The “Valley Megapolitan” (3-County) Model

The Valley Megapolitan model is an econometric, general equilibrium, structural model specifically designed to produce long term projections. The model is an outgrowth of economic models for metro Phoenix and metro Tucson that have been used over the past three decades to produce forecasts four times per year for the economies of Arizona’s largest two metro areas.

The Valley model introduces a new innovation by embedding a full-blown cohort-survival component into the model’s structure.

The economic component consists of nearly 100 endogenously determined measures of economic activity. Included are personal income by industry and source, wage rates in the private and government sectors, employment at the NAICS two-digit level (and three-digit level where estimation is enhanced), and various measures of economic activity such as consumer spending, residential building permits, gasoline sales, etc.

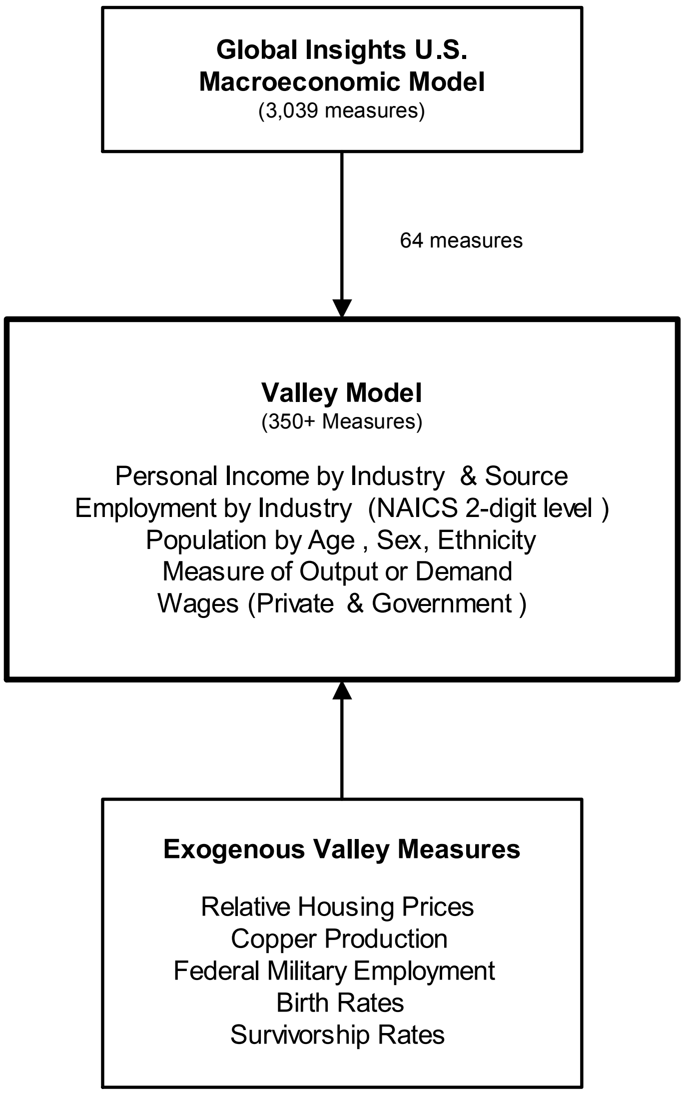
The cohort survival component adds nearly 250 additional measures of population and migration flows. Included are cohorts by age (in 5-year increments, 0-4, 5-9, … , 85&), sex, and ethnicity (Hispanic and non-Hispanic).

A listing of all endogenous measures (variables forecast by the model) is included in Appendix A.

The model is “driven” by some 64 exogenous measures that are forecast by Global Insight, a global leader in economic forecasting. Included are *nationwide* measures such as interest rates, industrial production, consumer spending, employment, homebuilding activity, inflation, relative prices, consumer sentiment, etc. In addition, survivorship rates and birth rates are determined outside the model. A listing of these exogenous measures is included in Appendix B.

