

**12th National Convention on Statistics (NCS)**

EDSA Shangri-La Hotel, Mandaluyong City

October 1-2, 2013

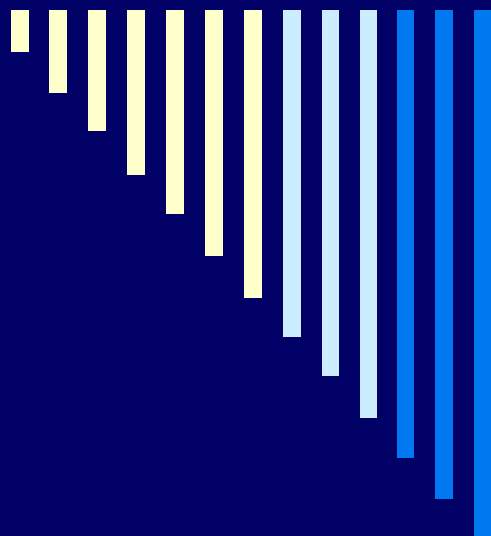
**HIGH-MIXED-FREQUENCY DYNAMIC LATENT FACTOR FORECASTING MODELS FOR  
GDP AND INFLATION IN THE PHILIPPINES**

by

Roberto S. Mariano

For additional information, please contact:

|               |   |
|---------------|---|
| Author's name | : Roberto S. Mariano  |
| Designation   | : Professor Emeritus of Economics   |
| Affiliation   | : Singapore Management University   |
| Address       | : Singapore Tel no. : 65-6828-0832  |
| Affiliation   | : University of Pennsylvania  |
| Address       | : Philadelphia, Pennsylvania, USA   |
| Tel. no.      | : 215-898-7701  |
| E-mail        | : <a href="mailto:rsmariano@smu.edu.sg">rsmariano@smu.edu.sg</a> ; <a href="mailto:mariano@econ.upenn.edu">mariano@econ.upenn.edu</a> |



# High-Mixed-Frequency Dynamic Latent Factor Forecasting Models for GDP in the Philippines

**Roberto S. Mariano\***

University of Pennsylvania & Singapore Management University

*12<sup>th</sup> National Convention on Statistics – Manila, Philippines*

*\* Partial support from the Sim Kee Boon Institute (SKBI) at Singapore Management University is gratefully acknowledged . Earlier version of this paper was presented at the Asian Meeting of the Econometric Society, held in Singapore on August 2013 and at ADB Conference in May 2012*



# Introduction

- Main discussion: modeling and technical issues involved in using data at high and mixed frequencies to forecast economic activity in the Philippines
  - Focus on dynamic models that combine latent factors with a parsimonious set of indicators that are observable at different frequencies
  - Particularly useful for policy makers who need to monitor the state of the economy in real time as well as watchers of financial market developments who would be interested in timely use of high-frequency indicators
-



# Alternative Approaches to Handle Mixed Frequency Data

- **Bridge equations** – e.g. Current Quarterly Modeling (**CQM**)
  - **Mixed Data Sampling (**MIDAS**)**  
regression  
and recent extensions
  - **State-Space Formulation**
    - Mixed Frequency VAR
    - Mixed Frequency Factor Models
-



# 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
-



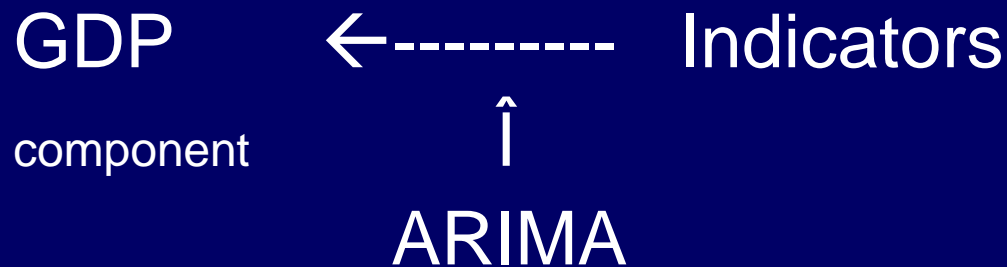
# Bridge Modelling

- Data intensive
  - Bridge equations link the low-frequency variables and time-aggregated high-frequency indicators
  - Not standard macroeconomic structural models
  - Specific indicators are chosen not because of causal relations but because of statistical evidence that they contain timely observable information about the target variable
-



# Two Model Ingredients

- ARIMA model of the indicator variables to project their values over the forecasting period
- ARIMAX bridge equations relating the target variable to the aggregated values of the indicators





