CHAPTER V

LAND ACQUISITION MODEL PRESENTATION AND ESTIMATION

The model presented here includes three basic equations for (1) land acquisition, (2) population change, and (3) tobacco prices. Land acquisition can serve as a dependent variable affected by such factors as population changes and tobacco prices, but land acquisition can in turn act as an independent variable affecting population changes and tobacco prices. Even more importantly for the purposes of this study, tobacco prices and population changes can have a quite significant and immediate impact on each other. Before we can properly apply multiple regression techniques to an analysis of land acquisition, we must insure that, at any given time, there is no "simultaneous" interaction between the three variables, that we are truly able to distinguish the direction of the causal arrow. If we were not certain of the direction, then we would have to use more sophisticated statistical tools than multiple regression. Thus models of population change and tobacco prices will also be developed but, unfortunately, not as rigorously as the model of land acquisition because of a shortage of information about demographic processes and the tobacco market in the 17th century Chesapeake. For further explanation of the statistical techniques employed in this paper, see Appendix II.

Model of Land Acquisition in the Colonial Chesapeake

Based on the two theories--staples and Malthusian--and a general analysis of the associated factors and available data, the basic demand function for land has been written in terms of lagged price and lagged inverse population density and a time trend. The time trend will be used to capture linear changes in expected monetary return, effective supply price, and attitudes towards mobility for lack of more specific information; non-linear changes in these variables will be captured in the residual.

The basic demand function for land is thus:

ACRES = a0 + a1*TOBO-2 + a2*DENS-2 + a3*YEAR + u (1)

where

ACRES = new land patent acreage

TOBO-2 = nominal farm tobacco prices lagged by two years (pence sterling per pound of tobacco)

DENS-2 = TACR-2/POP-2

TACR-2 = cumulative patent acreage lagged by two years

POP-2 = tithable population lagged by two years

YEAR = patent year with 1664 as year 1

The base patent acreage was back-calculated from the 1704 quit-rent totals, including Northern Neck grants listed in the patents, but sensitivities were run based on alternative methods for determining base patent acreage. I also tested variations on this model using TOBO-2, DENS-2, TACR-2, and POP-2 in different combinations to test other plausible ways of deriving econometric models from the two theories. It is expected that the signs of TOBO-2 and POP-2 will be positive, while DENS-2, TACR-2, and YEAR will be negative. The time trend will be negative due to the reduced desirability of land further and further from navigable waters and due to reduced willingness to migrate as Chesapeake society matured. If the staples hypothesis is correct, we would expect to find a significant positive coefficient for tobacco price (TOBO-2). If the Malthusian hypothesis is correct, in this specification of the model, we would expect to find a significant negative coefficient for inverse population density (DENS-2).

Model of Population Change in the Colonial Chesapeake

In this era before the emergence of a stable native society, changes in population were mostly due to immigration from England. There is much disagreement among historians over how the labor market for indentured servants worked in the 17th century Chesapeake, especially over whether "push" or "pull" factors were more important. However, as with the question over the correct model for land acquisition, since most historians accept that both "push" and "pull" forces were at work, these forces shoud be tested within the context of one model. Using a lagged English wage index as a proxy for "push" factors and lagged tobacco prices for "pull" factors, an equation for changes in population, plus a time trend to account for population changes due to natural increase, was developed:

$$DPOP = b0 + b1*WAGE{-}1 + b2*TOBO{-}1 + b3*YEAR + v \quad (2)$$

where

DPOP = change in tithable population from previous year

WAGE-1 = the Wrigley-Schofield English wage index lagged by one year

TOBO-1 = nominal farm tobacco prices lagged by one year

YEAR = tithable population year with 1664 as year 1.

The independent variables are lagged by one year because changes in population were due to the previous year's migration. The timing of harvests in both England and the Chesapeake and the physical time of transport across the Atlantic dictated a one year lag. However, the time lag of tobacco prices is not so clear cut because there are still a lot of unanswered questions about the 17th century European tobacco market. If English tobacco merchants were the major entrepreneurs in recruiting and marketing indentured servants and they foresaw increased demand for tobacco, there might be no time lag effect on prices. If Chesapeake tobacco planters were the major entrepreneurs in the servant trade, a tobacco price lag of two years might not be unreasonable. If emigration from the Chesapeake was a major factor in population changes, then no time lag might be reasonable as transient people chose to leave or stay based on current tobacco prices. Thus, alternative models with no time lag (TOBO-0) and a lag of two years (TOBO-2) on tobacco prices were also estimated. It is expected that the sign of TOBO and TIME (due to natural increase) will be positive, while the sign of WAGE-1 will be negative.

