Pre Analysis Plan : Social capital and violence during the French Revolution

Capucine FERRE

March 2024

1 Introduction

This project aims to study the impact of social capital on violence during the French Revolution. We will use the *Cahiers de Doléances* of 1789 to build proxy measures of social capital and estimate how they relate to the ensuing revolutionary violence. The analysis of the contents of the *Cahiers* will be based on two different methods: first, we will draw on Shapiro et al. (1998) database on the content of grievances which were coded by humans through a thorough process. Then, we will use machine learning techniques such as Topic modelling and Sentiment analysis. We will conduct an analysis with national coverage using the so-called *Cahiers de Synthèse* and at a more local level through a case study of Normandy using the parish-level *cahiers* (the so-called primary *cahiers*).

Studying the causes of the French Revolution, a major violent event throughout France, and especially the determinants of violent events disparities across the kingdom at the time will yield valuable insights regarding modern revolutionary violence and shed some light on the onset of this crucial event. The impact of social capital on conflict risk has gathered little attention, because of the difficulty of controlling for reverse causality (since the impact of conflict on social capital is pretty well documented), and of finding an exogenous source of variation of social capital. Focusing on social capital during the French Revolution is a novel and relevant way to better understand this conflict and, more generally, conflict onset. Data regarding violent events before the Revolution are available, and thus we will be able to control for pre-existing violence. Plus, the *Cahiers de*

Doléances were written just before the revolution and contain rich insights on a locality's grievances and mindset: we will thus be able to build proxy measures of social capital thanks to this extensive information.

Even though the *cahiers* were sporadically studied before, Hyslop (1934) was one of the first to work quantitatively on the *Cahiers* to study nationalism. A database was born from this work, the first to attempt to gather as many *cahiers* as possible with an analysis of their content. Shapiro et al. (1998) pushed the analysis even further with a detailed content analysis of each cahier de synthèse and a representative sample of parish *cahiers* building a database coded via the innovative use of computers. Since then, quantitative works on the analysis of the *cahiers* have been few and far between.

Thus, we will contribute to the literature on the analysis of the *Cahier de Doléances* by proposing an innovative analysis of the *cahiers* using machine learning techniques. A first analysis, called Sentiment analysis, will allow us to study the tone of the texts. A second analysis, called Topic modelling, will make it possible to reconstitute the topics discussed in the *cahiers* without having to use manually constructed databases, which are more subject to bias. This method yields a new and less subjective understanding of the content of the *Cahiers* and allows for a less time-consuming analysis of a vast amount of information.

Social capital can be summarised simply as the propensity of individuals to cooperate based on trust. It is measured through indicators of trust, social cohesion, in short, prosocial behaviours that favour cooperation (Paldam, 2000). This concept is increasingly seen as relevant to explain the consequences of post-conflict and might have an influence on the conflicts themselves (Bellows and Miguel, 2009; Blattman, 2009; Bauer et al., 2016). We make a distinction between bridging and bonding social capital:

- Bonding social capital is a form of social capital focused on homogeneous groups (trust and cooperation with similar people)
- Bridging social capital is focused on the links between groups (trust and cooperation with dissimilar people)

The difference between bonding and bridging social capital is a contribution we owe mainly to Putnam (2000). Other ways of describing these concepts already existed, in

the form of "strong ties" for bonding social capital and "weak ties" for bridging social capital (Granovetter, 1973) in social network analyses. The economic literature often prefers terms such as intra-group or inter-group. While the assumptions associated with these terms may vary, the concept remains broadly unchanged.

In studying the impact of social capital on violence during the French Revolution, we are also interested in distinguishing between the two types of social capital, as they could arguably have a different impact.

2 Research Strategy

2.1 The Setting

2.1.1 History of the *cahier* and Revolution violence

A *cahier* of a given zone is a written document containing grievances from local people. These *cahiers* were to be used in the assembly of the Estate General to present wishes and grievances from all three Estates of France (nobles, clergy and commoners). The aim was to suggest actions and reforms to the King. The Estate General of 1789 marks the end of this century-old custom, which dates back to the 14th century.

In the summer of 1788, a fiscal crisis shook the French Monarchy and put in motion the summoning of the Estate-General for May 1789. This event occurred against a background of a long series of recurrent fiscal crises and the inability of the Monarchy to solve them, especially its failure to tax the nobles. The Estate-General last calling was in 1614, so people from all of France took the matter seriously, which yielded a broad set of *cahiers* from all French territories.

