University of Pretoria
Department of Economics Working Paper Series

Is a DFM Well-Suited For Forecasting Regional House Price Inflation?

Sonali Das
CSIR
Rangan Gupta
University of Pretoria
Alain Kabundi
University of Johannseburg
Working Paper: 2008-14
June 2008

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413
Fax: +27 12 362 5207

Is a DFM Well-Suited for Forecasting Regional House Price Inflation?

Sonali Das*, Rangan Gupta** and Alain Kabundi#

Abstract

This paper uses the Dynamic Factor Model (DFM) framework, which accommodates a large cross-section of macroeconomic time series for forecasting regional house price inflation. As a case study, we use data on house price inflation for five metropolitan areas of South Africa. The DFM used in this study contains 282 quarterly series observed over the period 1980Q1-2006Q4. The results, based on the Mean Absolute Errors of one- to four-quarters-ahead out of sample forecasts over the period of 2001Q1 to 2006Q4, indicate that, in majority of the cases, the DFM outperforms the VARs, both classical and Bayesian, with the latter incorporating both spatial and non-spatial models. Our results, thus, indicate the blessing of dimensionality.

Journal of Economic Literature Classification: C11, C13, C33, C53.

Keywords: Dynamic Factor Model, VAR, BVAR, Forecast Accuracy.

*Senior Researcher, Logistics and Quantitative Methods, CSIR Built Environment, P.O. Box: 395, Pretoria, 0001, South Africa. Phone: +27 12 841 3713, Fax: +27 12 841 3037, Email: SDas@csir.co.za.

**Associate Professor, University of Pretoria, Department of Economics, Pretoria, 0002, South Africa, Email: Rangan.Gupta@up.ac.za. Phone: +27 12 420 3460, Fax: +27 12 362 5207.

#To whom correspondence should be addressed. Senior Lecturer, University of Johannesburg, Department of Economics, Johannesburg, 2006, South Africa, Email: akabundi@uj.ac.za. Phone: +27 11 559 2061, Fax: +27 11 559 3039.

1. Introduction

This paper investigates whether the wealth of information contained in the Dynamic Factor Model (DFM) framework, developed by Forni et al. (2005), can be useful in forecasting regional house price inflation. To illustrate, we use the DFM to predict house price inflation in five metropolitan areas of South Africa, namely, Cape Town, Durban, Johannesburg, Port Elizabeth and Pretoria, using quarterly data over the period of 1980Q1-2006Q4. The panel data comprises 282 quarterly series for the South African economy, a set of global variables such as commodity industrial inputs price index and crude oil prices, and time series of major trading partners , namely, Germany, the United Kingdom (UK), and the United States (US) of America.. The forecast performance of DFM is evaluated in terms of the Mean Absolute Errors (MAEs), by comparing it with Spatial Bayesian Vector Autoregressive (SBVAR) models, based on the First-Order Spatial Contiguity (FOSC) and the Random Walk Averaging (RWA) priors,. The performace is also compared to non-spatial models like the unrestricted classical Vector Autoregressive (VAR) model and Bayesian Vector Autoregressive (BVAR) models with the Minnesota prior which are estimated based on only the house price inflation .

The motivation for investigating regional house price inflation is to answer two related questions : First, why is forecasting house price growth an important exercise? And, second, why look at regional data for this purpose? The importance of predicting house price inflation is motivated by recent studies that conclud that asset prices help forecast both inflation and output (Forni, Hallin, Lippi, and Reichlin, 2003; Stock and Watson, 2003). Since large amount of individual wealth are imbedded in houses, similar to other asset prices, houses are thus important in signaling inflation as well.[1] As such, models that forecast house price inflation can give policy makers an idea about the direction of CPI inflation in the future, and\ hence, can provide a better control for designing of appropriate policies. Secondly, we use regional data on house prices as in this paper we are trying to investigate whether the rich information environment of the DFM can be used to improve the prediction of house price inflation, relative to other standard spatial models of forecasting besides the non-spatial VAR and BVARs,. More importantly, the need to use regional data is simply to account for possible heterogeneity and segmentation that might exist in the housing market. Herein then also comes the justification of modeling house-prices separately based on size of house.[2]

At this juncture, we elaborate on why we use the South African housing market as a case study. The reasons are twofold: Firstly, of the easy access and availability of such regional data, and; Secondly, and perhaps more importantly, the choice being driven by the existence of a recent study by Gupta and Das (2008) on forecasting house price inflation in the metropolitan areas of South Africa based on SBVARs. In this paper, we thus aim to compare the performance of our benchmark model, with spatial models, similar to those used by Gupta and Das (2008). Both the paper of Gupta and Das (2008), as well as the paper by Burger and van Rensburg (2007), provide an idea about the segmented nature of the housing market in South Africa. Gupta and Das (2008) estimated SBVARs, based on the FOSC and the RWA priors, for six metropolitan areas of South Africa, using monthly data over the period of 1993:07 to 2005:06, and then forecasted one- to six-months-ahead house prices over the forecast horizon of 2005:07 to 2007:06. It must be noted, that unlike Gupta and Das (2008), we use quarterly data on all variables, including house prices to compute house price inflation. And further, using a DFM also allows us to incorporate a wide variety of variables that can possibly affect the housing market, unlike that of the VAR, BVARs and SBVARs, which were estimated based on data on house price inflation only. It must be emphasized, though, that our analysis, eithe using the DFM, or the spatial and non-spatial models, is a general one, and these techniques can be used to forecasts regional variable(s) of any economy. In our case, it happens to be the South African economy due to the availability of relevant data and the pre-existing of similar studies.

The rationale for a DFM to forecast house prices inflation emanates from the fact that a large number of economic variables help in predicting housing price growth (Cho, 1996; Abraham and Hendershott, 1996; Johnes and Hyclak, 1999; and Rapach and Strauss, 2007). For instance, income, interest rates, construction costs, labor market variables, stock prices, industrial production, and consumer confidence index – which are included in the DFM,, are potential predictors. In addition, given that movements in the housing market are likely to play an important role in the business cycle, not only because housing investment is a very volatile component of demand (Bernanke and Gertler, 1995), but also because changes in house prices tends to have important wealth effects on consumption

[1] Gupta and Das (2007) point out that, in South Africa, housing inflation and CPI inflation tend to move together, though the former, understandably, is more volatile.

[2]See Kang and Stulz (1997), Choe et al. (1999), Dahlquist et al. (2003), Christoffersen et al. (2006), Burger and van Rensburg (2007) and Gupta and Das (2008) for further details.

(International Monetary Fund, 2000) and investment (Topel and Rosen, 1988), the importance of forecasting house price inflation is vital. The housing sector thus plays a significant role in acting as leading indicator of the real sector of the economy, and as such, predicting it correctly cannot be overemphasized, especially in the light of the recent credit crunch in the U.S. that started with the burst of the housing price bubble which, in turn, transmitted to the real sector of the economy driving it towards an imminent recession.

Note, in a DFM, each time series in the panel is represented as the sum of two latent components: a common component which captures most of the multivariate correlation, and an idiosyncratic component which is poorly cross-sectionally correlated. The rational behind factor analysis is that common components are driven by a few common shocks, and as such, the low dimensionality implies that common components can be consistently estimated and forecasted on the basis of few factors only. The estimation process involves the construction of aggregates of variables that capture relevant information in the cross-section, since the idiosyncratic components, which are poorly correlated vanish by the law of large number. There are several empirical researches that provide evidence of improvement in forecasting performance of macroeconomic variables using factor analysis (Gupta and Kabundi, 2008; Giannone and Matheson, 2007; Van Nieuwenhuyze, 2007; Cristadoro et al., 2005; Forni et al., 2005; Schneider and Spitzer, 2004; Kabundi, 2004; Forni et al., 2001; and Stock and Watson, 2002a, 2002b, 1999, 1991, and 1989). But, to the best of our knowledge, this is the first attempt to compare the forecasting performances of a full-fledged DFM with spatial and non-spatial econometric models in terms of predicting regional house price inflation. We must, however, point out of two related study by Rapach and Strauss (2007a, 2007b).[3] In the first paper, the authors used an autoregressive distributed lag (ARDL) model framework, containing 25 determinants, to forecast real housing price growth for the individual states of the Federal Reserve's Eighth District. Given the difficulty in determining a priori the particular variables that are the most important for forecasting real housing price growth, the authors also use various methods to combine the individual ARDL model forecasts, that resulted in better forecasts of real housing price growth. While, Rapach and Strauss (2007b) looks at doing the same for 20 largest US states based on ARDL models containing large number of potential predictors, including state-level, regional and national level variables. Again, the authors reach similar conclusions as far as the importance of combining forecasts are concerned. Given that, in practice forecasters and policymakers often extract information from many series than the ones included in smaller models like the ones used by Rapach and Strauss (2007a, 2007b), the role of a large-scale DFM cannot be ignored. In addition, one should not condone the fact that the main problem of small models, as seen above from the studies by Rapach and Strauss (2007a, 2007b), is in the decision regarding the choice of correct potential predictors to be included.