Model of Tobacco Prices in the Colonial Chesapeake

The price of tobacco is a function of both supply side and demand side factors, complicated by two markets separated by an ocean. Unfortunately, there is not enough information to determine independent equations for supply and demand. Insufficient data is available for overall tobacco production and for European tobacco prices. Most historians have assumed price-inelastic demand and income-inelastic demand for tobacco. Indeed, until there is better information on the 17th century tobacco market, the best proxy for tobacco demand is a simple time trend. Although transportation costs and labor productivity may have changed over the course of the 17th century, there are likewise no better estimates for these data than linear time trends. Thus the best model we have for farm tobacco prices includes lagged cumulative acreage and tithable population for supply side changes and a time trend for demand side and other changes:

TOBO = c0 + c1*TACR-3 + c2*POP-1 + c3*YEAR + w (3)

where

TOBO = farm tobacco prices

TACR-3 = cumulative patent acreage lagged three years

POP-1 = total tithable population lagged one year

YEAR = tobacco price year with 1664 as year 1

Also tested was a variation of this model with the Wrigley- Schofield English wage index lagged by one year (WAGE-1) as a proxy for demand changes.

The population is lagged by one year because this year's tobacco price is based on last year's harvest. A time lag of three years for cumulative patent acreage is used as an approximation of how long it took for new patent acreage to be put into production. Having no better information for the antebellum South, Gavin Wright tried lags of two, three, and four years with little statistical difference (115). Douglass C. North noted "there was a lag of approximately four years between the peak in land sales and a large increase in cotton production" in the antebellum South (73). Lacking any better information for the colonial Chesapeake, lags of three years (TACR-3) and four years (TACR-4) were tested. It is expected that the signs of TACR-L and POP-1 will be negative, while the signs of WAGE-1 and YEAR will be positive (reflecting the increased demand for tobacco over time).

Estimation

Since the resultant system of equations involve lagged, predetermined variables, there is no problem of simultaneity. We may thus estimate the coefficients using ordinary least squares (OLS). To determine whether there are any autocorrelation problems normally associated with such time series data, a Durbin-Watson (D.W.) statistic is reported for each model.

Results

Land Acquisition (ACRES)

Table XIII contains the results of the OLS regression on ACRES using nominal farm tobacco prices. Based on F-statistics, all of the models show statistical significance at the 5% level of significance.[1] The signs of the coefficients are all as expected, except for the time trend in Models 1 & 2, where it is statistically insignificant. Thus, the positive sign on the time trend of Models 1 & 2 is less certain than the negative sign on Models 3, 4 & 5. This would tend to support the belief that land acquisition was declining due to reduced desirability of land, even after taking into account both tobacco prices and population density.

The coefficients on population (POP) in Models 1 & 2 have the correct sign (increased population led to increased demand for land and vice-versa) but are statistically insignificant. Models 1 & 2 indicate

[1] Using Sargan's test (Maddala 210), the possible autocorrelation in Model 4 was determined to be an error specification problem, not a model specification problem. The error problem in Model 4 was corrected using an autoregression with a one year lag (Model 4AR1) and the results likewise presented in Table XI.

that increased cumulative land patent acreage was more important in inhibiting further land acquisition than increased population was in promoting land acquisition.

A Chow test[2] was run to test the contention that there was a transformation in the tobacco economy of late 17th century Virginia, that there was a different function for land acquisition in the first half (1664-1685) and the second half (1686-1706) of the time period under study. For all of the models in Table XIII, there was no statistical difference between the two data sets at the 5% level of significance, thus rejecting the contention of a transformation.

The effect of tobacco prices on land acquisition is rather difficult to determine. Based on the models which combine both staples and Malthusian factors (1 & 3), tobacco prices would be dismissed as statistically insignificant. Dropping tobacco prices in Model 2 barely alters the results of Model 1. However, Models 4 & 4AR1 with land acquisition simply a function of tobacco prices and a time trend, essentially the "staples" model, indicate statistical significance for the tobacco price coefficient at the 5% level of significance. In such a situation, where the hypotheses are "nonnested since the explanatory variables under one of the hypotheses are not a subset of the explanatory variables in the other," the Davidson and MacKinnon's J-test is appropriate for testing the two hypotheses (Maddala 443-445). This test basically confirmed the results of Models 1 & 3 and led to rejection of

[2] A Chow test is an "analysis-of-variance test" which tests whether the calculated coefficients are constant over the entire time period under study. For more information, see Maddala 130-137.

the staples thesis and acceptance of the Malthusian thesis. Thus, the Malthusian hypothesis seems well-supported by the study of land acquisition in the colonial Chesapeake.

Rejecting the staples thesis, either Malthusian Model 2 or Model 5 could be considered appropriate. Model 5, the simplest of the Malthusian models, expressing land acquisition as a function of only inverse population density and a time trend, is in many respects the best model. The intercept is quite precisely determined and the time trend shows statistical significance. For statistical as well as aesthetic reasons, Model 5 is the preferred model.

