The Arlington Model

To help understand and forecast how Arlington's growth and development will impact its bottom line, Arlington Analytics is developing a microsimulation tool similar to those used by the Congressional Budget Office, University of Pennsylvania, and other organizations for analyzing various components of federal fiscal policy. The key advantage of microsimulation models is that they offer effective ways to provide very detailed forecasts and analyses of the effects of policy proposals.

Unlike most models of this type, the Arlington Model simulates properties and not people. Furthermore, while most other models work with samples to derive their insights, the Arlington Model keeps track of every single property in the county. Simulating every property in Arlington allows us to simulate Arlington in an incredible amount of detail. For example, not only do we produce population estimates and school enrollment projections, but we can predict on an extremely granular level to predict where people are going to live, which parks they will use, and what schools they will attend. We can use the model to estimate future needs for public safety resources by neighborhood, student transportation, and how many people will have local access to Arlington's parks and other facilities.

We make these inferences by linking our census of all properties in the county with numerous other data sources about schools, population, crime, and development to develop a comprehensive look at the many facets of county finances and growth. In this paper, we describe the sources of data and how we use them to project what the county will look like in the next 10 years. Feedback and suggestions are what help make this project better, especially for the many additional features that are still in development.

In this white paper, we describe:

- The Data Sources
- Analyses of Current Economic and Fiscal Conditions
- Forecasting Growth and Finances in Arlington

The Data Sources

Arlington Open Data

We make extensive use of Arlington's Open Data portal. We start by merging together all of the assessments data from 2015, 2016, 2017, 2018, 2019, and 2020. Combining these data sources creates a complete assessment history for all properties in Arlington over the last five years.

We merge the list of properties and their assessments with three other datasets on the Open Data portal. First, the Improvement Interior data contains information about property improvements that are instrumental in estimating student generation factors, population generation factors, and the probabilities of redevelopment for

detached single-family homes. Second, the Real Estate Sales History dataset to provide information about recent sales for the properties. Third, we combine our data so far with a more general dataset called Property that provides additional location data for these properties.

Arlington GIS and other Mapping Data

ArlGIS contains map information about a variety of information that allows us to use mapping tools to locate these pieces of information for each property. First, we match the data on each property in the prior section to each property's boundary the boundaries for every property in the county. We locate each property in the county in each of four different geographical boundaries:

- Civic or Citizens Association
- Voting Precincts
- APS Planning Unit
- U.S. Census Block

When property boundaries were available for the property, we checked several points on and in the polygon to locate the property in the correct location. When the actual property was not listed in the ArlGIS database, we used the coordinates provided in the Open Data portal.[1] In a number of cases, particularly for larger and multi-building apartment complexes, the properties straddled more than one geographical area. In those cases, we had to use numerous other data sources to locate the property and, in some cases, break it up into separate property records and make estimates about how many units were located in each geographical region. We used:

- County information on affordable housing
- Individual property websites
- Data in the assessments and property databases
- Realtor.com (for unit counts per building)
- Apartments.com (for unit counts per building) and other apartment listing services
- White page directories such as Spokeo.com
- Google Maps and Satellite Imagery

Adding together the properties by these boundaries allows us to link property data by geographic area to other datasets that aggregate data over those geographic boundaries.

Schools

We three main pieces of school data for analysis: the school boundary maps, the Fall 2018 Elementary Boundary Process planning documents, and the Bus Eligibility Zones.

The Fall 2018 Elementary Boundary Process planning documents provide planning-unit details on elementary school students. These statistics allow us to estimate our own student generation factors and to forecast school populations at a planning unit level.[2]

[1] In some cases, the coordinates for each property from the Open Data website appeared to be somewhat far removed from the property street address. To minimize errors, we used the property polygon dataset in ArlGIS, which seemed to be more accurate, and used the Open Data information, Google Maps, and our own judgment as backup approaches to locating properties inside the correct boundaries.

[2] APS produced similar documents for the most recent middle school boundary process. They censored student populations by planning unit for student enrollment figures fewer than 10. Unfortunately, they disaggregated the students into individual grades 6, 7, and 8, which meant most of the planning units were simply listed as having fewer than 10 students. This makes the sample unsuitable for statistical analysis similar to the the variety that we conducted on the elementary schools.