# CQM: Motivation and Origins

- Klein and Sojo (1989) and Klein and Park (1993, 1995) – for the US economy
  - Now, weekly updates of US forecasts using CQM - Klein and Ozmucur; also Kumasaka
  - CQM models for updating quarterly forecasts in China, Japan, Hong Kong, Russia, Mexico, and other countries
  - CQM for Philippines, Malaysia, Thailand have been constructed - Kumasaka & Ozmucur (2012,2013)
-





# CQM Methodology

- Main objective is to forecast the national income components – typically available quarterly
  - Alternative sides of GDP accounting
  - Main approach – relate the GDP components to indicator variables
  - Indicator variables are observable, with sufficient correlation to the GDP component; and with enough lead time relative to the GDP components
-



# Treatment of GDP Components

- Use ARIMA models to relate GDP components to quarterly and monthly “indicator” variables
  - If no indicators are available, estimate an ARIMA model for the GDP component
-



# Treatment of Indicator Variables

- For monthly indicators, use averages over the quarter – update averages (or estimates of them) as more monthly observations become available
  - Use ARIMA models to forecast monthly and quarterly indicators
-



# Principal Components and Structural Variables

- When there is a large number of indicators, most likely with a high degree of multicollinearity, data reduction techniques, such as principal component analysis, are employed.
  - Also can include some structural variables in explaining some of the GDP components
-



# MIDAS

- Initial reference: Ghysels, Santa-Clara, and Valkanov (2004)
- See relevant MATLAB toolbox
- More parsimonious parametrization of distributed lag structures to model the relation of current and future GDP to current and lagged indicators
- Exponential Almon lag structure

$$\sum_k c_k L^k, \text{ where}$$
$$c_k = \exp(\theta_1 k + \theta_2 k^2) / \sum_k \exp(\theta_1 k + \theta_2 k^2)$$

- Extend to Autoregressive-MIDAS – add a lagged  $y$  to the regressors
- Early applications – financial; now also used to forecast macroeconomic time series



# Multi-Frequency VAR

- Treat each series as generated at the highest frequency, but may have periodically missing elements
  - E.g. – with monthly as the highest frequency, real GDP would be available only at time  $t = 3, 6, 9, \dots$  - e.g., for Singapore, Abeyasinghe (1998, 1999)
  - A VAR model for target and indicator variables can be re-cast with a state-space representation – Zadrozny (1988); Murasawa and Mariano (2010)
  - Kalman filtering technology can then be applied to the state-space formulation to estimate the model
-



# Dynamic Latent Factor Modelling

Main objective remains the same: forecast monthly prices and economic activity (target variables)

- Basic philosophy: macroeconomic fluctuations are driven by a small number of common shocks or factors and an idiosyncratic component peculiar to each economic time series
  - Can introduce another feature: use of mixed-frequency data - at both lower and higher frequencies than monthly.
  - This complicates, but also enriches, the analysis
-



# A Bit of Background

- Seminal papers: Sargent and Sims (FRB Minneapolis 1977), Stock and Watson (NBER 1989)
  - More recently revived for forecasting purposes in the U.S. and larger European countries – Foroni & Marcellino (WP 2012, WP 2013).
  - Earlier work (e.g., Stock and Watson) develop single factor models to construct composite indices of economic activity based on a handful of coincident indicators. More recent studies use the model to extract unobserved common factors from a large collection of observable indicator variables.
-





# A Bit of Background ...

- Another (related) application has dealt with combining mixed frequencies in constructing composite indices – e.g., Mariano and Murasawa (2003 JAE), Aruoba/Diebold/Scotti (JBES 2009)
  - Furthermore, the estimated factor model, properly validated, also may be used to forecast macroeconomic variables of interest - e.g., Mariano and Murasawa (OBES 2010) – and the application to the Philippine case in this paper.
-



# Model Formulation

- Common factors are latent, explained by their joint dynamics and possibly, interactions with observable indicators
  - Dynamics of the target variables depends on own lags, the unobservable common factors, and, possibly, exogenous factors
  - The system may also have other observable variables that serve as indicators for the latent common factors
-



# Relation to Business Condition Indices

A similar modeling approach is used in

- Mariano and Murasawa (2003, 2010) in constructing an improved coincident economic index indicator in the US using mixed frequencies

  - GDP – quarterly

  - Employees on non agriculture payrolls – monthly

  - Personal income less transfer payments – monthly

  - Index of industrial production – monthly

  - Manufacturing and trade sales - monthly

- Aruoba, Diebold & Scotti (JBES, 2009) in constructing a “real-time” (daily) BCI for the US, using four indicators