The gathering allowed for the representation of almost all regions of France. As far as the Third Estate is concerned, we can distinguish between the so-called *cahiers de synthèse* and the primary *cahiers*. The *cahiers de synthèse* came from a *bailliage* (a legislative district): they were the *cahiers* sent to Versailles. Deputies tasked to bring the *cahiers* to Versailles and to participate in the Estate General were elected during *bailliage* assemblies. At a more micro level, primary *cahiers* came from the primary assemblies of parishes or towns. Those were responsible for voting and sending a representative to sit in

the assembly of the *bailliage* for the elaboration of the *Cahiers de synthèse* and election of deputies. Turning to the Nobility and the Clergy, they mainly produced *cahiers de synthèse* because of their smaller population size.

Deputies were tasked during the *bailliage* assemblies to bring the *Cahier de synthèse* to Versailles for the grand assembly of the General Estate, where all three states would be present. The number of deputies of each *bailliage* was based on the population size, the Third Estate being granted twice as many deputies per *bailliage* than the Nobility and Clergy.

Most of the population at the time were permitted representation. All male registered taxpayers and at least 25-year-olds were called to participate in the assemblies of the Third Estate. Likewise, all male members of the Clergy and Nobility regardless of their rank were authorised to participate. Despite some deviations from those rules, such as the heavy pressure from several provinces institutions to directly elect their representants without following the usual process of secondary assemblies (the monarchy yielded to those demands for Béarn, Navarre and Dauphiné only), the process was mostly respected in all French territories.

The following meeting of the Estate-General at Versailles, starting on the 5th of May, was a significant event of the French Revolution. A consensus on taxes failed to emerge during the following weeks, leading the Third Estate to declare itself the National Assembly of France on the 17th of June. Popular demonstrations and mutinies among the guards and the army eventually led to the King's capitulation on the 27th of June. He commanded the members of the first and second Estates to join the third in the National Assembly, indirectly recognizing its legitimacy.

The start of revolutionary violence really started on the 12th of July when rumours stating that the King was planning to use the Swiss Guards to force the new National Assembly to close and thus retake power circulated among the Assembly and Parisian people. Those rumours brought crowds of protestors into the streets. Soldiers then refused to disperse them (Schama, 2004). The violence of these angry crowds of demonstrators culminated on the 14th of July with the storming of the Bastille, a royal fortress used as a prison and a storage area for arms and ammunition. Many soldiers joined the mob in attacking the fortress, capturing it after several hours of fighting. This symbol of the

King's power was then quickly torn down. Its destruction is considered a major milestone of the French Revolution. Revolutionary violence was then restricted to Paris.

Then, following the storming of the Bastille, violence spread throughout the Kingdom, both in rural areas and in cities. This key period of violence known as "The Great Fear" took place at the very beginning of the Revolution between 22 July and 6 August 1789 (Merriman, 2009). This period of great panic and riot was fueled by rumours of a conspiracy led by the King and the privileged to starve the Third Estate. Those rumours mostly came from fear of revenge due to the taking of the Bastille and constant worry regarding harvest and grain prices. This paranoia resulted in the formation of militia in self-defence. Eventually, it led to violence against nobles and sometimes even clergymen, specifically by attacking their property and destroying feudal documents (Lefebvre, 1983). This blatant breakdown of law and order and frequent attacks on aristocratic property led much of the nobility to flee abroad. The National Assembly abolished the feudal regime to appease fearful commoners.

A period of social unrest, change and violence ensued, both in cities and in rural areas. Protests against the prices of grain and violence against authorities and its symbol of power frequently occurred at the time.

2.1.2 Hypothesis and primary outcomes

Our dependent variable is conflict during the revolution (1789-1799). We are interested in violence that stems from collective action rather than criminality. Our main independent variable will be social capital. It is of interest to note that we assume the reference group used to define both bridging and bonding social capital is the geographical area of the *cahier* (here parishes or *bailliages*) in most of our measures of social capital. So, the bonding part of social capital is related to interaction within this local zone, while the bridging part is related to interactions between this local zone and the state, here in the Kingdom of France. The group of reference will be the Estate for others measures: for a Third Estate *cahier*, bonding social capital is related to its own Estate and bridging social capital is related to others Estate such as the Nobility and the Clergy.

Several hypotheses can be developed depending on the types of social capital. First, bridging social capital is expected to be negatively related to violence during the Revo-

lution due to a better capacity to negotiate with others, better cooperation with various groups such as the Nobility and the elites, and/or better links via trade, for example. The idea is quite intuitive and documented by some works, especially in the literature studying the link between trade and war (Rohner et al., 2013; Martin et al., 2008). This literature, however, also points to a possible positive effect of trade on the risk of war, if the ability to trade with many partners reduces the cost of initiating conflict with one of them (Martin et al., 2008). A similar effect may exist at a more local level: with many ties and wide cooperation with others, a conflict with one of them seems less problematic.