Finally, we outline a few facts about the South African house price data. Burger and van Rensburg (2007) show that products sold at different regions can only be comparable when a clear definition of the product is provided at the outset. Thus, as in Burger and van Rensburg (2007) and Gupta and Das (2008), we do not consider the residential market in general, rather we subdivide the market in terms of sizes and prices of the houses. Specifically, we use the ABSA[4] Housing Price Survey, which distinguishes between three price categories as --- luxury houses (R 2.6 million to R9.5 million), middle-segment houses (R226,000 to R2.6 million) and affordable houses (R226,000 and below with an area in the range of 40 m^2 -79 m^2); and further subdivides the middle segment category based on the square meters of house area into small (80 m^2 -140 m^2), medium (141 m^2 -220 m^2) and large (221 m^2 -400m^2). Given that regional house price data is only available for middle-segment houses, we restrict our study to this category. Also, though the ABSA Housing Price Review reports data for both metropolitan and non-metropolitan areas, the availability is limited and also lacks clarity regarding the area of coverage, especially for the rural areas. We thus limit our analysis to the five major metropolitan areas of South Africa.

The remainder of the paper is organized as follows: In section 2 we lay out the DFM, Section 3 discusses the data used to estimate the DFM. Section 4 outlines the basics of the VAR and Minnesota-type BVARs, and SBVARs based on the FOSC and the RWA priors, and Section 5 presents the results from the forecasting exercise. Finally, section 6 concludes and lays out the areas of further research.

2. The Model

This study uses the Dynamic Factor Model (DFM) developed by Forni et al. (2005) to extract common components between macroeconomics series, which are then used to forecast metropolitan house price

[3] See Dua and Smyth (1995) and Dua et al. (1999) for papers that use Bayesian methods to forecast home sales.
[4]ABSA is one of the Leading Private banks of South Africa.

inflation for the South African housing market. In the VAR models, since all variables are used in forecasting, the number of parameters to be estimated depends on the number of variables n. With such a large information set, n, the estimation of a large number of parameters leads to a curse of dimensionality. The DFM uses information set accounted by few factors $q << n$, which transforms the curse of dimensionality into a blessing of dimensionality. The DFM expresses individual times series as the sum of two unobserved components: a common component driven by a small number of common factors and an idiosyncratic component, which are specific to each variable. Forni et al. (2005) demonstrated that when the number of factors is small relative to the number of variables and the panel is heterogeneous, the factors can be recovered from the present and past observations.

Consider an $n \times 1$ covariance stationary process $Y_t = (y_{1t},...,y_{nt})'$. Suppose that X_t is the standardized version of Y_t, i.e. X_t has a mean zero and a variance equal to one. Under the DFM proposed by Forni et al. (2005), X_t is described by a factor model, and can be written as the sum of two orthogonal components:

$$x_{it} = b_i(L) f_t + \xi_{it} = \lambda_i F_t + \xi_{it} \quad (1)$$

or, in vector notation:

$$X_t = B(L) f_t + \xi_{it} = \Lambda F_t + \xi_{it} \quad (2)$$

where f_t is a $q \times 1$ vector of dynamic factors, $B(L) = B_0 + B_1 L + ... + B_s L^s$ is in an $n \times q$ matrix of factor loadings of order s, ξ_t is the $n \times 1$ vector of idiosyncratic components, F_t is $r \times 1$ vector of static factors, with $r = q(s+1)$. However, in more general framework we use $r \geq q$, instead of $r = q(s+1)$, which is too restrictive.

Let f_t and ξ_t be mutually orthogonal stationary processes and let $\chi_t = B(L) f_t$ be the common component. In factor analysis jargon $X_t = B(L) f_t + \xi_{it}$ is referred to as the dynamic factor model, and $X_t = \Lambda F_t + \xi_{it}$ as the static factor model. Similarly, f_t is regarded as the vector of the dynamic factors while F_t as the vector of the static factors. Since dynamic common factors are latent, they need to be estimated. Forni et al. (2005) estimate dynamic factors through the use of dynamic principal component analysis. It involves estimating the eigenvalues and eigenvectors decomposition of the spectral density matrix of X_t, which is a generalization of the orthogonalization process in case of static principal components. The spectral density matrix of X_t, which is estimated using the frequency $-\pi < \theta < \pi$, can be decomposed into the spectral densities of the common and the idiosyncratic component:

$$\Sigma(\theta) = \Sigma_{X}(\theta) + \Sigma_{\xi}(\theta) \quad (3)$$

where $\Sigma_{\chi}(\theta) = B(e^{-i\theta})\Sigma_f(\theta)B\left(e^{-i\theta}\right)'$ is the spectral density matrix of the common component χ_t and $\Sigma_{\xi}(\theta)$ is the spectral density matrix of the idiosyncratic component ξ_t. The rank of $\Sigma_{\chi}(\theta)$ is equal to the number of dynamic factors, q. Similarly, the covariance matrix of X_t can be decomposed as:

$$\Gamma_k = \Gamma_k^{\chi} + \Gamma_k^{\xi} \quad (4)$$

where $\Gamma_k^{\chi} = \Lambda \Gamma_k^F \Lambda'$, Γ_k^F is the covariance matrix of F_t at lag k and Γ_k^{ξ} is the covariance matrix of ξ_t at lag k. The rank of Γ_k^{χ} is equal to r; the number of static factors.

The forecast of the i^{th} variable *h*-steps ahead is not feasible in practice since the common factors are unobserved. However, if data follow an approximate dynamic factor model, the set of common factors F_t can be consistently estimated by appropriate cross-sectional averages, or aggregators in the terminology of Forni and Reichlin (1998) and Forni and Lippi (2001). The rational is that using the law of large numbers, only the pervasive common sources survive the aggregation, as the weakly correlated idiosyncratic errors are averaged out. Building on Chamberlain and Rothschild (1983), Forni et al. (2000) and Stock and Watson (2002a) have shown that principal components of the observed variables X_t, are appropriate averages. That is, the common component can be approximated by projecting either on the first r principal components of the covariance matrix (see Stock and Watson (2002a)) or on the first q dynamic principal components (see Forni, Hallin, Lippi, and Reichlin (2000)).

Empirically, we estimate the autocovariance matrix of standardized data, $\hat{X}_t = (\hat{x}_{1t},...,\hat{x}_{nt})'$ by:

$$\hat{\Gamma}_k = \frac{1}{T-k-1}\sum_{t=k}^{T}\hat{X}_t\hat{X}_t' \tag{5}$$

where T is the sample size. Following Forni et al. (2005) the spectral density matrix will be estimated by averaging a given number m of autocovariances:

$$\hat{\Sigma}(\theta) = \frac{1}{2\pi}\sum_{k=-m}^{m} w_k \hat{\Gamma}_k e^{-i\theta k} \tag{6}$$

where w_k are Barleet-lag window estimator weights $w_k = 1 - \frac{|k|}{m+1}$. The consistent estimates are ensured, provided that $m \to \infty$ and $\frac{m}{T} \to 0$ as $T \to \infty$. In the empirical section we will use $m = \sqrt{T}$, which satisfies the above asymptotic requirements.

