# Crops Vs. Fences: Did Homesteaders and Purchasers Follow Different Investment Strategies in Their Land?

Katherine Hauck*

October 9, 2024

**Abstract**

The Homestead Act of 1862 opened up the western half the United States to low-cost settlement and farming by individuals by giving away more than 10% of the land area in the country. This Act created a substantially cheaper way of acquiring farmland, raising the following question: To what extent and through what mechanisms did the Homestead Act change agricultural productivity and farmer types? This paper explores the causal effect of farmer selection into different methods of acquiring land on farm productivity at the individual level in the 1860's, 1870's, and 1880's in Kansas. Selection at the farmer level into each method of land acquisition - purchasing versus homesteading - complicates the causal analysis. I model this question using an instrumental variable approach, exploiting the unique administrative requirements of homesteading. I use a newly digitized dataset of individual-level land acquisition decisions and farm production in eastern Kansas and match these data to the full count U.S. population census to demonstrate how demographic characteristics at the individual level relate to farming decisions. I find that purchasers and homesteaders use different farming strategies: homesteaders initially invest in crops and livestock, while purchasers initially invest in durable improvements like fences. While homesteaders initially produce more agricultural output, after about eighteen months, purchasers surpass homesteaders.

*I am grateful to Laura Davidoff-Taylor, Price Fishback, Ashley Langer, Juan Pantano, Rob Porter, Andrew Stevens, Evan J. Taylor, Chris Udry, Tiemen Woutersen, and Mo Xiao for their advice, suggestions, and guidance. Valuable feedback was provided by seminar participants at the Young Scholars' Institute, particularly Laura Maravall and the Association for Mentoring and Inclusion in Economics Applied Microeconomics Conference. I further thank Carter Hale, Joseph Bickle, and participants at MTurk for digitization assistance, and Jeremy Atack for providing some data for this project. This work was funded by the University of Arizona, the Economic History Association, and the National Science Foundation (#2117324).

# 1 Introduction

Under the Homestead Acts, beginning on January 1st, 1863, more than 270 million acres of land in the western part of the United States were given out by the federal government to individuals to be farmed. This amount represents more than 10% of the land area in the United States, and the lost-cost nature of the Homestead Act materially shifted the nature of land acquisition and land use in the American West. The Homestead Act created a substantially cheaper way of acquiring farmland, raising the following question: To what extent and through what mechanisms did the Homestead Act change agricultural productivity and farmer types? This paper tackles this issue, focusing on selection into homesteading, and how that selection affects subsequent productivity in Kansas, one of the states most heavily affected by homesteading.

In the late 1800's, individuals had two main ways to acquire land from the federal government: direct purchase and homesteading. Purchasing land was much more expensive but had few administrative requirements; homesteading was so inexpensive as to be nearly free but required the farmer to improve the land before getting the title. Applying for a homestead required three steps: 1) filing an application which cost 10 dollars, 2) improving the land, and 3) applying for a patent (deed to the land) which cost 4 dollars. Improving the land was defined by living on the land, building a house and other buildings, and cultivating and farming the land for a minimum of 5 years and a maximum of 7 years of residence. Historians estimate that nationwide only about 45 percent of homesteads were successfully proved after 5 to 7 years (Gates (1968)). In general, the same public land was available for purchase and for homesteading, and I limit the sample to this land by excluding railroad grants. All land in this sample was subject to the same minimum price of 1.25 dollars per acre, regardless of quality, location, or any other factors. Purchasing 160 acres of land at 1.25 dollars per acre cost settlers 200 dollars, more than ten times the cost of homesteading the same amount.

This paper is concerned with the question of how the Homestead Act influenced agricultural productivity and farmer types. An important challenge for identifying the causal effect purchasing versus homesteading on land outcomes such as farm production is that there may be selection into each group. The fact that purchasers paid more than ten times as much per acre as homesteaders indicates that that the two groups of farmers may have been quite different. Purchasers were plausibly wealthier on average at the time of acquisition, and wealth at the time of acquisition is generally unobservable. This additional wealth may have allowed purchasers to put more capital into their farm, leading to the farm being more valuable in the future. I model this selection issue with an instrumental variable approach by exploiting a unique administrative feature of homesteading. Homesteaders and purchasers both were required to get their land from the land office; however, homesteaders were required to go the land office twice while purchasers only had to go once. Using an IV approach based on the distance to the land office allows me to decompose the differences in outcomes between purchasers and homesteaders into selection effects and direct effects caused by the method of land acquisition.

First, I present a simple model of selective purchasing and use it to develop an instrument for identifying the causal effect of the method of acquisition on subsequent farm productivity. Second, I provide evidence of selection into purchase and homesteading on demographic characteristics and discuss the consequences for previous findings in this area. Third, I show that different farming strategies are correlated with purchase and homesteading, but that purchasers and homesteaders employed the same farming strategies in the years immediately after their land acquisition using the instrument. This result implies that while their production strategies were different, the causal mechanism was not the method of acquisition. Finally, I develop a model with heterogeneous production functions and show that this leads to selection into purchase and homesteading because of the administrative requirements of both. I show there exists heterogeneity in production functions between purchasers and homesteaders, which can lead to the different farming strategies observed.

I find that homesteaders initially invest in crops and livestock, while purchasers initially invest in durable improvements to their farm like fences and buildings. While homesteaders initially produce more agricultural output, after about eighteen months, purchasers surpass homesteaders in terms of crop and livestock production. Meanwhile, purchasers who remain on their farms past eighteen months have a more valuable farm in terms of durable improvements than homesteaders who remain on their farm a comparable amount of time. Purchasers initially investing in durable improvements while homesteaders produce crops and livestock represent different farming strategies. However, I find that these different farming strategies are not caused by the method of acquisition itself (i.e., purchasing vs. homesteading), but instead caused by selection of different types of farmers with different demographic characteristics and farming abilities into purchasing or homesteading. I estimate the production functions for purchasers and homesteaders separately and find a statistically significant difference how purchasers and homesteaders use inputs into the production function to create value.

This analysis in general is made possible by a combination of rich, newly digitized, individual-level data sources. Much previous work on homesteading has been conducted at the county level, leading to a lack of individual-level demographic control variables. I use the individual Kansas Agricultural Censuses of 1860, 1870, and 1880, which record individual farm production, assessed value, fencing, and numbers of livestock and bushels of crops by owner, matched to the United States Population Censuses of the same years and to the Bureau of Land Management tract books, which record exact the location, owner, type, and date of each land acquisition from the federal government. Unlike county-level data, using individual-level data allows the econometrician to estimate the selection into homesteading based on individual attributes.

Much of the previous literature related to homesteading focuses on how the Homestead Act impacted long-term land development. In particular, Leonard and Allen (2021) and Mattheis and Raz (2021) study the impact of the initial purchase-homestead decision on economic development on the land by the twentieth century. Smith (2020) focuses on how the initial decision to homestead land relates to the rate of land consolidation. Hansen and Libecap (2004) examine how the Homestead Act created small farms which exacerbated the Dust Bowl. Anderson and Hill (1990) create a theoretical model of how the decision to purchase, homestead, or preempt land relates to when the land is put into farming production. Allen (1991) and Allen (2019) both focus on how the Homestead Act intersects with property rights in the West historically. Because it has been mostly conducted at the county level, this literature has been largely unable to control for individual initial farmer demographics, and the implicit assumption made is that, after some controls, farmers homestead and purchase exogeneously and there is no selection into either group based on unobservables.

Selection on wealth into purchasing versus homesteading would cause an over-estimation the effect of purchasing a farm on its subsequent value. In fact, several previous papers have found that purchased land is more economically valuable today than homesteaded land, and have attributed this causally to the method of acquisition (Leonard and Allen (2021), Smith (2020), Allen (2019)). However, purchasers and homesteaders may be different in unobservable ways, including wealth, motivation, future goals, ability, and health. While selection at the individual level has been often overlooked in the homesteading literature, this type of selection may cause an endogeneity issue on unobserved individual-level characteristics associated with purchasing or homesteading. Therefore, in order to identify any causal effect of the decision to purchase or homestead after factoring out selection, I exploit the geographical and time variation in the distance from the acquired land to nearest land office. I use this variation in administrative costs as an instrument for the decision to purchase or homestead at the individual level. Purchasers had to go to the land office only once, while homesteaders had to go twice. I show that increased distance to a land office is a significant predictor of purchase. I show that increased distance to a land office is a significant predictor of purchase. The identifying assumption is that the distance to a land office has no impact on farming strategies or agricultural productivity except through the decision to purchase or homestead. If this assumption holds, the distance to the nearest land office is a valid instrument to estimate the causal effect of the initial decision of how to acquire land on subsequent farm productivity and value. The most significant issue with the IV is settlers' ability to select their distance from the land office by choosing their plot of land; however, the fact that land offices were more likely to be in large towns mitigates this issue because purchasers were more likely to settle near large towns but less likely to settle near land offices.

The paper is organized as follows. Section 2 describes the model. Section 3 describes the data. Section 4 describes the results. Section 5 discusses the historical context informed by the model. Section 6 concludes.

# 2 Model

This section outlines the models used in the paper. First, I create a simple model of selection on wealth at the individual level into purchasing versus homesteading land. Second, I use that model of selection to detail my empirical specification based on the distance to the land office instrument. Third, I expand the selection model to describe the production function I estimate for purchasers and homesteaders.

## 2.1 Simple Selection Model

The decision to purchase over homesteading may select on wealth: only people who were wealthier were able to purchase and this wealth may have allowed them to put more capital into their farm, leading to it being more productive in subsequent years. To formalize this idea, I create a simple model that causes wealthier individuals to purchase land and less wealthy individuals to homestead land, based on Black et al. (2015).

Suppose the value of the farm in 1870 is a function $\upsilon$ of the improvements on the farm prior to 1870, $I$, and the land characteristics, $\ell$, which are fixed. The improvements, $I$, depend on the wealth at the time of acquisition, $\omega$, which is unobserved but exogeneously determined. Land characteristics, $\ell$, do not depend on $\omega$ because all land was available for both purchase and homestead. The value of the farm is given by

$$\textit{Value of the farm in 1870} = \upsilon(I(\omega), \ell) \tag{1}$$

The following characterizes $I$ and $\upsilon$. For any $\omega_1 < \omega_2$, $I(\omega_1) < I(\omega_2)$ (i.e., wealthier people put in more improvements), and $\upsilon$ is strictly increasing in $I$ and $\ell$ (i.e., the farm is strictly more valuable as the improvements and land characteristics increase). Therefore $\upsilon_I(I(\omega), \ell) > 0$, $\upsilon_\ell(I(\omega), \ell) > 0$, and $\upsilon_\omega(I(\omega), \ell) > 0$ (since $I$ is strictly increasing in $\omega$). Finally, $\upsilon(I(\omega), \ell)$ is a strictly concave function and $\upsilon_{I,\omega}(I(\omega), \ell) > 0$ (i.e., the marginal value of improvements is higher for wealthier farmers).

Now let $r$ be the market revenue yield to the value of the farm and $c$ be the cost per unit of farm improvement. Then individuals maximize the following:

$$U(I) = rv(I(\omega), \ell) - cI \tag{2}$$

Farmers solve $rv_I(I^*(\omega), \ell) = c$ to select their optimal level of improvement.

Now I add to the model the decision to purchase or homestead, which occurs before the improvement decision. Because purchasing the land meant that the farmer acquired the title to the land more quickly than by homesteading the same land, farmers may expect a higher market revenue yield to the value of the farm from purchasing the land than from homesteading it. I denote this difference in administrative requirements by $r_P > r_H$. Essentially, I am incorporating time discounting into the market revenue yield to the value of the farm to account for the purchasers acquiring the title to the land more quickly than homesteaders. Let $\kappa$ be the amount above the cost of homesteading that the farmer has to pay to purchase ($\kappa =$ *cost of purchasing - cost of homesteading*). Therefore, the farmer compares utility of purchase to utility of homesteading as follows:

$$U_P^* = r_P v(I_P^*(\omega), \ell) - cI_P^* - \kappa \tag{3}$$

$$U_H^* = r_H v(I_H^*(\omega), \ell) - cI_H^* \tag{4}$$

Equations (3) and (4) mean that purchasers and homesteaders choose different optimal levels of improvements, $I$. For any $\omega$, $I_P^*(\omega) > I_H^*(\omega)$, meaning that, for any level of wealth, the optimal level of improvement invested in the farm by purchasers is higher than the optimal level of improvement invested in the farm by homesteaders. This prediction of the model is consistent with what we see in the data: purchasers put in more improvements on their farms than do homesteaders, though the wealth level is unobserved.

Recall that the model predicts that wealthier individuals purchase land while less wealthy individuals homestead it. Equations (3) and (4) demonstrate this result. Following Black et al. (2015), I let $\hat{\omega}$ denote the wealth level which makes an individual indifferent between purchasing and homesteading. Such an $\hat{\omega}$ sets (3) and (4) equal. When an individual, before acquiring land, has a level of wealth $\omega$ such that $\omega \geq \hat{\omega}$, then the individual will purchase the land. Otherwise, the individual will homestead it.

