FEDERAL RESERVE BANK OF ATLANTA

Economic Review

Number 3, 2009

Is More Still Better? Revisiting the Sixth District Coincident Indicator

Pedro Silos and Diego Vilán

The *Economic Review* of the Federal Reserve Bank of Atlanta presents analysis of economic and financial topics relevant to Federal Reserve policy. In a format accessible to the nonspecialist, the publication reflects the work of the bank's Research Department. It is edited, designed, and produced through the Public Affairs Department.

To sign up for e-mail notifications when articles are published online, please visit www.frbatlanta.org and click the "Subscribe" link on the home page. For further information, contact the Public Affairs Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, Georgia 30309-4470 (404.498.8020).
ISSN 0732-1813

FEDERAL RESERVE BANK OF ATLANTA

Economic Review

Volume 94, Number 3, 2009

Is More Still Better? Revisiting the Sixth District Coincident Indicator

Pedro Silos and Diego Vilán*

Assessing the state of an economy is not an easy task and generally involves interpreting myriad and sometimes contradictory indicators. In 2007 the authors unveiled a dynamic common factor model, dubbed the D6 Factor, for the economy of the Sixth Federal Reserve District. This model combined disaggregated information for each of the six states in the Southeast and provided an estimate of an unobserved common component that would account for major shifts in the region's economic activity. The D6 Factor proved superior to the traditional practice of averaging state-level factors because it was able to filter out idiosyncratic shocks that could disproportionately affect one state in the sample.

This article presents an updated version of the D6 Factor that improves upon the original model in several ways. While the original D6 based its estimation on twenty-five distinct data series, the new version uses forty-eight. In addition, the revised model expands the sample estimation period by a decade. These changes provide the updated model with substantially more information while reducing the incidence that certain key series (like housing) had in the original common factor movement. The longer data set also allows for historical comparisons across several business cycles.

Another feature of the new D6 enables it to handle data at both monthly and quarterly frequencies, a feature that greatly increases researchers' options.

The authors find that, when compared to the original D6, the updated model does a better job of describing contemporary economic activity because it significantly reduces noise in the estimation.

JEL classification: C11, C32
Key words: coincident index, dynamic factor model

* Silos is a research economist and assistant policy adviser in the Atlanta Fed's research department. Vilán is a former economist at the Atlanta Fed.

Is More Still Better? Revisiting the Sixth District Coincident Indicator

Pedro Silos and Diego Vilán
Silos is a research economist and assistant policy adviser in the Atlanta Fed's research department. Vilán is a former economist at the Atlanta Fed.

When trying to assess the overall state of an economy, one can usually find a plethora of different and sometimes even contradictory indicators. The unemployment rate, industrial production, inflation, or perhaps a broader measure like gross domestic product (GDP) could all be used to sketch a picture of how business conditions are evolving in general.

However, it is not clear which measure is the right one to focus on since each of these statistics has some relevant information; yet none encompasses everything that we are looking for. Additionally, more often than not, these measures can give conflicting signals about where the economy is in the business cycle, creating confusion and leading to misguiding interpretations and suboptimal recommendations.

One solution to this problem is to combine several measures into a composite index of current economic activity. That objective was the main reason for developing the dynamic common factor model (D6) for the Sixth Federal Reserve District described in Silos and Vilán (2007). That model sought to combine disaggregated information for each of the six states within the district and provide an estimate of an unobserved common component that would account for the major shifts in economic activity in it. In that study we also showed that such an indicator would yield better results than the traditional practice of averaging state-level factors because the D6 Factor was able to filter out idiosyncratic shocks that would disproportionately affect one of the states in the sample.

In this article we continue to build on our original model and seek to improve it in several ways. While the original D6 based its estimation on twenty-five distinct data series, the new version uses forty-eight. Moreover, the sample period for estimation was increased by a decade. Both changes provide the model with substantially more information and at the same time reduce the incidence that certain key series (like housing) had in the original common factor movement. Furthermore, having a longer data set allows for historical comparisons because the model is now being estimated across several business cycles. Last, the new model has the capacity to handle data at both the monthly and quarterly frequencies, a feature that greatly increases the options available to the researcher.

