Mixed Frequency Models

Advanced Time Series Econometrics

Scottish Graduate Programme in Economics

Introduction

- Time series data comes at different frequencies (e.g. quarterly, monthly, daily)
- Most models of interest involve more than one variable (e.g. regression inolves a dependent variable and explanatory variables)
- Conventional time series modelling involves all variables at same frequency
- E.g. a VAR where all dependent variables are at monthly frequency
- Recently great interest in models with variables at different frequencies
- This lead to mixed frequency models
- Reading: Ghysels and Marcellino chapter 13

Introduction

- One of main methods for mixed frequencies is MIDAS (MIxed DAta Sampling)
- Eric Ghysels (http://www.unc.edu/~eghysels/) is pioneer in this field (and co-author of textbook for this course)
- The `midasr` package is a rich tool that allows for estimating MIDAS models in **R**
- In this course, you will not be asked to estimate mixed frequency models in computer labs

Why Mixed Frequencies?

- A lot of data at various frequencies is available
- More data means more information, so we should use it
- Key macro variables (e.g. GDP growth) available quarterly
- Other key macro variables: (e.g. inflation, industrial production) are monthly
- Google trends data is available weekly
- Financial variables (stocks and bonds, exchange rates) available daily
- Regional UK Gross value added (GVA) data available annually

Why Mixed Frequencies?

- Post financial crisis macroeconomists criticized for not including financial sector in their model
- Alessi, Ghysels, Onorante, Peach & Potter (2014) "Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences" Journal of Business & Economic Statistics.
- Andreou, Ghysels & Kourtellos (2013) "Should Macroeconomic Forecasters Use Daily Financial Data and How?" Journal of Business & Economic Statistics

The Importance of Timeliness

- Policymakers want to know key macro variables and what they will be in the future
- Use GDP as an example in this lecture, but can be any key macro variable
- Initial release of 2020Q4 GDP for UK will not be made until April 2021
- We do not even know what it is now
- Growing field of “nowcasting”
- Forecasting what is happening now

Digression: Real Time Data Flow

- Important but no time to cover in this course (so rest of lecture ignores this issue)
- Initial estimate of GDP will be updated in subsequent quarters
- Such “data revisions” may be large and important
- For each observation (e.g. 2015Q4) there will be several “vintages”
- Initial release: available in 2016Q2
- Second vintage: available in 2016Q3
- etc.
- Final vintage: estimate of 2015Q4 GDP available at the present time

Digression: Real Time Data Flow

- Terminology: Real time = forecast using vintage of data which would have been available to forecaster at the time forecast is made
- E.g. when making a forecast in 2015Q4 of 2016Q1 or later, forecaster would have available:
- No value for 2015Q4 GDP
- Initial estimate of 2015Q3 GDP
- Second vintage of 2015Q2 GDP
- etc.
- Final vintage forecasting: forecast using most up to data
- Final vintage: best guess at the true value of GDP
- Real time: information which would have been available at the time
- For some purposes (e.g. DSGE modeling) final vintage is better, for others (e.g. testing how well a forecasting model would have worked if used in practice) real time data is better

The Importance of Timeliness

- Many possible predictors for GDP (quarterly) observed at higher (monthly, weekly, daily) frequencies
- Financial data (e.g. stock prices) available immediately
- “Soft” variables (e.g. surveys) available quickly and at monthly frequency
- E.g. Markit’s Purchasing Manager’s Index
- Consumer confidence measures
- “Hard” variables (e.g. industrial production) available fairly quickly and at monthly frequency
- Policymakers want to update nowcasts/forecasts quickly and want to use this data

The Importance of Timeliness

- Example: Updating forecasts on 1 October, 2008 (just after Lehmann Brothers bankruptcy)
- Everyone knows: financial crisis is happening and something bad is going to happen to GDP growth
- What about the time series econometrician?
- Model 1: AR(4) for GDP growth
- Model 2: Mixed frequency regression containing quarterly GDP growth and daily financial variables

The Importance of Timeliness

- Model 1: Forecasts for 2008Q4 released on 1 October depend on last four quarters of GDP growth (and AR coefficients estimated using long stable period of data)
- Some of these quarters had pretty good GDP growth so 2008Q4 forecast likely will still look good
- Due to release delays 2008Q3 GDP data not yet available
- In practice, AR(4) forecasts will be very slow to adjust and pick up that econonomy entering a recession.

