REGIONTRACK
Regional Economic Forecasting and Analysis

Oklahoma Regional Economic Forecasting Model

This report provides an overview of RegionTrack's *Oklahoma Regional Economic Forecasting Model*, an economic analysis and forecasting tool developed over more than a decade as a modeling platform for studying the behavior of the Oklahoma economy. The model is intended to provide both the public and private sectors with timely evidence on current and expected future economic conditions.

The following sections of the report describe the broader framework of the model, structural linkages within the model, data requirements, underlying data estimates, the approach to equation specification, and model estimation and testing procedures.

<u>Model Framework</u>

The model is a large-scale simultaneous equation econometric model built largely in the Cowles tradition. The state model is constructed bottom-up at the 2-digit NAICS industry level and consists of approximately 230 equations and 75 identities. Local area models are also constructed bottom-up at roughly the 2-digit NAICS industry level and consist of approximately 80 equations and 35 identities. The state model can be viewed as a satellite model driven by exogenous forecasts of U.S. economic behavior. The local models are similarly driven by the state and U.S. forecasts. The structure of the model follows from its primary task of providing short- to intermediate-run regional forecasts. Dependent variables are differenced (typically once) unless stationary in levels. Lagged endogenous variables are used to capture both partial adjustment effects and expectational effects though no effort is made to partition or identify the effects. The core of the model defines relationships among employment, income, and output at the industry level; these endogenous variables in turn drive behavior throughout the remainder of the model.

Estimation: The model is estimated in reduced form using a standard OLS system solution within EViews (www.eviews.com). Most explanatory variables are dropped from equations if highly insignificant or display the wrong theoretical sign. Efforts are made to remove serial correlation with additional lagged explanatory variables, but no attempt is made to correct for heteroskedasticity, multicollinearity, or other common classical estimation problems. Seasonal moving average terms are added to some equations (esp. employment) when highly significant. Add factors are rarely used to augment model forecasts though the model will automatically process any user specified adjustment to the forecast.

Regional Aggregation: The state of Oklahoma is the top-level, and counties the bottom-level, geographic unit underlying the structure of the model. Metropolitan area data is formed by aggregating across counties. Custom geographical units such as energy counties, agricultural counties, manufacturing counties, etc… are also created by aggregating across counties.

Industry Aggregation: Income, employment, and output are disaggregated at the 2-digit NAICS level. This level of industry detail is believed to provide sufficient granularity for determining the primary sources of growth and change in the state economy. An alternative industry structure available within the model uses 3-digit NAICS data to create sets of cointegrated industries for employment, income, and output.

Forecast Horizon and Data Frequency: The model is structured around a quarterly dataset and is designed to provide short-run forecasts of 1 to 12 quarters. This forecast horizon captures current-cycle dynamics and is believed useful for internal monetary policy and regional economic discussions focusing on short-run cyclical dynamics. Most equations in the model are estimated using quarterly data; the quarterly estimates can subsequently be used as drivers for annual equations for common low frequency data series – e.g. population components. The forecast is generally solved forward 10 years, though forecasts beyond two years are not scrutinized for reasonableness.

