High-Mixed-Frequency Forecasting Models for GDP in Selected Southeast Asian Countries

Roberto S. Mariano & Suleyman Ozmucur
University of Pennsylvania

For presentation at the United Nations DESA Expert Group Meeting on the World Economy (LINK Project) – October 21-23, 2015, New York City, USA
Main Topic: technical and practical issues involved in the use of data at mixed and high frequencies (quarterly and monthly and, possibly, weekly and daily) to forecast monthly and quarterly economic activity in a country

Renewed interest in this topic – for timely utilization of high-frequency indicators to update market assessments and forecasts – e.g.,
- Government policy planners
- Financial market analysts
Consider alternative high-frequency forecasting models for GDP growth in the Philippines and other selected Southeast Asian countries, utilizing indicators that are observable at different frequencies.

With particular focus on forecasting performance of dynamic time-series models that involve latent factors (DLFM), compared to alternative approaches such as:
- AR and VAR benchmark models
- Mixed Data Sampling (MIDAS) Regression
- Current Quarterly Modeling (CQM) with Bridge Equations.
Practical and Technical Issues

- Importance of using a parsimonious set of observable indicators
- Combination of mixed-frequency data and latent factors in the dynamic model introduces additional complexities in the estimation of the model
- Data reduction techniques when dealing with a large number of variables in the data set
Mixed-Frequency Data Set

- In general, the data set may include quarterly, monthly, weekly, and daily observations.

- In this paper,
  - target variables – quarterly
  - indicator variables - monthly
Alternative Forecasting Models

- **Quarterly Models**
  - Use observed quarterly values of target variables
  - Aggregate over the quarter for the monthly indicators,
    - Average over the quarter for a stock variable
    - Sum for a flow variable
    - Calculate growth rates from the aggregated series

- **Monthly Models**
  - Treats all the data series as generated at the highest frequency (monthly, in our case), but some of them are not observed
  - Variables observed at the low frequency (quarterly) are treated as having periodically missing or unobserved data points
Quarterly Models

- Benchmark ARMA and VARMA (no indicators used)
  - $Y_{tq} \sim \text{ARMA (p,r) or VARMA (p,r)} + \text{error}$

- Bridge Equations (w/ distributed lags)
  - $Y_{tq} \sim [\text{ARMA (p,r), DL(Z}_{tq})] + \text{error}$

- Bridge – PCA (principal components)
  - $Y_{tq} \sim [\text{ARMA}(p,r), \text{DL(PCA}_{tq})] + \text{error}$

- CQM – bridge modelling for high-frequency updates of forecasts of GDP and its components
Monthly Models

- Mixed-Frequency Vector Autoregressive (MF-VAR)
  - $Y_{tm} \sim [\text{VAR}(p), \text{DL}(Z_{tm})] + \text{error}$
  - Has a state-space model formulation
  - Can use Kalman filtering methods to estimate the model and calculate forecasts
- Mixed Data Sampling (MIDAS)
- Mixed Frequency Dynamic Latent Factor Model (MF-DLFM)
- All these models also provide estimates and forecasts of the target variables disaggregated at the high frequency (monthly)
MIDAS (1)


- Early applications – financial; now also used to forecast macroeconomic time series

- More parsimonious parametrization of distributed lag structures to model the relation of GDP to current and lagged indicators at the monthly frequency

\[ Y_{tm} \sim DL(Z_{tm}) + \text{error} \]

- Estimation method – Nonlinear Least Squares using actual observed data at mixed frequencies
MIDAS (2) - Lag Structures $\Sigma_k c_k L^k$

- **Exponential Almon**
  \[ c_k = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\Sigma_k \exp(\theta_1 k + \theta_2 k^2)} \]

- **Beta Lag**
  \[ c_k = \frac{f(k/K; a, b)}{\Sigma_k f(k/K; a, b)} \]
  \[ f(x; a, b) = x^{a-1} (1-x)^{b-1} \frac{G(a+b)}{G(a)G(b)} \]

- **Linear**
  \[ c_k = \frac{1}{K} \]

- **Hyperbolic**
  \[ c_k = \frac{g(k/K; a)}{\Sigma_k g(k/K; a)} \]
  \[ g(k; a) = \frac{G(k+a)}{G(k+1)G(a)} \]

