The Yeoman's Portfolio: Measuring Historical Risk Preferences Using Crop Choice*

Remy Levin
University of Connecticut

Daniela Vidart
University of Connecticut

March 26, 2026

Abstract

We design a new method for measuring the risk preferences of agents in the past using data on agricultural production decisions. We employ it to estimate risk preferences for contiguous U.S. counties decennially from 1890 to 1920. Our results reveal substantial cross-sectional heterogeneity in preferences centered at moderate risk aversion, and temporal variation in response to aggregate shocks. We show that measured risk aversion is negatively correlated with financial risk-taking including purchases of WWI bonds, is representative of the underlying household preference distribution, and that individuals in counties with higher exposure to the American frontier were more risk-taking.

Keywords: Risk Preferences, Historical Measurement, Agricultural Decision-Making, U.S. History 1890–1920

JEL Codes: D81, N51, N52, Q12

*Email: remy.levin@uconn.edu and daniela.vidart@uconn.edu. We would like to thank Leah Boustan, Katherine Hauck, David Lagakos, Omer Moav, Allen Peters, Steve Ross, and Wayne Sandholtz for their insightful feedback on this paper. We would also like to thank Price Fishback, Paul Rhode, and Ken Sylvester for kindly sharing data.

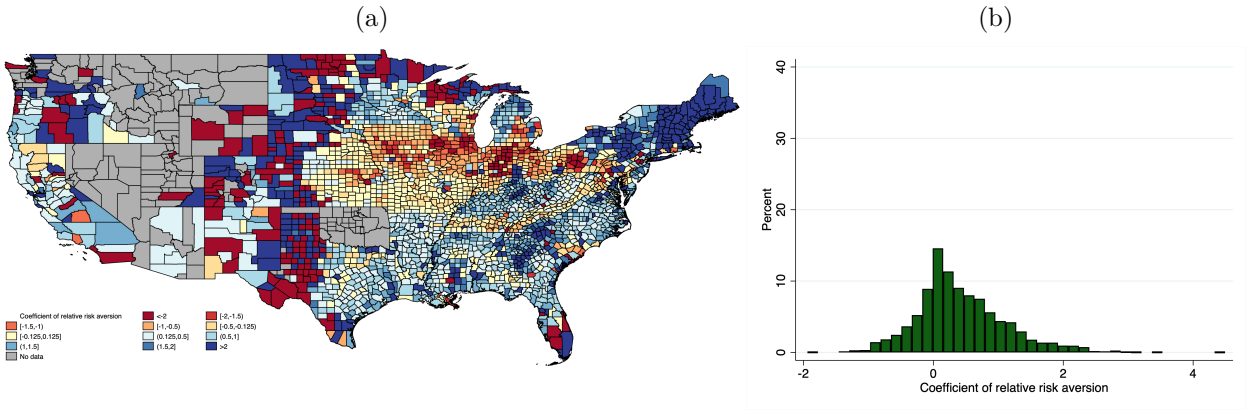
1 Introduction

Risk preferences are a key driver of economic decision-making, shaping investment behavior and responses to uncertainty in a broad range of domains. Understanding how these preferences vary across populations and over time is essential for explaining patterns of economic behavior and their evolution (Barsky et al. (1997), Malmendier and Nagel (2011), Herranz et al. (2015), Falk et al. (2016), Anderson and Mellor (2008), Jaeger et al. (2010), Levin and Vidart (2025)). While contemporary data have enabled economists to measure individual risk attitudes with increasing precision, evidence about historical variation in these preferences is practically nonexistent. This scarcity reflects a fundamental measurement problem: the most widely used methods for eliciting risk preferences—choices over incentivized or hypothetical lotteries—are modern innovations, with no analog in the historical record. Alternative measures based on observed financial behavior, such as stock market participation or portfolio allocations, are similarly limited to recent decades and to the small subset of the population with access to financial markets. As a result, we know very little about the distribution of risk preferences in the past, their geographic variation, and their responsiveness to major economic and environmental shocks.

In this paper we develop a new method for measuring the risk preferences of historical populations using agricultural production decisions. Agriculture offers a uniquely promising setting for this measurement. Farming has historically been the dominant form of economic activity, employing the majority of the global population well into the twentieth century. Production decisions in agriculture are inherently risky due to the volatility of yields and prices, and they are often undertaken without formal insurance or other mechanisms for smoothing shocks (Moschini and Hennessy (1999), Karlan et al. (2014), Cole et al. (2017)). Moreover, the basic elements of agricultural production (land allocation, crop yields, prices) as well as major sources of production risk like weather variation are often available in the historical record.

Our method centers on crop choice, a component of farming that has long been understood in agricultural and development economics to be akin to allocating a portfolio to safe and risky assets (Anderson et al. (1977), Fafchamps (1992), Rosenzweig and Binswanger (1993), Dercon (1996)). We start by building a parsimonious model of crop choice as a static portfolio allocation problem in which farmers maximize expected utility over consumption, face exogenous price and yield risk, and allocate land among different crops. The model yields a tractable estimating equation in which the coefficient of relative risk aversion can

Figure 1.1: Mean coefficient of relative risk aversion across counties, 1890–1920



Panel (a) displays the geographic distribution across counties of estimated mean relative risk aversion (RRA) from 1890 to 1920. Blue counties are risk averse; red are risk-seeking; and yellow are risk neutral. Panel (b) displays the cross-sectional distribution of mean RRA across counties, winsorized at the 5% level to increase legibility. Here a value of 0 corresponds to risk neutrality, and higher numbers imply higher measured risk aversion. Decennial estimates of RRA and time trends are available in Section 7.

be recovered from data on land shares, expected crop revenues, and farmer beliefs about the variance-covariance structure of these revenues. We operationalize this framework using U.S. census of agriculture data on acreage and yields for thirteen major crops from 1880 to 1920, a harmonized panel of state-level crop prices, and high-resolution historical climate data. We impute annual, county-level crop revenue histories between census years, and use their moments as empirical proxies for farmer beliefs, in line with evidence that agents' expectations are shaped by personal experience (Malmendier and Nagel (2011), Malmendier and Nagel (2016), Kuchler and Zafar (2019)).

Applying our method, we estimate county-level risk aversion for approximately 93% of contiguous U.S. counties decennially from 1890 to 1920. The resulting cross-sectional distribution of estimated preferences has a mean reflecting moderate risk aversion, is positively skewed, and exhibits substantial geographic variation (Figure 1.1). We find considerable temporal variation in risk preferences, with risk aversion rising nationally during the lean years in the agricultural sector around the turn of the twentieth century, and declining during the boom of the 1910s. We also document robust correlations between our measure and local economic conditions that align with decision-theoretic predictions: counties with greater agricultural wealth and income were more risk-seeking, while those with smaller farms and a less developed agricultural sector were more risk averse.

We complement the main analysis with four additional analyses aimed at testing our mea-

sure's behavioral relevance, assessing its representativeness, and studying its relationship with contemporaneous historical trends. Employing the census of agriculture in 1910 and 1920, we show that farmers in more risk-seeking counties took on more mortgage debt throughout the teens to participate in the speculative frenzy that gripped U.S. farming during World War I (Rajan and Ramcharan (2015)). Using data from Hilt and Rahn (2020) and Brunet et al. (2025) we show that more risk-seeking counties had higher subscription rates in 1919 to WWI Liberty Bonds, indicating that our measure also captures relevant investment behavior outside the agricultural domain. Using a rare farm-level panel dataset from Kansas (Sylvester and Rupley (2012)) we examine the relationship between our estimates at both the farm and township levels, and conclude that our county-level measure is likely to represent the median of the underlying farm-level risk preference distribution. Finally, we show that counties with greater historical exposure to the American frontier were more risk-seeking, consistent with the view that frontier conditions selected for and reinforced individual traits favoring risk-taking (Bazzi et al. (2020)).

Our contribution is threefold. First, we introduce a method for inferring risk preferences from historical production data that can be applied in a wide range of settings where agricultural records survive. Second, we provide, to our knowledge, the first large-scale map of the spatial distribution of risk preferences in historical America, together with evidence on their evolution through major macroeconomic and sector-specific shocks. Third, we demonstrate that these preferences are meaningfully related to both contemporaneous financial events and long-run historical experiences, offering a bridge between micro-level economic theory and macroeconomic and historical analyses of behavior under uncertainty.

The remainder of the paper is organized as follows. Section 2 presents the conceptual framework motivating our focus on crop choice as a window into historical risk preferences. Section 3 introduces the model and derives the estimating equation. Section 4 describes the U.S. agricultural setting from 1880 to 1920. Section 5 details our data sources. Section 6 outlines our methodology for imputing yield histories, constructing belief measures, and estimating the risk aversion parameter. We present the results of our main analysis in Section 7 and of the additional analyses in Section 8. Section 9 concludes.

2 Conceptual framework

Consider the problem of measuring the risk preferences of individuals in the deep past.[1] A natural starting point is to adopt the methods used today to measure risk aversion. The current gold standard for measuring risk preferences over income, used in experimental settings, is observing multiple choices between objective income lotteries. Measures of this nature have been deployed in field settings with some success, but only since the 1990s, which limits their utility for our purposes.[2] The next best candidate is data on risky investment decisions, like stock market participation, or portfolio allocations between stocks and bonds (Malmendier and Nagel (2011)). These measures of behavior are more widely available than experimental measures, and are tightly linked to well-established theories of risk-taking. However, they also suffer significant limitations. Data on financial choices is only available widely for the last few decades, and, more importantly, is only available for a highly selected sample: the small share of individuals in advanced economies that had sufficient wealth to invest in financial instruments.

The ideal data for our purposes would involve choices over a highly risky income process, taken by a large share of the population, which are observable (at least in principle) in the historical record. Agriculture is the natural candidate. Agricultural production is inherently risky due to a variety of factors, most notably the vagaries of the weather (Moschini and Hennessy (1999)). Since the Neolithic Revolution, the vast majority of humanity has been employed in agriculture, a situation that has persisted until quite recently. And since agriculture has been so central to the economy for thousands of years, data on it often exists in material, narrative, and environmental records.

The next step is to determine which aspect of the agricultural production process to focus on. Farming is complex and involves many decisions: which land to cultivate, when to plant, what to plant, when to harvest, which inputs to use, and more. Of these, the most promising

[1] For the purposes of this paper, the *deep past* refers to the time period that extends from that studied by economic historians—decades to centuries ago—all the way back to the neolithic revolution and the advent of agriculture 12,000 years ago.

[2] A classic implementation of this method is to ask an agent for a choice between, say, $100 for sure, and a 50-50 gamble over $50 and $200 dollars. This choice is repeated with varying amounts or odds of the lotteries, which allows researchers to pin down the risk aversion coefficient of the agent (Holt and Laury (2002)). There are many variants of this method (e.g. Gneezy and Potters (1997), Eckel and Grossman (2008)), but all involve choices between income lotteries at their core. Instruments of this nature, typically non-incentivized, have also been deployed successfully in field settings, in both developed and developing countries (Falk et al. (2016), Levin and Vidart (2025)), with earlier examples in the U.S. Health and Retirement survey and the NLSY.

candidate for our purposes is the choice of which crop to plant. Crop choice is a universal feature of agriculture, regardless of the time period or technology, and is well understood as a risky decision. A large literature in agricultural and development economics models crop choice as a portfolio allocation problem, with land as the stock of wealth to be allocated and crops as risky assets that can be "purchased" by planting (Anderson et al. (1977), Dercon (1996)). Finally, data on crop choices is more likely to survive in the historical record. Timing decisions involve *when* to grow, while input and technological decisions involve *how* to grow, both difficult to observe in data-poor environments. Crop choice, on the other hand, requires information on *what* is grown, which is much easier to observe.

Since our ultimate goal is to measure agents' risk preferences, we use a structural model to recover the behavioral parameter of interest from the crop choice data. The guiding principle we employ in designing this model is parsimony. Since historical settings are data-poor, the method must work with the kinds of data that are most likely to be preserved in the historical record, and with as little data as possible. The model we present in the next section achieves this by boiling down the portfolio allocation problem to one that requires only data on cropland shares, prices, and yields for estimation.

3 Model

In this section we develop a model of crop choice as a standard portfolio allocation problem faced by the farmer, designed for structurally estimating risk aversion from crop choices. We make the following assumptions about the production process:

1. **The farmer is hand-to-mouth.** Therefore consumption is equal to income in every period. This also implies that the farmer lacks an effective savings technology.[3]
2. **The crop choice decision is period-by-period static.** We abstract from dynamic considerations, such as crop rotation, fallowing, and storage.
3. **The only sources of risk facing the farmer are exogenous price risk and**

[3]Under this assumption the model can be applied to both pure commercial agriculture and pure subsistence agriculture, treating income and consumption as measured in calories in the latter case. The case where the production process involves a mix of subsistence and commercial outputs is more challenging. As Fafchamps (1992) shows, when the farmer is both a consumer and seller of his agricultural products, the simplest structural model of crop choice requires expenditure shares and income elasticities of substitution for each crop, in addition to the usual portfolio allocation objects (risk aversion, expected revenues, and their expected variance-covariance matrix). Consumption data of that kind, as opposed to production data, is not widely available in the historical record before the 1930s.

yield risk. We abstract away from other sources of risk the farmer might face, such as technological risk and policy risk.[4]

4. **All crops have constant factor ratios.** That is, other factors of production such as labor and capital are used in the same proportions across crops. This allows us to separate the land allocation problem from the labor and capital allocation problem and model it independently.