Model 5 can be used to determine "optimum density," the inverse population density at which there would be zero demand for land. Setting YEAR=1, the optimum density for 1664 was 199 acres per tithable; for YEAR=43, the optimum density for 1706 was 107 acres per tithables. The strong linear decline in optimum density supports the hypotheses that optimum density was not a constant in the colonial Chesapeake and that attitudes towards mobility were becoming more restrictive with time.

Using deflated tobacco prices had negligible effect on the results (Table XIV). Excluding Northern Neck land grants did reduce the significance of the coefficients, but not enough to change the conclusions of this study. (Population estimates for the area south of the Rappahannock were based on Menard's annual tithable series corrected by linearly interpolated population ratios estimated from available county tithable lists.) For a proper analysis of land acquisition in colonial Virginia, the Northern Neck appears to play a significant role and the patent data should be enlarged to include available Northern Neck patent data.

The inverse population densities in Table XIII were calculated based on Craven's headrights for the years 1634-1658. Alternatively, density was calculated based on the lower quit-rent acreage and the results are listed in Table XIV and shows little statistical difference over the base case. This is to be expected as changing the base acreage merely shifts the cumulative acreage curve up and down and does not change its general shape.

Lapsed land acquisition, listed in Table XIV, shows a similar result to the other models, although the overall correlation is weaker. Population density, although barely insignificant at the 5% level of significance, still far outweighs the effect of tobacco prices. Interestingly, there is no significant time trend, indicating that the increases in lapsed land acquisition over time noted earlier were due mostly to declines in inverse population density.

The contention that the behavior of small patentees differed from large patentees was also tested and the results are also listed in Table XIV. The patents were divided into three groups: greater than or equal to 1000 acres, less than 1000 acres, and less than 500 acres. Although the significance of the coefficients declined with each sub-group, each sub-group fits the general model. If anything, large patentees seem to be the least predictable group. Overall, though, it appears that the small planter/speculator responded to the same pushes and pulls as the large planter/speculator.

Population Changes (DPOP)

Table XV contains the results of the OLS regression on population changes (DPOP) using nominal farm tobacco prices and the English wage index. The results are also presented graphically in Figure IV. Based on F-statistics, all of the models show statistical significance at the 5% level of significance. The signs of the coefficients are all as expected, but not always statistically significant. Both graph and table indicate that tithable population grew most rapidly when English wages were low and tobacco prices high and tithable population grew least rapidly when English wages were high and tobacco prices low. All of the models show a statistically significant positive time trend which would be expected if natural increase was a growing part of population increase. Although colonial death crises may have significantly impacted the rise in tithable population in some years, the effect of such crises was not significant enough to distort the relationship between population change, English wages, and tobacco prices.[3]

The significance of the coefficients of the lagged English wage index and lagged tobacco prices varies with each model. The

[3] However, as determined by a Park-Glejser test (Maddala 162-167), the models are all heteroskedastic with time which may weaken the results. This heteroskedasticity in the residuals is due to greater precision in the population data at the end of the time period. The early population figures are based on interpolated averages but the later figures are based on actual colony-wide census records which show greater fluctuations.

theoretically most believable model, Model 2, also has the highest R^2 and indicates statistical significance for both variables lagged one year. Further lags of tobacco price decrease the statistical significance of tobacco prices. This analysis indicates that "push" and "pull" factors were both important factors of population increase.

As with the land acquisition analysis, a Chow test[4] was run to test the contention that there was a different function for population increase in the first half (1664-1685) and the second half (1686-1706) of the time period under study. For all of the models in Table XV, there was no statistical dif-ference between the two data sets at the 5% level of significance, thus rejecting the contention of a transformation.

Nominal Farm Tobacco Prices (TOBO)

Table XVI contains the results of the OLS regressions of lagged cumulative acreage (TACR) and population (POP) on TOBO. Based on F-statistics, all of the models show statistical significance at the 5% level of significance. The intercept is statistically significant and the time trends are statistically insignificant at the 5% level of significance. There is no significant difference between a three or four year lag in cumulative patent acreage. However, all of the models suffer from strong autocorrelation.

The signs on the coefficients of cumulative acreage (TACR) are correctly negative, but the signs for cumulative population (POP) are correct only for Models 1 & 2. The coefficient on the English wage index is statistically significant but the sign is wrong since WAGE-1 is supposed to be an indicator of wage-induced demand for tobacco in England. If tobacco was wage inelastic, or the English tobacco demand was not a large part of the market, or the English wage index was not a

[4] See fn. 2.

good proxy for European wages, then we might expect this coefficient to be statistically insignificant. But this is not the case. The problem is believed to arise due to the strong effect of wages on immigration from England; lagged English wages are a much better indication of supply-side changes than demand-side changes. Thus, including WAGE-1 in the model obscures the effect of the other supply-side variables.

