Two Different Types of Settlers: Effects of Selection into Purchase and Homesteading
Katherine Hauck

Under the Homestead Acts, beginning on January 1st, 1863, about 3 million 160-acre plots of land in the western part of the United States were given out by the federal government to individual settlers to be farmed. Settlers 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 settler to improve the land before getting the title. In general, the same land was available at any time for both methods of acquisition. 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 settler types? This paper tackles this issue, focusing on selection into homesteading, and how that selection affected subsequent productivity in Kansas.

An important challenge for identifying the causal effect of purchasing versus homesteading on farm outcomes is that there may be selection into each group. Nearly all previous literature has assumed purchasers and homesteaders were of the same "type" – type here defined by having the same characteristics, production functions, and goals. Because of this assumption, previous literature has ascribed the casual effect of long-term differences in economics outcomes on land originally homesteaded and originally purchased to the *method* of acquisition. However, this literature has been largely unable to control for individual initial settler demographics, and the implicit assumption made is that, after some controls, settlers homestead and purchase exogenously and there is no selection into either group based on unobservables.

However, the fact that purchasers paid more than ten times as much per acre as homesteaders indicates that that the two groups of settlers may in fact 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. This selection 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. However, this conclusion may be too hasty without a thorough examination of the selection issue. My approach to identifying the effect of the Homestead Act on agricultural productivity explicitly considers the idea that different types of settlers may have selected into purchasing and homesteading.

This analysis in general is made possible by a combination of rich, individual-level data sources which previous literature has not utilized because it has not been digitized. Much previous work on homesteading has been conducted at the county level, leading to a lack of 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 records exact the location, owner, type, and date of each land acquisition from the federal government. The created dataset is then matched to soil quality measures, slope, gradient, historical precipitation, historical water access, and historical town and railway access, all at the individual aliquot level. Unlike county-level data, using individual-level data allows the econometrician to estimate the selection into homesteading based on individual settler attributes.

Because the effect of settler selection into either purchasing land or homesteading it has been largely overlooked by the previous literature, this paper begins by demonstrating evidence of settler selection into purchasing and homesteading based on settler characteristics. 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 only 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 settlers who always purchase is higher than the expected value of the farm in 1870 for settlers 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 settlers who always homestead is lower than the expected value of the farm in 1870 for settlers who only homestead because they are moved by the instrument. These two predictions of the selection model hold, indicating that there does exist settler selection into purchase and homesteading.

I further test the presence and effect of selection using an OLS 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 settler-level controls one at a time to show that the coefficient on the method of land acquisition decreases as settler 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 settler controls are not included. Previous literature has been unable to use such settler-level controls because the analysis has been largely conducted at the county level. Together, these results indicate that settler selection into purchasing or homesteading is a significant issue which complicates the empirical setting.

Because settler selection into purchase or homesteading is likely, I use the distance to the land office instrument 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 settler 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 initial method of land acquisition – purchase or homestead – on subsequent farm outcomes. Then I use those estimates to analyze differences in farming strategies. Investment into either farm production (crops and livestock) or farm value (fences and buildings) represents different farming strategies, particularly because durable investments like fences and buildings were sold with the farm, while production was not. I show that farming strategies differed between purchasers and homesteaders.

I estimate the relationship between the initial decision to purchase or homestead and the production per acre or value per acre of the farm in the census year. The OLS results show highly significant differences between purchasers and homesteaders: homesteaders start out by farming and producing crops and livestock right away, while purchasers begin by making durable improvements to their land like fences. Therefore, purchasers are initially not as productive as homesteaders, even though their farms are more valuable. Within about 18 months, however, purchasers become more productive than homesteaders. This difference between purchasers and homesteaders in terms of a trade-off between production and value is what I refer to as different farming strategies, based on the different investment decisions of purchasers and homesteads.

However, the IV results, which estimate the causal effect of the *method* of land acquisition – purchase or homestead – on the farm production and value, do not show different farming strategies between purchasers and homesteaders. The nature of the instrumental variable factors out the effect of selection into purchase or homesteading, and 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 IV results indicate that there are different farming strategies; however, the causal mechanism behind this difference is not the method of acquisition itself. This stands in contrast to previous literature which has ascribed differences in long-term economic activity, land use, and land development to the initial decision to purchase or homestead. However, previous literature has not used individual-level data, and as therefore been unable to control for settler characteristics. The OLS estimates show that farming strategies are heterogeneous: 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 settler selection, do not reflect this. Therefore, the causal mechanism behind the

differences in farming strategies between purchasers and homesteaders was not the method of acquisition itself.