This selection result makes estimating the causal impact of the decision to purchase or homestead on subsequent land outcomes more complex. For example, consider the outcome variable $y$, which denotes agricultural production in 1870 at the farm level. If the data shows the purchasers have a higher level of agricultural production than homesteaders, this could be due to the decision to purchase or homestead, or it could be due to the higher level of wealth at the time of acquisition held by purchasers. Here, I explicitly consider wealth, but other demographic characteristics at the individual level could also pose an endogeneity concern by generating both selection into purchasing versus homesteading and different subsequent land outcomes at the farm level.

In order to deal with this endogeneity concern, I use $\kappa$, the difference in the administrative costs between purchasing and homesteading. Recall that $\kappa =$ *cost of purchasing - cost of homesteading.* Therefore, a change in $\kappa$ shifts the threshold between purchasing and homesteading: when $\kappa$ increases, individuals are more likely to homestead. When $\kappa$ decreases, individuals are more likely to purchase. To purchase, individuals had to go to the local land office once. To homestead, they had to go twice. When $\kappa$ decreases - i.e., when the land office is farther away from the selected land - purchasing becomes more attractive relative to homesteading.

Further, I test for how selection in to purchasing versus homesteading based on wealth might cause an endogeneity problem when considering subsequent farm outcomes like agricultural production. In the simple model above, after selecting a plot of land, individuals decide whether to purchase or homestead it based on their wealth level. Therefore, following Black et al. (2015), I divide the plots of land into two groups: those that are farms "near" the local land office and those that are farms "far from" the local land office. When the plot of land is far from the land office, individuals must be relatively more wealthy to homestead that plot than to homestead a plot near to the land office because the farmer has to travel to the land office twice when they homestead land, and that travel is relatively more expensive when their farm is farther from the land office than when it is closer. Let $\omega_N$ be the wealth level needed to purchase land when individuals choosing farms near the local land office, and let $\omega_F$ be the wealth threshold to purchase for individuals choosing farms far from the local land office, such that $\omega_F > \omega_N$ because of travel costs. Then I can divide the population that acquired land into three groups: 1) farmers with wealth below $\omega_N$ will never purchase/will always homestead (group $H$); 2) farmers with wealth above $\omega_F$ will always purchase/will never homestead (group $P$); and 3) farmers with wealth $\omega_F > \omega > \omega_N$ will purchase if they are on land far from the land office and will homestead if they are on land near to the land office (group $C$).

Finally, to test for selection into purchasing versus homesteading at the individual level based on wealth, I let $X = 1$ denote the decision to purchase and $X = 0$ denote the decision to homestead. Recall that $y$ denotes the land outcome variable, so call $Y_{X=1}$ the value of the farm in 1870 if the individual purchases, and call $Y_{X=0}$ the value of the farm in 1870 if the individual homesteads. To clarify, $E[Y_{X=1}|P]$ represents the expected value of the farm in 1870 conditional living close to a land office but still purchasing, and $E[Y_{X=0}|H]$ represents the expected value of the farm in 1870 living far from the land office but still homesteading. $E[Y_{X=1}|C]$ represents the expected value of the farm in 1870 if they purchased for those whose acquisition decision is changed by the instrument (distance to the land office), and $E[Y_{X=0}|C]$ represents the expected value of the farm in 1870 if they homesteaded for those moved by the instrument. I estimate $E[Y_{X=1}|P]$ and $E[Y_{X=0}|H]$, and I can recover $E[Y_{X=1}|C]$ and $E[X_{X=0}|C]$ under the following assumptions from Black et al. (2015): 1) the instrument (being close to a land office) shifts some individuals on the margin to switch their decision about how to acquire land from purchasing to homesteading (because homesteading becomes relatively cheaper as an individual lives closer to a land office); and 2) being close to a local land office is statistically independent the joint land outcome variable $(Y_{X=1}, Y_{X=0})$.

Under the above assumptions, I can test two predictions from the model:

$$E[Y_{X=1}|C] < E[Y_{X=1}|P] \tag{5}$$

$$E[Y_{X=0}|H] < E[Y_{X=0}|C] \tag{6}$$

Equation (5) considers only farmers who purchased land ($X = 1$), while equation (6) considers only farmers who homesteaded land ($X = 0$). For farmers who purchased land, equation (5) predicts that farmers who live close to a land office produce a higher subsequent value of agricultural output on their farm than farmers who would have purchased land (instead of homesteading it) if they had lived closer to the land office. This is because, of the subset of the population that purchases, those who live closer to a land office should, on average, have higher wealth than those who live far from a land office ($\omega_F > \omega > \omega_N$). Likewise, for farmers who homesteaded land, equation (6) predicts that farmers who live far from the nearest land office produce a lower value of agricultural output on their farm than farmers who would have homesteaded if they had lived farther away. This is because, of the subset of the population that homesteads, those who live far away from a land office should, on average, have lower wealth than those who live near a land office ($\omega_F > \omega > \omega_N$). If these two testable predictions of the model hold, it would indicate that there is in fact selection into homesteading and purchased based on unobserved wealth at the individual level.

## 2.2 Instrumental Variable Specification

In this section, I outline the empirical specifications I use to estimate the causal effect of the method of acquisition (i.e., whether the farm was purchased or homesteaded) on the subsequent farm productivity and value at the individual level. I use the simple selection model above to create an instrument for the decision to purchase or homestead based on the distance to the land office.

To represent subsequent farm output, I use two dependant variables: value per acre and production per acre. Value per acre is the assessed value of the farm by the Kansas Agricultural Census divided by the acres of the farm. Production per acre is defined as follows:

$$Prod_i = \frac{livestock\ value_i + (bushels\ of\ crops_i \times crop\ price)}{acres_i} \tag{7}$$

Production per acre measures the amount of farm output and consumable investment like crops and livestock created by the farm. The value per acre is a function of the assessed value of the farm, which measures durable investment like fences and buildings. Investment in either farm production or farm value represents different farming strategies, particularly because durable investments like fences and buildings were sold with the farm, while production was not.

Consider a linear model in which the independent variable is the decision to purchase or homestead and the dependent variable is subsequent farm output. The primary relationship I will be estimating is that of originally purchasing or homesteading farm $i$ on the type and extent of farm production of farm $i$ in 1870 and 1880. The level of observation is the farm. The model is of the form:

$$Y_{ijkt} = \alpha_{jt} + \gamma_{kt} + \beta(Purchase_i) + X_i\delta + \epsilon_{ijkt} \tag{8}$$

$Y_{ijkt}$ is a continuous variable which represents total production per acre or total value per acre on farm $i$ in year $t$ (either 1870 or 1880). $Purchase_i$ is a binary variable indicating if farm $i$ was purchased (= 1) or homesteaded (= 0), $\alpha_{jt}$ are county-year fixed effects, and $\gamma_{kt}$ are township-range-year fixed effects. $X_i$ is the vector of controls discussed above, including soil type, farm value, days since acquisition, distance to town, population of town, gender, race, age, the presence of water, speculation in the county, unsold land in the township-range, census enumerator, latitude and longitude of the farm, acres of the farm, distance to the nearest land office, years since the survey, and the tax rate in the county. Standard errors are clustered by township-range.

$\beta$ is the parameter of interest. It estimates the relationship between purchasing a farm and the value of products produced on the farm in subsequent years. Farm productivity is a function of land characteristics at the farm level and farmer characteristics at the individual level. In the linear model, the most significant threat to identification is that a higher production or value on homesteaded or purchased farms may reflect selection into that group. By using the IV approach generated by the simple selection model, I can greatly reduce this issue by exploiting variation in administrative travel costs created by the differing administrative requirements of purchasing and homesteading. I use the distance to the nearest land office active at the time as an instrument for the decision to purchase or homestead because homesteaders were required to go to the land office twice while purchasers only had to go once. The average distance to the nearest land office in this sample was 28 miles, which represented a significant travel cost in the late 1800's because it took more than 5 hours to travel that far on horseback. Therefore, going to the land office was at least a one or two day trip. The unconditional correlation cofficient between the method of acquisition (where $Purchase = 1$) and $LandOfficeDist_{it}$ is 0.79: when farms are farther away from the land office, the individual is more likely to purchase.

The identifying assumption of the IV model is that the total value or production per acre of the farm is only related to the distance to the nearest land office through the decision to purchase or homestead, after controlling for land and farmer characteristics. The biggest concern about the validity of the IV is individuals' ability to select their distance from the land office by choosing their plot of land; however, the fact that land offices were more likely to be in large towns mitigates this issue because purchasers were more likely to settle near large towns but less likely to settle near land offices.

The level of observation is the farm. I estimate the first stage of my IV model as follows:

$$Purchase_{ijkt} = \alpha_{jt} + \gamma_{kt} + \beta_1(LandOfficeDist_{it}) + X_i\delta_1 + \eta_{ijkt} \tag{9}$$

$\beta_1$ represents the relationship between farm $i$'s acquisition method (purchase or homestead) and the distance to the land office. $Purchase_{ijkt}$ is a binary variable which predicts the method of acquisition of farm $i$. $X_i$ is a vector of controls including soil type, days since of acquisition, distance to town, population of town, gender, race, age, place of origin, the presence of water, speculation in the county, unsold land in the township-range, latitude and longitude of the farm, year of acquisition, acres of the farm, distance to the nearest land office, years since the survey, precipitation, farm value, census enumerator, and tax rate in the county by decade. County-year and township-range-year fixed effects are represented by $\alpha_{jt}$ and $\gamma_{kt}$, respectively. Standard errors are clustered by township-range.

To estimate the second stage, I use the first-stage model to derive the predicted method of acquisition of each farm $i$ from its surrounding farms and use it to create the variable $Pur\hat{c}hase_{it}$. This variable represents the estimated method of acquisition of farm $i$ from the surrounding previously acquired farms. Then I estimate the following second stage model using the predicted method of acquisition of farm $i$:

$$Y_{ijkt} = \alpha_{jt} + \gamma_{kt} + \beta_2(Pur\hat{c}hase_{it}) + X_i\delta_2 + \epsilon_{ijkt} \tag{10}$$

The primary effect I am estimating is the effect of the initial decision to purchase or homestead on the production per acre or value per acre in 1870 or 1880 at the individual farm level. This effect is captured by $\beta_2$. $Y_{ijkt}$ is the production per acre or value per acre in 1870 or 1880 of farm $i$.

## 2.3 Heterogeneous Production Functions

In this section, I develop a model of how heterogeneous production functions among farmers leads to both different investment decisions on the farm and different land acquisition decisions. The general idea of this model is to show why some farmers initially invest in improvements and why others initially invest in crops and livestock, and how that exogenous difference in production functions leads to selection into purchase or homesteading. This model expands on the selection model from Section 2.1.

This model shows that the exogenous, innate difference in production and value functions among farmers can cause them to select into purchase and homesteading. Once they did so, the same exogenous difference caused them to choose specific types of investment - either investment into improvements, $I$, or investment into consumables, $s$.

I first build a model of the two ways farmers can accrue returns to their farm: 1) physical improvements such as fences and buildings, and 2) production such as crops and livestock. Farmers each have a value function $\upsilon$ and a production function $\rho$. The value of the farm is a function $\upsilon$ of the improvements to the farm such as fences and buildings ($I$) and the land characteristics ($\ell$).

$$\textit{Value of the farm} = \upsilon(I, \ell) \tag{11}$$

The production of the farm is a function $\rho$, which depends on $\ell$, and crop and livestock-specific investments such as seeds $s$.

$$\textit{Value of production} = \rho(\ell, s) \tag{12}$$

Both $\upsilon$ and $\rho$ are strictly increasing and strictly concave. Now let $r_1$ be the market revenue yield to the value of the farm and $c_1$ be the cost per unit of farm improvement. Let $r_2$ be the market revenue yield to production and let $c_2$ be the cost per unit of crop and livestock-specific investments.

$$\textit{Utility of farm} = U(I, \ell) = r_1 \upsilon(I, \ell) - c_1 I \tag{13}$$

$$\textit{Utility of production} = U(\ell, s) = r_2 \rho(\ell, s) - c_2 s \tag{14}$$

Intuitively, $r_1$ represents the returns to selling the farm (the market value of the farm), and $r_2$ represents the returns to growing and selling or eating crops and livestock (the market value of crops and livestock). Therefore, the farmer has the following total utility function, which is the sum of the utility they get from the value of the farm and the utility they get from the production on the farm:

$$U(I, \ell, s) = U(I, \ell) + U(\ell, s) \tag{15}$$

The farmer has a limited amount of wealth $w$ and can either invest in improvements like fences and buildings (invest in $I$) or in production like crops and livestock (invest in $s$) or in a combination.

$$\textit{Budget constraint} = B(w, \boldsymbol{c}) = w - c_1 I - c_2 s \tag{16}$$

The farmer chooses the ratio of the level of investment in $I$ and $s$ that maximizes $U(I, \ell, s)$ subject to their budget constraint $B(w, \boldsymbol{c})$. Therefore, the farmer solves the constrained maximization problem:

$$\max_{I,s} U(I, \ell, s) = U(I, \ell) + U(\ell, s) = r_1 \upsilon(I, \ell) - c_1 I + r_2 \rho(\ell, s) - c_2 s \tag{17}$$

$$s.t.$$

$$B(w, \boldsymbol{c}) = w - c_1 I - c_2 s \tag{18}$$

$$c_1 > 0, c_2 > 0, I \geq 0, s \geq 0 \tag{19}$$

The farmer chooses $I^*$ and $s^*$ in order to maximize their total utility function subject to their budget constraint. Both $I$ and $s$, which are units of investment, are functions of wealth, $\omega$: $I(\omega)$ and $s(\omega)$. Let $D_I = c_1 I$ be the dollar amount invested in $I$ and let $D_s = c_2 s$ be the dollar amount invested in $s$. Let $D$ be the level such that $D_s = D_I$.