The Importance of Timeliness

- Model 2: Forecasts for 2008Q4 released on 1 October will reflect information in financial variables in late September
- Late September financial variables (esp. various asset markets: credit default swaps, etc.) will reveal that financial crisis has hit
- 1 October forecast will reflect this and show sharp downturn in GDP growth
- Mixed frequency models are much quicker to pick up changes like this

Regression with Mixed Frequency Data

- Y is dependent variable
- Begin with single explanatory variable X
- Notation relating to timing crucial (and different papers use different notation)
- Assume Y is quarterly (e.g. GDP growth)
- $t = 1, .., T$ denotes time at quarterly frequency
- X is daily (e.g. financial variables)
- $n = 1, .., N$ denotes number of days in quarter
- Typically $N = 66$ (22 trading days each month), but quarters can differ by a day or two (ignore this)
- $X_{n,t}$ = observation for day n in quarter t

Regression with Mixed Frequency Data: Bridge Equations

- Simplest thing to do:
- Construct quarterly financial variable out of daily observations
- New variable

$$X_t^q = \frac{X_{N,t} + X_{N-1,t} + .. + X_{1,t}}{N}$$

- Use conventional regression model

$$Y_t = \alpha + \beta X_t^q + \varepsilon_t$$

- Or, when forecasting h periods ahead

$$Y_{t+h} = \alpha + \beta X_t^q + \varepsilon_t$$

- Link from high to low frequency called "bridge", hence bridge equations
- This was old way of treating mixed frequencies
- Problems: misses timeliness issues
- All daily observations weighted equally

Regression with Mixed Frequency Data: MIDAS

- Why not simply run regression putting all the daily observations on right hand side?
- Use forecasting horizon $h = 1$ to illustrate ideas (methods work for any h)

$$Y_{t+1} = \alpha + \beta_1 X_{1,t} + ... + \beta_N X_{N,t} + \varepsilon_t$$

- When N is small this often done (e.g. quarterly/monthly frequency mis-match has $N = 3$)
- This is unrestricted MIDAS (U-MIDAS)
- If N is large run into Fat Data problems
- Proliferation of parameters

MIDAS

- MIDAS uses distributed lag specifications to solve proliferation of parameters problem
- $$Y_{t+1} = \alpha + \beta \sum_{j=0}^{N-1} w_{N-j}(\theta) X_{N-j,t} + \varepsilon_t$$
- $w_n(\theta)$ are weights given to each daily observation (sum to one)
- θ are parameters to be estimated
- Explanatory variable is weighted average of all daily observations in quarter
- Bridge equation has $w_n(\theta) = \frac{1}{N}$ for each day
- MIDAS weights differ and are estimated

MIDAS

- MIDAS uses distributed lag specifications to estimate $w_n(\theta)$ for $n = 1, .., N$
- Terminology: DL-MIDAS
- Many DL forms are possible (see Ghysels and Marcellino chapter 12.3 for more examples)
- E.g. Almon lag, Beta, polynomial specification
- E.g. exponential Almon

$$w_n(\theta) = \frac{\exp(\theta_1 n + \theta_2 n^2)}{\sum_{j=1}^{N} \exp(\theta_1 j + \theta_2 j^2)}$$

- θ_1 and θ_2 are parameters to estimate (only 2 of them so no proliferation of parameters problem)
- Graphs on next slide give examples of implications for weights

Figure 4

Beta Polynomial Weighting Function

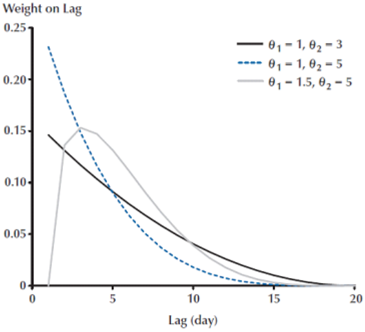
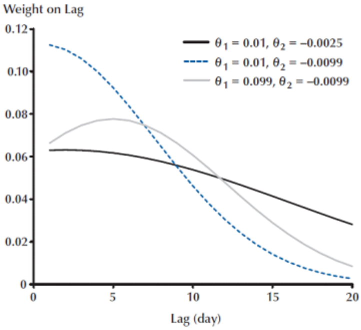