Secondly, bonding social capital is frequently associated with the "dark side of social capital". It can promote a lack of trust in other groups and withdrawal, particularly regarding identity or ethnicity, which can then increase the chances of conflict (Montalvo and Reynal-Querol, 2005; Grosjean, 2014; Satyanath et al., 2017). We thus assume a positive relationship between this type of social capital and violence during the Revolution. It could also be argued that bonding social capital has a negative effect on conflict and violence because of its protective aspect against expropriation, which would discourage attackers and, knowing this, pacify the bonded group. Several examples of this protective aspect of bonding social capital exist, notably in the form of community or neighbourhood protection groups (Bellows and Miguel, 2009; Sawyer, 2005; Jones and Munoz, 2010).

We expect these hypotheses to apply especially in conflict events related to opposition to authorities. For instance, effects could be less important regarding violence linked to the prices of grain, another usual type of violence during the revolution.

It is possible that representativeness (or "voice") plays a key role in this analysis. It portrays how well an area was represented during the whole process of the Estate General. Representativeness captures how people were treated during the making of the *Cahier* and could very possibly have an interaction with social capital. With no interaction with social capital, we expect more representativeness to be negatively related to acts of violence: people who have the feeling that their issues were adequately heard are less likely to resort to violence. High representativeness combined with high bridging social capital (with our main assumption of a negative relationship between bridging and violence), would yield an even more negative relationship with violence. In the same manner, high representativeness combined with high bonding social capital (with our main assumption

of a positive relationship between bonding and violence), would attenuate the positive relationship with violence of high bonding capital.

2.1.3 Measuring social capital before the revolution

How to measure social capital is a central question here. Overall, there are three trends in the economic literature on the measurement of social capital (Bauer et al., 2016):

- A first trend based on the concept of trust measured in surveys such as the World Value Survey.
- A second focused on prosocial behaviour during experimental games (such as the dictator game, the ultimatum game, etc.).
- A third focused on directly observable behaviours (such as Putnam's measure : the density of voluntary organizations, but also abstention in elections, participation in community activities, etc.).

Trust seems the best way to capture the essence of social capital as well as being measurable through widely used surveys, but it has been criticized for its self-reported nature. Glaeser et al. (1999) shows that the answers of individuals to this type of question mainly predict their degree of trustworthiness rather than their degree of trust in others. The use of experimental games based on trust and cooperation is an interesting alternative. Yet, these experiments are costly and time-consuming to set up and yield relatively small samples.

While easier to measure, it is not always clear which component of social capital the third trend captures. Putnam's measure has been criticized because it does not adequately capture informal forms of social capital. In particular, (Fafchamps, 2006) argues that generalized trust can make interactions through clubs less necessary. Plus, when thinking of social capital in a violence-prone setting, one can fear that many measures assume infrastructures (such as voting or voluntary organization indicators) likely to be destroyed by conflicts.

Since we are measuring social capital on the eve of the Revolution, most of these usual measures are unavailable to us. However, we draw from this literature to build our own

measures to capture social capital. Our main idea is to use the *cahiers* to capture both bridging and bonding social capital.

Even though our focus is on Third Estate *cahiers*, Nobility and Clergy *cahiers* will also be used as our second set of measures of social capital related to trust can be built using those *cahiers*.

The *cahiers* will provide a measure mainly akin to the third trend based on directly observable behaviours. Participating in the making of the *cahiers* and being involved in it is similar to participating in communities' activities and voting. It shows a willingness to participate in the society and cooperate. A first class of measures coming from this trend can be used [1]:

- First, the care put into the making of the *cahier* will be explored. This can be captured through several indicators from the *cahiers*, such as the number of grievances, the total length of the *cahier*, the level of detail in expressed grievances, or the number of people in attendance (relative to the population of the area). It is unclear a priori whether those can be classified as a bonding or a bridging type of social capital. Other indicators should allow us to add this nuance.

- Then, the content of the grievance itself can be used. More local grievances might reveal a more bonding type of social capital in an area, while more national-level grievances shall reflect more bridging social capital. Again, it shows an interest in participating in society, whether with similar individuals or others. Plus, grievances demanding more representation can also reveal a desire to participate in society hence capturing more bridging social capital.

We can also extract from the *cahiers* measures akin to the first trend, related to trust:

- First, disagreement or agreement between the Nobility/Clergy *cahiers* and Third Estate *cahiers* of an area hint at cooperation issues and thus low or high bridging social capital. It might be interesting to distinguish between agenda disagreement (if something should be considered a problem) and program disagreement (agreement on the problem but not on how to solve it) as Shapiro and Markoff did. The first one might be more informative on cooperation issues.

[1] See Appendix for precision

- In line with this first indicator, we plan to use sentiment analysis to better understand the tone (positive or negative) of each *cahier* regarding some categories of people, such as the Nobility, the King and local communities... to measure, depending on the group, bonding or bridging social capital.