- **Geometric**
  \[ c_k = \frac{a^k}{\Sigma_k a^k} \]
MIDAS (3) - Extensions

- Autoregressive MIDAS
- Unrestricted MIDAS
- Nonlinear MIDAS
- Asymmetric MIDAS
- Smooth Transition MIDAS
- Markov-Switching MIDAS
- Factor-MIDAS
MF-DLFM (1)
One Common Factor

t = time index for the highest frequency
x_t = latent common factor at time t

y_{t^i} = ith business / economic variables at time t (covers both target and indicator variables
z_t^k = kth exogenous variable at time t

y_{t^\sim^i} = ith observable business / economic indicator at time t
MF-DLFM (2) Model

1. Model for latent factor $x_t$ : AR($p$) + error

$$
\rho(L) x_t = \varepsilon_t, \quad \varepsilon_t \sim \text{iid } N(0,1),
$$

$$
\rho(L) = 1 + \rho L + \rho^2 L^2 + \ldots + \rho^p L^p
$$

2. Model for variables $y_{t}^{i}$ (NOT fully observed!)

$$
y_{t}^{i} = \chi_i + \beta_i x_t + \sum_{k} (\delta_{ik} z_{t}^{k}) + \gamma(L) y_{t}^{i} + u_{t}^{i}
$$

$$
= [\text{AR}(r), x_t, z_t] + \text{error}
$$
State Space Formulation

**Measurement Eq:** \( y_t = Z_t \alpha_t + \Gamma \omega_t + \varepsilon_t ; \varepsilon_t \sim (0, H_t) \)

**State Eq:** \( \alpha_{t+1} = T \alpha_t + R v_t ; v_t \sim (0, Q) \)

- \( y_t \): vector of FULLY observed variables
- \( \alpha_t \): vector of state variables
- \( \omega_t \): vector of predetermined variables such as constant term, trends, exogenous factors, and lagged dependent variables
- \( \varepsilon_t \): measurement shocks
- \( v_t \): transition shocks

Mariano & Murasawa (JAE 2003, OBES 2010)
Aruoba, Diebold & Scotti (JBES 2009)
Empirical Results in the Paper - Philippines

- Two Target Quarterly Variables
  - Real GDP Growth Rate
  - GDP Deflator Growth Rate

- Estimation Period
Philippine GDP & GDP Deflator Growth Rates, 2000-2015

Real Gross Domestic Product growth rate (year-on-year)

GDP Deflator growth rate (year-on-year)
Indicators for Philippine Real GDP Growth Rate (Y52)

- **Y-o-y growth rates of monthly**
  - Industrial production index (Y01)
  - Merchandise emports (Y02)
  - Merchandise exports (Y03)
  - Real government expenditure (Y04)
  - Real money supply (M1) (Y05)
  - Gross international reserves (Y06)
  - Real stock price index (Y07)
  - Real exchange rate (Y08)

- **Y-o-y difference of monthly**
  - Time deposit rate – savings deposit rate (Y09)
  - Treasury bill rate (91-day) – US treasury bill rate (3-month) (Y10)
Indicators for Philippine GDP Deflator Growth Rate (Y53)

- **Y-o-y Growth Rates for**
  - Consumer price index Y(21)
  - Producer price index (Y22)
  - Wholesale price index, Metro Manila Y(23)
  - Retail price index Y(24)
  - Exchange rate Y(25)
  - Money supply (M1) Y(26)

- **Y-o-y differences for**
  - Time deposit rate – savings deposit rate (Y09) or Y(29)
  - Treasury bill rate (91-day) – US treasury bill rate (3-month) (Y10) or Y(30)
Forecasting Models Estimated for the Philippines

- AR
- VAR
- LEI
- Bridge
- Bridge – PCA
- PCA with Two Groups
- MIDAS
- MIDAS – PCA
- DLFM
Estimated Forecasting Models (1)

- Estimated quarterly AR models:
  - AR(2) for real GDP growth rate
  - ARMA(3,3) for GDP deflator growth rate
- Estimated quarterly VAR for Y52 and Y53 with lags 1 and 2.
- LEI includes the leading economic indicator index and its lags as additional regressors in the individual quarterly AR models. The estimated models show 4 lags of the index of leading economic indicators in the real GDP equation, and 2 lags in the GDP deflator equation. In both cases, the estimated AR order is one.
Estimated Forecasting Models (2)

- BRIDGE and BRIDGE-PCA equations are estimated separately for the two target variables. These are quarterly data regressions of target variables on the indicators, with monthly indicators converted to quarterly by averaging.