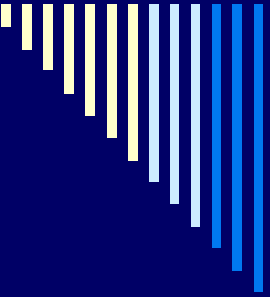
  - GDP – Quarterly

  - Employment – Monthly

  - Initial jobless claims – Weekly

  - Yield curve premium rate - Daily

---



# Model Framework - One Common Factor

$x_t$  = latent common factor at time  $t$

$y_t^i$  =  $i$ th business / economic indicator at time  $t$

$w_t^k$  =  $k$ th exogenous variable at time  $t$

$\tilde{y}_t^i$  =  $i$ th observable business / economic indicator at time  $t$

---



# Model

1. Dynamic latent factor model for  $x_t$

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

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

2. Model for Indicator  $y_t^i$  (NOT fully observed!!!)

$$y_t^i = \chi_i + \beta_i x_t + \sum_k (\delta_{ik} w_t^k) + \gamma(L) y_t^i + u_t^i$$



# State Space Formulation of the Model

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)**

---



# State Space Formulation ...

- “Missing” observations need to be factored in constructing the observation matrices in the state-space formulation
  - Distinguish treatment of stock and flow variables
  - The linear state-space formulation is only an approximation to the true relationship – would need nonlinear filtering procedures, typically through stochastic simulations, to get exact solution
  - But linear approximations may suffice
  - Other possible approaches: EM algorithm, Ghysels’ MIDAS, Bayesian, Forecast combination (e.g., Abeyasinghe & Tay)
-



# Dynamic Factor Model for Singapore (Chow & Choy, 2009)

- Builds a dynamic factor model for Singapore to analyze business cycles and to forecast real economic activity and price inflation in Singapore
  - Uses a large panel data set of Singapore's macroeconomic variables and global economic indicators relevant to macroeconomic fluctuations in Singapore
    - 41 global and 136 national variables, over 62 quarters (to accommodate availability of data on electronics)
    - Quarterly and monthly data; monthly data were aggregated or averaged
    - Almost all variables are transformed into Y-O-Y growth rates. Exception: seasonally adjusted data are used for interest rates, exchange rates, unemployment, and business expectations
-





# The Dynamic Factor Model

$$y_{it} = \lambda_i(L) x_t + \varepsilon_{it}, \quad i = 1, 2, 3, \dots, N; t = 1, 2, 3, \dots, T$$

$$\lambda_i(L) = \lambda_{i0} + \lambda_{i1}L + \dots + \lambda_{is}L^s$$

$y_{it}$ : 1x1 observation on the  $i$ th indicator variable at time  $t$

$x_t$ :  $q \times 1$  vector of latent common factors;  $q \ll N$

$\varepsilon_{it}$ : idiosyncratic component of  $i$ th indicator, which can have limited serial correlation and cross-correlation

$s > 0$  produces the dynamic version of the model

---



# Reformulation of the Model

$$Y_t = \Lambda F_t + \varepsilon_t$$

$\Lambda$  =  $N \times r$  matrix of factor loadings

$F_t = (f'_{t,1}, f'_{t,2}, \dots, f'_{t,s})$ , stacked common factors,  $r \times 1$

$$r = q(s+1)$$

Can apply principal component analysis to extract the common factors. The column space spanned by the common factors can be consistently estimated (as  $N$  and  $T$  go to infinity jointly) by taking principal components of the covariance matrix of  $X$ .

Also can use recent methodology proposed by Bai and Ng (2007) to determine the number of common factors present in the data.



# Empirical Results in the Singapore Study

- Estimated factor model explains well the observed fluctuations in real economic activity and price inflation
  - Four latent common factors are found and broadly interpreted as cycles in
    - World
    - Regional
    - Electronics
    - Domestic economy
  - Forecasts generated by the estimated common factors are generally more accurate than the predictions of univariate models VARs that employ leading indicators
-



# Empirical Results - Europe

- Main Reference – Foroni & Marcellino (2012, 2013)
  - Discusses the forecasting performance of the alternative procedures outlined in this presentation
  - Target variable for forecasting: quarterly growth rate of Euro area GDP
  - Indicator variables: large set of monthly observations
-



# Empirical Result: U.S.

- From Mariano and Murasawa (2010)
  - Component indicators
    - Real GDP, quarterly
    - Employment nonagriculture
    - Personal income less transfer payments
    - Index of industrial production
    - Manufacturing and trade sales
  - MF-VAR and MF-Factor models are estimated
-



# Empirical Result: U.S.

Compare

- Smoothed estimate of monthly GDP growth rates from the estimated MF-VAR model – VAR(1)
  - Smoothed estimate of monthly GDP growth rates from the estimated MF-Factor model – (two-factor model)
  - smoothed estimate of the common factor component of monthly real GDP in the estimated two-factor model
-