Figure 5

Exponential Almon Polynomial Weighting Function



Generalization of MIDAS models

- DL-MIDAS contains:
- no lagged dependent variables
- precisely one quarter of daily observations as explanatory variables
- Often want to relax these two assumptions
- autoregressive distributed lag MIDAS (ADL-MIDAS):

$$Y_{t+1} = \alpha + \sum_{i=0}^{p_y-1} \rho_{i+1} Y_{t-i} + \beta \sum_{i=0}^{p_x-1} \sum_{j=0}^{N-1} w_{N-j+i*N}(\theta) X_{N-j,t-i} + \varepsilon_t$$

- p_y lags of dependent variable
- p_x lags of weighed average (over quarter) of daily observations

Econometrics of MIDAS Models

- Straightforward methods for estimating distributed lag models
- For U-MIDAS OLS estimation
- For ADL-MIDAS nonlinear least squares can be used
- Information criteria (AIC, BIC) can be used to select p_y, p_x or for choice of weights

Multiplicative MIDAS Models

- ADL-MIDAS model involved a single β (coefficient on weighted average of all daily observations
- Alternative (less parsimonious) is to first create quarterly explanatory variable using DL weighting then put in regression
- Let

$$X_t^Q = \sum_{j=0}^{N-1} w_{N-j}\left(\theta\right) X_{N-j,t}$$

- Then run regression:

$$Y_{t+1} = \alpha + \sum_{i=0}^{p_y-1} \rho_{i+1} Y_{t-i} + \sum_{i=0}^{p_x-1} \beta_{i+1} X_{t-i}^Q + \varepsilon_t$$

- Ghysels calls this ADL-MIDAS-M

MIDAS with Factors

- Also possible to augment the MIDAS model with factors
- E.g. dependent variable: quarterly GDP growth
- daily financial variables
- lots of other quarterly macro variables (e.g. from the FRED data set) used for constructing factors (f_t)
- Factors could be principal components or can be treated as states (see lecture on Factor Models)
-

$$
\begin{aligned}
Y_{t+1} &= \alpha + \sum_{i=0}^{p_y-1} \rho_{i+1} Y_{t-i} + \sum_{i=0}^{p_f-1} \phi_{i+1} f_{t-i} + \\
& \beta \sum_{i=0}^{p_x-1} \sum_{j=0}^{N-1} w_{N-j+i*N}(\theta) X_{N-j,t-i} + \varepsilon_t
\end{aligned}
$$

MIDAS with Factors

- Or can construct factors out of high frequency variables
- E.g. dependent variable: quarterly GDP growth
- hundreds of daily financial variables (e.g. stocks, bonds, derivatives, exchange rates, commodity prices, etc. etc.)
- $f_{n,t}$ = factor for day n in quarter t

$$Y_{t+1} = \alpha + \sum_{i=0}^{p_y - 1} \rho_{i+1} Y_{t-i} + \beta \sum_{i=0}^{p_x - 1} \sum_{j=0}^{N-1} w_{N-j+i*N}(\theta) f_{N-j,t-i} + \varepsilon_t$$

MIDAS with Leads

- With MIDAS can constantly be updating nowcasts/forecasts
- E.g. every day release a new nowcast of 2019Q1 GDP growth as new data comes in
- This is done using MIDAS with leads

$$
\begin{aligned}
Y_{t+1} &= \alpha + \sum_{i=0}^{p_y-1} \rho_{i+1} Y_{t-i} + \\
&\quad \beta[\sum_{j=0}^{L-1} w_{L-j}(\theta) X_{N-j,t+1} + \\
&\quad \sum_{i=0}^{p_x-1} \sum_{j=0}^{N-1} w_{N-j+i*N}(\theta) X_{N-j,t-i}] + \varepsilon_t
\end{aligned}
$$

- Note: this is for standard ADL-MIDAS, can also have version for multiplicative MIDAS or MIDAS with factors