2.2 Data sources

The *cahiers de synthèses* are freely accessible online from the archive work contained in *Archives Parlementaires de 1787 à 1860: États généraux, Cahiers des sénéchaussées et bailliages* (Laurent and Mavidal, 1879). Almost all of the known *cahiers de synthèse* are present there. The *cahiers de synthèse* are well preserved on the whole French territory. Those *cahiers* are to be used in the topic analysis part for a national coverage.

We will also perform a topic modelling analysis for regional coverage. Indeed, disparities of parish *cahiers* across France make it difficult to obtain national coverage at the parish level. We have chosen Normandy for this analysis for several reasons. First, data are of good quality as almost all parish *cahiers* from Normandy (roughly 1800) are available. Plus, most were published (meaning their content was written down using a writing machine, so we do not have to decipher handwriting) and are accessible online. Some *cahiers* will still require deciphering handwriting through OCR (optical character recognition). For them, we will use the Transkribus software. Using deciphered *cahiers*, we will train the software to recognize handwriting of the time, correcting it if needed. A list of the parish *cahiers* of Normandy, specifying where to find each of them, is available in Maneuvrier-Hervieu (2021).

Data collection and analysis will follow three steps depending on the data type due to how time-consuming processing certain types will be. First, we will collect and analyse published *cahiers* available online. This represents approximately 400-500 *cahiers* from 5 *bailliages* in Normandy. Power analysis will be performed afterwards to better understand how likely the occurrence of a type 2 error is and how important it is to build a bigger sample. Then, if needed and feasible, we may collect and analyse *cahiers* not available online. Those published *cahiers* are only available in physical forms and must be scanned and processed through OCR. This should give us a sample of roughly 1000 *cahiers*. Again power analysis will be performed. Finally, depending on the outcomes of the previous

two steps, we will consider collecting and analysing handwritten *cahiers*, which yields the entirety of the *cahiers* from Normandy (1800). The aim is to eventually collect all *cahiers* from Normandy. However, due to time limitations, analyses will first be performed on a subset of the entirety of the *cahiers* from Normandy while taking into account statistical power to keep in mind if the subset is big enough to allow detection of an effect.

Thus, we can explore the text of both the *cahiers de synthèse* with a national coverage and the parish *cahiers* with a regional coverage for Normandy for the topic analysis.

In addition, we will also use the dataset of Shapiro et al. (1998) that does not contain the text of the *cahiers* but precisely describes the content of each grievance. It is the most complete database to date. It includes a coding system indicating for each *cahier* its precise demands, namely the subject of the demand in question (tax, constitution, social structure...), its nature (abolish, suppress, maintain...), as well as a nuance (local demand, subject to a condition...). The database contains the totality of the known *cahiers de synthèse* of the Third Estate, that is 198, and of the Nobility, that is 166 (the Clergy is not included), and 748 parish *cahiers* from 46 *bailliages*. Given that this coding was done manually and that the parish *cahiers* are numerous, a sampling was carried out for these parish cahiers. Parish *cahiers* are supposed to be representative of the *cahiers* of the time. These data can be found through Degrave et al. (2023) work, who graciously imported the dataset in a comprehensive way and made it openly accessible online.

While these data are useful, they have several shortcomings. First, Clery *cahiers* were not taken into account both due to a lack of resources and a lack of interest since Nobility *cahiers* were thought to be more informative (Nobility is the first privileged Estate that comes to mind when thinking about the French Revolution) and too similar to Clergy *cahiers*. To date, no quantitative analysis has been carried out on Clergy *cahiers*. Such an analysis would be quite informative to compare Nobility and Clergy *cahiers*, and better understand the relationship between the Clergy and the Third Estate. Secondly, again due to a lack of resources, it was impossible to use each parish *cahier* from all of France, so a sample was carried out. This does not allow us to conduct a very micro-level analysis. Finally, these data were human coded, so a human had to read every *cahier* to understand the content of each grievance. This raises some questions regarding possible bias. To complement the data of Shapiro and Markoff, we will carry out a computer-

centred analysis of *cahiers*, adding Clergy *cahiers* and parish *cahiers* from Normandy for micro-level analysis.

Data on violence come from the Historical Social Conflict Database (HiSCoD) (Chambru and Maneuvrier-Hervieu, 2022), which lists the location of violent events between the 12th century and the late 19th century. Violence types are encoded through broad categories (political violence, religious violence...), and then precisions are added through a quick summary of the event. The number of attendees and geocoded data are also available. The database is sizable containing 15 813 events in France. We only intend to use events from before the revolution (9,542 events in the database) and events during the revolution (2,143 events). Data are available online. We will use data regarding all types of social conflict, excluding only smuggling-related violence, which is more linked to criminality and is not the type of violence we are interested in.

