MACROECONOMIC ADVISERS
By IHS Markit
Macro Focus

Ben Herzon
ben.herzon@ihsmarkit.com

Joel Prakken
joel.prakken@ihsmarkit.com

June 27, 2018

MA's Second-Generation Dynamic Factor Model

A little over a year ago, we introduced and incorporated into our GDP tracking system our dynamic factor model (DFM).[1] We have been using this model to assess the risks to our assumptions for key monthly indicator series that feed into our GDP tracking. Where appropriate, we've made adjustments to these assumptions and, hence, our GDP tracking to manage these risks. We have since made several improvements to our DFM and have begun incorporating the improved, "second-generation" model into our GDP tracking. Key among the improvements is the explicit inclusion of monthly GDP in the DFM, which allows us to assess the risks to our GDP tracking directly (as opposed to indirectly through risks to key indicator series that feed into GDP). In this Macro Focus, *we summarize the changes we've made to our DFM and illustrate how we are now using it.*

The Updated Model

To review, a dynamic factor model (DFM) is a statistical representation of a panel of data in which each data series is assumed to be a function of a relatively small number of latent common factors and an idiosyncratic error term. The dynamics of the system are captured primarily in the dynamics of the common factors, which are typically assumed to evolve as a vector autoregression.

In our updated model, an individual data series is assumed to be a function of the current and up to two lags of the vector of common factors. Furthermore, the idiosyncratic error term is assumed to follow a moving-average error process with order as high as three.[2] The equations that express the observed data as a function of the common factors and idiosyncratic error terms are referred to as "observation equations." In matrix notation, the vector of n observation equations is given by

(1) $$z_t = \Phi_0 f_t + \Phi_1 f_{t-1} + \Phi_2 f_{t-2} + e_t + \Theta_1 e_{t-1} + \Theta_2 e_{t-2} + \Theta_3 e_{t-3}$$

where z_t is an nx1 vector of observed data, f_t is a px1 vector of common factors, e_t is an nx1 vector of idiosyncratic error terms, Φ_i (i=0, 1, 2) are nxp matrices of parameters, and Θ_i (i=1, 2, 3) are nxn, diagonal matrices of parameters.[3]

As in our first-generation DFM, we assume that the idiosyncratic error terms are mean zero, have a constant variance, and are uncorrelated across equations and over time.[4]

Just as in our prior model, the common factors evolve according to

(2) $$f_t = A_1 f_{t-1} + A_2 f_{t-2} + \cdots + A_q f_{t-q} + \nu_t$$

[1] See "MA's Dynamic Factor Model," *Macroeconomic Advisers Macro Focus*, May 18, 2017.

[2] In our previous model, each individual data series was assumed to be a function of only the current vector of common factors (no lags) and an idiosyncratic error term that was assumed to exhibit no serial correlation.

[3] The data in z_t have been transformed to be stationary and normalized to have zero mean and unit standard deviation.

[4] In this Macro Focus, we highlight the differences between our current and prior DFM's and, as such, have left out some of the details of the model that are common to both. We encourage our readers to review our Macro Focus cited above for an exposition that includes the details common to both models.

where A_i ($i = 1...q$) are pxp matrices of parameters and v_t is a px1 vector of factor error terms. These equations are referred to as "state" or "transition" equations. As in our first-generation model, we assume the factor error terms are mean zero, have a constant variance, and are uncorrelated over time. *Unlike* our first-generation model, we now assume they are *uncorrelated* contemporaneously; i.e., their covariance matrix is diagonal. Our second-generation DFM reduces to our prior DFM when we assume

(3) $$\Phi_1 = \Phi_2 = \Theta_1 = \Theta_2 = \Theta_3 = 0$$

and when the covariance matrix of the factor residuals is not restricted to be diagonal.

The key innovation in this second-generation model is the expansion of the observation equations to include lags of the common factors and moving-average error terms. There were two motivations for making these changes. First, inspection of the regression residuals in many of the observation equations from our prior model revealed serial correlation that, in most cases, called for modeling the error terms as following first-order moving average processes. Generalizing the model to allow for moving-average error terms, then, significantly improves the near-term forecasting performance of the model. Second, we wanted to include monthly GDP in the DFM, and regressions of monthly GDP on the common factors from our prior model revealed that doing so successfully would require allowing for lags of the factors and moving-average error terms in the observation equations.[5]

A Few Notes on Specification and Estimation

The astute reader will note that generalizing the model to allow for lags of the common factors and moving-average error terms in the observation equations greatly increases the potential number of parameters to be estimated in a model that ultimately projects the same number of variables.[6] Increasing the number of parameters in a model can introduce (or exacerbate) problems due to overfitting. To address these issues, we implemented a strategy of imposing zero restrictions on parameters where indicated. In short, we imposed zero restrictions in both the state and observation equations were the common factors did not statistically help the equation.[7] As a result, rather than dramatically increasing the number of parameters to be estimated in the model, we actually *reduced* the number of parameters by about 30%. We wound up with a model that we believe is better specified and will therefore produce better projections.

