FORECASTING AND MONETARY POLICY ANALYSIS. NEW EMPIRICAL EVIDENCE

by

ROQUE E. MONTERO

A Dissertation Submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Norman Swanson

And approved by

New Brunswick, New Jersey

May 2016

ABSTRACT OF THE DISSERTATION

Forecasting and Monetary Policy Analysis.

New Empirical Evidence

By ROQUE E. MONTERO

Dissertation Director:

Professor Norman Swanson

This dissertation presents new empirical evidence in two specific fields in economics: Forecasting and Monetary Policy Analysis. The dissertation comprises two separate but related papers, each of one tackles one of these two fields. The main objective aims to provide new lights and insights about specific questions that have been studied before in the literature, but using alternative methodological approaches to improve our understanding of the topics.

In Chapter 2, I argue that the out-of-sample forecast performance of nonlinear models for the conditional mean has being underestimated in the literature because these models are highly parameterized and hence parameter estimation error can easily offset their predictive gains. Thus, I consider restricted versions of nonlinear models that are commonly used in forecast competition between linear and nonlinear models. The restrictions aim to reduce the number of parameters to estimate allowing the specification of parsimonious nonlinear models. This setting explores more deeply the space of nonlinear models in order to find a suitable specification able to boost the performance of this type of models. The empirical evaluation is conducted using a linear benchmark and both global and local test of forecast predictive accuracy. The main results can be summarized as follow. First, if forecast comparison between

linear and nonlinear models excludes restricted nonlinear models then results are in line with previous findings. However, results change dramatically in some cases when restricted versions of nonlinear models are incorporated. In particular, I spot cases on which the mean square forecast error decreases by almost fifty percent relative to the benchmark model. These results give us new lights about the performance of nonlinear models and challenge the conventional view that the literature has about them because they show that their predictive gains may be elusive but that a simple exploration of their functional form may reveal significant predictive gains.

In Chapter 3, I investigate the propagation of a foreign monetary policy shock over a small open economy, in particular over the Chilean economy. This is an joint research project with Jorge Fornero and .Andrés Yany from the Central Bank of Chile. Our motivation is base on the ongoing period of monetary normalization already started by the Fed. We follow Canova (2007) and compare the impulse response functions of Structural VAR models and a DSGE model tailored for the Chilean economy. We use the recursive VAR model of Sims (1980a) and an extension of the "agnostic" VAR model of Uhlig (2005) and Arias et al. (2014) for small open economies following Koop and Korobilis (2010). The results suggest that the recursive VAR model does not properly identify the shock and its implications are counterintuitive. On the contrary, beyond the quantitative differences, we find that the responses of the "agnostic" VAR model are in line qualitatively with those of the DSGE model except for output. However, the transmission of the shock to the local economy is limited but more persistent according to the DSGE model. Finally, we spot different policy implication arising from both models. According to the "agnostic" VAR model, the central bank do not need to rise its policy rate because the drop in activity offsets any burst of inflation; whereas in the DSGE model the rise in prices is partially accommodated by an increase in the policy rate. Thus, this comparison motivates an interesting discussion for the policy maker.

Acknowledgements

First of all, I would like to thank my advisor, Professor Norman Swanson. It has been a real honor to write my dissertation under the guidance of such great mentor and I could not have imagined having a better advisor than him. His constant support over the years was crucial, especially during tough times when all hope seems lost.

I am also grateful to the rest of the members of my committee, Professors John Landon-Lane and Bruce Mizrach for their time, support and valuable suggestions provided to this thesis. I am also indebted to my outside member, Professor John Chao from University of Maryland for his suggestions and comments made to this thesis. Finally, I thank CONICYT for the financial support given over the years through the program BECAS CHILE.

My experience at Rutgers was also marked by extremely good teachers, especially Professors Roberto Chang, Richard McLean, Roger Klein, Hugh Rockoff and Eugene White. I deeply enjoyed their classes and the rigurosity and expertise of each of them in their fields.

I also want to thank to my co-authors of my third chapter, Jorge Fornero and Andres Yany from the Central Bank of Chile for their valuable efforts and time dedicated to this research project that started in a summer internship that I did in the Central Bank of Chile a couple of years ago. I am greatful for this joint research project since it has allowed me to attend to many international conferences which in turn helped me a lot to improve the quality of this chapter.

I also want to thank to my classmates Atisha Ghosh, Vladimir Kolchin and Jessica Scheld for helped me enduring the first year of the program. I will always remember the endless study hours in the basement of the Economics Department. Also, I want to thank to Antonio Cusato, Yuliyan Mitkov, Georgia Bush, Geoffrey Clarke, Ryuichiro Izumi and Freddy Rojas for their companionship and friendship over the years. I will always remember the coffee talks and lunches on which we shared laughs, experiences, frustrations and research ideas. I could not have imagined have survived all of these years without them.

Finally, I would like to thank my family for their love and unconditional support throughout my doctoral studies and my life in general. This PhD degree is not only filled with my personal sacrifices, it is also filled with sacrifices made by many other people, in particular over the past two years. I thank to God for the family that I have. My greatest admiration and gratitude goes to my mother for her unflagging strength; her sacrifices are the sole reason why I am here today finishing this chapter of my life.

Dedication

To my parents, Isabel and Juan

Contents

List of Tables

List of Figures

Chapter 1
Introduction

This dissertation will present empirical research in two specific fields in empirical macroeconomics: Policy Analysis and Forecasting. Each field by it self is important. Policy Analysis helps economists to quantify and understand how the economy will react after a given impulse (monetary policy shock, productivity shock, uncertainty shock, credibility shock, etc). Unveil the driving forces of the economy, such as propagation or amplification mechanism, is extremely important to any policy maker with some objective in mind, such as in inflation targeting, full employment or consumption smoothing. Finally, forecasting is also a crucial input to policy makers, financial institutions or other agents alike. The reason is quite simple and appealing to understand: good forecasts lead to good decisions. The two fields usually come together quite frequently. Huge efforts and resources are oriented to understand and predict the state of the economy at a given horizon of time.

In the second chapter of this dissertation entitled "*A comparison of the predictive performance of restricted nonlinear models, linear models and forecast combination schemes*". I study the predictive performance of nonlinear models for the conditional mean to challenge the conventional view that the literature has about their predictive accuracy[1]. In particular, there are two main results that can be found in this literature: (1) nonlinear models are not able to outperform simple linear models and even if they do, the forecast gains are small or not statistically significant; and (2) forecast combination schemes usually outperform single (non)linear models. Thus, the main objective of this chapter is to challenge these two key results.

