Abstract

The time it takes for the completion of major development in areas that changed from rural to urban and are added to urban areas is examined for 59 large urban areas in the United States. Discrete-time survival models are used to predict the probability of the areas reaching completion of development over time and are estimated using logistic regression. This probability declines over time and is best predicted using the inverse of time. The percent change in the number of housing units in the entire urban area in each decade is significantly related to the probability of completing major development and therefore time to completion of major development, with more rapidly growing areas taking longer. Predictions of the model are used to estimate the median time to completion of major development. Estimates of the median are made for values of percent change in the urban area 1 standard deviation below and above the mean for all areas and decades. The predicted median time to completion of major development for more slowly growing areas is 30.9 years while the predicted median for more rapidly growing areas is 34.2 years.

Introduction

When land at the fringe of an urban area sees significant new development and an increase in density and is added to an urban area, the process of development is far from over. Rather, this is merely the start of a period of development that will typically continue over several decades until the major development of the area has been completed. This development over an extended period of time has been described for large urban areas in the United States from 1950 to 2010 in an earlier paper (Ottensmann 2019).

In an early study Blumenfeld (1954) described the successive zones of the expanding Philadelphia area as going through a sequence of first slow growth, then more rapid growth, followed by slower growth again. He called the zone of most rapid growth “the crest of the wave” of metropolitan expansion. Boyce (1971), Hart (1991),
and Peiser (2001) have all adopted the metaphor of the wave to describe the process of development at the urban fringe.

Harvey and Clark (1965) and Clawson (1971) have observed that the scattered, leapfrog development in newly urbanized areas, frequently referred to as sprawl, was often followed by later infill development. Clawson noted that little information was available on the rate at which this occurred but that some case studies indicated that this might take a long time. Bruegmann (2005) suggested that “scattered development often results, in the end, in densities higher than those that would have been achieved with continuous development because it allows for infill at higher densities in the second and third waves of urban growth.” Clawson (1971) and Schmid (2015) have made similar arguments.

Studies of development at the urban fringe have tended to focus on individual areas because of issues related to the assembly of the necessary data. Irwin and Bockstael (2004) and Irwin, Bell, and Geoghegan (2006) used survival analysis methods including Cox proportional hazards models to look at development at the parcel level in Calvert County, Maryland, over an 8-year period. Templeton, Hite, and Sohngen (2006) took a similar approach for development over 10 years in Delaware County, Ohio.

Survival analysis is particularly appropriate for questions involving the length of time it takes most of the development in an area to be completed once it has become part of an urban area. This paper uses data on the number of housing units in each census year up to 2010 in areas added to 59 large urban areas in 1960, 1970, and 1980. Discrete-time survival models are used to examine the time to the completion of major urban development in those areas and the relationship to the rate of growth of the entire urban area in each decade. The urban patterns dataset with data on housing units by census tract from 1950 to 2010 and the delineation of urban areas for each census year is described in the following section. This is followed by the descriptions of the areas being considered as having become urban and having been added to the urban area in each year along with the description of the development of these areas over time. The study requires the determination of the time at which each of these areas concludes the period of major development and the specification of discrete time survival models for the analysis, which is addressed next. The first results focus on the role of time alone in predicting the probability of reaching the end of major development in each decade, which is called the baseline hazard. This is followed by models incorporating the effect of the percent change in the urban area in each decade on these probabilities and thus time to completion of major development. Extensive graphical interpretations of these results are presented as are several models exploring the robustness of the results. An appendix provides more detail on how the time to the conclusion of major development was determined.
Urban Patterns Data

The current research is part of a larger research project that looks at patterns of development within 59 large urban areas in the United States from 1950 to 2010. These areas, the urban areas delineated for each census year, form the context within which the analysis of time to completion of major development is being undertaken. This section starts by identifying the urban areas included in the dataset, describes the housing unit data for census tracts that is the core of the data, and discusses how these data have been used to define urban areas from 1950 to 2010.

Urban Areas Included

This research uses a dataset for the analysis of urban patterns over time that was developed with data on numbers of housing units in census tracts for large urban areas in the United States from 1950 to 2010. The tracts for urban portions of metropolitan areas were identified within the Combined Statistical Areas (CSAs) as delineated by the Office of Management and Budget for 2013 (U.S. Bureau of the Census 2013). CSAs were used rather than the more commonly employed Metropolitan Statistical Areas (MSAs) as it was felt they better represented the full extent of the metropolitan areas, including those instances in which 2 or more MSAs should more properly be considered to be parts of a single area (Ottensmann 2017). For those MSAs which were not incorporated into a CSA, the MSA was used.

The 59 CSAs and MSAs with 2010 populations over one million were selected for the creation of the dataset. A number of these areas had multiple large centers associated with separate urban areas that had grown together. This posed the issue of identifying those cases in which a second or third urban area could be considered sufficiently large in relation to the largest area to be included as an additional center around which urban development occurred. The decision was made by comparing the populations of census Urbanized Areas (either from the current census or the last census in which the areas were separate) with the largest area. An area was considered to be an additional center if its population were greater than 28 percent of the population of the largest area. The three areas included with the lowest percentages were Akron (with Cleveland), Tacoma (with Seattle), and Providence (with Boston). Areas with multiple centers have each center included in the name of the urban area.

Housing Unit Data for Census Tracts

The primary data source for this research was the Neighborhood Change Database developed by the Urban Institute and Geolytics (2003). This unique dataset provides census tract data from the 1970 through 2000 censuses, with the data for 1970
through 1990 normalized to the 2000 census tract boundaries. Population and housing unit data from the 2010 census were added by aggregating the counts from the 2010 census block data (U.S. Bureau of the Census 2012).

Housing units and housing unit densities—the numbers of housing units divided by the land areas of the tracts in square miles—are used in this research rather than the more commonly employed population and population density measures for two reasons. Housing units better represent the physical pattern of urban development as they are relatively fixed, while the population of an area can change without any changes in the stock of housing. Other studies of urban patterns have made similar arguments for choosing housing units over population, for example Galster, et al. (2001), Theobald (2001), Radeloff, Hammer, and Stewart (2005), and Paulsen (2014).

Using housing units also allows the extension of the analysis to census years prior to 1970. The census provides data on housing units classified by the year in which the structure was built, and these data are included in the Neighborhood Change Database. The 1970 year-built data can be used to estimate the numbers of housing units present in the census tracts for 1940, 1950, and 1960. Several prior studies have used the housing units by year-built data to make estimates for prior years in this manner, though they have used more recent census data to make the estimates, not the earlier 1970 census data (Radeloff, et al. 2001; Theobald 2001; Hammer, et al. 2004; Radeloff, Hammer, and Stewart 2005).

Sources of error in these housing unit estimates for earlier years from the year-built data arise from imperfect knowledge of the year in which the structure was built and from changes to the housing stock due to demolitions, subdivisions, and conversions to or from nonresidential uses. These errors increase for estimates farther back in time. Numbers of housing units for 1970 to 1990 were estimated from the 2000 year-built data and compared with the census counts in the Neighborhood Change Database. The judgment was made that estimates 2 decades back involved acceptable levels of error, but this was not the case for 3 decades back. As a result, the decision was made to use the housing unit estimates for 1950 and 1960 but not for 1940.

**Defining Urban Areas**

Urban areas have been defined for the broader urban patterns research for each census year since 1950 consisting of those tracts contiguous to an urban center meeting a minimum housing unit density threshold. (This is comparable to the way in which the census defines Urbanized Areas using blocks and larger units and Paulsen (2012) defined urban areas using block groups.) For the definition of Urbanized Areas for the 2000 and 2010 censuses, a minimum population density of 500 persons per square mile was required for an area to be included (U.S. Bureau of the Census 2002, 2011). Using the ratio of population to housing units for the nation as a whole in both 2000 and 2010
of 2.34 persons per unit, a density of 500 persons per square mile is almost exactly equivalent to 1 housing unit per 3 acres or 213.33 units per square mile. This was used as the minimum urban density threshold. Note that this is a measure of gross density, not lot size, as the areas of roads, nonresidential uses, and vacant land are included.

To provide for a set of urban areas that represents the cumulative expansion of the urban areas over time, a further condition was imposed that if a census tract does not exceed the minimum housing unit density and has not been included in the urban area in any given year, it will not be included in urban areas delineated in earlier years even if the density exceeds the minimum. The rule has been imposed in this direction—if rural, then not urban earlier—rather than in the opposite direction—if urban, then urban later—because the more recent data are considered to be generally more accurate.\(^1\)

**Development over Time**

This section sets the context for addressing the question of the length of time it takes for newly urbanized areas to become largely developed. The first part describes the definition of the areas being considered, the portions of each urban area considered to have been developed as urban and added to the urban area in each year. This is followed by the description of how those areas developed, the densities and changes in densities in the decades following their being developed as urban. Finally, a brief mention is made of the relationship of that development to the percentage changes in the entire urban areas. This section relies on the analysis and findings from the earlier paper on development over time (Ottensmann 2019).

**Defining the Areas Developed as Urban**

The urban patterns dataset described in the previous section has census tract data with the number of housing units for 59 large areas from 1950 to 2010. For each census year the urban areas have been identified consisting of those contiguous tracts with housing unit densities exceeding 213.33 units per square mile. As the urban areas expand, new tracts are added for each decade. So for each census from 1960 to 2010, the set of tracts added to the urban area can be identified. These tracts represent the starting point for establishing the areas that have been developed as urban for each year.

