University of Connecticut

Department of Economics Working Paper Series

Was the Recent Downturn in US GDP Predictable?

Mehmet Balcilar
Eastern Mediterranean University

Rangan Gupta
University of Pretoria

Anandamayee Majumdar
University of North Texas Health Science Center

Stephen M. Miller
University of Nevada-Las Vegas
University of Connecticut

Was the Recent Downturn in US GDP Predictable?

Mehmet Balcilar
Department of Economics
Eastern Mediterranean University
Famagusta, NORTHERN CYPRUS, via Mersin 10, TURKEY

Rangan Gupta
Department of Economics
University of Pretoria
Pretoria, 0002, SOUTH AFRICA

Anandamayee Majumdar
Department of Biostatistics
University of North Texas Health Science Center
School of Public Health
Fort Worth, Texas, 76107, USA

Stephen M. Miller*
Department of Economics,
University of Nevada, Las Vegas
Las Vegas, Nevada, 89154-6005 USA

Abstract:

This paper uses small set of variables-- real GDP, the inflation rate, and the short-term interest rate -- and a rich set of models -- athoeretical and theoretical, linear and nonlinear, as well as classical and Bayesian models -- to consider whether we could have predicted the recent downturn of the US real GDP. Comparing the performance by root mean squared errors of the models to the benchmark random-walk model, the two theoretical models, especially the nonlinear model, perform well on the average across all forecast horizons in out-of-sample forecasts, although at specific forecast horizons certain nonlinear athoeretical models perform the best. The nonlinear theoretical model also dominates in our *ex ante* forecast of the Great Recession, suggesting that developing forward-looking, microfounded, nonlinear, dynamic-stochastic-general-equilibrium models of the economy, may prove crucial in forecasting turning points.

Keywords: Forecasting, Linear and non-linear models, Great Recession

JEL classification: C32, E37

* Corresponding author

1.   Introduction

This paper analyzes whether researchers could have predicted the recent downturn of the US real Gross Domestic Product (GDP), using a small set of variables and a rich set of models. We estimate a wide range of econometric models that include atheoretical linear models (classical and Bayesian vector autoregressive models), atheoretical nonlinear models (time-varying parameter, Markov-switching, smooth transition vector autoregressive, and artificial neural network models), nonparametric and semi-parametric atheoretical models, and linear and nonlinear micro-founded theoretical models [Dynamic Stochastic General Equilibrium (DSGE) models based on Kalman and particle filters]. Our restricted data set includes real GDP, the rate of inflation of the GDP implicit deflator, and the three-month Treasury-bill rate.

Although researchers widely use these alternative models to predict key macroeconomic variables,[1] this paper brings these models together to compare their forecasting ability simultaneously and, hence, covers the entire spectrum of currently popular forecasting methods ranging from linear and nonlinear (with known and unknown functional forms) theoretical and atheoretical models.[2] In addition, the studies cited in footnote 2 and others, barring the analyses that focus on time-varying vector autoregressive models and Camacho's (2004) paper, who

---

[1] See, for example, Stock and Watson (1996), Clements and Smith (1997), Clements and Krolzig (1998), Clements, *et al*. (2003), van Dijk and Franses (2003), Camacho (2004), Korobilis (2006), Altavilla and Ciccarelli (2007), Del Negro, *et al*. (2007), Marcellino (2007), Pichler (2008), D' Agostino, *et al*. (2011), Arora, *et al*. (2011), Gupta *et al.* (2011), Korobilis (2011), Del Negro and Schorfheide (2012), Koop and Korobilis (2012), Bekiros and Paccagnini (2012), Balcilar *et al.* (forthcoming), and the references cited in these papers.

[2] This paper ignores one line of forecasting and excludes models involving large data sets and estimated using factors or Bayesian shrinkage with either constant or time-varying parameters. We omit such models because we want to ensure that all our models use the same information set on the three variables. For a detailed discussion of such models, see Barnett, *et al*. (2012). Another related area of research involves incorporating information from large data sets into Dynamic Stochastic General Equilibrium (DSGE) by adding dynamic factors in DSGE-Dynamic Factor Models (DSGE-DFM). See, for example, Consolo, *et al.* (2009); Paccagnini (2011), and the references cited therein for further details. We also ignore these models to do our analysis over a common three-variable data set.

considers the role of leading indicators in forecasting US GDP growth using multivariate parametric nonlinear models, use univariate versions of the nonlinear parametric, nonparametric, and semi-parametric models.[3] More importantly, no study, barring Del Negro and Schorfheide (2012), focus on predicting downturn(s) in the economy *ex ante*.[4]

We focus on the most recent financial crisis and Great Recession, as it proved atypical of post-WWII recessions. That is, financial-crisis-induced recessions exhibit more depth and length than typical recessions (Reinhart and Rogoff, 2009). Reinhart and Rogoff (2009) examine the responses of 18 countries to financial crisis.[5] They characterize the typical recession emanating from a financial crisis as entailing, on average, (i) a fall in real GDP of 9.3 percent over 2.9 years; (ii) an average unemployment rate above 7 percent for 4.8 years; (iii) a fall in stock prices of 55.9 percent over 3.4 years; and (iv) a fall in real housing prices of 35.5 percent over 6 years. The recent US financial crisis and Great Recession experienced to date (i) a fall of 4.7 percent in real GDP over 1¾ years; (ii) an average unemployment rate of 8.3 percent over the last 4 2/3 years; (iii) a fall in stock prices of 53.8 percent over 1 year and 5 months; and, finally, (iv) a fall of 32.4 percent in housing prices over the last 5 years.[6]

---

[3] Another notable exception is Balcilar *et al.,* (forthcoming), where the authors use both multivariate parametric nonlinear and multivariate nonparametric models to forecast out of sample as well as to predict ex ante the gross gaming revenue and taxable sales for the state of Nevada.

[4] Del Negro and Schorfheide (2012) perform a post-mortem of DSGE model forecasts of the Great Recession, showing that forecasts from a version of the Smets and Wouters (2003) model augmented by financial frictions and with interest rate spreads as an observable compare favorably to the Blue Chip forecasts in predicting the GDP growth rate.

[5] The 18 countries include what they call the big five developed countries (Spain 1977, Norway 1987, Finland 1991, Sweden 1991, and Japan 1992), the countries directly involved in the 1997–1998 Asian crisis (Hong Kong, Indonesia, Korea, Malaysia, the Philippines, and Thailand), Colombia 1998, Argentina 2001, two earlier episodes where we know housing prices (Norway 1899 and United States 1929), and three on-going (at the time) crises in Iceland, Ireland, and Hungary. The averages reported in the text exclude the on-going crises.

[6] The decrease in real GDP transpired from 2007:Q4 through 2009:Q2. We averaged the unemployment rate over December 2007 through August 2012. As an alternative, the average from December 2008 through August 2012

Our analysis unfolds through the following steps. First, we estimate the wide range of econometric models noted previously over the 1979:Q3 to 1999:Q2 period, using data on the detrended logged real GDP, quarter-on-quarter inflation based on the GDP deflator, and the three-month Treasury-bill rate. Second, we forecast one- to eight quarters-ahead of detrended logged real GDP over an out-of-sample forecast horizon from 1999:Q3 to 2006:Q4, estimating each of these models recursively over the out-of-sample horizon. Third, we choose the model within each category (theoretical, linear and nonlinear versions of atheoretical, and semi-parametric or nonparametric versions of atheoretical models) that produces the minimum average root mean square errors (RMSEs) relative to the benchmark random walk model as the "best" model for a specific category. Fourth, we use the best model within each category to forecast the level of logged real GDP *ex ante* (without updating the parameter estimates of the optimal models) over 2007:Q1 to 2011:Q2, adding the estimate of the trend for the logged real GDP at 2006:Q4 to the forecasts of the detrended GDP in logs over this period.

The choice of the in-sample and out-of-sample periods emanate from the following considerations. First, the in-sample excludes the pre Volcker-Greenspan-Bernanke era. Second, the endpoint of the in-sample gives us 80 observations, which An and Schorfheide (2007) regard as a realistic in-sample size. Third, since we adopt Pichler's (2008) linear and nonlinear DSGE models, we also adopt his in-sample period and choices and transformations of the variables in the models. Fourth, the end point of the out-of-sample period occurs one year in advance of the starting date of the downturn. Thus, our *ex ante* forecasts begin one year before the beginning of

yields 9.0 percent. The decrease in stock prices occurred from October 9, 2007 through March 9, 2009. Finally, the housing price decrease occurred from 2006:Q1 through 2011:Q1. As an alternative, housing prices decreased by 30.7 percent from 2006:Q1 through 2009:Q1, or 3 years.

the recession and extend beyond the end of the recession through 2011:Q2, which was the last observation available when we constructed our data set.

To make our analysis realistic, we carry out the out-of-sample forecasting exercise based on the vintage of these variables available on October 27, 2006, which corresponds to the first release of GDP estimates for 2006:Q3. The trend estimate of logged real GDP that we use to detrend the data over the in-sample comes from the vintage of January 31, 2007. Once we forecast the detrended logged real GDP from the different "best" models, we add back this estimate of the trend for the period of 1979:Q3 to 2006:Q4 to obtain a forecast for the logged real GDP in log-levels, and compare this with the actual values of the logged real GDP available on July 27, 2012, which corresponds to the first release of GDP estimates for 2012:Q2. Since the data for the GDP implicit deflator depends on nominal and real GDP estimates, we follow a similar approach with the nominal GDP data over the period of 1979:Q3 to 2006:Q4. Since the three-month Treasury bill rate is available at a weekly frequency, we take the averages of the weekly values over each quarter to generate our quarterly series. Also, the estimate for the quarterly Treasury bill rates for 2006:Q3 and 2006:Q4 use the vintages for October 2, 2006 and January 3, 2007, respectively. Finally, we compute the RMSEs, using the vintage dates mentioned above that corresponds to the first release of the three variables for 2006:Q4.

The rest of the paper is organized as follows: Section 2 outlines the basics of the different models used for the forecasting exercise. Section 3 discusses the data and presents the forecasting results and the *ex ante* prediction of the real US GDP. Finally, Section 4 concludes.

2.   Models:

This section describes the atheoretical linear models, atheoretical nonlinear models, nonparametric and semi-parametric atheoretical models, and linear and nonlinear micro-founded theoretical models.

Atheoretical Linear Models:

The Vector Autoregressive (VAR) model, though 'atheoretical,' is particularly useful for forecasting purposes.[7] An important drawback of VAR models is that they require the estimation of many parameters, where many are potentially insignificant. This problem of over-parameterization, resulting in multicollinearity and loss of degrees of freedom, leads to inefficient estimates and large out-of-sample forecasting errors. One solution, often adopted, simply excludes the insignificant lags based on statistical tests. Another approach uses near VAR models, which specify unequal number of lags for the different equations.