Our aim is that through a thorough understanding of the dynamics behind this common factor, academics, policymakers, and businesspeople will be able to make better diagnoses of the condition of the region's economy. Furthermore, when compared to models for the nation or other Federal Reserve districts, we believe our model could assist in identifying crucial differences and similarities used to develop more accurate diagnostics and in turn support monetary policy formulation.

The methodology

In the late 1980s James Stock and Mark Watson developed an econometric model that estimated changes in the underlying state of an economy. Naturally, these fluctuations are never observed directly but rather are reflected in a wide array of indicators such as industrial production, the unemployment rate, the housing market, and so on. Using the estimated changes in the underlying conditions of the economy, Stock and Watson constructed a coincident indicator for the national U.S. economy.

The model presented in this study remains heavily based on the coincident indicator approach pioneered by Stock and Watson (1989). However, given the size of our data set, we continue to follow closely the methodology employed by Otrok, Silos, and Whiteman (2003), in which the estimation is done sequentially rather than in a one-block approach. Yet perhaps the biggest methodological contribution has been to allow for the inclusion of quarterly data into the coincident index.

When working with state-level data, a researcher faces constraints that do not generally arise at a national level simply because fewer monthly data series are available at the subnational level. Consequently, as we aimed to improve our estimation by expanding the number of series employed, we sought out ways to articulate state-level quarterly data into our model. We achieved this by following Chow and Lin's (1971) proposed method for interpolating lower-frequency into high-frequency data series. (Refer to the appendix for technical details.)

The setup

We continue to model the economic activity in the Sixth District as being driven by an unobserved common factor. Economic activity in this case will be measured by a large set of monthly economic statistics. Disaggregated information for each state is thus incorporated into a model from which the common component is estimated. In such a view, the model as well as the methodology greatly resembles the one described in Silos and Vilán (2007).

There are n observed variables denoted $y_{it}/i = 1,\ldots,n$ that reflect economic activity (employment, income, housing, etc.) in period $t = 1,\ldots,T$. Note that each i refers to a specific data series; for example, $i = 1$ could be housing starts in the state of Georgia while $i = 2$ could be housing starts in Alabama. It should also be noted in the case of a quarterly data series, such as personal income, one should perform the Chow-Lin decomposition prior to performing the estimation. In other words, all data series included in y_{it} should be monthly.

A single common factor, F_t, is assumed to account for all comovement among the n variables. Furthermore, the factor is assumed to be latent (unobserved) and related in a linear fashion to the proposed observables:

$$y_{it} = \gamma_i F_t + \varepsilon_{it},$$

where the error terms follow an autoregressive process of the type

$$\varepsilon_{it} = \varphi_{i,1}\varepsilon_{i,t-1} + \varphi_{i,t-2} + \upsilon_{i,t};\ \upsilon_{i,t} \sim N(0,\sigma_t^2).$$

The equation describing the common factor dynamics has an autoregressive structure as well:

$$F_t = \phi_1 F_{t-1} + \phi_2 F_{t-2} + \omega_t;\ \omega_t \sim N(0,1).$$

Data description

The forty-eight series used to perform the estimation are classified into four groups: employment, housing, industrial activity, and income statistics. Data are monthly, from January 1980 to December 2008. Contrary to the original version of the model, we have included no series for which data are not available for every state in the district. 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, from the U.S. Bureau of Labor Statistics (BLS), are monthly and are not seasonally adjusted. The series includes payroll data from construction, trade, government, and transportation and utilities, among other important sectors of the economy. The unemployment data per state, also from the BLS, are also not seasonally adjusted. This statistic tracks the proportion of the labor force sixteen years old and over who were available for work and 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 5 percent of overall GDP), we increased housing's representation in the model by including an additional data series. 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 are provided by the U.S. Census Bureau and refer 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 households' 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, published by the U.S. Bureau of Economic Analysis (BEA), is reported quarterly. Note that following Crone (2000), we exclude transfer payments from our measure of personal income because 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. 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 U.S. Department of Energy (DOE), measured in megawatts per hour (MWh), and is not seasonally adjusted. Data available on the DOE Web site went back until January 1990, so the first ten years of our data set needed to be backcast based on the eighteen years of available data.