All or nearly all of the tracts contiguous to the original urban area at the start of the decade will have seen development that has moved them from below urban density levels to above. If their densities had exceeded the minimum urban density threshold

\(^1\) More detail on the construction of the dataset and the delineation of the urban areas is provided in Ottensmann (2014).
earlier, they would have been added to the urban area at that time. This will not have been the case for all of the tracts added to an urban area over the course of any decade, however. As the urban area expands with development in contiguous tracts increasing their densities, the area may become contiguous to tracts that already exceeded the urban density minimum at earlier census years but were not added to the urban area because they were not then contiguous to the urban area. These tracts could have just achieved urban densities at the previous census or could have been urban for many decades. They range from isolated tracts to small towns to sometimes large urban areas that have now become contiguous to the main urban area and are being merged into it. These previously urban tracts may have experienced all manner of development histories that would not reflect patterns of development of newly urban areas. Therefore tracts that were added to the urban area that exceeded the minimum urban density of 213.33 units per acre at the previous census are not included in the area being considered to have been developed as urban.

A generally much smaller number of tracts may be added to the urban area in any year that do not meet the minimum urban density threshold. These additions may have been made under several conditions. First, if a tract or group of tracts with below urban densities is completely surrounded by urban tracts and has a total land area of less than 5 square miles, the tracts are considered to be urban and are added to the urban area. Larger enclosed areas are not added. This parallels a rule used by the Census in defining Urbanized Areas. In addition, if such an area is surrounded by a combination of newly urban tracts and water and if over half the perimeter is adjacent to the urban tracts, the area is added as urban. In a few instances enclosed areas larger than 5 square miles have been added when satellite imagery showed the areas to be nearly completely developed. These were typically very large industrial areas.

The second condition involves tracts containing major airports. Again, the Census considers the areas of major airports to be urban and includes them in Urbanized Areas. For the urban patterns dataset, tracts that included airports with at least 250 emplanements per year and that had an area of less than 10 square miles were included in the urban area if adjacent. In a handful of cases, larger tracts were added that included extremely large airports developed recently, where the tract was completely or largely surrounded by other urban tracts, and where satellite imagery showed most of the remainder of the tract to be developed.

In defining the urban areas, tracts separated by water were considered to be contiguous, again following the Census (think Manhattan and Brooklyn, San Francisco and Oakland, for example.) The practice in delineating census tracts around major rivers flowing through urban areas was somewhat inconsistent, however. In many instances, urban density tracts on both sides of the river extended to the middle of the river and the tracts were contiguous. But in some instances areas including the river and adjacent low-lying lands were included in separate tracts that did not meet the density
minimum. Tracts that were long and narrow, extending along the river contiguous to urban density tracts on both sides were included in the urban area. This was the case for Kansas City, for example. On the other hand, if the tract extended far from the urban area, it was not considered urban but the urban tracts on either side were considered to be contiguous and part of the urban area. This resulted in the discontinuous urban area including Memphis and West Memphis, separated by the Mississippi River.

The areas developed as urban for each census then consisted of those tracts that were added to the urban area since the previous census, that had a housing unit density of less than 213.33 units per square mile (non-urban) at the previous census, and had a density exceeding this level at the time they were added to the urban area. These areas were identified for each of the 59 urban areas for each census from 1960 through 2010. Given how the areas have been defined, the areas developed as urban for the different censuses include different census tracts and are distinct, non-overlapping areas. In 2 instances, no tracts met the criteria and no areas developed as urban were identified, for Dayton in 2000 and Rochester in 2010. So only 58 areas are included for each of those years. This gives 352 areas developed as urban for the 6 censuses and 59 urban areas, not including those 2 for which no areas were developed as urban.

How Areas Developed Over Time

The total number of housing units and the total land area has been aggregated for each of the areas developed as urban for each census from 1950 to 2010. Magnitudes obviously vary widely depending both on the overall size of the urban area and the extent of new development and urban expansion of the area in each decade. Housing unit densities in units per square mile have therefore been calculated to provide for comparisons across areas and over time. In addition to the densities for each census, the change in density and the percent change in density are also considered.

To see how the areas developed as urban developed over time, we consider the densities of the areas beginning with the decade before they were developed as urban and for all of the decades following. Note that since the data for all of the areas extends only to 2010, the areas developed as urban in different years will have differing numbers of decades of densities after development as urban. The areas developed as urban in 1960 will have densities for 5 decades after while the areas developed as urban in 2000 will have densities only for 1 decade after, for 2010 (and the areas developed as urban in 2010 will have none).

To summarize these data, the mean densities for all of the areas developed as urban was calculated for each decade relative to the year developed as urban. For example, for the first decade relative to development as urban, this includes for the areas developed as urban in 1960 the densities in 1970, for those areas developed as
urban in 1970, the densities in 1980, and so forth. These mean densities were plotted against the decade relative to the year developed as urban in the graph in Figure 1.

From the decade before development as urban (-1) to the decade developed as urban (0), housing unit density necessarily increased from below the urban density threshold of 213 units per square mile to a level well above. But that hardly constituted the completion of development of the area. Mean density continued to increase steadily over the entire 5 decades (50 years) relative to the time the area was developed as urban. The development of these areas was a long-term process that continued long after an area had crossed the urban density threshold and was been added to the urban area.

The graph combines the data for areas developed as urban in the different years. The densities for the later years reflect progressively fewer densities because only the areas developed as urban in the earlier years have data for this many decades after. But the general pattern of these increases in density over time hold for areas developed as urban in each year, to the extent of the data available. More detail is provided in the previous paper describing development over time for these areas (Ottensmann 2019).

The density of an area at each decade relative to the year developed as urban is, of course, the sum of all of the development that occurred in the previous years (less, of course, demolitions or conversions). A further perspective on development over time

Figure 1. Mean Housing Unit Density for Areas Developed as Urban 1960-2010 by Decade Relative to Year Developed as Urban.
is provided by looking at the amount of development in each decade after the time the area was developed as urban. Because the areas vary greatly in size, it is appropriate to look not at the change in the number of housing units but the change in the density. Figure 2 graphs the mean change in density for each decade following development as urban for all of the areas for which the data are available.

Development not only was not completed in the decade prior to the development of the area as urban, that was not even the decade having the greatest amount of development. On average, the increase in density was slightly higher in the first decade after the area became urban. That was the peak. In the following decades, the amount of new development (the increase in housing unit density) became steadily smaller, declining to under 100 units per acre by decades 4 and 5.

An alternative to the increase in density, the amount of new development, is the increase relative to the existing level of development. Figure 3 shows the mean percentage change in housing unit density relative to the density at the beginning of each decade. (Because the land areas of the areas developed as urban are fixed, this is also the percentage change in housing units.) Of course this shows a decline over time as well, but it is a sharper drop that comes close to flattening out in the latter decades. This will be important for the subsequent analysis.

Figure 2. Mean Change in Housing Unit Density for Areas Developed as Urban 1960-2010 by Decade Relative to Year Developed as Urban.
As mentioned in the introduction, a major question to be addressed in this research is the effect of the rate of growth in the urban area as a whole on the length of time it takes for areas added as urban to become largely developed. Several alternative hypotheses about the nature of this relationship are presented here.

The prior paper (Ottensmann 2019) did include some exploration of the relationship of rate of urban area growth to development over time. Quite high, statistically significant correlations were found between the percent change in housing units in the area developed as urban and the percent change in the number of housing units in the urban area as a whole. This was the case for every decade relative to development as urban and for areas developed as urban in each year. But while this shows the expected relationship between the growth of the urban area as a whole and the growth in the areas developed as urban at the various times, it does not provide insight into the relationship between urban area growth and the length of time it takes for an area developed as urban to complete the major part of its development.

Perhaps most obvious in the hypothesis that higher rates of growth in the urban area as a whole would be associated with the faster development of areas added as urban, resulting in lower times to the completion of major development in those areas. The rationale is pretty straightforward: Faster development in the earlier decades will
lead the development of most of the available land in these areas and thus an earlier conclusion to major development.

However, equally reasonable are several hypotheses that would lead to exactly the opposite conclusion: Higher rates of growth in the urban area could be associated with development in those areas continuing over a longer period of time, with longer times to the completion of major development. One basic explanation is simply that faster growth of the urban area in later years would create pressure for more growth in all areas, including those developed as urban in earlier years. This is basically the inverse of the preceding hypothesis.

A more nuanced hypothesis involves consideration of the behavior of landowners and developers looking at expectations associated with returns to development in future years. At each point in time as an area develops, landowners will be balancing the profit they could make by selling for development at that time versus what they might expect to make if they waited and sold at a later time. This, of course, starts long before an area is urbanized. Expecting future urban development, a landowner must compare what they might receive from selling at the current time compared with waiting for the higher prices that will inevitably be offered as the area urbanizes. The comparison will necessarily include considerations of the time value of money, holding costs, and assessment of risk. Furthermore, landowners will vary with respect to their forecasts of future returns and their personal situations.

This assessment of current versus future returns does not end with the initial urbanization and the incorporation of the land into the urban area. It may be reasonable to expect that undeveloped land could command a still higher prices in the future as demand emerges for higher density development that did not exist at the time of initial urbanization. So the landowner must continue to balancing current returns against the possibility of selling the land for a higher price in the future, taking into account once again holding costs, time costs of money, and risk. The extent to which prices might be higher in the future will depend on the future demand. That will depend on the future growth of the urban area. This is, of course, unknown. But the expectation of the rate of future growth may not unreasonably be based on the current rate of growth of the urban area. Faster growing urban areas might be expected to continue to do so in the future with inverse expectations for areas growing more slowly. Think Las Vegas versus Buffalo.