An alternative approach to overcoming over-parameterization, as described in Litterman (1981), Doan *et al.* (1984), Todd (1984), Litterman (1986), and Spencer (1993), uses a Bayesian VAR (BVAR) model. Instead of eliminating longer lags, the Bayesian method imposes restrictions on the model's coefficients by assuming that these coefficients more likely approach zero than the coefficients on shorter lags. If strong effects from less important variables exist, the data can override this assumption. The researcher imposes restrictions by specifying normal prior distributions with zero means and small standard deviations for all coefficients with the standard deviations decreasing as the lag length increases. The researcher sets the coefficient on the first own lag of a variable equal to unity, unless the variable is mean reverting or stationary. Generally, following Litterman (1981), the constant exhibits a diffuse prior. This specification of

---

[7] Refer to Korobilis (2011) for further details.

the BVAR prior is popularly called the 'Minnesota prior' due to its development at the University of Minnesota and the Federal Reserve Bank at Minneapolis.

In addition to the shrinkage approach of the Minnesota-type BVAR models, numerous other efficient methods exist to prevent the proliferation of parameters and eliminate parameter or model uncertainty [e.g., variable selection priors (George *et al*. 2008), steady state priors (Villani, 2009), Bayesian model averaging (Andersson and Karlsson, 2008), and factor models (Stock and Watson, 2005)]. Following Korobilis (2011), we compare the forecasting performances of the classical and the Minnesota-type BVAR models with those of linear (fixed-parameter) and nonlinear (time-varying parameter [TVP]) VARs involving a stochastic search algorithm for variable selection, estimated using Markov Chain Monte Carlo (MCMC) methods. The "stochastic search" approach means that when the model space is too large to assess in a deterministic manner, the algorithm looks for only the most probable models. Two main benefits occur from using this approach over shrinkage method. First, variable selection occurs automatically. That is, along with the parameter estimates, we also see the associated probabilities of inclusion of each parameter in the "best" model. This allows one to select among all possible VAR models without estimating each and every one of these models. Second, this Bayesian variable selection procedure does not depend on the prior assumptions about the parameters. The decision to use the stochastic search variable selection algorithm proposed by Korobilis (2011) over other available ones, such as those developed by George *et al*. (2008) or Korobilis (2008), reflects our ability to apply the current algorithm to variable selection non-linear (time-varying) VAR models.

We can represent a reduced form VAR using following linear regression specification:

$$Y_{t+1} = Bx_t + \varepsilon_{t+1} \tag{1}$$

where Y_{t+1} denotes an (m x 1) vector of dependent variables (i.e., detrended natural logarithm of real GDP, the GDP deflator inflation rate, and the Treasury bill rate) from time $t = 1, ..., T$; x_t denotes a (k x 1) vector, which may include lags of the dependent variables, intercepts, dummies, trends, and exogenous regressors; B denotes an (m x k) vector of VAR coefficients; and $\varepsilon_t \sim N(0,\Sigma)$, where Σ denotes a ($m \times m$) covariance matrix.

We can rewrite equation (1) as a system of seemingly unrelated regressions (SURs) as follows, where different equations in the VAR can include different explanatory variables:

$$Y_{t+1} = z_t \beta + \varepsilon_{t+1} \tag{2}$$

where Y_{t+1} and ε_t are defined in equation (1); $z_t = I_m \otimes x_t'$ is a (m x n) matrix vector; and $\beta = vec(B)$ is an (nx1) matrix. When no parameter restrictions exist, equation (2) is an unrestricted VAR model. We embed Bayesian variable selection, therefore, in equation (2) by an indicator variables: $\gamma = (\gamma_1,...,\gamma_n)$ such that $\beta_i = 0$ if $\gamma_i = 0$ and $\beta_i \neq 0$ if $\gamma_i = 1$. We treat the indicator variables as random variables by assigning a prior on them and allowing the data likelihood to determine their posterior values. We explicitly insert these indicator variables multiplicatively into the VAR model using the form:

$$Y_{t+1} = z_t \theta + \varepsilon_t \tag{3}$$

where $\theta = \Gamma\beta$; Γ denotes an (n x n) diagonal matrix with $\Gamma_{jj} = \gamma_j$ ($j=1,2....,n$) elements on its main diagonal; and for $\Gamma_{jj} = 0$, $\theta_j = \Gamma_{jj}\beta_j = 0$, where θ_j is restricted while for $\Gamma_{jj} = 1$, $\theta_j = \Gamma_{jj}\beta_j = \beta_j$ so that we can explore all possible 2^n specifications and variable selection is equivalent to model selection in this case. We use Gibbs sampling to estimate these parameters

by conditioning on the data and Γ. Adopting the independent Normal-Wishart prior produces standard-form densities of β and Σ.

The restriction indices γ add one more block to the Gibbs sampler of the unrestricted VAR model. If needed for the restriction indicators, we sample the n element in the column vector $\gamma = (\gamma_1,....\gamma_n)'$ and we recover the diagonal matrix $\Gamma = diag\{\gamma_1.....,\gamma_n\}$. Simplified derivations emerge, however, if the indicators γ_j do not depend on each other. We can define the priors as follows:

$$\beta \sim N_n(b_0, V_0); \tag{4}$$

$$\gamma_j \mid \gamma_{\backslash-j} \sim Bernoulli(1, \pi_{0j}); \text{ and} \tag{5}$$

$$\Sigma^{-1} \sim Wishart(\alpha, S^{-1}), \tag{6}$$

where b_0 is *(n x 1)* vector, V_0 is *(n x n)* matrix, $\pi_0 = (\pi'_{01},......,\pi'_{0n})$ is *(n x 1)* vector, Ω is *(m x m)* matrix, and α is a scalar.

We argue that this form of variable selection may apply to many non-linear extensions of the VAR as compared to stochastic variable selection algorithms for VAR models. Adopting variable selection in a TVP-VAR model, therefore, simply extends the VAR model with constant parameters, where we replace equation (7) below with equation (3) and variables are as explained in equation (3) explains the variables while equations (4) through (6) identify the priors, except now $\beta_0 \sim N_n(b_0, V_0)$.

Specifically, we use the following priors for the restricted VAR with variable selection (VAR-VS): $\gamma_j / \gamma_{\backslash-j} : Bernouilli(1, 0.5)$ for all $j = 1,...,n,$ and $\beta_j \sim N(0, 10^2)$, if β_j is an intercept, and $\beta_j \sim N(0, 3^3)$, otherwise. For the benchmark VAR, we adopt the same priors as

the VAR-VS, except that we restrict $\gamma_j = 1$ for all j. For the BVAR based on the Minnesota prior

(VAR_MIN), the means and variances of the Minnesota prior for β take the form

$\beta \sim N(b^{min}, V^{min})$ where $V_{i,l}^{min} = g_1/p^2$ and $g_3 \times s_i^2$ applying to parameters on own lags and

for intercepts respectively, while $V_{i,l}^{min} = (g_2 \times s_i^2)/(s_l^2 \times p^2)$ is for parameters j on variable

$l \neq i, l, i = 1, ..., m.$ s_i^2 is the residual variance from the p-lag univariate autoregression for variable

i. After experimenting to produce the best possible forecast, we set the hyperparameters to the

following values: $g_1 = 0.01$, $g_2 = 0.0025$, and $g_3 = 100$. Since we transform the variable used in

the forecasting exercise to induce stationarity, we set the prior mean vector b^{min} equal to zero

for parameters on the lags of all variables, including the first own lag (Banbura *et al*. 2010). The

forecasts from the VAR models with and without variable selection, as well as the Minnesota

prior run 30,000 draws from the posterior, discarding the first 2,000 draws. Also, we set the lags

in these models to 2, determined by the Bayesian information criterion (BIC).

Atheoretical Nonlinear Models:

Macroeconomic time series contain structural breaks due to major political and economic

changes. That is, changes in the economic policy, financial and economic crises, and business

cycle shocks create breaks in the observed series. The structural breaks contribute significantly to

the forecasting failures of macroeconomic time series. No consensus exists, however, among the

macroeconomists about the significance and relevance of structural breaks and whether

forecasters need to go beyond traditional models and construct models that include structural

breaks or regime switches. See, for example, Cogley and Sargent (2001) and Sims (2001).

Neftci (1984) notes that dynamic properties of major economic time series differ over time, particularly across the different phases of the business cycles. Neftci reports that recessions generate much sharper declines than increases during expansions, implying asymmetric adjustment. Hamilton (1989) also argues that the US gross national product (GNP) exhibits different dynamics during episodes of slower growth and faster growth. He emphasizes that the traditional linear models fail to capture such shifts in the dynamic behavior of the series and forecasters need nonlinear models to capture departures from linearity. Neftci (1984) and Hamilton (1989) argue that regime switches occur systematically enough to enter the probabilistic structure of a model. That is, Neftci (1984) states that the systematic switches occur frequently enough such that the data can endogenously capture the dynamic behavior across the regimes, which different markedly in their responses.

Clements and Hendry (1999) develop a theory of forecasting in the presence of deterministic structural breaks. For recurrent, systematic breaks that affect not only the mean growth rate but also the dynamics of the underlying series, however, a model that incorporates such time varying dynamics and models with endogenous switches may produce superior forecasts to linear models with a fixed dynamic process (i.e., regime-switching models). Regime switching models that can generate sharp drops during recessions and slow gradual growth during expansions will naturally fit the data better around turning points (Neftci 1984). We consider four classes of nonlinear multivariate models that successfully model regime-switching time series -- Markov-switching (Neftci 1984; Hamilton, 1989; Krolzig, 1997), smooth transition (Granger and Terasvirta 1993; Terasvirta 1998; Potter, 1999), artificial neural network (McCulloch and Pitts 1943; Minsky and Papert 1969; White 1988, 1989), and time-varying

parameter vector autoregressive models (D'Agostino *et al.* 2011; Korobilis, 2011; Koop and Korobilis, 2012).

Lucas (1977) emphasizes the co-movement of macroeconomic variables such as the output, inflation, unemployment, interest rates, consumption, investment, and so on. Diebold and Rudebusch (1996) identify two important features of business cycle models: (1) co-movement of economic variables and (2) persistence of the states. In the following section, we consider only dynamic nonlinear models that involve both features. A univariate model can possess persistence of states with or without dynamic structures, but not co-movement of economic variables. To permit further dynamic adjustment through autocorrelations, we consider dynamic multivariate nonlinear models for forecasting the US real GDP series. Empirically, one expects switches between a normal regime (expansion) and a recession regime will lead to different (asymmetric) dynamic links between the macroeconomic variables that co-move, which nonlinear models can capture more accurately.

Markov-Switching Vector Autoregressive Model. Markov-switching (MS) models are one of the most popular nonlinear models for the analysis of regime switching time series such as the business cycle. Structural change models admit only occasional, exogenous changes and, thus, structural breaks or regime shifts are deterministic. MS models, on the other hand, include a endogenous mechanism, where an unobservable state variable controls the switches through a first-order Markov chain. By allowing switches between the states with different parameter sets, which may involve parameters on variable lags, MS models can capture complex dynamic structures.