In order to determine the causal mechanism behind the fact that homesteaders initially invest in crops and livestock and purchasers initially invest in fences and buildings, I develop a model of heterogeneous production and value functions. This model shows why some settlers initially invested in improvements and why others initially invested in crops and livestock, and how that exogenous difference in production functions led to selection into purchase or homesteading.

I first build a model of the two ways settlers can accrue returns to their farm: 1) physical improvements such as fences and buildings, and 2) production such as crops and livestock. Settlers each have a value function υ and a production function ρ. The value of the farm is a function υ of the improvements to the farm such as fences and buildings (I) and the land characteristics (ℓ). The production of the farm is a function ρ, which depends on ℓ, and crop and livestock-specific investments such as seeds s. Therefore, *value of the farm* $= \upsilon(I, \ell)$ and *production of the farm* $= \rho(\ell, s)$. Both υ and ρ are strictly increasing and strictly concave. Let r_1 be the market returns to the value of the farm and c_1 be the cost per unit of farm improvement. Let r_2 be the market returns to production and let c_2 be the cost per unit of crop and livestock-specific investments. This yields:

$$Utility of the farm = U(I, \ell) = r_1 \upsilon(I, \ell) - c_1 I$$
$$Utility of production = U(\ell, s) = r_2 \rho(\ell, s) - c_2 s$$

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 settler 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. The settler solves the standard constrained maximization problem to choose I^* and s^* in order to maximize their total utility function subject to their budget constraint.

Now consider the initial decision to purchase or homestead. Assume the c_1 and c_2 are the same for purchasers and homesteaders, meaning that they face the same market costs. Likewise assume that r_2 (the market returns to selling crops and livestock) is the same for purchasers and homesteaders. Recall the legal, administrative requirement that purchasers gain the title more quickly than homesteaders. This means that $r_1^P > r_1^H$: purchasers gain the title more quickly than homesteaders, enabling them to sell or mortgage their farm more quickly, so purchasers' returns to durable farm improvements are higher than homesteaders'. Essentially, I am implicitly modeling time discounting in their returns to improvements. Finally, let κ be the cost of purchasing minus the cost of homesteading.

Both I and s, which are units of investment, are functions of wealth, ω: $I(\omega)$ 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 $D_I = D_s$.

Not all farmers have the same production function or value function. This idea reflects individual's differing abilities, goals, and resources. Production and value functions are exogenous to the model and cannot be chosen by the settler. Consider a farmer i with $\upsilon_i(I, \ell) > \rho_i(\ell, s)$: at any given level of monetary investment, this farmer is better at building fences than planting crops. This farmer i has the following utility function:

$$U_i(\ell, D) = r_1 \upsilon_i(\ell, D) - D + r_2 \rho_i(\ell, D) - D$$

This model shows that the exogenous, innate difference in production and value functions among settlers caused 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.

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* υ *and* ρ *both be strictly increasing in D and* ℓ *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. *Settler i will purchase, meaning* $U_i^P(\ell, D_P^*) > U_i^H(\ell, D_H^*)$
2. $I_P^*/s_P^* > I_H^*/s_H^*$, *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 homestead and choose to invest more money in fences than seeds. Having different exogenous production and value functions *causes* settlers 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 from previous literature that differences between purchased and homesteaded land is caused by the initial method of acquisition. This model includes and specifically highlights the effect of selection and how it can partition settlers based on truly exogenous factors: innate production and value functions.

After developing this model of heterogeneity in production and value functions, I test it in several ways. I use the unique data provided by the 1880 Kansas Agricultural Census to calculate each settler's production and value functions using a standard Cobb-Douglas production function. The 1880 Census provides information on the inputs to the production function (acres of each crop planted, fertilizer, machine value, and labor) and outputs to the production function (bushels of each crop produced). Taking the natural log of the Cobb-Douglas production function gives:

$$ln(Y_{it}) = \beta_0 + \beta_1 I_{it} + \beta_2 s_{it} + \omega_{it} + \varepsilon_{it}$$

where Y_{it} is farm production per acre from the 1880 Census, I_{it} are durable improvements such as fences, s_{it} are crop and livestock-specific investments like seeds, and ω_{it} represents farm-specific productivity factors that are observed by the farmer but not by the econometrician. Because ω_{it} may be correlated with the inputs I_{it} and s_{it}, I use instruments P_I and P_s, which are the prices of I_{it} and s_{it} respectively, to identify the coefficients of interest, β_1 and β_2. These coefficients represent the causal effect of improvements and crop and livestock specific investments on productivity. Under the assumption of no monopoly power for each small farm and that ω_{it} is taken as given by the farmer, the input prices P_I and P_s are correlated with the inputs I_{it} and s_{it} but orthogonal to ω_{it} and ε_{it}.