5. **Returns to scale in land across crops are constant.** This implies that diversification of land-shares is driven solely by hedging considerations.

We assume that the agent is an expected utility maximizer with twice-continuously differentiable utility and subjective beliefs about the means and variance-covariance matrix of prices and yields (and therefore revenue). Given these assumptions, the model follows a standard asset portfolio decision problem with uncertainty over both the prices and yields of crops, where the farmer aims to maximize the utility derived from crop production:

$$max_{c,\{l_i\}_{i=1}^N} \mathbb{E}\left[U(c)\right] \quad \text{s.t} \quad c = \sum_{i=1}^{N} p_i q_i l_i \bar{L} \quad \text{and} \quad \sum_{i=1}^{N} l_i = 1. \tag{1}$$

c denotes consumption, N is the total number of crops available for the farmer to plant, and $\bar{L}$ is the total amount of land that can be planted by the farmer. l_i denotes the share of the farmer's land dedicated to crop i, while p_i and q_i capture the crop's price and per-acre yield, respectively.

In order to solve the farmer's problem in Equation (1), we set a given crop as numeraire so that the revenues of all crops are now relative to those of this crop. Taking first order conditions and applying standard linearization techniques (see Appendix A for details), yields the following estimating equation that we can bring to the data:

$$\mathbb{E}(p_i q_i) - 1 = \frac{-U''(\mathbb{E}(c))}{U'(\mathbb{E}(c))} \mathbb{E}(c) \frac{\sum_{j=1}^{N} Cov(p_j q_j, p_i q_i) l_j}{\sum_{j=1}^{N} \mathbb{E}(p_j q_j) l_j}. \tag{2}$$

From this equation we can see that the farmer's coefficient of relative risk aversion, $\frac{-U''(\mathbb{E}(c))}{U'(\mathbb{E}(c))}\mathbb{E}(c)$, can be recovered using only data on cropland allocations and expectations of per-acre yields

[4] See Moschini and Hennessy (1999) for a review of the sources of risk in agriculture, encompassing price risk, yield risk, technological risk (the risk of stranded capital due to technological change) and policy risk (the risk arising from changes in governmental policy). Our empirical setting is one in which technological risk and policy risk are likely to be modest.

and crop prices. Note that this does not require any functional form assumptions about the farmer's utility function, since we never have to compute moments of U directly.

Next, we turn our attention to using this result to measure risk preferences empirically, starting with choosing an amenable historical setting for the analysis.

4 Historical setting: U.S. agriculture 1880-1920

Our theoretical result entails a basic data requirement for the historical setting to which we apply our method: there must be data available on cropland shares, per-acre yields, and agricultural prices.

Several practical considerations also apply. First the setting must feature some diversity of crops grown, since estimating a portfolio allocation decision benefits from observing multiple asset allocations. Second, prices and yields should reflect the true income risk faced by farmers, which might not be the case when crop insurance, governmental subsidies, or other agricultural market interventions are present. Third, farming should be mostly localized, with farmers owning and working land in a single location. Because comprehensive land ownership data is unlikely to exist in the historical record, the widespread presence of large absentee land-owners that can hedge risk across far-flung locations is a threat to identification when using location-specific outcomes. Finally, we would ideally like a setting in which a substantial share of the working population is employed in agriculture, because that would mean that our method would directly measure the risk aversion of a large share of the population.

Given these criteria, we focus on the United States around the turn of the twentieth century, from 1880 to 1920. Rich nationwide agricultural data exists for this period thanks to the decennial U.S. agricultural census. Most importantly for our purposes, starting in 1880, the agricultural census records both acres planted in various crops and per-acre yields. The census records these values for a broad range of crops, covering a substantial share of the farmland and agricultural output of the country.

Other features of the U.S. agricultural sector in this period are also conducive to our analysis. Government intervention in the sector was relatively minimal before the Great Depression, and especially before the 1920s.[5] Agricultural production was largely small-scale and local

[5]Before the 1920s, the federal government's involvement in agriculture was largely indirect, through land grants, infrastructure construction, tariffs, trade policy, and monetary policy. Direct intervention began in

in this period, with a large share of the labor force employed in agriculture, and a majority of the population living in rural areas.[6]

From a technological perspective, our period of study was characterized by a brisk pace of structural change. Biological innovation in the form of new seed varieties expanded the range and productivity of crops such as wheat and corn, offering farmers more viable planting options across a broad swath of the American landscape. Mechanization proceeded first through the adoption of large horse-drawn combines in the late nineteenth century, and then through the invention and widespread adoption of the petroleum tractor after 1900 (Olmstead and Rhode (2018)).

From a macroeconomic perspective, the pre- and post-1900 periods were markedly different. The late nineteenth century was characterized by persistent deflation and a strong dollar following the demonetization of silver and the Panic of 1873. This hurt the profits of farmers (who were often debtors and exporters), and led to substantial political upheaval (Frieden (1997)). Inflation returned after 1900, and farm profits soared over the next two decades and especially during World War I, when national and global demand for agricultural products outstripped supply, driving substantial increases in commodity prices.

5 Data

Our empirical analysis requires four types of data: cropland shares by crop, per-acre yields for each crop, price data for each crop, and climate data for imputing yield histories. In this section we describe our sources for each of these data. In all cases, we restrict our analysis

1922 with the Capper-Volstead act, which granted anti-trust exemptions to farming cooperatives, followed by failed attempts at direct price support later in the decade (the McNary-Haugen Farm Relief Bills, vetoed by president Coolidge in 1927 and 1928) (U.S. Senate Committee on Agriculture (1998)). Direct federal intervention in the farm economy expanded massively during the Great Depression and the New Deal. However, since our period of study ends in 1920, we can generally consider our setting as free from most forms of direct government intervention in agriculture, including insurance, price supports, and collectivization.

[6] In 1880, 41.6% of the U.S. labor force was employed in agriculture, including 24.8% in operator or supervisory positions. These numbers equal 35.3% (20%) in 1900; 30.5% (16.6%) in 1910; and 24.9% (16.3%) in 1920 (Katz and Margo (2014)). In 1880, 71.2% of the U.S. population lived in rural areas (towns of under 2,500 inhabitants); 64.9% in 1890; 60.4% in 1900; 54.4% in 1910; and 48.8% in 1920 (Kim and Margo (2004)). Our measure of risk aversion (constructed from 1890-1920) can be interpreted as a direct measure of the risk preference of farmers making the crop choice decision, or as a proxy for the risk preferences of the broader rural (and perhaps total) population. By the narrowest interpretation, this would mean that we are measuring the risk preferences of approximately 1/6 to 1/4 of the population over the period, while the broader interpretation implies that we are measuring the preferences of 1/2 to 3/5 of the population.

to the contiguous United States.

5.1 The census of agriculture

Our data on total farmland, planted acreage, and crop yields come from the U.S. Census of Agriculture (hereafter Ag Census), a comprehensive farm-level national survey administered decennially by the Census Bureau from 1840 to 1920.[7] The census was conducted during the census year,[8] but output and acreage data referred to the previous calendar year (e.g. 1899 information was collected in the 1900 wave). Although the underlying data is at the farm level, this information is not readily available. Instead, we use the county-level data compiled by Haines et al. (2018).[9]

Between 1840 and 1870 the Ag Census collected only crop output data. Starting in the 1880 wave, it also collected acreage planted for a set of thirteen core crops which remain available and consistent over time: barley, buckwheat, corn, cotton, hay, oats, potatoes, rice, rye, sugar cane, sweet potatoes, tobacco, and wheat. We harmonize the survey information about these crops across waves.[10]

County-level data on total farmland acreage and acreage planted in each core crop is available directly in the Ag Census. On average, the core set of crops accounted for roughly 60% of all cultivated land in each wave and encompassed the majority of cultivated land in most counties, as shown in Figure 5.1.[11] We calculate per-acre yields by dividing the total yield reported for each crop in the Ag Census by its planted acreage. To derive the land portfolio share l of each core crop we compute

$$l_{i,c,t} = \frac{L_{i,c,t}}{\sum_{i=1}^{N} L_{i,c,t}}$$

where $L_{i,c,t}$ captures the total number of acres of crop i planted in county c in year t, and $\sum_{i=1}^{N} L_{i,c,t}$ is the total number of acres planted across all core crops.

[7] Starting in 1925, the Ag Census was conducted every five years. In 1997, responsibility for its collection was transferred from the Census Bureau to the USDA.

[8] The reference date for enumeration was typically June 1, except in 1910, when it was April 15, and in 1920, when it was January 1.

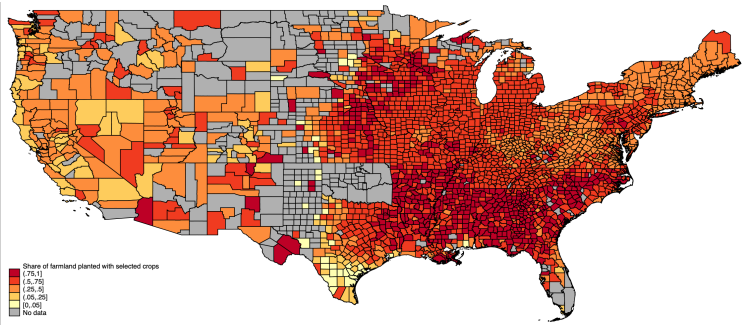
[9] Since county boundaries have changed throughout time in the U.S., we maintain consistent county definitions using 1900 boundaries, and link other years using the crosswalk built by Eckert et al. (2020).

[10] We visualize cultivation patterns over time in Appendix B.1 by mapping the core crop with the largest cropland share in each county and wave.

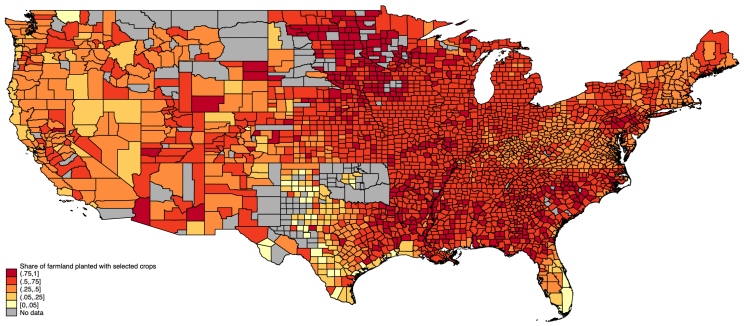
[11] Average cropland shares for each core crop in each wave across counties are reported in Appendix B.3.

Figure 5.1: Share of farmland planted with core crops in 1880–1920

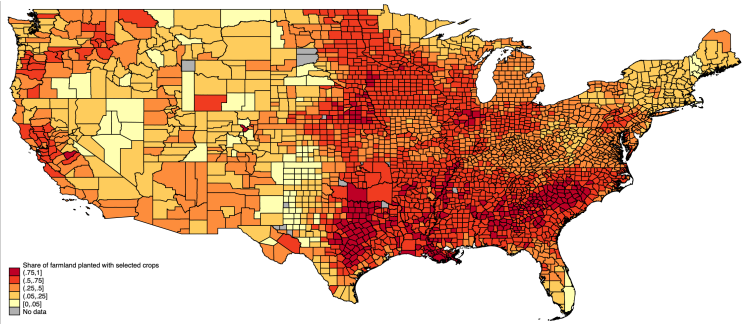
(a) 1880



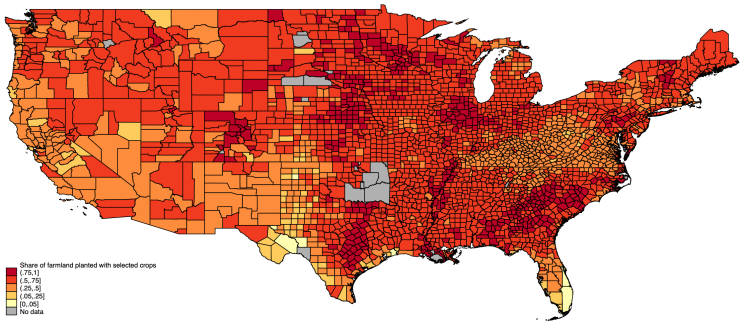
(b) 1890



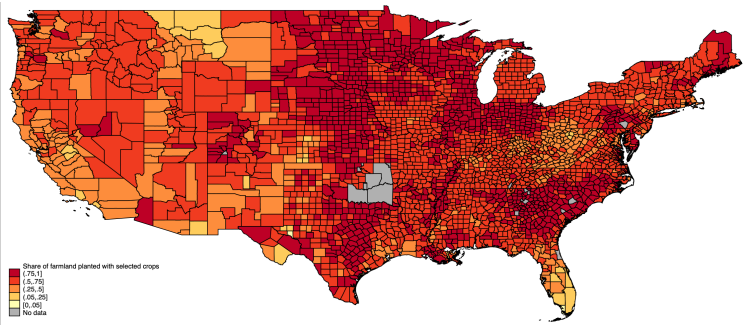
(c) 1900



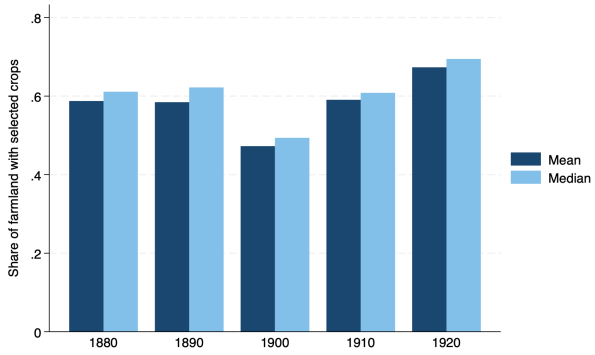
(d) 1910



(e) 1920



(f) Average county share of improved land in core crops



Notes: These maps plot the share of total farmland planted with the core set of crops in each county in 1880, 1890, 1900, 1910, and 1920; along with the corresponding aggregate mean and median in each year. Total farmland is measured as the number of acres of improved land held by farms in each county (see Appendix B.2 for a full definition of improved land).