Now consider the initial decision to purchase or homestead. Recall that $\kappa$ is the difference between the cost of purchasing and the cost of homesteading, i.e., $\kappa =$ *cost of purchase - cost of homestead*. This assumption reflects that fact that purchasing was more expensive than homesteading. Further, recall the administrative requirement that purchasers gain the title more quickly than homesteaders. This means that $r_1^P > r_1^H$: the returns to investing in improvements like fences and buildings are higher for purchasers than for homesteaders. This is because purchasers gain the title more quickly, enabling them to sell or mortgage their farm more quickly than homesteaders, so purchasers' returns to durable farm improvements are higher than homesteaders'. Essentially, I am implicitly modeling time discounting in their returns to improvements. I assume the $c_1$ and $c_2$ are the same for purchasers and homesteaders, meaning that they face the same market costs. Likewise, I assume that $r_2$ (the market revenue yield to selling crops and livestock) is the same for purchasers and homesteaders. This assumption is based on the fact that there were many small farmers who could not individually impact the market price.

Not all farmers have the same production function or value function. This heterogeneity reflects individual's differing abilities, goals, and resources. Note that a farmer's ability to grow crops or build fences may proxy for other exogenous qualities like wealth or goals, but here I will refer to this whole category as ability. Production and value functions are exogenous to the model and cannot be chosen by the farmer. Consider a farmer $i$ with $\upsilon_i(\ell, D) > \rho_i(\ell, D)$: at any given level of investment, this farmer is better at building fences than planting crops. For any dollar amount farmer $i$ invests in $I$, they will get a higher value of farm products than if they invest the same dollar amount in $s$ because they are relatively more efficient at building durable investments than consumables. This farmer $i$ has the following utility function:

$$U_i(\ell, D) = r_1 \upsilon_i(\ell, D) - D + r_2 \rho_i(\ell, D) - D \tag{20}$$

**Proposition 1** *Let $\upsilon_i(\ell, D') > \rho_i(\ell, D)$ for all $D' \geq D$ and let $r_1^P > r_1^H$. Let $\upsilon$ and $\rho$ both be strictly increasing in $D$ and $\ell$ and strictly concave. Let $r_1^P \upsilon_i(\ell, D) > D + \kappa$, $r_1^H \upsilon_i(\ell, D) > D$, and $r_2 \rho_i(\ell, D) > D$ for all $D$. Then:*

1. *Farmer $i$ will purchase, meaning $U_i^P(\ell, D_P^*) > U_i^H(\ell, D_H^*)$*
2. *$\frac{I_P^*}{s_P^*} > \frac{I_H^*}{s_H^*}$, meaning that the ratio of improvements to seeds is greater for purchasers than homesteaders*

Intuitively, this proposition means that if farmer $i$ is better at building fences than planting crops and if the returns to investing in improvements like fences are higher for purchasers than for homesteaders, then farmer $i$ will choose to purchase and choose to invest more money in fences than seeds. Proofs are in the appendix.

Having different exogenous production and value functions causes farmers to select into purchasing or homesteading. Then, once they have selected into purchasing or homesteading, those same factors cause them to separate into those who invest in durable investments (purchasers) and those who invest in consumable investments (homesteaders). Therefore, this model provides an explanation for why purchased and homesteaded land may differ in future periods without relying on the explanation that differences between purchased and homesteaded land are caused by the initial method of acquisition. Instead, this model includes and specifically highlights the effect of selection and how it can partition farmers based on truly exogenous factors (innate production and value functions).

I can use a standard Cobb-Douglas production function to test this idea of heterogeneous production and value functions. First, I estimate the following production function separately for purchasers and homesteaders:

$$Y_{it} = AI_{it}^{\beta_1} S_{it}^{\beta_2} e^{\omega_{it}} \tag{21}$$

where $I_{it}$ represents improvements and $S_{it}$ represents crop and livestock-specific inputs such as seeds. Taking the natural log yields:

$$y_{it} = \beta_0 + \beta_1 improvements_{it} + \beta_2 seeds_{it} + \omega_{it} + \epsilon_{it} \tag{22}$$

where $y_{it} = \ln(Production)$, $\beta_0$ is a constant common to all farms, $\omega_{it}$ are farm-level controls like land fertility which are observed by the farm, $improvements_{it}$ are improvements to the farm such as fences and buildings, and $seeds_{it}$ are the inputs such as seeds and fertilizer that are crop and livestock-specific.

I use the following linear specification to test if the coefficients of interest, $\beta_1$, $\beta_2$, and $\beta_3$, are statistically different for purchasers and homesteaders by running the following regression separately for each group:

$$\ln(Prod_{ijkt}) = \beta_0 + \beta_1 FarmValue_i + \beta_2 \frac{FertilizerCost_i}{Acres_i} + \beta_3 MachineryValue_i + \delta X_{ijkt} + \theta W_i + \alpha_{jt} + \gamma_{kt} + \epsilon_{ijkt} \tag{23}$$

$FarmValue_{ijkt}$ is the assessed value of farm $i$ in the 1870 or 1880 Census. It is used as a measure of $I_{it}$. As with the IV regressions, $Prod_{ijkt}$ is defined as in equation 7. $X_{ijkt}$ is a vector of farm characteristics including wages paid, $W_i$ is a vector of individual demographic characteristics, and $\alpha_{jt}$ and $\gamma_{kt}$ are county-year and township-range-year fixed effects. Together, $W_i, X_{ijkt}, \alpha_{jt}$ and $\gamma_{kt}$ represent $\omega_{it}$ in equation 22. $\frac{FertilizerCost_i}{Acres_i}$ and $MachineryValue_i$ represent $seeds_{it}$ in equation 22 because these are inputs only used when planting crops.

The above regression estimates the production functions of purchasers and homesteaders separately, enabling a test to determine if the coefficients of interest, the $\beta_j$'s, are statistically different between purchasers and homesteaders. If $\beta_j$'s, are statistically different, this result would indicate that purchasers and homesteaders do in fact have different production functions. If the $\beta_j$ differ between purchasers and homesteaders, it indicates that the two groups have ways of producing crops and livestock on their farms, meaning they have different production functions.

I use a similar method to test if purchasers and homesteaders have heterogeneous value functions. Again, I use a standard Cobb-Douglas production function:

$$V_{it} = AI_{it}^{\beta_1} L_{it}^{\beta_2} e^{\omega_{it}} \tag{24}$$

$I_{it}$ again represents improvements like fences and $L_{it}$ represents land characteristics. Taking the natural log yields:

$$v_{it} = \beta_0 + \beta_1 improvements_{it} + \beta_2 land_{it} + \omega_{it} + \epsilon_{it} \tag{25}$$

More formally, I can use the following linear specification to test if the $\beta_j$'s are statistically different for purchasers and homesteaders:

$$\ln(FarmValue_{ijkt}) = \beta_0 + \beta_1 \frac{TilledAcres_i}{Acres_i} + \beta_2 \frac{PastureAcres_i}{Acres_i} + \beta_3 \frac{FenceCost_i}{Acres_i} + \delta X_{ijkt} + \theta W_i + \alpha_{jt} + \gamma_{kt} + \epsilon_{ijkt} \tag{26}$$

For the value function above, the assessed value of the farm is the output of the value function, while tilled acres, pasture acres, fence costs, and the $X_{ijkt}$ and $W_i$ controls are the input variables, including wages paid. The above equation describes how the improved acres of the farm and the fences create value. If the $\beta_j$ differ between purchasers and homesteaders, that indicates that purchasers and homesteads have different ways that they create value using the same variables, i.e., they have different value functions.

# 3 Data

I use three main individual-level data sources: the individual Kansas Agricultural Census, which details individual farm production, the Bureau of Land Management tract books, which record individual purchase and homestead decisions, and the United States Population Census, which records individual demographic characteristics. Using individual-level farm data as opposed to the county-level farm data generally used in previous literature allows me to estimate the heterogeneity in production functions of farmers. Linking to the individual United States Population Census allows for many more individual demographic controls than any previous literature in the area, and also allows me to empirically demonstrate selection into purchase and homesteading on observable characteristics.

Kansas was chosen for analysis in part because of the availability of the Kansas Agricultural Censuses. These censuses are state-level censuses, separate from the U.S. Agricultural Census. They are notable because the Kansas Agricultural Census was conducted at the individual farm level, with the name of each farm owner, whereas the federal agricultural censuses are county-level.

The Kansas Agricultural Census contains the following variables: name, township, county, the exact date that the census was taken, enumerator, acres of land on the farm (subdivided into improved and unimproved acres), farm value (including land, fences and buildings), farm machinery value, livestock value, wages for farm labor in the year, bushels of different types of crops produced, and number of different types of livestock. It does not include demographic controls, such as age, race, or wealth of the owners.

The Census took place in June through August of the census year, and the crop data reflect the previous year harvest. Notably, of farmers who acquired land prior to the census year, only successful farmers were included in the Census because the farmers who were unsuccessful would have resold or abandoned their land.

The Bureau of Land Management (BLM) keeps the original scanned tract books used at each land office in Kansas to record the acquisition of land. The tract books record all original acquisitions of land from the federal government. Like the Kansas Agricultural Census, the level of observation is the farm. The tract books record both successful and unsuccessful homestead claims and purchases, including the original date filed, the date the patent was applied for, the date the patent was acquired, the name of the owner, the size of the acreage, the price per acre, the Act it was filed under, and the state, county, county-section, township-range (PLSS), section, and quarter-section[1]. The purchases in the tract books contain only new, unbroken land sold, not land already used for farming that changed hands. Since homestead land was all new, unbroken land, it is not comparable to land that had been farmed for several years.

Finally, the above two datasets are linked together and then to the United States Population Censuses. These Censuses detail individual-level information about age, race, wealth, occupation, place of origin, family structure, and other demographic characteristics in every decade.

Additional information about these data and controls such as farm-level historical water sources, farm-level soil quality classification, historical precipitation level, distances to and populations of surrounding towns, and other land characteristics can be found in Appendix A.

# 4 Results

This section details the results, breaking them down into three parts. First, I demonstrate selection into purchasing and homesteading at the individual level on demographic characteristics. Second, I present the results from the IV model. Finally, I estimate the Cobb-Douglas production functions separately for purchasers and homesteaders to show that they produce value and goods differently. Additionally, I provide a robustness check on my empirical results using another instrument for farmer $i$'s method of acquisition.

## 4.1 Evidence of Selection into Homesteading

I begin by testing the predictions of the simple selection model outlined in section 2.1 in order to determine if selection into homesteading is a significant issue. From the model, I expect that:

$$E[Y_{X=1}|C] < E[Y_{X=1}|P] \tag{27}$$

$$E[Y_{X=0}|H] < E[Y_{X=0}|C] \tag{28}$$

where $Y$ represents the farm output per acre in 1870 and $X$ represents the decision to purchaser or homestead. If these inequalities hold, it would indicate that there exists selection into purchase and homesteading. We expect that among individuals who purchase ($X = 1$), those who live close to a land office should have higher subsequent farm productivity than individuals who live far from land offices who would have purchased if they had lived closer. Conversely, we expect that among individuals who homestead ($X = 0$), those who live far from a land office should have lower subsequent farm productivity than individuals who live near a land office.

In order to test the predictions, I define the value for always purchasers (purchase even when they are near to a land office) as

$$E[ValuePerAcre|\textit{Land Office Dist} \leq 28, Purchase = 1] = 24.40 = E[Y_{X=1}|P] \tag{29}$$

and the value for never purchasers (homestead even when they are far from a land office):

$$E[ValuePerAcre|\textit{Land Office Dist} > 28, Purchase = 0] = 15.69 = E[Y_{X=0}|H] \tag{30}$$

[1]A township-range is 36 square miles, with 36 sections in it. Each of these sections can be sub-divided, and a quarter-section is 160 acres. Sections 16 and 36 were reserved for future sales to support schools.

The average distance from a farm to a land office is 28 miles, and 28 miles is used as the cut off for being "near" or "far" from a land office.