Crime

We take information from Arlington's Open Data on incidents reported by the police, which will be used in the future to estimate the effects different types of properties have on crime, estimate the long-run public safety budgetary effects of different types of development, and provide projections for the future demands on public safety services.

Development

Starting with our database of all of the properties in Arlington, we merge this data with several other sources of data to help us forecast properties in the county:

- Quarterly Development Tracking Reports (QDTR)
- APS Housing Unit Forecasts (see Attachment B)
- Development Project Pages (example)
- Our own statistical model of detached single-family home (SFH) redevelopment
- Our own rough estimate of additional building starting in six or more years likely being developed, but not approved or listed in the reports above

We combine the QDTRs with the APS Housing Unit Forecasts because there are a number of developments in the latter which are not listed in the former. When not listed explicitly in the documents themselves, we use maps to locate the planning units, census blocks, and civic / citizens' associations for each of these developments to the extent possible.[3] We exclude detached single-family homes listed in those reports because those reports do not anticipate development beyond a few months.

Population

There are about 180 census blocks in Arlington County. The Census Bureau releases detailed population statistics on all of these blocks. We use the 2010 census—-the last release for which fully detailed population counts are estimated—-combined with our own estimates of the properties located in each census block to estimate population generation factors. This gives us a foundation to construct our own population projections, which we use to support spending forecasts over the next ten years.

We use data from the 2017 American Community Survey to identify Spanish-speaking regions within the county. We merge this data together to allow us to analyze current economic and fiscal conditions in the county.

Analyses of Current Economic and Fiscal Conditions

New Development Assessment

When new properties are built, they typically have a higher assessment than existing properties of the same type. Therefore, we need to develop a method for valuing these new properties. We break down each type of new property into one of eight categories: detached single-family house (SFH), condominium, townhome, hotel, office, market rate apartment, committed affordable (CAF) rate apartment, and commercial space. For townhomes, condos, and detached single-family homes, the entire property is valued at once. Hotels and apartments are valued by taking an estimate of the average assessment per unit and multiplying that by the

[3] Several of the properties in the Housing Unit Forecasts list planning units but not a street address, so they are assigned to the most likely census block.

number of units. Commercial and office space is valued per square foot of new space. The values of these new builds are summarized in Table 1.

Table 1: New Development Assessment Values

Type	Measure	Percentile
SFH	$ (separate estimates for 2-, 3, 4, 5+ bedrooms)	90%
Condo	$ (separate estimates for 1-, 2, 3+; elevator, mid, garden)	75%
Townhomes	$ (separate estimates for 2-, 3, 4+)	80%
Hotel	$ / room	80%
Office	$ / sq ft	75%
Apt – Market	$ / unit	95%
Apt – CAF	$ / unit	71% of market
Commercial	$ / sq ft	80%

In order to evaluate the likely value of a new build, we build a sample of nearby "comps", or similar properties. We start assembling the properties in the same civic association, if there are not enough properties within that civic association, then we add in properties from neighboring civic associations. If there still are not enough properties within the nearby civic associations, then we extend our range to the entire county. For most residential properties, there are plenty of "comps" within the same civic association. For larger buildings such as hotels or commercial buildings, we often need to look beyond the borders of the civic association in order to have enough properties to generate a believable estimate.

In our analysis, we find that a new property is worth more than most of the "comps". We examine a number of new properties built in the last few years and compare them against their comps. A new single-family home, on average, is worth more than 90% of existing single-family home "comps" with the same number of bedrooms. A new hotel is typically worth about more than about 80 percent of comparable hotels on a per unit basis. Office space is typically worth more than 70 percent of comparable nearby office space.

For example, a new five bedroom house would be assessed at $2,072,780, more than 90% of the five bedroom houses in Lyon Village. A new 200 unit hotel in Courthouse would be valued at $57.4 million (a little under $290K per unit), more valuable per unit than about 80 percent of hotels countywide.

