THE BLESSING OF DIMENSIONALITY IN FORECASTING REAL HOUSE PRICE GROWTH IN THE NINE CENSUS DIVISIONS OF THE US

Forthcoming in *Journal of Housing Research*

Sonali Das[1], Rangan Gupta[2] and Alain Kabundi[3]

[1]CSIR, Pretoria
[2]University of Pretoria, Pretoria
[3]University of Johannesburg, Johannesburg

7^{th} Triennial, Kolkata, 2009

Outline

Main question

Is the wealth of information contained in 126 monthly series more useful in forecasting regional real house price growth rate for the 9 census divisions of the US?

- Why is forecasting real house price growth important?
  - Asset prices help to forecast both inflation and output
  - Models that forecast real house price inflation can be indicative of overall inflation
  - (Forni et al., 2003; Stock and Watson, 2003; Gupta and Das, 2008a,b; Das *et al.*, 2008a, b)
- Why use large scale models?
  - Large number of economic indicators help in predicting real house price growth
  - (Cho, 1996; Abraham and Hendershott, 1996; Rapach and Strauss, 2007, 2008)
- Why use regional data?
  - Segmented nature of the housing market
  - Non-uniform economic conditions across regions
  - (Carlino and DeFina, 1998, 1999; Burger and van Rensburg, 2008; Gupta and Das 2008a; Vargas-Silva 2008a, b)

Our previous work

Forecasting Small scale models, Large scale models (DFM)

Turning Point Predicting downturns

Spatial Models Effect of spatial neighbors

Bayesian Models

Outline

Unrestricted classical Vector Autoregressive (VAR) model (*Sims, 1980*)

Model

- $y_t = C + A(L)y_t + \epsilon_t$, where $A(L)$ is a $n \times n$ polynomial matrix in the backshift operator L with lag p
- *atheoretical* though particularly useful for forecasting
- Equal lag length for all variables
- Many parameters to estimate
- Possible large out-of-sample forecast errors

Solution

- Exclude insignificant lags
- Specify unequal number of lags for different equations

Bayesian Vector Autoregressive (BVAR)

- An alternative to overcome overparameterization

  (Litterman (1981, 1986), Doan *et.al* (1984), Todd (1984)and Spencer (1993))
- Instead of eliminating longer lags, impose restrictions on these coeff. $\rightarrow$ more likely to be near zero than the coeff. on shorter lags
- If strong effects exist from less important variables, data can override this assumption
- Restrictions: normal prior distributions with zero means and small standard deviations for all coeff.One Exception: coeff. on the first own lag of a variable has a mean of 1
- Popularly referred to as the *Minnesota Prior*

Minnesota Prior

Notationally,

- $\beta_i \sim N(1, \sigma^2_{\beta_i})$ and $\beta_j \sim N(0, \sigma^2_{\beta_j})$
- To circumvent overparametrization, *Doan et al (1984)* suggest a formula to generate standard deviations (*s.d.*) as a function of a few hyperparameters: w, d and a weighting matrix $f(i, j)$
- *s.d.* of variable j in equation i and lag m is given as

$$\sigma_{ijm} = \{w \times m^{-d} \times f(i, j)\} \times \frac{\hat{\sigma}_j}{\hat{\sigma}_i}$$

$\hat{\sigma}_i$: Standard error of univariate autoregression for variable i

w: Overall tightness

d: Decay factor

$f(i, j)$: Tightness of variable j in equation i relative to variable i

$f(i, j) = 1$ if $i = j$ and k_{ij} otherwise; $(0 \leq k_{ij} \leq 1)$, and; $d > 0$.