To find the probability of being in each group (group $C$, group $P$, and group $H$), I use the first stage of the IV regression, as defined in equation 9. The coefficient on *Land Office Distance* is equal to $P_C$, the probability of being in group $C$. The probability of being in group $P$ is given by the constant term in the first stage.[2] Therefore, $P_C = 0.42$ and $P_P = 0.26$. The nature of the instrument assumes that all individuals are either in group $P$, group $H$, or group $C$; therefore, $P_H = 1 - P_P - P_C = 0.32$.

Next, I back out $E[Y_{X=1}|C]$ and $E[Y_{X=0}|C]$ as follows:

$$E[ValuePerAcre|\textit{Land Office Dist} > 28, Purchase = 1] = 19.49 = \frac{P_P \times E[Y_{X=1}|P] + P_C \times E[Y_{X=1}|C]}{P_P + P_C} \tag{31}$$

$$E[ValuePerAcre|\textit{Land Office Dist} \leq 28, Purchase = 0] = 17.18 = \frac{P_H \times E[Y_{X=0}|H] + P_C \times E[Y_{X=0}|C]}{P_H + P_C} \tag{32}$$

This gives $E[Y_{X=1}|C] = 16.45$ and $E[Y_{X=0}|C] = 18.12$. $E[Y_{X=1}|C]$ is the value per acre for those moved by the instrument (distance to the land office) who purchase and $E[Y_{X=0}|C]$ is the value per acre for those moved by the instrument who homestead. Therefore, I can now test the predictions of the selection model:

$$16.45 = E[Y_{X=1}|C] < E[Y_{X=1}|P] = 24.40 \tag{33}$$

$$15.69 = E[Y_{X=0}|H] < E[Y_{X=0}|C] = 18.12 \tag{34}$$

Since the inequalities predicted by the model hold, there is likely to be selection into purchase and homesteading based on wealth. The selection model predicts that those who always purchase should have a higher value of their farm than those who only purchase because they are far from a land office, and that those who always homestead should have a lower value of their farm than those who only homestead because they are close to a land office. This is consistent with the results from the data.

The same test can be used with $Y$ equal to the production per acre of the farm in 1870. I find similar results, again indicating selection:

$$17.72 = E[Y_{X=1}|C] < E[Y_{X=1}|P] = 23.06 \tag{35}$$

$$7.81 = E[Y_{X=0}|H] < E[Y_{X=0}|C] = 10.18 \tag{36}$$

These results together support the idea that there exists individual selection into purchasing and homesteading, likely based on initial wealth or other unobservable individual characteristics at the time of land acquisition. This selection presents a challenge in identifying the causal impact of the method of land acquisition on subsequent farm outcomes.

I further explore the effect of individual characteristics by demonstrating that adding individual-level demographic characteristics as controls changes the coefficients of interest. The previous literature has been

[2] The coefficients used here differ from the results in Table 3 because Table 3 uses the distance to the land office as a continuous instrument and here I use it as a discrete instrument where the cutoff is 28 miles (the average distance to the land office in the data).

largely unable to control for individual characteristics such as family size or birthplace because it has used county-level data. Therefore, it was limited to county-level controls such as the percent of immigrants in the county or the average age in the county. This method of controlling for demographic characteristics does not allow for the idea that individuals with different demographic characteristics selected into purchasing and homesteading, and that those same demographic characteristics may effect their subsequent farm outcomes.

I demonstrate the effect of individual characteristics on farm outcomes by adding individual-level demographic characteristics one at a time as controls. I use the following linear model:

$$Y_{ijkt} = \alpha_{jt} + \gamma_{kt} + \beta Purchase_i + \delta X_{ijkt} + \theta W_i + \epsilon_{ijkt} \tag{37}$$

$Y_{ijkt}$ indicates the value per acre or production per acre of farm $i$ in 1870. $\beta$ is the coefficient of interest, which gives the relationship between the initial decision to purchase or homestead land ($Purchase_i = 1$) on the dependent variable. $X_{ijkt}$ is a vector of land controls and $\alpha_{jt}$ and $\gamma_{kt}$ are county-year and township-range-year fixed effects. $W_i$ is a vector of individual controls which are added one by one to determine how their inclusion changes $\beta$.

Table 1: Individual Controls in OLS Estimation of Acquisition Method on Production and Value Per Acre

| | Production | Value | Production | Value | Production | Value | Production | Value |
|---|---|---|---|---|---|---|---|---|
| Purchase=1 | 6.62*** | 6.30** | 4.37** | 3.59** | 3.81* | 3.94** | 4.89** | 4.52*** |
| | (1.08) | (2.83) | (1.02) | (1.22) | (2.06) | (1.80) | (2.72) | (1.96) |
| Female=1 | No | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Immigrant=1 | No | No | No | No | Yes | Yes | Yes | Yes |
| Family Size | No | No | No | No | No | No | Yes | Yes |
| Observations | 1516 | 1516 | 1516 | 1516 | 1516 | 1516 | 1516 | 1516 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From Table 1, it is clear that adding individual characteristics $W_i$ as controls changes the coefficient on $Purchase_i$, the relationship between the initial land acquisition decision and the farm outcomes in 1870. This change indicates that individual characteristics are a significant predictor of farm outcomes. Demographic characteristics may or may not be a causal mechanism behind differing farming outcomes: individual characteristics may cause farms to have a higher or lower value, or it may simply be that some individual characteristics are correlated with the initial decision to purchase or homestead. Being unable to control for such characteristics, as in the previous literature which uses county-level data, casts doubt on the ability to identify the causal mechanism behind subsequent land outcomes such as farm value or productivity. Moreover, the coefficient of interest decreases when adding individual characteristics, indicating that those characteristics are absorbing some of the effect of the decision to purchase or homestead on farm outcomes. This result suggests selection.

Above, I have demonstrated that the coefficient of interest may change by adding observable individual characteristics at the individual level. This is consistent with the hypothesis of individual selection into purchase and homesteading based on individual characteristics. These characteristics, such as age or birthplace, can be controlled for by adding them to a regression, thus eliminating their selection effect. However, more important are the unobservable individual characteristics which are not included in the U.S. Population Census, such as wealth directly before land acquisition, health, goals, and motivation, because these characteristics cannot be controlled for and thus may create a selection effect. There exists selection based on observable individual characteristics, and I cannot rule out that there exists selection on unobservable individual characteristics as well. This possibility necessitates the use of the distance to land office IV strategy described in section 2.2 to be able to identify the effect of the method of acquisition on farm outcomes.

## 4.2 IV Estimation Results

In this section, I presents results of the estimation of the causal effect of the method of acquisition on farm productivity and value using the distance to the nearest land office as an instrument for the decision to purchase or homestead land. First, I demonstrate that different farming strategies are associated with purchase and homesteading using a linear model. Then, I show that theses different strategies are not directly caused by the *method* of acquisition itself by using IV estimation to control for selection. OLS estimates show differences in farming strategies, but IV estimates do not, indicating that another mechanism besides the acquisition method drives this difference.

Table 2 presents results from a linear model which relates the decision to purchase and homestead to farm production per acre or value per acre at the individual level. This is the model from equation 8, which includes controls at the individual level for soil type, farm value, days since acquisition, distance to town, population of town, gender, race, age, the presence of water, speculation in the county, unsold land in the township-range, census enumerator, latitude and longitude of the farm, acres of the farm, distance to the nearest land office, years since the survey, and the tax rate in the county. From Table 2, purchasing is a significant predictor of having a more valuable farm, no matter when the farm was purchased. For farms acquired well before the Agricultural Census (acquired in 1863-1868 or 1871 to 1878), purchased farms were both more productive and more valuable than homesteaded farms by the time both reached the Census year. By necessity I observe only moderately successful farmers who survived until 1870. Both purchased farms and homesteaded farms in columns 1, 2, 5, and 6 of Table 2 were by definition moderately successful because they survived multiple years to the Census year. For these successful farmers who remained on the land past eighteen months, purchasers were more productive than homesteaders.

However, when considering only newly acquired farms which may or may not become established and successful in the future (i.e., farms acquired just before the census year), homesteads were in fact more productive than purchased farms, even though these homesteads had a lower assessed farm value per acre (less investment in fences and buildings). This lower level of investment in durable farm improvements by homesteaders persisted past the first few years for the homesteaders and purchasers that remained on the land. However, in this time, the purchasers became more productive in terms of farm output than the homesteaders, reversing their initial trend.

To summarize, consistent with the idea that purchasers and homesteaders were fundamentally different types of people with different goals, purchasers initially invested in buildings and fences and homesteaders initially invested in crops and livestock. After several years, the purchasers who remained on the land were both more productive and had a more valuable farm than the homesteaders who remained on the land.

Table 2: OLS Estimation of Acquisition Method on Production and Value Per Acre

| | Acquired 1863 to 1868 | | Acquired 1869 and 1870 | | Acquired 1871 to 1878 | | Acquired 1879 and 1880 | |
|---|---|---|---|---|---|---|---|---|
| | Production | Value | Production | Value | Production | Value | Production | Value |
| Purchase=1 | 4.52*** | 4.30*** | -3.46** | 2.23** | 8.98** | 12.11** | -6.18*** | 8.35** |
| | (1.16) | (1.61) | (1.04) | (0.75) | (3.94) | (4.13) | (2.09) | (3.38) |
| Observations | 479 | 479 | 383 | 383 | 398 | 398 | 256 | 256 |

Standard errors in parentheses

Includes individual level demographic and land controls

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Successful purchasers who stayed on the land had a different strategy than successful homesteaders.

However, there could exist selection into purchasing or homesteading on unobservables which impact farm value or production, such as wealth. Therefore, I used the IV method generated from the simple selection model.

Table 3: First Stage

| | Distance to Land Office | Neighbors' Purchase Decisions |
|---|---|---|
| | Purchase=1 | Purchase=1 |
| Instrument | 0.0078** | 0.827*** |
| | (0.001) | (0.030) |
| F-stat | 30.88 | 101.44 |
| Observations | 956 | 848 |

Standard errors in parentheses

Includes individual level demographic and land controls

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 presents the results from equation 9, which describes the first stage of the IV estimation and includes controls for soil type, farm value, days since acquisition, distance to town, population of town, gender, race, age, the presence of water, speculation in the county, unsold land in the township-range, census enumerator, latitude and longitude of the farm, acres of the farm, distance to the nearest land office, years since the survey, and the tax rate in the county. From Table 3, we can see that both instruments are significant (F-stat > 10), but that the neighbors' purchase decisions instrument (discussed in the appendix) is a stronger instrument. However, the distance to the land office instrument is likely to have fewer defiers and is therefore more theoretically justified. I use the neighbor instrument to provide robustness checks.

While we see highly significant differences between purchasers and homesteaders in terms of farm value per acre and farm production per acre in the linear regression, we fail to find those same differences in the IV estimation (Table 4), which factors out the effect of selection. Note that the results from Table 4 include controls for soil type, farm value, days since acquisition, distance to town, population of town, gender, race, age, the presence of water, speculation in the county, unsold land in the township-range, census enumerator,

Table 4: Distance IV Estimation of Acquisition Method on Production and Value Per Acre

| | Acquired 1863 to 1868 | | Acquired 1869 and 1870 | | Acquired 1871 to 1878 | | Acquired 1879 and 1880 | |
|---|---|---|---|---|---|---|---|---|
| | Production | Value | Production | Value | Production | Value | Production | Value |
| Purchase=1 | 3.36 | 2.18 | 4.90 | 1.87 | -5.24 | 10.09 | -4.68 | 8.31 |
| | (6.92) | (7.21) | (3.77) | (6.48) | (8.64) | (7.09) | (11.13) | (8.10) |
| Observations | 479 | 479 | 383 | 383 | 398 | 398 | 256 | 256 |

Standard errors in parentheses

Includes individual level demographic and land controls

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

latitude and longitude of the farm, acres of the farm, distance to the nearest land office, years since the survey, and the tax rate in the county at the individual level. None of the coefficients from the distance to the nearest land office IV are significant, and several change sign from the OLS results. The IV estimation does not support the idea that the method of acquisition - purchase or homesteading - causes the differences in farming strategies seen in the linear model. If there exists selection into purchasing or homesteading on unobservables, the IV estimates will be smaller than the OLS estimates, which is what we see in Table 4. While the OLS estimates indicate that purchasers initially invest in durable investments such as buildings and fences, while homesteaders initially invest in consumable products like crops and livestock, the IV estimates do not support this difference in farming strategies between the two groups. While there exists differences in farming strategies and investment decisions for purchasers and homesteaders, this difference disappears after factoring out the effect of selection into either group.

Further, I provide a robustness check for the IV results described here in the appendix. I develop another instrument for farmer $i$'s decision to either purchase or homestead land. This instrument is the average acquisition method of the farms immediately surrounding and preceding farm $i$. Farm plots were structured in a grid, so I use the method of acquisition (purchase or homestead) of the 8 surrounding farms to predict the method of acquisition of the 1 farm in the middle. This instrument provides results consistent with the results from the distance to the land office instrument: the neighbor IV estimates do not show different farming strategies between purchasers and homesteaders. The IV estimates are smaller than the OLS estimates, indicating selection. While there exists differences in farming strategies and investment decisions for purchasers and homesteaders, this difference disappears after factoring out the effect of selection into either group.