Now if the landowner believes the land can be sold at a price that is sufficiently higher in the future based on expectations of higher future growth, he or she may be more likely not to sell and wait for the higher future return. And by not selling the land for development at an earlier time, the time to completion of major development in the area will be greater for areas with higher rates of growth in the urban area.

This argument that higher rates of growth in the urban area would lead to the withholding of land from development and later development at higher densities was
laid out in my earlier article, “Urban sprawl, land values, and the density of
development” (Ottensmann 1977). Similar arguments have also been made by Ohls and
Pines (1975) and Peiser (1989).

**Time to Completion of Major Development**

The data on development over time after areas had been added to urban areas
leads to the consideration of the time to the completion of major development in those
areas. This is the subject of the current research. How this development time is
conceptualized and measured is the first topic addressed in this section. Then given
these times, the model to use the information to predict the probability of development
in the decades after initial urbanization is described.

**Measuring Time to Completion of Major Development**

This research is attempting to investigate the time it takes for an area to be
developed after it has reached the minimum density threshold and has been added to
the urban area. Two things are of central importance to keep in mind in considering
this. First, given data are available only every 10 years, at the time of each census, so
any measure of time to development will necessarily be very crude. And second, as seen
when looking at actual levels of development over time in the last section, development
doesn’t neatly proceed for some period of time and then abruptly cease. Rather, it tapers
off to a very low level of activity that can continue for a long period of time. Therefore,
it is necessary to consider the time to the completion of major development in those
areas. And this raises the question of determining the point at which major
development has ended.

A first thought might be that major development has been completed when the
amount of undeveloped land drops below some level. The census or other sources of
data collected regularly at the national level do not include this information. The
National Land Cover Dataset (Multi-Resolution Land Characteristics (MRLC)
Consortium 2020; Yang, *et al*. 2018) provides data for recent decades that might be used
to estimate undeveloped land. But it is important to point out that land cover is not land
use. A pixel classified as grasslands could be undeveloped or could be part of a golf
course or someone’s backyard. Methods have been developed to address these issues
but they are far from perfect.

In addition, some land that has not been developed may not be available for
development. Physical characteristics may render the land undevelopable, though the
extent to which areas cannot be developed depends on the demand and the cost one
might be willing to incur. The Back Bay area of Boston was originally under water, as
the name suggests, before being filled in and developed. Legal barriers to development
may exist, such as conservation easements. All of these issues suggests that
determination of the amount of land remaining for development may require local
knowledge and may be difficult to accomplish for a nationwide sample of urban areas.

This leaves the determination of the time at which major development has been
completed to examination of the extent and pattern of development over time. Once
development of an area drops to a very low level, it may be reasonable to conclude that
the period of major development has been completed. Then it becomes a matter of
considering what measure of development over time to consider and what threshold
should be used to identify the end of major development.

The graphs in the previous section illustrating development over time for the
areas developed as urban show the decline in the decades after the area had become
urban. The percent change in housing unit density (which is also the percent change in
housing units) displayed in Figure 3 showed the most dramatic decrease, dropping to a
quite low level in the later decades. For this reason, percent change in housing units
over time was selected as the metric to consider.

Then comes the issue of determining the threshold for percent change in housing
units below which we can conclude that major development has been completed. The
best situation would be if the pattern to the decline showed a change that might suggest
the end of major development—one pattern during the period of major development
followed by a different pattern thereafter. Examination of the data for all of the areas
added to the urban areas in 1960, 1970, and 1980 suggested such a pattern. Percent
change in housing units generally declined fairly steadily to a level of around 10
percent, at which point the percent change seemed like it might be leveling off. Based
on this, a cutoff of 10 percent decline in the percent of housing units added was taken to
establish the end of the period of major development. The appendix provides more
detail on the analysis supporting this choice.

For each of the areas added to the urban areas in 1960, 1970, and 1980, the
percent change in housing units for each succeeding decade was computed. The end of
the decade at which the percent change in housing units permanently dropped below
10 percent was then taken as the time to completion of major development for that area.
The condition “permanently” was included as there were a very small number of cases
where the percent change dropped below 10 percent in one decade and then increased
above that level in a subsequent decade. These earlier drops below the cutoff were not
considered to represent the conclusion of major development.

A few additional comments on these data. Only those areas added in 1960
through 1980 were considered. Areas added in later decades did not have sufficient
numbers of decades of data after they had become urban to establish meaningful
estimates of time to completion of major development. Not all areas added as urban in
1970 and 1980 saw the percent change in housing units drop below 10 percent and
therefore do not have a time to completion of major development. Finally, especially for
the areas developed as urban in the later years, data are not available for as many years after they became urban, so there is a slightly greater probability that the percent change might have dropped below 10 percent but again rose in a decade after which data were available. Since there were so few instances where this actually occurred (8 out of 177 cases) this was not considered to be a serious problem.

*Discrete survival models predicting time to completion of development*

Given the estimates of the time to completion of major development after an area had become urban, the next issue is to identify ways to measure the effect of other characteristics on that time. The primary question here relates to the effect of the rate of change in housing units in the urban area as a whole. For a number of reasons traditional statistical methods are not appropriate for addressing these types of questions. A body of techniques variously called survival analysis or event history analysis has evolved for the analysis of these types of questions.

Standard linear regression analysis cannot be effectively used to simply predict the time to an event such as the completion of major development in these situations. Some of the problems are shared with other situations in which linear regression is not ideal—distributions that are not normal and dependent variables which can have a limited range of values. With the current problem, time to completion of major development obviously cannot be negative. Two additional difficulties are present for these kinds of analysis that have required the development of completely new methods. These are the problems of censorship and time-varying covariates.

In looking at problems involving the time until an event occurs, it frequently happens that some of the cases will not experience the event during the period of observation for which data are available. In the current research, for those areas that were added to the urban area in 1970 and 1980, not all of the areas saw the percentage increase in housing units fall below the 10 percent threshold and therefore were not considered to have reached the completion of major development. Thus no time to completion of major development is available for those areas. The term used to describe such cases is that they have been censored, that is, the limitation on the length of data available makes it impossible to know time to the event. This poses problems for conventional statistical analysis. If the dependent variable is time to the event, then you have cases for which that information is missing. For those cases that are censored, for which the time to the event is not available, some information is available about the time to the event, that is, the time is greater than the last observation. Dropping these cases for the analysis will produced biased estimates. Methods of survival analysis are designed to address the issue and incorporate those cases.

The second problem with using conventional regression methods to predict the time to the event is that it does not provide a means for incorporating time-varying
covariates. (The term covariate is generally used in survival analysis rather than equivalent terms such as predictor or independent or explanatory variable and thus is used here.) The current research question involves such a covariate, percent change in housing units in the urban area as a whole. This information is available for each decade and obviously varies from decade to decade. Each of these percentages might be included as separate covariates in a conventional regression model, but this would not be correct. For the areas that became urban in 1960, for example, 5 decades of percent change in the urban area after that year are available, from 1960 to 2010. But having the 2000-2010 percent change included in a model would be inconsistent with those areas that have experienced completion of major development in an earlier year, perhaps by 1990 or 2000.

Models for survival analysis differ in how they consider the time variable, which is obviously central to the models. One approach treats time to the event as a continuous variable, with time theoretically measurable with fine precision. These methods are the most commonly presented and used, with Cox regression being the prime example. The alternate approach considers time to be measured in discrete intervals, with events assumed to occur at the ends of those intervals. The current problem is obviously one involving discrete intervals, with measurements from the census at decade intervals. Time to the completion of major development is measured as the number of decades after the area had become urban. Therefore, discrete time survival or event history models are most appropriate and are used here.

Research with survival models focuses most on 2 distributions of events with respect to time. For discrete time models these are the hazard probabilities and the probabilities of survival as functions of time. Both are conditional probabilities, with \( t \) being any time interval and \( T \) being the interval in which the event occurs. Then the discrete time hazard probability function \( h(t) \) is

\[
h(t) = Pr(T = t | T \geq t)
\]

This is the probability of the event occurring in the interval \( t \), assuming that the event did not occur before \( t \). For the current research, this is the probability that an area sees the completion of major development in decade \( t \), given that it had not reached that point earlier. This will be the primary focus in the research and the discrete time model: How is the hazard probability affected by other factors such as the percent change in the urban area?

The second major function is the probability of survival, which is also conditional on the time the event occurs and is

\[
S(t) = Pr(T > t | T \geq t)
\]
This is the probability of the event not having occurred through interval \( t \), of the entity “surviving” and not experiencing the event until sometime later. This is the probability for any decade that an area is still seeing major development and has not reached the time of completion. The survival function starts with the value of 1 (all entities surviving, no events having occurred) and necessarily monotonically decreases over time. (A note on the description of this as the “survivor” function and of the general methods as “survival analysis:” Much of the original development in the field occurred in biostatistics where a major question involved how long organisms survived, with death being the event to which time was measured.)

The hazard and survival functions are simply different ways of looking at the same information on the duration of time until an event occurs. One function can be derived from the other. Additional probabilities are sometimes also of interest. An example is the (unconditional) probability of an event occurring at time \( t \), the distribution the events across all intervals. This distribution, generally denoted \( f(t) \), would be the proportions of the areas seeing completion of development across the various decades. Note that these probabilities do not necessarily sum to 1 if some entities have not experienced the event and are therefore censored.