We set corn as the numeraire for the purpose of the model's calculation, as it has the broadest geographic and temporal coverage of all crops, with 92%–99% of counties growing corn in 1880–1920.

5.2 Price data

We compile a panel of state-level annual prices for our core set of crops from 1880 to 1920 using several sources. Our main source is the state-level price histories from 1866–1930 constructed by (Cooley et al. (1977)). Because this dataset contains gaps for some crops, years, and states, we supplement it with two national-level datasets (US Bureau Of Labor Statistics (1927), US Census Bureau (1975)).[12]

We use two procedures to combine the datasets. First, in cases where the Cooley et al. (1977) data has a gap for a given crop in a state-year but includes prices immediately before and after, we compute the ratio between state-level and national-level prices for all crops with available state-level data that year, and impute the missing data by multiplying this ratio by the national price of the missing crop. Second, in cases where the Cooley et al. (1977) data coverage ends before 1920, we extend the series by applying the national-level percentage change in prices for subsequent years to the last available state-level price.

We deflate all prices in the harmonized panel using the national wholesale price index in each corresponding year, with a base year of 1913. Figure C.1 plots the resulting price time series for our set of crops, revealing substantial variation aross crops, time, and space.[13]

5.3 Climate data

Our source of historic climate data is the NOAA/CIRES/DOE 20th Century Reanalysis (V3) dataset (hereafter 20CRV3).[14] 20CRV3 is a gridded climate dataset that reconstructs records for dozens of climate variables based on historic measurements of atmospheric pressure and sea ice, combined with state-of-the-art climate models (Slivinski et al. (2019)). It is available as grid cells globally at the .7° by .7° spatial resolution for every 3-hour interval starting in 1806.

In our analysis we use the 6-hour distribution of the 20CRV3, aggregated up to the monthly level. We use nine variables from those available, aiming to capture key moments in the temperature and precipitation distributions: minimum temperature, maximum tempera-

[12] Full details of these datasets are in Table C.1.

[13] We provide summary statistics of the price time-series for each crop and decade in Table C.2.

[14] Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Earth System Research Laboratory Physical Sciences Laboratory. NOAA/CIRES/DOE 20th Century Reanalysis (V3) data provided by the NOAA PSL, Boulder, Colorado, USA, from their website at https://psl.noaa.gov.

ture, mean temperature, and standard deviation of temperature; as well as total monthly precipitation, maximum precipitation, minimum precipitation, mean precipitation, and the standard deviation of precipitation. We associate each county with a single grid cell using the minimum distance between county and grid cell centroids. On average each grid cell is assigned to 3.35 counties (4.2 east of the Mississippi, 2.4 west of the Mississippi).

6 Methodology

In addition to cropland shares, recovering the risk aversion parameter from Equation (2) requires measures of farmers' beliefs about crop revenues, which are not directly observable in the historical record. In this section, we describe our method for constructing credible empirical proxies for these expectations, and how we use them to estimate risk preferences. Our method involves three steps. First, we impute an annual county-level panel of per-acre crop yields using the yields observed in the Ag Census waves and higher-frequency climate data. Second, we use the yield panel and our price data to construct county-year beliefs about the means and covariances of per-acre crop revenues, assuming these beliefs are based on revenue experiences in each location over the previous decade. Finally, we combine these belief measures with cropland shares to estimate Equation (2) and recover risk aversion.

6.1 Imputing yield histories

We can construct per-acre yields for our core crops in Ag Census years, but not for the intervening years. To fill these gaps, we regress per-acre yields in census years on a set of climate variables, and then use climate observations in non-census years to predict per-acre yields.

We start by constructing a set of climate predictors of observed crop yields using our month-by-county panel of nine climate variables in the 20CRV3 dataset. For each census year, we generate a set of 24 monthly observations of each of the nine climate variable, encompassing the weather in the measured census year and the preceding year (e.g. for the 1900 census wave, which measures yields in 1899, we use weather realizations for every month in 1898 and 1899). This results in a set of 216 climate predictors for each crop in every county-year.

To avoid overfitting, we run a series of standard Lasso regressions to select a subset of these predictors for each crop. Since fixed effects will be included in the final prediction exercise, we also incorporate them in the Lasso regressions to ensure that the selected climate

variables have marginal predictive power above and beyond that of the fixed effects. Our unit of analysis is a crop-county-year. Since we include county fixed effects and thus must include multiple time periods, we run the regressions for each pair of successive census years (e.g. 1880–1890, 1890–1900, 1900–1910, etc.) This allows the climate variables and their estimated effects to vary across decades, accommodating structural changes that may alter the relationship between climate and yields. Furthermore, by estimating a separate regression for each crop, we capture heterogeneity in how different crops respond to climate variation, year and state idiosyncrasies, and county characteristics.

Formally, let n be a county-year that grows crop i in county c in state s in year $t \in t_{\text{pair}}$, where $t_{\text{pair}} \in \{1880-1890, 1890-1900, 1900-1910, 1910-1920\}$.[15] Let $q_n^{i,t_{\text{pair}}}$ be the per-acre yield of for crop i in county-year n for the two census years in t_{pair}. We then run Lasso regressions for each i in each t_{pair} by minimizing the following expression:

$$\min_{\{\gamma_C^{i,t_{\text{pair}}}\},\,\{\alpha_{all}^{i,t_{\text{pair}}}\}} \quad \frac{1}{N}\sum_{n=1}^{N}\left[q_n^{i,t_{\text{pair}}} - \alpha^{i,t_{\text{pair}}} - \sum_{k=1}^{9}\sum_{m=1}^{24}\gamma_C^{i,t_{\text{pair}}}C_{k,n,m} - \alpha_c^{i,t_{\text{pair}}} - \alpha_t^{i,t_{\text{pair}}} - \alpha_{s,t}^{i,t_{\text{pair}}}\right]^2 + \frac{\lambda^{i,t_{\text{pair}}}}{N}\sum_{k=1}^{9}\sum_{m=1}^{24}\left|\gamma_C^{i,t_{\text{pair}}}\right|,$$

where N is the cardinality of n, C denotes climate variables indexed by k with lags indexed by m, and $\alpha_{all}^{i,t_{\text{pair}}} = \{\alpha^{i,t_{\text{pair}}}, \alpha_c^{i,t_{\text{pair}}}, \alpha_t^{i,t_{\text{pair}}}, \alpha_{s,t}^{i,t_{\text{pair}}}\}$ is the set consisting of the regression constant plus county, year, and state-by-year fixed effects. The γ's are the regression coefficients for the climate variables and λ's are the Lasso penalty weights, chosen by minimizing the Extended Bayesian Information Criterion (EBIC).

After selecting the set of climate variables $\mathcal{S}^{i,t_{\text{pair}}}$ for each crop and year pair using the Lasso regressions, we then regress the per-acre yields for each crop in each county-year on its selected set of climate variables, along with county, year, and state-by-year fixed effects:

$$q_n^{i,t_{\text{pair}}} = \alpha^{i,t_{\text{pair}}} + \sum_{C_h \in \mathcal{S}^{i,t_{\text{pair}}}} \beta_h^{i,t_{\text{pair}}} C_h^{i,t_{\text{pair}}} + \alpha_c^{i,t_{\text{pair}}} + \alpha_t^{i,t_{\text{pair}}} + \alpha_{s,t}^{i,t_{\text{pair}}} + \varepsilon_n^{i,t_{\text{pair}}},$$

where h indexes the elements of $\mathcal{S}^{i,t_{\text{pair}}}$.

Finally, we use the estimated decadal β coefficients, interacted with the selected climate variables in each year to predict per-acre yields for the years between the decennial census

[15] Since not every county grows each crop in every year pair, we include only those n's for which this is the case.

waves. We also linearly interpolate the fixed effects for each prediction.[16] For example, to estimate per-acre yields for crop i in county c in 1883, we use the model estimated from the 1880-1890 census-year pair and compute:

$$y_{i,c,1883} = \alpha^{i,1880\text{–}1890} + \sum_{C_j \in \mathcal{S}^{i,1880\text{–}1890}} \beta_j^{i,1880\text{–}1890} C_j^{1883} + \alpha_c^{i,1880\text{–}1890} + \left(0.7 \times \alpha_{1880}^{i,1880\text{–}1890} + 0.3 \times \alpha_{1890}^{i,1880\text{–}1890}\right) + \left(0.7 \times \alpha_{s,1880}^{i,1880\text{–}1890} + 0.3 \times \alpha_{s,1890}^{i,1880\text{–}1890}\right)$$

Here, the weights 0.7 and 0.3 reflect the distance of 1883 from the 1880 and 1890 census years, respectively.

The final result of this analysis is an annual, county-level panel of per-acre crop yields for 1880 to 1920. We provide summary statistics for this time series for each crop and decade in Table D.1.

6.2 Constructing expectations

The next step in our analysis is to construct proxies for farmer beliefs about the means and variance-covariance matrix of crop revenues. We assume that farmers learn about these moments from their experiences with realized revenues (Malmendier and Nagel (2011), Levin and Vidart (2025)). Accordingly, we use the location-specific histories of these moments as empirical measures of the beliefs of farmers in these locations.

To construct annual, county-level revenue histories, we combine our per-acre yield panel with our price panel by taking the product of per-acre yield and price for each crop in every year and county. Using these histories we construct the perceived expected revenue for crop i in county c at time t, $\mathbb{E}(p_{i,c,t}q_{i,c,t})$, as the average revenue observed in c over the 10 years prior to t:

$$\mathbb{E}(p_{i,c,t}q_{i,c,t}) = \frac{1}{10}\sum_{s=1}^{10} p_{i,c,t-s}q_{i,c,t-s}.$$

Similarly, we construct the perceived covariance of revenue for each crop-pair $\{i, j\}$ in county c at time t, $Cov(p_{j,c,t}q_{j,c,t}, p_{i,c,t}q_{i,c,t})$, by computing the covariance of the revenues of crops i and j observed in c during the 10 years prior to t:

[16] Our results are robust to using an alternative interpolation method that uses the year and state-by-year fixed effects corresponding to the year *closest* to the year of interest. Results using this approach are available upon request.

$$Cov(p_{j,c,t}q_{j,c,t}, p_{i,c,t}q_{i,c,t}) = \frac{1}{10}\sum_{s=1}^{10} p_{j,c,t-s}q_{j,c,t-s}, p_{i,c,t-s}q_{i,c,t-s} - \left[\frac{1}{10}\sum_{s=1}^{10} p_{j,c,t-s}q_{j,c,t-s} \times \frac{1}{10}\sum_{s=1}^{10} p_{i,c,t-s}q_{i,c,t-s}\right].$$

Although we use the 1880 yields when constructing yield histories for 1880–1890, we lack the corresponding yield histories for 1870–1880. As a consequence, we cannot generate belief measures for 1880. Our expectations data thus covers 1890–1920, which is the time period used in our main analysis.

6.3 Empirical specification

With our expectation measures and cropland shares in hand, we can now recover the coefficient of relative risk aversion from our model. For crop i in county c and year t, let $X_{c,i,t} = \frac{\sum_{j=1}^{N} Cov(p_{j,c,t}q_{j,c,t}, p_{i,c,t}q_{i,c,t})l_{j,c,t}}{\sum_{j=1}^{N} \mathbb{E}(p_{j,c,t}q_{j,c,t})l_{j,c,t}}$ and $Y_{i,c,t} = \mathbb{E}(p_{i,c,t}q_{i,c,t}) - 1$. Define $\rho_{c,t}$ as the coefficient of relative risk aversion. Then Equation (2) can be expressed as:

$$Y_{i,c,t} = \rho_{c,t}X_{i,c,t} + \epsilon_{i,c,t}. \tag{3}$$

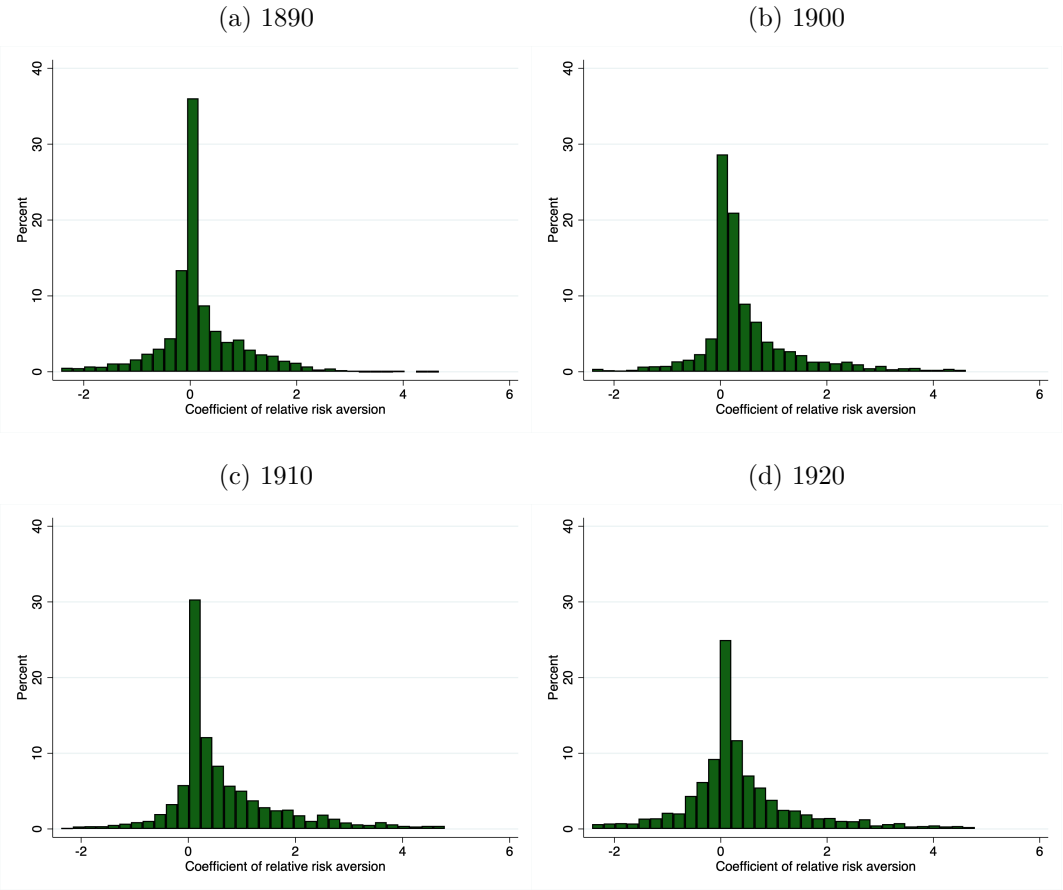
We estimate $\rho_{c,t}$ through OLS using data for all crops grown in county c in year t.