The procedure of by Forni et al. (2005) consists of two steps. The first step is the problem of the spectral density matrix, defined, at a given frequency θ, as:

$$\hat{\Sigma}(\theta)V_q(\theta) = V_q(\theta)D_q(\theta) \tag{7}$$

where $D_q(\theta)$ is a diagonal matrix having the diagonal on the first q largest eigenvalues of $\hat{\Sigma}(\theta)$ and $V_q(\theta)$ is the $n \times q$ matrix whose columns are the corresponding eigenvectors. Let X_t is driven by q dynamic factors, the spectral density matrix of the common component is given by:

$$\hat{\Sigma}_\chi(\theta) = V_q(\theta)D_q(\theta)V_q(\theta)' \tag{8}$$

The spectral density matrix of the idiosyncratic part is estimated as a residual:

$$\hat{\Sigma}_\xi(\theta) = \hat{\Sigma}(\theta) - \hat{\Sigma}_\chi(\theta) \tag{9}$$

The covariance matrices of common and idiosyncratic parts are computed through an inverse Fourier transform of spectral density matrices:

$$\hat{\Gamma}_k^\chi = \frac{2\pi}{2m+1}\sum_{j=-m}^{m}\hat{\Sigma}_\chi(\theta_j)e^{ik\theta_j} \tag{10}$$

$$\hat{\Gamma}_k^\xi = \frac{2\pi}{2m+1}\sum_{j=-m}^{m}\hat{\Sigma}_\xi(\theta_j)e^{ik\theta_j} \tag{11}$$

where $\theta_j = \frac{2\pi}{2m+1}j$ and $j = -m, ..., m$

In a second step, the estimated covariance matrix of the common components is used to construct the factor space by r contemporaneous averages. These r contemporaneous averages are solutions from the generalized principal components (*GPC*) problem:

$$\hat{\Gamma}_0^\chi V_{rg} = \hat{\Gamma}_0^\xi V_{rg} D_{rg} \tag{12}$$

$$\text{s.t. } V_{rg}'\hat{\Gamma}_0^\xi V_{rg} = I_r$$

where D_{rg} is a diagonal matrix having on the diagonal the first r largest generalized eigen values of the pair $\left(\hat{\Gamma}_0^\chi, \hat{\Gamma}_0^\xi\right)$ and V_{rg} is the $n \times r$ matrix whose columns are the corresponding eigenvectors.

The first r GPCs are defined as:

$$\hat{F}_t^g = V_{rg}'\hat{X}_t \tag{13}$$

The off-diagonal elements of $\hat{\Gamma}_0^\xi$ are set to zero to overcome the problem of instability that is common in the generalized principal component methodology. With such restrictions, the generalized principal components can be seen as static principal components computed on weighed data, in that these weights are inversely proportional to the variance of the idiosyncratic components. Such a weighting scheme should provide more efficient estimates of the common factors.

$$\hat{\chi}_{iT+r|T} = \hat{\Gamma}_{i.,r}^\chi V_{rg}(V_{rg}'\hat{\Gamma}_0 V_{rg})^{-1}V_{rg}'\hat{X}_T \tag{14}$$

and

$$\hat{\xi}_{iT+r|T} = \left[\hat{\Gamma}_{ii,r}^\xi, ..., \hat{\Gamma}_{ii,r+p}^\xi\right]W_{i,k}^{-1}\left[\hat{x}_{iT}, ..., \hat{x}_{iT-p}\right]' \tag{15}$$

where $W_{i,k} = \begin{bmatrix} \hat{\Gamma}_{ii,0} & \cdots & \hat{\Gamma}_{ii,-(k-1)} \\ \ldots & \cdots & \cdots \\ \hat{\Gamma}_{ii,k-1} & \cdots & \hat{\Gamma}_{ii,0} \end{bmatrix}$

The forecast of $y_{i,T+h|T}$ is computed as follows:

$$\hat{y}_{i,T+k|T} = \hat{\sigma}_i \left(\hat{\chi}_{i,T+k|T} + \hat{\hat{\xi}}_{i,T+k|T} \right) + \hat{\mu}_i \qquad (16)$$

3. Data

It is imperative in factor analysis framework to extract common components from a data rich environment. After extracting common components of house price inflation in five metropolitan areas of South Africa, we make out-of-sample forecast for one, two, three, and four quarters ahead. The data set contains 282 quarterly series of South Africa, ranging from real, nominal, and financial sectors. We also have intangible variables, such as confidence indices, and survey variables. In addition to national variables, we use a set of global variables such as commodity industrial inputs price index and crude oil prices. The data also comprises series of major trading partners such as Germany, the United Kingdom (UK), and the United States (US) of America. All series are seasonally adjusted and covariance stationary. The more powerful *** (DFGLS) test of Elliott, Rothenberg, and Stock (1996), instead of the most popular *** (ADF) test, is used to assess the degree of integration of all series. All nonstationary series are made stationary through differencing. The Schwarz information criterion is used in the selecting the appropriate lag length so that no serial correction is left in the stochastic error term. Where there were doubts about the presence of unit root, the KPSS test proposed by Kwiatowski, Phillips, Schmidt, and Shin (1992), with the null hypothesis of stationarity, was applied. All series are standardized to have a mean of zero and a constant variance. The in-sample period contains data from 1980Q1 to 2000Q4, while the out-of-sample set is 2001Q1-2006Q4.[5]

There are various statistical approaches in determining the number of factors in the DFM. For example, Bai and Ng (2002) developed some criteria guiding the selection of the number of factors in large dimensional panels. The principal component analysis (PCA), where the a number of factors q be based on the first eigenvalues of the spectral density matrix of X_t, can also be used in establishing the number of factors in the DFM. Thereafter, the principal components are added till the increase in the explained variance is less than a specific α, say equal to 0.05. The Bai and Ng (2002) approach proposes five static factors, while Bai and Ng (2007) suggests two primitive or dynamic factors. Similar to the latter method, the principal component technique, as proposed by Forni et al. (2000) suggests two dynamic factors, with the first two dynamic principal components explaining approximately 99 percent of variation, while the eigenvalue of the third component is 0.002 < 0.05.

4. Alternative Forecasting Models

In this study, the DFM is our benchmark model. To evaluate the forecasting performance of the DFM, we require alternative models, and in our case, these are namely, the unrestricted classical VAR, BVARs based on the Minnesota Prior, and the SBVARs based on the FOSC and RWA priors. This section outlines the basics of the above-mentioned competing models.

4.1 The Vector Autoregressive (VAR) model

The Vector Autoregressive (VAR) model, though 'atheoretical', is particularly useful for forecasting. A VAR model can be visualized as an approximation of the reduced-form simultaneous equation structural model. An unrestricted VAR model, as suggested by Sims (1980), can be written as follows:

$$y_t = A_0 + A(L) y_t + \varepsilon_t \qquad (17)$$

where *y* is a ($n \times 1$) vector of variables being forecasted; *A(L)* is a ($n \times n$) polynomial matrix in the backshift operator *L* with lag length *p, i.e., A(L)* = $A_1 L + A_2 L^2 + ... + A_p L^p$; A_0 is a ($n \times 1$) vector of constant terms, and ε is a ($n \times 1$) vector of error terms. In our case, we assume that $\varepsilon \sim N(0, \sigma^2 I_n)$, where I_n is a $n \times n$ identity matrix. Note, the VAR model generally uses equal lag length for all the variables of the model. A drawback of VAR models is that many parameters need to be estimated, some of which may be insignificant. This problem of overparameterization, resulting in multicollinearity and a loss of degrees of freedom, leads to inefficient estimates and possibly large out-of-sample forecasting errors. A solution often adapted, is simply to exclude the insignificant lags based on statistical tests. Another approach is to use a near VAR, which specifies an unequal number of lags for the different equations.

An alternative approach to overcome overparameterization, as described in Litterman (1981), Doan *et al* (1984), Todd (1984), Litterman (1986), and Spencer (1993), is to use a Bayesian VAR (BVAR) model. Instead of eliminating longer lags, the Bayesian method imposes restrictions on these

[5]Details about data and their statistical treatment are available upon request.

coefficients by assuming that they are more likely to be near zero than the coefficients on shorter lags. However, if there are strong effects from less important variables, the data can override this assumption. The restrictions are imposed by specifying normal prior distributions with zero means and small standard deviations for all coefficients, with the standard deviation decreasing as the lags increase. The exception to this is that the coefficient on the first own lag of a variable has a mean of unity. Litterman (1981) used a diffuse prior for the constant. This is popularly referred to as the 'Minnesota prior' due to its development at the University of Minnesota and the Federal Reserve Bank at Minneapolis.