Either Model 1 or Model 2, then, is the preferred model, although both are rather poor. A Sargan's test[5] indicated there were model specification problems at the 5% level of significance so a modified model was tested which included, as independent variables, tobacco prices lagged by one year (TOBO-1), average lagged cumulative patent acreage (TACRA), average lagged population (POPA) and a time trend (YEAR). After correcting this model for error specification problems, all of the coefficients were found to be insignificant.

Because using cumulative acreage and population might not capture the short-run fluctuations in supply-side changes, the model for tobacco prices was recalculated using annual patent acreage lagged three and four years (ACRE-3 and ACRE-4) and annual population changes lagged one year (DPOP-1) and the rsults reported in Table XVII. Based on F-statistics, all of the models show statistical significance at the 5% level of significance. All of the intercepts are statisticaly significant. The time trend is negative but statistically significant at the 5% level of significance only in Model 1. Unlike the cumulative totals, annual changes show a statistical improvement in increasing the lag from three

---

[5] See fn. 1.

to four years.

However, all of the signs are incorrect and all of the models still indicate strong autocorrelation. Inclusion of the English wage index has noticeably less effect with annual changes than with cumulative totals reflecting the annual nature of the index. In an attempt to salvage this model of tobacco prices, the same tests and corrections were run as previously done for the other tobacco price models, with similar negative results. The final corrected model had no significant coefficients. Further work on the 17th century tobacco market will be required in order to resolve the problems presented by these tobacco price models.

CHAPTER VI

ALTERNATIVE MODELS

This study has not determined the exact role of fluctuations in the tobacco economy on growth and development of the colonial Chesapeake. Although rigorous hypothesis testing indicates that the importance of short-term fluctuations in tobacco prices has been over-emphasized, alternative models show significant effects of tobacco prices which at present cannot be explained. Table XVIII tests for the effect of various patent time lags and shows that with a four or five year time lag, tobacco prices had a significant effect on demand for land. Although there is no evidence that the patent process took anywhere near four years, decisions to patent land could have preceded formal patent application by several years and thus a patent lag of four years is not out of the question. However, the similarity of results for large patentees (who might have been more responsive to prices) and small patentees tends to make such time lags unlikely.

This analysis is confirmed by the sample of headright certificate totals from Accomack, Nortumberland, Lancaster and York counties, tested against the land demand model and presented in Table XIX. The similarity in the interweaving effect of density and tobacco prices with different time lags confirms the conclusions of Table VI that acquisition of headrights preceded land acquisition by 0-1 years. Thus demand for headrights either coincides with or slightly lags behind demand for land (if one accepts a 1-2 year delay in the land patent process).

A much better explanation of the lagged effect of tobacco prices on land acquisition would be that tobacco prices have a lagged effect on the factors which affect demand for land. Although regression analysis shows no statistically significant lagged relationship between tobacco prices and population density, perhaps tobacco prices affect other non-Malthusian factors of land demand. The lagged 4-5 year effect of tobacco prices on land acquisition could be seen as support for Clemens's hypothesis of a two-stage process of importing servants when tobacco prices were high and then acquiring additional land perhaps a couple of years later. Possibly the planter already had sufficient land from the previous cycle of servant importation and land acquisition to delay the acquisition of new land for 2-3 years.

However, other tests (presented below) show no such positive correlation between tobacco prices and servant importation. Using a model similar to that used to study population change (but with no time lags on the wage index), emigration from Bristol, England during the years 1654-1686 was found to be negatively correlated with lagged tobacco prices, as shown in Table XX. This tends to support the conclusion of Lewis C. Gray, mentioned earlier, that planters were less concerned with making quick profits than with maintaining a certain income and they imported servants, not when tobacco prices were high, but when tobacco prices were low. This conclusion runs counter to the findings listed in Table XV for tithable population changes which were positively correlated with tobacco prices. As Menard found in his recent study of the Bristol emigration list, I also found a significant negative relationship between the English wage index and emigration (Menard,1988,108n). The relationship between immigration, tobacco prices, and English wages is shown in Figure V.

This negative relationship between tobacco prices and English emigration might be dismissed as coincidental, except that the same negative correlations may also be found upon examination of Chesapeake unindentured servant registrations. Menard and Walsh have found positive correlation with tobacco prices when examining individual counties (Menard,1973,326-8; 1977,363-5; 1988, 115-117; Walsh,1977, 26-27), but when all these servant registrations are examined in aggregate the results are quite different, as shown in Table XXI. In both Maryland and Virginia, masters were required to bring their unindentured servants (those who did not obtain indentures before leaving England) into the county court soon after acquiring them, and it was in the masters' best interest to do so quickly to insure the maximum allowable term of indenture. Thus, a 0-1 year lag is probably the proper lag, but there are no independent estimates of this lag and longer lags are quite possible. If a 0-1 year lag is correct, then tobacco prices have an insignificant effect on servant registrations. But for longer lags, the tobacco prices have a significant and negative impact on servant registration. Servant registration correlates negatively with English wages regardless of the lag used, which supports the findings of Tables VI and XX.[6]

[6] Using a Chow test, the only model which shows significant difference between the first half of the period (1660-1685) and the second half (1686-1706) is the 0 year lag case in which tobacco prices are positive and significant only in the second half.