[1]see Diebold and Nason (1990), De Gooijer and Kumar (1992), Stock and Watson(1999), Clements, Franses and Swanson (2004), Marcellino, Stock and Watson (2006), Bates and Granger (1969), Clemen (1989), Deutsch, Granger and Teräsvirta (1994), Swanson and Zeng (2001), Hendry and Clements (2004), Elliot and Timmermann (2005), Stock and Watson (2004) and Frances (2011) for more details

The conjecture can be briefly summarized as follow, I argue that the out-of-sample forecast performance of nonlinear models has being underestimated in previous research due to a deficient specification in the functional form of these models. In particular, nonlinear models are highly parameterized and hence parameter estimation error can easily offset their predictive gains. Thus, I incorporate restrictions into the functional forms of nonlinear models in order to decrease the number of parameters and at the same time to increase their predictive accuracy. I call these models "restricted nonlinear models". I consider restricted versions of three type of nonlinear models that are commonly used in this literature: The Self-Exciting Threshold Autoregressive Model (SETAR), The Markov Switching Autoregressive Model (MSAR) and The Multilayer Perceptron Network (MLP):

The comparison includes the prediction from forty one models, including the linear AR model which is used as the benchmark model. Thus, these experiments allow me to tackle the first main result of the literature. To tackle the second result, I also consider several forecast combination schemes. Thus, the comparison includes the predictions of forty one models and thirteen combinations schemes for three forecast horizons ($1, 3, 12$-step-ahead forecast).

To this end, I use three key macroeconomic variables: the percent change in the industrial production index (IP), the percent change in the consumer price index (CPI) and the difference of the 10 year Treasury-bond yield (r10); and I use the Giacomini and White (2006) and Giacomini and Rossi (2010) tests of global and local predictive ability.

The main finding of this paper is to provide empirical evidence supporting the view that traditional forecast comparisons between linear and nonlinear models can be misleading since the predictive gains of nonlinear models may be elusive but that a simple exploration of the functional form may reveal significant predictive gains.

In the third chapter of this dissertation entitled "*Reassessing the effects of foreign* *monetary policy on output: new evidence from structural and agnostic identification procedures*", I investigate the propagation of a foreign monetary policy shock over the Chilean economy. This is an joint research project with Jorge Fornero and .Andrés Yany from the Central Bank of Chile. Having said that, I want to emphasise that discussion and interpretation of the main results are more biased toward my own interpretation of them.

From a policy point of view, this chapter is well-timed since the ongoing period of monetary normalization already started by the Fed. According to Canova (2007), the conventional approach to tackle these questions is to use: (1) recursive VAR models on which identification is achieved using a specific order of the variables of the system; and (2) a theoretical model on which the complete structure of the economy is specified (i.e., a DSGE model). Intuitively; recursive VAR models should be able to uncover the main dynamics of variables after a shock and therefore they should provide an empirical benchmark to real structural model (i.e. the DSGE model) in order to assess its potential bias or misspecification.

However, what the literature has found is that identification of monetary policy shocks through recursive VAR models has always been a subject of debate and different specifications and models may lead to different responses under a monetary policy shock, see Christiano et al. (1998); in addition the identification of the shock is full of puzzling results, such as the "price puzzle" explained in Sims (1992), Eichenbaum (1992), Bernanke et al. (2005), Weber et al. (2009) and Castelnuovo (2015).

Therefore, we extend the conventional approach and include a novel "agnostic" approach that combines the works of Uhlig (2005) and Arias et al. (2014) with Koop and Korobilis (2010). We call this model the "agnostic" VAR model. In this setting, the identification of the shock is done by restricting the impulse responses functions directly according to conventional wisdom without relying in a particular order of the VAR system. Thus, this approach offers a challenging benchmark that allows to reassess the propagation of foreign monetary policy shocks and the policy implication derived from DSGE models. In particular, we analyze the propagation of this shock with a DSGE model tailored to the Chilean economy.

Our results suggest that recursive VAR models do not identify properly the shock and its implications are counterintuitive. On the contrary, beyond the quantitative differences, we find that the responses of the "agnostic" VAR model are in line qualitatively with those of the DSGE model except for output. However, the transmission of the shock to the local economy is limited but more persistent according to the DSGE model; which in turn will imply different policy implication for the central bank. Thus, the comparison unveils an interesting discussion for the policy maker and it highlights the benefits of using alternative approaches to better understand the propagation of shocks through the economy.

Chapter 2
A comparison of the predictive performance of restricted nonlinear models, linear models and forecast combination schemes

2.1 Introduction

Over the last three to four decades, a lot of research has been conducted analyzing the out-of-sample forecast performance of nonlinear models for the conditional mean. The motivation for this line of research was supported under the premise that nonlinear models should be able to better capture the nonlinearities that naturally arise in economic data[1] in order to generate superior forecasts compare to the ones obtained from linear models. However, the statistical evidence systematically showed that these type of models were not able to forecast best than a simple AR model. In particular, two main results can be found in this literature: (1) nonlinear models are not able to outperform simple linear models and even if they do, the forecast gains are small or not statistically significant; see Diebold and Nason (1990), De Gooijer and Kumar (1992), Stock and Watson(1999), Clements, Franses and Swanson (2004) and Marcellino, Stock and Watson (2006); (2) forecast combination schemes usually outperform single (non)linear models; see Bates and Granger (1969), Clemen (1989), Deutsch, Granger and Teräsvirta (1994), Swanson and Zeng (2001), Hendry and Clements (2004), Elliot and Timmermann (2005), Stock and Watson (2004) and Frances (2011). These results are counterintuitive since nonlinear models arise naturally in time series econometrics as a promising framework to exploit the predictive

[1]For instance, financial variables can be characterized by a two-regime process: low and high volatility, output growth as a two-regime process (recessions and expansions), inflation as a two-regime process in countries on which the Central Bank explicitly defines a band for inflation (i.e. inside and outside the band), zero-lower bound for the interest rates, exchange rate bands, structural breaks, time varying pass-through of exchange rates, etc.

content of nonlinearities in the data. Moreover, this nonlinear dependency between economic variables is also suggested by structural models, such as RBC models.

The first result is tackled comprehensively by Diebold and Nason (1990). In particular, they identify four possible explanations: (1) The degree of nonlinearity in the data; if nonlinearities are not significant in the data generating process (DGP) then linear models should forecast best. This view favors the parsimonious argument in time series forecasting. (2) False rejection of the null hypothesis of linearity. The rejection of the null may be due to outliers or structural breaks which in turn will lead the econometrician to specify a nonlinear model instead of a linear one. (3) Nonlinearities may exist in higher even-ordered conditional moments which can not be exploited efficiently by focusing in the first moment (i.e., the conditional expectation). (4) Misspecification; statistical tests may be able to reject linearity but they are not informative about how you should proceed afterwards; a poor specification of the nonlinear model will have a significant impact in its performance.

The second result has been reported in many empirical applications since the seminal work of Bates and Granger (1969)[2]. In an ever evolving DGP, forecast combination arises as an alternative method to produce superior forecasts when all the models are misspecified. However, at the same time it implies a consistent inability of the forecaster to specify a suitable approximation to the DGP which may be a questionable implication of this literature.