Dynamic Factor Model

- Can cope with many variables without running into degrees of freedom problems
- Common factors can affect variables not only contemporaneously, but also with lags
- Advantage over VAR models where one chooses variables

Dynamic Factor Model

- Each time series represented as sum of two latent components: common (capturing multivariate correlation) and idiosyncratic
- Let $X_t = (x_{1t}, ..., x_{nt})'$ be a standardized stationary process
- In terms of a DFM, X_t can be written as:

$$X_t = B(L)f_t + \xi_t \quad (1)$$
$$= \Delta F_t + \xi_t \quad (2)$$

- $B(L) = B_0 + B_1 L + \ldots + BsL^s$ a $n \times q$ matrix of factor loadings of order s
- f_t is a $q \times 1$ vector of dynamic components
- F_t is the $r = q(s+1) \times 1$ vector of static factors, i.e., $F_t = (f_t', f_{t-1}', \ldots, f_{t-s}')$
- Δ is the matrix of factor loadings

Dynamic Factor Model

Determination of number of dynamic factors

- Informal criteria based on prop. of variance explained

  (Bai and Ng, 2005; Stock and Watson, 2005)

- Principle component (Forni et al, 2004)

Estimation of dynamic factors - Frequency domain

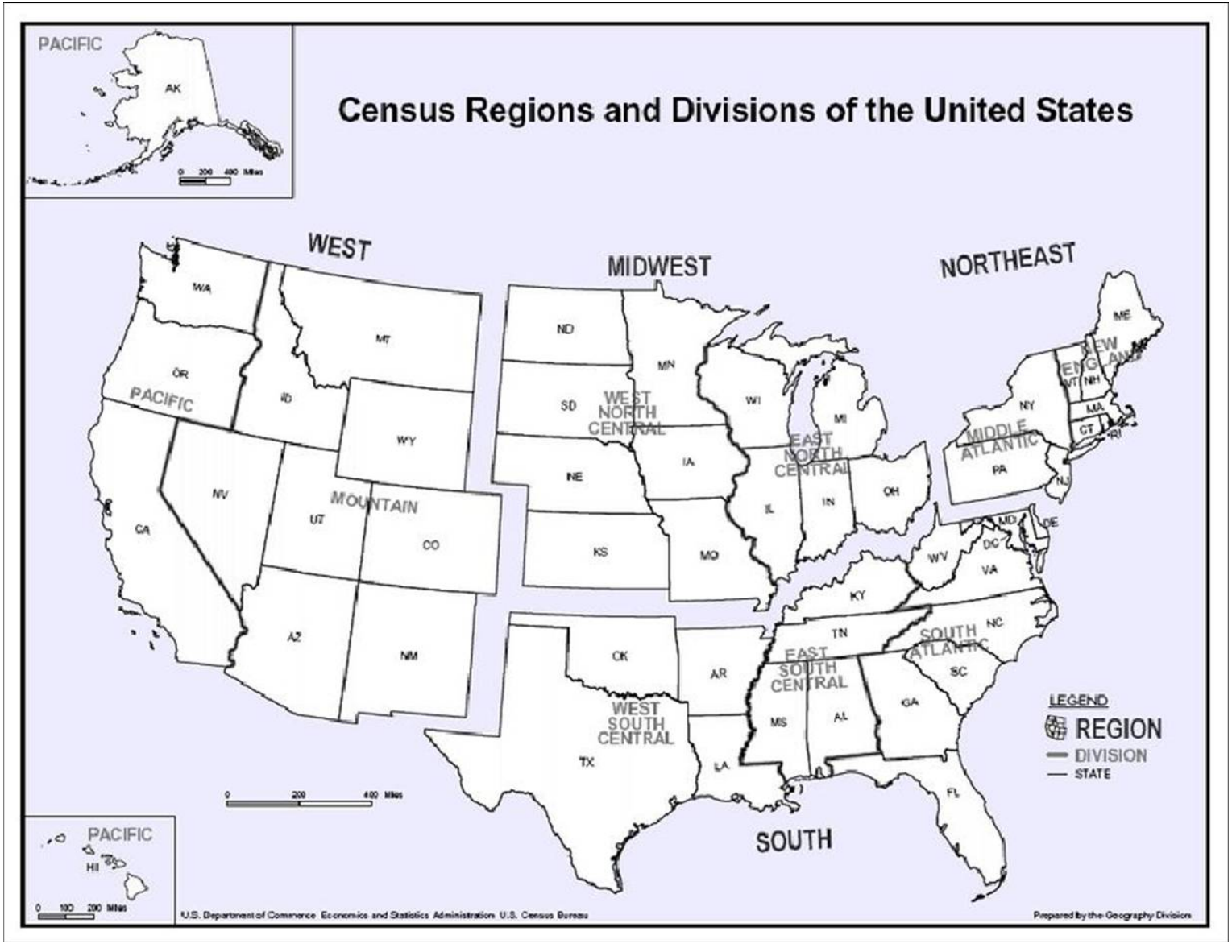
(Forni et al, 2000, 2002)