An alternative model for determining tobacco prices which shows great promise, although there is no theoretical basis to justify it, is simply a function of the Bristol emigration list totals to the Chesapeake (1654-1686) lagged by five years, as listed in Table XXII. The relationship is also pictured in Figure VI.[7] Along with a strong negative time trend, the negative correlation between this five-year lagged Bristol emigration and Chesapeake tobacco prices helps explain many of the fluctuations of tobacco prices in this time period. Interestingly, as might be expected from the positive correlation between population change and tobacco prices, this five year lagged Bristol emigration also shows a strong negative correlation with population changes, as shown in Table XXIII.

Although Bristol emigration could easily represent a proxy for all English emigration and this emigration may be a better proxy for supply changes than the population and acreage totals used above, why lagged by five years? One possible explanation is the normal four or five year term of indenture. Perhaps indentured servants did not reach their productive capacity until after they were freed and working for themselves, although there is no evidence to support such a conclusion. Perhaps indentured servants generally left the county upon gaining their freedom or else competition for land between freedmen and other farmers caused a general exodus from the county. The freedmen may have left the colony for good

---

[7] Using Sargan's test, all except the five-year lagged model was shown to have model specification errors at the 5% level of significance. The autocorrelation of the five-year lag model was determined to be an error specification problem and an autoregression with a one year lag produced model 5AR1, also listed in Table XX.

or otherwise avoided being recorded or for other reasons simply were not recorded on county tithable lists. Although all of this is pure conjecture and the relationship may be totally coincidental, the strength of the fit (especially in comparison to other tobacco price models tested) demands additional study of the relationship between immigration and the tobacco economy.

But if immigration to the Chesapeake is negatively correlated with tobacco prices, and immigration was the greatest source of population increase in the 17th century, why is there a strong positive correlation between population increase and tobacco prices? Indeed, as shown in Table XXIV, this strong positive correlation holds for the entire time period 1630-1730 although the relationship weakens significantly with time. (English wages only seem to be significant for the time period covered by the major part of this study.) Possibly the role of other components of population increase were more important than suspected. Most likely, people were less willing to emigrate from the colony when tobacco prices were high. Perhaps mortality was reduced when times were "good." This analysis certainly indicates the need to take a closer look at population changes in the 17th century on a local and colony-wide level.

CHAPTER VII

CONCLUSIONS

Land acquisition in the colonial era has been relatively ignored by economic and social historians. Land was simply abundant and was not the limiting factor of production in a perpetually labor-short colonial society. Colonists could go out and get land when they wanted it. But this abundance is deceiving because new land was made available to the people only by the slow, methodical process of patenting. And, as this study has tried to show, the decision to patent land was not taken lightly.

This thesis has shown the value of using social science theory to structure historical analysis. Theory organizes research agendas by identifying historical topics which need more research and by providing a formal basis to develop hypotheses which can be systematically tested. The great advantage of a study of land acqusition, to economists, demographers, and historians, is that it allows a chance to test social science theory much better than does most historical data. Land data is generally the best and most complete data from historical periods because of the ever-present concern over property rights. The ability to test such theories should help us to develop an understanding of historical causation, of why things happened the way they did.

Also I have strived to show how comparative analysis with the work done on other regions and time periods (for this study, colonial New England and the antebellum North and South) also can benefit studies of the colonial Chesapeake. With further study, the trend might be reversed so that other regions could benefit from work on the colonial Chesapeake. A comparison between two closely related historical periods can offer great insights to both periods.

Comparative analysis especially offers much hope for explorations into the interrelationships between the colonial and antebellum South. The economic studies of land acqusition in the antebellum South should be re-examined with consideration of demographic factors. The two time periods should be treated as more continuous than discontinuous. Historians of both periods at least should no longer ignore the important work being done in the other.

There is little statistical evidence to support transformation theories in the late 17th century Chesapeake. Neither land acquisition, annual population changes, unindentured servant registrations, nor immigration showed any significant changes over the time period 1660-1706. However, there is some evidence of a "transformation" in the early 18th century provided by the study of annual population changes. Population change in the early 18th century became less dependent on fluctuations in the tobacco economy and English wage conditions, which could possibly reflect the growth of importance of natural increase and/or the major shift to slave labor, as Kulikoff has suggested.