7 Main results

Using our method, we estimate the coefficient of relative risk aversion (RRA) for counties in the contiguous U.S. decennially from 1890 to 1920. In total, we are able to estimate RRA for 2,551 counties in 1890 (93.2% of all counties); 2,600 (92.7%) in 1900; 2,608 (93.2%) in 1910; and 2,622 (93.6%) in 1920. These counties comprise almost all counties across 40 states, but do not include (for the most part) counties in Arizona, Idaho, Montana, Nevada, Oklahoma, Utah, Washington, and Wyoming.[17]

[17] County-years are excluded from the analysis for one or more of five reasons: 1. they do not grow corn (the numeraire crop), or have missing information for corn; 2. they grow only one crop (or only one crop plus corn), which makes estimating Equation (3) impossible; 3. they have missing yield data for the crops they grow; 4. they have missing climatic information in all of the 10 years prior to the Ag Census wave; 5. they have missing price information for all crops in a given year. Of these, the most common issue with the western states missing form the analysis is the absence of corn cultivation.

Figure 7.1: Relative risk aversion coefficients across counties, 1890–1920

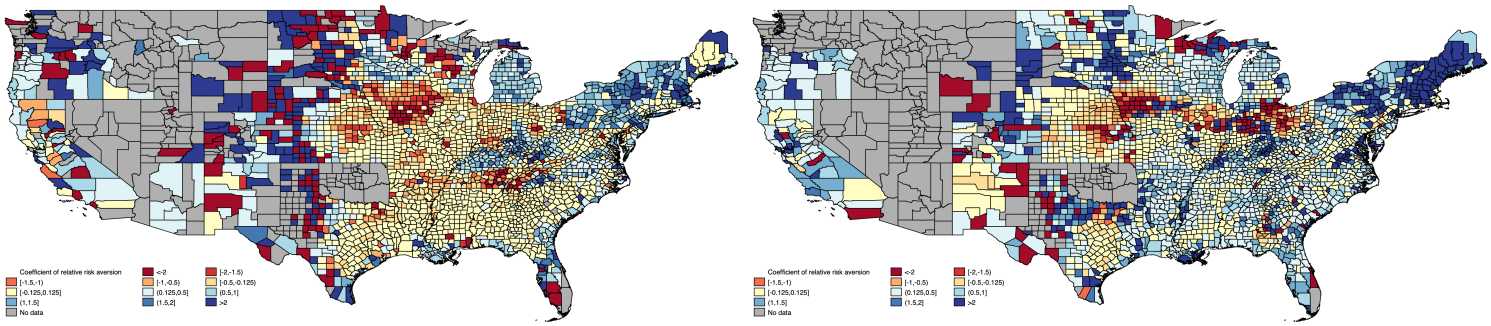


Notes: These figures plot the histograms of the relative risk aversion coefficients across all counties in 1890–1920. A value of 0 corresponds to risk neutrality, and higher numbers imply higher measured risk aversion. The distributions are trimmed at the 5% level to increase legibility.

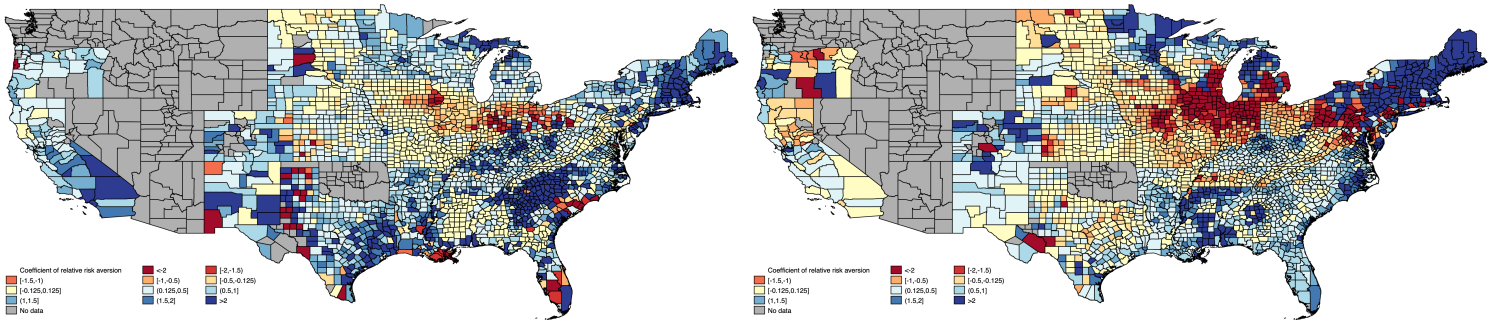
We begin by presenting the cross-sectional distribution of RRA across counties and years (Figure 7.1). Several patterns emerge. First, the distributions are unimodal, with positive means reflecting moderate risk aversion. Second, there is substantial variation in both directions around the mean, with some counties estimated as risk-taking, but with skew in each wave towards risk aversion. Third, the vast majority of RRA estimates lie between 2 (a common benchmark for risk-averse agents in macroeconomic models), and -2 (its risk-seeking counterpart). Overall, these results suggest that our estimates are at first blush reasonable, with the measured preference distribution being (skew-)normal, centered around moderate levels of risk aversion, and with values similar to modern estimates of risk aversion. The results also indicate that there is substantial variation in risk aversion across counties and years.

Figure 7.2: Relative risk aversion coefficients across counties, 1890–1920

(a) 1890 (b) 1900



(c) 1910 (d) 1920



Notes: These maps plot the RRA coefficient values in each county decennially from 1890–1920. Yellow to red shades correspond to risk-seeking preferences, and light blue to navy shades to risk aversion.

Next we examine the variation in risk aversion across space. Figure 7.2 maps the estimated RRA across counties for each census wave. Several notable patterns emerge. First, there is substantial fine-grained variation in RRA: many counties have neighbors with sharply different RRA estimates. At the same time, there are clear regional differences in risk aversion: the Northeast is the most consistently risk averse region; the Midwest is the most consistently risk-seeking region; and the south tends to be somewhere in between, with many counties approximately risk-neutral, except in the 1910 wave, where a cluster of counties in the traditional cotton belt are more risk averse than the nation as a whole. Finally, counties on the Great Plains are strikingly polarized in their risk preferences, forming a mosaic of the most risk-seeking and most risk averse areas.[18]

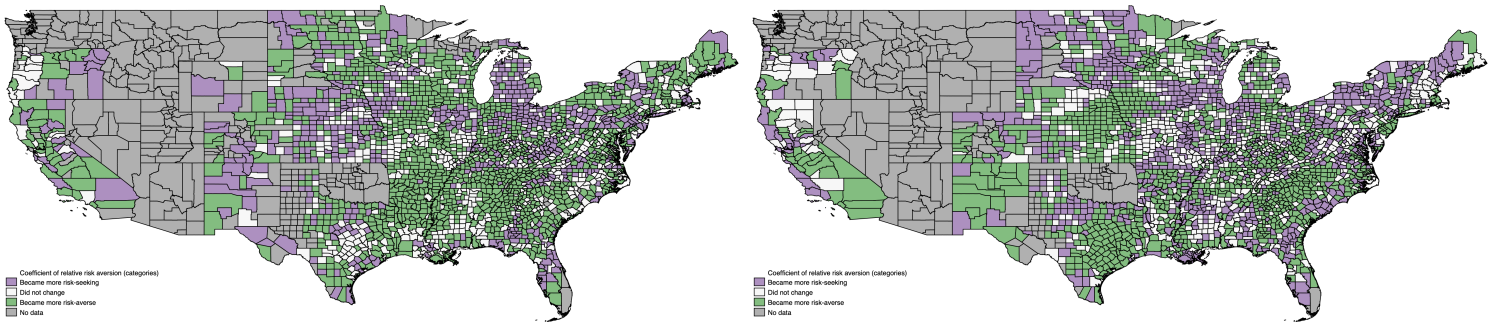
The cross-sectional patterns above suggest substantial temporal variation in RRA as well. We illustrate this variation in Figure 7.3, which maps the decadal change in RRA across

[18] We return to this latter pattern in Section 8.4, where we examine the relationship between risk aversion and exposure to the frontier.

Figure 7.3: Change in relative risk aversion coefficients across counties, 1890–1920

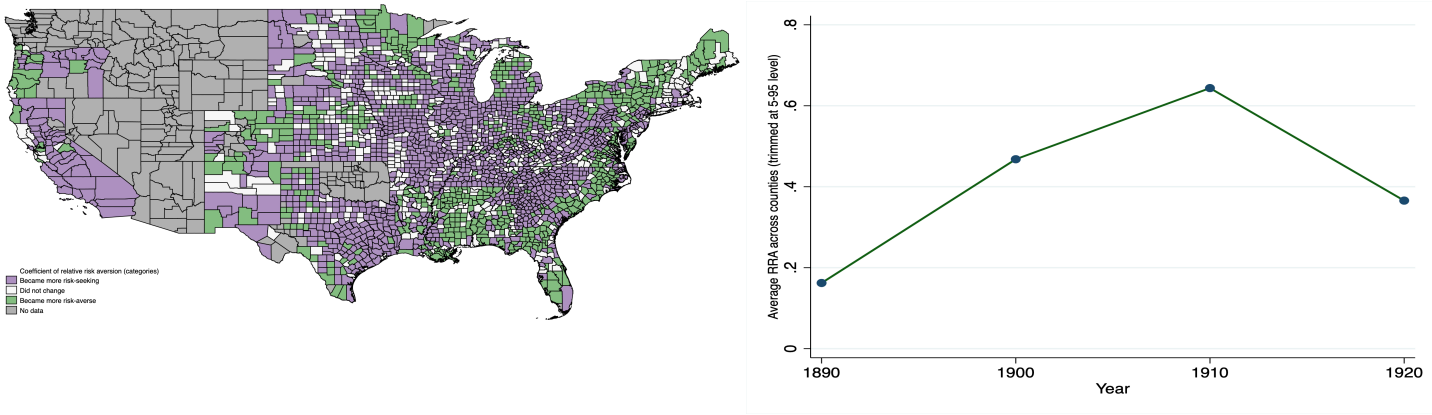
(a) Change in 1890–1900

(b) Change in 1900–1910



(c) Change in 1910–1920

(d) Change in county mean, 1910–1920



Notes: These maps plot the changes in the RRA coefficient values in 1890–1900, 1900–1910, and 1910–1920 in each county, and on average across the waves.

counties and in the national average. These maps reveal significant heterogeneity in the direction and magnitude of changes in risk aversion across counties. In addition, the national trend shows that from 1890 to 1910, farmers collectively became more risk averse, perhaps as a consequence of the same economic hardships that inspired the agrarian populist movement of the late nineteenth century (Goodwyn (1978)). Conversely, from 1910 to 1920, average risk aversion declined, possibly driven by the sharp increase in agricultural commodity prices and the accompanying rise in speculative activity in the agricultural sector (Rajan and Ramcharan (2015)).

7.1 Correlation of risk aversion with observed characteristics

We next examine the correlations between our risk aversion measure and other observable county-level characteristics. Our goal is two-fold. First, for covariates with a theoretically

strong relationship with risk aversion (such as wealth or income), correlations can provide supporting evidence that our measure of risk aversion captures the underlying preference parameter. Second, for covariates where this relationship is not established, this exercise provides an exploratory analysis.