I estimate the production function above separately for purchasers and homesteaders and show that the coefficients of interest β_1 and β_2 are significantly different between the two groups. This result means that purchasers and homesteaders have different ways of using the same inputs to create production on their farms. This provides evidence in support of the heterogeneous production functions model. I use a similar method to estimate heterogenous value functions.

Finally, to empirically test the effect of exogenously determined production and value functions that are innate and unique to each settler on subsequent farm outcomes, I estimate the production and value functions on a subset of the sample which includes only settlers who acquired more than 160 acres. Ideally, we could perform a randomized controlled trial by exogenously assigning production and value functions to settlers at birth and observe which acquisition decision – purchase or homestead – they subsequently chose. Instead, we can use settlers who acquire multiple pieces of land.

Legally, after the first 160 acres, all pieces of land acquired from the federal government were purchased, not homesteaded. I determine a settler's production and value function from their second (purchased) piece of land, and I assume that is the production and value function that is innate to that settler. Since all pieces of land after the first 160 acres were purchased, there is no variation in investments in durable or consumable goods that comes from the method of acquisition. There is no variation in the method of acquisition – purchase or homestead – in second or subsequent land because all that land is purchased. Therefore, all the variation in what is invested in the farm comes from the production and value functions innate to the settler. I use the second and subsequent pieces of land to estimate the production and value functions of the settler, and then use those estimated exogenous value and production functions to identify their causal effect on selecting into purchasing or homesteading for their first piece of land.

In order to use the settlers' second piece of land to estimate the causal effect of their production and value functions on their initial decision to purchase or homestead, I use the following fixed-effects

model on settlers' second and subsequent pieces of land to estimate each settlers' ratio between production and value:

$$Z_{ijkt} = \alpha_{jt} + \gamma_{kt} + \delta X_{ijkt} + \theta W_i + \eta_{ijkt}$$

Z_{ijkt} represents how good settler i is at creating durable improvements on their land to creating farm output by using the ratio $ValuePerAcre_{it}/ProductionPerAcre_{it}$. $ValuePerAcre_{it}$ and $ProductionPerAcre_{it}$ are both obtained from the Kansas Agricultural Censuses, so they represent the value of the farm and the production on the farm in the Census years. Therefore, the ratio gives the amount of value in fences and building to the amount of production in crops and livestock created per acre for each farm. X_{ijkt} is a vector of land characteristics including a measure of how long the settler has been on the land prior to the Census, and α_{jt} and γ_{kt} are county-year and township-range-year fixed effects. W_i is a vector of settler characteristics such as birthplace. The fixed effects regression above estimates the settlers' abilities to create value versus farm products in the amount of time they have owned their second piece of land.

Then I use the following logit model:

$$Y_{ijkt} = \alpha_{jt} + \gamma_{kt} + \beta Z_{ijkt} + \delta X_{ijkt} + \theta W_i + \varepsilon_{ijkt}$$

Y_{ijkt} is a binary variable where $Y_{ijkt}=1$ represents a settler's the decision to purchase their first piece of land and $Y_{ijkt}=0$ represents the decision to homestead their first piece of land. β is the coefficient of interest, which estimates the causal effect of the settler's innate production and value functions (estimated from their second piece of (purchased) land) on their original decision to purchase or homestead. We expect β to be positive: under the identifying assumptions, a positive coefficient on Z_{ijkt} means that being better at creating value via fences and buildings than producing crops and livestock caused settlers to select into purchasing as opposed to homesteading. Under the assumption that production and value functions are exogenous after controlling for land characteristics, this estimates the causal effect of a settler's production and value functions on their decision to purchase or homestead. I find that having a higher value function caused settlers to select into purchasing and having a higher production function caused settlers to select into homesteading.

The Homestead Acts are regarded by historians to be one of the most far-reaching land laws in U.S. History; however, they were superimposed on top of a pre-existing land system. In considering their impact on farm outcomes, we need to be aware of potential selection into homesteading as opposed to other methods of land acquisition available at the time and how it may bias our estimates of the effect of the Homestead Acts. For that reason, I use individual-level data as oppose to county-level data to control for settler-level characteristics. I create a simple model of selection into purchasing and homesteading land and show that adding individual settler controls decreases the effect of the initial decision to purchase or homestead on subsequent farm outcomes.

This paper documents the effects of different farming strategies between purchasers and homesteaders: homesteaders start growing crops and livestock right away, will purchasers build fences. While homesteaders initially produce more farm output, after about 18 months, 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 settler. Each settler has a heterogeneous, innate ability to grow crops and build fences – to produce farm goods and to create farm value. Differences in the legal and 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 settlers' differing innate production and value functions.