A number of studies successfully use Markov-switching models to analyze aggregate output and business cycles (e.g., Hamilton 1989; Lam 1990; Goodwin 1993; Diebold, *et al*.

1994; Durland and McCurdy 1994; Filardo 1994, Ghysels 1994; Kim and Yoo 1995; Filardo and Gordon 1998; and Kim and Nelson 1998). Following Hamilton (1989), these studies employ various forms of univariate MS autoregressive (MS-AR) models. Krolzig (1997) further generalizes the MS-AR model to vector-autoregressive (VAR) time-series models, a most extensively used nonlinear multivariate model dynamic. Krolzig (1997) generalizes the linear vector autoregressive (VAR) models of Hamilton (1989) by allowing parameters of the VAR model to switch across regimes in response to a hidden first-order Markov chain.

Recognizing how nonlinearity and asymmetry importantly affect the VAR forecasting model, we employ a Markov-switching vector autoregressive (MS-VAR) model to address the nonlinearity in an explicit and formal way. The MS-VAR models fit the data better than their linear counterpart VAR models. The superior in-sample fit does not usually generalize to superior forecasting performance, as noted by Clements and Krolzig (1998), Dacco and Satchell (1999), and Krolzig (2000). We examine whether the superior fit generalizes to superior forecasting performance and if it does not, then where does the MS-VAR model fail.

The real GDP series that we forecast frequently switches between contractions and expansions. The other two series, the inflation and Treasury bill rates, also co-move with the real GDP series. MS-VAR models, which are nonlinear with regime dependent parameters, can naturally study whether such regime switching better represents the data generating process (DGP). The MS-VAR model that we use not only allows regime shifts in the detrended logged real GDP but also fully allows all parameters of the VAR model and its variance matrix to depend on the regime.

Let y_t denote real detrended logged GDP, r_t denote the Treasury bill rate, and π_t denote the inflation rate. Define the time-series vector X_t up to and including period t as

$X_t = [y_t, r_t, \pi_t]'$ and let $\mathfrak{I}_t = \{X_\tau | \tau = t, t-1, ..., 1-p)$, where p is a nonnegative integer. For the vector valued time series X_t of random variables, assume that there exists a density (probability) function $f(X_t | \mathfrak{I}_{t-1}, \theta)$ for each $t \in \{1, 2, ..., T\}$. The parameters and the parameter space are denoted by θ and Θ, respectively. The true value of θ is denoted by $\theta_0 \in \Theta$. Let the stochastic variable $S_t \in \{1, 2, \rightleftharpoons, q\}$ follow a Markov chain with q states. Then we can write the MS-VAR model that allows all parameters to depend on the S_t regime as follows:

$$X_t = \mu_{S_t} + \sum_{k=1}^{p} \Phi_{S_t}^{(k)} X_{t-k} + \varepsilon_t,\tag{7}$$

where p is the order of the MS-VAR model, $[\varepsilon_t | S_t \sim N(0, \Omega_{S_t})]$, and Ω_{S_t} is positive definite. The random state or regime variable S_t, conditional on S_{t-1}, is unobserved, independent of past Xs, and assumed to follow a q-state Markov process. In other words, $\Pr[S_t = j | S_{t-1} = i, S_{t-2} = h_2, \rightleftharpoons, \mathfrak{I}_{t-1}] = p_{ij}$, for all t and h_l, regimes $i, j = 1, 2, ..., q$, and $l \geq 2$. More precisely S_t follows a q state Markov process with transition matrix given by

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1q} \\ \vdots & \vdots & & \vdots \\ p_{q1} & p_{q2} & \cdots & p_{qq} \end{bmatrix}, \qquad \sum_{j=1}^{q} p_{ij} = 1.\tag{8}$$

Thus, p_{ij} is the probability of being in regime j at time t, given that the economy was in regime i at time $(t-1)$, where i and j take possible values in $\{1, 2, ..., q\}$. We estimate the parameters of Markov-switching model via the expectation maximization (EM) algorithm (Lindgren 1978; Hamilton 1990, 1994) assuming that the conditional distribution of X_t given $\{\mathfrak{I}_t, S_t, S_{t-1}, \rightleftharpoons, S_0; \theta'\}$ is normal. We numerically approximate the likelihood function using the EM algorithm and we find the ML estimates using the Broyden–Fletcher–Goldfarb–Shano

(BFGS) optimization algorithm. We compute the asymptotic standard errors of estimates from the inverse of the empirical Hessian. The order of the MS-VAR model for each case is 2 according to the BIC.

The empirical procedure for building suitable MS-VAR models starts with identifying a possible set of models to consider. We determine the order p of the MS-VAR model using the BIC in a linear VAR(p) model. The models differ in terms regime numbers (q) and the variance matrix specification. We consider both regime-independent variance models, MS(q)-VAR(p), and regime-dependent variance (heteroscedastic) models, MSH(q)-VAR(p). Once a specific MS model is estimated, we next test for the presence of nonlinearities in the data. When testing the MS model against the static (linear) alternative, or a q regime model against a (m-1) regime model, we follow Ang and Bekaert (2002) and use the likelihood-ratio statistic (LR), which is approximately $\chi^2(q)$ distributed, where q equals the number of restrictions plus the nuisance parameters (i.e., free transition probabilities) that are not identified under the null. We use p-values based on the conventional χ^2 distribution with q degrees of freedom and also for the approximate upper bound for the significance level of the LR statistic as derived by Davies (1987). Once we establish nonlinearity, we choose the number of regimes and the type of the MS model based on both the likelihood-ratio statistic and the Akaike information Criterion (AIC).[8] The procedure selects a heteroscedastic model with all parameters in the VAR being regime dependent based on both the Davies (1987) test and AIC. We maintain the MSH(2)-VAR(2) afterwards.

---

[8] Krolzig (1997) and Psaradakis and Spagnolo (2003) suggest selecting the number of regimes and the MS model using AIC, and using Monte Carlo experiment Psaradakis and Spagnolo (2003) show that AIC generally yields better results in selecting the correct model.

Computing multi-step forecasts from MS-VAR models as well as all nonlinear time series models prove complicated because no ordinary method of computing the future path of the process exists. Good forecasts require that we know the future path of the process, since the forecasts depends on the regime. Using Monte Carlo simulations, Clements and Smith (1999) and Pesaran and Potter (1997) show that the forecasting performance of the regime-switching models do depend on the regime at the time of the forecast. We cannot resolve the dependence of the forecasts on the future path of the process cannot by simply substituting the expected values of the future shock into the conditional mean function of the model. The problem is usually solved by Monte Carlo or Bootstrap simulation techniques. Although MS-VAR models are nonlinear, a number of authors (Krolzig, 2000; Clements and Smith, 1999; Pesaran and Potter, 1997) note that analytical formula for forecasting from these models exist, at least in simple cases. In this paper, we use the method proposed in Krolzig (2000) to obtain the multi-step forecasts from the MS-VAR model.

Smooth Transition Autoregressive Model Identification. Recent empirical studies show that smooth–transition-autoregressive (STAR) models can successfully model economic time series that move smoothly between two or more regimes (e.g., recession to expansion). When considering the joint dynamic properties of the real detrended logged GDP, the inflation rate, and the interest rate, it is natural to consider vector STAR (VSTAR) models. Van Dijk, *et al*. (2002), among many others, discuss VSTAR models. Montgomery, *et al*. (1998) and Marcellino (2002) report favorable forecasting performance for LSTAR forecasts, while Stock and Watson (1999) show that linear models generally dominate nonlinear models in terms of forecasting performance. Despite specification difficulties, such as the appropriate transition variable, the number of regimes, the type of transition function, and so on, STAR models prove useful for

state dependent multivariate relationships. Recent applications (e.g., Rothman, *et al.* 2001; Psaradakis, *et al.* 2005; Tsay 1998; De Gooijer and Vidiella-i-Anguera 2004) find that VSTAR models successfully model nonlinear economic time-series data.

Define $X_t = (x_{1t}, x_{2t}, \cdots, x_{nt})'$ as a $(k \times 1)$ time-series vector. In our case X_t is defined as $X_t = [y_t, r_t, \pi_t]'$ and a (3×1) time-series vector. We specify the k-dimensional VSTAR model as follows:

$$X_t = (\Theta_{1,0} + \sum_{j=1}^{p} \Theta_{1,j} X_{t-j}) + (\Theta_{2,0} + \sum_{j=1}^{p} \Theta_{2,j} X_{t-j}) G(s_t; \gamma, c) + \varepsilon_t, \qquad (9)$$

where $\Theta_{i,0}$, $i = 1,2$, are $(k \times 1)$ vectors, $\Theta_{i,j}$, $i = 1,2$, $j = 1,2,\cdots,p$, are $(k \times k)$ matrices, and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{kt})$ is a k-dimensional vector of white noise processes with zero mean and nonsingular covariance matrix Ω, $G(\cdot)$ is the transition function that controls smooth moves between the two regimes, and s_t is the transition variable.

The VSTAR model in equation (8) defines for two regimes, one associated with $G(s_t; \gamma, c) = 0$ and another associated with $G(s_t; \gamma, c) = 1$. The transition from one regime to the other occurs smoothly, depending on the shape of the $G(\cdot)$ function. In this paper, we consider a logistic transition function

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(s_t - c)/\hat{\sigma}_s\}}, \qquad \gamma > 0, \qquad (10)$$

where $\hat{\sigma}_s$ is the estimate of the standard deviation of transition variable s_t. The threshold parameter c determines the midpoint between two regimes at $G(c; \gamma, c) = 0.5$. The parameter γ determines the speed of transition between the regimes with higher values corresponding to faster transition.

16

To specify the VSTAR model, we follow the procedure presented in Terasvirta (1998) (see, also, Van Dijk, *et al*. 2002; Lundbergh and Terasvirta 2002). First, we specify the lag order of $p = 2$.

Second, we test linearity against the VSTAR alternative. Since the VSTAR model contains parameters not identified under the alternative, we follow the approach of Luukkonen *et al*. (1988) and replace the transition function $G(\cdot)$ with a suitable Taylor approximation to overcome the nuisance parameter problem. The testing procedure selects a logistic VSTAR model with a single threshold, which we maintain for the univariate case as well.