Beyond the addition of these specification tests, we estimated the parameters of the model in essentially the same way as we did before. From the panel of 36 stationary, normalized data series, we extracted the first 10 principal components and used these components as stand-ins for the common factors in OLS regressions. In the observation equations, we modelled the error terms as moving-average error processes with order determined by the partial autocorrelation functions of the initial regression residuals.

Table 1 compares the r-squared and Durbin-Watson (DW) statistics across the 35 observation equations common to both DFM's. For reference, the total number of independent variables in these 35 equations in the first-generation DFM is 350 (35 equations x 10 contemporaneous common factors per equation). In the second-generation DFM, the total number of independent variables, including the lagged white-noise error terms for equations with moving-average errors, is 340. Despite having fewer total independent variables, the r-squared averaged across these 35 equations is three points higher in the second-generation DFM (0.73) than in the first-generation DFM (0.70). Also, the r-squared is higher in the second-generation DFM in 29 of these 35 equations. Furthermore, in the first-generation DFM, just over one-half of the equations had DW stats more than 0.3 point different from 2.0,

[5] To get the model in state-space form, we redefine the observation residuals as state variables, recasting the observation equations as identities. The Kalman Filter algorithm is somewhat altered to include an observation covariance matrix that is trivially equal to zero.

[6] The number of observation equations is the same as before. In addition to adding monthly GDP to the model, we have dropped the labor-force participation rate.

[7] As a reminder, we also reduced the number of parameters in the covariance matrix of the factor residuals by assuming zero contemporaneous correlation.

indicating serial correlation in the error terms. In the second-generation DFM, where we model the error terms as moving-average processes where indicated, none have DW stats more than 0.3 point different from 2.0.

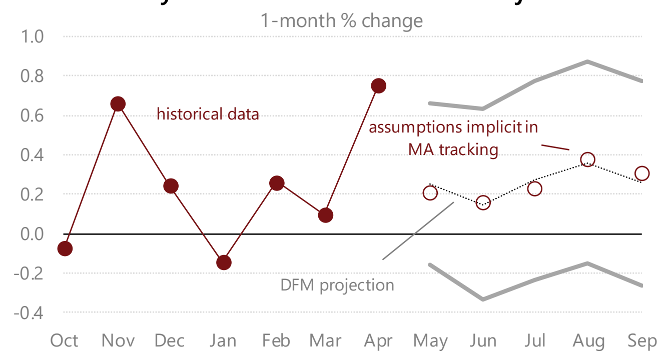
Table 2 compares the r-squared and adjusted r-squared across the 10 state equations in the first- and second-generation DFM's. Across the 10 equations, the first-generation DFM has a total of 300 independent variables (three lags of each of the 10 factors in each equation), while the second-generation DFM has a total of 168 independent variables, or slightly over half of the first-generation. Averaged over the 10 state equations, the r-squared of the second-generation DFM is only 0.03 point lower than that of the first, and the adjusted r-squared is identical. This suggests that the zero-restrictions we imposed were appropriate, and we believe the second-generation DFM will yield better out-of-sample forecasts than the first.

Using the Second-Generation DFM

We are using the second-generation DFM the same way we used the previous one; we gauge the risks to our assumptions for key indicator series by comparing them to the DFM's projections. What's new with the second-generation DFM is the addition of monthly GDP. Not only can we gauge the risks to key indicators that feed into our tracking, but we can also gauge the risks to our GDP tracking directly. This is because in the normal course of our tracking, we maintain a projection for monthly GDP, whose growth rate at the quarterly frequency is identical to our GDP tracking forecast, and whose monthly values are updated whenever we issue a GDP tracking update. Every week, when we run the DFM, we can now compare the projection of monthly GDP from the DFM with the monthly GDP assumptions implicit in our GDP tracking.

The nearby chart illustrates. The solid marked line shows the historical data on the one-month percent change of monthly GDP through April. The unfilled circles show our assumptions for May through September. These are the assumptions implicit in our latest tracking forecast of 4.6% GDP growth in the second quarter and our latest model-based forecast of 3.1% growth in the third quarter. The gray dotted line

Figure 1
Monthly GDP: Historical Data and Projections



Source: Macroeconomic Advisers by IHS Markit

shows the projection from the second-generation DFM when we ran it last Friday. The DFM's projection suggests the risks to our GDP tracking forecast are balanced. Indeed, the DFM projection for monthly GDP gives 4.6% quarterly growth in the second quarter, identical to our tracking, and 3.2% growth in the third quarter, only one-tenth different from our forecast.