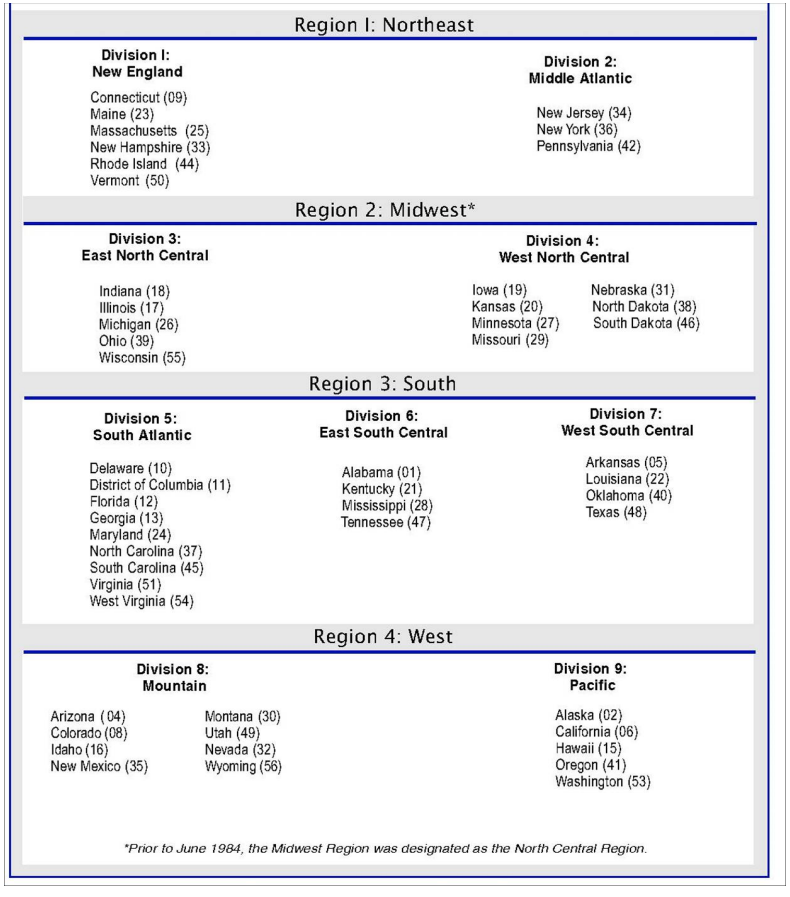
- Spectral density matrix of X_t decomposed as $\Sigma(\theta) = \Sigma_\chi(\theta) + \Sigma_\xi(\theta),\ -\pi < \theta < \pi$
  - Rank of $\Sigma_\chi(\theta) = q$, is the number of dynamic factors
- Covariance of X_t can be decomposed as $\Gamma_k = \Gamma_k^\chi + \Gamma_k^\xi$
  - Rank of $\Gamma_k^\chi = r$, the number of static factors

DFM specifications considered

- UFAVAR: includes one of the variables of interest and the obtained number of common static factors;
- MFAVAR: includes all the nine real house price growth rates and the common static factors;
- UBFAVAR: uses one of the variables of interest and the common static factors, and which, in turn, are estimated based on Bayesian restrictions discussed in the previous subsection;
- MBFAVAR: with a specification similar to the MFAVAR, except that the current model applies Bayesian restrictions on lag of the variables based on the Minnesota prior.