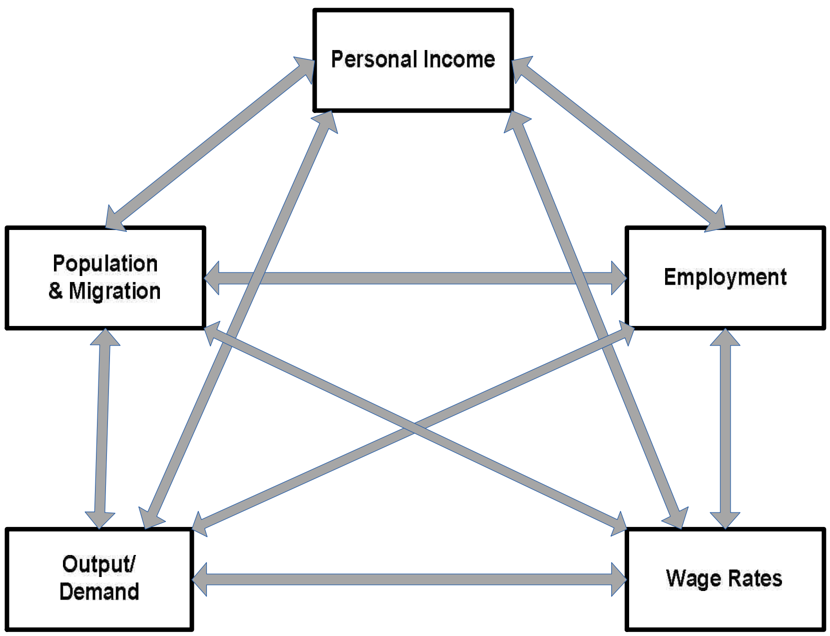
Exhibit 1 shows the relationship between the model and measures that are determined exogenously.

Exhibit 1 – The "Valley Megapolitan Model"



In a general equilibrium model of this type, measures for the various blocks are simultaneously determined, just as economic decisions are determined in "real time." For example, if a large high-tech manufacturer establishes a new plant in the Valley, it will hire as many workers as it can find from the local area and then recruit from other parts of the country. To attract these workers, they will need to offer higher wage rates. The new recruits will need housing which boosts local homebuilding activity and spurs demand for more construction workers and building materials. Higher wages and more workers lift personal income, which in turn generates additional demand for all products and services, which in turn boosts employment throughout all sectors. Higher wages relative to other parts of the country make the area less attractive for future companies looking to open facilities. The simultaneity of these actions is portrayed in Exhibit 2.

Exhibit 2 – Measures are Simultaneously Determined



The model resides in AREMOS, a software package provided by Global Insight. AREMOS is the modeling package of choice for serious econometricians. It is specially designed for time series data, and provides data basing capabilities with full documentation, several econometric methods for estimating coefficients, model building features including a Tarjan ordering algorithm, and a module using the Gauss-Seidel method to simultaneously solve models. It also includes powerful report writing capabilities, presentation quality graphics, and advanced programming capabilities that allow automation of tasks and operations on lists of variables. It also provides a bridge to Microsoft Excel spreadsheets, the format that is used to provide data for this project.

Model Specification

Employment in export based sectors, such as mining and aerospace manufacturing, are modeled using national demand drivers (that are forecast by Global Insight). Non-basic sectors such as health care, trade, state & local government, and construction are modeled using local demand measures. The general specification of these labor demand equations is a function of output and relative prices. Where measures of output are not available, a measure of the level of sectoral activity or primary determinants of output may be used. A detailed description of specification issues may be found in Charney (1983). The following two paragraphs provide examples.

Employment in a basic industry like computer and electronic product manufacturing, for example, is estimated as a function of nationwide employment in that same category, relative wage rates (which is not statistically significant, but is left in because doing so doesn't bias the remaining coefficients), and the ratio of nonfarm employment locally relative to nationwide (which adjusts for timing differences between the local and national business cycles). An autoregressive correction is added to capture any missing information. The equation explains 97% of the variation of the dependent variable.

```
V_EMANDCOMPU
 Cochrane-Orcutt
 ANNUAL data for   16 periods from 1991 to 2006
 Date: 11 SEP 2008

 log(v_emandcompu)

  =    1.01009 * log(emd334) + 0.00027 * v_wrtlp/(ypcompwsdp/eeap)
     (11.6289)              (0.47990)

    + 0.02920 * v_enf.1/eea.1 + 2.77431
     (2.66607)                (5.15707)

 Sum Sq     0.0027   Std Err    0.0156   LHS Mean   3.9316
 R Sq       0.9832   R Bar Sq   0.9771   F  4, 11  160.656
 D.W.( 1)   1.5113   D.W.( 2)   2.1326

 AR_0 =  + 0.65525 * AR_1
        (2.29257)
```