We construct county-year covariates using demographic data from the National Historical Geographic Information Database (NHGIS). and farm characteristics from the Ag Census. Demographic variables include total population and the shares of the population that are rural, female, non-white, and foreign born. Farm characteristics include the number of farms, the total value of farmland (a measure of agricultural wealth), the value of crop output (a measure of agricultural income), and the share of farms in different size bins, which reflects the wealth distribution and industrial organization of the local farm economy. We also construct two measures of concentration in agricultural production: the Hirschman-Herfindahl Index (HHI) of land planted in different core crops, and the HHI of crop value based on measured revenue.[19]

We present the results of regressing RRA on all covariates in Table 7.1. Column 1 presents the raw marginal correlations, while Columns 2–5 add different combinations of county, year, and state-by-year fixed effects. It is important to note that the inclusion of different fixed effects changes the interpretation of the coefficients: county fixed effects remove time-invariant cross-sectional variation, year fixed effects remove aggregate temporal variation, and state-by-year fixed effects remove local temporal variation. While these fixed effects are useful for causal identification, their utility in the context of a correlational analysis depends on which dimensions of the relationship of RRA with the covariates are of interest. In some cases, the regression coefficients with fewer fixed effects can be informative (e.g., when considering the cross-sectional relationship between preferences and demographics.)

Of the covariates we examine, wealth and income have the strongest theoretical link to risk aversion, with higher levels typically associated with lower risk aversion. Our regressions confirm this: aggregate farmland and crop output values are significantly and negatively correlated with RRA in nearly every specification. The relationship with farm size is even more pronounced. Counties with higher shares of small farms, which have lower wealth and income, are significantly more risk averse than those with larger farms, with a clear

[19] Crop value HHI corresponds to the measure of agricultural diversity in Fiszbein (2019), but restricted to our 13 core crops. Correlations between our RRA measure and the two HHI measures reveal that the RRA measure captures something different from pure agricultural diversity: its correlation with cropland HHI is 0.21, and with crop value HHI -0.02.

Table 7.1: Correlation between measured RRA and county-level characteristics

Dependent variable: RRA	(1)	(2)	(3)	(4)	(5)
Total population (millions)	0.11	-0.24	0.26	-0.29	-0.11
	(0.26)	(0.55)	(0.27)	(0.51)	(0.41)
Share of rural population	-0.14	0.016	-0.089	0.30	0.10
	(0.11)	(0.27)	(0.11)	(0.24)	(0.21)
Share of female population	-0.071	-4.05**	0.66	-2.14	-3.67**
	(1.42)	(1.79)	(1.45)	(1.79)	(1.87)
Share of non-white population	-0.24	-0.38*	-0.15	-0.56**	-0.0067
	(0.16)	(0.21)	(0.16)	(0.23)	(0.21)
Share of foreign-born population	0.88**	-1.05	0.98***	-0.30	-0.67
	(0.38)	(0.88)	(0.37)	(0.94)	(0.89)
Share of land in farms	-0.32**	-0.14	-0.41***	-0.53**	-0.29
	(0.15)	(0.24)	(0.14)	(0.24)	(0.22)
Number of farms (thousands)	0.13***	0.36***	0.11***	0.25***	0.16***
	(0.029)	(0.058)	(0.030)	(0.058)	(0.045)
Value of farmland (mil. of dollars of 1913)	-0.0095***	-0.010***	-0.012***	-0.0071**	-0.0022
	(0.0024)	(0.0027)	(0.0026)	(0.0029)	(0.0024)
Value of crop output (mil. of dollars of 1913)	-0.099***	-0.11***	-0.060**	-0.0064	0.0080
	(0.025)	(0.033)	(0.025)	(0.038)	(0.034)
Share of farms with 0-9 acres	2.05***	3.89***	1.52**	2.48***	3.46***
	(0.67)	(0.93)	(0.64)	(0.88)	(0.85)
Share of farms with 10-19 acres	1.49*	1.14	1.61**	0.71	1.95**
	(0.77)	(1.06)	(0.73)	(1.03)	(0.91)
Share of farms with 20-49 acres	1.09*	3.22***	1.22**	3.07***	1.77**
	(0.59)	(0.83)	(0.58)	(0.82)	(0.71)
Share of farms with 50-99 acres	-0.20	0.68	-0.26	0.60	1.07
	(0.53)	(0.88)	(0.54)	(0.88)	(0.85)
Share of farms with 100-499 acres	0.77*	1.57***	0.82**	1.60***	1.44***
	(0.40)	(0.57)	(0.41)	(0.55)	(0.55)
Share of farms with 500-999 acres	0.28	0.34	0.18	0.40	0.76
	(0.57)	(0.72)	(0.58)	(0.71)	(0.65)
Share of owner-operated farms	0.16	-0.081	0.41	0.56	1.32***
	(0.28)	(0.37)	(0.26)	(0.35)	(0.33)
Cropland HHI	2.48***	0.63*	2.43***	0.50	-0.096
	(0.40)	(0.37)	(0.39)	(0.36)	(0.34)
Crop value HHI	-1.35***	0.11	-1.37***	-0.11	0.050
	(0.30)	(0.27)	(0.30)	(0.28)	(0.26)
Observations	10,084	10,014	10,084	10,014	10,014
R-squared	0.127	0.453	0.140	0.462	0.544
County FE		Y		Y	Y
Year FE			Y	Y	Y
State x Year FE					Y

Notes: dependent variable is county-level estimated coefficient of relative risk aversion (RRA), with higher values indicating higher risk aversion. Sample of counties is pooled across 1890–1920 waves of the Ag Census. Excluded category for farm size is share of farms over 1000 acres. Conley standard errors (75 mile cutoff) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

gradient across the farm size distribution. In Column 5, a 1% higher share of farms under 10 acres is associated with a 3.46 point increase in RRA (relative to the excluded group of farms greater than 1,000 acres), a 1% higher share of farms 10–19 acres with a 1.95 point increase, 20-49 acre farms with a 1.77 point increase, and 100-499 acre farms with a 1.44 point increase.

We also observe several notable correlations between RRA and county demographics. The share of the female population is significantly and negatively associated with RRA, but only

when controlling for county fixed effects, implying that counties where the female share increases over time become more risk seeking. The share of foreign born residents shows the opposite pattern: a positive and significant correlation with RRA appears only in specifications without county fixed effects, suggesting that counties with higher foreign-born shares are more risk averse, but changes over time in this share are not associated with changes in risk aversion. Finally, the share of the population that is non-white is negatively associated with RRA, though significance varies by specification.

We also find significant correlations between RRA and the organization of the agricultural sector within counties. In the specification with county, year, and state-by-year fixed effects, the share of owner-operated farms is significantly and positively correlated with risk aversion. Risk aversion is also positively and robustly correlated with the number of farms in a county (holding the farm size distribution constant), and negatively correlated with the share of the county land that is in farms. These latter results indicate that less developed counties, such as those that are on the frontier, have individuals that are more risk-seeking.[20]

Finally, in specifications without county fixed effects, RRA is positively correlated with cropland HHI, and negatively correlated with crop value HHI. Taken together, these two correlations tell a story about how cross-sectional risk aversion is related to diversification in agriculture. Farmers who were more risk-seeking had a more diversified portfolio of cropland shares, but a less diversified distribution of agricultural revenue. This likely indicates that risk-seeking farmers planted several smaller crops that were high-risk, high-return; and that most of these crops did not deliver high returns, thereby concentrating revenue in a few "winning" crops.

8 Additional analyses and results

In this section, we present results from four additional analyses we conduct in subsamples of our data. The first two analyses test whether our measure truly captures risk aversion by studying its correlation with financial risk-taking both within and outside agriculture. The third analysis examines the degree to which our measure is representative of the underlying preference distribution by studying the mapping from farm-level risk aversion to the aggregate measure, using a dataset of farm-level crop choices in Kansas during our sample period. Finally, the fourth analysis examines the relationship between our measure of risk aversion

[20] We return to this relationship using direct measures of frontier exposure in Section 8.4.

and exposure to the American frontier at the turn of the twentieth century, which prior literature has shown shaped the long-run preferences of the American population.

8.1 Farm debt

Our ability to validate our measure of risk preference against other contemporaneous measures of risk-taking is inherently limited, since if such measures existed at scale, there would be no need for the present analysis. However, certain subsamples of the data contain measures of financial risk aversion. In this section we examine the correlation between such a measure–farm mortgage debt burden–and measured RRA in the latter part of our sample.

A well-documented feature of American agriculture in the 1910s was the widespread speculative activity within the sector. As agricultural commodity prices rose, peaking during World War I, new farmers entered the sector and existing farmers increased their land holdings. This expansion was largely financed by debt and therefore highly risky. This risk came to bear during the 1920s, when commodity prices became persistently depressed, triggering a wave of credit distress and bankruptcies across the sector (Rajan and Ramcharan (2015)).

In the 1910 and 1920 waves, the Ag Census recorded several measures of farm debt. We examine the relationship between our RRA measure and three of these variables: (i) the share of farms with mortgages; (ii) average mortgage debt across all farms; and (iii) average mortgage debt across farms with mortgages. To do this, we regress RRA on each of the variables separately within the 1910 and 1920 waves, and within the pooled sample:

$$\begin{aligned} MortgageVar_{c,1910} &= \beta RRA_{c,1910} + \delta X_{c,1910} + \alpha_s + \epsilon_{c,1910} \\ MortgageVar_{c,1920} &= \beta RRA_{c,1920} + \delta X_{c,1920} + \alpha_s + \epsilon_{c,1920}. \end{aligned} \tag{4}$$

$$MortgageVar_{c,t} = \beta RRA_{c,t} + \delta X_{c,t} + \alpha_c + \alpha_t + \alpha_{s,t} + \epsilon_{c,t}. \tag{5}$$

In these equations, $MortgageVar_{c,t}$ is one of the three mortgage debt variables in county c and year t. Controls $X_{c,t}$ include all the county-level covariates from Table 7.1. We include state fixed effects in the cross-sectional regressions, while in the pooled regression, we include county, year, and state-by-year fixed effects.

We present the results for the regressions in Table 8.1. We find no significant correlation

Table 8.1: Farm mortgage debt and RRA, 1910 and 1920

Dependent variables:	**Share of farms w/ mortgages**			**Avg mortgage debt (all)**			**Avg mortgage debt**		
	1910	1920	Pooled	1910	1920	Pooled	1910	1920	Pooled
RRA	0.00085	0.000068	0.0010	-1.22	-21.4**	-8.21*	-25.1	-58.0**	-39.5***
	(0.0016)	(0.0015)	(0.00080)	(5.54)	(8.78)	(4.60)	(23.5)	(24.8)	(12.8)
Observations	2,535	2,557	5,032	2,535	2,557	5,032	2,535	2,555	5,028
R-squared	0.721	0.808	0.949	0.700	0.726	0.908	0.588	0.734	0.923
Pop. & farmland controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Farm income & wealth	Y	Y	Y	Y	Y	Y	Y	Y	Y
Farm size & ownership & HHI	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y		Y	Y		Y	Y	
Year FE			Y			Y			Y
County FE			Y			Y			Y
State x Year FE			Y			Y			Y

Notes: *RRA* is winsorized at the 5% level. Average mortgage debt is in constant 1913 dollars. Conley standard errors (75 mile cutoff) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

between RRA and the share of farms in a county with mortgages in any specification, and no correlation between RRA and average mortgage debt in 1910. However, RRA is significantly and negatively correlated with average mortgage debt both in 1920 and in the pooled sample from 1910 to 1920 (which can be interpreted as the change between the periods). Specifically, a one-point increase in RRA is associated with $21.4 lower mortgage debt per farm in 1920, and with a $8.21 decline in mortgage debt per farm from 1910 to 1920. Likewise, for farms with mortgages, a one-point increase in RRA corresponds to $58 less mortgage debt in 1920; and a $39.5 decline from 1910 to 1920.

Comparing these values to the medians of the dependent variables helps put their magnitudes in context. In 1920, the median mortgage value across all farms (both mortgaged and not) was $459, while the median change from 1910 to 1920 was $210. Our estimates therefore imply that a one-point higher county RRA is associated with a 4.6% lower mortgage value in 1920 and a 3.9% smaller increase over the period. For farms with mortgages, the median value in 1920 was $2,001, while the median change over the decade was $990, implying a 2.9% lower value in 1920 and a 4% smaller increase over the period for each 1 point higher county RRA.

We draw several conclusions from these estimates. First, estimated risk aversion correlates negatively with risk-taking behavior in the same agricultural domain, but measured from a different kind of behavior. This correlation is not mechanical–farmers with mortgages could choose to grow safe crops. Second, the estimated magnitudes are substantial, especially given that RRA is measured from the behavior of the entire county, not just mortgaged farmers,

and that both mortgage values and RRA are likely measured with noise. Taken together, these results support the validity of our measure of risk aversion as capturing underlying risk preferences.

The findings also shed light on the agricultural production environment of the 1910s. Interpreting RRA as a measure of financial speculation, we can infer that farm mortgage debt in 1910 was not driven by speculative motives, whereas the sharp run-up in the following decade was. Further, the patterns documented suggest that financial speculation in agriculture operated mainly on the intensive margin of debt, rather than the extensive margin, since RRA is not correlated with the number of farms with mortgages, but is correlated with average mortgage size.

8.2 World War I Liberty Bonds

Another direct measure of financial behavior available for the latter part of our period is subscription rates to World War I Liberty Bonds. The U.S. government financed part of its wartime effort through four public bond campaigns conducted between 1917 and 1919 (Hilt and Rahn (2020), Brunet et al. (2025)). These drives were heavily promoted and the bonds were widely adopted, with approximately two-thirds of middle-income households participating. Although marketed by the government as a safe asset, Liberty bonds were a novel financial instrument for most purchasers, and their real returns proved volatile after the war due to Federal Reserve policy (Hilt and Rahn (2020)).