The notation for the hazard and survival functions follows that presented by Mills (2011). Details on the mathematical relationships among the functions are clearly presented by Box-Steffensmeier and Jones (2004) and Tekle and Vermunt (2012). These treatments have the advantage that these authors are focusing exclusively on discrete-time survival models. Many discussions of survival models place primary emphasis on the continuous time models, giving most attention to the functions for those (which are somewhat different) and devote less attention to the discrete-time functions.

Discrete-time survival models require that the data be formatted in a unique manner. This will be described next, as seeing this organization of the data helps to motivate the models that are used for the analysis. Most datasets are formatted with one record for each entity, for each person, area developed as urban in a specific decade, or whatever. The discrete-time models have the data in what is often called a person-period format or what would be, in the current context, an area-decade format. That is, each entity in the dataset would in general have multiple records, one for each decade up through the time interval of the event or, in the case of censoring, the last decade for which data are available. This dataset would have an event variable \( y_{it} \) indicating whether or not the event occurred for entity \( i \) at time \( t \) with the value 1 if the event occurred and 0 if it did not.

For example, for the current dataset, consider first an area that became urban in 1960 and that experienced the end of major development 4 decades later during the interval from 1990 to 2000. That area would have 4 records, for decades 1 through 4. The event variable would have values of 0 for the first 3 decades while major development was continuing and a value of 1 for decade 4 marking the end of major development.
Because major development had ended in decade 4, that area would have no record in the dataset for interval 5. Now take the case of an area that became urban in 1980 and did not reach the end of major development during the following 3 decades up to 2010, the last time for which data were available. It would have 3 records, one for each of those decades, with the event variable having the value of 0 in each case. The presence of the 0 in the final record for the decade 2000-2010 would indicate that the time to the conclusion of major development is unknown and that this event is censored. The dataset in this format has the complete information on the time to completion of major development for those areas that have reached that state. For those that have not, it would have information on how many decades of information were available and therefore the time of censoring.

Time-varying covariates are readily included in a dataset organized in this manner. For example, for that portion added to the Albuquerque urban area in 1960, the percent increase in housing units for the entire Albuquerque area from 1960 to 1970 can be included in the decade 1 record for the area, the percent increase from 1970 to 1980 in the decade 2 record, and so forth. Any covariates that do not vary over time, for example the region in which an area is located, would have the same value entered for each decade.

The discrete-time model predicts the hazard probability as a function of time and a set of one or more covariates. Data for the hazard probabilities are not available to be used as a dependent variable. Rather, we have the information in the dataset described above as to whether or not an area reached the end of major development in each time period. The event variable $y_{it}$ has the value of 1 if the end of major development occurred in interval $t$ and a value of 0 if it had not occurred. This is analogous to a situation that arises in conventional statistical analysis. A binary dependent variable has 2 values that can be denoted as 0 or 1, and the desire is to predict the probability of having the value 1 as a function of a set of covariates. This is exactly the situation here for each time interval $t$. The difference is that here that situation obtains for multiple time intervals.

Probabilities for problems with binary dependent variables can be estimated using several different approaches. Logistic regression is most familiar and is and is most often used for discrete-time survivor models. The method is generalized for discrete time application here. Let $h_{it}$ be the value of the hazard probability for area $i$ at time interval $t$. Then taking the logit of $h_{it}$ produces this model for the hazard $h_{it}$ for case $i$ at time $t$ as a function of time and the covariates:

$$\text{logit} ( h_{it} ) = \log \left( \frac{h_{it}}{1 - h_{it}} \right) = \alpha (t) + \beta' \mathbf{x}_{ti}$$
where \( \alpha(t) \) is some function of time \( t \), \( x_{ti} \) is a vector of 1 or more covariate values for area \( i \) at time \( t \), and \( \beta \) is a vector of regression coefficients to be estimated. The model is estimated using as the dependent variable the event variable \( y_{it} \), indicating whether or not area \( i \) experienced the completion of major development in each time interval \( t \).

The function \( \alpha(t) \) is the baseline hazard, the value of the hazard separate from the covariates. The form of the function and its parameters can be estimated in the course of estimating the model. The baseline hazard function can vary from something as simple as a constant or a constant plus some coefficient times \( t \) to polynomial functions of time, and functions involving transformations of time. The most flexible alternative, suggested by some authors as the starting point for the baseline hazard, is to include dummy variables for each time interval \( t \) (or dummies for all but one interval plus the constant term). While this can reflect any possible distribution for the baseline hazard, it is not necessarily an ideal choice. First of all, this can use a lot of degrees of freedom, especially when there are many time intervals and could result in overfitting the model. For certain datasets, some of the dummy variables may produce perfect predictions for their time interval meaning a coefficient cannot be estimated and they must be dropped from the model. And finally, if there is any interest in interpreting the form of the baseline hazard functions, having large numbers of coefficients for many intervals may make the process more difficult than with a more parsimonious function of time.

In the model presented above, the effect of the covariates is independent of and additive to the baseline hazard function of time to produce the predicted logit of the hazard. This means that for the odds of the hazard, the effect of the covariates is proportional to the hazard. The model can be described as a proportional hazard model. This is also the case for Cox regression, the most widely used method for continuous-time models. In this discrete-time model, the possibility of the effect of the covariates depending on time can be addressed by including interaction terms between the covariates and time.

The model presented above can be estimated like any other logit model using standard logistic regression procedures available in virtually all statistical software. The usual statistical tests employed with logistic regression may be used here, including the Wald tests of the significance of regression coefficients and likelihood ratio tests for comparing the significance of models. As in standard logistic regression, the estimated regression coefficients describe the effect of a unit change in a covariate on the logit of the hazard probability, not a generally intuitive concept. Interpretation of the coefficients requires either conversion so they become measures of the change in the odds of change in the hazard or prediction of specific values of the logit of the hazard and conversion of these to probabilities that can be compared and interpreted in that context. The latter is the approach taken here.

The presentation of the logit form of the discrete hazard model here follows most closely that of Steele (2005) and was also heavily influenced by Allison (1982).
Describing Time to Completion of Major Development

Given the presence of censoring, traditional descriptive statistics are inappropriate for describing the time for the areas to reach completion of major development after the areas have been added to the urban area. For example, it is obvious that the mean of the observed times can be an underestimate of the true mean if the areas that have not achieved completion of major development will have (unknown) times that are greater. Description of the data are accomplished by using what is called a life table, which accounts for the cases and changes in each interval of time. This allows the calculations of the hazard and survival proportions from the data and the estimation of the median time to completion of major development.

Table 1 presents the results for the areas that achieved urban density and were added to their urban areas in 1960, 1970, and 1980. Each row of the table represents a decade after the area became urban, so row 1 has information for the areas that were developed as urban in 1960 for the decade from 1960 to 1970, information for the areas developed as urban in 1970 for 1970 to 1980, and so forth. The next column of the table gives the number of areas at risk during the decade, the number of areas that have not yet reached the point of completion of major development but that had the potential to do so during that decade. For decade 1, this is 177, the total number of areas in the dataset, for each of the 59 urban areas with the areas added in the 3 years of 1960, 1970, and 1980.

None of the areas added to the urban areas reached completion of major development during the first decade. Therefore the full set of 177 areas was again at risk for completion of major development in decade 2. That decade did see 1 area reach the threshold of completion of major development, reducing the number of areas at risk in

Table 1. Life Table for Time to Completion of Major Development.

<table>
<thead>
<tr>
<th>Decade after Year Developed as Urban</th>
<th>Areas at Risk at Start</th>
<th>Areas Completing Major Development</th>
<th>Areas Lost (Censored)</th>
<th>Hazard (Proportion Completing Development)</th>
<th>Survival (Cumulative Proportion not Completing Development)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>177</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>177</td>
<td>1</td>
<td>0</td>
<td>0.006</td>
<td>0.994</td>
</tr>
<tr>
<td>3</td>
<td>176</td>
<td>59</td>
<td>48</td>
<td>0.335</td>
<td>0.661</td>
</tr>
<tr>
<td>4</td>
<td>69</td>
<td>45</td>
<td>17</td>
<td>0.652</td>
<td>0.230</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
decade 3 to 176. Note that no areas had been censored during the first 2 decades because a minimum of 3 decades of data are available for all of the areas. The third decade is the first time significant numbers of areas reach completion of major development 59 out of the 176 areas. This is also the first decade that saw censoring of the data. For decade 3, 48 of the areas that were developed as urban in 1980 had not reached completion of major development, and the third decade for those areas was 2000-2010, the last decade for which data are available. These 48 areas are censored at this point, lost from further consideration for possible completion of major development.

The number at risk for decade 4 is now the number from decade 3, 176 areas, less the number of areas that reached completion of major development in decade 3, 59 areas, and also less the number of areas that were lost and censored because it was their final year of data and they had not yet seen completion of development. So only 69 areas remained at risk for the fourth decade. Of these areas, 45 reached completion of major development and 17 were censored. Those censored were areas that had been added to their urban area in 1970 for which the decade 4 was the final year for which data were available. Finally, only 7 areas remained at risk, not having reached completion of major development by the beginning of the fifth decade. These were all areas that had been added to their urban area in 1960, for these were the only areas for which the full five decades of information were available. All 7 of these are reached completion of major development in the fifth decade, so none from the set of areas added in 1960 were censored.