3 Empirical Analysis

3.1 Variables

Revolution violence (1789-1799) is our dependent variable. We will first use a dummy variable equal to one if violence took place in the area, zero otherwise, and second a discrete variable (number of violent events in area i). Data on previous violence (from before the Estate-General of May 1789) will be introduced on the right-hand side to address reverse causality issues by controlling for pre-existing violence. Social capital is our main explanatory variable. Two methods will be used to create measures of social capital based on the contents of the *cahiers*: a Topic modelling method and a Sentiment analysis. In addition, we will use the Shapiro and Markoff database, which contains information on grievance contents.

The Topic modelling approach allows the classification of a text into different topics and subtopics. This can be done through unsupervised learning, where the program recognizes topics without help, and through supervised learning with human help. This way, we can classify each grievance into a topic. We will also run a Sentiment analysis to classify the tone of each grievance. However, it will not allow us to recreate the grievance content with the same precision as the database from Shapiro and Markoff regarding the

action to be done (for instance, a grievance regarding tax with an overall negative tone does not necessarily mean it demands an abolition of the tax). It is nonetheless helpful information for the analysis: we gain insight into Clergy *cahiers*, an analysis on all parish *cahiers* from Normandy and more importantly a more objective analysis of contents.

An index of "voice" for *bailliage* (showing how well represented a *bailliage* was) can be built using the distance to Versailles. An index of "voice", capturing how well-represented parishes were during the making of the *cahiers* can also be built. First, we can use disparities in the *bailliage* size. Indeed, some *bailliage* are small and others are sizable. Thus, the *cahier de synthèse* of a small *bailliage* might better represent all areas than the *cahier* of a bigger *bailliage* (voices of a parish will be drowned out among the numerous other parishes). We then create an index of representativeness which will be based on the size of the *bailliage* in question, its population, or the number of parishes in the *bailliage*.

However, with this index, all parishes from the same *bailliage* will be characterized by the same level of voice. A parish-specific index, which will be particularly useful for the parish-level analysis, can also be built using the distance to the nearest important town or the transport cost. The idea is that if it is difficult for a parish to bring its *cahier* back into town or to send representatives at the *bailliage* assemblies, it is more likely the parish was not well represented in the higher assemblies. We can also use the similarities of parish *cahier* to the corresponding *bailliage cahier* to see how much specific grievances from a parish were represented at a higher level. It will be measured using how many grievances were still mentioned in the corresponding *bailliage cahier*.

Several standard controls will be used such as income (Béaur and Minard, 1997), population or geography (Motte and Vouloir, 2007). Controlling for bad harvests using climate data (Luterbacher et al., 2004) (Pauling et al., 2006) or the price of grain (Béaur and Minard, 1997) is an important control to add, given the frequency of violence linked to a scarcity of food. "*Pays d'état*" (local autonomy), "*Pays d'élection*" (centralized power) and "*Pays d'imposition*" (very centralized power) are different mean of administrating a region that also should be taken into account in controls (Mousnier, 1974). In "*Pays d'état*", taxes were decided by a regional assembly of all Estates that were summoned frequently, whereas in "*Pays d'élection*" taxes were decided by a royal officer on-site. On the other hand, in "*Pays d'imposition*", taxes were directly decided by the King himself

with no royal officer on-site. The "*Cinq grosses fermes*" free trade zone should also be considered (Panckoucke, 1784). A measure of Protestantism for an area will also be taken into account (Béaur and Minard, 1997). The global tone of a *cahier*, such as a threatening tone or a soft tone, can be computed using sentiment analysis: they should be added in control. Industrialization of an area (Béaur and Minard, 1997), inheritance customs (Béaur and Minard, 1997), proximity to Paris, urbanization, height (in cm) of conscripts (as a mesure of welfare) (Komlos, 2003), mortality (Séguy, 2001), literacy (Clout, 2013) presence of a court in an area (Béaur and Minard, 1997) are all various controls we will take into account.

3.2 Specifications

First, we will study the *cahiers de synthèse* for all of France with a cross-section approach. Second, we will focus on a case study in Normandy to study its primary *cahiers* (or parish *cahiers*) as all its parish *cahiers* are available. The general equation to be estimated will be:

$$Y_{i,t} = \beta_0 + \beta_1 K_i + \beta_2 Y_{i,t-1} + \beta_3 V_i + \beta_4 V_i * K_i + \beta_5 X_i + \epsilon_i \tag{1}$$

Where i is the *bailliage* or the parish, t represents the Revolution period, and t-1 represents the period from before the Revolution. The dependent variable for area i $Y_{i,t}$ is Revolution violence, K_i is social capital for area i (bridging or bonding), $Y_{i,t-1}$ is pre Revolution violence for area i, V_i stands for "voice" for area i and X_i is a control vector for area i. It will allow us to compare areas with similar past violence but different levels of social capital to see whether they participated differently in revolutionary violence. We will start with a linear estimator. Then, depending on the nature of the outcome variable (dummy or not), a probit or a logit will be used, Poisson otherwise. This generic specification will be declined in three different fashions, depending on the level of analysis and coverage. We further detail each of these three empirical models.