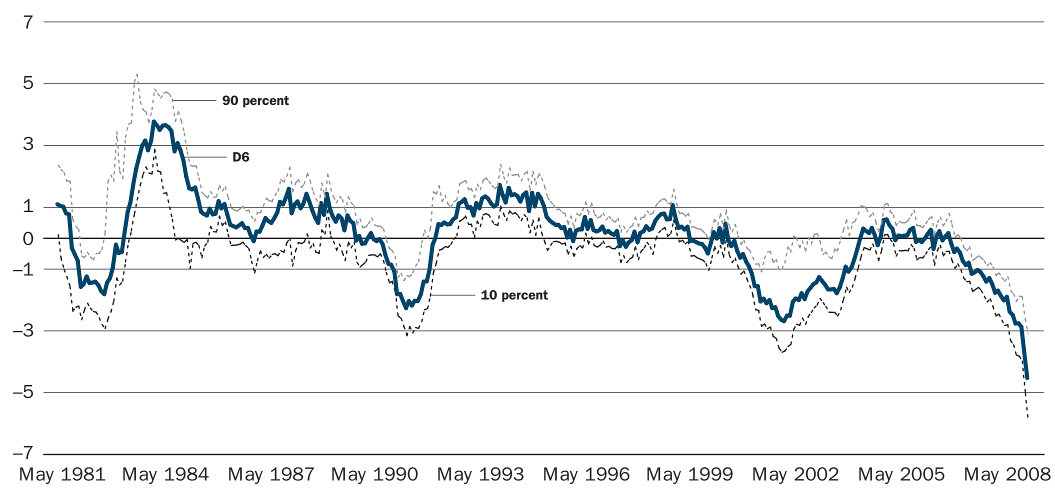
Estimation results

To summarize the results of our model, we first describe the evolution of the unobserved component for the Sixth District and compare the predicted business cycles to those of the national economy as defined by the National Bureau of Economic Research (NBER). Second, we compare the original model with the new one and trace out differences and similarities. Finally, we study the effects that changes to the model have had on the mean factor loadings. Both the greater number of data series and the longer data set had significant impact on the way the model succeeded in fitting the data.

Figure 1 shows the median of the estimated common component along with its tenth and ninetieth percentiles. Given that the common factor is, in essence, a random variable, we should keep in mind that it will have a distribution at each point in time. The percentiles plotted along the median are an indication of the uncertainty surrounding such a distribution.

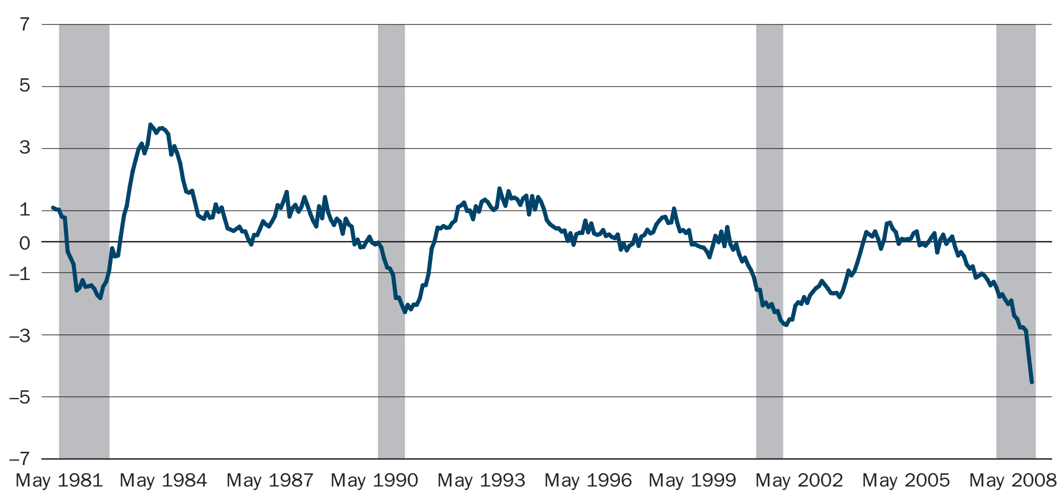
Figure 2 again plots the median of the common factor together with national recessions as established by the business cycle dating committee of the NBER. Periods of national economic downturns seem to be well matched by the model, controlling for the particularities of the southeastern economy. As such, four of the factor's biggest dips coincided with the recessions

Figure 1
Distribution of the model's common factor



Note: Data are from May 1981 to April 2009.