Non basic sectors such as retailing of building materials and garden supply are estimated as a function of real disposable income locally (to capture remodeling and home improvement projects) and residential building permits (new construction). A two-period moving average is used for both to capture the length of time required to complete projects. Over 95% of the variation of the dependent variable is explained.

```
V_ERTBLDGMAT
 Ordinary Least Squares
 ANNUAL data for   24 periods from 1983 to 2006
 Date: 11 SEP 2008

 log(v_ertbldgmat)

  =    0.72946 * log(m2(v_ydp/jpc)) + 0.25638 * log(m2(v_hutot))
     (15.1354)                    (7.01904)

    - 5.08135
     (13.3374)

 Sum Sq     0.1029   Std Err    0.0700   LHS Mean   2.4513
 R Sq       0.9561   R Bar Sq   0.9519   F  2, 21  228.669
 D.W.( 1)   0.4633   D.W.( 2)   0.9409
```

Population by age, sex and ethnicity requires identities for each cohort or cell. For example, the number of female Hispanics age 30-34 in time period T is found by surviving the age group by one year, subtracting the number of 34 year olds who move into the next age category, adding the number of surviving 29 year olds from the prior year that move into the age group, and finally adding the number of net migrants (equation below).

```
v_popFH30_34  = 0.8 * (v_popFH30_34[-1]  * v_survrFH30_34[-1]) + 0.2 * (v_popFH25_29[-1]  * v_survrFH25_29[-1]) +
                v_nmigFH30_34;
```

Survival rates for each cohort are set exogenously. Net migration is allocated to individual cells in a top down approach from the aggregate number of net migrants. The allocation is based on each cell's share of the total as of the most recent data period.

Net migration is modeled in two components: retirement related and all other. Retirement migration (population age 65 and over) is explained by the number of people in that same age group nationwide, relative housing prices in Los Angeles and Phoenix, and the employment-to-population ratio, Valley relative to U.S. The relative housing price measure enters with a lag of three years (once the gap becomes large enough, Californians cash out and move to Arizona). The equation explains over 77% of the variation in net migration.

```
Ordinary Least Squares
ANNUAL data for   26 periods from 1981 to 2006

v_nmig65&/v_pop.1

  =   0.15664 * np65a.1/np.1 + 0.00183 * applosangeles$.3/appphx$.3
     (2.42069)              (2.47634)

    + 0.04384 * (v_enf.1/v_pop.1)/(eea.1/np.1)
     (4.56293)

    - 0.00601 * spike(100,1) - 0.06444
     (6.65698)              (5.73187)

Sum Sq     0.0000   Std Err    0.0009   LHS Mean   0.0017
R Sq       0.8113   R Bar Sq   0.7754   F  4, 21  22.5734
D.W.( 1)   1.5109   D.W.( 2)   2.1048

V_NMIG65&=(??)*v_pop.1
```

Development of demographic data needed for modeling

One of the important design features of the co-joined cohort-survival/economic model is the link between migration and needs of the labor market. To model the two components of migration – work-related and retirement-related -- the characteristics of the migrating population must be known. This is not provided on a year-by-year basis but can be created by aging the population (by age, sex, and ethnicity) in traditional cohort-survival fashion from one period to the next, and then comparing the result to the known distribution for the next time period. The difference between the two is net migration. This results in estimates for each cell in the age-sex-ethnicity matrix. In the end we needed a *historical time series year-by-year* of retirement related migration and work related migration to adequately model these two components. These data are for the 3-county aggregate region.

Sources of data include the US Census Bureau (Census) and the Arizona Department of Health Services (ADHS). The Census provides population counts by county for each year from 1990 – 2006 by age (19 categories – 5-year intervals), sex, and ethnicity (from which we identify 2 groups – Hispanic and non-Hispanic). Unfortunately, while population by age and sex are provided for the decade of the 1980s, ethnicity is not available so special efforts were required to derive these data.