Summing Up

We have made a few key improvements to our dynamic factor model: we have reconfigured the model to allow for moving-average error terms in the observation equation; we have allowed for lags of the common factors to appear in the observations equations; we have included monthly GDP in the model; and we have implemented new statistical procedures to reduce the number of parameters in the model and improve the forecasting performance.

Table 1: Comparative Data & Observation Equation Statistics

Stationarity Transformation: L=level, D=first difference, X=mo/mo % change

		R^2		DW	
Business and Consumer Sentiment		Gen1	Gen2	Gen1	Gen2
L	ISM Manufacturing PMI: Composite Index	0.73	0.89	1.1	1.9
L	Philly Fed Mfg Business Outlook: Current Activity Diffusion Index	0.60	0.77	1.1	2.0
X	University of Michigan Consumer Sentiment	0.47	0.47	2.2	2.1
Financial					
D	CBOE Market Volatility Index: VIX	0.74	0.78	2.0	2.1
D	Spread of Moody's Baa Corporate Bond Yield over 20-year Treasury	0.66	0.71	1.8	2.0
X	Stock Price Index: Standard & Poor's 500 Composite (month ave.)	0.76	0.76	1.8	2.0
Inflation					
X	PPI: All Commodities	0.85	0.85	2.3	2.0
X	PPI: Finished Goods	0.83	0.86	2.4	2.0
X	CPI-U: All Items	0.90	0.90	2.0	2.0
X	CPI-U: All Items Less Food and Energy	0.74	0.79	2.1	2.0
X	PCE Chain Price Index	0.90	0.90	2.2	2.0
X	PCE less Food & Energy Chain Price Index	0.78	0.78	2.2	2.2
Labor Market					
D	Nonfarm Payroll Employment	0.72	0.77	1.7	2.0
D	Unemployment Rate	0.69	0.52	2.3	2.0
D	Participation Rate	0.83	N/A	2.3	N/A
Consumption					
X	Retail Sales & Food Serv Excl Auto, Gas Stations & Building Materials	0.52	0.55	2.6	1.9
D	Light Vehicle Sales	0.74	0.78	2.3	2.0
X	Real Personal Consumption Expenditures	0.87	0.87	2.3	2.3
Equipment					
X	Mfrs' New Orders: Nondefense Capital Goods ex Aircraft	0.84	0.91	2.4	2.0
X	Mfrs' Shipments: Nondefense Capital Goods ex Aircraft	0.65	0.71	2.7	2.0
X	Mfrs' Unfilled Orders: Nondefense Capital Goods ex Aircraft	0.67	0.72	1.5	1.8
Housing and Construction					
X	1-Family Housing Units Authorized	0.64	0.57	2.3	2.0
X	1-Family Housing Starts	0.60	0.69	2.5	2.0
X	Construction Spending Less Res. Improvements and Federal	0.64	0.60	1.9	1.9
Inventories					
X	Manufacturers' Inventories: Durable Goods (Census)	0.71	0.71	1.7	1.9
X	Mfrs' Inventories: Nondurable Goods (Census)	0.62	0.76	1.8	1.9
X	Merchant Wholesalers Inventories (Census)	0.63	0.68	2.0	2.1
X	Retail Inventories (Census)	0.62	0.67	2.0	2.1
X	Real Nonfarm Inventories (BEA)	0.76	0.92	1.6	1.9
International Trade					
X	Nominal Goods Exports (Census)	0.52	0.66	2.6	1.9
X	Nominal Services Exports (BEA)	0.58	0.61	2.5	1.9
X	Nominal Goods Imports (Census)	0.52	0.61	2.6	2.0
X	Nominal Services Imports (BEA)	0.61	0.63	2.5	2.0
Production					
X	Industrial Production: Total Index	0.86	0.88	1.6	1.8
D	Capacity Utilization: Total Index	0.88	0.89	1.9	2.0
D	IP: Motor Vehicle Assemblies	0.52	0.47	2.3	2.0
X	Monthly GDP	N/A	0.86	N/A	2.0

Table 2: State Equation Statistics

	R^2		Adj-R^2	
equation	Gen1	Gen2	Gen1	Gen2
1	0.73	0.71	0.70	0.69
2	0.38	0.37	0.32	0.34
3	0.27	0.23	0.20	0.19
4	0.30	0.24	0.23	0.21
5	0.21	0.16	0.12	0.12
6	0.40	0.36	0.34	0.32
7	0.19	0.17	0.11	0.14
8	0.36	0.32	0.30	0.28
9	0.15	0.15	0.06	0.12
10	0.19	0.19	0.11	0.13
average	0.32	0.29	0.25	0.25