Laying out the data in this manner, looking at the number of areas reaching completion of major development in each decade and the number censored in each decade allows the determination of the numbers at risk in each decade which forms the basis for estimating the hazard and survivor proportions. The hazard probability is the probability of an area reaching completion of major development in an interval given that it had not done so in an earlier interval. In the life table, the hazard proportion for each decade is then simply the number of areas completing major development divided by the number at risk for that decade. The hazard proportion is very low, 0 or close to 0 for the first 2 decades. Nearly all of the areas required at least 3 decades to reach completion of major development. In the following 2 decades, the hazard proportion continued to climb, reaching 1 for the fifth decade. If an area had not reached final development before, it did then. The simple line graph in Figure 4 shows this pattern for the hazard proportion over time. Thinking in terms of probabilities, an area had a very low probability of reaching completion of major development during the first 2 decades with the probability increasing in each of the following decades. That the pattern of the hazard proportion shows this pattern seems quite reasonable. As shown in the earlier plots of densities and changes over time, significant development continues for at least several decades after an area becomes urban, so the low initial
proportions were not unexpected. But then the percentage change in density starts dropping and areas reach completion of major development. It is also not surprising that major development does not continue forever. At some point development will decline and the completion of major development will be reached.

This takes us to the final column of the life table, the survival proportion. The survival probability is the probability at each interval that the area has not reached completion of major development given that it had also not done so earlier. So the survival proportion is the cumulative proportion not completing development in a given interval. The survival proportion is 1 minus the hazard proportion (the proportion not completing major development) times the survival proportion for the previous interval. In terms of probabilities, it is the probability of surviving to the start of the interval times the probability of surviving during the interval. For the first decade, the probability at the start is 1—nothing has happened, no areas could have completed development, so all survive. Since the hazard proportion for that decade is 0, 1 minus that is 1 and the survival proportion remains 1. In the second decade, there is a small hazard proportion so the survival proportion is reduced by a small amount to just under 1. Then in the third interval, the larger hazard proportion produces a greater reduction in the survival proportion. And by the final decade, since all of the areas at risk have reached completion of major development there are no “survivors” continuing major development and the survival proportion is 0.
A plot of the survival proportion over time is presented in Figure 5. By convention, survival functions are plotted as stair step graphs to emphasize the drops in the numbers surviving at each point in time. This clearly shows the decline in the numbers continuing major development during the third, fourth, and fifth decades.

Given censoring, it is not possible to calculate a mean time to the completion of major development for this dataset. However, the survival curve provides a way of estimating the median time to completion. The median is, of course, the value for the case in the middle of a distribution. So the median time is the time at which half of the cases survive, where the survival value is 0.5. The life table gives only a few survival values, including 0.66 for decade 3 and 0.23 for decade 4. So the median time to completion of major development is between 3 and 4 decades. Linear interpolation between those 2 points can be used to estimate a more precise value for the median time. Doing so gives a median time to completion of major development of 3.37 decades or 33.7 years.

**Baseline Hazard for Time to Completion of Major Development**

The estimation of the model begins with the exploration of the baseline hazard time to completion of major development. This is a model with only the function of
time, without any covariates included. The objective is to understand the options for the form of this function.

Various models with different functions of time were estimated using standard logistic regression routines, in this case using Stata. The results for the most important models illustrating the exploration are presented in Table 2. The first model included simply the time variable, the number of decades since the area became urban, with values ranging from 1 to 5. The first column of results gives the estimates for this model, including the regression coefficients for time and the constant term. The coefficient for time was positive and highly significant. This was not surprising given the increase in the hazard over time illustrated in Figure 4.

The next step was to explore the performance of a polynomial function of time. A time-squared variable was added along with time, producing the results shown in the next column of Table 2. The coefficients for both time and time squared were statistically significant. The coefficient for time remained positive. The coefficient for time squared was negative, indicating a function concave downwards. The log likelihood for this model of -167 was substantially larger than value of -172 for the model using only time. A likelihood ratio test comparing this model with the time-only model produced a $p$-

### Table 2. Alternative Models Predicting Baseline Hazard for Time to Completion of Major Development (Standard errors in parentheses).

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Base Model with Time</th>
<th>Polynomial Model Including Time Squared</th>
<th>Model with Inverse of Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (Decades since Area Became Urban)</td>
<td>2.209 ***&lt;br&gt;(0.212)</td>
<td>7.379 ***&lt;br&gt;(1.808)</td>
<td>—</td>
</tr>
<tr>
<td>Time Squared</td>
<td>—</td>
<td>-0.799 **&lt;br&gt;(0.261)</td>
<td>—</td>
</tr>
<tr>
<td>Inverse of Time</td>
<td>—</td>
<td>—</td>
<td>-20.734 ***&lt;br&gt;(2.466)</td>
</tr>
<tr>
<td>Constant</td>
<td>-7.681 ***&lt;br&gt;(0.672)</td>
<td>-15.757 ***&lt;br&gt;(3.086)</td>
<td>6.100 ***&lt;br&gt;(0.794)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-172.32</td>
<td>-166.81</td>
<td>-165.435</td>
</tr>
<tr>
<td>Model Chi-squared</td>
<td>235.44 ***</td>
<td>246.47 ***&lt;br&gt;(2.22)</td>
<td>249.22 ***</td>
</tr>
</tbody>
</table>

** Differences significant at 0.01 level
*** Differences significant at 0.001 level
value of less than 0.001, allowing a strong conclusion that the polynomial model performed significantly better.

When a third-order polynomial model was attempted, adding time cubed, the logistic regression procedure failed to converge and could not produce estimates for the model. This indicates a limitation with this dataset, which includes just 5 intervals for the time with 2 of those intervals having no variation in the development variable. (No areas reached the end of major development in the first interval and all of the remaining areas concluded major development in the fifth interval.

The second order polynomial model showed that a curvilinear function of time performed better in predicting the logit of the hazard probability than a linear function. This led to the exploration of alternative transformations of the time variable. Three transforms of time were considered: the natural logarithm of time, the square root of time, and the inverse of time. The log of time performed poorly, with a log likelihood lower than the initial model with time in its natural form. The model using the square root of time did somewhat better than the model with time in natural form, with higher log likelihood of -170, still somewhat lower than the -167 for the polynomial model. However this model has one less coefficient to be estimated. Because the square root and polynomial models are not nested, their performance cannot be compared using a likelihood ratio test. The Akaike Information Criterion (AIC) is one metric for making a comparison in such situations. Examining those values, the polynomial model still performed better than the square root model.

The final transformation of time employed was the inverse of time. The results for that model is presented in the final column of Table 2. The coefficient for the inverse was significant. The log likelihood of -165 was the highest of any of the models considered, slightly better than the polynomial model. And of course the AIC for this model was better (lower) than for the polynomial model since the AIC discounts the log likelihoods using the number of degrees of freedom used by the model. One final comparison was undertaken using a model including both time and the inverse of time. The regression coefficient of time was not significant in this model while the coefficient of the inverse was highly significant. The likelihood ratio test comparing this model with the first model using only time showed this model doing significantly better, no surprise. A likelihood ratio test could also compare this model with the final model in Table 1 using only the inverse of time. The model including time along with the inverse was not significantly better than the model with the inverse alone. It is reasonable to conclude that of the models examined, the model using the inverse of time for the baseline hazard performed best. The performance of the polynomial model was also better than the other models considered.

One further possibility was considered. The presentation of the logistic discrete-time model mentioned that the use of time dummies for each interval for the baseline hazard could be considered the most flexible alternative. As usual with dummy
variables representing a variable with multiple values, either 1 has to be excluded to avoid collinearity or the model must be estimated without a constant. Both methods are equivalent. When estimating the model using the 5 time dummy variables with no constant, Stata reported that the time dummies for intervals 1 and 5 predicted the values perfectly (obviously, because all the values for time 1 were 0, no areas reaching the end of major development, and all the values for time 5 were 1, all areas concluding major development). Because of this, the observations for times 1 and 5 were dropped and the model was estimated using only the observations for times 2, 3, and 4, with coefficients reported for those time dummy variables. The number of observations had been reduced from the original count of 606 to 540.\(^2\) One does not want to lose the information from the first and last intervals from the analysis, so the model using time dummy variables does not work here.

The Effect of Change in the Urban Area on Completion of Major Development

Discrete-choice logit models showing the effect of the change in housing units in the entire urban area on the length of time it takes new urban areas to complete major development are presented. The first part shows the results for the major models estimated in the process of arriving at the final model. Predictions developed using that model illustrate the nature of that relationship between urban area change and time to completion of development and are shown graphically. In the final section the robustness of the results is examined by comparing estimates for subsets of the data.

Estimating Models of the Effect of Urban Area Growth

The covariate percent change in housing units for the entire urban area in each decade are added to the baseline hazard model. While the investigation of the functional form for the baseline hazard suggests that the inverse of time would work best, different alternatives were again considered to see if the presence of the covariate produced any changes to that.

Table 3 shows the results for 4 of the models considered. The first starts again using time, the number of decades for the hazard function and the percent change in housing units in the entire urban area for each decade. For the areas that were added as urban in 1960, the first interval included the percent change from 1960 to 1970, for the areas added as urban in 1970, this would be the percent change from 1970 to 1980, and

\(^2\) When estimating the model with all of the time dummies with the constant term, Stata dropped an additional dummy (for interval 4) because of collinearity, standard procedure in such instances. But then the observations for interval 4 were also dropped in addition to those for intervals 1 and 5. This reduced the number of observations for the final model that was estimated from the original count of 606 to only 153.
Table 3. Models of the Effect of the Percent Change in the Number of Housing Units in the Urban Area on Time to Completion of Major Development (Standard errors in parentheses).