Third, we select the transition variable s_t. To identify the appropriate transition variable, we run the linearity tests for several candidates, $s_{1t}, s_{2t}, \rightleftharpoons, s_{mt}$, and select the one that gives the smallest *p*-value for the test statistic. Here, we consider lagged values of all variables for lags 1 to 2 as the candidate transition variable. Let $s_t = x_{i,t-d}$, where x is any of the three variables $\{y, r, \pi\}$. We test linearity with these variables for delays $d = 1, 2$. We obtain the smallest *p*-value with $s_t = \pi_{t-d}$ and $d = 1$. Analytical point forecasts do not exist for non-linear (V)AR models with a Gaussian disturbance term even when $h \geq 2$, as $E[f(x)] \neq f[E(x)]$, where h is the number of steps-ahead for the forecasts.[9]

<u>Artificial Neural Network Model Identification.</u> Artificial–neural-network (ANN) models (McCulloch and Pitts, 1943; Minsky and Papert, 1969) perform well in forecasting nonlinear and chaotic time series (Lachtermacher and Fuller 1995). As for the STAR models, we consider only

---

[9] Details of the bootstrapping procedure are available upon request from the authors. We implement all computations of the STAR models with the RSTAR package (Version 0.1-1) in R developed by the one of the authors of this paper.

multivariate autoregressive ANN (MAR-ANN) models. Lisi and Schiavo (1999) use ANN models for predicting European exchange rates, finding that they perform as well as the best model, a chaos model. Using statistical tests, Lisi and Schiavo (1999) discover no significant difference between the ANN and chaos models. Stern (1996) applies ANN models to several simulated data from autoregressive models of order 2, AR(2), with various signal-to-noise ratios. The results show that ANN models do not generate good predictions with a small signal-to-noise ratio. Thus, ANN models seem most suitable for forecasting time series with small signal-to-noise ratios, given sufficient data and appropriate data transformations. Success of ANN models in forecasting nonlinear time series reflects their universal function approximation capability (White 1988, 1989). This includes any linear or nonlinear function (Cybenko 1989; Funahashi 1989; Hornik, *et al*. 1989; and Wasserman 1989). Because of this approximation capacity, neural network models offer several potential advantages over alternative methods for non-normal and non-linear data (Hansen, *et al*. 1999).

Researchers use a variety of neural-network architectures for time-series prediction. The most widely used architecture is the multilayer perceptron (MLP) (also known as a feed-forward neural network) (Sarle 2002). The MLP can resolve a wide variety of problems (Bishop 1995; Kaastra and Boyd 1996). In this paper, we also prefer the MLP network for VAR-ANN based forecasting. In an MLP network, we partition units into layers. Usually, the MLP network contains an input and an output layer, and one or more hidden layers of neurons between the input and output layers. In the MLP architecture, data always transmit from the input layer to the output layer. In our case, each input neuron represents one of the lagged values, while the output neuron(s) represent dependent variable(s) or MLP network forecasts. The MLP network links each unit in the k-th layer directly and only to units in the $(k + 1)$-st layer. In VAR-ANN models,

the lags of variables enter as inputs to the first layer, and outputs from the network appear in the last layer. A weight ("connection strength") associate with each link, and a network is trained ("learned") by modifying these weights, thereby modifying the network function that maps inputs to outputs. We use the VAR-ANN model with q-hidden layers, which we write as follows:

$$X_t = \sum_{i=1}^{q} \beta_i' G(\Theta_{i,0} + \sum_{j=1}^{p} \Theta_{i,j} X_{t-j}) + \varepsilon_t, \tag{11}$$

where X_t is as before, β_i, $i=1,2,\cdots,q$, are parameters called weights or connection strengths, $\Theta_{i,0}$, $i=1,2,\cdots,q$, are $(k \times 1)$ vectors, $\Theta_{i,j}$, $i=1,2,\cdots,q$, $j=1,2,\cdots,p$, are $(k \times k)$ matrices, $G(\cdot)$ is the "squashing (activation) function" called the "hidden unit", and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{kt})$ is a k-dimensional vector of white-noise processes with zero mean and nonsingular covariance matrix Ω.

When using ANNs for forecasting time series, researchers usually subdivide the sample into three sets -- training, validation, and test sets (Bishop 1995; Ripley 1996). The training set constructs the network, the validation set obtains forecast performance measures, and the test set checks for generalization capacity of the network. This method can usefully construct networks with good generalization capability that performs well with new cases. During the network's training stage, the weights iteratively adjust, using an algorithm such as the back propagation of Rumelhart *et al*. (1986), on the basis of the training set's values to minimize the error between the network's predicted output and the actual (desired) output. We use sum-of-squared errors (SSE) as a criterion to determine the optimal weights based on the training set. Nevertheless, training the ANN model using the training set may lead to overfitting. To avoid overfitting, the validation set controls the learning process. We evaluate an ANN model's performance by

changing the number of hidden layers and type of activation function at hidden and output layers, using the mean squared error (MSE) obtained from the trained ANN forecasts in the validation set. Finally, the test set, which does not depend on the data set, provides an unbiased estimate of the generalization error or forecasting performance.

No optimal rules exist to select the size of each data set, although by general agreement, the training set is the largest. In this paper, we use data from 1979:Q3 to 1994:Q2 as the training set (60 observations, 55%), data from 1994:Q3 to 1999:Q2 as the validation set (20 observations, 18%), and data from 1999:Q3 to 2006:Q4 as the test set (30 observations, 37%). We evaluate a network's performance based on the 1-to-8 step-ahead forecasts in the validation period and we select the best performing network based on the minimum MSE. Then, we use the validation set to select the network used to forecast the test period.[10] The *ex-ante* forecast over the 2007:Q1 to 2012:Q2 maintains the same ANN architecture, but extends the in-sample period to 2006:Q4.

Creating an MLP network involves five sets of parameters: the learning rule, network architecture, learning-rate and momentum factor, activation function of the hidden and output layers, and number of iterations. Over the years, researchers develop many methods to train an ANN model (see Fine 1999). MacKay (1992) proposes a Bayesian framework, called the Bayesian regularization, to overcome the problems in interpolation of noisy data. Bayesian regularization facilitates the selection of parsimonious models as well as maximum likelihood estimation. It also advantageously expands the cost function to search not only for the minimal error, but also for the minimal error using the minimal weights. In the Bayesian regularization approach, one determines a set of smaller models nested within a larger model and the algorithm

---

[10] See Section 5 for further details.

chooses one of these smaller models, providing a method to select parsimonious models. The procedure first assigns prior probabilities to each of the smaller models and then determines the model that posts the highest posterior probability. Following the recommendation in Foresee and Hagan (1997), we fit the models using the Levenberg–Marquardt algorithm.

In this paper, the MLP architecture uses three lags of each variable as inputs for the VAR-ANN model. An MLP network's capacity to learn depends on the number of hidden neurons. Despite its significant role, no statistical criteria exist to select the optimum number of hidden neurons. We select the best ANN with Bayesian regularization, bearing in mind the overfitting issue, based on its MSE in the validation set, using the least possible number of hidden neurons (Masters 1993; Smith 1993; Rzempoluck 1998). We try ANN models with maximum q set to 9. We obtain the best performing VAR-ANN model with $q = 2$.

In our study, the input layer neurons use a linear activation function, while the hidden and output layer neurons use a sigmoid activation function, $G(\cdot)$. Two sigmoid functions widely used in MLP are the logistic (providing continuous values between 0 and 1) and hyperbolic tangent sigmoid, called tansig, functions (providing continuous values between -1 and 1). In this study, we use the tansig function in the hidden and output layers of the MLP networks, since it allows much faster learning in comparison to the logistic function (Fahlman 1988; Fausett 1994). We scale our data onto -1 and 1, which is the range covered by the tansig function.

The learning-rate parameter plays a crucial role in the training process of MLP networks. The learning rate controls the change in the weights in each step of the iteration of training. In order to obtain optimum weights, researchers should avoid both too-small and too-large changes in the weights. We use a learning rate of 0.25, which provides good results in most practical cases (Rumelhart, *et al*. 1986). We can increase the speed of learning by filtering, based on the

past changes, the oscillations caused by the learning rate. The momentum factor parameter controls the effect of past changes, which should be a number close to 1. In this study, we use a momentum factor equal to 0.85.[11]

Time-Varying Vector Autoregressive Model. Modern macroeconomic applications increasingly involve the use of VARs with time-varying mean regression coefficients and covariance matrices, which implies a nonlinear VAR model. A time-varying parameter VAR with constant variance (Homoscedastic VAR) takes the following form:

$$Y_{t+1} = z_t \beta_t + \varepsilon_{t+1} \tag{12}$$

$$\beta_t = \beta_t + \eta_t \tag{13}$$

where z_t, $x_t^{'}$, Σ, and ε_t are defined as before in equation (1), β_t is an *(n x 1)* vector of *t=1,....T* parameters, and $\eta_t \sim N(0, Q)$ with Q as a *(n x n)* covariance matrix. The implied priors for β_1 to β_t are of the form $[\beta_t|\beta_{t-1}, Q \sim N(\beta_{t-1}, Q)]$. Since the covariance matrix Q is unknown, we postulate its prior as $Q^{-1} \sim Wishart(\xi, R^{-1})$. To avoid explosive behaviour (which might affect forecasting negatively) of the random walk assumption on the evolution of β_t, we must restrict its covariance Q. As such, we subjectively choose the hyper-parameters for the initial condition β_0 and the covariance matrix Q to form a tight prior. The hyper-parameters influence the performance of variable selection, which affects the mean and variance of the mean coefficients β. For the VAR case, when $\gamma_j = 0$ and β_j is restricted, we take a draw from each prior, implying that the prior variance V_0 cannot be very large (that is go to ∞) since it implies the

selection of no predictors. The hyper-parameter of the Bernoulli prior of γ_j also affects variable selection.

For time-varying parameters model with variable selection (TVP-VAR VS), a prior on the initial condition is of the form $\beta_0 \sim N(0, 4^2 V^{\min})$, with $[\gamma_j | \gamma_{\backslash -j} \sim \textit{Bernoulli}\ (1, 0.5)]$. The time-varying VAR without variable selection (TVP-VAR) uses a prior as in TVP-VAR VS with the restriction $\gamma_j = 1$ for all $j = 1,..., n$ imposed. The covariance Q of the time-varying coefficients in the TVP-VAR VS uses the prior $Q^{-1} \sim \textit{Wishart}(\xi, R)$, where $\xi = n+1$ and $R^{-1} = 0.0001(n+1)V^{\min}$ and $V^{\min}$ is the matrix defined in equation (9). The forecasts from the TVP-VAR models with and without variable selection use a run of 30,000 draws from the posterior, discarding the first 2,000 draws, with lags equal to 2.