- MIDAS – unrestricted MIDAS regressions are estimated separately for the two target variables. For MIDAS-PCA principal components of the indicators are utilized in the regressions.
Estimated DLFM

- The two groups of indicators are combined into one and, because of data issues (mostly, timeliness), real government expenditures and the difference between the time deposit rate and the savings rate are excluded.

- Separate monthly DLFM models are estimated for the two target variables. All indicators are monthly, all missing data are in the target variables which are available only quarterly. In both estimated models, it is assumed that there is only one latent factor.

- Estimation is implemented through EVIEWS
DLFM One-Step Ahead Forecasts – Y52 & Y53

One-step-ahead Y52

One-step-ahead Y53
Comparison Results for the Philippines (1)

- Our results, based on one-period-ahead forecasts, indicate that dynamic factor model has the lowest mean absolute error and root mean square error.

- Mean absolute error for real GDP is 0.33, and 0.28 for the GDP deflator. Corresponding statistics are 0.46, and 0.33 for MIDAS, which ranks the second. Principal components, and bridge equations follow these two models; the benchmark AR and VAR models show the biggest errors.
## Mean Absolute Errors over 1999.Q1 - 2015.Q2

### 1999.Q1 - 2015.Q2

<table>
<thead>
<tr>
<th>Model</th>
<th>Real GDP</th>
<th>GDP Deflator</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>0.88</td>
<td>0.71</td>
</tr>
<tr>
<td>VAR</td>
<td>0.83</td>
<td>0.73</td>
</tr>
<tr>
<td>LEI</td>
<td>0.83</td>
<td>0.71</td>
</tr>
<tr>
<td>BRIDGE</td>
<td>0.73</td>
<td>0.47</td>
</tr>
<tr>
<td>BRIDGE-PCA</td>
<td>0.70</td>
<td>0.47</td>
</tr>
<tr>
<td>PCA-2 Groups</td>
<td>0.67</td>
<td>0.47</td>
</tr>
<tr>
<td>MIDAS</td>
<td>0.46</td>
<td>0.33</td>
</tr>
<tr>
<td>MIDAS-PCA</td>
<td>0.70</td>
<td>0.43</td>
</tr>
<tr>
<td>DLFM</td>
<td>0.33</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Source: Table 1a - Condensed
Comparison Results for the Philippines (2) Diebold-Mariano Test

- Diebold-Mariano statistics were calculated to test the forecast accuracy of DLFM relative to the other models, one at a time. Test results show significantly lower errors for DLFM.

- One exception: MIDAS model for the GDP deflator growth rate, where the DM statistic is -1.08 for differences in squared errors. Although errors are lower for DLFM, the difference is not significant enough at the 5% level.

- The bivariate tests done here can be extended to a multivariate test comparing MF-DLFM with the alternative methods taken together - see Mariano & Preve (2012)
## Diebold-Mariano Statistics for Forecast Accuracy

### Alternative Model Versus DLFM

Based on Squared Forecast Errors


<table>
<thead>
<tr>
<th></th>
<th>Real GDP</th>
<th>GDP Deflator</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>2.50</td>
<td>4.13</td>
</tr>
<tr>
<td>VAR</td>
<td>5.56</td>
<td>4.24</td>
</tr>
<tr>
<td>LEI</td>
<td>5.74</td>
<td>4.69</td>
</tr>
<tr>
<td>BRIDGE</td>
<td>3.58</td>
<td>2.96</td>
</tr>
<tr>
<td>BRIDGE-PCA</td>
<td>4.38</td>
<td>3.78</td>
</tr>
<tr>
<td>PCA-2 Groups</td>
<td>4.51</td>
<td>2.59</td>
</tr>
<tr>
<td>MIDAS</td>
<td>3.77</td>
<td>1.08</td>
</tr>
<tr>
<td>MIDAS-PCA</td>
<td>6.94</td>
<td>3.31</td>
</tr>
</tbody>
</table>

Source: Table 2 - Condensed
Comparison Results for the Philippines (2) – Turning Point Analysis

- All models do relatively well, if the prediction is for the level of GDP, real GDP or the GDP deflator. However, not all of them fare well in predicting the turning point in the growth rate of these indicators (Table 3).