Formally, as discussed above, the means and variances of the Minnesota prior take the following form:

$$\beta_i \sim N(1, \sigma^2_{\beta_i}) \text{ and } \beta_j \sim N(0, \sigma^2_{\beta_j}) \tag{18}$$

where β_i denotes the coefficients associated with the lagged dependent variables in each equation of the VAR, while β_j represents any other coefficient. In the belief that lagged dependent variables are important explanatory variables, the prior means corresponding to them are set to unity. However, for all the other coefficients, β_j's, in a particular equation of the VAR, a prior mean of zero is assigned to suggest that these variables are less important to the model. The prior variances $\sigma^2_{\beta_i}$ and $\sigma^2_{\beta_j}$, specify uncertainty about the prior means $\overline{\beta}_i$ = 1, and $\overline{\beta}_j$ = 0, respectively. Because of the overparameterization of the VAR, Doan *et al.* (1984) suggested a formula to generate standard deviations as a function of small numbers of hyperparameters: *w*, *d*, and a weighting matrix *f(i, j)*. This approach allows the forecaster to specify individual prior variances for a large number of coefficients based on only a few hyperparameters. The specification of the standard deviation of the distribution of the prior imposed on variable *j* in equation *i* at lag *m*, for all *i, j* and *m*, defined as $S_1(i, j, m)$, can be specified as follows:

$$S_1(i, j, m) = [w \times g(m) \times f(i, j)] \frac{\hat{\sigma}_j}{\hat{\sigma}_i} \tag{19}$$

with *f(i, j) = 1*, if *i = j* and k_{ij} otherwise, with ($0 \le k_{ij} \le 1$), *g(m) =* $m^{-d}, d > 0$. Note that $\hat{\sigma}_i$ is the estimated standard error of the univariate autoregression for variable *i*. The ratio $\hat{\sigma}_i / \hat{\sigma}_j$ scales the variables to account for differences in the units of measurement and, hence, causes specification of the prior without consideration of the magnitudes of the variables. The term *w* indicates the overall tightness and is also the standard deviation on the first own lag, with the prior getting tighter as we reduce the value. The parameter *g(m)* measures the tightness on lag *m* with respect to lag 1, and is assumed to have a harmonic shape with a decay factor of *d*, which tightens the prior on increasing lags. The parameter *f(i, j)* represents the tightness of variable *j* in equation *i* relative to variable *i*, and by increasing the interaction, i.e., the value of k_{ij}, we can loosen the prior.[6]

Note, the overall tightness (*w*) and the lag decay (*d*) hyperparameters used in the standard Minnesota prior have values of 0.1 and 1.0, respectively, while k_{ij} = 0.5, implies a weighting matrix (*F*) with 1.0 on the diagonals and 0.5 as the off-diagonal elements.. Given that the Minnesota prior treats all variables in the VAR, except for the first own-lag of the dependent, in an identical manner, several attempts have been made to alter this fact. Usually, this has boiled down to increasing the value for the overall tightness (*w*) hyperparameter from 0.10 to 0.20, so that the larger value of *w* can allow for more influence from other variables in the model. In addition, as proposed by Dua and Ray (1995), we also try out a prior that is even more loose, specifically with *w* = 0.30 and *d* = 0.50. Alternatively, LeSage and Pan (1995) have suggested the construction of the weight matrix based on the First-Order Spatial Contiguity (FOSC), which simply implies the creation of a non-symmetric *F* matrix that emphasizes the importance of the variables from the neighboring states/provinces more than that of the non-neighboring states/provinces. Lesage and Pan (1995) suggests the use of a value of unity on not only the diagonal elements of the weight matrix, as in the Minnesota prior, but also in place(s) that correspond to the variable(s) from other state(s)/province(s) with which the specific state in consideration have common border(s). However, for the elements in the *F* matrix that corresponds to variable(s) from state(s)/province(s) that are not immediate neighbor(s), Lesage and Pan (1995) proposes a value of 0.1.

Referring to the provincial map of South Africa given in Figure 1, the design of the *F* matrix

[6] For an illustration, see Dua and Ray (1995).

based on the FOSC prior, given the alphabetical ordering[7] of the five metropolitan areas as the Eastern Cape Metropolitan area (Port Elizabeth/Uitenhage), Greater Johannesburg, the Kwa-Zulu Natal Metropolitan area (Durban Unicity), Pretoria and the Western Cape Metropolitan area (Cape Town), can be formalized as follows:

$$F = \begin{bmatrix} 1.0 & 0.1 & 1.0 & 0.1 & 1.0 \\ 0.1 & 1.0 & 0.1 & 1.0 & 0.1 \\ 1.0 & 0.1 & 1.0 & 0.1 & 0.1 \\ 0.1 & 1.0 & 0.1 & 1.0 & 0.1 \\ 1.0 & 0.1 & 0.1 & 0.1 & 1.0 \end{bmatrix} \quad \ldots\ldots\ldots\ldots \quad \ldots(20)$$

The intuition behind this asymmetric *F* matrix is based on our lack of belief on the prior means of zero imposed on the coefficient(s) for house price inflations(s) of the neighboring province(s). Instead we believe that these variables do have an important role to play, hence, to express our lack of faith in the prior means of zero, we assign a larger prior variance, by increasing the weight values, to these prior means on the coefficients for the variables of the neighboring states. This, in turn, allows the coefficients on these variables to be determined based more on the sample and less on the prior.

LeSage and Krivelyova (1999) have put forth an alternative approach to remedy the equal treatment nature of the Minnesota prior, called the "Random-Walk Averaging" (RWA) prior. that involves both the prior means and the variances based on a distinction made between important variables (like house price inflation(s) of neighboring province(s)) and unimportant variables (like house price(s) of non-neighboring province(s)) in each equation of the VAR model. To understand the motivation behind the design of the prior means, consider the weight matrix *F* for the VAR consisting of house price inflation of the five metropolitan areas. Retaining the ordering of the five metropolitan areas as outlined in the FOSC prior, the weight matrix contains values of unity in positions associated with the house price inflation(s) of neighboring province(s), i.e., for important variables in each equation of the VAR model, while, zero values are assigned to the unimportant variables, i.e., house price(s) of non-neighboring province(s). However, as with the Minnesota prior, we continue to have a value of one on the main diagonal of the *F* matrix, simply to emphasize our belief that the autoregressive influences from the lagged values of the dependant variable (house price of a specific metropolitan area) are important.[8]

[INSERT Figure 1 HERE]

$$F = \begin{bmatrix} 1.0 & 0 & 1.0 & 0 & 1.0 \\ 0 & 1.0 & 0 & 1.0 & 0 \\ 1.0 & 0 & 1.0 & 0 & 0 \\ 0 & 1.0 & 0 & 1.0 & 0 \\ 1.0 & 0 & 0 & 0 & 1.0 \end{bmatrix} \quad (21)$$

The weight matrix given above is then standardized so that the rows sums to unity. Formally, we can write the standardized *F* matrix, *C,* as follows:

$$C = \begin{bmatrix} 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.50 & 0 & 0.50 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0.50 & 0 & 0 & 0 & 0.50 \end{bmatrix} \quad \ldots\ldots\ldots\ldots. \quad (22)$$

The matrix *C*, standardized along the rows, allows us to consider the random-walk with drift, which averages over the important variables in each equation *i* of the VAR. Formally,

$$y_{it} = \delta_i + \sum_{j=1}^{n} C_{ij} y_{jt-1} + u_{it} \quad \ldots\ldots\ldots\ldots \quad (23)$$

where in our case $n = 5$. On expanding equation (23), we observe that multiplying y_{jt-1} containing the house price growth rates of five metropolitan areas at *t-1* by the matrix *C* would produce set of

[7] It must, however, be pointed out that alternative ordering of the six metropolitan areas do not affect our final results in any way.

[8] However, using a value of one on the main diagonal element of the *F* matrix, under the RWA prior, is not always an obvious choice. See LeSage and Krivelyova (1999) for an alternative exposition, where autoregressive influences are considered to be important only for certain variables.

explanatory variables for each equation of the VAR equal to the mean of observations from the important variables (neighboring house prices) in each equation *i* at *t-1*.[9] This also suggests that the prior mean for the coefficients on the first own-lag of the important variables is equal to $1/c_i$, with c_i being the number of important variables in a specific equation *i* of the VAR model. However, as in the Minnesota prior, the RWA prior uses a prior mean of zero for the coefficients on all lags, except for the first own lags. At this juncture, it is important to point out that RWA approach of specifying prior means require the variables to be scaled to have similar magnitudes. This is simply because, it does not make much sense intuitively otherwise to suggest that the value of a variable at *t* was equal to the average of values from the important variables at *t-1*. This transformation is not much of an issue as the data on the variables, in our case the house prices, can always be expressed as percentage change or annualized growth rates, thus meeting the similar magnitudes requirements of the RWA prior.