3.2.1 *Bailliage* level and national coverage

First, we will do an analysis at the *bailliage* level (B for *bailliage*) with national coverage, such as:

$$Y_{B,t} = \beta_0 + \beta_1 K_B + \beta_2 Y_{B,t-1} + \beta_3 V_B + \beta_4 V_B * K_B + \beta_5 X_B + \epsilon_B \tag{2}$$

Where every variable is aggregated at *bailliage* level. The most adapted measure of voice in this case is the one that captures the *bailliage*'s representativeness as our outcome is at the *bailliage* level, hence the distance to Versailles. Given the relatively small number of *bailliages*, it is possible that this approach does not yield enough statistical power to detect a significant link between social capital and violence.

3.2.2 Parish level and national coverage

To alleviate this concern, we will do a second analysis at the parish level with national coverage, such as:

$$Y_{P,t} = \beta_0 + \beta_1 K_B + \beta_2 Y_{P,t-1} + \beta_3 V_P + \beta_4 V_P * K_B + \beta_5 X_P + \epsilon_P \tag{3}$$

$$Y_{P,t} = \beta_0 + \beta_1 K_B + \beta_2 Y_{P,t-1} + \beta_3 V_B + \beta_4 V_B * K_B + \beta_5 X_P + \epsilon_P \tag{4}$$

Where violence variables are at the parish level, but the social capital variable is still at the *bailliage* level. This shall give us more statistical power (each *bailliage* is akin to a treatment group) while allowing for national analysis.

Voice measures that will be used are only those regarding parish representativeness as our outcome is at the parish level, and can be both at the *bailliage* (for equation 4) or parish level (for equation 3). So we can use *bailliage* size/number of parishes in a *bailliage* in equation 4 and distance to the nearest important town, transport cost or similarities of parish *cahier* to the corresponding *bailliage cahier* in equation 3. For this parish-level analysis, standard errors will be clustered by *bailliage*.

3.2.3 Parish level and Normandy coverage

Finally, a third analysis at the parish level, covering Normandy only, will be carried out such as:

$$Y_{P,t} = \beta_0 + \beta_1 K_P + \beta_2 Y_{P,t-1} + \beta_3 V_P + \beta_4 V_P * K_P + \beta_5 X_P + \epsilon_P \quad (5)$$

Where all variables are at the parish level. Dummy variable for *bailliage* will be introduced (equal to 1 if inside the *bailliage*, 0 otherwise). Thus, voice measures that will be used are only those regarding parish representativeness and must only be at the parish level, so we can use distance to the nearest important town, bilateral transport cost or similarities of parish *cahier* to the corresponding *bailliage cahier*. For this parish-level analysis, standard errors will be clustered by *bailliage*.

Even though fragmentation of parish *cahiers* across France make it difficult to obtain national coverage, we can use all parish *cahiers* from Shapiro and Markoff (1998) to obtain a larger, albeit spotty coverage. This will allow us to check Normandy results while keeping data limitations in mind.

3.2.4 Complementary analyses

Political expression: An interesting addition would be to study how the exercise of political expression in itself shaped the link between social capital and violence. Political expression is different from how well a parish was represented, rather we are interested in how the making of a *cahier* in a community affected violence intensity. As all the *cahiers* were not written at the same time, the idea is to take the year 1789 and compare violence intensity in parishes that had already expressed themselves with others that did not in a panel analysis. Heterogenous effect due to social capital type and level will be studied to understand better which mechanisms of social capital are related to violence. A specification such as this one will be used[2]:

$$Y_{P,t} = \beta_0 + \beta_1 E_{B,t} + \beta_2 E_{B,t} * K_B + \beta_3 X_{P,t} + \lambda_P + \lambda_t + \epsilon_{P,t} \quad (6)$$

For a parish level analysis with a national coverage with parish and time fixed effect.

[2]Pre-revolutionary violence to account for past violence should be captured by fixed effect. However, we will test if adding it considerably change results. To do this we will take into account violence before political expression (one week for instance).