MIDAS with Leads

- Look carefully at the new term

$$\sum_{j=0}^{L-1} w_{L-j}(\theta) X_{N-j,t+1}$$

- L is number of leads (note $t+1$ subscript)
- E.g. $L = 22$ at the end of January, 2019 for nowcasts of 2019Q1
- Daily observations on financial variables for January included as exp. vars ($X_{N-j,t+1}$)
- $w_{L-j}(\theta)$ is weight function (usually distributed lag)
- As each trading day goes by, L increases by 1 and new nowcasts using updated info produced

Other Extensions of MIDAS

- I will not cover many other extensions of MIDAS such as:
- Markov switching MIDAS
- Threshold MIDAS
- See Chapter 12.3.5 of Ghysels and Marcellino textbook

Empirical Illustration

- The textbook by Ghysels and Marcellino (Chapter 12.6) has empirical example on "Nowcasting US GDP Growth"
- Example shows how MIDAS used for nowcasting, forecasting and predicting recessions
- I encourage you to read it to get good understanding of MIDAS methods
- But too long (20 textbook pages) to cover in this lecture
- Instead I will take an example from Armesto, Engemann and Owyang "Forecasting with Mixed Frequencies" published in Federal Reserve Bank of St. Louis Review

Empirical Illustration

- Forecasting (quarterly) US GDP growth using (monthly) employment growth data
- See Armesto et al paper for details about data and specification (e.g. lag length) choices
- Armesto et al has more variables and more models and forecasting horizons
- Here I will focus just on MIDAS and Bridge Equation and $h = 1$
- They use the exponential Almon weighting function in MIDAS

Digression: Evaluating Forecasts

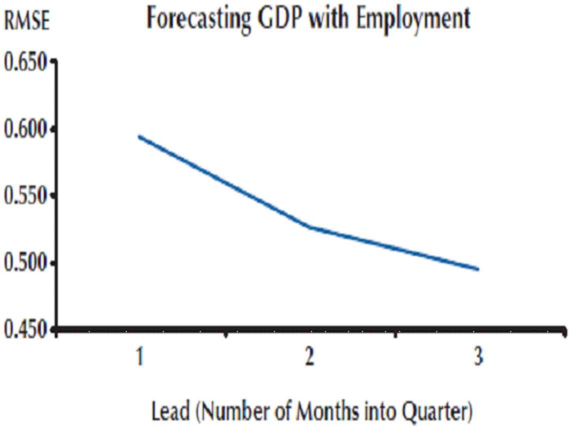
- Root mean squared forecast error (RMSFE) is common way of measuring forecast performance
- Choose forecast evaluation period (e.g. $\tau_0, .., T$)
- Produce forecasts of Y_τ using data available at time $\tau - 1$, call them $\widehat{Y}_\tau^F$ for $\tau = \tau_0, .., T$
- $$RMSFE = \sqrt{\frac{\sum_{\tau=\tau_0}^{T} \left(Y_t - \widehat{Y}_\tau^F\right)^2}{T - \tau_0 + 1}}$$
- Model with lower RMSFE is better

Empirical Illustration

- With bridge equation, we only have one forecast per quarter (at end of quarter)
- X_t^q is calculated for each quarter (once 3 months of quarter are complete)
- With MIDAS we have 3 forecasts per quarter (one at the end of each month)
- But if we compare the end of quarter forecasts of MIDAS with bridge equation we obtain:
- RMSFE for MIDAS: 0.5669
- RMSFE for bridge equation: 0.5972
- MIDAS forecasting a bit better than bridge equation

Empirical Illustration

- Advantage with MIDAS is that we can update our forecasts of GDP growth as each new month of employment data is released
- Figure on next slide shows:
- RMSFE for forecasts made in first month of each quarter (i.e. using data through end of previous quarter)
- RMSFE for forecasts made in second month (also including employment data for first month of each quarter)
- RMSFE for forecasts made in third month (also including employment data for second month of each quarter)
- RMSFE is dropping as new data on employment growth is included


RMSE
Forecasting GDP with Employment
0.650
0.600
0.550
0.500
0.450
1
2
3
Lead (Number of Months into Quarter)