To summarize, the OLS estimates show that farming strategies do differ between purchasers and homesteaders: homesteaders initially invest in crops and livestock, while purchasers initially invest in fences and buildings. However, the IV results, which factor out the impact of individual selection, do not reflect this. While there are differences in farming strategies between purchasers and homesteaders, those differences do not remain after controlling for selection, and in fact those differences are likely to be a result of selection into land acquisition decisions. Therefore, the causal mechanism behind the differences in farming strategies between purchasers and homesteaders was not the method of acquisition itself. The fact that the OLS estimates show significant differences between purchasers and homesteaders while the IV results do not indicates that the causal mechanism behind the difference in farming strategies occurred before the purchase-homestead decision.

## 4.3 Heterogeneous Production Function Results

In the previous section, I showed that purchasers initially invest in durable investments such as buildings and fences, while homesteaders initially invest in consumables like crops and livestock, but that the decision to purchase or homestead is not the causal reason for this difference. In this section, I present results from the heterogeneous production functions model from section 2.3. That model predicts that farmers who are better at producing value on their farm by growing crops will select into homesteading and farmers who ware better at producing value on their farm by building fences will select into purchasing. Note that a farmer's ability to grow crops or build fences may proxy for other exogenous qualities like wealth or goals, but here I will refer to this category as ability. I model how exogenous differences in ability to grow crops (create consumable goods) or build fences (create durable goods) lead to selection into different methods of land acquisition.

Table 5 reports the results from equations 22 and 26, which estimate how purchasers and homesteaders create production and value on their farms. I estimate these production functions separately for purchasers and homesteaders and test to see if the coefficients of interest are significantly different between purchasers and homesteaders. If the coefficients of interest on how individuals create value on their farm are significantly different between purchasers and homesteaders, this result indicates purchasers and homesteaders have different production functions, i.e., different methods of creating value on their farm. Note that the results presented in Table 5 include individual level controls for soil type, days since of acquisition, distance to town, population of town, gender, race, age, place of origin, the presence of water, speculation in the county, unsold land in the township-range, latitude and longitude of the farm, year of acquisition, acres of the farm, distance to the nearest land office, years since the survey, precipitation, farm value, census enumerator, and tax rate in the county by decade.

Table 5 shows that the various improvements are significant predictors of farm value and that elements used in production are significant predictors of production value, which is what we would expect. More importantly, the coefficients of interest $\pi_j$ are significantly different at the 5 percent level between purchases and homesteads. This result means that purchasers and homesteaders do in fact have different production and value functions: they use the same inputs to their farm differently in creating farm products and farm value. When given the same amounts and types of inputs into the farm (i.e., fertilizer, wages, etc.), purchasers and homesteaders use these inputs to create different amounts of outputs (i.e., crops, livestock, and assessed farm value). This result is consistent with the heterogeneous production and value functions model that I have described.

Further, I present results from the second implication of the model: The second result of the heterogeneous production and value functions model is $\frac{I_P^*}{s_P^*} > \frac{I_H^*}{s_H^*}$, meaning that the ratio of improvements to seeds is greater for purchasers than homesteaders. To test this prediction, I estimate the following ratio separately for purchasers and homesteaders:

$$\frac{fences_i}{fertilizer_i} \tag{38}$$

The model designates $I$ and $s$ each as a unit of improvements and crop-specific investments. In the Kansas Agricultural Census, $fences$ and $fertilizer$ are both already in dollars spent in the year prior to the

Table 5: Heterogeneous Production and Value Functions

| | Purchase | | Homestead | |
|---|---|---|---|---|
| | Farm Value | Production | Farm Value | Production |
| Tilled Acres / Acres =1 | 7.47** | | 10.45*** | |
| | (2.33) | | (2.61) | |
| Pasture Acres / Acres | 6.02** | | 8.71** | |
| | (2.03) | | (3.10) | |
| Fence Cost / Acres | 4.81*** | | 2.53** | |
| | (1.07) | | (1.32) | |
| Fertilizer Cost / Acres | | 2.30** | | 4.92*** |
| | | (1.14) | | (1.93) |
| Machinery Value | | 5.60*** | | 7.97*** |
| | | (1.84) | | (2.62) |
| Observations | 676 | 676 | 781 | 781 |

Standard errors in parentheses

Includes individual level demographic and land controls

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Census. This gives use the ratio of spending (investment) on fences to fertilizer separately for purchasers and homesteaders, which can be tested to see if it is significantly different between groups. The model predicts that the ratio will be higher for purchasers.

More specifically, I estimate the following, where $j$ can denote either purchasers or homesteaders:

$$E_j[\frac{fences}{fertilizer}|X, W] \tag{39}$$

To control for land and individual demographic characteristics in this ratio, I create ratios for each intersection of controls $X$ and $W$ and take the weighted average. Therefore, any difference in the ratio of fences to fertilizer between purchasers and homesteaders reflects a true difference between those two groups and is not simply driven by correlates. I estimate the ratio between fences and fertilizer separately for purchasers and homesteaders.

I use the fence cost on the farm to proxy for the durable investments $I$ and the fertilizer cost on the farm to proxy for the consumable investments $s$. Fences are resold with the farm, whereas fertilizer is an input used only in growing crops. I use the intersection of land controls $X$ and demographic controls $W$ at the individual level to insure that any difference in the ratio of fences to fertilizer reflects the true difference between purchasers and homesteaders. The results are as follows:

$$E_P[\frac{fences}{fertilizer}|X, W] = 4.09 > 2.71 = E_H[\frac{fences}{fertilizer}|X, W] \tag{40}$$

Above, there is a significant difference between the ratio of fence cost and fertilizer cost for purchasers and homesteaders, after controlling for land and farmer characteristics. This result indicates that the second prediction of the heterogeneous value and production functions model holds: purchasers invest more in durable improvements compared to consumable goods than homesteaders.

It is notable that I use variables which represent the *decisions* that the farmers make about their farm inputs, not their farm outputs such as value per acre or production per acre. Farm outputs have some stochastic element which the farmers cannot control, like weather. In section 5, I demonstrated that there exists a difference between purchasers and homesteaders in terms of farm output strategies: purchasers initially have a higher valued farm, but homesteaders initially produce more farm output like crops. However, farmers are able to completely control their *inputs*, such as fences and fertilizer. Therefore, the ratio in equation 39 describes the decisions made by purchasers and homesteaders, not the results of those decisions. This result further supports the idea that purchasers and homesteaders had different goals.

The results are consistent with the heterogeneous production and value functions model: we observe 1) different farming strategies between purchasers and homesteaders, 2) different production and value functions between purchasers and homesteaders, and 3) different ratios of investment in durables and consumables between purchasers and homesteaders.

# 5 Historical Context Informed by the Model

Although the United States owned the land in the interior of the continent, with few American citizens living on it, there was a significant fear that the land could not be controlled by the United States because of the native peoples already residing there. The Homestead Act was used to encourage (largely white, largely American citizens) to quickly settle permanently in the interior to establish meaningful property rights and American sovereignty in an area which was realistically only tenuously controlled by the United States government (Allen, 2019). Because of this goal, the Homestead Act was partially designed for colonization in the short-term, not just for farming in the long-term.

Because the Homestead Act was not actually focused on creating long-term farmers, it produced largely low-skill farmers on land that was unsuitable for farming. This paper shows that high-ability farmers selected into purchasing land and low-ability farmers selected into homesteading it, even on land that was available for both purchasing and homesteading. This selection complicates the potential long-term impacts of the Homestead Act because the selection makes it difficult to compare the outcomes of the two groups. I provide evidence of selection into purchasing versus homesteading which decreases the effect of the method of land acquisition as a causal determinant of short-term land outcomes, such as the value of the crops grown on the farm in the next decade. This result casts doubt onto the method of land acquisition being a significant causal determinant of long-term land outcomes, such as the level soil erosion today. Even prior to the method of land acquisition, purchased land and homesteaded land were placed on two different trajectories.

I estimate the relationship between the initial decision to purchase or homestead and the production per acre or value per acre in the census year. The OLS results show highly significant differences between purchasers and homesteaders: homesteaders start out by farming and producing farm output right away, while purchasers begin by making durable improvements to their land, so they are initially not as productive as homesteaders, even though their farm is more valuable. Within about eighteen months, however, purchasers become more productive than homesteaders. This difference between purchasers and homesteaders results in different farming strategies. The OLS estimates show that there is a relationship between the method of acquisition and this heterogeneity in farming strategies.

However, this same result is not visible in the IV estimation. The IV results do not show different farming strategies between purchasers and homesteaders. The IV results estimate the causal effect of the method of land acquisition - purchase or homestead - on the farm production and value. Under the identifying assumptions, the nature of the instrumental variable factors out the effect of selection into purchase or homesteading. Therefore, without the effect of selection, the difference in farming strategies between purchasers and homesteaders observed in the OLS results disappears.

Taken together, the OLS and the IV results indicate that there are different farming strategies between purchasers and homesteaders; however, the causal mechanism behind this difference is not the method of acquisition itself. This stands in contrast to the previous literature which has ascribed differences in long-term economic activity, land use, and land development to the method of acquisition - the initial decision to purchase or homestead. However, the previous literature has generally not used individual-level data, and has therefore been unable to control for individual characteristics. While this paper does not replicate results of the previous literature with individual-level data, the fact that differences in short-term farm outcomes between purchased land and homesteaded land cannot be causally ascribed to the method of acquisition makes it unlikely that differences in long-term outcomes should be causally ascribed to the initial decision to purchase or homestead.

Historians have focused on the demographic heterogeneity of farmers and ranchers in the West (Edwards, 2008; Cook, 2014). In this paper, I document another type of heterogeneity. Purchasers and homesteaders invested different inputs into their farms: purchasers invested more into durable improvements and homesteaders invested more into crops and livestock. I model this as a result of the different incentives faced by purchasers and homesteaders: purchasers could more easily resell their land because they already owned the title, whereas homesteaders had to pay additional money to acquire the title in the first five years.

Since durable improvements like fences were sold with the farm, purchasers received a higher return from investing in durable improvements than did homesteaders. This higher return itself creates an incentive for purchasers to resell their land more quickly than homesteaders. I provide empirical support for this result in Figure 1: 30% of purchasers resold their land in the first year, as opposed to less than 5% of successful homesteaders. Even after for accounting for the length of time it takes to get the title to the farm, only 10% of proved homesteads are sold directly after acquiring the title, as compared to the 30% of purchased farms. This different rate of resale further puts homesteaded and purchased land on different long-term trajectories.

Disentangling the causal mechanism behind this faster resale by purchasers remains an avenue for future research: are purchasers able to resell more quickly because they have put more valuable improvements on their land or do purchasers choose to put more valuable improvements on their land because they want to resell quickly? Either way, the above results related to to farmer heterogeneity and selection into purchasing and homesteading and their subsequent production and resale decisions can inform the future work of historians of the American West. If purchasers and homesteaders each had their own motives and goals for acquiring land in the West, had different rates of land resale, or had different production functions and abilities - i.e., they were two different "types" of farmers - then historical research should treat the two groups differently.

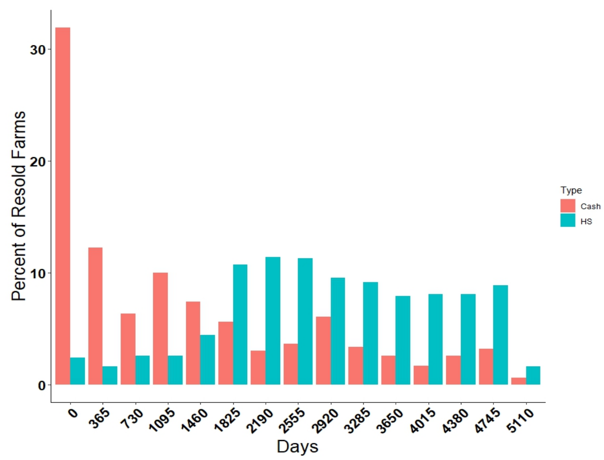


Figure 1: 30% of purchasers resold their land in the first year, as opposed to less than 5% of homesteaders

# 6 Conclusion

The effect of individual selection into either purchasing land or homesteading it has been largely overlooked by the previous literature. Instead, previous literature ascribes the causal mechanism behind differences in farming outcomes to the method of land acquisition, i.e., purchasing land causes that land to later become more economically developed than homesteading the same land (Allen, 1991; Allen and Leonard, 2019; Libecap and Hansen, 2004; Smith, 2020). However, the analysis in this literature is generally conducted at the county-level, meaning that it cannot control for individual demographic characteristics of the owners and farmers. Being unable to control for individual demographic characteristics has led most previous literature to make the strong assumption that purchasers and homesteaders were the same "type" of farmers and did not differ in any observable or unobservable ways that could have affected future land outcomes. However, the different administrative requirements unique to purchasing and homesteading make it unlikely that individuals did not systematically select into the two methods of land acquisition. Purchasing cost ten times more per acre than homesteading, making it likely that there existed selection into purchasing based on wealth. This higher wealth could also lead to better farm outcomes for purchasers.