Where p is a parish and t is a day. Thus, an observation is a parish at time t. Here political expression variable $E_{B,t}$ (equal to 1 once the *cahier* of the area is written, 0 otherwise) is at the *bailliage* level as social capital (captured by the writing of the *cahiers*) is at the *bailliage* level. An event study specification will also be used:

$$Y_{P,t} = \beta_0 + \sum_{j>j'} \gamma^j E^j_{B,t} + \sum_{j>j'} \gamma^j E^j_{B,t} * K_B + \beta_3 X_{P,t} + \lambda_P + \lambda_t + \epsilon_{P,t} \tag{7}$$

Where j are points in time used to measure the effects of the writing of a *cahier* as they manifest over time. Social capital does not change in time as it is considered roughly constant during the year and was only measured through the *cahiers*[3]. Standard errors will be clustered by parish and *bailliage* to account for serial autocorrelation. The following specification will also be used:

$$Y_{P,t} = \beta_0 + \beta_1 E_{P,t} + \beta_2 E_{P,t} * K_P + \beta_3 X_{P,t} + \lambda_P + \lambda_t + \epsilon_{P,t} \tag{8}$$

And through an event study specification:

$$Y_{P,t} = \beta_0 + \sum_{j>j'} \gamma^j E^j_{P,t} + \sum_{j>j'} \gamma^j E^j_{P,t} * K_P + \beta_3 X_{P,t} + \lambda_P + \lambda_t + \epsilon_{P,t} \tag{9}$$

For a parish-level analysis for Normandy with parish and time-fixed effect. Standard errors will be clustered by parish to account for serial autocorrelation.

CGF regression in discontinuity: Another idea would be to take advantage of the *Cinq Grosses Fermes*. At the time, there was an internal free trade zone known as the *Cinq Grosses Fermes* (CGF). The idea was to bolster supply to Paris by making trade easier in those areas. Johnson (2015) shows that regions just inside those areas were more likely than regions just outside to identify themselves with national, as opposed to local institutions. This hints at differences in social capital inside those areas. We can do a simple regression in discontinuity at the border of the *Cinq Grosses Fermes* to see if there are any differences regarding violence just inside the CGF and outside.

[3]Social capital is then captured by our fixed effects. We only need an interaction term with social capital.

Validity of the *cahiers*: An important concern regarding the *cahiers* is their reliability, and especially to what extend they represent the wishes of common people and not just the elite preferences. This is central to the analysis of Shapiro et al. (1998).

Comparing similarities between Third Estate *cahiers* and Clergy *cahiers* would be an important analysis to carry out to make sure that *cahiers* grievances were not driven by religious elites. In the same manner, comparing *cahiers* near Paris and far away ones could reveal a possible impact of centralized power on popular grievances. Those analyses would strengthen Shapiro and Markoff's claim that the Third Estate *cahiers* are representative of the common people and not driven by elites.

Climate and social capital: Differences in social capital type are likely to affect how an area can withstand a shock. Climate shocks are known to exacerbate acts of violence during the French Revolution (Waldinger, 2024), but little is known regarding how social capital mitigates or exacerbates a shock impact on violence. In the literature on the effect of climate shocks on violence, ethnicity-related variables were found to have an effect (Almer et al., 2017; Harari and Ferrara, 2018). Ethnicity is likely to be linked to different level of social capital and different social capital type. It would be interesting to add an interaction between climate shock and social capital in our baseline specification to specifically study this.

References

Almer, C., Laurent-Lucchetti, J., and Oechslin, M. (2017). Water scarcity and rioting: Disaggregated evidence from sub-saharan africa. *Journal of Environmental Economics and Management*, 86:193–209.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., and Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives*, 30(3):249–274.

Bellows, J. and Miguel, E. (2009). War and local collective action in sierra leone. *Journal of public Economics*, 93(11-12):1144–1157.

Blattman, C. (2009). From violence to voting: War and political participation in uganda. *American political Science review*, 103(2):231–247.

Béaur, G. and Minard, P. (1997). *Atlas de la Révolution française*. Paris, Éditions de l'École des Hautes Études en Sciences Sociales.

Chambru, C. and Maneuvrier-Hervieu, P. (2022). Introducing hiscod: a new gateway for the study of historical social conflict. *American Political Science Review*, pages 1–8.

Clout, H. D. (2013). *Themes in the historical geography of France*. Elsevier.

Degrave, A. et al. (2023). Local rule, elites, and popular grievances: Evidence from ancien régime france. *Journal of Historical Political Economy*, 3(1):1–29.

Fafchamps, M. (2006). Development and social capital. *The Journal of Development Studies*, 42(7):1180–1198.

Glaeser, E. L., Laibson, D., Scheinkman, J. A., and Soutter, C. L. (1999). What is social capital? the determinants of trust and trustworthiness.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.

Grosjean, P. (2014). Conflict and social and political preferences: Evidence from world war ii and civil conflict in 35 european countries. *Comparative Economic Studies*, 56:424–451.

Harari, M. and Ferrara, E. L. (2018). Conflict, climate, and cells: a disaggregated analysis. *Review of Economics and Statistics*, 100(4):594–608.

Hyslop, B. F. (1934). French nationalism in 1789, according to the general cahiers. *(No Title)*.

Jones, S. G. and Munoz, A. (2010). *Afghanistan's Local War: Building Local Defense Forces*. RAND National Defence Research Institute.