Figure 2
Common factor versus NBER-dated U.S. recessions

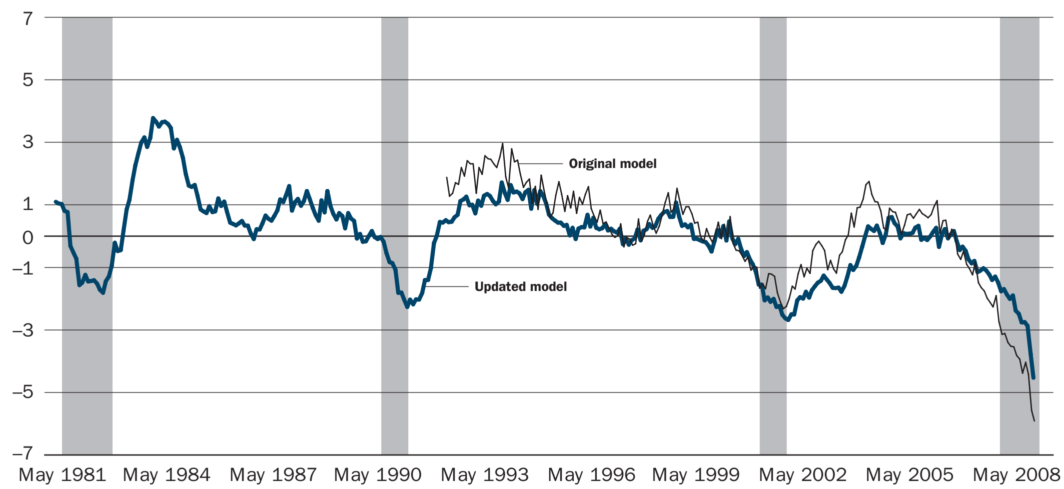


Note: The figure shows the median value for the common factor—the same plot as in Figure 1. The gray bars represent recessionary periods as defined by the NBER.

experienced by the United States since 1981. The model also does a good job at matching the fluctuations during the period usually referred to as the Great Moderation with business cycles of reduced volatilities. Finally, the ongoing financial crisis of 2007–09 is also well portrayed by the strong dive the common factor experienced in mid-2007.

To compare the new with the original version of our model, Figure 3 plots the series of both median values against the NBER recession bars. It is easy to appreciate that both models appear to be very consistent with each other. However, visual inspection reveals a larger variance of the original version of the D6 at high frequencies. In other words, the original indicator looks choppier than the current version. Quantitatively, we can assess this difference by computing the volatility

Figure 3
The two versions of the D6



Note: The gray bars represent recessionary periods as defined by the NBER.

of the two series after having isolated the variation at those high frequencies. We achieve this by subtracting a three-month moving average from the original series and computing the standard deviation of that residual. Doing this for the two series (for those years in which both series are available) shows that the standard deviation at high frequencies of the original D6 is 56 percent larger than that of the new model. This result implies that month-to-month variations in the new model give a clearer signal about the state of the economy because the amount of noise has been significantly reduced. Thus, it is relatively easier to infer the state of the economy by observing the current version of the D6 rather than its predecessor.