This thesis shows that the interrelationship between economic and demographic forces in the 17th century Chesapeake were indeed complex. Although the econometric analysis of land acquisition finds that the Malthusian hypothesis is superior to the staples thesis, the effect of short-term fluctuations in the tobacco economy cannot be ignored. The analysis of alternative models indicates that further work is necessary before there will be a clear explanation of the economic behavior of tobacco planters and merchants. Additional effort should be devoted to continuing the work of Yasuba, Forster and Tucker, Rutman, and Easterlin in identifying why population density is such a strong explanatory variable of early American economic and demographic development.

Perhaps lower tobacco prices caused small freeholders and landless tenants, men with small ties to the local economy, to migrate out of the colony, thus retarding population growth. But planters with great capital invested in the Chesapeake would have been unable to migrate and would even have been forced to import additional servants and slaves to maintain their cash flow, pay off their debts, and make up for the loss in the labor force due to out-migration. The further study of the interrelationship between immigration to the Chesapeake and the tobacco economy offers great promise. It is possible that all of the ups and downs of the tobacco economy were driven by changes in the immigration flow which were totally independent of planter behavior, or perhaps even exacerbated by planter behavior (if planters brought over more servants when tobacco prices were low).

Perhaps the Malthusian and staples hypotheses represent merely two sides of a complex, multi-faceted theory of human behavior. The Malthusian theory could explain intra-colonial migration and the staples theory could explain inter-colonial migration. The staples model reflects differences in opportunity between the tobacco economy and other colonial economies; the Malthusian model reflects differences in opportunity within the tobacco economy. Migration between England and the colonies might be explained by a combination of the two theories, since the colonies were both a part of and separate from the English economy.

Neither theory is complete in and of itself. Although Malthusian theory might predict the level of land acquisition or intra-colonial migration, this theory tells us little about the direction of that acquisition or migration in the colonial Chesapeake. For the staples theory, 17th century planters often did not respond as quickly or in the way that staples theory would predict. With more studies such as this, it may be possible for a more generalized theory to be developed which subsumes both the Malthusian and staples hypotheses, combines equally economic and demographic factors, and explains both individual and aggregate behavior.

APPENDIX I

NAMECODING SYSTEM AND ANALYSIS

Because of the vagaries of 17th century English orthography and the peculiarities of 17th century English surnames, any computer data base analysis of individual land acquisition must include some sort of namecoding system. In trying to match individuals between two different records (for example, patent and patent or rent roll and patent), I wanted to be sure that I did not miss possible individual matches yet, at the same time, it would have been too costly to go through all possible matches by hand to ensure that the matches were realistic.

If the data base is small enough, one simply could simply standardize the spelling of each name when entered. However, when the data base gets rather large, this practice becomes inefficient and impractical. Also, for the 17th century, I would be very hard pressed to determine the standard spelling of any surname due to the wide variety of surnames which are very uncommon in the modern United States.

The most common method of standardizing names is to use a coding system which reduces names to an alphanumeric code based on similar sounds of various letters or combinations of letters. The most often used coding system is the Soundex system developed for genealogical tracing of individuals in the 1880, 1900, and 1910 censuses. In the Soundex system, "the surname is indexed by using the initial letter; the letter is followed by three numerical digits based on the consonants that follow the surname initial. Consonants that sound somewhat alike, though widely separated in the alphabet, are drawn together under one numeric code." The consonant groupings are: b,p,f,v; c,s,k,g,j,q,x,z; d,t; l; m,n; r. Vowels, unless initial, are not coded (Helmbold 54-55). Bouchard and Pouyez present an interesting description of alternative namecoding systems developed for French Quebec names (119-125).

Although I could have used the Soundex system as is, I found that it was both too much and too little. At least for 17th century English surnames, the Soundex too often combined unsimilar names and missed similar names. The major problems were the loss of information with vowel elimination and the peculiar pronunciation of various consonant-consonant and vowel-consonant combinations in 17th century English speech.

Like the Soundex, I realized that the key to a good namecoding system was to envision each surname as code representing a combination of sounds. The written name, especially in the 17th century, is merely an approximation of an oral name and it is the oral name we need to try to recreate. Dictionaries of English pronunciation (Noory; Wright) were examined to identify problematic consonant and vowel sound groupings. Various 16th and 17th century English spelling-books (Fox and Hookes; Price; Bullokar) were studied to determine both 17th century "standard" English spelling and pronunciation and typical English spelling problems.

The most troublesome problem was consonants which changed their sound in various vowel-consonant and consonant-consonant combinations. This includes the standard English "soft" and "hard" consonants like "c","g", and "ch" but also some non-standard English variable-sound consonants. For anyone familiar with traditional southern speech, 17th century English pronunciation is not totally foreign. The most notorious consonant is "r" which is sometimes pronouced (after consonants and before vowels) and sometimes is not (before consonants and after vowels). The "r" also tends to convert all preceding vowels into a schwa. Similar problems arise for other consonants like "l", "m", and "n" in various combinations with other consonants and vowels.