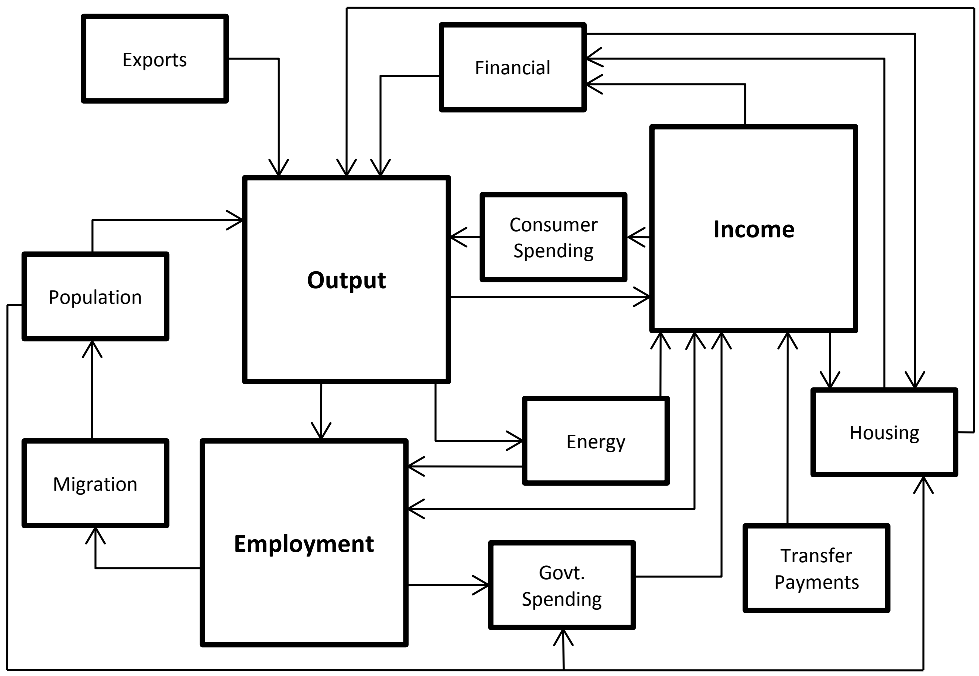
Exogenous Drivers: The state model is a "satellite" model in the sense that it is driven by forecasts of various national variables which are treated as exogenous to the underlying system of equations. Exogenous variables are estimated internally by RegionTrack and derived from numerous sources. Other sources include Federal Reserve surveys and internal forecasts; other private sector forecasts; federal agencies such as U.S. Department of Energy (DOE), Census Bureau, Bureau of Labor Statistics (BLS), and Bureau of Economic Analysis (BEA).

County-Level Models: Smaller bottom-up models for each county are estimated along with the broader state model. County-level variables include wage & salary employment, personal income, gross state product, and LAUS unemployment variables. The bottom-up county level equations can be driven by both state and national variables.

Model Structure and Data

Figure 1 describes the major causal linkages in the model.

Figure 1



1. *Output* – Output is measured using quarterly Gross Domestic Product (GDP) as defined by the Bureau of Economic Analysis (BEA). Nominal output is estimated at the 2-digit NAICS level consistent with industry definitions used in the BEA total employment series (SA25). Output is modeled on a per worker basis and then multiplied by total employment estimates for each sector to determine total output in each sector. Total output at the region level is the sum of output across industries. Estimates of total real GDP (as well as for the mining and manufacturing sectors) are determined within the model by forecasting the BEA state GDP deflator and multiplying it by total nominal output derived in a bottom-up fashion. Comparable U.S. price deflators are used as exogenous variables for the regional deflators.

2. *Employment* – Employment is estimated bottom-up using data disaggregated at the NAICS 2-digit level. Changes in employment produce changes in income and government spending (through unemployment compensation) and can induce migration flows. Four measures of employment are used within the model: 1) BLS Current Employment Statistics (BLS CES) by industry 2) BLS LAUS employment, 3) BLS Quarterly Census of Employment and Wages (BLS QCEW) employment by industry, and 4) BEA total employment (BEA SA25) by industry. The model also provides estimates of farm wage and salary employment,

proprietor employment, and military personnel. CES wage and salary employment data provide the primary measure of employment for the state and major metro areas; QCEW is the primary measure at the county level. BEA data on farm, proprietor, and military employment are used to supplement the CES and QCEW data. QCEW data can be used as the underlying source data for estimating missing CES data at the state level. The model also provides estimates for the labor force, civilian population, number unemployed, and unemployment rate from the LAUS survey. The historical BEA total and proprietor (BEA SA25) employment estimates are raised from annual to quarterly frequency using optimization procedures.

3. *Income* – Income (total and disposable) in the household sector is estimated bottom up based on the earnings of employed workers. Total real income is calculated by adjusting nominal estimates using estimated price indices for the state. The State Quarterly Personal Income (BEA SQPI) survey produced by BEA is the primary income data source. The components of personal income, including earnings (BEA SA05), compensation (BEA SA06), and wage and salary income (SA07), are estimated in the model. The components of personal income are estimated along with earnings by industry at the 2-digit NAICS level. Total earnings (BEA SA05) are estimated for farm and non-farm proprietors. Total earnings by industry are modeled using the average earnings per worker by industry and then multiplying these estimates by employment at the industry level. The industry estimates are summed to form estimates of total income by industry. Estimates of disposable income are formed using BEA estimates of personal taxes (converted from annual to quarterly).