As proposed by LeSage and Krivelyova (1999), a flexible form in which the RWA prior standard deviations ($S_2(i,j,m)$) for a variable *j* in equation *i* at lag length *m* is as follows:

$$
\begin{aligned}
&S_2(i,j,m) \sim N(\frac{1}{c_i},\sigma_c);\ j \in C; m=1; i,j=1,....,n \\
&S_2(i,j,m) \sim N(0,\eta\frac{\sigma_c}{m});\ j \in C; m=2,....,p; i,j=1,....,n \qquad (24) \\
&S_2(i,j,m) \sim N(0,\rho\frac{\sigma_c}{m});\ j \neg\in C; m=1,....,p; i,j=1,....,n
\end{aligned}
$$

where $0<\sigma_c<1; \eta>1$ and $0<\rho\le 1$. For the variables *j = 1,…..,n* in equation *i* that are important in explaining the movements in variable *i* i.e., $j \in C$, the prior mean for the lag length of 1 is set to the average of the number of important variables in equation *i* and to zero for the unimportant variables, i.e., $j \neg\in C$. With $0<\sigma_c<1$, the prior standard deviation for the first own-lag imposes a tight prior mean to reflect averaging over important variables. For important variables at lags greater than one, the variance decreases as *m* increases, but the restriction of $\eta>1$ allows for the zero prior means on the coefficients of these variables to be imposed loosely. Finally, we use $\rho\sigma_c/m$ for lags on unimportant variables, which has prior means of zero, to indicate that the variance decreases as *m* increases. In addition, with $0<\rho\le 1$, we impose the zero means on the unimportant variables with more certainty.

5. Evaluation of Forecast Accuracy

The BVARs and the SBVARs, based on the FOSC and the RWA priors, are estimated using Theil's (1971) mixed estimation technique, which essentially involves supplementing the data with prior information on the distribution of the coefficients. The number of observations and degrees of freedom are increased by one in an artificial way, for each restriction imposed on the parameter estimates. The loss of degrees of freedom due to over- parameterization associated with a classical VAR model is, therefore, not a concern in the BVARs and SBVARs.

Given the specification of the priors above, we estimate a VAR, BVARs and two SBVAR models each for small, medium and large middle-segment houses, based on the FOSC and the RWA priors, for the Eastern Cape Metropolitan area (Port Elizabeth/Uitenhage), Greater Johannesburg, the Kwa-Zulu Natal Metropolitan area (Durban Unicity), Pretoria and the Western Cape Metropolitan area (Cape Town) over the period of 1980:Q1 to 2000:Q4, using quarterly data. Then we compute the out-of-sample one- through four-quarters-ahead forecasts for the period of 2001:Q1 to 2006:Q4, and compare the forecast accuracy relative to that of the forecasts generated the benchmark DFM model. Note the variables included in the VARs, classical and Bayesian (spatial and non-spatial) are the house price inflation (percentage change in the house prices) of the above mentioned five metropolitan areas. All data on house prices are seasonally adjusted, before being converted to house price inflation, in order to, *inter alia*, address the fact that, as pointed out by Hamilton (1994:362), the Minnesota-type priors are not well suited for seasonal data. Again recall, the house price data are obtained from the latest ABSA Housing Price Review.

In each equation of the different types of VARs, there are 41 parameters including the constant, given the fact that the model is estimated with 8 lags[10] of each variable. Note Sims *et al.* (1990) indicates that with the Bayesian approach entirely based on the likelihood function, the associated

[9] Just as with the constant in the Minnesota Prior, δ is also estimated based on a diffuse prior.

[10] The choice of 8 lags is based on the unanimity of the sequential modified LR test statistic, Akaike information criterion (AIC), and the final prediction error (FPE) criterion.

inference does not need to take special account of nonstationarity, since the likelihood function has the same Gaussian shape regardless of the presence of nonstationarity. Given this, we do not need to bother about ensuring the stationarity of house price inflation in the Bayesian models.

The five-variable VAR, and the BVAR and the SBVAR models for an initial prior, are estimated for the period of 1980:Q1 to 2000:Q4 and, then, we forecast from 2001:Q1 through to 2006:Q4. Since we use eight lags, the initial eight quarters of the sample, 1980:Q1 to 1981:Q4, are used to feed the lags. We generate dynamic forecasts, as would naturally be achieved in actual forecasting practice. The models are re-estimated each quarter over the out-of-sample forecast horizon in order to update the estimate of the coefficients, before producing the 4-quarters-ahead forecasts. This iterative estimation and 4-steps-ahead forecast procedure was carried out for 24 quarters, with the first forecast beginning in 2001:Q1. This experiment produced a total of 24 one-quarter-ahead forecasts, 24-two-quarters-ahead forecasts, and so on, up to 24 4-step-ahead forecasts. We use the algorithm in the Econometric Toolbox of MATLAB[11], for this purpose. The MAEs[12] for the 24, quarter 1 through quarter 4 forecasts are then calculated for the house price inflation of the five metropolitan areas. The values of the MAE statistic for one- to four-quarters -ahead forecasts for the period 2001:Q1 to 2006:Q4 are then examined. The model, DFM or any of the VARs, that produces the lowest average value for the MAE is selected, as the 'optimal' model for a specific metropolitan area corresponding to a specific size of the middle-segment houses.

To evaluate the accuracy of forecasts generated by the DFM, we need alternative forecasts. To make the MAEs comparable with the DFM, we report the same set of statistics for the out-of-sample forecasts generated from an unrestricted classical VAR, BVARs, and SBVAR based on the FOSC and RWA[13] priors. In Tables 1 to 3, we compare the average MAEs of one- to four-quarters-ahead out-of-sample-forecasts for the period of 2001:Q1 to 2006:Q4, generated by the unrestricted DFM, VAR, the BVARs and the SBVARs.[14] The conclusions from these tables can be summarized as follows:

[INSERT TABLES 1 THROUGH 3]

(i) Large Middle-Segment Houses: As can be seen from the average MAE values for one- to four-quarters-ahead forecasts, reported in Table 1, for this category of middle-segment houses, the DFM outperforms the all the other models, except for Eastern Cape and the Western Cape metropolitan area, which, in turn, were forecasted best by the SBVAR model based on the FOSC, respectively. Amongst the alternative VARs, the SBVAR model does the best for all the metropolitan areas, except for Durban Unicity, under the Kwa-Zulu Natal metropolitan area, which, in turn, is forecasted with the lowest errors by the BVAR model with $w = 0.1$, $d = 1.0$.

(ii) Medium Middle-Segment Houses: As reported in Table 2, the DFM produced the minimum one- to four-quarters-ahead average MAE values for all of the four metropolitan areas, except for the Western Cape, when compared to all the competing models. The BVAR model with $w = 0.2$, $d = 1.0$, does the best for Western Cape. Amongst the VARs, the BVAR with $w = 0.2$, $d = 1.0$, stands out for Johannesburg, by producing the lowest average MAEs. Durban Unicity and Pretoria are forecasted with the lowest average MAEs by the SBVAR based on the FOSC. While, the BVAR with $w = 0.1$, $d = 1.0$, is best-suited for forecasting the Eastern Cape and Western Cape metropolitan areas.

---

[11] All statistical analysis was performed using MATLAB, version R2006a.

[12] Note that if A_{t+n} denotes the actual value of a specific variable in period $t + n$ and ${}_tF_{t+n}$ is the forecast made in period t for $t + n$, the MAE statistic can be defined as $(\frac{1}{N}\sum abs(A_{t+n} - {}_tF_{t+n})) \times 100$, where *abs* stands for the absolute value. For $n = 1$, the summation runs from 2001:01 to 2006:04, and for $n = 2$, the same covers the period of 2001:02 to 2006:04, and so on.

[13] Note, the SBVAR model based on the RWA prior that did best amongst other SBVAR models with the RWA prior, consistently for all house sizes and majority of the metropolitan areas, had the following values of the hyperparameters: $\sigma_c = 0.3; \eta = 8$ and $\rho = 1$. Note the values for these hyperparameters are based on the ranges suggested by LeSage (1999).

[14] However, the MAEs for each of the steps of the one- to four-quarters-ahead for all the seven models have been reported in Tables A1 through A7 in the appendix of the paper. We have now abbreviated the metropolitan areas as: ECAP, JOBU, KWAZ, PRET and WCAP for Eastern Cape, Johannesburg, Kwa-Zulu Natal, Pretoria and Western Cape.

(iii) Small Middle-Segment Houses: As can be seen from Table 3, except for the Pretoria and Western Cape Metropolitan areas, the DFM stands out as the best-suited model for forecasting house price inflation. For the Eastern Cape metropolitan area, SBVAR based on the FOSC performs best, while for Pretoria the BVAR model with w = 0.1, d = 1.0 is the preferred model. Amongst the VARs, the SBVAR based on the FOSC produces the lowest average MAEs for the two Capes, while the BVAR model with w = 0.1, d = 1.0 outperforms the other VAR models for Johannesburg and Pretoria. The BVAR model with w = 0.2, d = 1.0 does the best for Durban.