Detached Single-Family Home Redevelopment

Arlington Analytics evaluates the likelihood of property redevelopment of detached single-family homes using a basic probit model. We define a property redevelopment as an increase in their assessed value of 30 percent or more. The likelihood of a property being redeveloped depends on how much less the assessment is relative to the newest / most valuable properties in the civic association and the size of the plat.[4]

The further a property's assessed from the most expensive properties, the more likely the property is to be redeveloped. Developers appear most likely to redevelop the least expensive, least valuable properties. These properties often have not been sold for a long time and may not meet the needs of prospective homebuyers. Furthermore, the large margin between the prospective sale price of the existing property and a new

[4] Another good explanatory variable is the age of the building, however, the data does not contain information about when the building was built prior to its most recent development.

development makes it more likely that developers can make enough money to recover its investment in the property.

In most cases, the range of probabilities for a property redeveloping is between zero and three percent. In total, our model predicts that about 200-250 single family houses get redeveloped per year, which is broadly in line with the developments reported in the Quarterly Development Tracking Reports. We predict that most of the future development is concentrated in areas that are already experiencing a high degree of redevelopment.

Student Enrollment

We estimate student generation factors using a statistical model that takes, as inputs, the number of elementary students from each planning unit and the census of residences per planning unit generated by the Arlington Model.[5] We estimate how each type of residence contributes to APS enrollment.[6] Reflecting findings in a 2016 consultant's study, we estimate the effect of newly-renovated / newly-redeveloped detached single family homes separately from the existing stock. Furthermore, reflecting trends in published student generation reports, we separate out the enrollment contributions from apartments near Columbia Pike, which generate enrollment at a rate that far exceeds the enrollment rates of apartments elsewhere in the county. For more information and the results of these estimates please reference the white paper on student enrollment projections.

Population and Development

With a process that's nearly identical to the process used to estimate student enrollment by planning units, we estimate population generation factors for each unit of housing. Based on population statistics from the 2010 Census, we similarly estimate how many people each type of housing produces.

We present the population generation factors, or the predicted average number of people living in each type of property, in Table 2. Unfortunately, there are fewer census blocks than planning units, which means we have to combine several types of residences to get more sensible results. The population generation factors—estimated on 2010 Census data and estimates of the 2010 residence count—are below. When it becomes available, we will update our estimates of population generation factors with data the 2020 Census, which should improve the estimates.

[5] A planning unit is the base geographical unit used by APS during its planning processes. Each planning unit is assigned to an elementary school, a middle school, and a high school. During boundary planning periods, whole planning units are attached to schools.

[6] We use a non-linear, least-squares model with a few restrictions. We restrict houses, condos, duplexes, and townhomes to generate at least as many students on average as the same properties with fewer bedrooms. For example, a 3-bedroom detached single-family home generates at least as many kids as a 2-bedroom detached single-family home on average.

Table 2: Population Generation Factors

Type		Garden	Mid	Elevator
Apt (Market)		1.7	1.5	1.1
Apt (CAF)		2.4	2.4	2.4
Condo (1 bed)		1.2	2.2	1.0
Condo (2 bed)		1.8	2.2	2.0
Condo (3 bed)		4.2	2.2	4.6

Type	2- bed	3 bed	4 bed	5+ bed
SFH	2.7	2.7	2.7	3.8
Town	1.3	2.1	3.4	3.4
Duplex	2.7	2.7	2.7	2.7