Therefore, the main objective of this paper is to challenge these two key results

[2]The main contribution of this paper was to derive the combination weights using an optimality condition to minimize the overall variance of the forecast errors. Since their seminal paper, several combination methods have been proposed. For instance, Deutsch et al. (1994) analyse the performance of nonlinear forecast combination methods on which the weights are chosen according to an observable state variable. More recently, Elliot and Timmermann (2005) study the forecast gains derived from markov switching combination schemes. Other authors have tackled the combination problem in a model selection framework, as in Swanson and Zeng (1997) and Frances (2011). In these setting, only the best models are chosen to construct the weights due to the collinearity of the forecasting models. A complete review of this literature is far from the scope of this paper, but the interested reader can refer to two comprehensive literature reviews done by Clemen (1989) and Diebold and Lopez (1996).

from the literature. I argue that the out-of-sample forecast performance of nonlinear models has being underestimated in previous research due to a deficient specification in the functional form of these models. Nonlinear models are highly parameterized[3] and hence parameter estimation error can easily offset their predictive gains. However, the literature has not addressed this point properly in previous forecast comparisons exercises and hence the empirical results that support the current view about their performance are incomplete. Thus, I consider restricted versions of three type of nonlinear models that are commonly used in this literature: The Self-Exciting Threshold Autoregressive Model (SETAR), The Markov Switching Autoregressive Model (MSAR) and The Multilayer Perceptron Network (MLP). The restrictions aim to reduce the number of parameters to estimate allowing the specification of parsimonious nonlinear models. The complete set of models (henceforth models or methods) are evaluated with respect to a linear univariate benchmark (i.e. the autoregressive model). This model rich environment allow us to explore more deeply the space of nonlinear models in order to find a suitable specification able to boost the performance of these type of models[4]. In addition, this framework may lead to a dynamic model able to outperform traditional forecast combination schemes. On the contrary, if this search over the space of nonlinear models fails, then it constitutes strong evidence against these types of models and confirms the previous findings of this literature.

My approach relates to two of the four points developed by Diebold and Nason (1990); more precisely with points (1) and (4) outlined earlier: the degree of nonlin-

[3]Even the simplest nonlinear model can easily have twice as many parameters than a linear one and the degrees of freedom vanish quite rapidly when more complex nonlinear models are take it into account. This is known as the the curse of dimensionality.

[4]To motivate the main argument, considered the following example. For instance, in the Markov Switching Autoregressive Models (MSAR), the parameters of an autoregressive model change between states according to a hidden markov process. However, if the only parameter that changes between the states is the constant term and this restriction is not included in the model, then parameter estimation error increases and it may dominate any predictive gains from this type of model. The inclusions of these restrictions may offer significant predictive gains.

earity in the data and misspecification. I do not focus my attention in the remaining two points since in-sample tests do not necessarily imply good out-sample performance and because the focus of this paper is on the out-of-sample performance of the models for the conditional mean, i.e. the first moment.

To this end, I use three key macroeconomic variables that the Federal Reserve considers to conduct their monetary policy according to Armah and Swanson (2011): the percent change in the industrial production index (IP), the percent change in the consumer price index (CPI) and the difference of the 10 year Treasury-bond yield (r10); similarly to Rossi and Sekhposyan (2010) but I enhance their work by incorporating nonlinear models and different forecast horizons. The forecast comparison is performed using two tests: (1) the unconditional Giacomini and White (2006) test (GW) of predictive ability to evaluate the average (or global) performance of the models; (2) the fluctuation test (FL) of Giacomini and Rossi (2010) (GR) to evaluate their local performance since the GW test may hide important information about the performance of the models through time. The FL test may be able to identify windows of opportunities as it was pointed out by Clements, Franses and Swanson (2004). The models are also compared with different forecast combination schemes, as in Rapach et al. (2010), to show that in some cases a careful forecaster should be able to identify a suitable forecast model able to outperform these combination schemes. This comprehensive evaluation framework along with the model rich environment, will give us new lights and insights about the performance of nonlinear models and enables me to challenge the results from this literature. This novel approach to assess the predictive performance of nonlinear models has not been considered in the empirical literature so far, or at least to this level.

In addition, the average performance of the models is evaluated in two subsamples, before and after the financial crisis of 2008, to identify breaks in the performance of the models due to this unique event. Finally, In order to compare my findings with previous results from the literature and to stress my argument, I perform the forecast comparison between linear and nonlinear models for the case on which the econometrician fails to consider the restricted versions of nonlinear models. This exercise is extremely helpful since it provides a benchmark to judge the empirical results and conclusions from this paper.

The main findings can be summarized as follows. First, results of forecast comparison between linear and univariate nonlinear models show results in line with the two key findings outlined earlier. But, the results change dramatically in some cases when restricted nonlinear models are included. The global evaluation shows interesting results for the industrial production index (IP), I spot cases on which the relative mean square error with respect to the benchmark model (rMSFE) decreases between thirty and forty five percent (30% - 45%) depending on the forecast horizon. In particular, the multivariate restricted versions of the markov switching model show the highest performance and they are also able to outperform the forecast combination schemes. For the interest rate (r10) the evidence is less conclusive since the predictive gains are not as sharp as before. The greatest gains for nonlinear models are reported at longer forecast horizons but linear models forecast best at shorter horizons. The local evaluation for IP shows that most of the models have an overall improvement with respect to the benchmark model after 2008, whereas for r10 the performance of the models fluctuates considerably through time. Regarding the consumer price index (CPI), the global and local evaluation show that the benchmark model consistently outperforms the rest of the models. However, the local evaluation shows that at longer forecast horizons the multivariate versions of the threshold model report promising results and further research in this line may reveal significant predictive gains from these type of models. Finally the sub-sample evaluation shows that the results for IP do not change significantly. This is not the case for r10 since several models are able to outperform the benchmark model before 2008 but not after[5]. In conclusion, this paper has provided empirical evidence supporting the view that forecast comparisons between linear and nonlinear models can be misleading since the predictive gains of nonlinear models may be elusive but that a simple exploration of the functional form may reveal significant predictive gains.

Other authors have tackled the model specification issue before in this context but from a different approach. For instance Swanson and White (1995, 1997a, 1997b), Teräsvirta et al. (2005) and Teräsvirta (2006), show that adaptive procedures on which the optimal number of regressors is chosen each time a new observation becomes available from the evaluation window, can offer significant predictive gains for nonlinear models. Alternatively, other authors have applied a different strategy to keep parameter estimation error under control by first testing the linearity hypothesis in the data, as in Teräsvirta (1998). However, in-sample properties do not necessarily imply good out-of sample performance. Instead, I follow a different approach and explore the predictive gains due mainly to the functional form of the models.