MIDAS with More than One Explanatory Variable

- Preceding material describes MIDAS with one high frequency explanatory variable
- What if you have several?
- Add more weighted averages of them into the MIDAS regression
- E.g. if $X_{n,t}$ and $Z_{n,t}$ are two high frequency variables:

$$Y_{t+1} = \alpha + \beta_1 \sum_{j=0}^{N-1} w_{N-j}^{X}(\theta) X_{N-j,t} + \beta_2 \sum_{j=0}^{N-1} w_{N-j}^{Z}(\theta) Z_{N-j,t} + \varepsilon_t$$

- $w_n^X(\theta)$ are DL weights for X
- $w_n^Z(\theta)$ are DL weights for Z
- But a common alternative is to work with 2 MIDAS models and then average results

Model Averaging with MIDAS

- Before (single model) we had forecasts $\widehat{Y}_{\tau}^{F}$ for $\tau = \tau_0, .., T$
- Now suppose $m = 1, .., M$ different MIDAS models
- E.g. M different financial variables to forecast GDP growth
- Each individual model parsimonious
- $\widehat{Y}_{m,\tau}^{F}$ is forecast of model m
- Overall forecast:

$$\widehat{Y}_{\tau}^{F} = \sum_{m=1}^{M} v_{m,\tau} \widehat{Y}_{m,\tau}^{F}$$

- $v_{m,\tau}$ is weight attached to model m at time τ
- $\sum_{m=1}^{M} v_{m,t} = 1$

Model Averaging with MIDAS

- How do we calculate weights? Several ways
- Ghysel's Matlab code has:
- Equal weights:

$$v_{m,\tau} = \frac{1}{M}$$

- BIC weights: each model receives weight proportional to exp of BIC
- MSFE weights: each model receives weight inversely proportional to MSFE
- MSFE = mean squared forecast error
- Discounted MSFE weights: as preceding but with more recent forecast performance getting more weight
- Weights can change over time
- Which models get most weight?
- Those which have forecast best in the past

Mixed Frequency Models as State Space Models

- MIDAS is most popular method for dealing with mixed frequencies
- State space methods are also sometimes used
- More details will be given below in relation to mixed frequency VARs, but the general strategy is:
- Assume all variables are at the higher frequency
- E.g. Y_t is GDP for day t
- X_t is a financial variable for day t
- Then build a conventional time series model (e.g. regression or VAR)

Mixed Frequency Models as State Space Models

- Trouble: daily GDP is not observed, only quarterly
- Y_t is only observed once every 66 days
- Rest of days missing value for Y_t
- Treat missing value as unobserved states
- Estimate using state space model methods (e.g. Kalman filter)
- Mixed frequency problem is turned into a missing value problem
- State space methods often used for missing value problems of various sorts

Mixed Frequency Models as State Space Models

- MIDAS and state space approach are closely related
- Ghysels' research has shown:
- MIDAS regression can also be viewed as a reduced form representation of the state space model approach
- In some cases the MIDAS regression is an exact representation of the Kalman fillter (in other approximation errors are small)
- MIDAS has some advantages
- MIDAS does not require setting out a full state space system of equations
- Writing out full state system usually more prone to specification error and less parsimonious
- MIDAS computationally easier (nonlinear least squares versus Kalman filtering/state smoothing)

Mixed Frequency VARs

- VARs were covered in Econometrics 2
- One of the most popular tools of modern macroeconomics and finance
- VAR(p) can be written as

$$y_t = \sum_{j=1}^{p} \Theta_j y_{t-j} + \varepsilon_t$$

- y_t is a vector of M variables (e.g. inflation, unemployment, industrial production)
- ε_t is i.i.d. $N(0, \Omega)$
- Ω is $M \times M$ covariance matrix

Mixed Frequency VARs

- To do impulse response analysis, usual work with structural VAR

$$Ay_t = \sum_{j=1}^{p} B_j y_{t-j} + u_t$$

- u_t is i.i.d. $N(0, I)$
- From Econometrics 2 you will know how to estimate VARs
- Econometric estimation of mixed frequency VARs builds on methods you will know from previous study

Mixed Frequency VARs

- Several mixed frequency VARs exist
- I will discuss two main approaches
- Stacked VAR (Ghysels and Marcellino call this the "Observation Driven Approach")
- State space VAR (G+M call this the "Parameter Driven Approach")