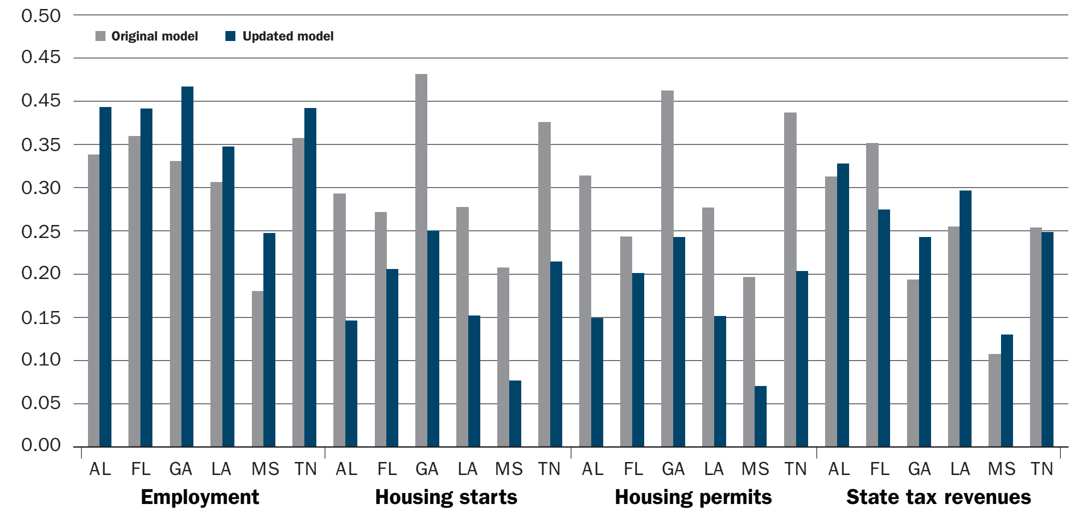
Additionally, we compared the factor loadings in an attempt to assess some of the effects that a longer and richer data set had in the estimation. The factor loadings relate each individual variable with the common factor; they are given by the regression coefficient (γ_i) in the original setup. A positive factor loading implies a positive relationship between a given variable and the D6 common factor. Moreover, the larger the factor loading for a given observable, the more related that observable is to the D6. A comparison of the factor loadings for the subset of variables common to the two data sets is shown in Figure 4. We observe that for most of the series that appeared in both models, the mean loadings remain almost unchanged. Nonetheless, a noticeable variation is the decrease in the relationship between the housing variables (permits and starts) in the new version of the D6. The average size of the factor loadings for housing permits and starts decreases by about 50 percent (from a value of 0.3 to 0.17) from the original to the new version of the model.

Conclusion

Assessing the state of an economy is not an easy task and generally involves the interpretation of several data series, each describing a particular area of the economy. This article attempts to improve a model capable of extracting a common signal from a large array of time series representing different economic activity indicators. This is done with a particular focus on the Sixth Federal Reserve District.

All in all, when comparing the southeastern business cycles (as measured by the D6) with those of the national economy (as defined by the NBER) one can observe that the model does a

Figure 4
Factor loadings



Note: The figure shows the factor loadings for the set of variables common to the two versions of the dynamic factor model.

pretty good job of matching expansions and recessions. Moreover, when compared to the original version of the D6, the current one does a better job of describing contemporary economic activity since the amount of noise present in the estimation has been significantly reduced. Finally, having a greater and longer number of observables allows for a reduction in the factor loadings of the housing market, which tended to dominate in the original version of the model.

Appendix

A primer on Chow-Lin interpolation[1]

A recurrent problem in empirical macroeconomics is the desire to employ high-frequency data when the researcher can only really depend on lower-frequency data. For example, one would like to have an estimate of monthly GDP, yet GDP is released only once a quarter by the BEA.

One way of solving this issue is to take advantage of the relationship between those series released at lower frequencies and those released at higher ones. For example, one could use monthly data on consumption, industrial production, and employment, which are greatly correlated with GDP, to infer what monthly GDP might have been.

A classic paper by Chow and Lin (1971) proposed a method for doing just this. In fact, under their assumptions, their method produces the best linear, unbiased estimate of the high-frequency data. Without loss of generality we will refer here to an interpolation of a quarterly time series into the monthly frequency. But the same approach could be used to interpolate an annual time series into the quarterly frequency and so on.