At this stage, refer to the results of Gupta and Das (2008), to put the results of this paper into perspective. The authors observed that, though there did not exist a specific model that performed outright best, in terms of forecasting house prices of different sizes in the six metropolitan areas of South Africa the spatial models in general, tended to outperform the other models for large middle-segment houses; while the unrestricted VAR and the BVAR models tended to produce lower average out-of-sample forecast errors for middle and small middle segment houses, respectively. In our case though, in general, and especially for the large middle-segment houses, the SBVAR model based on the FOSC, tends to stand out,[15] when we take the DFM out of consideration. The next best performing model is the BVAR model with w = 0.2, d = 1.0. In addition, the VAR, the BVAR with w = 0.2, d = 0.5, and the SBVAR based on the RWA prior are the worst performing models. However, what we can say is that, in general, the DFM, performing best in three, four, and three cases for the large-, medium-, and small-middle segment houses, respectively, clearly is the overwhelming favorite in forecasting regional house price inflation in South Africa over the period of 2001:Q1 to 2006:Q4.[16]

6. Conclusions

This paper analyzes whether the wealth of information contained in the DFM framework can be useful in forecasting regional house price inflation. As a case study illlustration, we use the DFM to predict house price inflation in five metropolitan areas of South Africa, namely, Cape Town, Durban, Johannesburg, Port Elizabeth and Pretoria, using quarterly data over the period of 1980Q1-2006Q4. The in-sample period contains data from 1980Q1 to 2000Q4 and the out-of-sample forecasts are based on one- to four-quarter-ahead forecasts over a 24-quarter forecasting horizon covering 2001Q1 to 2006Q4. The forecast performance of DFM is evaluated in terms of the MAEs, by comparing it with SBVAR models, based on the FOSC and the RWA priors, besides the non-spatial models like the VAR model and BVAR models with the Minnesota prior, estimated merely based on the house price inflation of the five abovementioned metropolitan areas of South Africa.

Our results indicate that the data-rich DFM, in general, is best suited in forecasting regional house price inflation, when compared to the alternative VARs. Clearly then, the role of a DFM containing a wide range of data for an economy, besides also including a set of global variables of major trading partners, cannot be underestimated in predicting regional house price inflation. In addition, given that there are at least two major limitations to using a Bayesian approach for forecasting,: Firstly, as it is clear from Tables 1 to 3, the forecast accuracy is sensitive to the choice of the priors. So if the prior is not well specified, an alternative model used for forecasting may perform better. Secondly, in case of the Bayesian models, one requires to specify an objective function, for example the average MAEs, to search for the 'optimal' priors, which, in turn, needs to be optimized over the period for which we compute the out-of-sample forecasts. However, there is no guarantee that the chosen parameter values specifying the prior will continue to be 'optimal' beyond the period for which it was selected. As such, the DFM is, perhaps, a better model to base ones' forecasts on.

An immediate extension of the current study would be to put the DFM to test against a full-fledged model of house prices, based on proper theoretical considerations of the demand and supply factors affecting the housing market. In this regard, as in Rapach and Strauss (2007a, 2007b), one can also consider using a small-scale DFM that incorporates the essential fundamentals affecting the housing market and variables specific to this market.

[15] Gupta and Das (2008) point out that this might be due to the importance of spatial correlations in the determination of the prices of large-sized houses possibly because, there exists less heterogeneity in the supply of these kind of housing, or, alternatively, wealthier customers tend to have similar characteristics, thus, causing the prices to cluster around some values.

[16] In all the case where the DFM are outperformed, it is the third-best model.

References

Abraham JM, Hendershott PH. Bubbles in Metropolitan Housing Markets. *Journal of Housing Research* 1996; **7**: 191–207.

Bai J, Ng S. Determining the Number of Primitive Shocks in Factor Models*. Journal of Business and Economic Statistics* 2007; **25**:52–60.

Bai J, Ng S. Determining the Number of Factors in Approximate Factor Models. *Econometrica* 2002; **70**:191–221.

Bernanke B, Gertler M. Inside the Black Box: the Credit Channel of Monetary Transmission. *Journal of Economic Perspectives* 1995; **9**:27–48.

Burger P, Van Rensburg LJ. Metropolitan House Prices in South Africa: Do they Converge?, *Working Paper* 2007 ; Department of Economics, University of Free State.

Christoffersen P, Chung H, Errunza V. Size matters: The impact of financial liberalization on individual firms. *Journal of International Money and Finance* 2006; **25**: 1296-1318.

Cho M. House Price Dynamics: A Survey of Theoretical and Empirical Issues*. Journal of Housing Research* 1996; **7**: 145–172.

Choe H, Kho BC, Stulz RM. Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial Economics* 1999; **54**:227-264.

Cristadoro R, Forni M, Reichlin L, Veronese G. A Core Inflation Indicator for the Euro Area .*Journal of Money, Credit and Banking* 2005; **37**: 539–560.

Dahlquist M, Pinkowitz L, Stulz RM, Williamson R. Corporate governance and the home bias. *Journal of Financial and Quantitative Analysis* 2003; **38**:87-110.

Doan T, Litterman RB, Smis C. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 1984; **3**:1-144.

Dua P, Miller SM, Smyth DJ. Using Leading Indicators to Forecast U. S. Home Sales in a Bayesian Vector Autoregressive Framework. *Journal of Real Estate Finance and Economics* 1999; **18**: 191-205.

Dua P, Ray SC. A BVAR Model for the Connecticut Economy. *Journal of Forecasting* 1995; **14**:167-180.

Dua P, Smyth DJ. Forecasting U.S. Home Sales using BVAR Models and Survey Data on Households' Buying Attitude*. Journal of Forecasting* 1995; **14**: 217-227.

Elliott G, Rothenberg TJ, Stock J. Efficient Tests for an Autoregressive Unit Root. *Econometrica* 1996; **64**: 813–836.

Forni M, Hallin M, Lippi M, Reichlin L. The Generalized Dynamic Factor Model: identification and estimation. *Review of Economics and Statistics* 2005;**82**:540-554.

Forni M, Hallin M, Lippi M, Reichlin L. Do financial variables help forecasting inflation and real activity in the euro area*? Journal of Monetary Economics* 2003; **50**: 1243-1255.

Forni M, Lippi M. The Generalized Dynamic Factor Model: representation theory. *Econometric Theory* 2001;**17**:1113-1141.

Forni M, Hallin M, Lippi M, Reichlin L. Coincident and leading indicators for the EURO area. *The Economic Journal* 2001, **111**: 62-85.

Forni M, Hallin M, Lippi M. Reichlin L. The Generalized Dynamic Factor Model: identification and estimation. *Review of Economics and Statistics* 2000; **82**: 540–554.

Forni M, Reichlin L. Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies* 1998; **65**: 453-473.

Giannone D, Matheson TD. A New Core Inflation Indicator for New Zealand. *International Journal of Central Banking* 2007; **3**: 145-180.

Gupta R. Das S. Spatial Bayesian methods of forecasting house prices in six metropolitan areas of South Africa. Mimeo, 2008.

Gupta R, Kabundi A. A Dynamic Factor Model for Forecasting Macroeconomic Variables in South Africa. Mimeo, 2008.

Hamilton J D. Time Series Analysis. Second Edition. Princeton: Princeton University Press, 1994.

International Monetary Fund. World Economic Outlook: Asset Prices and the Business Cycle, 2000;May.

Johnes G, Hyclak T. House Prices and Regional Labor Markets. *Annals of Regional Science* 1999; **33**: 33–49.

Kabundi A. Synchronization between South Africa and the U.S.: A Structural Dynamic Factor Analysis. Economic Research Southern Africa, University of Cape Town, *Working Paper* 2007; No. 50.

Kabundi A. Estimation of Economic Growth Using Business Survey Data. International Monetary Fund, *Working Paper*, 2004; No. 04/69.

Kang JK, Stulz RM. Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. *Journal of Financial Economics* 1997;**46**:2-28.

Kwiatowski D, Phillips PCB, Schmidt P, Shin Y. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics* 1992; **54**: 159–178.

LeSage JP, Pan Z. Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models. *International Regional Science Review* 1995; **18**: 33-53.