<u>Nonparametric and Semi-Parametric Models.</u> We now consider nonparametric and semi-parametric regression approaches for forecasting detrended logged real GDP. We consider two competing multivariate models, and examine their forecasting abilities. These specifications are as follows:

Model 1: Nonparametric regression model (NP model)

$$y_t = f\left(y_{t-1}, y_{t-2}, r_{t-1}, r_{t-2}, \pi_{t-1}, \pi_{t-2}\right) + \varepsilon_{yt}; \tag{14}$$

$$r_t = f\left(y_{t-1}, y_{t-2}, r_{t-1}, r_{t-2}, \pi_{t-1}, \pi_{t-2}\right) + \varepsilon_{rt}; \text{ and} \tag{15}$$

$$\pi_t = f\left(y_{t-1}, y_{t-2}, r_{t-1}, r_{t-2}, \pi_{t-1}, \pi_{t-2}\right) + \varepsilon_{\pi t}. \tag{16}$$

Model 2: Semi-parametric regression model (SP model)

$$y_t = \alpha_{0y} + \alpha_{1y} y_{t-1} + \alpha_{2y} y_{t-2} + g(r_{t-1}, r_{t-2}, \pi_{t-1}, \pi_{t-2}) + \varepsilon_{yt}; \tag{17}$$

$$r_t = \alpha_{0r} + \alpha_{1r} r_{t-1} + \alpha_{2r} r_{t-2} + g(y_{t-1}, y_{t-2}, \pi_{t-1}, \pi_{t-2}) + \varepsilon_{rt}; \text{ and} \tag{18}$$

23

$$\pi_t = \alpha_{0\pi} + \alpha_{1\pi} \pi_{t-1} + \alpha_{2\pi} \pi_{t-2} + g(y_{t-1}, y_{t-2}, r_{t-1}, r_{t-2}) + \varepsilon_{\pi t}. \tag{19}$$

Here, $f(.)$ and $g(.)$ denote unknown functions that the data estimate. The ε_{ik}, $i=y$, r, π, are mean-zero errors with unchanged variance over the entire data set. The parameters α_{0i}, α_{1i}, and α_{2i}, $i=y$, r, π, are constants estimated from the data. Therefore, we can also describe the semi-parametric model as a partially linear nonparametric model.[12]

In the time-series context, nonparametric regressions can lead to issues with correlated errors (e.g., Opsomer, *et al.* 2001). For instance, the data-driven band-width selection techniques in the kernel smoothing methodology can break down in this context. In such cases, we could use a correlation-corrected method called CDPI to yield stable results. In our case, for Models 1 and 2, two lags guarantee the absence of autocorrelation. As a result, the responses in equations (13) to (15) and (16) to (18) exhibit uncorrelated errors. Also, stationarity checks ensure constant variances in each model. Finally, we compare such models based on their prediction errors or forecast performances.

We check the goodness of fit using Bootstrap testing and find *p*-values close to 1 for the models used. When estimating the unknown functions $f(\cdot)$ and $g(\cdot)$ in case of the nonparametric models, we use a local linear regression using AIC_c bandwidth selection criterion. In this case, we also examine all options for the choice of kernels and find that the Gaussian kernel of order 2 works the best yielding highest R-squared values and smallest MSE. We use the optimum bandwidth chosen by the software. In case of the semi-parametric modeling, we first compute data-driven bandwidths of the kernels to use in the $f(\cdot)$ and $g(\cdot)$ parts of the model, since bandwidth selection for lower levels of tolerance takes an extremely long time. We use a local-

---

[12] We use the *np* package in *R* to carry out the regressions outlined above.

linear, and not local-constant, regression type, as the local-linear type yields smaller R-squared values [13] Again, for the $f(\cdot)$ and $g(\cdot)$ parts of the model, we use Gaussian kernels of order 2, because they yield the highest R-squared values and the lowest MSE. We generate the forecasts from the NP and SP models using a recursive algorithm. That is, the forecast from origin n is generated for period $n+1$, and forecast values for period $n+1$ is inserted for unobserved values when forecasting for period $n+2$, and so forth.

Linear and Nonlinear Dynamic Stochastic General Equilibrium Models. Dynamic stochastic general equilibrium (DSGE) models not only carry out business cycle analysis, but also forecast macroeconomic variables. Since DSGE economies generally lack an analytical solution, economists work with numerical approximations to their theoretical models. Researchers linearize most models around a non-stochastic steady state to compute such approximations. This approach is appealing from an econometric perspective, since it allows the use of Kalman filtering techniques to build the likelihood function implied by the approximate model, and construct out-of-sample forecasts. Linearization, however, may prove problematic, especially when nonlinearities are important or when significant shocks move the economy far from the steady state – as in during the recent downturn. Moreover, recent work by Fernandez-Villaverde and Rubio-Ramirez (2005) and Fernandez-Villaverde et al. (2006) points out that estimating DSGE models using linearized solutions will generally lead to biased parameter estimates. Hence, Fernandez-Villaverde *et al*. (2006) suggest moving to second-order approximations when

---

[13] The decision to use the local linear regression method instead of the kernel smoother methods adopted by Arora *et al.* (2011) in forecasting US GNP based on nonparametric method, emanates from the fact that the former does not suffer from the problem of biased boundary points.

taking DSGE models to the data. This approach, however, comes with computational costs, since it requires the use of Monte Carlo methods for constructing the likelihood function.

We use the framework developed by Pichler (2008), essentially a relatively small new-Keynesian monetary economy featuring monopolistic competition, capital accumulation, and price and capital adjustment costs characterizing the rigidities in the economy. The economy includes a representative household, a representative finished-goods-producing firm, a continuum of intermediate-goods-producing firms, and a central bank following the basic Taylor rule, which smoothes the interest rate over time and reacts to deviations of output and inflation from their target values, in setting the interest rate. Households face adjustment costs to convert final goods to productive capital, while Rotemberg (1982)-type price adjustment costs confront intermediate goods producers. In this model, technology shocks, preference shocks, and monetary shocks drive macroeconomic fluctuations. A system of 13 non-linear equations characterize the model's symmetric equilibrium, which, in turn, can be summarized as follows:

$$E_t R(f_{t+1}, f_t, s_{t+1}, s_t, \varepsilon_{t+1}; \Theta) = 0, \tag{20}$$

where R is a nonlinear operator, the vectors s_t and f_t include the model's state and control variables, respectively, ε_{t+1} summarizes the exogenous disturbances, and Θ collects the structural parameters.

The model's solution, in turn, is given by decision rules $\Psi(s_t, \varepsilon_{t+1})$ and $\Phi(s_t)$ that satisfy the following condition:

$$E_t R(\Phi(\Psi(s_t, \varepsilon_{t+1})), \Phi(s_t), \Psi(s_t, \varepsilon_{t+1}), s_t, \varepsilon_{t+1}; \Theta) = 0 \tag{21}$$

26

for all s_t, ε_{t+1}, and t. Collecting state and decision variables in a vector $X_t = (s_t', f_t')'$, we can write this solution compactly as $X_{t+1} = H(X_t, \varepsilon_{t+1}; \Theta)$, where H is a nonlinear function depending on Φ and Ψ, respectively. For the economy under consideration, we cannot compute the functions Φ and Ψ (and, therefore, H) analytically, but must approximate by numerical methods.

Assume that we observe N_Y macroeconomic time series of length T, collected in $y^T = \{y_t\}_{t=1}^T$. Together, with the data, our DSGE model forms the following nonlinear state-space system:

$$X_{t+1} = H(X_t, \varepsilon_{t+1}; \Theta), \text{ and} \tag{22}$$

$$y_t = GX_t + v_t. \tag{23}$$

The vector v_t denotes a normally distributed and uncorrelated measurement errors [i.e., $v_t \sim N(0, \Sigma_v)$ with a diagonal Σ_v]. We assume that the model's variables relate to the observable data series in a linear way, as represented by the matrix G.

From the state-space system in equations (22) and (23), we denote the likelihood of the sample data y^T conditional on our model H with parameters $\Omega = \{\Theta, \Sigma_v\}$ as $L(y^T / \hat{H}, \Omega)$. Unfortunately, since we cannot compute the model solution H analytically, we cannot evaluate $L(y^T / \hat{H}, \Omega)$ exactly. Filtering techniques, however, allow us to approximate the likelihood based on approximate decision rules $\hat{H}$. In this paper, we compute the likelihood based on linear and nonlinear approximations to the model's decision rules, denoted by $\bar{H}$ and $\tilde{H}$. As in Pichler (2008), we derive these approximations using first and second-order

perturbation methods as described by Klein (2000) and Schmitt-Grohe and Uribe (2004), respectively.

Based on a first-order approximation of the model, we can construct the likelihood function in a straightforward way. Since $\bar{H}$ is linear and the innovations ε_{t+1} are Gaussian, the densities in $L(y^T \, / \, \hat{H}, \Omega)$ are normal, such that we can use the Kalman filter to analytically build function $L(y^T \, / \, \bar{H}, \Omega)$. For estimating the model based on its quadratic approximation $\tilde{H}$, however, the Kaman filter is no longer available. In particular, since $\tilde{H}$ is nonlinear, we cannot evaluate the densities in closed form. To estimate parameters based on the quadratic model, we, thus, resort to Monte Carlo methods, precisely the basic particle filter suggested by Fernandez-Villaverde and Rubio-Ramirez (2007).

Having constructed the likelihood function from either the linearized or the quadratic model, we can compute maximum likelihood estimates numerically as follows: $\bar{\Omega}_{ML} = argmax_{\Omega} L(y^T \, / \, \bar{H}, \Omega)$ and $\tilde{\Omega}_{ML} = argmax_{\Omega} L(y^T \, / \, \tilde{H}, \Omega)$, respectively. Deriving these estimates, however, proves a non-trivial exercise, as several complications arise:

First, the likelihood function is almost flat with respect to some parameters. Hence, we calibrate these parameters rather than estimate them via maximum likelihood. Specifically, we set the elasticity of output with respect to capital at 0.36, the depreciation rate at 0.025, and the capital adjustment costs parameter at 10. For similar reasons, we calibrate the mark-up parameter at 6. Furthermore, we choose the parameters corresponding to leisure and real money balances in the utility function (equal to 2.4) such that the household spends 30 percent of its time working in the steady state, and to match the steady state ratio between real balances and quarterly output, respectively. Finally, as in Pichler (2008), we calibrate the measurement error variances to 10

percent of the variance of the respective data series. The DSGE model possesses a parameter space consisting of 19 parameters. Thus, using 2 lags, based on the BIC, in the different other competing models keeps the size of the parameter space amongst all the models comparable. Second, the likelihood function features many local maxima and minima. Finally, when we use the particle filter to construct the likelihood based on the nonlinear solution, the resulting likelihood function is not continuous with respect to parameter vector Ω. To address the latter two problems, we employ a simulated annealing approach instead of gradient-based methods for maximizing the likelihood function (Pichler, 2008).