Outline


Census Regions and Divisions of the United States
WEST
MIDWEST
NORTHEAST
SOUTH
PACIFIC
MOUNTAIN
WEST NORTH CENTRAL
EAST NORTH CENTRAL
MIDDLE ATLANTIC
NEW ENG LAND
WEST SOUTH CENTRAL
EAST SOUTH CENTRAL
SOUTH ATLANTIC
LEGEND
REGION
DIVISION
STATE
U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau
Prepared by the Geography Division



Region I: Northeast

Division I: New England

Connecticut (09)
Maine (23)
Massachusetts (25)
New Hampshire (33)
Rhode Island (44)
Vermont (50)

Division 2: Middle Atlantic

New Jersey (34)
New York (36)
Pennsylvania (42)

Region 2: Midwest*

Division 3: East North Central

Indiana (18)
Illinois (17)
Michigan (26)
Ohio (39)
Wisconsin (55)

Division 4: West North Central

Iowa (19)
Kansas (20)
Minnesota (27)
Missouri (29)
Nebraska (31)
North Dakota (38)
South Dakota (46)

Region 3: South

Division 5: South Atlantic

Delaware (10)
District of Columbia (11)
Florida (12)
Georgia (13)
Maryland (24)
North Carolina (37)
South Carolina (45)
Virginia (51)
West Virginia (54)

Division 6: East South Central

Alabama (01)
Kentucky (21)
Mississippi (28)
Tennessee (47)

Division 7: West South Central

Arkansas (05)
Louisiana (22)
Oklahoma (40)
Texas (48)

Region 4: West

Division 8: Mountain

Arizona (04)
Colorado (08)
Idaho (16)
New Mexico (35)
Montana (30)
Utah (49)
Nevada (32)
Wyoming (56)

Division 9: Pacific

Alaska (02)
California (06)
Hawaii (15)
Oregon (41)
Washington (53)

*Prior to June 1984, the Midwest Region was designated as the North Central Region.

- The bureau recognizes four census regions
- Further organizes them into nine divisions
- For the presentation of data
- *Not to be construed as necessarily being grouped owing to any geographical, historical, or cultural bonds.*

- In-sample analyses period: 1991(02) to 2000(12)
- Out-of-sample forecast evaluation period: 2001(01) to 2005(06)
- Out-of-sample forecast is done for one to twelve months ahead
- Choice of 2001:01 as the onset of forecast horizon motivated from Iacoviello and Neri (2008)

- Small-scale VARs, both the classical and Bayesian variants
  – Only the nine variables of interest, namely, real house price growth rates of the nine census divisions of the US
- Large-scale BVARs and the DFM:
  – Based on 126 monthly series (9 + 1 + 116).

The nominal house price figures for these nine US census divisions and for the whole of US were obtained from the Office of Federal Housing Enterprize Oversight (OFEO), and were converted to their real counterpart by dividing them with the personal consumption expenditure deflator.

Remaining 116 variables from dataset of Boivin et al. (2008) which contains a broad range of macroeconomic variables such as industrial production, income, employment and unemployment, housing starts, inventories and orders, stock prices, exchange rates, interest rates, money aggregates, consumer prices, producer prices, earnings, and consumption expenditure

Balanced panel of 126 monthly series from 1991:02-2005:06

All data transformed to induce stationarity

Outline

'*optimal*' model based on minimum average RMSE for real house price growth rate

'*optimal*' model based on minimum average RMSE for real house price growth rate

- For all the 9 census divisions, some large model outperforms small-scale models

  UVFAVAR for East South Central and West South Central, UVBFAVAR (w=0.1, d=2) for East North Central and Mountain, LBVAR (w=0.1, d=1) for Middle Atlantic, MVBFAVAR (w=0.2, d=2) for New England, MVBFAVAR (w=0.1, d=2) for South Atlantic and West North Central, and MVFAVAR for Pacific

- Even the second-best performing model are large-scale in nature

  barring West South Central for which the SBVAR (w=0.1, d=1) comes second

- There always exists at least one small-scale model that outperforms the large-scale BVAR

  exceptions being Middle Atlantic under w=0.3, d=0.5, w= 0.2, d=1, w=0.1, d=1 and w=0.1, d=2

Outline

FAVARs, in their various forms, are standout performers

Data rich environment of the FAVAR models that include a wide range of macroeconomic series of the US economy, besides the house price growth rates of the census divisions, more informative when forecasting the house price growth rate of the nine census regions

Role of fundamentals in affecting the housing market cannot be underestimated

First - detailed look at the factors to determine the dominant macroeconomic variables that comprise these factors;

Second - incorporate role of the house price growth rate of the neighboring division(s) in the forecasting process of a particular census division, by developing spatial versions of the above models.