Ethnicity estimates for the decade of the 1980s were derived by interpolating between 1980 and 1990 each cell's share of the total population in that age-sex cohort. For example, the Hispanic share of males age 20-24 grew from 17.1% in 1980 to 24.0% in 1990. The straight line interpolation of the shares was used to derive an estimate of the number of Hispanic males in that age group for each year from 1981-1989.

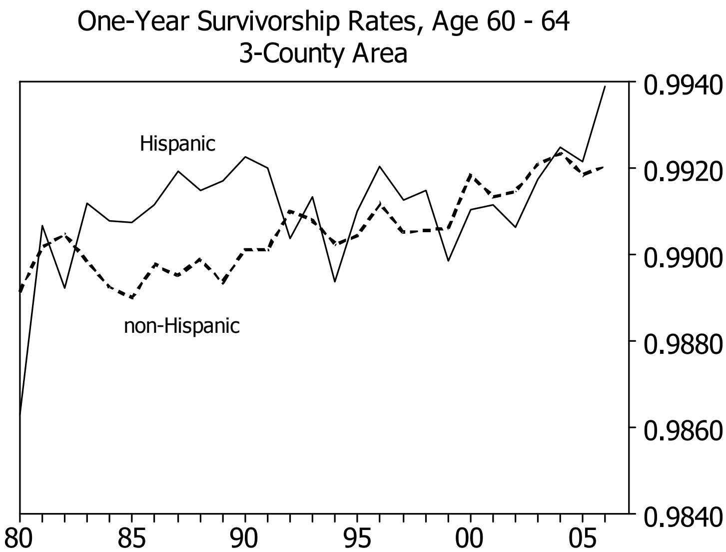
The Census also provides county *aggregate* numbers (totals) for population, deaths, births, and net migration for the period 1970 through 2006. Unfortunately, these totals did not match the totals from the cohort data for some years. The two data sets matched for the current decade, but not for prior years. For those years, the cohort data for each cell was therefore adjusted using the ratio of aggregate to cohort total. These adjustments were relatively small, but were necessary to ensure that the cohort estimates added up.

ADHS provides births (by age of mother and ethnicity) by county from 1980 to 2007. Data for 1991- 2006 comes from their annual reports *Arizona Health Status and Vital Statistics*, while 1980-90 and 2007 are from special tabulations. From these estimates, birth rates for Hispanic and non-Hispanic mothers of child-bearing age were calculated.

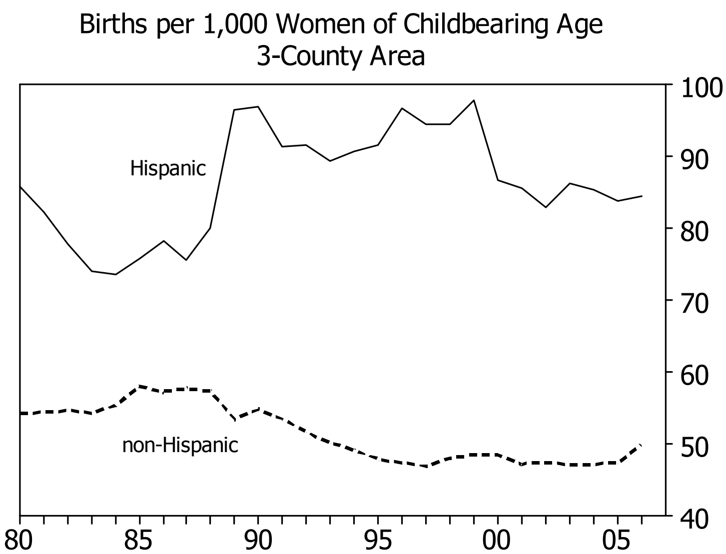
ADHS provides county-by-county deaths for the three categories for the period 1980 - 2007. Again, 1992-2006 comes from their books, while prior years and 2007 are from special tabulations. From these data, survival rates for each cohort were calculated.

Once birth rates, survival rates and the existing population were in place, it was possible to calculate net migration for each cohort. Starting with the known distribution in 1980, the population was survived and births were added to determine the 1981 "natural increase" population. The difference between the derived result and the known distribution for 1981 is net migration. This process was repeated for each subsequent year.