To explore the relationship between Liberty Bond subscription rates and risk aversion, we use county-level data from Hilt and Rahn (2020) on the share of subscribers in the fourth bond drive, which took place in 1919, as a share of the 1920 county population. We regress this participation rate on counties' estimated RRA, along with the set of controls included in Table 7.1 and state fixed effects:

$$AvgBondParticipationRate_c = \beta RRA_{c,1920} + \beta X_{c,1920} + \alpha_s + \epsilon_{c,1920}. \tag{6}$$

We present the results from these regressions in Table 8.2. We find a significant negative correlation between RRA and subscription rates to the fourth Liberty Bond drive: a one-point higher RRA corresponds to a 0.7–0.8 percentage point lower participation rate. Given the median participation rate of 16.1%, this represents a 5% percent reduction in participation for each additional point of risk aversion.

Table 8.2: Liberty bond subscriptions and RRA, 1920

Dependent variable: Participation rate in 4th bond drive	(1)	(2)
RRA	-0.0081***	-0.0070***
	(0.0024)	(0.0022)
Observations	1,192	1,192
R-squared	0.632	0.672
Pop. & farmland controls	Y	Y
Farm income & wealth	Y	Y
Farm size & ownership & HHI	Y	Y
State FE		Y

Notes: *RRA* is winsorized at the 5% level. County-level liberty bond subscription rate from Hilt and Rahn (2020). This specification corresponds to that of Equation (6). Conley standard errors (75 mile cutoff) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

This finding is a notable validation of our RRA measure, as it captures the correlation between risk aversion measured using agricultural information in our method and contemporaneous financial behavior in an entirely different domain. As with farm mortgage debt, the coefficient is likely attenuated by measurement noise. The sign of the effect implies that more risk-seeking counties had higher Liberty Bond participation rates. One possible explanation is that risk-seeking agents were more likely to adopt novel financial instruments, as observed in the latter half of the twentieth century (Malmendier and Nagel (2011)). Another possible explanation is that risk-seeking farmers, having taken on greater risk in their farming activities, may have placed greater value on the hedging benefits of the bonds, which offered an income stream independent of agricultural returns.

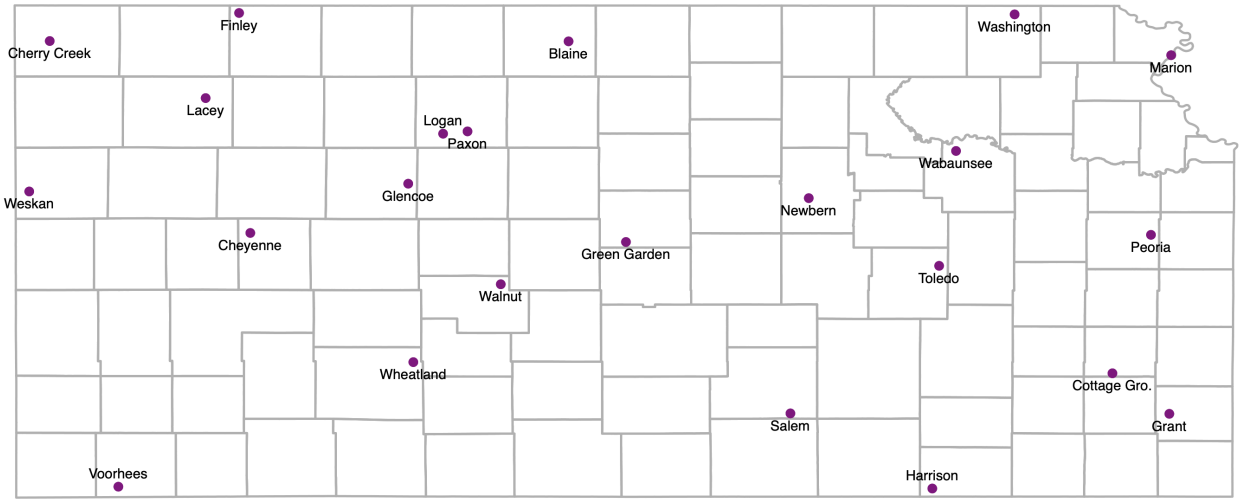
8.3 Kansas farm-level analysis

An important open question in our analysis up to this point is whether our RRA measure, which is estimated at the county level, can capture the individual or household risk preferences of the underlying population. At its core, this is a question about preference aggregation: are the estimated preferences of the county-level representative agent truly representative?

It is difficult to test this question at scale, since widespread farm-level data on agricultural production is not readily available. There is, however, one partial source for such data: the dataset compiled by Sylvester and Rupley (2012), which contains detailed agricultural records for all farms in 23 Kansas townships for the years 1875, 1880, 1885, 1895, 1905, 1915, and 1920.[21] In this section, we address the aggregation question by applying our

[21] Although the data cover all farms in each township-year, individual farms cannot be reliably tracked over

Figure 8.1: Location of townships in the Kansas farm-level dataset



Notes: This map shows the location of the townships included in the Kansas farm-level dataset.

method separately to the farm-level and to aggregated township-level data in each year and comparing the resulting RRA measures.

Figure 8.1 displays the names and locations of the Kansas townships in the dataset. In our analysis, we use data on 10 of the core crops in the county-level analysis, excluding only cotton, rice, and sugar cane which were not cultivated in Kansas during the periods studied.[22] As in the main analysis, we use corn as the numeraire crop due to its widespread cultivation across farms in the sample. Since per-acre yields can only be constructed in 1880 (the sole wave with both output and acreage data), we use the county-level per-acre yield histories corresponding to each township that we constructed previously,[23] along with crop prices for Kansas from the main price dataset.

We recover the relative risk aversion (RRA) coefficient for each farm f and township w in year t by estimating Equation (3) separately for each unit, using information on the crops grown in f or w at time t. To assess the relationship between farm- and township-level RRAs, we compute two statistics: (1) the difference between the mean RRA of all farms in a township-year and the township-level RRA in that year, $Dist_{mean,w,t} = Mean(RRA_{f,t}) - RRA_{w,t}$;

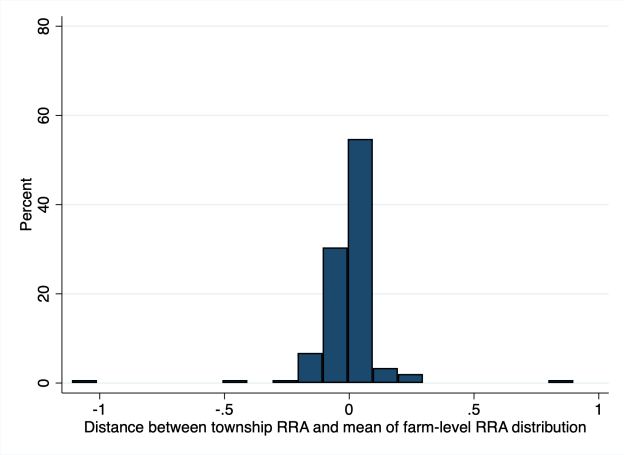
time due to changes in proprietorship, mergers, and other turnover. We therefore treat the dataset as a series of repeated cross-sections. The 181 township-year cells each contain between 6 and 798 farms, with the median cell including 101 farms.

[22] Table E.1 reports summary statistics on average acreage shares by crop and year, as well as the average share of land planted with the selected crops.

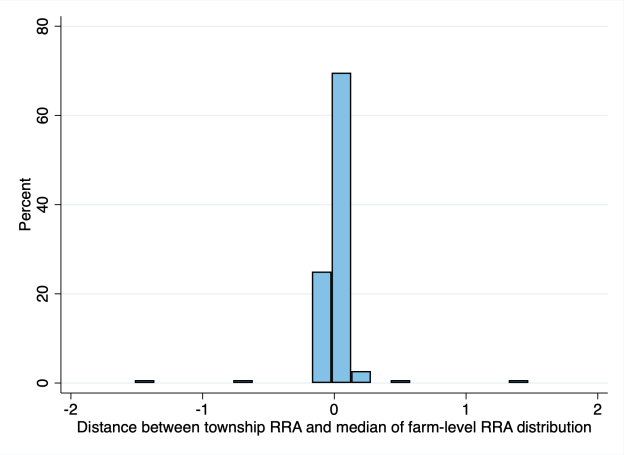
[23] The correlation between county-level yields and township-level yields constructed from individual farm data for all crops in 1880 is 0.83.

Figure 8.2: Distance between risk aversion coefficients at the farm and township levels

(a) Difference between the mean farm RRA and township RRA



(b) Difference between the median farm RRA and township RRA



Notes: These figures plot the histograms of (1) the difference between the mean RRA of all farms in a township-year and the township-level RRA in that year; and (2) the difference between the median RRA of all farms in a township-year and the township-level RRA in that year. The analysis is conducted using farm-level data from Sylvester and Rupley (2012) for 23 Kansas township with farm-level data in 1875, 1880, 1885, 1895, 1905, 1915, and 1920.

and (2) the difference between the median RRA of all farms in a township-year and the township-level RRA in that year, $Dist_{median,w,t} = Median(RRA_{f,t}) - RRA_{w,t}$. We plot the distributions of $Dist_{mean,w,t}$ and $Dist_{median,w,t}$ across all township-years in panels (a) and (b) of Figure 8.2, respectively.[24]

The results show that the township-level RRA measures closely match the average RRA across farms within each township. Both the mean and median distance distributions are highly concentrated around zero, with approximately 85% of township mean RRAs and 95% of township median RRAs in within 0.1 of the aggregate measure. While there is a moderate amount of skew and a few outliers in each distribution, these findings suggest that our county-level RRA measures likely reflect the average RRA of farmers within them.

8.4 Frontier experience

Recent work in economic history has shown that exposure to the American frontier helped shape long-run preferences for "rugged individualism" among those exposed to it (Bazzi et al. (2020)). In this section, we examine whether the legacy of the frontier also shaped risk-taking, a trait related to individualism, and another prominent feature of the mythos of

[24] As in the main analysis, the farm- and township-level RRA measures are winsorized at the 5% level.

Table 8.3: Relationship of total frontier experience and RRA measure in 1890

Dependent variable: Total frontier experience	(1)	(2)
RRA	-0.074***	-0.022
	(0.019)	(0.015)
Observations	2,513	2,513
R-squared	0.144	0.423
Pop. & farmland controls	Y	Y
Farm income & wealth	Y	Y
Farm size & ownership & HHI	Y	Y
State FE		Y

Notes: *RRA* is winsorized at the 5% level. Total frontier experience measures the number of decades that each county spent in the frontier from 1790 to 1890, from Bazzi et al. (2020). Conley standard errors (75 mile cutoff) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

the American West (DeArment (1990)).

To address this question, we regress measured RRA in 1890 on the county-level measure of total frontier experience (TFE) from Bazzi et al. (2020), which is defined as the number of decades the county was on the frontier between 1790 and 1890, along with a set of control variables and state fixed effects:

$$TFE_c = \beta RRA_{c,1890} + \beta X_{c,1890} + \alpha_s + \epsilon_{c,1890}. \tag{7}$$

We present the results of this regression in Table 8.3. We find that counties with higher total frontier experience exhibit higher measured risk-taking. The relationship is highly significant excluding state fixed effects, and is nearing standard levels of significance (p = 0.13) in the specification that includes state fixed effects, where identification comes solely from variation between counties within the same state.

9 Conclusions

In this paper we develop a new empirical strategy for measuring the risk preferences of individuals in the past, using crop choice as a proxy grounded in a standard portfolio allocation model. Leveraging U.S. Census of Agriculture data from 1880 to 1920, harmonized crop price series, and historical climate information, we recover county-level estimates of relative risk aversion and document substantial spatial heterogeneity and temporal variation. These patterns are consistent with longstanding theoretical and empirical work linking agricultural production choices to risk attitudes and offer a rare window into the historical distribution of preferences under uncertainty.

We further demonstrate the risk preference measure's internal and external validity through complementary analyses. In particular, we show that measured risk aversion was negatively correlated with financial risk-taking both within and outside of agriculture, by examining its relationship with speculative farm debt and WWI Liberty Bond subscriptions in the 1910s. In addition, we provide evidence that the aggregate county-level measure is representative of the underlying preference distribution of farms using detailed farm-level data from Kansas. Finally, we connect the geography of risk attitudes to experiences of the American frontier, finding that counties with greater frontier exposure were more risk-seeking, consistent with the hypothesis that frontier conditions selected for and reinforced tolerance for risk. Taken together, these results show that our method is an effective tool for recovering meaningful, behaviorally relevant measures of historical risk preferences, and that it can be applied broadly to study the long-run evolution of attitudes toward risk in diverse historical contexts.