The rest of the paper is organized as follows; the next section describes the unrestricted and restricted nonlinear models used to predict the target variables. Section three describes the forecast exercise. Section four reports the results and the last section concludes.

2.2 Linear and Nonlinear Models

Define object y_t as the endogenous or target variable of a dynamic model; ε_t as a noise component, θ as a vector of parameters and $w(z_t, x_t) = (z_t', x_t')'$ as a vector of exogenous or predetermined variables with $z_t \in R^p$, $x_t \in R^k$ and $w \in R^{p+k}$; where $z_t = (y_{t-1}, ..., y_{t-p})'$ is a vector that contains the autoregressive part of the model (i.e., lags of the endogenous variables) and $x_t = (x_{1t}, ..., x_{kt})'$ a vector of strongly

[5]This results may be due to the unconventional monetary policies conducted after 2008 (zero lower bound) since interest rates were following a complete different dynamic after 2008

exogenous variables. Without loss of generality a dynamic model with an additive noise component can be defined as follows:

$$y_t = g\left(w(z_t, x_t); \theta\right) + \varepsilon_t \tag{1}$$

Where $g(\cdot)$ is a generic function that maps the exogenous variables and parameters to the output variable y_t. Depending on the vector x_t, the univariate or multivariate version of the model can be specified. I consider three types of models: The Self-Exciting Threshold Autoregressive Model (SETAR), The Markov Switching Autoregressive Model (MSAR) and The Multilayer Perceptron Network (MLP). These models are special cases of (1) and they have been used many times in the empirical literature. The next sections describe the unrestricted and restricted versions of these models[6].

2.2.1 Unrestricted models

The linear AR model (henceforth benchmark model) is defined as:

$$y_t = \beta_0 + \beta' w(z_t, x_t) + \varepsilon_t \tag{2}$$

With $\beta_0 \in R^1$ and $\beta \in R^{p+k}$. The parameters are estimated by minimizing the sum of squares of the residuals (SSR) of the model. One of the first extension to this model was the inclusion of threshold variables to allow for different regimes for the target variable. The m-regime SETAR model (henceforth threshold model) incorporates a regime switching process for y_t that depends on its past values (i.e., the regime is observable at time period t). Following the notation of Hansen (2000), the model is

[6] A detailed description of each nonlinear model is available in Appendix 2.A for the interested reader.

defined as follows:

$$y_t = \sum_{i=1}^{m} (\beta_{0,i} + \beta_i' w(z_t, x_t)) I_{it}(\gamma, d) + \varepsilon_t \quad (3)$$

Where $\beta_{0,i} \in R^1$, $\beta_i \in R^{p+k}$ for $i = 1, ..., m$ and $\varepsilon_{t,} \sim N(0, \sigma^2)$. The object $I_{jt}(\gamma, d) = I(\gamma_{j-i} < y_{t-d} \leq \gamma_j)$ is the indicator function that takes the unit value if the internal condition holds and zero otherwise, with d as the delay parameter ($d < t$) and $\gamma = (\gamma_1, ..., \gamma_m)$ with $\gamma_1 < \gamma_2 < ... < \gamma_m$ as the vector of threshold parameters. The parameters are estimated using a sequential approach to minimize the SSR of the model.

One of the assumption behind the threshold model is that the state variable that defines the regime of the process is observable at time t. Hamilton (1989) introduces a generalization of this model, the m-regime MSAR model. In this model the regimes (or states) are unobservable and the model uses the data to make an inference about the state at time t, this is known as filtering. The m-states MSAR model is defined as:

$$y_t = \beta_{0,s_t} + \beta_{s_t}' w(z_t, x_t) + \varepsilon_{t,s_t} \quad (4)$$

Where $\beta_{0,s_t} \in R^1$ and $\beta_{s_t} \in R^{p+k}$ for $s_t = 1, ..., m$; the state variable s_t evolves following an independent markov chain and $\varepsilon_{t,s_t} \sim N(0, \sigma^2_{s_t})$. The parameters are estimated by quasi-maximum likelihood.

Finally, neural networks are a class of nonlinear model that are flexible enough to approximate any unknown function given enough structure. This is known as Universal Approximation Theorem, see Cybenko (1989). For a comprehensive discussion of this type of models, see Kuan and White (1994). In addition, interesting application can be found in Swanson and White (1995, 1997a, 1997b). Thus, these type of models provide a flexible functional form to approximate the conditional mean of a target variable. Neural networks decompose (1) into two components; a linear and nonlinear component. Following the notation of McNelis (2005), the one layer Multilayer

Perceptron Network (MLP) with v^* hidden units is defined as:

$$y_t = \beta_0 + \sum_{i=1}^{p+q} \beta_i w_i + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(w(z_t, x_t)) + \varepsilon_t \tag{5}$$

$$N_{v,t}(w(z_t, x_t)) = \frac{1}{1 + e^{-n_{v,t}}}$$

$$n_{v,t} = \omega_{v,0} + \sum_{i=1}^{p+q} \omega_{v,i} w_i$$

Where w_i is the ith component of the vector $w(x_t, z_t) \in R^{p+q}$, $\beta_0 \in R^1$ and each β_i and γ_v are in R^1. In addition, $N_{v,t}(\cdot)$ is the logistic squashing function with $\omega_{v,0} \in R^1$ and each $\omega_{v,i} \in R^1$. The parameters are estimated by minimizing the SSR. In this paper I use a two step procedure to minimize this function; in the first stage the function is optimized using a genetic algorithm, whereas in the second stage the function is optimized using a Quasi-Newton algorithm that uses the output from the previous stage as the starting values for the algorithm.

2.2.2 Restricted nonlinear models

Two constraints are imposed over the threshold models: (1) the number of regimes is set to two ($m = 2$) and (2) each regime has the same set of exogenous variables ($w(z_t, x_t)$). This setting defines the 2-regime SETAR model:

$$\text{SETAR:} \quad y_t = (\beta_{0,1} + \beta_1' w(z_t, x_t)) I_{1t}(\gamma, d) + (\beta_{0,2} + \beta_2' w(z_t, x_t)) I_{2t}(\gamma, d) + \varepsilon_t$$

The MSAR type of models are restricted in two ways: (1) the number of regimes is set to two ($m = 2$) and (2) restrictions over the parameters that are state dependent. The unrestricted MSAR is defined as:

$$\text{MSARc(1):} \quad y_t = \beta_{0,s_t} + \beta_{s_t}' w(z_t, x_t) + \varepsilon_{t,s_t}$$

Whereas the restricted versions of this model are defined as follow:

$$\text{MSARc(2):} \quad y_t = \beta_{0,s_t} + \beta' w(z_t, x_t) + \varepsilon_t$$

$$\text{MSARc(3):} \quad y_t = \beta_0 + \beta'_{s_t} w(z_t, x_t) + \varepsilon_t$$

$$\text{MSARc(4):} \quad y_t = \beta_0 + \beta' w(z_t, x_t) + \varepsilon_{t,s_t}$$

$$\text{MSARc(5):} \quad y_t = \beta_0 + \beta'_{s_t} z_t + \beta' x_t + \varepsilon_t$$

Note that in the MSARc(2) specification only the constant term is state dependent whereas in the $MSARc(3)$ all the parameters of the conditional mean are state dependent but the constant term. In the $MSARc(4)$ model only the noise component is state dependent. Finally in the $MSARc(5)$ specification, only the autoregressive part of the model is state dependent.