Komlos, J. (2003). An anthropometric history of early-modern france. *European Review of Economic History*, 7(2):159–189.

Laurent, E. and Mavidal, J. (1879). *Archives Parlementaires de 1787 à 1860: Etats généraux, Cahiers des sénéchaussées et bailliage*. Paris : Librairie Administrative P. Dupont.

Lefebvre, G. (1983). *The Great Fear of 1789: rural panic in revolutionary France*. Princeton University Press.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H. (2004). European seasonal and annual temperature variability, trends, and extremes since 1500. *Science*, 303(5663):1499–1503.

Maneuvrier-Hervieu, P. (2021). La normandie dans l'économie atlantique au xviii e siècle. production, commerce et crises. In *Annales historiques de la Révolution française*, number 4, pages 203–214. Cairn/Publilog.

Martin, P., Mayer, T., and Thoenig, M. (2008). Make trade not war? *The Review of Economic Studies*, 75(3):865–900.

Merriman, J. (2009). *A history of modern Europe: from the Renaissance to the present*, volume 1. WW Norton & Company.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review*, 95(3):796–816.

Motte, C. and Vouloir, M.-C. (2007). Le site cassini. ehess. fr: Un instrument d'observation pour une analyse du peuplement. *Bulletin du Comité français de cartographie*, 191:68–84.

Mousnier, R. (1974). *Les institutions de la France sous la monarchie absolue, 1598-1789*. PUF, Paris.

Paldam, M. (2000). Social capital: one or many? definition and measurement. *Journal of economic surveys*, 14(5):629–653.

Panckoucke, C.-J. (1784). *Encyclopédie méthodique: Finances...*, volume 1. chez Panckouke.

Pauling, A., Luterbacher, J., Casty, C., and Wanner, H. (2006). Five hundred years of gridded high-resolution precipitation reconstructions over europe and the connection to large-scale circulation. *Climate dynamics*, 26:387–405.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and schuster.

Rohner, D., Thoenig, M., and Zilibotti, F. (2013). War signals: A theory of trade, trust, and conflict. *Review of Economic Studies*, 80(3):1114–1147.

Satyanath, S., Voigtländer, N., and Voth, H.-J. (2017). Bowling for fascism: Social capital and the rise of the nazi party. *Journal of Political Economy*, 125(2):478–526.

Sawyer, A. (2005). *Social capital, survival strategies, and their potential for post-conflict governance in Liberia*. Number 2005/15. WIDER Research Paper.

Schama, S. (2004). *Citizens: A chronicle of the French Revolution*. Penguin UK.

Séguy, I. (2001). *La population de la France de 1670 à 1829: l'enquête Louis Henry et ses données*. Ined.

Shapiro, G., Markoff, J., Tackett, T., and Dawson, P. (1998). Revolutionary demands: A content analysis of the cahiers de doléances of 1789. *(No Title)*.

Waldinger, M. (2024). "let them eat cake": drought, peasant uprisings, and demand for institutional change in the french revolution. *Journal of Economic Growth*, 29(1):41–77.

4 Appendix

Variable type	*Social capital type*	*Source*	*How ?*
Directly observable behaviors : Willingness to participate in society			
Care put into the making of the cahiers			
Total length of the cahier	No nuance	My work (so with synthèse + parish + all estates)	Counting number of words of the cahier. Possible as we have the written cahier.
Number of grievances	No nuance	My work + S&M (synthèse only + no clergy)	Counting number of articles in the text using words signaling for start of an article ("Art.", "article" or "[number]°"). Or through S&M: a grievance is one obs.
Level of details in expressed grievances	No nuance	S&M	Using "SPEC" code
Content of the grievances			
Local centred/national centred grievances	Bonding (local) and Bridging (national)	My work + S&M	Using topic modelling, it will be possible to classify grievances depending on their topics as some words linked to each topic will be more likely to appear. So, we can identify local vs centred grievances. Using S&M, grievances are very detailed, so we can distinguish which grievances are about local vs national issues.
Grievances demanding more representation	Bridging	My work + S&M	Same idea
Trust			
Disagreement or agreement between Nobility/Clergy cahiers and Third Estate cahiers			
Agenda disagreement (if something should be considered a problem)	Bridging	My work + S&M	First using Topic modelling, it will be possible to find if a certain subtopic is mentioned in a third estate cahier and not in a nobility cahier for instance. Using S&M, detail on grievance content is available so I can retrieve differences in grievances between estates.
Program disagreement (how to solve the problem)	Bridging	My work + S&M	Possibly more difficult to find in topic modeling, too precise to find. Can be found in S&M
Tone of Third Estate cahiers regarding:			
Nobility	Bridging	My work	Sentiment analysis in sentences talking about nobles
Clergy	Bridging	My work	Same idea
Government/King	Bridging	My work	Same idea
Local communities	Bridging	My work	Same idea