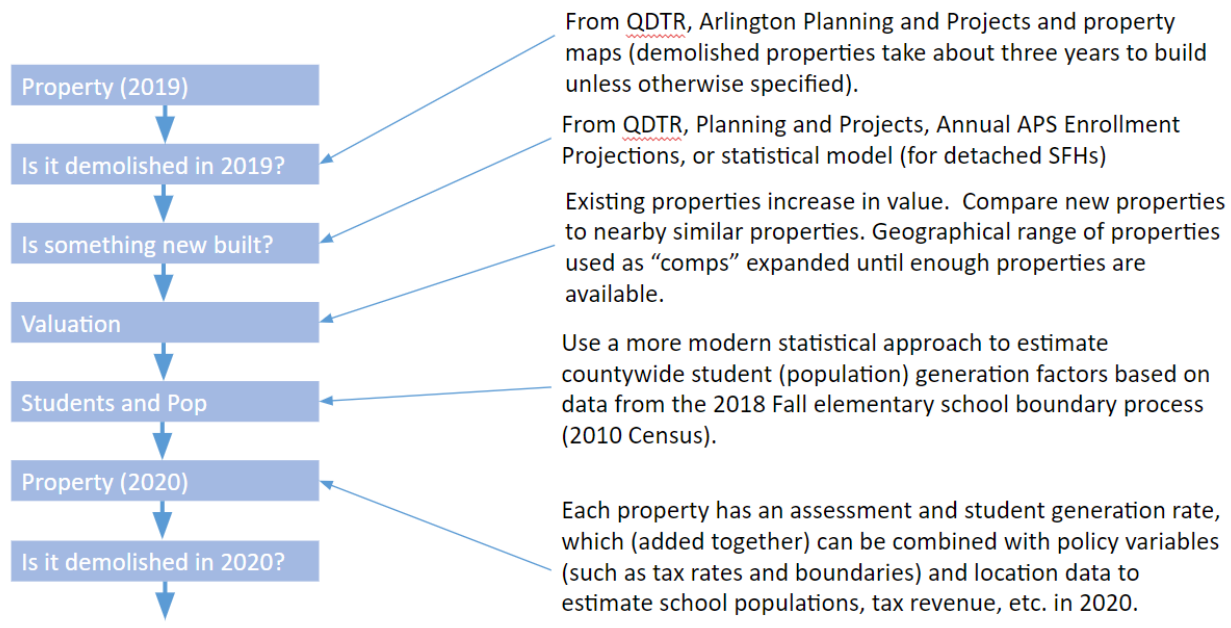
Future Projects

We look forward to conducting numerous analyses relating crime rates, spending, car ownership, capital projects, and other statistics to the property database. Learning about these relationships will help forecast the need for spending on services and infrastructure investment and revenue other categories outside of the personal property tax, which will help us more fully and accurately understand the effects of development on the county's finances.

Forecasting Growth and Finances

We have a database of all of our properties, and we use the analysis described in the previous description to develop a property-by-property simulations. Each property is simulated in each year for about 10 years using the following procedure:

Figure 1: Simulating Properties



Assessments

New construction is valued using the approach described above. Arlington Analytics does have an option that allows it to adjust the assessed value of a new property should we know something specific about that property. For example, luxury hotels tend to generate a higher assessed value than the average hotel. In the vast majority of cases, the existing framework for valuing new properties will likely generate plausible values of new developments, but in those cases in which the property significantly deviates from the characteristics of its local "comps", the Arlington Model has the ability to increase or decrease its value judgmentally.

The assessments on existing construction are increased by a fixed each year. Arlington Analytics chooses an increase in assessments each year by property type.[7] In Table 3, we present the average growth in assessments for existing development by type:

Table 3: Average Assessment Growth Excluding New Development

Type	Avg 2015-2019 Change	Forecast
SFH	2.5 percent	Starts at 3.5%, tapers to 2.0%
Condo	1.5 percent	Starts at 2.5%, tapers to 2.0%
Townhomes	2.1 percent	Starts at 2.5%, tapers to 2.0%
Duplex	1.3 percent	Starts at 3.0%, tapers to 2.0%
Hotel	5.0 percent	Starts at 4.5%, tapers to 3.0%
Office*	5.5 percent	Starts at 4.5%, tapers to 3.5%
Apt – Market	2.9 percent	Starts at 3.5%, tapers to 2.0%
Apt – CAF	2.9 percent	2%
Commercial	4.6 percent	Starts at 4.0%, tapers to 3.0%

* The 2015-2019 change includes effects from changes in vacancy rates. The forecast excludes vacancy rate changes. We project that vacancy rates will decrease by about 4 percent over the next 10 years, which leads to slightly higher growth in office assessments for existing properties.