Finally, the restricted MLP networks consider different combinations of variables for the linear and nonlinear components of the network. The unrestricted network is defined as:

$$\text{MLP(1):} \quad y_t = \beta_0 + \textstyle\sum_{i=1}^{i^*} \beta_i w(z_t, x_t) + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(w(z_t, x_t)) + \varepsilon_t$$

Whereas the restricted networks are defined as follow:

$$\text{MLP(2):} \quad y_t = \beta_0 + \textstyle\sum_{i=1}^{i^*} \beta_i w_i + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(x_t) + \varepsilon_t$$

$$\text{MLP(3):} \quad y_t = \beta_0 + \textstyle\sum_{i=1}^{i^*} \beta_i w_i + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(z_t) + \varepsilon_t$$

$$\text{MLP(4):} \quad y_t = \beta_0 + \textstyle\sum_{i=1}^{i^*} \beta_i x_t + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(z_t) + \varepsilon_t$$

$$\text{MLP(5):} \quad y_t = \beta_0 + \textstyle\sum_{i=1}^{i^*} \beta_i z_t + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(x_t) + \varepsilon_t$$

$$\text{MLP(6):} \quad y_t = \beta_0 + \textstyle\sum_{i=1}^{i^*} \beta_i z_t + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(z_t) + \varepsilon_t$$

$$\text{MLP(7):} \quad y_t = \beta_0 + \textstyle\sum_{v=1}^{v^*} \gamma_v N_{v,t}(z_t) + \varepsilon_t$$

The first two specifications differ in the variables included in the squashing function $N_{v,t}(\cdot)$. The networks MLP(4) and MLP(5) restrict the vector of exogenous variables of the model and considerer different combination of exogenous variables for the linear and nonlinear part of the model. The MLP(6) specifies a univariate network, since only the lags of the dependent variables are used to predict the conditional mean. Finally, MLP(7) is a restricted version of the MLP(6) network on which the linear component is shut out.

2.3 Forecasting

The h-step-ahead point forecast for variable y_{t+h} from model i is defined as:

$$\hat{y}^{i}_{t+h|t} = g\left(w(z_t, x_t); \hat{\theta}_t\right)$$

Where $\hat{\theta}_t$ is the vector of parameter estimates obtained using information up to time t. Note that the point forecast is computed directly; the literature refers to this approach as the direct forecast method. An alternative way to produce the h-step-ahead point forecast is the recursive method; in this setting the dynamic model is used to compute the one-step ahead forecast and the h-step-ahead forecast is computed by the iteration of the original model. The main advantage of the direct over the recursive method is that it avoids the numerical issues involved in the multi-step ahead forecast (i.e. $\hat{y}_{t+h|t}$ for $h \geq 2$) in the context of nonlinear models, see Stock and Watson (1999), Teräsvirta et al. (2005), Marcellino et al. (2006) and Bredahl and Teräsvirta (2010) for a detailed explanation of this issue.

The pseudo out-of-sample predictions are obtained for three key economic variables for the American economy: the percent change in the industrial production index (IP), the percent change in the consumer price index (CPI) and the difference of the 10 year Treasury-bond yield (r10). Similarly to Rossi and Sekhposyan (2010) but I enhance their work by incorporating (un)restricted nonlinear models and different forecast horizons (h = 1, 3, and 12). The data are monthly observations covering the period from January 1960 to December 2014. The model selection works as follows. The order of the autoregressive part of the model (z_t) is chosen according to the Bayesian information criterion (BIC) or Schwarz criterion[7] for each forecast horizon

[7] Where $BIC = T \cdot ln(\hat{l}) + k \cdot ln(T)$; with $\hat{l}$ as the log-likelihhod of the model, T the sample size and k the number of parameters of the model. Equivalently, under the assumption of normality and independence in the error term, the BIC can also be defined as:

$$BIC = T \cdot ln(\hat{\sigma}^2_{\varepsilon}) + k \cdot ln(T)$$

using data from the period 1960.01 to 1985.12 (estimation window).

Multivariate models incorporate predetermined variables and lag values of the target and predetermined variables; whereas univariate models include only lags of the target variable. To simplify the forecast comparison, I restrict the number of predetermined variables to three and define three type of multivariate nonlinear models. The first type of models (x1) include predetermined variables dated at time t (i.e., (x_t)), the second type of models (x2) also include the first lags of the predetermined variables (i.e., (x_t, x_{t-1})) and the third type of models include the first and second lags of the predetermined variables (i.e., (x_t, x_{t-1}, x_{t-2})). See Table B.2.1 for a detailed description of the predetermined variables chosen for each target variable. In particular some variables are chosen due to the close link suggested by economic theory and others due to the information they have about the state of the economy. For instance, DSGE models suggest a close connection between interest rates and activity through a Taylor rule or between unemployment and inflation through a Phillips curve. Finally, the leading economic and supplier deliveries indexes are diffusion index hence contain information from many sectors of the economy and thus they could contain valuable information to predict the target variables[8].

I follow this approach because the main objective of this paper is to explore the predictive gains due mainly to the functional form of the model. In general terms, the pseudo-out of sample evaluation works as follow. The full sample is divided in two windows: the estimation window (R) and the evaluation window (P). The models are estimated using the estimation window and forecasts are made for the evaluation windows. The models are then re-estimated each time new information is added from the evaluation to the estimation window and new forecasts are produced. The process continues until the evaluation window is exhausted. Finally, models are

Where $\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$.

[8]This argument do not imply that these are the best predetermined variables to predict the target variables. It just clarifies the selection rule. The inclusion of more variables could potentially improve the performance of the models discussed in this paper.

compared using statistical tests or statistics to summarize their properties.

Forecasts for each model are computed as follow. For the MSAR type of models, the h-step-ahead forecast is computed as a convex combination of the predictions of the states: $\hat{y}_{t+h|t} = h'_k \hat{\xi}_{t+k|t}$; where h_k is the vector of forecasts and $\hat{\xi}_{t+k|t}$ the vector with the states probabilities. Forecasts for the SETAR models are computed directly since the regime is known at time t, and forecasts for the linear models and neural networks are computed directly. Nonlinear models may be quite sensitive to outliers or other features of the data and thus in some experiments these models may produce extreme predictions that a rational forecaster would never use. To mimic this process, an automatic trim procedure was included in the MSAR and MLP type of models in the same spirit of Swanson and White (1995,1997a,1997b); whenever the model produces a prediction larger that any previously observed value over the estimation window, I keep the last forecast (i.e. no change in forecast). The full forecast comparison incorporates the predictions from forty one models[9] for each variable at each forecast horizon.

