

[15] Hung, Ho-Fung (2011) *Protest with Chinese characteristics: Demonstrations, riots, and petitions in the Mid-Qing dynasty*, Columbia University Press.

[16] GuwenBert (2023), BERT language model pre-trained on classical Chinese language, https://github.com/Ethan-yt/guwenbert

[17] Jia, Ruixue (2014), “Weather shocks, sweet potatoes and peasant revolts in historical China”, The Economic Journal, 124, 92-118.

[18] Johnson, Neil, and Mark Koyama (2023), “States and economic growth: Capacity and constraints”, Explorations in Economic History.

[19] Keller, Wolfgang, Carol H. Shiue, and Xin Wang (2021), “Capital Markets in China and Britain, 1770-1860: Evidence from Grain Prices”, American Economic Journal: Applied Economics Vol. 13(3), pp. 31-64.

[20] Kuhn, Philip (1970), Rebellion and Its Enemies in Late Imperial China: Militarization and Social Structure, 1796 - 1864, Harvard University Press, Cambridge.

[21] Miller, I. M. (2013), “Rebellion, crime and violence in qing china, 1722-1911: a topic modeling approach”, Poetics, 41, 626-649.

[22] Mitchener, Kris (2015), “The 4D Future of Economic History: Digitally-Driven Data Design”, Journal of Economic History, published online Dec 16, 2015.

[23] Nunn, Nathan, and L. Wantchekon (2011), “The Slave Trade and the Origins of Mistrust in Africa”, American Economic Review Vol. 101, pp. 3221-52.

[24] Ownby, David (1996), Brotherhoods & Secret Societies in Early and Mid-Qing China, Stanford University Press, Stanford.

[25] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6), 386.

[26] Shiue, Carol H., and Wolfgang Keller (2007), “Markets in China and Europe on the Eve of the Industrial Revolution”, American Economic Review Vol. 97(4), pp. 1189-1216.

[27] Skocpol, Theda (1985), “Bringing the State Back In: Strategies of Analysis in Current Research”, in *Bringing the State Back In* (eds. P. Evans, D Rueschemeyer„ and T. Skocpol), Cambridge: Cambridge University Press.

[28] Sparck Jones, Karen (1972), “A Statistical Interpretation of Term Specificity and its Application in Retrieval”, Journal of Documentation Vol. 28(1), pp. 11-21.

[29] State Meteorological Society (1981), Zhongguo jin wubai nien hanloa fenbu tuji (Collected Maps of Droughts and Floods in China in the Past Five Hundred Years.) Beijing: Ditu chu ban she.

[30] Xu, S., Li, Y., & Wang, Z. (2017), “Bayesian multinomial Naïve Bayes classifier to text classification”, Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2017 11 (pp. 347-352). Springer Singapore.

[31] Vaccaro, Andrea (2023), “Measures of state capacity: so similar, yet so different”, Quality and Quantity Vol. 57, pp. 2281-2302.

[32] Yang, Xinmei, Abhishek Arora, Shao-Yu Jheng, and Melissa Dell (2023), “Quantifying Character Similarity with Vision Transformers”, forthcoming, *Empirical Methods on Natural Language Processing.*

[33] Zhang, D. D., Zhang, J., Lee, H. F. and He, Y.-q. (2007), “Climate change and war frequency in eastern china over the last millennium”, Human Ecology, 35: 403-414.

A Supplemental Information

A. Data

The following summarizes existing studies that include analysis of unrest events of various kinds during the Qing dynasty.

Figure A11: Alternative Qing Protest Analyses: An Overview

Source	Time Period	Scale/Extent	Primary Source	Content/Methodology
Chan 1983	1796-1911	Province/Annual	PhD dissertation	Constructed Index
Hung 2013	1740-1839	National/Annual	Ming/Qing Shilu + Gazetteers	List
Jia 2014, Zhang 2007	1470-1900	Prefecture/Annual	China Military History	List
Miller 2013	1722–1911	National/Month	Qing Shilu	Topic Modeling
Chen's book	1600-1800		Qing Shilu	List
Our Classifier	1645-1910	Local/Month	Qing Shilu	Classifier

Notes: Summary of studies employing alternative measures of unrest during the Qing dynasty. See references for sources.

B Robustness

This section reports a number of robustness checks. First, we report evaluation results for a four-way classifier that directly classifies each entry into (i) No unrest, (ii) peasant unrest, (iii) militia unrest, and (iv) secret society unrest. That is, the two-step approach of first classifying into unrest versus not-unrest, followed by a three-category classification into peasant, militia, or secret society unrest is replaced by a one-step approach with four classes. Table A1 presents evaluation results for all three machine learning approaches.

Table A1: Four-way Classifier Evaluation Results

	Accuracy	Precision	Recall	F_1
Multinominal Naive Bayes	0.618	0.603	0.392	0.361
TF-IDF + Perceptron	0.874	0.848	0.832	0.839
GUWEN-BERT Classifier	0.922	0.905	0.895	0.899

Notice that for the four-way classifier differences in performance are substantial. In particular, the Multinomial Naive Bayes four-way classifier does not work in this setting, yielding an accuracy of around 0.6 and an F_1 of only 0.36. The TF-IDF Perceptron model achieves now performance metrics around 83 to 87%, compared to about 90% in our baseline approach (see Table 2). Our baseline GUWEN-BERT model performs best also in the four-way classification setting, with a macro-F_1 of just under 90% and an accuracy above 92%. This provides more evidence that the GUWEN-BERT classifier is the most promising in the present context.