- For the growth rates, DLFM appears to have a bigger edge over the other models in predicting turning points.

- DLFM correctly predicts 87% of turning points in real GDP, while MIDAS predicts 74% of them (Table 3). The ratio is 79% for the bridge equation model, and the PCA model. On the other hand, DLFM correctly predicts 89% of downturns, and 85% of upturns. Corresponding ratios for the MIDAS model are 74% and 68%.
### Turning Points

<table>
<thead>
<tr>
<th>Alternative models</th>
<th>n11</th>
<th>n12</th>
<th>n21</th>
<th>n22</th>
<th>correct total</th>
<th>correct downturn</th>
<th>correct upturn</th>
<th>Pearson $\gamma$2</th>
<th>Phi Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real Gross Domestic Product Growth (y-o-y)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>25</td>
<td>10</td>
<td>9</td>
<td>17</td>
<td>0.69</td>
<td>0.71</td>
<td>0.63</td>
<td>8.2</td>
<td>0.37</td>
</tr>
<tr>
<td>Bridge</td>
<td>26</td>
<td>9</td>
<td>4</td>
<td>22</td>
<td>0.79</td>
<td>0.74</td>
<td>0.71</td>
<td>20.7</td>
<td>0.58</td>
</tr>
<tr>
<td>Bridge_PCA</td>
<td>25</td>
<td>10</td>
<td>3</td>
<td>23</td>
<td>0.79</td>
<td>0.71</td>
<td>0.7</td>
<td>21.5</td>
<td>0.59</td>
</tr>
<tr>
<td>LEI</td>
<td>24</td>
<td>11</td>
<td>9</td>
<td>17</td>
<td>0.67</td>
<td>0.69</td>
<td>0.61</td>
<td>6.9</td>
<td>0.34</td>
</tr>
<tr>
<td>MIDAS</td>
<td>26</td>
<td>9</td>
<td>7</td>
<td>19</td>
<td>0.74</td>
<td>0.74</td>
<td>0.68</td>
<td>13.5</td>
<td>0.47</td>
</tr>
<tr>
<td>MIDAS_PCA</td>
<td>27</td>
<td>8</td>
<td>5</td>
<td>21</td>
<td>0.79</td>
<td>0.77</td>
<td>0.72</td>
<td>20.1</td>
<td>0.57</td>
</tr>
<tr>
<td>PCA</td>
<td>25</td>
<td>10</td>
<td>3</td>
<td>23</td>
<td>0.79</td>
<td>0.71</td>
<td>0.7</td>
<td>21.5</td>
<td>0.59</td>
</tr>
<tr>
<td>VAR</td>
<td>24</td>
<td>11</td>
<td>8</td>
<td>18</td>
<td>0.69</td>
<td>0.69</td>
<td>0.62</td>
<td>8.5</td>
<td>0.37</td>
</tr>
<tr>
<td>DLFM</td>
<td>31</td>
<td>4</td>
<td>4</td>
<td>22</td>
<td>0.87</td>
<td>0.89</td>
<td>0.85</td>
<td>32.7</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Concluding Remarks (1)

- We have considered alternative models for use of data at mixed frequencies (quarterly and monthly and, possibly, weekly and daily) to forecast monthly and quarterly economic activity in a country.

- While alternative models are mostly data-intensive, DLFM presents a parsimonious approach which depends on a much smaller data set that needs to be updated regularly. But it also faces additional complications in methodology and calculations as mixed-frequency data are included in the analysis.
Regarding high-frequency forecasting of GDP growth in the Philippines, our preliminary results based on static in-sample simulations and turning point analysis indicate that the Dynamic Latent Factor Model performs better than MIDAS regression, Bridge Equations and the benchmark Autoregressive models.
Concluding Remarks (3)

Further analysis and empirical applications are needed to determine robustness of these findings – especially in efforts to

- Introduce more elaborate error structures,
- Experiment further with multiple latent common factors,
- Consider other exogenous indicators in the high-frequency models
- Cover out-of-sample and dynamic simulations of the estimated models, and
- Expand the analysis to other Southeast Asian countries
THE END