The comparison is performed using the relative mean square forecast error with respect to a linear benchmark model, i.e. the AR model of order p. The mean square forecast error (MSFE) of model i at forecast horizon h is defined as:

$$MSFE_{i,h} = \sum_{t=R-h+2}^{T-h+1} \left(y_{t+h} - \hat{y}^{i}_{t+h} \right)^2$$

A relative mean square forecast error (rMSFE) greater than one indicates that the benchmark model outperforms the alternative model, whereas the opposite holds if the value is less than one. Two test are use to assess the statistical significance of the results: the Fluctuation test (FL) of Giacomini and Rossi (2010) (GR) and the unconditional Giacomini and White (2006) test (GW) of predictive abil-

[9] More precisely: four AR models, four SETAR models, twenty MSAR models and thirteen MLP networks.

ity. This framework requires two conditions: (1) stationarity of the target variable y_{t+h} and (2) rolling window estimation scheme. To address the first point the target variables are transformed as follows[10]: $y^{ip}_{t+h} = 1200ln\left(IP_{t+h}/IP_t\right)/h$, $y^{cpi}_{t+h} = 1200\left(ln\left(CPI_{t+h}/CPI_t\right)/h - ln\left(CPI_t/CPI_{t-1}\right)\right)$ and $y^{r10}_{t+h} = r10_{t+h} - r10_t$. For the second condition, forecasts are made following the procedure outlined earlier but the sample size of the estimation window is kept constant by dropping the first observation each time new information is added form the evaluation window.

Finally, the comparison includes several forecast combination schemes with fix and time-varying combination weights, thirteen combination schemes in total[11]. The latter approach requires the use of a training window to estimate the weights and therefore the evaluation window needs to be adjusted to take this into account.

2.3.1 Forecast combination schemes

The composite h-step-ahead forecast is defined as:

$$y^{c}_{t+h|t} = f\left(y^{1}_{t+h|t}, ..., y^{M}_{t+h|t}; \varpi\right)$$

Where $f(\cdot)$ is a (non)linear combination function[12], M the total number of models, $y^{i}_{t+h|t}$ is the conditional forecast from model i and ϖ is the vector of weights, see Bates and Granger (1969), Clemen (1989), Elliott and Timmermann (2005) and Frances (2011) for a comprehensive discussion of this literature.

Table 2.1 describes and summarizes the combinations schemes considered in the experiments. Two types of weights can be found in the literature: fix and time-varying combination weights. In the former type, the weights are independent of the forecast accuracy of the models (for instance the mean or median of the forecasts). In the

[10]The transformation for the CPI follows the work of Stock and Watson (2003) and assumes that the index is $I(2)$. The case of I(1) was also considered but the results did not change significantly. The results for this case are available at Table B.2.2 in Appendix section B.

[11]Thus, the full comparison includes the predictions from fifty four models.

[12]See Deutsch et al. (1994) for an application of nonlinear combination functions.

Table 2.1
Forecast combination schemes

Label	Description
Fix combination weights	
mean	Arithmetic mean all models.
med	Median all models.
mAR	Arithmetic mean AR models.
mST	Arithmetic mean SETAR models.
mMS	Arithmetic mean MSAR models.
mMLP	Arithmetic mean MLP networks.
Time-varying combination weights	
tmean	Arithmetic mean all models (trimmed by 20%).
best	Selects the best model based on the absolute prediction error from last period.
best5	Arithmetic mean of the five most accurate models, based on the absolute prediction error from last period.
MSEs90	Selects the best model from the set of models for which the rMSE < 90%; otherwise the arithmetic mean of all models is used.
MSEm90	Arithmetic mean of the five most accurate models, according to the MSE.
OLSr	OLS bias correction of the best model in terms of MSE. Weights are computed using the model: $y^c_{t+h} = \beta_0 + \beta_1 \hat{y}^{best}_{t+h\|t} + \varepsilon_{t+h}$
MSEw	Weighted average all models: $y^c_{t+h\|t} = \sum_{i=1}^{41} \varpi_i y^i_{t+h\|t}$ and $\varpi_i = \frac{MSE_i}{\sum_{i=1}^{41} MSE_i}$ where MSE_i is the MSE of model "i".

Where MSE is the mean square prediction error and rMSE is the relative MSE defined as the MSE of model "i" divided by the MSE of the benchmark model (i.e. the univariate AR model). The MSE for the time-varying combination schemes are computed using the out-sample training window (5 years of data).

latter type, the weights depend on their in/out-of sample performance of the models and thus they have to be estimated. In this paper, the time-varying combination weights are estimated according to their out-of sample performance by using a rolling training windows of five years of data starting with the period of 1986.01 to 1990.12. Therefore, the evaluation window for the experiments is adjusted to take this into account and thus the final evaluation sample covers the period of 1991.01 to 2014.12 (i.e., the last twenty four years of data).

2.3.2 Local and global test of average equal predictive ability

The performance of the models and forecast combination schemes is evaluated with respect to the benchmark model in terms of their local and global performance. Several statistical tests are available in the literature to test the hypothesis of global average equal predictive ability, like the Diebold and Mariano (1995) (DM) test which probably remains as one of the most used test to compare the forecast accuracy between models. According to Diebold (2013), the DM test was developed to compare two competing forecasts and it was not intended for the comparison of competing models in pseudo-out of sample experiments since more sophisticated methods exist for this purpose. In the DM framework, the forecast errors can have non-zero mean, serially or contemporaneously correlated and no distributional assumption is made over them. In addition, the asymptotic distribution of the test is the standard normal distribution. Thus, the elegance and simplicity of the procedure can help us to understand why it is still used in many empirical applications. Unfortunately the DM can not be used when the competing models are nested because in this case the limiting distribution of the test is not defined under the null hypothesis of equal predictive ability. Thus, we need to specify a different statistical procedure to evaluate the performance of the models since nonlinear models nest linear models, i.e. the benchmark model.

The forecast comparison between nested competing models has been investigated deeply in the literature, see for example Clark and McCracken (2001), Clark and West (2007), Giacomini and White (2006) and Clark and McCracken (2012). In this paper I use the Giacomini and White (2006) (GW) test because the limiting distribution under the null is standard and the test is robust to misspecification of the competing models. Thus, it follows the same spirit of the DM test but the GW test evaluates the forecasting models along with other choices that are involve in the prediction process, such as: estimation procedure or different data sets. Therefore, the test focuses more on the forecast method rather than the forecasting model which makes the comparison of nested and nonnested models feasible and straightforward since the limiting distribution of the test is standard[13]. The GW requires two conditions as it was explained at the beginning of this section: (1) stationarity of the target variable y_{t+h} and (2) rolling window estimation scheme.