Assume that y_t^Q is the quarterly time series that the researcher would like to use at a higher (in this case monthly) frequency. We assume a relationship of the following type: $y_t^Q = y_{t,1} + y_{t,2} + y_{t,3}$, where $y_{t,1}$ denotes the series in the first month of the quarter and so on. In the month i of quarter t, the researcher is nonetheless able to observe other variables that are assumed to be related to $y_{t,1}$ in the following manner:

$$y_{t,i} = \beta_1 x_{1,t,i} + \beta_2 x_{2,t,i} + \ldots + \beta_p x_{p,t,i} + u_{t,i}$$

and that

$$u_{t,i} = aLu_{t,i} + \varepsilon_{t,i},$$

where L denotes the monthly lag operator, which could be in the previous quarter, and $\varepsilon_{t,i}$ is independent and identically distributed with mean zero and variance σ^2. Accordingly, the $3T \times 3T$ variance-covariance matrix of monthly errors is

$$V = \begin{bmatrix} 1 & a & a^2 & \ldots & a^{3T-1} \\ a & 1 & a^2 & \ldots & a^{3T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a^{3T-1} & a^{3T-2} & a^{3T-3} & \ldots & 1 \end{bmatrix}.$$

Let y^Q and y^M denote the vectors of quarterly and monthly series of lengths T and $3T$, respectively. Then one could write $y^Q = Cy^M$, where

$$C = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & & & \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 \end{bmatrix}.$$

In that same fashion, we could form quarterly estimates with those series that are available on a monthly frequency. As such, let X^M be the $3T \times P$ matrix of the monthly variables and define $X^Q = CX^M$. With this expression, we can write our original equation in the form $y^M = X^M\beta + u^M$ and premultiply this by C to yield

$$y^Q = X^Q\beta + u^Q,$$

where $u^Q = Cu^M$. In this regression the variance-covariance matrix of errors is CVC'; multiplying this out, the first autocorrelation of the errors can be calculated as

$$\frac{a^5 + 2a^4 + 3a^3 + 2a^2 + a}{3 + 2a^2 + 4a}.$$

With all these building blocks, one could summarize the Chow-Lin interpolation procedure as follows:

First, construct the observed quarterly series from the observed monthly ones:

$$X^Q = CX^M.$$

Appendix (continued)

Second, obtain the OLS estimates from quarterly observed data:

$$\hat{\beta}_{OLS} = (X'^{Q} X^{Q})^{-1} X'^{Q} y^{Q}.$$

Next, calculate the first-order autocorrelation of the residuals from this ordinary least squares (OLS) regression and find the value of a that sets the value of CVC' to this autocorrelation. With this value of a, obtain an estimate of V named $\hat{V}$. Given this estimate, obtain a feasible generalized least squares (FGLS) estimate by

$$\hat{\beta}_{FGLS} = [X'^{Q} (C\hat{V}C')^{-1} X^{Q}]^{-1} X'^{Q} (C\hat{V}C')^{-1} y^{Q}.$$

Finally, obtain the corresponding residual, $\hat{u}^{Q}_{FGLS}$, and use this to obtain the monthly estimate. Chow and Lin show that this will be the best linear, unbiased estimator:

$$y^{M} = X^{M} \hat{\beta}_{FGLS} + \hat{V}C' (C\hat{V}C')^{-1} \hat{u}^{Q}_{FGLS}.$$

1. These notes draw heavily from Wright's (2009) summary on Chow and Lin's interpolation methodology.

References

Chow, Gregory C., and An-loh Lin. 1971. Best linear unbiased interpolation, distribution, and extrapolation of times series by related series. *Review of Economics and Statistics* 53, no. 4:372–75.

Crone, Theodore M. 2000. A new look at economic indexes for the states in the Third District. Federal Reserve Bank of Philadelphia *Business Review* (November/December): 3–14.

Otrok, Christopher, Pedro Silos, and Charles H. Whiteman. 2003. Bayesian dynamic factor models for large data sets: Measuring and forecasting macroeconomic data. University of Iowa unpublished manuscript.

Silos, Pedro, and Diego Vilán. 2007. When more is better: Assessing the southeastern economy with lots of data. Federal Reserve Bank of Atlanta *Economic Review* 92, no. 3:17–26.

Stock, James, and Mark Watson. 1989. The revised NBER indexes of coincident and leading indicators. In *NBER macroeconomics annual* 1989, edited by Olivier Blanchard and Stanley Fischer. Cambridge, Mass.: MIT Press.

Wright, Jonathan. 2009. Unpublished lecture notes, Johns Hopkins University.