4. *Real Estate*: The state housing market is assumed impacted by population changes, income, and lending activity. The model tracks state housing prices as defined by Federal Housing Finance Authority (FHFA) housing price indexes at both the state and non-metropolitan levels. Census estimates of residential building permits are used to model housing construction activity at the state level.

5. *Demographics:* The demographic component of the model provides estimates of population by component (births, deaths, and migration) and is tied to the economic blocks of the model. Migration flows are the major factor driving fluctuations in population, which in turn determine labor supply. The estimates are used in a top down manner to control estimates by age cohort group.

6. *Consumer Spending:* Estimates of aggregate taxable retail sales are created at the state and county levels using data from the Oklahoma Tax Commission. Other measures of retail activity include trade coverage area, retail leakage, retail buying power index, potential sales, and overall pull factor. Inflation adjusted estimates are also provided for total and per capita retail sales.

7. *Financial*: Major categories of commercial bank deposits and loans are forecasted along with 90-day mortgage delinquency rates and foreclosures.

8. *Energy*: Energy prices and production impact employment, income, and output. Supply side equations are specified for the state-wide rig count and production of crude oil and natural gas. The oil and natural gas equations specify relations for both quantity produced and the state price, which are then multiplied to obtain the value of production.

In order to close portions of the model with respect to output, the housing market, the financial sector, population, income, and exports drive changes in sectoral output. Exports are assumed exogenous to the model.

Large-Scale Econometric Models

Criticisms of large-scale simultaneous equation econometric models are numerous, including their performance relative to naïve models (e.g. random walk), concerns over their theoretical underpinnings (time series frameworks such as VAR and VECM), the high cost of maintaining large-scale models relative to naïve time series models, breakdown in forecasting ability at business cycle turning points, and their fragility as a policy analysis tool (Lucas Critique).

These models nonetheless remain in widespread use by private forecasters and central banks for several reasons. Most importantly is that naïve time series models simply cannot provide an explanation of the behavior underlying a forecast, and users of forecasts will continue to ask 'why.' Lucas (1976) notes that the features of large-scale simultaneous equation models "which lead to success in short-term forecasting are unrelated to quantitative policy evaluation, that the major econometric models are (well) designed to perform the former task only, and that simulations using these models can, in principle, provide no useful information as to the actual consequences of alternative economic policies." In short, the value of these models is largely confined to their capabilities as short-term forecasting tools.

Numerous approaches are used to amend and augment traditional large-scale econometric models to make them better mimic real-world adjustment processes under regime change. The Oklahoma model attempts to integrate many of the following efforts to improve the forecasting accuracy of large-scale regional econometric models:

1. Input-output relationships. Relationships from a custom input-output model of the state economy are used to select cross industry linkages (e.g. manufacturing drives activity in the transportation and warehousing equation) when specifying individual equations.
2. Co-integration methods are often used within regional forecasting models in two ways: 1) to determine the industry structure of the model by grouping co-integrated industries; and 2) to estimate relationships within a co-integrated framework (VECM). The addition of error correction terms in the Oklahoma model mimics the behavior of the VECM framework by capturing both short-run dynamics and long-run equilibrium relationships among variables.
3. Mimic general equilibrium relationships where possible (especially the labor market).
4. Exploit known export-base relationships when choosing driver variables for output equations.
5. Use national panel estimates where appropriate. The Oklahoma model uses equations derived from models of the energy-producing states in specifying the equations for the state's energy sector.
6. Use multiple models when forecasting and evaluate all bottom-up forecasts with top-down models. We use both top-down and bottom-up forecasts for all major aggregate variables in the model for all regions. A VECM of aggregate state employment, income, and output is similarly estimated to evaluate the overall forecast results.

7. Integrate microeconomic structure where possible: leading edge macroeconomic modeling frameworks such as Dynamic Stochastic General Equilibrium (DSGE) models typically lack the sort of industry disaggregation or broad variable set that is desired in many areas of regional analysis and forecasting. These techniques have not yet filtered down to regional modeling in any meaningful way. In practice, large scale DSGE models are difficult to estimate and their structure challenging to communicate. DSGE models, however, seem ideally suited as a source for key exogenous national driver variables underlying top-down regional models.