Define $f(\theta)$ and $f'(\theta')$ as two competing forecast models (or methods) for target variable y_{t+h}. The parameters of each model are estimated using rolling windows

[13]One of the main drawbacks of alternative tests to compare nested models is that their limiting distribution under the null hypothesis are model dependent.

of data. Define $\hat{\varepsilon}_{t+h,f(\theta)}$ and $\hat{\varepsilon}_{tt+h,f'(\theta')}$ as the forecast errors for each model. In particular, define $f(\theta)$ as the benchmark model and $f'(\theta')$ as the alternative model. Under a quadratic loss function the null hypothesis of equal predictive ability is defined as:

$$H_0 : E\left(\hat{\varepsilon}^2_{t,f'(\theta')} - \hat{\varepsilon}^2_{t,f(\theta)}\right) = 0 \ for \ t = R + h, ..., T$$

The test can be implemented using the statistic:

$$t = \frac{\Delta\bar{L}}{\hat{\sigma}_P/\sqrt{P}}$$

Where $\Delta\bar{L} = \frac{1}{P}\sum_{\tau=1}^{P}\left(\hat{\varepsilon}^2_{\tau,f'} - \hat{\varepsilon}^2_{\tau,f}\right)$; P as the sample size of the evaluation window; $\hat{\sigma}_P$ as a Heteroskedasticy and Autocorrelation Consistent[14] (HAC) estimator of the asymptotic variance of $\left[\sqrt{P}\Delta\bar{L}\right]$ and $t \sim N(0,1)$. A two-tailed test identifies which method forecasts best on average and under H_0 both models have the same predictive performance on average. However, the GW test may hide important information regarding the performance of the models. For instance, the test can reject the null hypothesis between two similar competing models if one them have a small subset of large prediction errors; or it may be the case that each of the competing models forecast best in one part of the evaluation window, thus the GW test may conclude that both models predict the target variable with the same accuracy[15].

Thus, to enhance the forecast comparisons between models, I also perform a local evaluation of the forecast accuracy in addition to the global evaluation. This framework provides a comprehensive evaluation of the predictive performance of the

[14]For example, as in Rossi and Sekhposyan (2010):

$$\hat{\sigma}_P = \sum_{i=-q(P)+1}^{q(P)+1}(1 - |i/q(P)|)P^{-1} \times \sum_{j=R+H}^{T}(\hat{\varepsilon}_{j,f} - \hat{\varepsilon}_{j,f'})(\hat{\varepsilon}_{j-1,f} - \hat{\varepsilon}_{j-1,f'})$$

Where $q(P)$ is a bandwidth that grows with the sample size of the evaluation window.

[15]Following Clements et al. (2004), nonlinear models may have windows of opportunity on which they can outperform linear specifications but the GW test may be unable to detect these cases since it focuses on the average performance of the models.

alternative models with respect to the linear benchmark model and it has not being explored before in the context of nonlinear models. The local average predictive ability is measured with the Giacomini and Rossi (2010) fluctuation test (FL). Under the null hypothesis both models have the same predictive ability at each point in time. The FL test follows the same spirit of the GW test since it measures the difference between the mean square forecast errors but using a moving window of m periods:

$$\Delta \bar{L}_t = \frac{1}{m}\left(\sum_{j=t-m/2}^{j=t+m/2} \hat{\varepsilon}^2_{j+h,f'} - \hat{\varepsilon}^2_{j+h,f}\right)$$

The FL test is defined as:

$$F_{t,m} = \hat{\sigma}^{-1} m^{-1/2} \times \left(\sum_{j=t-m/2}^{j=t+m/2} \hat{\varepsilon}^2_{j+h,f'} - \hat{\varepsilon}^2_{j+h,f}\right)$$

Where $\hat{\sigma}$ is HAC estimator of the asymptotic variance of the statistic. The test plots the sample path of $F_{t,m}$ along with critical values: $cv_{up} > 0$ and $cv_{lw} < 0$. If $F_{t,m}$ crosses one of the critical values then it signals that one of the model was outperformed by the other one at some point. In particular, if $F_{t,m} < cv_{lw}$ the alternative model forecasts best, if $F_{t,m} > cv_{up}$ the benchmark model forecasts best and if $F_{t,m} \in [cv_{lw}, cv_{up}]$ then both models have equal out-of-sample performance at each point in time. The main drawback of the FL test is that it does not have a standard distribution under the null hypothesis. Simulations for the critical values of the test are reported in the original paper of GR for several values of m (the size of the moving evaluation window). Following their results, m is set equal to $0.3P$ in order to maximize the power of the test.

2.4 Results

The results are divided in four sections for ease of exposition. The first two sections analyse the global average performance using the GW test. In particular, the first section performs the forecast comparison using univariate unrestricted nonlinear models and forecast combination schemes whereas the second section extends the analysis by including the full set of (un)restricted nonlinear models. This comparison unveils the potential bias from the literature. The third section evaluates the local average predictive performance of the models using the fluctuation test of GR. Finally, the last section shows the comparison of the models before and after the financial crisis of 2008 to analyze the robustness of the global evaluation to this unique event.

2.4.1 Univariate unrestricted nonlinear models

Following the work of Stock and Watson (1999), Teräsvirta et al.(2005), Teräsvirta et al.(2006), Marcellino (2008), Bredahl and Tersvirta (2010), among others; Table 2.2 shows the relative mean square forecast errors (rMSFE) for univariate unrestricted nonlinear models, forecast combination schemes and multivariate AR models with respect to the benchmark model for three forecast horizons. The table also highlights the cases for which the null hypothesis of equal global average predictive ability is rejected according to the GW test.

Table 2.2 shows discouraging results from this exercise for univariate nonlinear models, this is specially true for CPI and r10. that the benchmark forecasts best than the alternative nonlinear models, this is especially true for the CPI and r10 at any forecast horizon. For IP the results slightly differ at $h = 1$, since multivariate versions of the AR(p) model outperform the benchmark model. But, it is worth noting that the univariate neural network MLP(6) shows a significant decrease in the rMSFE of almost ten percent (10%). The results for the forecast combination schemes suggest that they may provide predictive gains in some cases. Thus, this

Table 2.2
Relative mean square forecast errors (rMSFE) and GW test
Univariate unrestricted nonlinear models