3.   Data and Results:

Our US data set includes quarterly time series on output, the inflation rate, and the nominal interest rate. The series come from the ALFRED (ArchivaL Federal Reserve Economic Data) database maintained by the Federal Reserve Bank of St. Louis. Output corresponds to seasonally adjusted at an annual rate quarterly gross domestic product (GDP) in billions of chained 2005 dollars (specifically, the GDPC1 series at ALFRED), whereby we remove a linear trend from the (logged) GDP series. We calculate the inflation rate as the relative change in the GDP deflator (GDPDEF in ALFRED) with a base year of 2005 and the nominal interest rates as the averaged weekly 3-Month Treasury bill rates (WTB3MS in ALFRED). We select the time period 1979Q3 through 2011Q2. This gives a total sample of 132 observations on each series with the first 80 (1979:Q3 through 1999:Q2) used for in-sample analysis, the next 30 (1999:Q3 through 2006:Q4) used for the out-of-sample forecasting, and the remaining 22 (2007:Q1 through 2012:Q2) used for predicting the real logged GDP *ex ante*. Barring the DSGE models, we took first differences of the variables to ensure stationarity, since we could not reject the null hypothesis of a unit-root

in all variables, using standard unit-root tests.[14] After generating the forecasts of the first-differences, we recover the forecasts of the levels by using the actual data of the previous periods.[15] Thus, we estimate the models in first-differences with one lag, since the BIC suggests 2 lags for a VAR estimated in levels.[16]

Out-of-Sample Forecasting:1999:Q3-2006:Q4

This subsection discusses the findings from the forecasting performance of the 12 best models in each category of specifications -- the VAR, BVAR1, BVAR2, ANN, VSTAR, MS-VAR, TVP-VAR1, TVP-VAR2, NP, SP, DSGE-Linear, and DSGE-Nonlinear models -- compared to the RW benchmark model. We first examine the raw root mean squared errors (RMSEs) of each specification (i.e., the RW model and the 12 best models in each category). Then we consider the RMSEs of the best 12 models to the RMSE of the RW model. Finally, we report whether the ratios of the 12 RMSEs to the RMSE of the RW model differ significantly from the ratios for the other specifications.

---

[14] These results are available upon request from the authors.

[15] Allowing one lag in the first-differenced atheoretical models implies that our in-sample estimation starts from 1980:Q1 – the same starting point recently used by Bekiros and Paccagnini (2012) when estimating DSGE and TVP-VAR models. Ireland (2004) indicates that significant changes occurred in US monetary and fiscal policy in 1980 and, thus, they constitute a major breakpoint. Further, 1980:Q1 roughly coincides with the end of the Volcker stabilization and disinflation era. Hence, the in-sample exhibits a more stable monetary and financial structure and a lower volatility of the macroeconomic variables. Also, Justiniano and Primiceri (2008) point out that researchers find structural breaks in mean and volatility by comparing the pre- and post-80 periods, while one cannot reject the null hypothesis of parameter stability in the post-80 period. Furthermore, Benati and Surico (2008) claim that if the US economy experienced an indeterminate equilibrium before 1980, then by estimating models before and after the 80s will mix two different regimes, thus obtaining biased estimates of the structural parameters. Finally, Herbst and Schorfheide (forthcoming) argue that as strong empirical evidence exists that monetary policy as well as the volatility of macroeconomic shocks changed in the early 1980s. Given that, the information set in estimating the models should provide relevant information to the exercise of contemporary policy making, thus a sample after the 1980s ensures better forecasting performance.

[16] The Trace and Maximum-eigenvalue tests of cointegration, which are available upon request from the authors, do not detect any evidence of cointegration amongst the three key variables. Hence, no explicit need exists to model the error-correction term, and we simply use the VAR versions of the different linear and nonlinear models.

When forecasting into the future, we anticipate, other things constant, that the RMSE will increase with the number of periods into the future that the model forecasts. The RW model follows this expectation. Examining the raw RSMEs across the 12 other best models, we see that the specifications divide into three categories, First, seven specifications -- the best ANN, VSTAR, MS-VAR, NP, SP, DSGE-Linear, and DSGE-Nonlinear models -- all exhibit monotonically rising RMSEs with the forecast horizon. Second, three specifications -- the best BVAR2, TVP-VAR1, and TVP-VAR2 models -- exhibit a falling RMSE across the forecast horizon, except for the horizon at seven quarters where the RMSE rises. Finally, the best VAR and BVAR1 models exhibit rising RMSEs over the initial horizons and then falls for horizons 6 and 7, and 5 to 8, respectively.

Table 1 reports the relative RMSEs of the 12 best specifications to the RMSE of the RW model across the eight forecasting horizons. Once again, the models divide into three categories of outcomes. Five specifications -- the best BVAR2, TVP-VAR1, TVP-VAR2, DSGE-Linear, and DSGE-Nonlinear models -- each exhibit monotonically decreasing RMSEs across the forecast horizons. The latter two models experience declining relative RMSEs compared to the RMSE of the RW model, even though they experience rising RMSEs over the forecast horizons, because their rising RMSEs do so more slowly than that of the RW model. Four specifications -- the best ANN, MS-VAR, NP, and SP models -- experience rising relative RMSEs over the initial four forecast horizons, experiencing some decreases in the relative RMSEs over the last four forecast horizons. Finally, the remaining three specifications -- the best VAR, BVAR1, and VSTAR models -- generally experience a fall in their relative RMSEs, except for some increases at the second and third forecast horizons.

When we average the relative RMSEs across the eight forecast horizons, the DSGE-Nonlinear model performs the best followed closely by the DSGE-Linear model.[17] By the end of the forecast horizon, the best BVAR2, TVP-VAR1, and TVP-VAR2 models enjoy the lowest relative RMSEs. They do not outperform the DSGE models averaged over all forecasting horizons because of their poor forecasting performance at the initial horizons. In sum, the two DSGE models perform well on the average across all forecast horizons because they do a reasonably good job of forecasting relative to the RW model at every horizon. On the other hand, the BVAR2, TVP-VAR1, and TCP-VAR2 models perform the worst at horizon one and then improve their forecasting performance across all eight horizons.

Table 2 reports the significance of differences between the RMSEs of various specifications using the Giacomini and White (2006) statistic for equal forecast accuracy. The findings confirm our observations from Table 1. First, the two DSGE models significantly outperform the BVAR2, TVP-VR1, and TVP-VAR2 models at forecast horizons one, two, and three, but the reverse occurs at horizons six, seven, and eight. In fact, the RW model significantly outperforms the BVAR2, TVP-VR1, and TVP-VAR2 models at forecast horizons one and two. The reverse occurs beginning at horizon four.

Predicting the Real Logged GDP Ex Ante: 2007:Q1-2012:Q2

Our final test examines the ability of the various models to predict the Great Recession *ex ante*. Figure 1 plots the forecast and actual values for real GDP. We include, in addition to the actual observations, the *ex ante* forecasts for the RW, VAR, TVP-VAR2, DSGE-Linear, and DSGE-

---

[17] For the individual horizons, the MS-VAR model performs the best at the 1st- and 2nd-quarter horizons, the DSGE-Nonlinear model, at the 3rd- and 4th-quarter horizons, and the BVAR2 model, at the 5th-, 6th-, 7th-, and 8th-quarter horizons.

Nonlinear models, since these specifications capture the scope of the forecasting performance across all specifications.

The actual series shows a peak in 2007Q4 followed by a trough in 2009:Q2, the time identified by the Business Cycle Dating Committee of the NBER. Only the two DSGE models capture a recession. The DSGE-Linear model forecasts two sharp drops in real GDP for one quarter in each case. The first drop exhibits a peak in 2008:Q4 and a trough in 2009:Q1 while the second, a peak in 2010:Q4 and a trough in 2011:Q1. The DSGE-Nonlinear model forecasts a recession with a peak in 2009:Q1 and a trough in 2009:Q3. The DSGE-Nonlinear forecast proves closer to the actual observations before the two steep drops in the forecast of the DSGE-Linear model. After the second drop, the two DSGE forecasts generally follow each other, but at a level closer to the actual observations than forecasted by the other models, either potted in Figure 1 or not. Each of these other series do not forecast a recession or steep drop, but rather continue to forecast rising real GDP. Nearly all these other series follow a path not too different from the forecast of the RW model.

4.   Conclusion

This paper uses small set of variables and a rich set of models to consider whether we could have predicted the recent downturn of the US real Gross Domestic Product (GDP). The rich set of models includes athoeretical and theoretical, linear and nonlinear, as well as classical and Bayesian models. Our restricted data set includes real GDP, the rate of inflation of the GDP implicit deflator, and the three-month Treasury-bill rate.

Our analysis considers the most recent financial crisis and recession, as financial-crisis-induced recessions exhibit more depth and length than typical recessions (Reinhart and Rogoff, 2009). First, we estimate the wide range of econometric models noted previously over the in-

sample period from 1979:Q3 to 1999:Q2. Second, we forecast one- to eight quarters-ahead of detrended logged real GDP over an out-of-sample forecast horizon from 1999:Q3 to 2006:Q4. Third, we choose the model within each category that produces the minimum average root mean square errors (RMSEs) relative to the benchmark random walk model as the "best" model for a specific category. Fourth, we use the best model within each category to forecast the level of real logged GDP *ex ante* (without updating the parameter estimates of the optimal models) over 2007:Q1 to 2011:Q2.

Comparing the performance by RMSEs of the models to the benchmark RW model, the two DSGE models perform well on the average across all forecast horizons because they do a reasonably good job of forecasting relative to the RW model at every horizon. On the other hand, the BVAR2, TVP-VAR1, and TVP-VAR2 models perform the worst at horizon one and then improve their forecasting performance across all eight horizons. The DSGE-Nonlinear model does by far the best job of forecasting the Great Recession in the *ex ante* forecasts. In sum, some atheoretical and theoretical models perform the best in the out-of-sample forecast exercise. The theoretical models, however, dominate in our *ex ante* forecast comparison, when trying to forecast the Great Recession. These findings support those in Gupta *et al*. (2011). They too find the superiority (relative to small and large-scale atheoretical linear models) of a DSGE model that explicitly incorporates the housing sector in predicting *ex ante* the downturn in real US house prices. They suggest, and we corroborate, that to forecast the downturn in a specific variable may require forward-looking, microfounded, DSGE models in the fundamental variables. Further, the fact that the DSGE-Nonlinear ex ante forecasts prove closer to the actual observations, also highlight the importance of second-order approximation of the model

economy around the steady-state to account for nonlinearities when significant shocks move the economy far from the steady state, as occurred during the Great Recession.

In sum, the nonlinear DSGE model performs the best overall in the out-of-sample RMSE averaged across all horizons as well as in tracking the turning point in the Great Recession. This occurs despite the limited economic structure consider in Pichler's (2008) model, essentially a relatively small new-Keynesian monetary economy featuring monopolistic competition, capital accumulation, and price and capital adjustment costs characterizing the rigidities in the economy. That is, the model includes a system of 13 non-linear equations characterize the model's symmetric equilibrium. A slightly more comprehensive model may provide even better out-of-sample forecast accuracy and the tracking of turning points.