The Stacked VAR

- Illustrate assuming one variable is monthly (X), the other quarterly (Z)
- $t = 1, .., T$ are quarters
- In any quarter, we have 3 monthly values of X:
- $X_{1,t}$,$X_{2,t}$, $X_{3,t}$
- The stacked VAR is a VAR with the vector of dependent variables being:
- $y_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \\ Z_t \end{pmatrix}$

The Stacked VAR

- In any quarter, month 1 occurs before month 2 which occurs before month 3
- It is common to work with a structural version of the stacked VAR which imposes this information
- $$A = \begin{bmatrix} a_{1,1} & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} & 0 \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}$$
- Structural VARs covered in Econometrics 2 so no new econometric issues

The Stacked VAR

- In one sense, stacked VARs are easy – they are just VARs
- But they can easily become Big VARs
- With one monthly and one quarterly variable, stacked VAR has 4 variables ($M = 4$)
- With two monthly and three quarterly variables, $M = 9$
- With one daily and one quarterly variable, $M = 67$
- etc..
- Bayesian methods often used with such Big VARs

The State Space VAR

- Illustrate assuming one variable is monthly (X), the other quarterly (Z)
- NOW I AM CHANGING TIMING NOTATION
- $t = 1, .., T$ are MONTHS
- X_t is known data
- Z_t is NOT known data
- Suppose we want to work with a VAR with dependent variables:

$$y_t = \begin{pmatrix} X_t \\ Z_t \end{pmatrix}$$

- Problem: Z_t is not known
- Solution: treat Z_t as states in a state space model

The State Space VAR

- Exactly what is known about Z_t depends on whether it is in (log) levels (e.g. GDP) or growth rates (e.g. GDP growth)
- Here assume (log) levels and the data is:
- Z_t^Q is total GDP produced over the quarter (i.e. over the last 3 months)
- At the end of any quarter, we will observe

$$Z_t^Q = Z_t + Z_{t-1} + Z_{t-1}$$

- When data is logged this relationship is approximate
- Can write as

$$Z_t^Q = \Lambda \begin{pmatrix} Z_t \\ Z_{t-1} \\ Z_{t-2} \end{pmatrix}$$

  where $\Lambda = (1, 1, 1)$
- Different data transformations (e.g. growth rates) and frequencies of data imply Λ has different form

The State Space VAR

- Z_t^Q is a quarterly variable so will be observed in March, June, September and December
- In other months, Z_t^Q is not observed
- Stats agency does tell us GDP for 1st quarter (January, February and March)
- Stats agency does NOT tell us GDP for other three monthly periods (February, March and April)
- Define M_t to capture this pattern
- $M_t = 1$ when t is at end of quarter (March, June, Sept, December)
- M_t is an empty matrix for all other months

The State Space VAR

- Combining the previous two slides we have the relationship:

$$Z_t^Q = \Lambda M_t \begin{pmatrix} Z_t \\ Z_{t-1} \\ Z_{t-2} \end{pmatrix}$$

- This is measurement equation in state space model capturing following ideas:
- Z_t, Z_{t-1} and Z_{t-2} are unknown states
- At the end of every quarter we observe their sum
- In all other months we do not observe anything
- ΛM_t is a system matrix (see page 55 of lecture slides on state space models)
- Measurement equation does not have any errors but this is fine (see page 53 of lecture slides on state space models and set $\Sigma_t = 0$)

State Space VAR: Summary

- State space methods can be used to create a mixed frequency VAR
- State equation is just a VAR

$$y_t = \sum_{j=1}^{p} \Theta_j y_{t-j} + \varepsilon_t$$

- But $y_t = \begin{pmatrix} X_t \\ Z_t \end{pmatrix}$ and Z_t are unobserved states
- Measurement equation captures relationship between Z_t and observed data (Z_t^Q)
- Standard econometric methods for state space models can be used

Summary

- Macroeconomists/financial researchers increasingly want to work with mixed frequencies
- This lecture covered mixed frequency regressions and mixed frequency vars
- These can be parameter-rich (Fat Data)
- MIDAS is a parsimonious (and computationally simple) way of working with mixed frequency data
- Stacked VAR is a simple way of building a VAR with mixed frequency data
- State space methods also be built into a VAR to handle mixed frequency data