In the post-WWII era through 2000, real estate prices for houses typically appreciated with or slightly above the rate of inflation. Arlington's average assessment increases from 2015-2019 are roughly in line or slightly higher than inflation for most residential categories, and significantly above inflation for business categories. With the recent Amazon price announcement, real estate prices have increased significantly in certain segments of the market. Already, Arlington has seen larger increases in areas such as Crystal City and Aurora Highlands. Therefore, overall assessment growth for existing properties is projected to exceed inflation for several years across most categories of property. We begin assessments at rates exceeding their historical averages. Certain categories—hotels, offices, and commercial—have a 2015-2019 change that is substantially higher than a long-run inflation rate. Therefore, we taper their rates of assessment growth to rates a bit higher than inflation, but lower than their recent historical averages. We update these forecasts as new data becomes available each January.

[7] The assessment increase for each property type is set for the entire county: houses in Green Valley will appreciate in assessment by the same percentage as houses in Dover Crystal. Using the assessment data from 2015-2019, we saw no clear differences in the total assessment excluding new construction across different civic associations. Aside from having a very small effect on the predicted house redevelopments, this assumption has very little effect on the overall county finances. In future iterations of the model, we may build in the ability to change assessments growth on a civic association basis.

As there are a wide range of plausible forecasts for assessment growth, we provide an application on the web that allows users to change these values (as well as several others). Changing the assessment rates will allow users to see which categories are the most important determinants of the revenue raised by the real estate tax.

Student Enrollments

The number of students has a very strong effect on county expenditures. Student growth increases the need for more educational facilities and professionals. Over the last few years, overall county expenditure increases over inflation and population growth has been driven mostly by increases in school expenditures.

Therefore, we project student enrollments to help adequately forecast county-wide, long-term expenditure growth. We use the student generation factors (SGFs) to forecast enrollment by planning unit. See the white paper on student enrollment for more information about how the SGFs are applied to the property simulation.

Population

The number of people determines the demand for county services, particularly with public safety, libraries, emergency response, utilities, and transportation. In order to construct better estimates of the cost of these services, we need projections for the number of people using those services.

Similar to the way that we use the SGFs to estimate student enrollment, we project population increases with population generation factors PGFs. Unlike SGFs, which have been rapidly increasing over the past 10 years for certain types of properties, PGFs do not appear to grow nearly as fast as SGFs, if at all. PGFs estimated from the 2010 census do a pretty good job of mirroring the 2019 Arlington estimated population. We slightly adjust our PGFs to be consistent with the growth we have historically seen and project in SGFs. In other words, even though the adults "generated" by residential units is fairly stable over time, the growing number of kids in the county increases the total PGFs by a small amount in all property categories except for duplexes.[8]

We look forward to developing this area of the model, especially as data from the 2020 U.S. Census is released.

Occupancy Rate

As Arlington continues to remain an attractive place for businesses, we lower our projected vacancy rate for office buildings over the next 10 years. Over the last few years, the Arlington office vacancy rate has hovered around 20 percent, with the latest calculations of about 18 percent in the 3rd quarter of 2018. A vacancy rate of around 10 percent would be close to the best in the nation. While Arlington continues to improve as a destination for businesses including, but not limited to Amazon, we predict that the vacancy rate will decline to an above-average 14 percent, several percentage points below the national vacancy rate of about 17 percent.

We include an option to adjust the expected vacancy rate on the website.

All of these assumptions put together allow us to evaluate a wide variety of public policy issues in the county. The white paper on the budget explains how the projections detailed in this report are translated into fiscal projections, which will inform decisions about budget and development priorities.

Notes and Acknowledgments

By Jon Huntley. Last updated, February 2020.

For additional tools, data, and analysis, please visit our site at https://www.arlington-analytics.com/.

[8] These estimates will be updated when the 2020 Census data is released.

Special thanks to Kody Carmody, who coded parts of the Arlington Model. Additional thanks to all of the people who have provided advice, particularly members of the APS Advisory Council on School Facilities and Capital Programs Steven Leutner and Lois Koontz. Thanks to Duke Banks, Lars Florio, Peter Rousselot, and John Vihstadt for providing feedback, advice, and recommendations. Thanks to Arlington Public Schools for providing a very rich and helpful dataset on student enrollment from the 2018 Boundary Process. Thanks to Jamie Lees, Arlington County's Chief Data Officer, and other Arlington county staff for filling various data requests, particularly for mapping information. All errors are the author's own.

Feedback is always welcome, and can be sent to analytics.arlington@gmail.com or jon.huntley@gmail.com.