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Base Model with Time</th>
<th>Polynomial Model Including Time Squared</th>
<th>Model with Inverse of Time</th>
<th>Model with Inverse of Time and Inverse of Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (Decades since Area Became Urban)</td>
<td>2.104 *** (0.222)</td>
<td>6.923 *** (1.864)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Time Squared</td>
<td>—</td>
<td>-0.739 ** (0.269)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Inverse of Time</td>
<td>—</td>
<td>—</td>
<td>-20.173 *** (2.538)</td>
<td>-21.392 *** (2.666)</td>
</tr>
<tr>
<td>Percent Change in Urban Area Housing Units</td>
<td>-0.0280 ** (0.0098)</td>
<td>-0.0246 * (0.0099)</td>
<td>-0.0242 * (0.0098)</td>
<td>—</td>
</tr>
<tr>
<td>Inverse Percent Change in Urban Area Units</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>8.071 *** (1.873)</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.812 *** (0.737)</td>
<td>-14.474 *** (3.199)</td>
<td>6.373 *** (0.829)</td>
<td>5.565 *** (0.840)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-167.67</td>
<td>-163.39</td>
<td>-162.105</td>
<td>-152.948</td>
</tr>
<tr>
<td>Model Chi-squared</td>
<td>244.74 ***</td>
<td>253.32 ***</td>
<td>255.88 ***</td>
<td>274.19 ***</td>
</tr>
</tbody>
</table>

* Differences significant at 0.05 level
** Differences significant at 0.01 level
*** Differences significant at 0.001 level

so forth. The coefficient for time was positive, nearly the same as in the baseline hazard model, and was statistically significant at the 0.001 level. The coefficient for percent change in the urban area was negative. The hazard for completion of development declines for more rapidly growing urban areas, meaning that the time to completion of major development will be greater. The coefficient was statistically significant at the 0.01 level.

The second model shown in the table uses the polynomial function of time, adding time squared to the model. The coefficients for both time and time squared were...
statistically significant and are again similar to those for the comparable baseline hazard model. The coefficient for percent change was only slightly smaller in magnitude than in the previous model and was significant, this time at the 0.05 level. The likelihood ratio test showed this model to be significantly better than the model with time alone.

The transforms of time were considered next. Once again, the inverse of time performed the best and that is the third model shown in Table 3. The coefficient for the inverse of time was almost the same as for the baseline model and was statistically significant. The coefficient for percent change was almost identical to the previous model using the polynomial function of time and had the same level of significance. The log likelihood is slightly higher for this model using the inverse of time than for the polynomial model and the AICs show an even greater advantage for this model. As with the baseline models, likelihood ratio tests using a model with both time and the inverse of time showed that to be significantly better than the model with time alone but not better than the model with only the inverse of time.

The distribution of the percent change in the urban area variable is significantly right-skewed, with smaller numbers of much higher values. Thus it would not be surprising if a transformation of that variable performed better than the variable in natural form. Three transforms were considered, the natural logarithm of percent change, the square root of percent change, and the inverse of percent change. All three measures performed better than the model with percent change in natural form, with the inverse of percent change doing the best. The results using the inverse of percent change are presented in the final column of the table. The coefficient for the inverse of time was again highly significant and nearly the same as in the previous model. The coefficient for the inverse of percent change in the urban area was now of course positive, reversing direction using the inverse, and was also highly significant at the 0.001 level. To do likelihood ratio test comparisons with the previous model, a model including both percent change in natural form and the inverse of percent change did significantly better than the model with only the natural form. And it was not significantly better than the model with the inverse of percent change alone. This was considered to be the tentative final model for the effect of percent change in the urban area on the logit of the hazard and ultimately the relationship to the time to the conclusion of major development.

A final check using the other functions of time with the inverse of percent change showed the model with the inverse of time and the inverse of percent change performed the best. There remained the question as to whether there might be any interaction between time and the measure of percent change. Interaction terms using different functional forms of time and percent change were included in the model and none were even close to being statistically significant. So the proportional hazard formulation is appropriate.
On a purely exploratory basis, several other covariates were included in the model to see if they had any predictive power. First was the size of the entire urban area in the year the area was added as urban, the total number of housing units. No particular hypothesis motivated this, only the fact that the size of an urban area is so frequently related to other aspects of urban patterns. Neither the number of housing units or its log was significant in the model. Likewise, the region of the country in which the urban area was located was also not significant in the models, despite its relationship to urban patterns investigated earlier.

The presence of barriers to the expansion of an urban area have been shown to be highly significantly related to various aspects of urban development in earlier research using these data. In particular, such barriers were found to be significant predictors of the change in density over the 3 decades following the year these areas were developed as urban (Ottensmann 2019). Thus it was natural to consider whether they might be related to the time to completion of major development. The barriers considered were whether urban expansion was physically constrained by the presence of wetlands, mountains, and lands protected from development as a result of public ownership, or whether expansion might be limited by constraints on the availability of water in arid regions. Presence of the first 3 barriers were determined by judgment, with arid regions being defined as having average annual rainfall less than 15 inches. Each was considered in turn by adding the dummy variable indicating the barrier to the model. Only the presence of mountains as a barrier to expansion was statistically significant, and that only barely at the 0.05 level. Given that in prior investigations with these measures, all were highly and consistently significant, and given that mountains, protected lands, and arid areas are highly correlated, one cannot conclude that barriers to urban expansion are related to the time to completion of major development.

The last model shown in Table 3, using the inverse of time and the inverse of percent change in housing units in the urban area, is considered to be the final model.

Illustrating the Prediction of Time to Completion of Major Development

The relationship of the inverse percent change in the number of housing units in the urban area to the logit of the hazard function is one of the least intuitive things to consider. This section illustrates the relationship of percent change in the urban area to development over time in more meaningful ways, displaying plots of functions associated with the predicted probabilities associated with the predicted logit of the hazards using the final model presented in Table 3.

The estimated model can be used to make predictions of the logit of the hazard for any value of time, using arbitrarily small increments to produce smooth curves when values are then graphed against time. The model was estimated using data limited to decade long increments of time. Some of the functions to be derived from the
predictions, such as the survival function, involve calculation of cumulative values over time. Performing such calculations for small increments of time produces biased values. Using 5-year increments for the logit hazard predictions seemed to be a reasonable compromise that provides somewhat more temporal resolution than making the predictions for decade-long intervals while minimizing this bias.

Predicting a value for the logit hazard requires specification not only of the time but values for the percent change in housing units in the entire urban area. For the initial examples comparing the predicted results with the functions from the life table data, the mean percent change for all urban areas for all decades from 1960 to 2010 was used. This value of 32.4 percent was used for the prediction for each time interval. Then, to examine the effect of percent change on the functions, values of percent change 1 standard deviation below and above this mean value were used, 8.4 percent and 56.4 percent.

The presentation begins with the function of the hazard probability over time. The final model using the inverse of time and the inverse of percent change was used to predict the values of the logit of the hazard using the mean percent change for 5-year intervals with time values from 0.5 to 5.0 (time in the model is in decades since the area became urban). These logits were converted to probabilities, which are plotted versus time as the red curve in Figure 6. This is compared with the actual hazard proportions.

Figure 6. Actual Hazard Proportion and Predicted Hazard Probability for Mean Percent Change in Urban Area Housing Units Calculated in 5-Year Increments.
calculated in the life table (Table 1), shown in light blue. Both hazard functions have very low values for the first 2 decades, beginning their climbs around year 20 to the highest values at year 50.

The predicted survival probabilities can then be calculated from the hazard probabilities. The predicted survival curve is shown in red in Figure 7 along with the actual survival proportions in light blue. The general patterns are similar. At some of the points the predicted survival probability does decline before the actual survival proportions. However this is at least in part due to the use of the 5-year intervals in making the predictions. The actual survival proportions can only decline at the ends of the decades.

As described for the survival proportions, the median time to completion of major development can be estimated from the survivor function as the time at which half of the areas remain in the process of major development, completing development later. Since the values for the predicted hazard probabilities are widely spaced, linear interpolation between the survival probabilities above and below 0.5 has been used to make the estimate. The median time to completion of major development estimated from the predicted survival probabilities is 33.8 years. This is virtually identical to the median value of 33.7 years estimated from the survival proportions calculated in the life table.

![Figure 7. Actual Survival Proportions and Predicted Probability of Survival for Mean Percent Change in Urban Area Housing Units Calculated in 5-Year Increments.](image-url)
A series of graphs are now presented to illustrate the effect of the percent change of housing units in the urban area on various functions associated with the time to the completion of major development. Each of these graphs will display 2 functions, predicted probabilities using percent change for the urban area 1 standard deviation below and above the mean for all areas and decades, 8.4 and 56.4 percent. The function for the high percent change in the urban area will be displayed in red, and the function for the low percent change will be light blue.

Figure 8 presents the plots of the predicted hazard probabilities over time for the low and high rates of change in housing units in the urban areas. Both are increasing over time in much the same fashion. But the increase in the hazard over time is slower for the higher percent change. At each point, an area experiencing faster growth in the urban area is predicted to be less likely to reach completion of major development. The difference is greatest, with hazard probabilities of 0.4 and 0.2 at the 35-year point.