References:

Altavilla, C., and Ciccarelli, M., 2007. Inflation forecasts, monetary policy and unemployment dynamics evidence from the US and the euro area. *Working Paper* No. 125, European Central Bank.

An, S., and Schorfheide, F., 2007. Bayesian analysis of DSGE models. *Econometric Reviews* 26, 113-172.

Anderson, M. K., and Karlsson, S., 2008. Bayesian forecast combination for VAR models. In: Chib, S., Grifitths, W., Koop, G., and Terrell, D., (eds.). *Bayesian Econometrics*, 501-524. *Advances in Econometrics* 23. Oxford: Elsevier.

Ang, A., and Bekaert, G., 2002. Regime Switches in Interest Rates. *Journal of Business and Economic Statistics* 20, 163-82.

Arora, S., Little, M. A., and McSharry, P. E., 2011. Nonlinear and nonparametric modeling approaches for probabilistic forecasting of the US Gross National Product. Available at: http://www.maxlittle.net/publications/SNDE_GNP.pdf

Balcilar, M., Gupta, R., Majumdar, A., and Miller, S. M., forthcoming. Forecasting Nevada gross gaming revenue and taxable sales using coincident and leading employment indexes. *Empirical Economics*.

Banbura, M., Giannone, D., and Reichlin, L., 2010. Large Bayesian VARs. *Journal of Applied Econometrics* 25, 71-92.

Barnett, A., Mumtaz, H., and Theodoridis, K., 2012. Forecasting UK GDP growth, inflation and interest rates under structural change: a comparison of models with time-varying parameters. *Working Paper* No. 450, Bank of England.

Bekiros, S., and Paccagnini, A., 2012. On the predictability of time-varying VAR and DSCE models. *Empirical Econom*ics, DOI 10.1007/s00181-012-0623-z.

Benati, L., and Surico, P., 2008. Evolving US monetary policy and the decline of inflation predictability. *Journal of European Economic Association* 6, 634–646

Bishop, C., M., 1995. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Camacho, M., 2004. Vector smooth transition regression models for US GDP and the composite index for leading indicators. *Journal of Forecasting* 23, 173-196.

Clements, M. P., and Hendry, D. F., 1999. *Forecasting Non-stationary Economic Time Series: The Zeuthen Lectures on Economic Forecasting*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Krolzig, H. M., 1998. A comparison of the forecast performance of Markov switching and threshold autoregressive models of US GNP. *Econometrics Journal* 1, C47–C75.

Clements, M. P., and Smith, J., 1996. The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting* 13, 463-75.

Clements, M. P., and Smith, J., 1999. A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics* 14, 124–141.

Clements, B., Bhattacharya, R., and Nguyen, T. Q., 2003. External debt, public investment, and growth in low-income countries. IMF *Working Paper* 03/249.

Cogley, T., and Sargent, T. J., 2001. Evolving post World War II US inflation dynamics. *NBER Macroeconomics Annual* 16, 331-373.

Consolo, A., Favero, C. A., and Paccagnini, A., 2009. On the statistical identification of DSGE models. *Journal of Econometrics* 150, 99-115.

Cybenko, G., 1989. Approximation by superposition of sigmoidal functions. *Mathematics of Control, Signals and Systems* 2, 303-314.

D'Agostino, A., Gambetti, L., and Giannone, D., 2011. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, DOI: 10.1002/jae.1257.

Dacco, R., and Satchell, S., 1999. Why do regime-switching models forecast so badly? *Journal of Forecasting* 18, 1–16.

Davies, R. B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33-43.

De Gooije, G., and Vidiella-i-Anguera, A., 2004. Forecasting threshold cointegrated systems. *International Journal of Forecasting* 20, 237-253.

Del Negro, M., Schorfheide, F., Smets, F., and Wouters, F., 2007. On the fit of new Keynesian models. *Journal of Business and Economic Statistics* 25, 123-162.

Del Negro, M., and Schorfheide, F., 2012. DSGE model-based forecasting, Staff Reports 554, Federal Reserve Bank of New York.

Deschamps, P. J., 2008. Comparing smooth transition and Markov switching autoregressive models of US unemployment. *Journal of Applied Econometrics* 23, 435-462

Diebold, F. X., and Rudebusch, G. D., 1996. Measuring business cycles: A modern perspective. *Review of Economics and Statistics* 78, 67–77.

Diebold, F. X., Lee, J.-H., and Weinbach, G. C., 1994. Regime switching with time-varying transition probabilities. In C. Hargreaves (ed.) *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302, Oxford: Oxford University Press.

Doan, T., Litterman, R., and Sims, A. C., 1984. Forecasting and conditional projection using realistic prior distributions. *Economic Reviews* 3, 1-100.

Durland, J. M., and McCurdy, T. H., 1994. Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business and Economic Statistics* 12, 279–288.

Fahlman, S. E., 1988. Faster-learning variations on back-propagation: An empirical study. In: Touretsky, D, Hinton, G. E., and Sejnowski, T. J., (eds.). Proceedings of the 1988 Connectionist Models Summer School. San Mateo: Morgan Kaufmann, 38–51.

Fausett, L., 1994. *Fundamentals of Neural Networks*. NJ: Prentice-Hall.

Filardo, A. J., 1994. Business-cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12, 299–308.

Fernandez-Villaverde, J., and Rubio-Ramirez, J. F., 2005. Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics* 20, 891-910.

Fernandez-Villaverde, J., and Rubio-Ramirez, J. F., 2007. Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies* 74, 1059-1087.

Filardo, A. J., and Gordon, S. F., 1998. Business cycle durations. *Journal of Econometrics* 85, 99–123.

Fine, T. L., 1999. *Feedforward Neural Network Methodology*. Berlin: Springer-Verlag.

Foresee F. D., and Hagan M. T., 1997. Gauss–Newton approximation to Bayesian regularization. *IEEE International Conference on Neural Networks* 3, New York: IEEE, 1930-1935.

Funahashi, K., 1989. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2, 183–192.

George, E. I., Sun, D., and Ni, S., 2008. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* 142, 553-580.

Ghysels, E,. 1994. On the periodic structure of the business cycle. *Journal of Business and Economic Statistics* 12, 289–298.

Giacomini, R., and White, H., 2006. Tests of conditional predictive ability. *Econometrica* 74, 1545-1578.

Goodwin, T. H., 1993. Business-cycle analysis with a Markov switching model. *Journal of Business and Economic Statistics* 11, 331–339.

Granger, C. W. J., and Terasvirta, T., (1993). *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Gupta, R., Kabundi, A., and Miller, S. M., (2011). Forecasting the US real house price index: Structural and non-structural models with and without fundamentals. *Economic Modelling* 28, 2013-2021.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

Hamilton, J. D., 1990. Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39-70.

Hamilton, J. D., 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hansen, J. V., McDonald J. B., and Nelson, R. D., 1999. Time series prediction with genetic-algorithm designed neural networks: An empirical comparison with modern statistical models. *Computational Intelligence* 15, 171-184.

Herbst, E., and Schorfheide, F., forthcoming. Evaluating DSGE model forecasts of comovements. *Journal of Econometrics*.

Hornik, K., Stinchombe, M., and White, H., 1989. Multilayer feed-forward networks are universal approximators. *Neural Networks* 2, 359–366.

Ireland, P., 2004. A method for taking models to the data. *Journal of Economic Dynamics and Control* 28, 1205–1226.

Justiniano, A., and Primiceri, G. E., 2008. The time-varying volatility of macro-economic fluctuations. *American Economic Review* 98, 604–641.

Kaastra, I., and Boyd, M., 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10, 215–236.

Kim, C. J., and Nelson, C. R., 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics* 80, 188–201.

Kim, M.-J., and Yoo, J.-S., 1995. New index of coincident indicators: A multivariate Markov switching factor model approach. *Journal of Monetary Economics* 36, 607– 630.

Klein, P., 2000. Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control* 24, 1405-1423.

Koop, G., and Korobilis, D., 2010. Large time-varying parameter VARs. *Working Papers 2012_04, Business School - Economics, University of Glasgow*.

Korobilis, D., 2006. Forecast comparison of nonlinear time series models of US GDP: A Bayesian approach, Unpublished Masters Dissertation, Erasmus University.

Korobilis, D., 2008. Forecasting in vector autoregressions with many predictors. *Advances in Econometrics* 23: Bayesian Macroeconometrics, 403-431.

Korobilis, D., 2011. VAR forecasting using Bayesian variable election. *Journal of Applied Econometrics,* DOI: 10.1002/jae.1271.

Krolzig, H. M., 1997. *Markov Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Berlin: Springer.

Krolzig, H. M., 2000. Predicting Markov-switching vector autoregressive processes. *Working Paper,* No. 38, Department of Economics, University of Oxford.

Lachtermacher, G., and Fuller, J. D., 1995. Backpropagation in time-series forecasting. *Journal of Forecasting* 14, 381–393.

Lam, P. S., 1990. The Hamilton model with a general autoregressive component. *Journal of Monetary Economics* 26, 409–432.

Lindgren, G., (1978). Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics* 5, 81-91.

Lisi, F., and Schiavo, R. A. 1999. A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics and Data Analysis* 30, 87–102.

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

Litterman, R. B., 1986. Forecasting with Bayesian vector autoregression: Five years of experience. *Journal of Business and Economic Statistics* 4, 25-38.

Lucas, R. E., 1977. Understanding business cycles, In: Brunner, K., and Metzler, A., (eds.). *Stabilization of the Domestic and International Economy*, 7–29, *Carnegie–Rochester Series on Public Policy* 5.

Lundbergh, S., and Terasvirta, T., 2002. Forecasting with smooth transition autoregressive models. In: Clements, M. P., and Hendry, D. F., (eds.). *A Companion to Economic Forecasting.* Oxford: Blackwell, 485–509.

Luukkonen, R., Saikkonen, P., and Terasvirta, T., 1988. Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.

MacKay, D. J. C., 1992. A practical Bayesian framework for backpropagation networks. *Neural Computation* 4, 448–472.

Marcellino, M., 2002. Instability and non-linearity in the EMU. *Discussion Paper* No. 3312, Centre for Economic Policy Research.

Marcellino, M., 2007. A comparison of time series models for forecasting GDP growth and inflation. Available at: http://www.eui.eu/Personal/Marcellino/1.pdf.

Marcellino, M., 2004. Forecasting EMU macroeconomic variables. *International Journal of Forecasting* 20, 359-372.

Masters, T., 1993. *Practical neural networks recipes in C++*. London: Academic Press.

McCulloch, W. S., and Pitts, W. H., 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115-133.

Milas, C., and Naraidoo, R., forthcoming. Financial market conditions, real time, nonlinearity and European Central Bank monetary policy: In-sample and out-of-sample assessment. *Computational Statistics and Data Analysis*.

Minsky, M., and Papert, S., 1969. *Perceptrons*. MIT Press, Cambridge, MA.