In studying English names, it was clearly obvious that the only important distinguishing vowel sounds were the initial vowel sound and a final, ending vowel sound (if the name ended in a vowel sound, such as BIGELOW). The prototype namecoding system included a vowel conversion routine which converted vowels in various vowel-vowel and vowel-consonant combinations to standard vowel sounds:

A = long A

E = short A,E,I,Y

I = long E

O = schwa, long O, short O,U

U = long U

Y = long I,Y

However, after various tests, this vowel conversion routine was found to create as many problems as it solved. In reality, there is too much variety in spellings of various vowel sounds, and too much overlap in vowel sound groupings. Thus, in my final namecoding system I settled on two vowel sounds "E" (for long and short A,E,I & Y) and "O" (for schwa, long and short O & U). I could have eliminated all vowel sounds like the Soundex, but the two-vowel system was found to be superior for reducing the number of misgroupings. Also more efficient than simply eliminating the vowel sound is to at least mark the location of the initial vowel sound to distinguish between such common initial consonant-vowel combinations as "BAL" and "BLA". All final, ending vowel sounds were simply reduced to "O".

My namecoding system, similar to the soundex, reduces consonants sounds down to numeric representation:

1 = B,P,F,V

2 = S,Z

3 = K,G,Q

4 = D,T

5 = L

6 = M,N

7 = R

8 = W

9 = H

Particular attention was paid to defining the "R","W", and "H" sounds.

After deriving the basic formula, the namecoding routine was tested using various lists of problematic colonial Virginia and English surnames (Green; Hall; Bardsley; Ewer) and then fine tuned. As an example, after running through this namecoding routine, the name "MOUNTFORD" is converted to "MO614". The initial consonant is maintained; the vowel "OU" is reduced to "O"; the consonant combination "NTF" is reduced to "61" due to the tendency in the 17th English language to drop the middle consonant in a three-consonant combination headed by "M" or "N"; the second vowel sound is dropped; and the final "RD" is converted to "4" due to the tendency for "R" after a vowel and before a consonant to blend into the final consonant. (This explanation is just a simplification of these conversions which occur in the namecoding routine, but give the basic gist.) In a similar way, the names "MANFORD," "MOMFORD," "MONTFORT," "MOUNFORD," "MOUNTFORT," "MUMFORD," "MUMFORT," "MUMFORT," "MUMPFORD," and "MUNFORD," all of which appear in the patent records, are likewise reduced to "MO614". This agrees with B.W. Green's statement that the Virginia names spelled Montford or Munford was actually pronounced Mumford (15).

Finally, all of the available surnames in the patent records from 1660-1706 were converted and the surname groupings analyzed for irregularities. These irregularities were then corrected using separate cross-reference tables for both names and codes. All given names and abbreviations of given names were standardized using a cross-reference table. For certain unusual given names, the patent records were searched to see if there were transcription or spelling errors. For all matching analysis, because of the uncertainty associated with surnames, it was absolutely critical to have standardized given names. The matches described in this report all required an exact match of both the given name and namecode (and sometimes county, date, or acreage). This double or triple matching eliminated most spurious matches since most given name-namecode combinations in 17th century Virginia were still rather unique due to the immigrant source and smallness of the population.

In practice, two different coding systems were developed, one for headright analysis and another for landowner analysis. The namecoding system described above was used for landowner analysis but a simplified, more generalized code was required for headright analysis due to greater variation in spelling of headright names. This was especially necessary for analysis of headrights who later became landowners to avoid missing a possible match. In this simplified code all vowels in the code were reduced to "E" and certain troublesome consonants ("W","H", final "S", final "R") were eliminated. Because this simplified code had a tendency to overmatch (to match unrelated names), this required a follow-up hand search through the derived name matches to eliminate unreasonable matches.

For the analysis of headright abuse of individual headrights, sensitivities were run eliminating namecodes of less than 3 or 4 characters (which were more likely to combine unrelated names), common namecodes like "S6E4" (Smith) and "B7E6" (Brown), and the most common given names (John, William, Thomas, Robert, Richard, Mary). The percentages were then prorated based on the size of the sample.

For the analysis of serial abuse in headrights, a headright was paired with the preceding headright (in the list contained within the patent) and then the paired headrights matched with similar paired headrights. Thus "Anne Green, Thomas Bowen" in one patent might be matched with "Ann Growen, Thom. Boon" in another patent. Each headright list was then examined for pattern of such matches, taking into account spelling or transcriptions errors (which were rampant in the headright lists) which might intermittently break the serial pattern.