The next way of looking at the differences associated with the low and high percent change in the urban area is with the survivor curves. Figure 9 displays the predicted survival probabilities, again with the red showing survival for the high percent change and light blue for low percent change. The survival curve for high percent change is consistently higher than the curve for low percent change. More rapidly growing areas “survive” longer, that is, are likely to take longer to reach the end of major development.

![Figure 8. Predicted Hazard Probability for Areas with Percent Growth in the Urban Area 1 Standard Deviation Below and Above the Mean.](image)
Figure 9. Predicted Probability of Survival for Areas with Percent Growth in the Urban Area 1 Standard Deviation Below and Above the Mean.

Again, estimates of the median survival time can be made from these survival probabilities. For predictions with the percent change in the urban area at 1 standard deviation below the mean, 8.4 percent, the predicted median time to completion of major development is 30.9 years. With percent change at 56.4 percent, 1 standard deviation above the mean, predicted median time to completion of major development is 34.2 years. The faster growing areas are expected to take over 3 years longer to reach the end of major development.

An additional way of looking at this information is to consider the unconditional probability of reaching the end of major development during each 5-year period. This shows how the distribution of times to completion are expected to be distributed across the periods. Note that because there is the probability that some areas may not reach the end of major development at the conclusion of the 50-year period, these probabilities will add up to less than 1. Figure 10 shows that the predicted probability of an area reaching the end of major development is higher in the earlier years with the low percent change in the urban area. This holds up through 35 years. Then in the later years, the probability of an area concluding major development is greater with the high percent change. The probability is less that areas with the low percent change “survive” and have not reached the end of major development in these later years. The difference between these distributions is greatest in the 5-year interval ending in year 30. The
probability of a low-growth area seeing completion of development in this interval is 0.31 while the probability of a high-growth area ending development is only 0.18.

**Considering the Robustness of the Results**

The previous section showed that the estimated hazard and survivor functions were very close to those developed from the data in the life table. Another approach to assess the robustness of these results is to compare models estimated for the areas developed as urban in the different years such as 1960 and 1970. We would not expect the results to be the same for such models. Too many factors varying over time can affect patterns of development, factors not included in the model. But if the estimates presented above are reasonable and provide meaningful insight into patterns of development over time and the time to completion of major development, some degree of consistency should be expected.

To do this assessment, 3 additional models have been estimated using subsets of the data. Each model used the same variables in the same form as the final model. The first 2 of these models are for the 59 areas that have been added to their urban areas after achieving urban density in 1960 and the 59 areas added in 1970. No model could
be estimated for the areas added in 1980. These had data for only 3 decades after that year. None of the areas reached the end of major development during the first 2 following decades. As a result, there was insufficient variation in the data to estimate a comparable model. To indirectly explore the effect of the 1980 data, a third model was estimated using the data for the areas added in the 2 earlier years, 1960 and 1970 together. The assumption was that if the data for the areas added in 1980 was severely inconsistent with the earlier years, its omission would produce a major change in the estimates from those with that data included.

Table 4 presents the results of the estimation of these models. First is the original model estimated using the data for all 3 years, areas added as urban in 1960, 1970, and 1980. Next are the models for the areas added as urban in 1960 and then for the areas added in 1970. The final model is for the areas added in 1960 and 1970 as the indirect check on the areas added in 1980. Some broad observations: All of the models were highly significant with the Chi-squared test showing significant at the 0.001 level in each case. The regression coefficients for both the inverse of time and the inverse of percent change in the urban area were likewise statistically significant in all of the

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Years Developed as Urban Included in Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2.666)</td>
</tr>
<tr>
<td>Inverse Percent Change in Urban Area Units</td>
<td>8.071 ***</td>
</tr>
<tr>
<td></td>
<td>(1.873)</td>
</tr>
<tr>
<td>Constant</td>
<td>5.565 ***</td>
</tr>
<tr>
<td></td>
<td>(0.840)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-152.95</td>
</tr>
<tr>
<td>Model Chi-squared</td>
<td>274.19 ***</td>
</tr>
</tbody>
</table>

* Differences significant at 0.05 level
*** Differences significant at 0.001 level
models, though the levels of significance did vary. The signs of the coefficients were consistent, and the coefficients were of the same order of magnitude.

Some differences can be seen, of course. The coefficient for the inverse of time for the areas added in 1970 was around -9, while the coefficients for all of the other models were reasonably close, varying only from -18 to -23. The coefficient for the inverse of percent change was definitely the highest at about 35 for the areas added in 1970. The 1960 and 1960 plus 1970 models had similar coefficients, 13 and 14. Interestingly, the effect of percent change was least for the model with all of the data, with a value of 8.

As already mentioned, other factors not included in the model would be expected to vary over time and affect patterns of development over time in different ways for areas added as urban in different years. One obvious factor expected to influence development would be general economic conditions. While these could certainly influence the percent growth of the urban areas, other different and potentially very significant effects on development might also be expected.

Differing economic conditions in the early decades considered here certainly suggest a possibly significant effect. After reasonably strong economic growth during much of the 1960s, the 1970s have been characterized as a period of stagflation—sluggish economic growth combined with high levels of inflation. After a brief but sharp recession at the start of the 1980s economic growth came back. The 1980s also saw extremely high interest rates in the first part of the decade as the Federal Reserve worked to bring down the inflation and brought on that recession. The prime rate peaked above 20 percent. The mortgage interest rate rose to the upper teens and continued at the double-digit level through most of the decade.

Now consider how this relates to the areas being considered. Those areas that were added to their urban areas in 1960 started with a good economic conditions, not facing the stagflation of the 1970s and the subsequent high interest rates until later. For those areas added to their urban areas in 1970, those economic conditions were present from the start, affecting the initial decade, generally the time of the greatest development after an area becomes urban. And of course the areas added in 1980 followed that period of stagflation, though they did face the early recession and the high interest rates. Given the difficult economic conditions faced immediately after becoming urban for those areas added in 1970, it is not surprising that patterns of development would have been different and that the estimated coefficient would vary most for the model estimated for those areas.

Conclusion

The first observation about this analysis is that it actually worked. When starting this work, expectations of success were tempered. Survival analysis was being attempted using observations made at 10-year intervals. Data were available for a
maximum of 5 intervals following the time an area was added to an urban area and even fewer for the areas added after 1960. The measure of the time to completion of major development was necessarily rather arbitrary. Despite these hurdles, a survival model was estimated that produced predictions consistent with the original data. Models estimated using subsets of the data for areas added in different years showed reasonable consistency.

Both the model and the original life table produced estimates of the median time to completion of major development of about 34 years. This is a striking value for the typical length of time significant development continues in areas after they have become “urbanized” and added to urban areas. This time is very consistent with the account of development over time described in the earlier paper (Ottensmann 2019).

The key question addressed was the relationship of the overall rate of growth in the urban area to the time to completion of major development. The estimated model found a statistically significant negative relationship between time to completion of major development and the percent change in housing units in each decade in the entire urban area. Analysis using the survival model represented the appropriate way of examining this question while accounting for the differences in the percent change in the urban area in each decade (that this was a time-varying covariate).

The research began with alternate hypotheses concerning the relationship between the rate of growth in the urban area and the time it takes for the development in an area to be completed after that area has been added to the urban area. One hypothesis was that the relationship would be positive, that faster urban area growth would produce faster growth in the areas added to the urban area, leading to completion of major development sooner. The results here suggest that this hypothesis can be rejected.

Several possible explanations were provided that might explain the opposite hypothesis that faster growth in the urban area would result in development continuing for longer periods of time in the areas added to the urban area. One suggestion was simply that more growth in the urban area meant more growth could be occurring in all portions of the area. Thus as the percentage increase in development dropped over time, that drop would occur more rapidly in the areas that were growing more slowly. And thus they would fall below the threshold that marked the end of major development sooner than the faster growing urban areas. But this leaves one question hanging: If the more slowly growing areas generally have lower levels of development and major development ceases earlier, does this mean that the areas added as urban are left less completely developed than those in more rapidly growing areas? This might be a possibility, but why would this happen? Why would development stop in such areas with opportunities for further development remaining?

An alternative explanation involves the role of landowner expectations involving returns to development at various times. Landowners in more rapidly growing urban
areas may reasonably expect these higher levels of growth to continue into the future. This may lead to expectations of greater returns in the future from development at higher densities that would exceed the holding costs including the time value of money. Thus more land would be withheld from development in the rapidly growing urban areas, leading to a longer time to the completion of major development in those added as urban.

The current analysis and data do not allow one to distinguish between these two explanations for the relationship between faster urban area growth and longer times to the completion of major development. More detail on what is happening in terms of development within the areas added as urban would be required to do so. And as discussed when describing the problem of estimating time to completion of major development, data limitations on the amount of land that has been developed and not developed and the problems with the use of such data limit the possibilities for using that.

While the data from available from the census are limited as has been noted, the data being used here have been aggregated from census tract data. Undertaking analysis using data at the tract level provides an opportunity to consider variation between tracts in development over time. Variables that have not been considered here, including individual tract characteristics, might be included such analyses.

Analysis using the census tract data could involve both conventional statistical analysis such as that reported in the earlier paper on development over time (Ottensmann 2019) as well as survival analysis comparable to that presented in the current paper. Tract-level survival analysis would of course still be limited by the small number of intervals for which data are available. But the larger numbers and greater variation might provide opportunities for considering the effects of other factors, if not simultaneously, at least individually.