Therefore, this paper begins by demonstrating evidence of individual selection into purchasing and homesteading based on unobservable individual-level characteristics such as initial wealth. I develop a simple theoretical model that predicts selection and use it to create an instrument, the distance to the nearest local land office, to identify the causal effect of the method of land acquisition on subsequent farm outcomes and productivity. The validity of this instrument hinges on the differing administrative requirements of purchasing and homesteading: purchasers went to the land office once, while homesteaders had to go twice. As the distance to the land office increases, the additional travel costs mean that homesteading becomes relatively more expensive to purchasing.

In order to empirically demonstrate evidence of selection, I test two predictions of the simple selection model. I show that the expected value of the farm in 1870 for individuals who always purchase is higher than the expected value of the farm in 1870 for individuals who only purchase because they are moved by the distance to the land office instrument. Likewise, the expected value of the farm in 1870 for individuals who always homestead is lower than the expected value of the farm in 1870 for individuals who only homestead because they are moved by the instrument. These two predictions of the selection model hold, indicating that there does exist individual selection into purchase and homesteading. I further test the presence and effect of selection using a linear model which estimates the relationship between the decision to purchase or homestead and the production per acre of the farm in 1870. I add individual-level demographic controls one at a time to show that the coefficient on the method of land acquisition decreases as individual characteristics are added to the regression. This indicates that estimating the causal effect of purchasing or homesteading land on subsequent land outcomes may lead to biased and overstated estimates if individual controls are not included. Together, these two results indicate that individual selection into purchasing or homesteading is a significant issue which complicates the empirical setting.

Because individual selection into purchase or homesteading is likely, I use the IV developed from the simple selection model to estimate the causal effect of the method of land acquisition on farm productivity and value. The identifying assumption is that the purchase-homestead decision occurs after the individual decides which piece of land to acquire on the causal pathway. The nature of the instrument factors out any potentially selection into purchasing and homesteading, leading to a causal estimate of the method of land acquisition - purchase or homestead - on subsequent farm outcomes.

I show that while OLS estimates indicate that different farming strategies are associated with purchasing and homesteading, IV estimates show no significant difference in farming strategies between purchasers and homesteaders. OLS estimates indicate that homesteaders immediately begin growing crops and livestock, while purchasers initially invest in durable improvements like fences and buildings. Therefore, initially, homesteaders are more productive than purchasers, but successful purchasers surpass successful homesteaders after about eighteenmonths in terms of productivity. However, IV estimates, which control for individual selection into purchase and homesteading, show no such difference in farming strategies between purchasers and homesteaders. This means that there does exist a difference in farming strategies, but it is not the decision itself to purchase or homestead which drives this difference. Therefore, there must be another mechanism besides the initial land acquisition method that causes heterogeneity in farmings strategies.

I propose a causal mechanism behind the fact that homesteaders initially invest in crops and livestock and purchasers initially invest in fences and buildings by developing a model of heterogeneous production and value functions. In this model, every farmer has exogenous, innate production functions and value functions which differ between farmers. This represents the idea of farmer ability or effectiveness. Then I show mathematically that the different administrative requirements of purchasing and homesteading cause farmers with specific production and value functions to select into purchasing and homesteading. Specifically, farmers who are better at making durable improvements to the land such as fences than they are at farming receive a higher payoff if they purchase than if they homestead. This means that farmers who can create more value by building structures than by planting crops will choose to purchase and then choose to spend more of their time and resources on durable improvements to the land than on farming because this maximizes their payoff. This model provides a plausible mechanism to explain why empirical evidence shows that purchasers initially invest more into fences and buildings and homesteaders initially invest more into crops and livestock, but that is not the method that drives this difference. Unlike previously proposed mechanisms of the method of land acquisition, this mechanism explicitly includes the effect of farmer selection and hinges on the historical requirements of purchasing and homesteading.

After developing this model of heterogeneity in production and value functions, I test it in several ways. I estimate the assessed value of the farm and the production of the farm separately for purchasers and homesteaders and find that the coefficients are statistically different between the two groups. Then I estimate the ratio $\frac{I^*}{s^*}$ separately for purchasers and homesteaders by estimating $\frac{fences_i}{fertilizer_i}$ separately for each group. Here again, I find that the ratio between durable investments and consumable investments is statistically different between purchasers and homesteaders. This provides further evidence in support of the heterogeneous production functions model because this result means that purchasers and homesteaders made different input decisions about how to create value on their farm.

This paper documents different farming strategies between purchasers and homesteaders: homesteaders start growing crops and livestock right away, while purchasers build fences. While homesteaders initially produce more farm output, after about eighteenmonths, purchasers surpass homesteaders in terms of farm production. However, IV results indicate that the causal mechanism driving this difference in farming strategies is not the initial decision to purchase or homestead land. Instead, I show evidence of selection into purchase and homesteading, and then use that selection to generate another mechanism besides the *method* of land acquisition which could explain the observed difference in farming strategies between purchasers and homesteaders.

The mechanism I propose is heterogeneous production and value functions which are innate to each farmer. Each farmer has a heterogeneous, innate ability to grow crops and build fences - to produce farm goods and to create farm value. Differences in the administrative requirements of purchasing and homesteading make homesteading more attractive to people with high abilities to grow crops, and make purchasing more attractive to people with high abilities to build fences. Finally, I test this heterogeneous production functions model mechanism empirically. Results show that there did exist selection into purchasing and homesteading *caused* by the farmers' differing innate production and value functions.

## References


[1] Allen, Douglas W. (2019). ”Establishing Economic Property Rights by Giving Away an Empire”. *The Journal of Law and Economics*, 62, no. 2 (May 2019): 251-280.

[2] Allen, Douglas W. (1991). ”Homesteading and Property Rights; Or, ”How the West Was Really Won””. *The Journal of Law and Economics*, 34, no. 1 (Apr., 1991): 1-23.

[3] ”An Act Making Further Provision for Sale of Public Land”. Act of April 21, 1820 (Land Act), Public Law Stat. 566; General Records of the United States Government; National Archives.

[4] ”An Act to secure Homesteads to actual Settlers on the Public Domain.” Act of May 20, 1862 (Homestead Act), Public Law 37-64; Record Group 11; General Records of the United States Government; National Archives.

[5] Anderson, Terry L. and Peter J. Hill (1990). ”The Race for Property Rights.” *The Journal of Law and Economics*, 33, no. 1 (Apr., 1990): 177-197.

[6] Atack, Jeremy (2013). ”On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited”. *The Journal of Economic History*, Vol. 73, No. 2 (June 2013).

[7] Atack, Jeremy, Fred Batemen, and Robert A. Margo (2010). ”Did Railroads Induce or Follow Economic Growth?” *Social Science History* 34:2 (Summer 2010).

[8] Atack, Jeremy, Fred Batemen, and Robert A. Margo (working paper). ”The Transportation Revolution Revisited: Towards a New Mapping of America’s Transportation Network in the 19th Century”.

[9] Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor (2015). “The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South.” *American Economic Review* 105(2): 477–503.

[10] Cook, Lisa D., Trevon D. Logan, and John D. Parman (2014). ”Distinctively black names in the American past”. *Explorations in Economic History*, Vol. 53, July 2014, 64-82.

[11] Edwards, Richard (2008). ”Why the homesteading data are so poor (and what can be done about it)”. *Great Plains Quarterly*, 28(3):181–190.

[12] Gates, Paul Wallace. *History of Public Land Law Development.* The Public Land Law Review Commission, Washington, D.C., 1968.

[13] Gates, Paul Wallace (1936). ”The Homestead Law in an Incongruous Land System”. *The American Historical Review*, 41(4):652-681.

[14] Gates, Paul Wallace (1977). ”Homesteading in the High Plains”. *Agricultural History*, 51(1):109-133.

[15] Hansen, Zeynep K. and Gary D. Libecap (2004). ”Small Farms, Externalities, and the Dust Bowl of the 1930s”. *Journal of Political Economy*, 112, no. 3 (June 2004): 665-694.

[16] Kohl, Edith E. *Land of The Burnt Thigh: A Lively Story of Women Homesteaders On The South Dakota Frontier*. Funk and Wagnalls, Inc., St. Paul, MN, 1938.

[17] Leonard, Bryan and Douglas W. Allen (2019). "How Many rushed during the Oklahoma land openings?" *Cliometrica*, 23 Sept 2019.

[18] Leonard, Bryan and Douglas W. Allen (working paper). "Property Rights and Path Dependence: 19th Century Land Policy and Modern Economic Outcomes".

[19] Peck, William J. *The Palimpest*. The State Historical Society of Iowa, Iowa City, IA, 1967:48(10).

[20] Smith, Cory (2020). "Land Concentration and Long-Run Development: Evidence from the Frontier United States".

[21] United States Department of Agriculture. National Agriculture Statistics Service, Subject Data: Livestock.

# Appendices

## A Data

I use three main individual-level data sources: the individual Kansas Agricultural Census, which details individual farm production, the Bureau of Land Management tract books, which record individual purchase and homestead decisions, and the United States Population Census, which records individual demographic characteristics. Using individual-level farm data as opposed to the county-level farm data generally used in the previous literature allows me to estimate the heterogeneity in production functions of farmers. Linking to the individual United States Population Census allows for many more individual demographic controls than any the previous literature in the area, and also allows me to empirically demonstrate selection into purchase and homesteading on observable characteristics.

Kansas was chosen for analysis in part because of the availability of the Kansas Agricultural Censuses. These censuses are state-level censuses, separate from the U.S. Agricultural Census. They are notable because the Kansas Agricultural Census was conducted at the individual farm level, with the name of each farm owner, whereas the federal agricultural censuses are county-level.

The Kansas Agricultural Census contains the following variables: name, township, county, the exact date that the census was taken, enumerator, acres of land on the farm (subdivided into improved and unimproved acres), farm value (including land, fences and buildings), farm machinery value, livestock value, wages for farm labor in the year, bushels of different types of crops produced, and number of different types of livestock. It does not include demographic controls, such as age, race, or wealth of the owners.

The Census took place in June through August of the census year, and the crop data reflect the previous year harvest. Notably, of farmers who acquired land prior to the census year, only successful farmers were included in the Census because the farmers who were unsuccessful would have resold or abandoned their land.

The Bureau of Land Management (BLM) keeps the original scanned tract books used at each land office in Kansas to record the acquisition of land. The tract books record all original acquisitions of land from the federal government. Like the Kansas Agricultural Census, the level of observation is the farm. The tract books record both successful and unsuccessful homestead claims and purchases, including the original date filed, the date the patent was applied for, the date the patent was acquired, the name of the owner, the size of the acreage, the price per acre, the Act it was filed under, and the state, county, county-section, township-range (PLSS), section, and quarter-section[3]. The purchases in the tract books contain only new, unbroken land sold, not land already used for farming that changed hands. Since homestead land was all new, unbroken land, it is not comparable to land that had been farmed for several years.

Finally, the above two datasets are linked together and then to the United States Population Censuses. These Censuses detail individual-level information about age, race, wealth, occupation, place of origin, family structure, and other demographic characteristics in every decade.

[3]A township-range is 36 square miles, with 36 sections in it. Each of these sections can be sub-divided, and a quarter-section is 160 acres. Sections 16 and 36 were reserved for future sales to support schools.

## A.1 Bureau of Land Management Tract Books

The Bureau of Land Management (BLM) tract books record all original acquisitions of land from the federal government, including the homesteads and purchased land used in this analysis. The level of observation is the aliquot. The tract books record both successful and unsuccessful homestead claims and purchases, including the original date filed, the date the patent was applied for, the date the patent was acquired, the name of the owner, the size of the acreage, the price per acre, the act it was filed under, and the state, county, township-range (PLSS), section, and aliquot. Aliquots may be any size, but the typical size is 160 square acres. The land in the tract books contain only new, unbroken land sold, not land already used for farming that changed hands.

The BLM already has a digitized database of about 50 percent of the homesteads from the federal government. However, the currently digitized BLM database does not include abandoned homesteads and it groups commuted homesteads with cash sales. Using these data compares successful homesteads to purchased farms which may or may not have been successful, and neglects approximately fifty percent of the homesteads altogether. Both abandoning and commuting (paying an extra 1.25 dollars per acre to acquire the title more quickly) was a common decision by homesteaders, so leaving those options out presents a skewed picture of land outcomes.