APPENDIX II

ECONOMETRIC ANALYSIS

This thesis employs multiple regression analysis described in any standard statistics textbook (e.g., Maddala). All models are expressed in the form of the algebraic equation:

$$Y = a0 + a1*X1 + a2*X2 + ... + an*Xn + u$$

where,

Y is the dependent variable (variable to be explained) on the left-hand side of the equation

X1,...,Xn are the independent or explanatory variables on the right-hand side of the equation

a0 is a constant coefficient (similar to the y-intercept in a simple algebraic equation)

a1,...,an are the coefficients of the independent variables

u is the error term

Multiple regression analysis determines the coefficients a0,...,an which provide the "best fit" upon inserting sets of data into the equations. Since the data involved in this analysis is annual time series data, the "set of data" is that unique combination of dependent and independent variables which occur in any particular year, say 1696. However, when a lagged effect occurs, say when tobacco prices in 1696 affect land patent acquisition in 1698, then lagged independent variables are used, represented in the form Xn-L where L is the number of years of lag. Thus, tobacco prices in 1696 lagged by two years (TOBO-2) are treated as if they were tobacco prices in 1698.

Since the calculated set of coefficients a0,...,an is only a "best fit," any particular set of data will rarely fit the equation perfectly so an error term u, which can be either positive or negative and differs for each set of data, is included in the equation.

The tables of multiple regression results list the values of the calculated coefficients a0,...,an, and list the t-statistics in parentheses underneath each coefficient. A t-statistic is the measure of confidence that the listed coefficient is not merely random (i.e., is significantly different than zero), and is calculated by dividing the coefficient by the standard deviation of the coefficient. For most analyses presented here, t-statistics of 2 or greater indicate that there is statistically less than a 5% chance that the coefficient is purely random. Lower t-statistics indicate a much greater chance of randomness and 5% is considered by most econometricians the maximum degree of chance acceptable when assessing statistical significance. Coefficients that have less than a 5% chance of being random and are thus statistically significant at the 5% level of significance are specially indicated by an *. T-statistics are listed as absolute values (without + or - signs) simply for ease of reading, because the t-statistic will always have the same sign as the coefficient.

A measure commonly presented in econometric analyses is the R^2 statistic, which is a measure of the fraction of variation in the dependent variable that is explained by the "best fit" equation. Normally the higher the R^2 the better the model, but the R^2 statistic can be quite deceiving and is not a proper statistic for judging the statistical significance of a model because it heavily dependent on the type of model and data being tested. A model which uses a lot of individual level data and has an R^2 of 0.15 may be much better than a model which uses aggregated data and has an R^2 of 0.90. However, when comparing two similar models using the same data, R^2 provides a quick check of which is the better model.

Because we are dealing with time series data which often tends to be cyclical in nature, one of the most important statistics in every table presented is the Durbin-Watson or D.W. statistic. A major assumption of multiple regression analysis is that the error term (u) for any set of data should be random. For annual time series data, however, this year's error term is often related to last year's error term. For example, if last year's prediction was high, then this year's prediction might tend to be high. If the model tends to overpredict for a few years and then underpredict for a few years, this is called positive autocorrelation which is the most common problem in normal time series data. (If the model bounces back and forth every year between overprediction and underprediction, this is called negative autocorrelation which is much less common.) Often the problem occurs because the independent variable cycles continually lag behind or lead the dependent variable cycles and so the cycles never synchronize. Autocorrelation could be due to problems either with the model (model specification error) or with the data (error specification error) and there are statistical ways to test for this, such as Sargan's test (Maddala 210). For most of the models presented here, the model is the problem. When dealing with lagged effects, the most likely cause of model error leading to autocorrelation is choosing the wrong lag time.

The Durbin-Watson coefficient is a measure of autocorrelation on a scale of 0 to 4 with 0 (perfect positive autocorrelation), 2 (no autocorrelation) and 4 (perfect negative autocorrelation). As with t-statistics, we are interested in when the odds of autocorrelation being problematic have less than a 5% chance of being random. This depends heavily on the length of the time series and the number of independent variables in the model. For the models and data examined in this paper, positive autocorrelation generally becomes problematic when the Durbin-Watson statistic falls below about 1.6.

Other statistical techniques are discussed in the appropriate footnotes.

VITA

Bruce Chandler Baird, Jr.

Born in New Orleans, Louisiana, November 22, 1956. Graduated from Clear High School in Houston, Texas, May 1974. Received B.S. in Chemical Engineering from Texas A&M University, May 1977. Worked for Atlantic Richfield Company from June 1977 to March 1982. Studied at University of Alaska at Fairbanks and received secondary teacher's certificate. M.A. candidate, College of William & Mary, with a concentration in American Studies.

In September 1989, the author entered the University of Florida as a Presidential Fellow in the Department of History.