I would not expect that analyses conducted using the census tract data would allow conclusive evidence to decide which of the explanations for longer time to development in faster growing urban areas is correct (if either). But I would hope that it could provide further insight on development over time that might at least be suggestive.

References


Appendix: Determination of the Cutoff for Completion of Major Development

The percentage change in the number of housing units in areas developed as urban tends to decline over time after the area has been added to the urban area. Once this percentage drops to a sufficiently low level, it may be reasonable to conclude that the period of major development of the area is over. This can then be used to establish for each area the number of decades it has taken for major development to have been completed. This appendix describes the methods used to determine this cutoff for percentage change of housing units for completion of major development.

Figure 3 above presents the mean percent change in density and housing units over time since an area became urban. The value declines at a decreasing rate over time, dropping to a fairly low level for decades 4 and 5 after development. The ideal situation for identifying the level that marks the completion of major development would be for the pattern of decline to change at that point. Percent change clearly declines at a decreasing rate in the early years and could be seen as the change that characterizes the period of major development. Then if there is a point at which the decline levels off and percent change continues at a lower but more steady level, this could mark the end of the period of major development and the start of the post-development period of modest increases in density.

Now if we had frequent observations of numbers of housing units and percent change for the urban areas, we could examine the pattern for change for any area and identify the time at which percent change leveled off. This would be the time to completion of major development for that area. If doing this for every area developed as urban was deemed too time-consuming, this could be done for a sample of areas. For each point at which percent change leveled off, the value of percent change could be recorded. One would hope these values would be similar for the various areas. The mean of those values of percent change for the conclusion of major development could then be used to identify the time to completion of development for the areas not in the sample.

Further insight is provided by considering how this might be done if the number of housing units were actually a continuous function of time. Then the rate of change in housing units would be the derivative of this function with respect to time. The goal is to determine when the change in that rate levels off. And the change in the rate of change with respect to time is the second derivative of housing units as a function of time. The time at which the second derivative leveled off would be the time to completion of major development (and would also be associated with the rate of development—the first derivative—at the completion of major development).

Of course we do not have the continuous function of housing units with respect to time. In fact, we do not have observations of housing units for many times at all. For the areas developed as urban in 1960, we have only 6 observations for housing units,
each decade to 2010, and five decades for the percent change in housing units. This does not provide the temporal detail to identify the time at which the decline in percent change levels off that would mark the completion of major development.

We do, however, have many observations for the percent change in housing units. the 59 areas developed as urban in 1960 each have percent change measures for the 5 decades, change from 1960 to 1970, 1970 to 1980, and so forth. In addition, we can calculate the change in percent change from one decade to the next, the difference between 1960-1970 change and 1970-1980 change, and so forth. There are 4 of these changes for each area, giving a total of 236 observations.

The argument presented above shows that the change in percent change (or the second derivative for continuous time) can be used to identify not only the time to the end of major development but also the percent change associated with that time (the first derivative for continuous time). We lack the fine-grained variation in time to do this with the current data. However, we do have many varying observations of percent change and change in percent change. By looking at the relationship of percent change to change in percent change, it becomes possible to identify the point at which change in percent change levels off and therefore the value of percent change to be associated with the completion of major development.

The analysis that follows uses the values of percent change in housing units and the change in percent change for the areas developed as urban in 1960, as these areas have data extending over the greatest period of time after urbanization. We look at the relationship between change in percent change and the percent change in the second decade. We do this because we want percent change closest to the point at which we observe the change in percent change leveling off.

The graph in Figure 11 below includes the scatterplot of change in percent change in housing units versus percent change in housing units. The range is restricted to those cases in which the percent change in less than 40 to allow focus on the area of interest which involves the low levels of percent change that would be associated with the end of major development. It is clear that change in percent change tends to decline as percent change decreases. And at low levels there is a blob of points near and below 0 change in percent change. But it is difficult from just the scatterplot to identify a level of percent change where the change in percent change could be seen as leveling off.

To assist in identifying this point, the values of change in percent change have been smoothed in the following manner: The cases were divided into 2-percent intervals for percent change, 0 to 2 percent, 2 to 4 percent, and so forth. The mean change in percent change is then calculated for the cases in each of these intervals. The values of those means are presented in Table 5, again limited to those intervals below 40 percent. While the pattern is far from perfect, there is a significant and generally consistent decline in the change in percent change as percent change decreases, at least until
Table 5. Mean Change in Percent Change in Housing Units for 2 Percent Intervals of Percent Change in Housing Units.

<table>
<thead>
<tr>
<th>Percent Change in Housing Units</th>
<th>Mean Change in Percent Change in Housing Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>-8 to -6</td>
<td>11.5</td>
</tr>
<tr>
<td>-6 to -4</td>
<td>—</td>
</tr>
<tr>
<td>-4 to -2</td>
<td>—</td>
</tr>
<tr>
<td>-2 to 0</td>
<td>7.9</td>
</tr>
<tr>
<td>0 to 2</td>
<td>10.7</td>
</tr>
<tr>
<td>2 to 4</td>
<td>6.9</td>
</tr>
<tr>
<td>4 to 6</td>
<td>4.5</td>
</tr>
<tr>
<td>6 to 8</td>
<td>10.9</td>
</tr>
<tr>
<td>8 to 10</td>
<td>6.4</td>
</tr>
<tr>
<td>10 to 12</td>
<td>18.4</td>
</tr>
<tr>
<td>12 to 14</td>
<td>13.1</td>
</tr>
<tr>
<td>14 to 16</td>
<td>12.0</td>
</tr>
<tr>
<td>16 to 18</td>
<td>19.3</td>
</tr>
<tr>
<td>18 to 20</td>
<td>32.4</td>
</tr>
<tr>
<td>20 to 22</td>
<td>45.7</td>
</tr>
<tr>
<td>22 to 24</td>
<td>51.5</td>
</tr>
<tr>
<td>24 to 26</td>
<td>38.9</td>
</tr>
<tr>
<td>26 to 28</td>
<td>42.6</td>
</tr>
<tr>
<td>28 to 30</td>
<td>31.6</td>
</tr>
<tr>
<td>30 to 32</td>
<td>61.9</td>
</tr>
<tr>
<td>32 to 34</td>
<td>31.3</td>
</tr>
<tr>
<td>34 to 36</td>
<td>58.8</td>
</tr>
<tr>
<td>38 to 40</td>
<td>48.0</td>
</tr>
</tbody>
</table>

around 10 percent. For the percent change interval from 8 to 10 percent, the change in percent change drops below 10 percent for the first time, to 6.4 percent. While the change in percent change goes up and down as percent change declines further, it never rises above 11.5 percent. And this is below the lowest value for change for any interval above 10 percent.

The pattern seen in the table is clearly shown on the graph in Figure 11. The red points, connected by the line, are the means of percent change in housing units from Table 4 for each interval. These have been plotted at the midpoints of the intervals, so the change for the interval from 8 to 10 percent is at 9 percent. Looking at the line starting at the right and moving left, to lower levels of percent change, while there are ups and downs, the general pattern is one of a steady decline to the point at 9 percent, the mean for the 8 to 10 percent change interval. Beyond that, the change tends more to bounce up and down rather than showing any clear trend. Also, one can see from the scatterplot that the higher values are associated with individual extreme values.

Based on the examination of the means in the table and on the graph, we conclude that the change in percent change in housing units levels off when one goes from the interval 10 to 12 percent change to the interval 8 to 10 percent change. From
Figure 11. Scatterplot of Change in Percent Change in Housing United versus Percent Change in Housing Units for Areas Developed as Urban in 1960 with Line Connecting Mean Change in Percent Change for 2 Percent Intervals.

this, we conclude that a value of 10 percent change in housing units can be associated with the end of major development. This value is used as the cutoff for establishing the time to the completion of major development.

Using Change in Density Rather Than Percent Change in Housing Units and Density

As seen in Figure 2, change in density also declines with time, so this might be used as an alternative for measuring the time to completion of major development. (Percent change in housing units and density declined more dramatically to a more consistently lower level, which is why it was originally selected for this purpose.) A procedure analogous to that described above for percent change was carried out for change in density. The resulting cutoff was used to estimate time to the conclusion of major development and these values were employed in analyses similar to those described in this paper. The final conclusion was that this approach was not successful.

As already noted, change in density did not decline as dramatically as percent change, so its utility for determining time to development was in question at the start. Proceeding with the analysis, it became clear that change in density had much greater variation than percent change in density. Comparison for each period after urbanization
showed the coefficients of variation of change in density to be higher than those for percent change.

It was much more difficult to identify an appropriate cutoff for the end of major development from the table and plot of mean change in density for intervals of change in density. A decision was ultimately made but with much less confidence than for percent change in housing units and density.

The change in density cutoff was then used to determine the time to end of major development for each urban area using the same procedure as with percent change in density. One issue that arises in this process are cases in which the change in density or percent change drops below the cutoff in one interval, only to go back above in a subsequent interval. In those instances where the pattern reversed, the earlier instance was ignored and not taken as the end of major development. The estimation of time to end of development using density had over twice as many of these reversals as when percent change in housing units was used, 18 versus 8. This raised further questions about the consistency of using change in density as the indicator of end of development.

Finally, the estimated times to development were used to estimate models of the effects of time and percent change in the urban area analogous to those reported above using percent change in housing units. The coefficient for the inverse of percent change in the urban area was very different than when using percent change to estimate time to development. For the best model, McFadden’s pseudo $R^2$ was substantially lower.