Montgomery, A., Zarnowitz, V., Tsay, R., and Tiao, G., 1998. Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association* 93, 478–493.

Neftci, S. N., 1984. Are economic time series asymmetric over the business cycle? *The Journal of Political Economy* 92, 307-328.

Newey, W. K., and West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Opsomer, J. D., Moisen, G. G., and Kim, J. Y., 2001. Model-assisted estimation of forest resources with generalized additive models. In: *Proceedings of the Section on Survey Research Methods*, American Statistical Association. Alexandria, VA.

Paccagnini, A., 2011. DGSE model evaluation and hybrid models: A comparison. *Working Paper* MWP 2011/11, European University Institute.

Pesaran, M. H., and Potter, S. M., 1997. A floor and ceiling model of US Output. *Journal of Economic Dynamics and Control* 21, 661–695.

Pichler, P., 2008. Forecasting with DSGE models: The role of nonlinearities. *The B.E. Journal of Macroeconomics*, 8(1), (Topics), Art. 20.

Potter, S., 1999. Nonlinear time series modelling: An introduction. *Journal of Economic Surveys* 13, 505-528.

Psaradakis, Z., Ravn M. O., and Sola, M., 2005. Markov switching causality and the money—output relationship. *Journal of Applied Econometrics* 20, 665-683.

Psaradakis, Z., and Spagnolo, N., 2003. On the determination of number of regimes in Markov-switching autoregressive models. *Journal of Time Series Analysis* 24, 237- 252.

Rzempoluck, E. J., 1998. *Neural Network Data Analysis using Simulnet*. New York: Springer.

Ripley, B. D., 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Rotemberg, J. J., 1982. Sticky prices in the United States. *Journal of Political Economy* 60, 1187-1211.

Rothman, P., 1998. Forecasting asymmetric unemployment rates. *Review of Economics and Statistics* 80, 164-168.

Rothman, P., van Dijk, D., and Hans, P., 2000. Multivariate STAR analysis of money-output relationship. *Macroeconomic Dynamics* 5, 506-532.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986. Learning internal representations by error propagation. In: Rumelhart, D. E., and McClelland, J. L., (eds.). *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 318–362.

Sarle, W. S., 2002. Neural network FAQ. Retrieved 28 May 2010, from ftp://ftp.sas.com/pub/neural/FAQ.html.

Schmitt-Grohé, S., and Uribe, M., 2004. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control* 28, 755–775.

Sims, C. A., 2001. Comment on Sargent and Cogley's' Evolving post World War II US inflation dynamics. *NBER Macroeconomics Annual* 16, 373-379.

Smets, F., and Wouters, R., 2003. An estimated stochastic dynamic general equilibrium model of the Euro area. *Journal of the European Economic Association* 1, 1123-1175.

Smith, M., 1993. *Neural Networks for Statistical Modeling*. New York: Van Nostrand Reinhold.

Spencer, D. E., 1993. Developing a Bayesian vector autoregression model. *International Journal of Forecasting* 9, 407-421.

Stern, H. S., 1996. Neural networks in applied statistics. *Technometrics* 38, 205–220.

Stock, J. H., and Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14, 11-30.

Stock, J. H., and Watson, M. W., 1999. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R., and White, H., (eds.). *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W. J. Granger*. Oxford: Oxford University Press, 1-44.

Stock, J. H., and Watson, M. W., 2005. Implications of factor models for VAR analysis. *Mimeo*, Princeton University.

Tsay, R. S., 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188-1202.

Terasvirta, T., 1998. Modelling economic relationships with smooth transition regressions. In: Ullah, A., and Giles, D. E. A., (eds.). *Handbook of Applied Economic Statistics*. New York: Marcel Dekker, 507-552.

Terasvirta, T., 2006. Forecasting economic variables with nonlinear models. In: Elliott, G., Granger, C., and Timmermann, A., (ed.). *Handbook of Economic Forecasting,* 1. Amsterdam: Elsevier 413-457.

Terasvirta, T., van Dijk, D., and Medeiros, M. C., 2005. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting* 21, 755-774.

Todd, R. M., 1984. Improving economic forecasting with Bayesian vector autoregression. *Quarterly Review*, Federal Reserve Bank of Minneapolis, Fall, 18-29.

van Dijk, D., Terasvirta, T., and Franses, P. H., 2002. Smooth transition autoregressive models - A survey of recent developments. *Econometric Reviews* 21, 1-47.

van Dijk, D., and Franses, P. H., 2003. Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. *Oxford Bulletin of Economics and Statistics* 65, supplement 724-744.

Villani, M., 2009. Steady state priors for vector autoregressions. *Journal of Applied Econometrics* 24, 630-650.

Wasserman, P. D., 1989. *Neural Computing: Theory and Practice*. New York: Van Nostrand Reinhold.

White, H., 1988. Economic prediction using neural networks: The case of IBM stock prices. Proceedings of the Second Annual IEEE Conference on Neural Networks, New York: IEEE Press. 17(II):451–458.

White, H., 1989. Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association,* 84:1003–1013.

Table 1: **Relative RMSEs: Out-of-Sample Forecasting Performance to Random Walk Model, 1999:Q3-2006:Q4**

Model	Forecast Horizon								Average
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Average**
RW	0.5314	0.7572	0.9804	1.1422	1.2725	1.4063	1.5474	1.692	1.1662
VAR	1.0412	0.9896	1.0165	0.9948	0.9225	0.8159	0.7187	0.6746	0.8967
BVAR1	1.1026	1.1325	1.0731	1.0311	0.9032	0.7684	0.6191	0.5508	0.8976
BVAR2	2.042	1.3828	1.0446	0.89	**0.7316**	**0.647**	**0.6001**	**0.4924**	0.9788
ANN	1.242	1.3998	1.4484	1.4971	1.556	1.5715	1.5639	1.5721	1.4814
VSTAR	1.2928	1.3166	1.3739	1.3614	1.299	1.2644	1.2065	1.1708	1.2857
MS-VAR	**0.8237**	**0.9623**	1.0057	1.0232	0.9893	0.9158	0.8384	0.7686	0.9159
TVP-VAR1	2.0445	1.3845	1.0479	0.8906	0.7341	0.649	0.6016	0.4947	0.9809
TVP-VAR2	2.0428	1.397	1.044	0.8933	0.7344	0.649	0.6011	0.4931	0.9818
NP	0.9248	1.1922	1.2297	1.3803	1.3304	1.2972	1.2319	1.2289	1.2269
SP	1.0424	1.4793	1.558	1.7361	1.8221	1.9035	1.8713	1.8542	1.6584
DSGE-Linear	1.0139	1.0071	0.926	0.8945	0.8439	0.8141	0.782	0.7672	0.8811
DSGE-Nonlinear	1.0197	0.9827	**0.8763**	**0.8395**	0.7738	0.7447	0.7084	0.6923	**0.8297**

Notes: RW: Random-Walk Model; VAR: Classical Vector Autoregressive Model; BVAR1 (BVAR2): Bayesian Vector Autoregressive Model based on the Minnesota-prior (Bayesian Vector Autoregressive Model based on Variable-Selection); ANN: Artificial Neural Network Model; VSTAR: Vector Smooth Transition Autoregressive Model; MS-VAR: Markov-Switching Vector Autoregressive Model; TVP-VAR1 (TVP-VAR2): Time-Varying Vector Autoregressive Model without Variable Selection (Time-Varying Vector Autoregressive Model with Variable Selection); NP: Nonparametric Regression; SP: Semi-Parametric Regression; DSGE-Linear (DSGE-Nonlinear): Dynamic Stochastic General Equilibrium Model Estimated with Kalman Filter (Dynamic Stochastic General Equilibrium Model Estimated with Particle Filter). The entries for the RW model correspond to the absolute RMSE from the model in percentages. The entries for the other models correspond the RMSE of the particular model relative to the RMSE of the RW model for a specific forecast horizon. The bolded entries identify the minimum relative RMSE for each horizon and the average across all horizons across the 12 models, excluding the RW model.

Table 2:   **Significance of Differences in Out-of-Sample Forecasting Performance, 1999:Q3-2006:Q4**

				Forecast Horizons				
Models	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
VAR versus RW	4.1225	-1.0376	1.6476	-0.5219	-7.7467	-18.4079**	-28.1291**	-32.5378*
BVAR1 versus RW	10.2631†	13.2470†	7.3086	3.1135	-9.6806	-23.1632**	-38.0883*	-44.9183*
BVAR2 versus RW	104.1983*	38.2849*	4.4562	-11.0020†	-26.8373**	-35.3050*	-39.9940*	-50.7551*
MS-VAR versus RW	-17.6329**	-3.7704	0.5725	2.3192	-1.0731	-8.4249	-16.1570**	-23.1354**
TVP-VAR1 versus RW	104.4525*	38.4482*	4.7936	-10.9358†	-26.5883**	-35.0967*	-39.8374*	-50.5303*
TVP-VAR2 versus RW	104.2771*	39.7008*	4.4049	-10.6658†	-26.5643**	-35.0988*	-39.8902*	-50.6880*
DSGE-Linear versus RW	1.3924	0.7138	-7.4031	-10.5477†	-15.6098†	-18.5853**	-21.7975**	-23.2838**
DSGE-Nonlinear versus RW	1.9674	-1.7296	-12.3668†	-16.0474**	-22.6222**	-25.5321**	-29.1604**	-30.7707*
DSGE-Nonlinear versus VAR	-2.0698	-0.6992	-13.7873†	-15.6070†	-16.1246**	-8.7315	-1.4349	2.6193
DSGE-Nonlinear versus BVAR1	-7.5235	-13.2247†	-18.3354**	-18.5823**	-14.3287†	-3.0830	14.4204†	25.6847**
DSGE-Nonlinear versus BVAR2	-50.0645*	-28.9363**	-16.1053**	-5.6691	5.7612	15.1061†	18.0543**	40.5815*
DSGE-Nonlinear versus TVP-VAR1	-50.1266*	-29.0201**	-16.3754**	-5.7392	5.4026	14.7368†	17.7470**	39.9428*
DSGE-Nonlinear versus TVP-VAR2	-50.0838*	-29.6565**	-16.0641**	-6.0242	5.3681	14.7405†	17.8504**	40.3903*

Note: See notes to Table 1. Negative (positive) entries indicate the gain (loss) in RMSEs from using the best model of a specific category relative to the RW model or the gain or loss from using the best DSGE-Nonlinear model relative to the best VAR, BVAR1, BVAR2, TVP-VAR1, or TVP-VAR2 models.

†, **, *   indicate that the forecast error significantly differs from the RW or the VAR, BVAR1, BVAR2, TVP-VAR1, or TVP-VAR2 models at 10%, 5% and 1% levels of significance respectively, based on the Giacomini and White (2006) statistic of equal forecast accuracy.

Figure 1: *Ex Ante* **Forecasts: 2007:Q1-2012:Q2**