References

Anderson, J. R., Dillon, J. L., and Hardaker, J. B. (1977). *Agricultural Decision Analysis*. The Iowa State University Press. 2, 6

Anderson, L. R. and Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*, 27(5):1260–1274. 2

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. *The Quarterly Journal of Economics*, 112(2):537–579. ISBN: 00335533. 2

Bazzi, S., Fiszbein, M., and Gebresilasse, M. (2020). Frontier Culture: The Roots and Persistence of Rugged Individualism in the United States. *Econometrica*, 88(6):2329–2368. 4, 28, 29

Brunet, G., Hilt, E., and Jaremski, M. (2025). Invest!: Liberty Bonds and Stock Ownership over the Twentieth Century. *Working paper*. 4, 25

Cole, S., Gin, X., and Vickery, J. (2017). How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment. *The Review of Financial Studies*, 30(6):1935–1970. 2

Cooley, T. F., DeCanio, S. J., and Matthews, M. S. (1977). An Agricultural Time Series-Cross Section Data Set. *NBER Working Paper Series*. 12, 37

DeArment, R. K. (1990). *Knights of the Green Cloth :The Saga of the Frontier Gamblers*. University of Oklahoma Press. 29

Dercon, S. (1996). Risk, Crop Choice, and Savings:Evidence from Tanzania. *Economic Development and Cultural Change*, 44(3):484–700. 2, 6

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study

using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17. 5

Eckert, F., Gvirtz, A., Liang, J., and Peters, M. (2020). A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790. *NBER Working Paper Series*. 10

Fafchamps, M. (1992). Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World. *American Journal of Agricultural Economics*, 74(1):90–99. 2, 6

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2016). The Nature and Predictive Power of Preferences: Global Evidence. *Quartely Journal of Economics*. 2, 5

Fiszbein, M. (2019). Agricultural Diversity, Structural Change and Long-run Development: Evidence from the U.S. Technical report. Publication Title: National Bureau of Economic Research Issue: July ISSN: 0898-2937. 20

Frieden, J. A. (1997). Monetary Populism in Nineteenth-Century America: An Open Economy Interpretation. *The Journal of Economic History*, 57(2):367–395. 9

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, (May). 5

Goodwyn, L. (1978). *The Populist Moment: A Short History of the Agrarian Revolt in America*. Oxford University Press. 19

Haines, M., Fishback, P., and Rhode, P. (2018). United States Agriculture Data, 1840 - 2012. 10

Herranz, N., Krasa, S., and Villamil, A. P. (2015). Entrepreneurs, Risk Aversion, and Dynamic Firms. *Journal of Political Economy*, 123(5):1133–1176. 2

Hilt, E. and Rahn, W. (2020). Financial Asset Ownership and Political Partisanship: Liberty Bonds and Republican Electoral Success in the 1920s. *The Journal of Economic History*, 80(3):746–781. 4, 25, 26

Holt, C. A. and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5):1644–1655. 5

Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., and Bonin, H. (2010). Direct evidence on risk attitudes and migration. *The Review of Economics and Statistics*, 92(3):684–689. 2

Karlan, D., Osei, R. D., Isaac, O.-A., and Udry, C. (2014). Agricultural Decisions after Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics*, 129(2):597–652. 2

Katz, L. and Margo, R. (2014). Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective. In *Human Capital in History*, pages 15–57. University of Chicago Press and NBER, Cambridge, MA. 9

Kim, S. and Margo, R. (2004). Historical Perspectives on U.S. Economic Geography. In *Handbook of Regional and Urban Economics*, volume 4. Elsevier North Holland, Amsterdam. 9

Kuchler, T. and Zafar, B. (2019). Personal Experiences and Expectations about Aggregate Outcomes. *The Journal of Finance*, 74(5):2491–2542. 3

Levin, R. and Vidart, D. (2025). Risk-Taking Adaptation to Macroeconomic Experiences. *Journal of Political Economy Microeconomics*. 2, 5, 15

Malmendier, U. and Nagel, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking? *The Quarterly Journal of Economics*, 126(1):373–416. ISBN: 00335533. 2, 3, 5, 15, 26

Malmendier, U. and Nagel, S. (2016). Learning from inflation experiences. *Quarterly Journal of Economics*, 131(1):53–87. 3

Moschini, G. and Hennessy, D. A. (1999). Uncertainty, risk aversion and risk management for agricultural producers. In *Handbook of Agricultural Economics*, volume 1A of *Handbooks in Economics 18*, pages 87–153. North-Holland. 2, 5, 7

Olmstead, A. L. and Rhode, P. W. (2018). Agriculture in American Economic History. In *The Oxford Handbook of American Economic History*, volume 1, pages 159–182. 9

Rajan, R. and Ramcharan, R. (2015). The Anatomy of a Credit Crisis: The Boom and Bust in Farm Land Prices in the United States in the 1920s. *American Economic Review*, 105(4):1439–1477. 4, 19, 23

Rosenzweig, M. R. and Binswanger, H. P. (1993). Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *The Economic Journal*, 103(416):56–78. 2

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brnnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., DomnguezCastro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyski, P. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 145(724):2876–2908. 12

Sylvester, K. M. and Rupley, E. S. A. (2012). Revising the Dust Bowl: High Above the Kansas Grasslands. *Environmental History*, 17(3):603–633. 4, 26, 28

US Bureau Of Labor Statistics (1927). Wholesale Prices 1890 to 1926. Technical Report Bulletin No. 440, Bureau of Labor Statistics. 12, 37

US Census Bureau (1975). Historical Statistics of the United States, Colonial Times to 1970. Technical report. Section: Government. 12, 37

U.S. Senate Committee on Agriculture, Nutrition, a. F. t. C. (1998). Chapter 3: From the "Golden Age" to the Great Depression: 1900-1929. Technical Report S. Doc. 105-24. 9

A Model derivation

In order to solve the farmer's problem in Equation (1), we first assume, without loss of generality, that crop 1 is the numeraire. Taking the first order conditions for all crops planted by the farmer yields the following series of equations:

$$\mathbb{E}\left[U'(c)\left(p_i q_i - 1\right)\right] = 0 \quad \forall i \in \{2, \ldots, N\} \,|\, l_i > 0 \tag{A1}$$

These equations capture that the expected marginal gain from planting one more acre of crop i needs to be equal to that of the numeraire, and therefore also that of other crops.

We can then linearize each of these equations using a first-order Taylor expansion of the expression inside the expectation around the expected value of consumption, $\mathbb{E}(c)$:[25]

$$U'(c)\left(p_i q_i - 1\right) \approx U'(\mathbb{E}(c))\left(p_i q_i - 1\right) + U''(\mathbb{E}(c))\left(p_i q_i - 1\right)\left(c - \mathbb{E}(c)\right) \tag{A2}$$

Notice that we can write:

$$c - \mathbb{E}(c) = \sum_{j=1}^{N} p_j q_j l_j \bar{L} - \sum_{j=1}^{N} \mathbb{E}(p_j q_j) l_j \bar{L} = \bar{L} \sum_{j=1}^{N} \left(p_j q_j - \mathbb{E}(p_j q_j)\right) l_j$$

Thus, if we multiply $(c - \mathbb{E}(c))$ by $(p_i q_i - 1)$ we get:

$$\left(c - \mathbb{E}(c)\right)\left(p_i q_i - 1\right) = \bar{L} \sum_{j=1}^{N} \left(p_j q_j p_i q_i - \mathbb{E}(p_j q_j) p_i q_i\right) l_j - \bar{L} \sum_{j=1}^{N} \left(p_j q_j - \mathbb{E}(p_j q_j)\right) l_j$$

Plugging this back in Equation (A2), taking expected value, and setting the expression equal to zero to match the first-order condition yields

$$\begin{aligned}
0 &= U'(\mathbb{E}(c))\left(\mathbb{E}(p_i q_i) - 1\right) + U''(\mathbb{E}(c))\mathbb{E}\left[\bar{L} \sum_{j=1}^{N} \left(p_j q_j p_i q_i - \mathbb{E}(p_j q_j) p_i q_i\right) l_j\right] - \mathbb{E}\left[\bar{L} \sum_{j=1}^{N} \left(p_j q_j - \mathbb{E}(p_j q_j)\right) l_j\right] \\
&= U'(\mathbb{E}(c))\left(\mathbb{E}(p_i q_i) - 1\right) + U''(\mathbb{E}(c))\left[\bar{L} \sum_{j=1}^{N} \left(\mathbb{E}\left(p_j q_j p_i q_i\right) - \mathbb{E}(p_j q_j)\mathbb{E}(p_i q_i)\right) l_j\right] \\
&= U'(\mathbb{E}(c))\left(\mathbb{E}(p_i q_i) - 1\right) + U''(\mathbb{E}(c))\left[\bar{L} \sum_{j=1}^{N} \left(Cov(p_j q_j, p_i q_i)\right) l_j\right]
\end{aligned}$$

[25] Using this Taylor expansion implies that the value of risk aversion we recover will be a local estimate around the expected value of consumption, and will rely on a linear approximation of the farmer's choices.

Rearranging, we get

$$\mathbb{E}(p_i q_i) - 1 = \frac{-U''(\mathbb{E}(c))}{U'(\mathbb{E}(c))} \bar{L} \sum_{j=1}^{N} Cov(p_j q_j, p_i q_i) l_j$$

In order to obtain the coefficient of relative risk aversion, we multiply and divide the right hand side of the equation by $\mathbb{E}(c) = \bar{L} \sum_{j=1}^{N} \mathbb{E}(p_j q_j) l_j$:

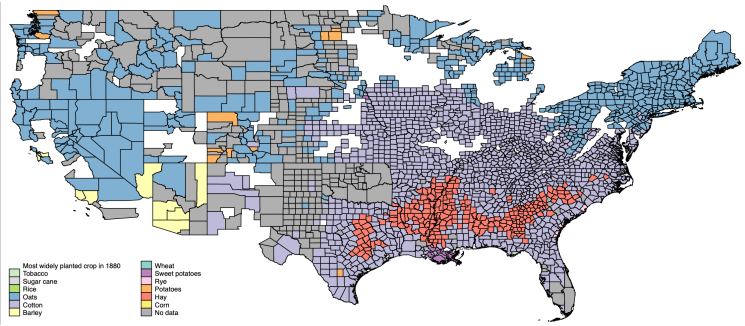
$$\mathbb{E}(p_i q_i) - 1 = \frac{-U''(\mathbb{E}(c))}{U'(\mathbb{E}(c))} \mathbb{E}(c) \frac{\sum_{j=1}^{N} Cov(p_j q_j, p_i q_i) l_j}{\sum_{j=1}^{N} \mathbb{E}(p_j q_j) l_j}.$$

B Additional information on census of agriculture

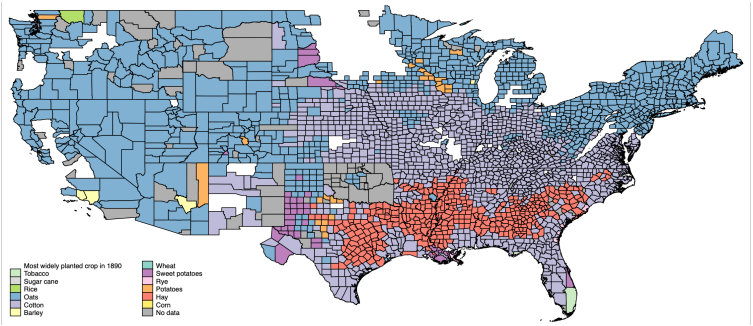
B.1 Cultivation patterns

Figure B.1: Most widely grown crop in 1880–1920

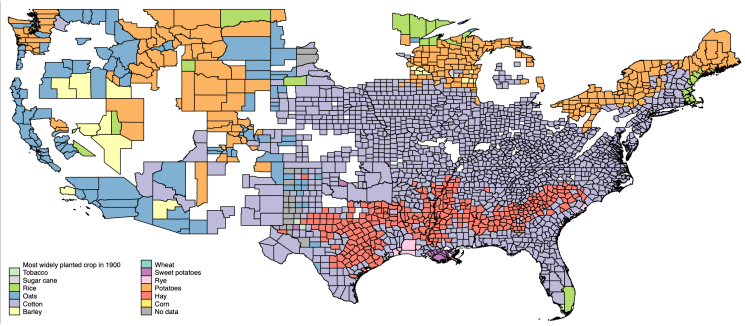
(a) 1880



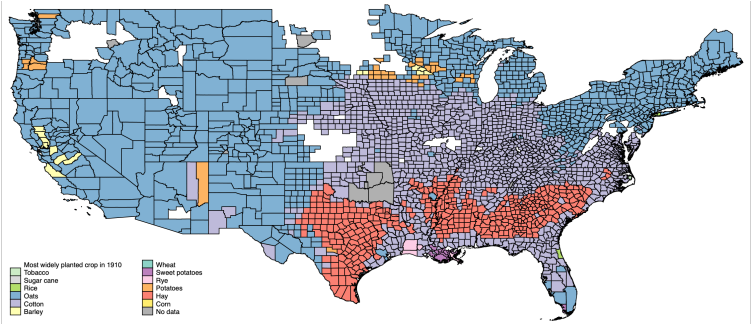
(b) 1890



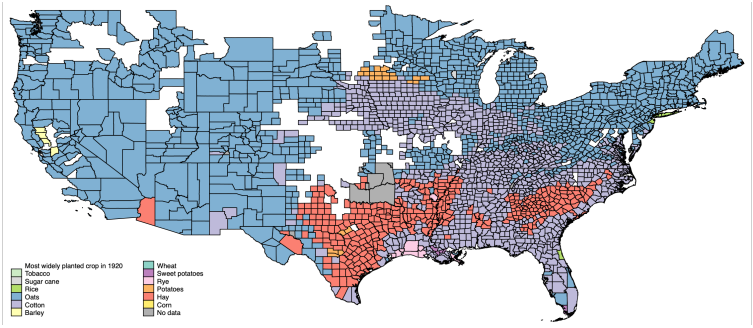
(c) 1900



(d) 1910



(e) 1920



Notes: This map plots the most widely grown crop out of the set of crops listed in Section 5.1 in each county in 1880, 1890, 1900, 1910, and 1920.