Table A2 presents grain price shock and extreme weather results for two alternative approaches. The first employs as the dependent variable the GUWEN-BERT output of the four-way classifier just discussed, see columns (5) to (8) in Table A2. The second alternative approach employs the probabilities instead of the binary (0/1) unrest outcomes of our GUWEN-BERT baseline approach; results are shown in columns (9) to (11) of Table A2. Baseline results from Table 3 in the text are presented again in columns (1) to (4) of Table A2 for convenience.

Table A2: Robustness: Grain Price Shocks, Extreme Weather, and the Incidence of Unrest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Baseline				Four-Way				Probabilities		
	All	Militia	Secret Society	Peasant	All	Militia	Secret Society	Peasant	Militia	Secret Society	Peasant
High Price	0.034^{**}	0.022^{**}	0.002^{*}	0.010^{**}	0.032^{**}	0.020^{**}	0.001^{*}	0.012^{**}	0.017^{**}	0.003^{*}	0.008^{**}
	(0.003)	(0.003)	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Extreme Weather	0.005^{**}	0.002	0.001^{**}	0.002^{**}	0.003^{**}	0.000	0.001^{**}	0.002^{**}	0.001	0.001^{**}	0.002^{**}
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Rsq	0.072	0.073	0.009	0.013	0.065	0.069	0.008	0.015	0.073	0.010	0.013
N	249,813	249,813	249,813	249,813	249,813	249,813	249,813	249,813	249,813	249,813	249,813

Notes: Dependent variable is GUWEN-BERT measure of unrest as described in the text. Estimation of equation (15) by OLS. All specifications also include a trend in form of a sixth-order polynomial in year-month at the national level, as well as an indicator for each grain (rice, wheat, and millet). Mixed specification for all unrest events not available. FE stands for fixed effects. Robust standard errors in parentheses; **/*/+ indicates significance at the 1%/5%/10% level.

We see that adopting a four-way classification approach leads typically to somewhat lower grain price and weather correlations than the baseline approach, although differences are small. In addition, the fraction of variation accounted for by grain price shocks and extreme weather is somewhat lower, see the R^2s in columns (5) and (1), respectively. This is consistent with the two-step baseline approach–first into unrest vs. not unrest, then into the particular unrest type–performing somewhat better than the one-step approach with four classes (see Tables 2 and A1, respectively). Employing probabilities instead of binary unrest outcome variables does not lead to major differences in the grain price and weather correlations, see columns (9) to (11) and columns (2) to (4), respectively.

B. Secret Societies during the Qing Dynasty

Our analysis focuses on three forms of unrest that are distinguished in the work by Chen et al. (1989). Militia (*wuzhuang*) incidents involve armed uprising struggles. The militia may be formed by farmers to protest high rents charged by landlords or high taxes charged by the government. The probability of a militia unrest event would also be affected by the population pressure and the quality of the harvest. This may make it challenging to distinguish militia unrest from another form of unrest, those involving peasants. From the point of view of the Qing rulers these forms of unrest because a food riot, especially involving an armed militia, might reduce the dynasty's power and legitimacy. At the same time, Qing rulers would not fundamentally view those involved as enemies of the state.

The third form of unrest on the other hand, involve secret activities such as associations and sects (*jieshe shejiao*). In addition to anti-Qing secret society chapters, they include brotherhoods, roving bands, triad pirates, and the amalgamation of alienated and embattled groups that subscribed to religious messianistic movements (Kuhn 1970, Ownby 1996). From the state's perspective, these individuals would be seen as rebels, bandits, criminals, or simply enemies of the state. The relevance of these jieshe protests for the Qing's state capacity is immediate.

The origin of secret societies lies centuries before the West's intervention in China during the 19th century. Specifically, the White Lotus rebellion with major uprising in Shandong and Henan provinces in 1775 is the re-emergence of the anti-Mongol revolt of the 14th century that led to the founding of the Ming Dynasty (1368-1644). Being ethnically Mongol themselves, the Qing rulers naturally feared anti-Mongol sentiment in China's population. Secret societies often grew out of informal cultural organizations–groups of people pooling money for a wedding, or for some financial expense on a revolving credit system.

Secret societies gained in importance during the Qing dynasty due to population growth and high prices, increased social mobility, and geographic dislocation. To some extent secret societies filled a vacuum left behind by the Qing state that provided fewer and fewer local public goods, to another extent the growth of secret societies and brotherhoods was stimulated by the incompetence of regular Qing troops. Secret societies also blossomed because Qing punishments of suspected secret society activity was harsh. An example of draconian Qing punishments is that Qing officials would burn the houses of an entire village on suspicion of one rebel (Ownby 1996). If one takes as the central indication of state capacity that the state has a monopoly on the use of physical force, that was lost due to local militarization in the border regions of China in the early 19th century, and the state monopoly on the use of force was lost by the 1850s also in the river valleys of China (Kuhn 1970).