



**Notes on preparing the data files and  
running the Sixth District common  
factor model v2.0**

**Last Updated: December 2009**

## Overview:

The Sixth District dynamic factor model (D6) is an estimate of the common (albeit unobserved) trend which is behind much of the fluctuations in the economic data series of the States that lie partly (Louisiana, Mississippi and Tennessee) or entirely (Alabama, Florida, Georgia) within the Federal Reserve's Sixth District.

This new version of the model (version 2.0) seeks to be an improvement over the original version (version 1.0) in several ways. First, while the original D6 based its estimation on 25 distinct data series, the new version does so in 48. This provides the model with substantially more information and, at the same time, reduces the incidence that certain key series (like employment) had in the common factor movement. Moreover, the sample period has been increased by a decade, again providing the model with more observations to perform the estimation as well as several other past experiences to be used as a comparison base.

All in all, the D6 still provides a broad measure of the business and economic conditions in our district that is timelier and available at higher frequency than other estimates provided by the BEA like GDP or industrial production. Moreover, unlike an average of the state-level GDPs or an average of state-level factors, the D6 is able to filter out idiosyncratic shocks that would otherwise disproportionately impact one of the states in the sample.

## The data series:

A total of 48 data series will be used to perform the model's estimation. The data set is comprised of 42 monthly and 6 quarterly data series. A Chow-Lin interpolation procedure is used to obtain the monthly estimates of the quarterly data series. This method is thoroughly explained in the Economic Review article. Moreover, contrary to the original version of the model, we have included no series for which data is not available for every state in the district. As such, the use of the *Georgia Purchasing Managers Index* produced by the National Association of Purchasing Management and Kennesaw State University has been discontinued due to the lack of equivalent statistics for the remaining States.

The relevant time series data could be classified into four groups: employment, housing, industrial activity and income statistics. The current data set begins January 1980 up to December 2008. Since most of the series used in this application are not seasonally adjusted we run the model in year-over-year growth rates to avoid problems with seasonality. A brief description of each series is offered below:

**Employment.** The employment statistics used include total nonfarm payroll employment and the unemployment rate for all six Southeastern states. The data on nonfarm employment is monthly, not seasonally adjusted and provided by the Bureau of Labor

Statistics (BLS). The series includes payroll data from construction, trade, government, transportation and utilities amongst other important sectors of the economy. The unemployment data per state is also not seasonally adjusted and provided by the BLS. This statistics track the proportion of the labor force 16 years and over, who were available for work, made specific efforts to find employment yet were unsuccessful at this search.

**Housing.** Given the relative importance of the housing sector (approximately one quarter of all investment spending and about five percent overall GDP), we sought to increase the representation it had in our model by including additional data series. Hence, the housing statistics employed include the number of housing starts as well as the number of housing permits awarded per month per state. The data on housing permits is provided by the Census Bureau and refers to the new privately owned housing units authorized by building permits in each state. The data on housing starts track the number of housing units that are under construction by purpose and design. Both series are not seasonally adjusted.

**Income.** Variations in the agents' disposable income will undoubtedly be governed by business cycle dynamics. States' sales tax receipts and personal income are the series used to account for variations in disposable income throughout the business cycle. State tax receipts are reported monthly by each state's Department of Revenue or Tax Commission and are an important indicator of each state's fiscal strength. Current month rather than year-to-date receipts are employed. On the other hand, personal income is a measure of individuals' purchasing power. The statistic is reported quarterly and is published by the Bureau of Economic Analysis (BEA). Note that following Crone (2000), we exclude transfer payments from our measure of personal income since transfer payments are typically insulated from business cycle dynamics.

**Industrial Activity.** Given the lack of state-level industrial production indexes, we employ two statistics to approximate the degree of monthly industrial activity: the average number of hours worked in manufacturing and the industrial electrical consumption per state. Average hours worked in manufacturing are reported monthly by the BLS and are not seasonally adjusted. This statistic tracks the average of weekly hours worked by production workers in manufacturing industries. If demand for production holds up, businesses will be forced to hire additional workers, signaling a strengthening economy. On the flip side, if demand for production slows, employers will ask workers to work fewer hours before laying them off presumably signaling a weakening economy. Industrial electrical consumption by state is published by the Department of Energy (DOE), measured in Mega Watts per hours (MWh) and is not seasonally adjusted. Data available on the DOE website went back until January 1990, so the first ten years of our data set needed to be backcasted based on the 18 years of available data.

The table below summarizes the data series used by each version of the model and where do these series can be found:

D6 v1.0	D6 v2.0	Source
Non-farm payroll employment	Non-farm payroll employment	Haver Analytics
Housing starts	Housing starts	Haver Analytics
Housing permits	Housing permits	Haver Analytics
Sales tax revenues	Sales tax revenues	Haver Analytics
GA – PMI	Avrg. hours worked in manufacturing	Haver Analytics
	Unemployment rate	Haver Analytics
	Industrial electrical sales	Dept of Energy
	Personal Income	Haver Analytics

### The data files:

The D6 version 2.0 feeds out of three data files which should be brought in text (txt) format into the relevant Matlab directory. Most of the series used will be monthly series, yet however, the model does incorporate data on personal income which can only be found at the quarterly frequency. In short, the data files are:

ysim.txt	Includes monthly, year over year changes of non-farm payroll employment, housing starts, housing permits, sales tax revenues, industrial electrical sales, average hours worked in manufacturing and unemployment rate per state.
datam.txt	Includes the monthly series (in levels) of non-farm payroll employment, unemployment rate and average hours worked in manufacturing.
dataq.txt	Includes the quarterly series (in levels) of personal income per state.

Note that in preparing these files it is important that:

1. Both the ysim.txt and datam.txt files should have the same number of rows. There will be occasions when some series in one file might be released prior to others in the other, yet files matrixes must have the same length for the program to work correctly.
2. There will be at occasions discrepancies between the length of *dataq* and *datam* (and hence *ysim*) because the former contains data released only once per quarter. The D6 v2.0 is able to realize this discrepancy and appropriately correct it by creating a monthly forecast of the quarterly series to match the length of the *ysim* matrix.
3. All three data files should contain the sates as columns and dates as rows.
4. It is paramount that both *dataq* and *datam*, share the same State ordering. That is to say, if GA is the first column in *dataq*, then it should also occupy the first three columns in *datam*. Otherwise, the program will be interpolating the quarterly data of one state with the monthly series of a different one.

## **Running the model:**

There are two relevant files: *runmodel.m* and *rungraph.m*.

The first file runs both version of the D6 common factor model. Upon execution, the file will ask the user which version of the model he would like to run and perform the corresponding estimation. Note that version 1.0 of the model only requires the *ysim.txt* file, yet version 2.0 requires all three files described above.

The second file creates a Matlab graph, which plots the evolution of the mean value of the unobserved common factor through time, together with the NBER recession bars for illustration purposes.

## **A note about the Department of Energy's series:**

Because the Department of Energy (DOE) is currently not updating its monthly electrical consumption series with the needed frequency, at times we'll need to augment the data by producing a contemporaneous forecast. For this we'll exploit the correlation between the electrical consumption in the national IP index and each state's industrial electrical consumption.

The relevant files are: *dataaugment.m* and *ip.txt*

Whenever DOE's series are found to be lagging the other series needed to compute the D6, prior to running the estimation routine one needs to: a) update the *ip.txt* file with the most current national energy IP information available and b) run the *dataaugment* routine to obtain estimates for each state. These estimates should be then included in the *ystim.txt* in lieu of the DOE's missing data.