B.2 Improved land definition

Improved land includes all land regularly tilled or mowed, land pastured and cropped in rotation, land lying fallow, land in gardens, orchards, vineyards, and nurseries, and land

occupied by farm buildings. Unimproved land encompasses any land which is not improved or is forested. Some examples of unimproved land are woodland (natural or planted, and which may later be used for firewood), brush land, rough or stony land, swamp, etc.

B.3 Cropland shares

Table B.1: Summary statistics of share of land planted in each crop in 1880–1920

	1880		1890		1900		1910		1920	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Share of barley	0.013	0.000	0.011	0.000	0.011	0.000	0.023	0.001	0.017	0.002
Share of buckwheat	0.004	0.000	0.003	0.000	0.003	0.000	0.006	0.000	0.005	0.000
Share of corn	0.375	0.412	0.330	0.357	0.346	0.372	0.341	0.371	0.301	0.317
Share of cotton	0.101	0.000	0.113	0.000	0.097	0.000	0.267	0.209	0.119	0.000
Share of hay	0.191	0.108	0.269	0.207	0.273	0.184	0.271	0.206	0.264	0.190
Share of oats	0.093	0.084	0.111	0.103	0.080	0.051	0.091	0.057	0.087	0.053
Share of potatoes	0.006	0.000	0.016	0.007	0.015	0.005	0.018	0.007	0.016	0.004
Share of rice	0.003	0.000	0.003	0.000	0.004	0.000	0.011	0.000	0.005	0.000
Share of rye	0.010	0.002	0.010	0.002	0.007	0.001	0.010	0.001	0.018	0.004
Share of sugar cane	0.004	0.000	0.004	0.000	0.005	0.000	0.018	0.001	0.004	0.000
Share of sweet potatoes	0.006	0.001	0.007	0.000	0.006	0.000	0.009	0.002	0.008	0.001
Share of tobacco	0.004	0.000	0.005	0.000	0.006	0.000	0.015	0.000	0.019	0.000
Share of wheat	0.187	0.154	0.125	0.068	0.149	0.086	0.106	0.046	0.164	0.108
Share of farmland w/ selected crops	0.590	0.612	0.585	0.622	0.620	0.645	0.581	0.600	0.622	0.646

Notes: This table reports the mean and median share of land allocated to each of the selected crops across all counties in 1880, 1890, 1900, 1910, and 1920. Shares are calculated relative to the total acreage planted with the full set of crops considered. The table also reports the mean and median share of total farmland devoted to these crops in each year.

C Additional information on price data

Table C.1: Data sources used to construct price time-series

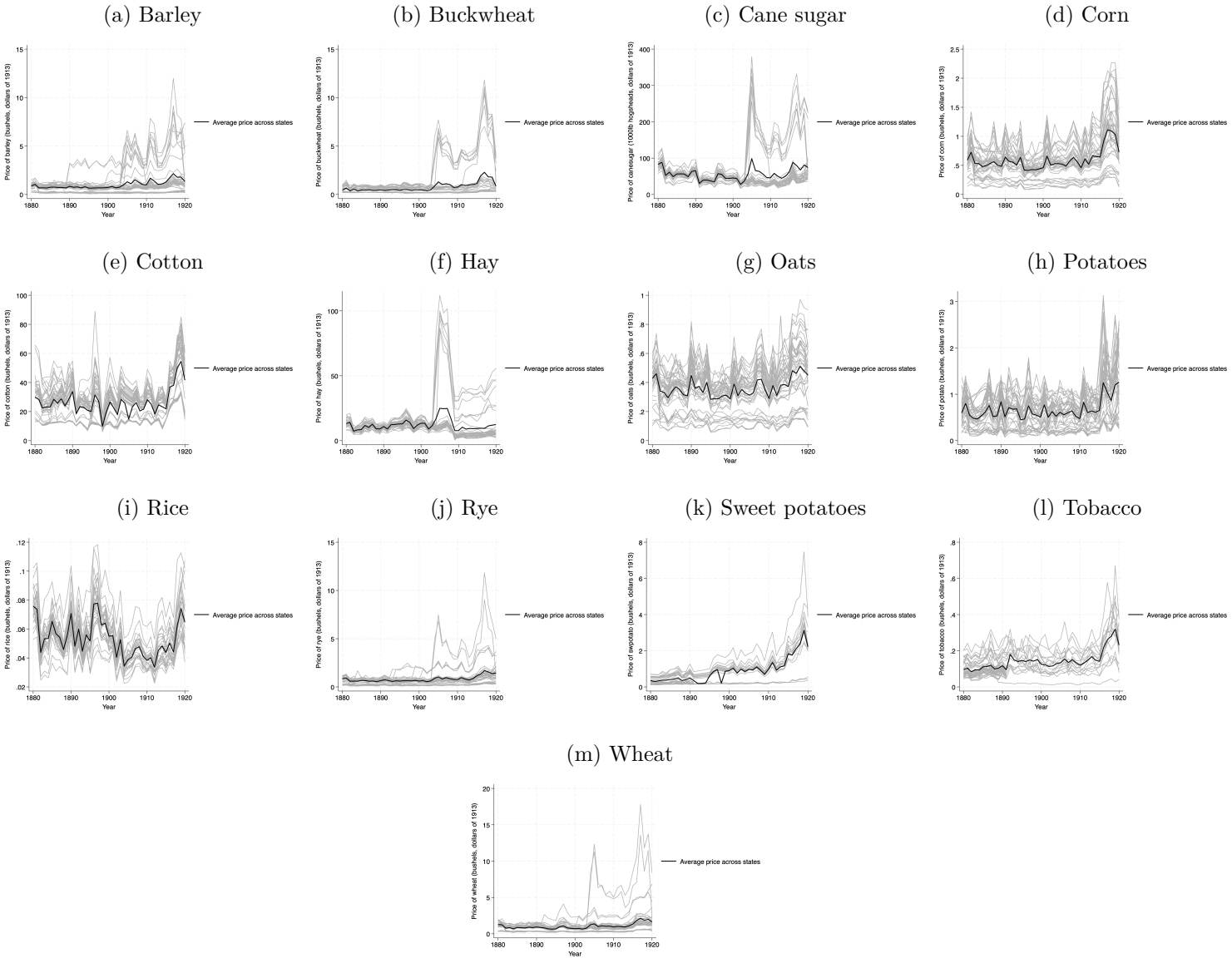
State-level prices (1880–1920)		
Data source	**Years & crops**	**Description**
Agricultural Time Series-Cross Section Data Set (Cooley et al. (1977))	1880–1920; barley, buckwheat, corn, cotton, hay, oats, potatoes, rice, rye, sweet potatoes, tobacco, and wheat	Agricultural Time Series-Cross Section (ATICS) dataset compiled using bulletins and circulars of the United States Department of Agriculture.
Auxiliary national-level data used to construct state-level prices		
Data source	**Years & crops**	**Description**
Wholesale Prices 1890 to 1926 (US Bureau Of Labor Statistics (1927))	1890–1920; barley, corn, oats, potatoes, rice, rye, sweet potatoes, cane sugar, and tobacco	Bulletin of the Bureau of Labor Statistics (BLS) with a record of nation-wide wholesale prices.
Historical Statistics of the United States, Colonial Times to 1970 (US Census Bureau (1975))	1880–1920; buckwheat, hay, and sweet potatoes	Compendium providing a statistical history of the U.S. on a variety of topics, including crop prices

Table C.2: Summary statistics of crop price time-series in each decade in 1880–1920, in dollars of 1913

	1880–1889		1890–1899		1900–1909		1910–1919	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Price of barley (bushels)	0.761	0.825	0.723	0.691	1.021	0.688	1.491	0.882
Price of buckwheat (bushels)	0.467	0.290	0.463	0.257	0.784	0.251	1.366	0.422
Price of cane sugar (1000lb hogsheads)	61.653	58.150	44.814	42.499	54.102	36.581	62.652	39.658
Price of corn (bushels)	0.546	0.562	0.505	0.490	0.547	0.543	0.772	0.671
Price of cotton (400lb bales)	26.164	26.737	22.705	22.804	23.342	24.098	32.220	29.170
Price of hay (tons)	10.500	10.080	12.148	11.832	16.413	11.103	9.757	5.055
Price of oats (bushels)	0.355	0.382	0.340	0.353	0.349	0.364	0.405	0.408
Price of potatoes (bushels)	0.587	0.579	0.622	0.630	0.587	0.601	0.817	0.715
Price of rice (pounds)	0.058	0.057	0.061	0.060	0.046	0.045	0.048	0.044
Price of rye (bushels)	0.665	0.682	0.646	0.670	0.777	0.653	1.122	0.820
Price of sweet potatoes (bushels)	0.393	0.290	0.596	0.597	0.918	0.975	1.675	1.441
Price of tobacco (pounds)	0.102	0.088	0.131	0.114	0.131	0.125	0.193	0.170
Price of wheat (bushels)	0.914	0.925	0.820	0.815	0.963	0.765	1.411	0.923

Notes: This table presents the mean and median of the prices of different crops across states in different decades. See text for the details about the construction of these price variables. All prices are deflated using the national wholesale price index, with a base year of 1913.

Figure C.1: Time series of crop prices



Notes: This figure plots the time series of prices for each of the selected crops.

D Additional information on yields data

Table D.1: Summary statistics of per-acre yields time-series in each decade in 1880–1920

	1880–1889		1890–1899		1900–1909		1910–1919	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Per-acre barley yield (bushels)	20.418	20.281	24.072	24.582	19.696	19.899	21.316	20.659
Per-acre buckwheat yield (bushels)	9.368	9.116	11.536	11.114	12.970	12.736	13.752	13.443
Per-acre cane sugar yield (1000lb hogsheads)[26]	0.204	0.011	0.113	0.008	0.037	0.000		
Per-acre corn yield (bushels)	23.012	22.240	22.905	21.802	16.678	14.542	31.658	31.003
Per-acre cotton yield (400lb bales)	0.301	0.282	2.192	0.326	0.411	0.000	0.356	0.327
Per-acre hay yield (tons)	1.153	1.118	1.243	1.209	1.171	1.114	1.317	1.252
Per-acre oats yield (bushels)	14.615	10.209	19.336	17.412	14.765	13.308	19.437	19.244
Per-acre potato yield (bushels)	352.804	161.367	91.855	89.511	70.111	66.463	75.592	73.632
Per-acre rice yield (pounds)	535.176	510.294	511.050	453.038	857.132	700.993	24.166	20.292
Per-acre rye yield (bushels)	9.047	8.704	11.289	11.016	11.698	11.582	11.032	10.698
Per-acre sweet potato yield (bushels)	63.808	63.417	69.303	69.140	49.604	46.466	104.164	101.510
Per-acre tobacco yield (pounds)	468.202	434.322	1,126.024	504.654	1,603.045	611.409	713.569	694.061
Per-acre wheat yield (bushels)	12.009	11.783	8.140	6.998	10.494	10.072	17.164	16.754

Notes: This table presents the mean and median per-acre yields for different crops across counties in different decades. See text for the details about the construction of these per-acre yield variables.

E Agricultural summary statistics for Kansas farm-level data

Table E.1: Summary statistics of share of land planted in each crop in the Kansas farm-level dataset

	1875		1880		1885		1895		1905		1915		1920	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Share of barley	0.017	0.000	0.008	0.000	0.025	0.000	0.024	0.000	0.028	0.000	0.013	0.000	0.025	0.000
Share of buckwheat	0.001	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Share of corn	0.525	0.528	0.460	0.429	0.539	0.527	0.511	0.518	0.374	0.365	0.261	0.226	0.219	0.167
Share of hay	0.270	0.000	0.262	0.107	0.276	0.146	0.273	0.234	0.498	0.500	0.623	0.643	0.602	0.626
Share of oats	0.095	0.024	0.040	0.000	0.095	0.054	0.128	0.089	0.063	0.000	0.078	0.000	0.096	0.050
Share of potatoes	0.029	0.008	0.017	0.003	0.023	0.006	0.022	0.004	0.021	0.004	0.013	0.002	0.011	0.001
Share of rye	0.029	0.000	0.003	0.000	0.027	0.000	0.015	0.000	0.003	0.000	0.005	0.000	0.006	0.000
Share of sweet potatoes	0.001	0.000	0.002	0.000	0.003	0.000	0.004	0.000	0.004	0.000	0.003	0.000	0.001	0.000
Share of tobacco	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Share of wheat	0.031	0.000	0.207	0.111	0.012	0.000	0.023	0.000	0.010	0.000	0.004	0.000	0.040	0.000
Share of farmland w/ selected crops	0.367	0.312	0.347	0.281	0.440	0.425	0.515	0.541	0.604	0.646	0.580	0.608	0.546	0.562

Notes: This table reports the mean and median share of land allocated to each of the selected crops across all farms in 1875, 1880, 1885, 1905, 1915, and 1920. Shares are calculated relative to the total acreage planted with the full set of crops considered. The table also reports the mean and median share of total farmland devoted to these crops in each year.

[26] For sugar cane, the measurable output is cane sugar. However, we lack production data for cane sugar in 1910 and 1920, which prevents us from calculating per-acre yields. Since we do have its acreage data and sugar cane represents a small share of total production, we keep it in our core set of crops.