This project has digitized 7011 unique homestead and cash purchase observations from two counties in Kansas: Bourbon County and Woodson County. These counties were selected because they lie in the eastern part of Kansas, east of the hundredth meridian, which is the dividing line between land suitable for small farms and arid land suitable for ranching and large farms. Land west of the hundredth meridian is not comparable to land east of the hundredth meridian because of the difference in aridity. These two counties have comparable latitudes, but Woodson County is slightly farther west than Bourbon County. This difference means that it was acquired by individuals from the federal government later than land in Bourbon County, reducing the likelihood that my results are driven by time-varying factors such as periods of drought or business cycles. The choice of counties allows us to keep the aridity, temperature, and weather conditions similar but still retain variation. While high ability farmers may be more likely to choose high quality land, the fact that my setting is geographically small mitigates this issue. Land within a small area is of similar quality and has similar market access.

I remove individuals who purchased more land than they could have homesteaded from the sample to eliminate large-scale land speculators who could not have legally homesteaded the land: this removes very few observations, as the mode acres for purchased land was 160 acres.

Table 6: Land Acquisition Decisions

| | | |
|---|---|---|
| Total Homesteads | | 3817 |
| Commuted | 532 | |
| Abandoned | 914 | |
| Total Purchased | | 3194 |
| Total | | 7011 |

From the summary table, there are roughly equal numbers of homesteads and purchases in the sample. About 24% of the homesteads were abandoned and a further 14% were commuted prior to the five year contract. Land in eastern Kansas was among the best quality land that was homesteaded in the United States, so this figure for the percent of homesteaders who abandoned represents a lower bound in the country as a whole.

In order to calculate distances (between farms, etc.), the BLM data had to be converted to geospatial data (latitude-longitude). The smallest unit of measure for the BLM data is the aliquot in the township-range (PLSS), which gives geospacial data to a precision of 18 decimal places. This level of precision is necessary for measuring the distance between quarter-section farms, the closest of which have a midpoint of half a mile apart. I used Earth Point to convert the township-range data to geospacial data.

## A.2 Kansas Agricultural Census

To provide agricultural data, I match the tract book observations to the Kansas Agricultural Censuses of 1860, 1870, and 1880. These censuses are state-level censuses, separate from the U.S. Agricultural Census. They are notable because the Kansas Agricultural Censuses were conducted at the individual farm level, with the name of each farm owner, whereas the federal agricultural censuses are county-level.

The Kansas Agricultural Censuses contain the following variables: name, township, county, the exact date that the census was taken, enumerator, acres of land on the farm (subdivided into improved and unimproved acres), farm value (including land, fences and buildings), farm machinery value, livestock value, wages for farm labor in the year, bushels of different types of crops produced, and number of different types of livestock. They do not include demographic controls, such as age, race, or wealth of the owners. The information contained in the Censuses can be grouped into two categories: 1) investments in durable goods (e.g.: fences) and 2) production or investments in consumables (e.g.: fertilizer or corn). These two categories represent different strategies of investment and indicated farmer's motivations and preferences for different methods of farming.

I have digitized the Kansas Agriculture Censuses for Bourbon and Woodson counties. I hand-match this dataset to the tract book dataset to provide information on the production and farm input decisions of farmers at the individual level. This matching allows me to control for the number of days the farmer has been working the farm before the date of the Census because both datasets include the day/month/year. I am then able to obtain the value of the crops produced on each farm by multiplying the number of bushels of each crop by the currency price per bushel in 1860, 1870, and 1880 in the United States (United States Department of Agriculture, 2013).

The Census took place in June through August of the census year, and the crop data reflect the previous year's harvest. Notably, of farmers who acquired land prior to the census year, only successful farmers were included in the Census because the farmers who were unsuccessful would have resold or abandoned their land. Therefore, I examine farms acquired in 1869 and 1870 and 1879 and 1880 and farms acquired outside of those periods separately.

## A.3 Land Controls

### A.3.1 Water

In order to determine whether each quarter-section farm in Kansas contained a water source, I define water as streams, rivers, ponds, lakes, swamps, and marshes, and waterfalls. I use historical county plat maps obtained through the Kansas Historical Society. It is necessary to use historical water maps, as the location of water may change over time. These maps, which date from between 1900 to 1906, were drawn to the scale of half-an-inch to the mile, which allows for detailed depictions of small streams, ponds, rivers, and swamps. The water maps are divided by township-range and section, which allows determination of each quarter-section. Therefore, using the township range, section, and quarter-section information from all of the Kansas farms in the dataset, I manually match each farm to the corresponding quarter-section on the water maps. This matching is direct because both datasets use PLSS and does not require geospacial conversion. Each individual farm in the dataset is thereby determined to either have a water source ($Water_i = 1$) or not have a water source ($Water_i = 0$).

### A.3.2 Soil Quality

A clear threat to identification of this model is that homesteaded and purchased land may have been of different quality, leading to different levels of productivity on the farms. Therefore, I control for soil quality at the quarter-section level using the National Commodity Crop Productivity Index from the Soil Survey Geographic (SSURGO) database. This dataset classifies 8 soil categories based on their effectiveness for farming. This measure is called the non-irrigated land capability class, where higher numbers indicate "greater limitations and narrower choices for practical use" (United States Department of Agriculture, 2013). The USDA classifies soil types 1 through 7 as suitable for grazing and ranching and soil types 1 through 4 as suitable for growing crops. Type 8 is suitable for neither.

Table 7: Soil Classification by Farm Type

| | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Homesteads | 413 | 3.16 | 0.61 | 1.50 | 4.74 |
| Pre-1863 Purchases | 297 | 3.20 | 0.57 | 1.75 | 5.25 |
| Post-1863 Purchases | 246 | 3.36 | 0.52 | 2.14 | 4.83 |

The summary table shows that homesteads are in general situated on slightly better land than purchases. This difference is statistically significant for all pairs except homesteads and purchases before 1863, though it may not represent economic significance because in Kansas nearly all soil is classified as suitable for some type of farming. The fact that homesteaders acquired better land than post-1863 purchasers might initially contradict intuition, but in fact, it is consistent with the result that homesteads generally preceded purchases in an area due to their lower risk from their lower cost. Because homesteads were less risky than purchases, homesteaders were willing to go farther into the frontier at any point, allowing them to select land first, leading to homesteaders having overall better quality land.

### A.3.3 Precipitation

In order to control for differing levels of precipitation between homesteaded and purchased farms that could lead to different productivity, I use the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information. Because precipitation data is only available for one weather station in Kansas before 1872 and because I need to exploit spacial variation in precipitation data, I use the

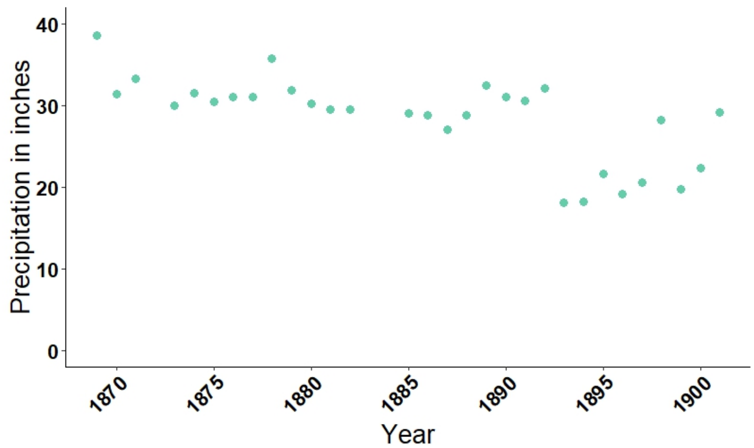


Figure 2: Average Rainfall in Kansas by Year

average precipitation level for each township-range from 1880 to 1900. After 1880, precipitation data are much more available and cover a much larger percentage of Kansas.

I assume that the precipitation in each township-range does not vary significantly by year and that therefore the average level of precipitation in an area from 1880 to 1900 is a good approximation of the level from 1856 to 1880. This assumption is supported by the available historical data: the level of precipitation remains quite constant over the available period, as shown by the yearly average in Figure 1.

### A.3.4 Towns and Land Offices

Towns represent differing access to markets which could affect farm productivity. I obtain historical population data for Kansas towns in every decade from the U.S. Population Censuses of 1850 to 1880. The Kansas Historical Society has the date of incorporation of 624 towns. Over half of the towns were under 250 people. I obtained the latitude and longitude of these towns from Open Data Soft's database of U.S. Zip Codes Latitude and Longitude. This database contains 43,191 cities in the United States, including the city name, state, zip code, and the latitude and longitude of each city to 6 decimal places.

Because I am interested in how individuals acquired land, I am only concerned with towns that were incorporated when the farmers bought or homesteaded land nearby. I use the date of incorporation for this. Then, for each decennial Population Census, I extract the population of the town. For each farmstead, I calculate the distance between the centroid of the farm and the nearest town at the time of acquisition. The number of towns in Kansas increased dramatically from 1860 to 1870.

The BLM provided the source for the local land office data. The land office was where the farmers acquired the land. Kansas contained between 12 and 16 local land offices in this period of time. I determine the date each land office was opened by the date of the first entry recorded in that office and the date each was closed by the date of the last entry recorded in that office. Each of the local land offices was located in a town, and I obtained the latitude and longitude of these towns to determine the distance of each farm to the nearest land office in operation at the time of acquisition.

### A.3.5 Other

I additionally control for the percentage of the township-range that had already passed into private ownership at the time of each acquisition, the amount of land in the county already owned by large speculators[4], whether the owner of the farm had previously acquired farmland from the federal government[5], whether the acquisition was part of a larger farm, the latitude and longitude of the centroid of the farm to flexibly control for potential variation across space within counties, the number of acres of each farm, and years between the original survey of that township-range and the date of acquisition of the farm. I also control for the enumerator (individual census-taker) from the Kansas Agricultural Census to allow for biases in their valuations. Additionally, I use the United States Population Censuses to control for the average property tax rate in the county by decade because owners of purchased farms had to pay taxes immediately after acquisition (because they acquired the deed to the land immediately), while homesteaders did not have to pay taxes for 5 to 7 years after acquisition (because that was when they acquired the title).

## A.4 Individual Demographic Controls

I match to the 1860, 1870, and 1880 United States Federal Population Census to control for time-invariant demographic characteristics at the individual level, such as gender, race, place of origin, and age. This matching allows for the analysis of individual-level demographic variables which may influence selection into purchasing or homesteading.

The U.S. Population Cenuses include multiple measures of wealth; however, such measure are not useable for this analysis because wealth in 1870 occurs after the decision to purchase or homestead in 1868 on the causal pathway. Therefore, wealth in 1870 is likely to be influenced by the individual's previous decision to purchase or homestead. It is not possible to match individuals to the U.S. Population Census before they acquire land because it is not possible to locate them geographically. Therefore, only time-invariant variables from the U.S. Population Census are suitable for use as controls.

# B Robustness Check: Neighbor IV

As a robustness check, I use a separate instrumental variable: the acquisition method of the previously acquired neighboring farms. I use a leave-one-out method: the PLSS structure sets up farms in essentially a grid, and the mode method of acquisition (purchase or homestead) of the 8 surrounding farms predicts the method of acquisition of the 1 farm in the middle. I create the variable $Neighbor_{it}$ by identifying all farms within half a mile of farm $i$ which had been acquired at any previous time to the acquisition date

[4] $PercentLargeLandOwners_{ijt-1}$ is a continuous variable which measures the percent of county $j$ that was purchased by individuals buying at least 5000 acres in year $t-1$. The 5000 acres do not have to be purchased at once, as long as they are purchased within 1 year. If multiple people purchased 1 piece of land or if the land was purchased in multiple counties, I average over the number of people or counties.

[5] $PreviousAcres_{it}$ is a continuous variable indicating the number of acres of the setter has previously acquired within a 2-mile radius of the centroid of farm $i$ in the 10 years prior to the acquisition of farm $i$. This measure can theoretically range from 0 to 12,800 prior acres. I use this measure to control for farmers with previous experience farming the area being more productive farmers. This control is important because farmers who purchased were more likely than homesteaders to have acquired other land in the area and therefore purchasers may have had more farming experience. Controlling for their previous acres of farmland eliminates this bias.

of farm $i$. Purchased farms (purchased before or after 1863) are classified as 1 and homesteaded farms are classified as 0. I average the acquisition method of the previously acquired farms within half a mile of farm $i$ to create $Neighbor_{it}$. This method creates a continuous variable which ranges between 0 and 1 for each farm $i$ in my sample. The unconditional correlation coefficient between the method of acquisition of farm $i$ and $Neighbor_{it}$ is 0.81.

The identifying assumption of this IV model is that the total value of production per acre of farm $i$ in 1870 or 1880 is only related to the method of acquisition of the *previously acquired* surrounding farms through the method of acquisition of farm $i$, after controlling for land and individual characteristics as above. As most farms in a township-range were acquired in a very narrow time frame, this assumption is reasonable: individuals in an area were likely to chose the same method of acquisition as their neighbors and predecessors due to shared information.