LeSage JP, Krivelyova A. A Spatial Prior for Bayesian Autoregressive Models.*Journal of Regional Science* 1999; **39**: 297-317.

LeSage JP. Applied Econometrics Using MATLAB, 1999. www.spatial-econometrics.com.

Litterman RB. A Bayesian procedure for forecasting with vector autoregressions. Federal Reserve Bank of Minneapolis, *Working Paper*, 1981.

Litterman RB. Forecasting with Bayesian vector autoregressions—five years of experience. *Journal of Business and Statistics* 1986; **4**: 25-38.

Rapach DE, Strauss JK. Forecasting Real Housing Price Growth in the Eighth District States. Federal Reserve Bank of St. Louis. *Regional Economic Development* 2007a; **3**: 33–42.

Rapach DE, Strauss JK. Forecasting Housing Prices with Many Potential Predictors. Mimeo, 2007b.

Schneider M, Spitzer M. Forecasting Austrian GDP using the generalized dynamic factor model. Oesterreichische National Bank, *Working Paper* 2004; No. 89.

Sims CA, Stock JH, Watson MW. Inference in linear time series models with some unit roots. *Econometrica* 1990; **58**:113-144.

Sims CA. Macroeconomics and Reality. *Econometrica* 1980; **48**: 1-48.

Spencer DE. Developing a Bayesian Vector Autoregression Model. *International Journal of Forecasting* 1993; **9**:407-421.

Stock JH, Watson MW. Forecasting Output Growth and Inflation: The Role of Asset Prices. *Journal of Economic Literature* 2003; **41**: 788–829.

Stock JH, Watson MW. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 2002a; **97**:147–162.

Stock JH, Watson MW. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economics Statistics* 2002b; **20**:147–162.

Stock JH, Watson MW. Forecasting Inflation. *Journal of Monetary Economics* 1999; **44**: 293-335.

Stock JH, Watson MW. A probability model of the coincident indicators. In leading economic indicators: New approaches and forecasting record (K. Lahiri, and G. Moore, eds.) Cambridge: Cambridge University Press, 1991; 63–95

Stock JH, Watson MW. New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual* 1989; 351-393.

Theil H. *Principles of econometrics*. John Wiley: New York, 1971.

Topel RH, Rosen S. Housing Investment in the United States. *Journal of Political Economy* 1988; **96**:718–740.

Van Nieuwenhuyze C. A. Generalized Dynamic Factor Model for the Belgian Economy Identification of the Business Cycle and GDP Growth Forecasts. *Journal of Business Cycle Measurement and Analysis* 2007; **2**: 213–248.

Table1. Average MAEs for Large Middle-Segment Houses (2001:Q1-2006:Q4)

Models	Eastern Cape	Johannesburg	Kwa-Zulu Natal	Pretoria	Western Cape
DFM	5.1261	**2.2191**	**4.9401**	**2.1836**	2.3428
VAR	9.1879	11.3401	7.0362	10.3206	4.3185
BVAR1	4.3975	6.2833	6.2342	6.2009	1.6372
BVAR2	6.4503	7.8983	6.4835	7.4639	2.6518
BVAR3	8.5352	10.4708	6.8107	9.6630	3.8612
SBVAR1	**3.5749**	4.7957	7.5275	4.2249	**1.4566**
SBVAR2	31.1068	82.8945	10.5748	37.1312	8.5379

Notes: BVAR(1):w=0.1,d=1.0;BVAR(2):w=0.2,d=1.0;BVAR(3):w=0.3,d=0.5, SBVAR1: FOSC, SBVAR2: RWA.

Table 2. Average MAEs for Medium Middle-Segment Houses (2001:Q1-2006:Q4)

Models	Eastern Cape	Johannesburg	Kwa-Zulu Natal	Pretoria	Western Cape
DFM	**3.3206**	**2.3328**	**3.5333**	**1.6659**	2.0668
VAR	7.1063	5.3399	5.7949	3.9338	3.2189
BVAR1	5.5432	5.4218	5.0300	3.6084	**0.3212**
BVAR2	5.8387	5.1316	5.4250	3.4440	0.9284
BVAR3	6.6339	4.8972	5.5663	3.5361	2.5226
SBVAR	5.7803	5.3247	4.0480	3.3879	0.9659
RWASVAR	78.2692	37.4065	16.3128	22.3649	26.5445

Notes: BVAR(1):w=0.1,d=1.0;BVAR(2):w=0.2,d=1.0;BVAR(3):w=0.3,d=0.5, SBVAR1: FOSC, SBVAR2: RWA.

Table 3. Average MAEs for Small Middle-Segment Houses (2001:Q1-2006:Q4)

Models	Eastern Cape	Johannesburg	Kwa-Zulu Natal	Pretoria	Western Cape
DFM	**4.6881**	**2.6770**	**4.1480**	2.4607	2.3597
VAR	12.7709	5.6551	5.4324	4.2063	5.7359
BVAR1	11.2287	3.1076	5.0095	**2.0019**	2.1255
BVAR2	12.1519	4.0998	4.3014	2.6640	3.2852
BVAR3	12.4742	5.4004	5.1237	3.8314	5.1758
SBVAR1	10.2354	4.1184	4.6609	2.4438	**1.7485**
SBVAR2	18.3635	5.8401	7.2547	7.7028	6.4930

Notes: BVAR(1):w=0.1,d=1.0;BVAR(2):w=0.2,d=1.0;BVAR(3):w=0.3,d=0.5, SBVAR1: FOSC, SBVAR2: RWA.

APPENDIX

Table: A1. VAR

	HORIZON				
	1	2	3	4	**AVERAGE**
PROVINCES	**LARGE**				
ECAP	15.0508	12.4597	5.9946	3.2464	9.187875
JOBU	10.0058	17.0889	9.5959	8.6696	11.34005
KWAZ	10.9667	6.4443	0.5436	10.1903	7.036225
PRET	1.2083	5.7533	13.065	21.2559	10.320625
WCAP	5.3392	3.2372	5.7853	2.9121	4.31845
MEDIUM					
ECAP	6.5993	9.4703	1.1398	11.2156	7.10625
JOBU	8.5337	2.4571	4.6054	5.7632	5.33985
KWAZ	1.1915	5.6245	9.2084	7.155	5.79485
PRET	4.3455	5.5938	4.0547	1.741	3.93375
WCAP	3.249	1.5139	6.4243	1.6884	3.2189
SMALL					
ECAP	13.6897	5.6866	13.5655	18.1417	12.770875
JOBU	1.0606	6.1381	3.1665	12.255	5.65505
KWAZ	5.6244	4.7593	4.5131	6.8327	5.432375
PRET	1.0269	2.1015	4.6619	9.0347	4.20625
WCAP	1.9862	10.1872	8.1482	2.6219	5.735875

Table: A2. BVAR1

[w=0.1, d=1.0]

	HORIZON				
	1	2	3	4	**AVERAGE**
PROVINCES	**LARGE**				
ECAP	10.0321	5.8801	1.2706	0.4071	4.397475
JOBU	3.9313	9.9808	5.705	5.5159	6.28325
KWAZ	7.959	6.1716	3.0805	7.7257	6.2342
PRET	1.4258	5.4672	6.2821	11.6288	6.200975
WCAP	2.6904	0.1238	2.8885	0.8459	1.63715
MEDIUM					
ECAP	4.1945	6.5177	1.5659	9.8946	5.543175
JOBU	4.6954	1.0591	5.6908	10.2419	5.4218
KWAZ	0.8303	4.6862	7.7882	6.8153	5.03
PRET	3.1483	3.8595	4.069	3.3567	3.608375
WCAP	0.2072	0.0455	0.7143	0.3176	0.32115
SMALL					
ECAP	6.0819	12.503	15.0107	11.3192	11.2287
JOBU	0.2011	0.1766	0.9689	11.0836	3.10755
KWAZ	2.3705	9.5199	1.9187	6.2289	5.0095
PRET	1.2747	0.3346	1.6375	4.7608	2.0019
WCAP	0.9446	4.1117	3.2231	0.2226	2.1255

Table: A3. BVAR2

[w=0.2, d=1.0]