# Empirical Result: U.S.

- The three estimates of monthly real GDP tend to be close to quarterly real GDP, in general
  - But the VAR and 2-factor estimates show more volatility and tend to be close to each other. While the common factor component is smooth, and at times differs from the other 2 substantially
  - Turning points in monthly GDP and its common factor component may differ, because the latter excludes the specific factor in GDP
  - The zig-zagging pattern in the 2-factor estimate indicates a negative serial correlation in the specific factor in monthly GDP – as observed in earlier studies.
-



# Preliminary Results - Philippines

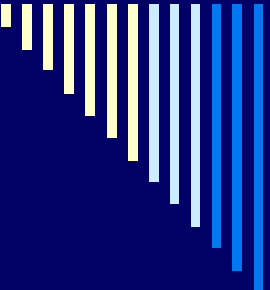
- Basic model:
    - One-factor model with  $AR(q)$  shocks for the indicator variables
    - $AR(p)$  process for the one common latent factor
  - Four indicators (seasonally adjusted with X12-ARIMA)
    - GDP at constant 2000 prices (quarterly)
    - Total employed persons (quarterly)
    - Volume of production index (monthly)
    - Volume of net sales index (monthly)
-





# Preliminary Results

- Algorithm used is developed by Murasawa and done through Ox
  - Applies a modified Kalman filtering procedure to get maximum likelihood estimates of the model parameters in the state space representation
  - The Akaike Information Criterion (AIC) and BIC are used to select the optimal values of  $p$  and  $q$ .
-



# Preliminary Empirical Results – Philippines

- At the estimation stage, there were instances of non-convergence of the MLE algorithm
  - The empirical results indicate that the underlying structure is more complicated than the basic one-factor model specified in this initial effort.
-



# Philippine Study – Moving Forward

- ❑ Improve the GDP equation by including (observable) explanatory variables
  - ❑ Expand the list of indicator variables, including inflation, Philippine stock market price index, Philippine peso-US \$ exchange rate, merchandise imports, merchandise exports, and electricity consumption
  - ❑ Experiment with two or more common factors
  - ❑ Compare forecasting performance with alternative procedures
-



# Concluding Remarks

- Temporal aggregation still the predominant technique – all data are sampled at the same lower frequency
  - Mixed-frequency data matter – procedures that allow use of different frequencies and the timeliness of the data improve the forecasts. But still not clear which procedure is superior.
  - Bridge models, where a dynamic equation is estimated between the low-frequency variable and time-aggregated high-frequency indicators, generally reduce the forecasting error, and are useful instruments, especially in nowcasting
-



# Concluding Remarks ...

- MIDAS, a mixed-frequency regression, explains a low-frequency variable with high-frequency indicators using parsimonious distributed lag models. Choice of weighting functions is not clear-cut and depends on the specific analysis.
  - MIDAS with an AR component performs better than the corresponding approach without AR.
  - MIDAS was initially applied to investigate stock market returns or future volatility but has recently been applied to forecast macroeconomic variables, with promising results for short-term forecasting.
-



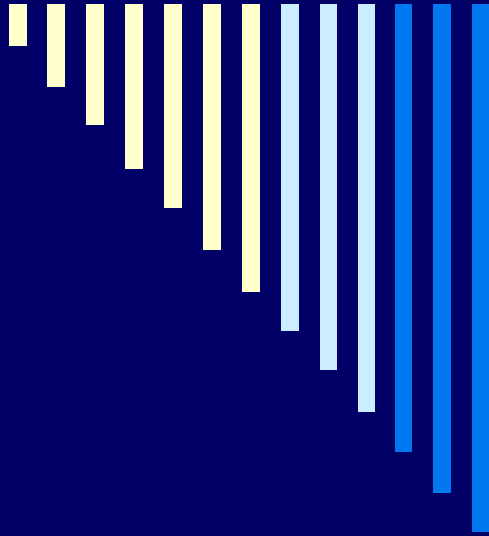
# Concluding Remarks ...

- State-space framework for mixed-frequency factor models and VAR representations. This treats the low-frequency variable as a high-frequency series with missing observations; and uses Kalman filters to implement real-time filtering
  - MF-VAR appears to outperform the MIDAS approach only at longer horizons
  - BENEFIT- allows estimation of missing high-frequency data
  - CAVEAT – this framework requires intensive computation and a relatively high number of parameters to be estimated – thus only models with few variables can be implemented
-



# Concluding Remarks ...

- MIDAS models appear to be more robust to estimation relative to bridge equation models and state-space models, and computationally simpler
  - RECENT INNOVATION: Merge mixed-frequency factor model and MIDAS into Factor- MIDAS –
    - Allows large high-frequency datasets
    - results are promising
-



*THE END*