Moreover, individuals within one township-range are not observably similar in terms of age, place of origin, or non-farming occupation, based on data from the U.S. Population Censuses. This heterogeneity makes it unlikely that there were wealthy areas where farmers who purchased were more likely to go. Farmers did not have unlimited choices of where to acquire land because of their method of arriving to Kansas: while all the land in Kansas was legally available for acquisition, it was not necessarily feasibly available. Individuals to Kansas arrived by railroad. Until the railroad was built to a certain location, that location was unavailable. Therefore, individuals were constrained in their ability to sort into wealthy locations, which is reflected by the fact the individuals in a given township-range were not observably similar. Finally, it is reasonable that the instrument affects all the individuals in the same direction: if most of their neighbors choose one method of acquisition, the owner of farm $i$ may change their method to conform to their neighbors' method, but it is unlikely that most of their neighbors choosing to purchase would make farmer $i$ choose to homestead when they were originally planning to purchase.

It is notable that I use only the surrounding farms that had been acquired *prior* to farm $i$ to predict the method of acquisition of farm $i$ because the productivity of farm $i$ could have influenced whether subsequent farms were purchased or homesteaded. However, it is not reasonable to assume that the productivity of farm $i$ acquired in year $t$ influenced the method of acquisition of surrounding farms in year $t-1$ because the method of acquisition of these farms would have already been chosen by the time farm $i$ was either homesteaded or purchased. For example, the productivity in 1870 of a farm acquired in 1868 would not have influenced the method of acquisition for a neighboring farm in acquired 1865. The identifying assumption implies that, after controlling for individual land and demographic characteristics, the production per acre in 1870 of farm $i$ acquired in 1868 is only related to whether the surrounding farms acquired before 1868 were purchased or homesteaded via whether farm $i$ was purchased or homesteaded. Contradicting this identifying assumption would mean that the decision of other nearby farmers to purchase or homestead multiple years ago had a direct impact on the production per acre of farm $i$ in 1870, even after controlling for shared land or demographic characteristics which may make a method of acquisition more likely.

Because I use the previously acquired surrounding farms to create $Neighbor_{it}$, there are a few instances in which farm $i$ was the first farm to have been acquired in the area. These farms are not included in the estimation of the IV model. However, this issue is uncommon because land had been being acquired in Kansas for many decades before the first farms in my sample. Therefore, there is little reason to believe it poses a significant selection issue.

The level of observation is the farm. I estimate the first stage of my IV model as follows:

$$Purchase_{ijkt} = \alpha_{jt} + \gamma_{kt} + \lambda_i + \beta_1(Neighbor_{it}) + X_i\delta_1 + \eta_{ijkt} \tag{41}$$

In order to control for the soil quality on the nearby farms, $\lambda_i$ represents the average soil type in the area encapsulated by $Neighbor_{it}$. $\beta_1$ represents the relationship between farm $i$'s acquisition method (purchase or homestead) and the method of acquisition of the surrounding farms that had been obtained before farm $i$. $Neighbor_{it}$ predicts the method of acquisition of farm $i$. $Purchase_{ijkt} = 1$ when the land is purchased and $= 0$ when it is homesteaded. $X_i$ is a vector of controls including soil type, days since of acquisition, distance to town, population of town, gender, race, age, place of origin, the presence of water, speculation in the county, unsold land in the township-range, latitude and longitude of the farm, year of acquisition, acres of the farm, distance to the nearest land office, years since the survey, precipitation, farm value, census enumerator, and tax rate in the county by decade. County-year and township-range-year fixed effects are represented by $\alpha_{jt}$ and $\gamma_{kt}$, respectively. Standard errors are clustered by township-range. Results from the first stage are presented in Table 3.

To estimate the second stage, I use the first-stage model to derive the predicted method of acquisition of each farm $i$ from its surrounding farms and use it to create the variable $\hat{Purchase}_{it}$. This variable represents the estimated method of acquisition of farm $i$ from the surrounding previously acquired farms. Then I estimate the following second stage model using the predicted method of acquisition of farm $i$:

$$Y_{ijkt} = \alpha_{jt} + \gamma_{kt} + \lambda_i + \beta_2(\hat{Purchase}_{it}) + X_i\delta_2 + \epsilon_{ijkt} \tag{42}$$

The primary effect I am estimating is the effect of the initial decision to purchase or homestead on the production per acre or value per acre in 1870 or 1880 at the individual farm level. This effect is captured by $\beta_2$. $Y_{ijkt}$ is the production per acre or value per acre in 1870 or 1880 of farm $i$.

Table 8: Neighbor IV Estimation of Acquisition Method on Production and Value Per Acre

| | Acquired 1863 to 1868 | | Acquired 1869 and 1870 | | Acquired 1871 to 1878 | | Acquired 1879 and 1880 | |
|---|---|---|---|---|---|---|---|---|
| | Production | Value | Production | Value | Production | Value | Production | Value |
| Purchase=1 | -3.41 | 4.98* | 2.92 | 1.28 | -2.74 | 8.70 | -4.33 | 2.58 |
| | 5.92 | (2.41) | (5.03) | (1.17) | (4.15) | (6.59) | (4.41) | (1.78) |
| Observations | 476 | 476 | 382 | 382 | 398 | 398 | 256 | 256 |

Standard errors in parentheses

Includes individual level demographic and land controls

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results from the neighbor IV are similar to the results from the distance IV: the IV estimates do not show different farming strategies between purchasers and homesteaders. Again, the IV estimates are smaller than the OLS estimates, indicating selection. While the OLS estimates indicate that purchasers initially invest in durable investments such as buildings and fences, while homesteaders initially invest in consumable products like crops and livestock, neither the distance IV estimates nor the neighbor IV estimates support this difference in farming strategies between the two groups. While there exists differences in farming strategies and investment decisions for purchasers and homesteaders, this difference disappears after factoring out the effect of selection into either group.

# C Proofs

Let $\upsilon_i(\ell, D') > \rho_i(\ell, D)$ for all $D' \geq D$ and let $r_1^P > r_1^H$. Let $\upsilon$ and $\rho$ both be strictly increasing in $D$ and $\ell$ and strictly concave. Let $r_1^P \upsilon_i(\ell, D) > D + \kappa$, $r_1^H \upsilon_i(\ell, D) > D$, and $r_2 \rho_i(\ell, D) > D$ for all $D$. Then:

1. Individual $i$ will purchase, meaning $U_i^P(\ell, D_P^*) > U_i^H(\ell, D_H^*)$
2. $\frac{I_P^*}{s_P^*} > \frac{I_H^*}{s_H^*}$, meaning that the ratio of improvements to seeds is greater for purchasers than homesteaders

We begin by proving part 1 of the proposition: $U_i^P(\ell, D_P^*) > U_i^H(\ell, D_H^*)$. The individual above will choose purchasing or homesteading based on the maximum of $U_i^P(\ell, D_P^*)$ and $U_i^H(\ell, D_H^*)$, where investments are chosen optimally in each case. Therefore, the individual will purchase if $U_i^{P*} > U_i^{H*}$.

$$U_i^{P*} = r_1^P \upsilon_i(\ell, D_P^*) - D_P^* + r_2 \rho_i(\ell, D_P^*) - D_P^* - \kappa \tag{43}$$

$$U_i^{H*} = r_1^H \upsilon_i(\ell, D_H^*) - D_H^* + r_2 \rho_i(\ell, D_H^*) - D_H^* \tag{44}$$

First, we want to show $D_P^* \geq D_H^*$.

$$\frac{\partial U_i^{*P}}{\partial D_P^*} = r_1^P \upsilon_i'(\ell, D_P^*) - 1 + r_2 \rho_i'(\ell, D_P^*) - 1 = 0 \tag{45}$$

$$\frac{\partial U_i^{*H}}{\partial D_H^*} = r_1^H \upsilon_i'(\ell, D_H^*) - 1 + r_2 \rho_i'(\ell, D_H^*) - 1 = 0 \tag{46}$$

$$r_1^P \upsilon_i'(\ell, D_H^*) + r_2 \rho_i'(\ell, D_P^*) = r_1^H \upsilon_i'(\ell, D_H^*) + r_2 \rho_i'(\ell, D_H^*) \tag{47}$$

$$\frac{r_1^P}{r_1^H} = \frac{r_1^H \upsilon_i'(\ell, D_H^*) + r_2 \rho_i'(\ell, D_H^*) - r_2 \rho_i'(\ell, D_P^*)}{r_1^H \upsilon_i'(\ell, D_P^*)} > 1 \tag{48}$$

$$r_1^H \upsilon_i'(\ell, D_H^*) + r_2 \rho_i'(\ell, D_H^*) > r_1^H \upsilon_i'(\ell, D_P^*) + r_2 \rho_i'(\ell, D_P^*) \tag{49}$$

Therefore, $D_P^* > D_H^*$ because $\upsilon$ and $\rho$ are both strictly concave. Because $\upsilon$ is strictly increasing in $D$ and $D_P^* > D_H^*$, we know:

$$\upsilon_i(\ell, D_P^*) > \upsilon_i(\ell, D_H^*) \tag{50}$$

We also know $r_1^P > r_1^H$, so:

$$r_1^P \upsilon_i(\ell, D_P^*) > r_1^H \upsilon_i(\ell, D_H^*) \tag{51}$$

$$r_1^P \upsilon_i(\ell, D_P^*) + r_2 \rho_i(\ell, D_P^*) > r_1^H \upsilon_i(\ell, D_H^*) + r_2 \rho_i(\ell, D_P^*) \tag{52}$$

Since $\rho$ is strictly increasing in $D$ and $D_P^* > D_H^*$, we have:

$$r_1^P \upsilon_i(\ell, D_P^*) + r_2\rho_i(\ell, D_P^*) > r_1^H \upsilon_i(\ell, D_H^*) + r_2\rho_i(\ell, D_H^*) \tag{53}$$

Since $r_1^P \upsilon_i(\ell, D) > D + \kappa$, $r_1^H \upsilon_i(\ell, D) > D$, and $r_2\rho_i(\ell, D) > D$ we know:

$$r_1^P \upsilon_i(\ell, D_P^*) + r_2\rho_i(\ell, D_P^*) - 2D_P^* - \kappa > r_1^H \upsilon_i(\ell, D_H^*) + r_2\rho_i(\ell, D_H^*) - 2D_H^* \tag{54}$$

Therefore, part 1 is proved: individual $i$ will purchase when $\upsilon_i(\ell, D') > \rho_i(\ell, D)$ for all $D' \geq D$ and $r_1^P > r_1^H$.

Next, we prove part 2 of the proposition: $\frac{I_P^*}{s_P^*} > \frac{I_H^*}{s_H^*}$. We will show that the ratio of durable investments to consumable investments is higher for purchasers than for homesteaders. We can write $U_i^{P*}$ and $U_i^{H*}$ as functions of $I, \ell.$ and $s$ as follows:

$$U_i^{P*}(I_P^*, \ell, s_P^*) = r_1^P \upsilon_i(I_P^*, \ell) - c_1 I_P^* + r_2^P \rho_i(\ell, s_P^*) - c_2 s_P^* \tag{55}$$

$$U_i^{H*}(I_H^*, \ell, s_H^*) = r_1^H \upsilon_i(I_H^*, \ell) - c_1 I_H^* + r_2^H \rho_i(\ell, s_H^*) - c_2 s_H^* \tag{56}$$

These are the same equations as above since $D = c_2 s = c_1 I$.

$$\frac{\partial U_i^{*P}}{\partial I_P^*} = r_1^P \upsilon_i'(\ell, I_P^*) - c_1 = 0 \tag{57}$$

$$\frac{\partial U_i^{*H}}{\partial I_H^*} = r_1^H \upsilon_i'(\ell, I_H^*) - c_1 = 0 \tag{58}$$

$$r_1^P \upsilon_i'(\ell, I_P^*) = r_1^H \upsilon_i'(\ell, I_H^*) \tag{59}$$

$$\frac{r_1^P}{r_1^H} = \frac{\upsilon_i'(\ell, I_H^*}{\upsilon_i'(\ell, I^*P} > 1 \tag{60}$$

$$\upsilon_i'(\ell, I_H^*) > \upsilon_i'(\ell, I_P^*) \tag{61}$$

Therefore, $I_P^* > I_H^*$ because $\upsilon_i$ is strictly concave.

$$\frac{\partial U_i^{*P}}{\partial s_P^*} = r_2\rho_i'(\ell, s_H^*) - c_2 = 0 \tag{62}$$

$$\frac{\partial U_i^{*H}}{\partial s_H^*} = r_2\rho_i'(\ell, s_H^*) - c_2 = 0 \tag{63}$$

$$r_2\rho_i'(\ell, s_H^*) = r_2\rho_i'(\ell, s_P^*) \tag{64}$$

$$\rho_i'(\ell, s_H^*) = \rho_i'(\ell, s_P^*) \tag{65}$$

Therefore, $s_H^* = s_P^*$.

Since $s_H^* = s_P^*$ and $I_P^* > I_H^*$, we have $\frac{I_P^*}{s_P^*} > \frac{I_H^*}{s_H^*}$. This proves part 2 of the proposition, meaning that the ratio of improvements to seeds is greater for purchasers than homesteaders. This is what we observe in the data in the above section: purchasers invest in more in fences and buildings and homesteaders invest more in crops and livestock.