PROVINCES	HORIZON 1	2	3	4	AVERAGE
	LARGE				
ECAP	12.0673	8.6699	3.275	1.7889	6.450275
JOBU	5.5166	12.3197	7.0058	6.751	7.898275
KWAZ	8.5688	6.0966	2.714	8.5545	6.483475
PRET	1.5411	5.5052	8.2946	14.5147	7.4639
WCAP	3.614	1.7174	4.5606	0.715	2.65175
MEDIUM					
ECAP	4.6092	7.6662	0.9685	10.111	5.838725
JOBU	5.3784	0.7535	5.259	9.1356	5.131625
KWAZ	0.9625	5.1783	8.4557	7.1036	5.425025
PRET	3.114	3.9278	3.878	2.8563	3.444025
WCAP	0.7087	0.3443	2.3682	0.2922	0.92835
SMALL					
ECAP	8.3782	11.1588	14.657	14.4135	12.151875
JOBU	1.3236	2.5438	0.256	12.2757	4.099775
KWAZ	4.0961	7.337	0.4965	5.2759	4.301375
PRET	0.7429	1.3416	2.3766	6.1947	2.66395
WCAP	0.1677	6.769	5.5159	0.6883	3.285225

Table: A4. BVAR3

[w=0.3, d=0.5]

PROVINCES	HORIZON 1	2	3	4	AVERAGE
	LARGE				
ECAP	14.3259	11.2806	5.4729	3.0613	8.535175
JOBU	8.7199	15.907	9.0779	8.1783	10.470775
KWAZ	10.2667	6.3273	1.0838	9.5651	6.810725
PRET	1.6494	5.7059	11.6689	19.6279	9.663025
WCAP	5.1252	2.823	5.7233	1.7734	3.861225
MEDIUM					
ECAP	6.0416	9.3684	0.5666	10.5593	6.633975
JOBU	7.3179	1.2121	4.6373	6.4215	4.8972
KWAZ	1.0925	5.426	8.8355	6.9113	5.566325
PRET	3.8311	5.1434	3.7428	1.4272	3.536125
WCAP	2.5125	1.1907	6.3641	0.0229	2.52255
SMALL					
ECAP	12.0705	7.0767	13.6923	17.0572	12.474175
JOBU	1.1382	5.6132	2.5776	12.2727	5.400425
KWAZ	5.6716	5.426	3.2538	6.1432	5.12365
PRET	1.316	1.7885	4.1368	8.0841	3.83135
WCAP	1.5112	9.3581	7.7059	2.1281	5.175825

Table: A5. SBVAR1

PROVINCES	HORIZON 1	2	3	4	AVERAGE
	LARGE				
ECAP	8.1351	5.3245	0.3423	0.4978	3.574925
JOBU	1.8976	7.2774	4.8673	5.1403	4.79565
KWAZ	10.4105	6.2443	3.2883	10.1669	7.5275
PRET	1.2296	3.3061	2.7382	9.6258	4.224925
WCAP	1.4089	1.9586	1.8549	0.6039	1.456575
MEDIUM					
ECAP	2.3531	7.1931	3.1303	10.4448	5.780325
JOBU	3.3057	2.388	5.1277	10.4774	5.3247
KWAZ	0.6009	3.7289	6.3767	5.4856	4.048025
PRET	2.6666	3.5549	4.1017	3.2287	3.387975
WCAP	1.0785	1.8961	0.4865	0.4024	0.965875
SMALL					
ECAP	5.2827	11.1244	15.517	9.0175	10.2354
JOBU	0.8871	3.7471	2.5432	9.296	4.11835
KWAZ	3.1231	8.457	2.9925	4.071	4.6609
PRET	2.8895	3.2949	0.2809	3.31	2.443825
WCAP	2.7179	0.8927	1.2679	2.1156	1.748525

Table: A6. SBVAR2

PROVINCES	HORIZON 1	2	3	4	AVERAGE
	LARGE				
ECAP	8.1346	28.9259	53.4731	33.8935	31.106775
JOBU	53.8578	65.2186	86.2647	126.237	82.894525
KWAZ	6.0048	17.2346	9.4617	9.5981	10.5748
PRET	17.4786	96.933	26.0426	8.0704	37.13115
WCAP	19.6173	3.3318	1.2938	9.909	8.537975
MEDIUM					
ECAP	134.4307	157.9479	16.647	4.0512	78.2692
JOBU	73.69	17.3448	47.3904	11.2008	37.4065
KWAZ	1.4714	1.8619	25.6476	36.2701	16.31275
PRET	35.0418	13.1685	30.8687	10.3808	22.36495
WCAP	29.7737	3.8096	6.0944	66.5002	26.544475
SMALL					
ECAP	19.477	22.1439	10.0257	21.8072	18.36345
JOBU	0.0463	1.4273	8.1476	13.739	5.84005
KWAZ	14.7464	3.2491	4.1831	6.8402	7.2547
PRET	16.4413	10.8168	0.4975	3.0555	7.702775
WCAP	5.4072	4.3506	3.7109	12.5032	6.492975

Table: A7. DFM

	HORIZON				
	1	2	3	4	**AVERAGE**
PROVINCES	**LARGE**				
ECAP	4.8165007	4.998561	5.193826	5.495563	5.12611289
JOBU	1.7741236	2.492483	2.402638	2.207314	2.21913958
KWAZ	4.6638863	5.145848	5.07201	4.878549	4.94007348
PRET	1.7914662	2.939627	1.939383	2.063738	2.18355355
WCAP	2.0844328	2.084066	2.563278	2.639538	2.34282873
MEDIUM					
ECAP	3.0095516	3.715013	3.253477	3.304407	3.32061216
JOBU	2.2333937	2.336845	2.390551	2.370381	2.33279262
KWAZ	2.9650327	3.761953	3.693652	3.712709	3.53333665
PRET	1.5771161	1.692194	1.708774	1.685557	1.66591036
WCAP	1.952452	1.994611	2.115071	2.205016	2.06678743
SMALL					
ECAP	4.5754356	4.984405	4.612619	4.579804	4.68806598
JOBU	2.3652301	2.292981	2.796089	3.253591	2.67697269
KWAZ	3.8410471	4.537948	4.086012	4.127189	4.14804915
PRET	2.2913425	2.734702	2.484763	2.331947	2.46068846
WCAP	2.0849579	2.359213	2.514553	2.48019	2.35972841

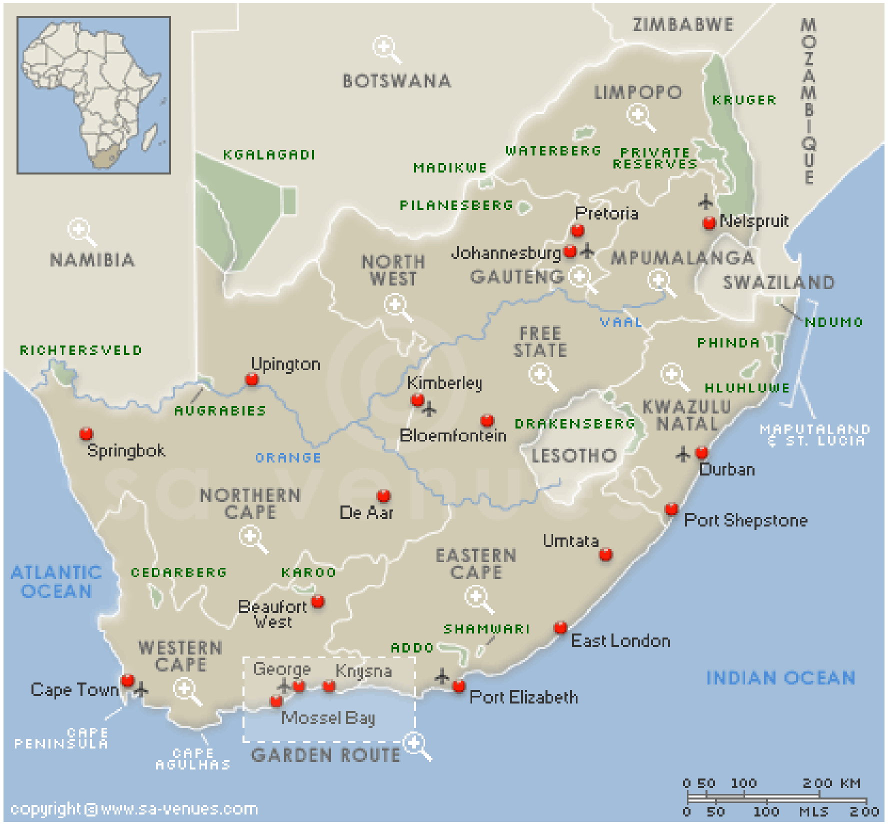

Figure 1. Provincial Map of South Africa
(**Source:** http://www.sa-venues.com/maps/south-africa-provinces.htm.)