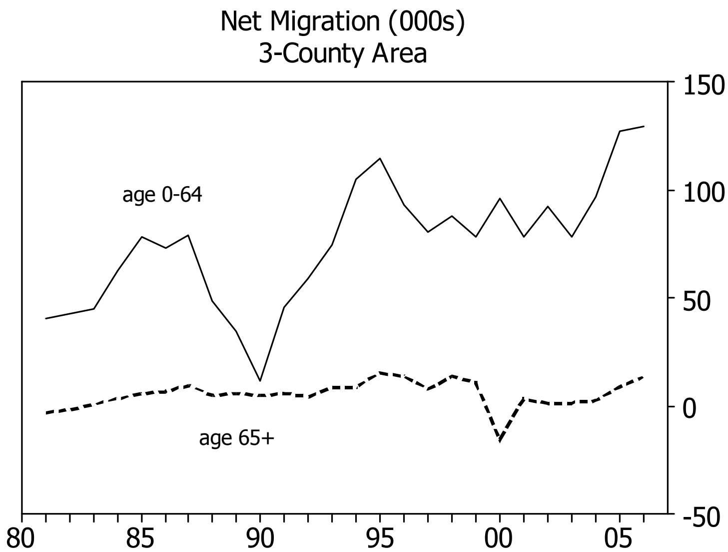
The resulting data set provides new windows on population dynamics over time for the 3-county area. For example, survival rates for older age cohorts have increased significantly in recent years, reflecting advances in detection of disease as well as drugs and procedures for treating various diseases.



Birth rates for Hispanic women were nearly double that of non-Hispanics during the 1990s. During the current decade, Hispanic birth rates have moved lower, while non-Hispanic rates have been stable. The gap was narrowest during the mid to late 1980s.



Migration patterns vary significantly over the business cycle for the "working age" population (and their dependents) while the much smaller "retirement-related" migration has held relatively steady over the past two-and-a-half decades.



This new database provides a treasure trove of information for future demographic research.

Development of economic data needed for modeling

For most of the nearly 100 economic measures, data for the 3-county area was calculated by adding together data for the two metropolitan areas that was already contained in EBR's data bases. Most but not all 2-digit NAICS sectors were available. A few 3-digit NAICS categories were included when doing so improved modeling accuracy.

Challenges arose in dealing with some categories for which Pima county data is not reported. For example, the educational services category is not reported as part of LMI's monthly CES-based estimates. Due to the growth of charter schools (and other privately owned schools), we've found that to successfully model employment of "teachers" one must combine public school employment (part of state and local government) with the private sector educational services category. Data from the QCEW (quarterly census of employment and wages), was used to construct the needed data for Pima County. Utilities were another unreported category, which we were able to generate from QCEW.

Additionally, EBR's data bases contain estimates for CES-reported NAICS categories going back into the 1960s and 1970s, depending on the category. When the SIC system was replaced with NAICS a few years ago, BLS only provided historical data back to 1990. In an extensive effort that used 2-digit SIC data from unemployment insurance records (to which EBR had access the mid-1970s through 1991), bridge tables provided by BLS, and intimate knowledge of the industrial composition of Arizona employment, we were able to create historical data prior to 1990 for Arizona, metro Phoenix and metro Tucson. Without the additional data developed by EBR, econometric modeling with annual frequency models would be severely limited due to the short history.

References

Charney, A. H. (1980). An econometric model of Arizona's economy. *Arizona Review.* vol 29, no. 1.

Charney, A. H. & C. A. Taylor (1983). *Integrated state-substate econometric modeling: design and utilization for long-run economic analysis*.

Taylor, C. A. (1979). Forecasting the Maricopa County economy. *Arizona Review*. vol 28, no. 1.

Taylor, C. A. (1977). Forecasting the Pima County economy. *Arizona Review*. vol 26, no. 2.

Taylor, C. A. & A. H. Charney (1983). *State-substate multiarea modeling: the case for non-rigid aggregation rules*. (Funding from National Science Foundation). Tucson, Arizona: UA Division of Economic and Business Research.