	IP			CPI			r10		
Method	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12
ARx1	0.921*	1.234	1.132	1.015	1.054	0.848	1.033	1.132*	1.240*
ARx2	0.762**	0.746	0.869	1.009	1.061	0.842	1.028	1.136	1.397*
ARx3	0.771**	0.713	0.855	1.005	1.099	0.859	1.072	1.087	1.295**
SETAR	1.032	1.353*	1.381*	1.068	1.195*	1.516*	1.042*	1.502*	2.227*
MSARc(1)	0.976	1.056	1.133	1.031	1.377	1.322	1.008	1.115*	1.277*
MLP(6)	0.906**	1.166	1.201*	1.029	1.678*	2.741*	1.017	1.114*	1.242*
Forecast combination schemes									
mean	0.762*	0.761*	0.772*	0.954	1.036	1.161	0.963	0.959	0.986
med	0.781*	0.753*	0.794*	0.946	1.065	1.283	0.956	0.982	0.971
mAR	0.801*	0.770	0.846	1.002	1.037	0.809	0.983	0.983	1.078
mST	0.883**	0.934	1.109	1.001	1.104*	0.968	1.113*	1.071	1.509*
mMS	0.751*	0.813*	0.736*	0.954	1.116	1.284	0.935*	0.991	0.941
mMLP	0.816*	0.752*	0.878	1.024	1.192	1.553	0.982	1.014	1.104
tmean	0.761*	0.755*	0.778*	0.953	1.042	1.177	0.957	0.954	0.983
best	1.080	1.595*	1.243	1.195**	1.729*	2.661*	1.053	0.998	0.981
best5	1.083	1.596*	1.243	1.195**	1.725*	2.661*	1.053	0.996	0.980
MSEs90	0.944	0.689**	0.890	0.973	0.959	0.910	0.973	1.081	0.941
MSEm90	0.818*	0.678*	0.750*	1.158	1.075	0.950	0.913*	1.029	0.890
OLSr	1.098	0.796	1.339	1.042	1.113	0.973	1.008	1.075	1.187
MSEw	0.758*	0.732*	0.764*	0.956	1.030	1.103	0.962	0.960	0.957

Notes: Numerical entries in this table are relative mean square forecast errors (rMSFE) with respect to the AR(p) model (benchmark model). Forecast are monthly for the period 1991.01 to 2014.12. (P = 244). Numerical values less than unity indicate that the alternative model forecasts best. Where "modxi" is the label for model "mod" and "xi" for $i = 1, 2, 3$, indicates the vector x_t^i included in the multivariate version of the model. If no "xi" is attached, then the label refers to the univariate version of the model. The composite forecasts were made using the entire set of models (i.e., forty one models in total).
(*) Denoting rejection at the 5% level.
(**) Denoting rejection at the 10% level.

exercise supports the view of the literature: (1) nonlinear models are not able to outperform simple linear models and even if they do, the forecast gains are small or not statistically significant; and (2) forecast combination schemes usually provide an overall improvement with respect to single models.

At this stage a forecaster would conclude that the predictive gains of nonlinear models to forecast these three target variables are limited and probably nonexistent. Thus predictions for these variables should be made using univariate or multivariate linear models or forecast combination schemes. In addition, alternative methodologies should also be explored and evaluated. But, unrestricted nonlinear models are highly parameterized and their poor performance may be due to parameter estimation error. Thus, the inclusion of restricted nonlinear models may lead to significant improvements in their performance and Table 2.2 is in some way showing evidence of this conjecture. The result for IP at $h = 1$ shows that the univariate neural network exhibits a small significant improvement with respect to the linear benchmark model. This result is interesting because neural networks are capable to approximate any unknown function given enough structure; thus this result may be signaling that an exploration of the functional form of nonlinear models may identify specifications capable to offer significant predictive gains for this target variable. The next section incorporates restricted versions of nonlinear models along with multivariate restricted versions of them, which allow us to analyze the robustness of results and conclusions from this part of the analysis.

2.4.2 A comparison of the predictive performance of restricted nonlinear models, linear models and forecast combination schemes

The full comparison includes the predictions from forty one models (one benchmark model and forty alternative models) and thirteen forecast combination schemes; a total of fifty four models for each target variable at each forecast horizon. Table 2.3 reports the relative mean square forecast errors (rMSFE) of each model with respect to the benchmark and it also highlights the cases for which the null hypothesis of equal global average predictive ability is rejected according to the GW test.

The main results can be summarized as follow. For IP, results change dramatically with the ones reported before since several restricted nonlinear models are able to outperform the benchmark model. The evidence for r10 is less conclusive and the greatest forecast gains for nonlinear models are in longer forecast horizons. Finally, for CPI the results show that the linear benchmark model outperforms alternative methods at any forecast horizon, but multivariate versions of the threshold model show promising results at longer forecast horizons. Thus, these results put in perspective the findings from the previous section by showing the potential predictive gains derived from restricting the functional form of nonlinear models in order to decrease parameter estimation error and boost their predictive performance.

Table 2.3
Relative mean square forecast errors (rMSFE) and GW test

	IP			CPI			r10		
Method	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12
ARx1	0.921*	1.234	1.132	1.015	1.054	0.848	1.033	1.132*	1.240*
ARx2	0.762**	0.746	0.869	1.009	1.061	0.842	1.028	1.136	1.397*
ARx3	0.771**	0.713	0.855	1.005	1.099	0.859	1.072	1.087	1.295**
SETAR	1.032	1.353*	1.381*	1.068	1.195*	1.516*	1.042*	1.502*	2.227*
SETARx1	1.073	1.401*	1.524*	1.071	1.219**	0.954	1.167*	1.197*	1.480*
SETARx2	0.880**	0.842	1.092	1.057	1.186**	0.935	1.292*	1.273*	1.762*
SETARx3	0.978	0.812	1.006	1.109	1.153	1.062	1.338*	1.191**	1.682*
MSARc(1)	0.976	1.056	1.133	1.031	1.377	1.322	1.008	1.115*	1.277*
MSARc(1)x1	0.911	1.246	0.903	1.024	1.333	1.306	1.011	1.170*	1.082
MSARc(1)x2	0.747**	0.771	0.826*	1.064	1.125	1.457	0.921*	1.101	0.972
MSARc(1)x3	0.732	0.646**	0.716*	0.954	1.138	1.319	0.913*	1.010	0.881**
MSARc(2)	0.993	1.010	1.014	1.033	1.036	1.349	1.019	1.132*	1.364*
MSARc(2)x1	1.603	1.325	0.830	1.024	1.302	1.390	1.027	1.853*	0.998
MSARc(2)x2	0.735	0.816	0.834*	1.023	1.305	1.466	1.024	1.228*	0.948
MSARc(2)x3	0.794	0.554**	0.611*	0.916	1.305	1.455	0.900*	1.625*	0.874
MSARc(3)	0.911*	1.040	1.006	1.048	1.224	1.442	0.993	1.117**	1.030
MSARc(3)x1	0.998	1.367	0.830	1.022	1.174	1.311	0.960	1.295*	0.915
MSARc(3)x2	0.717**	0.804	0.817*	0.998	1.175	1.456	0.915*	1.139*	0.911
MSARc(3)x3	0.679**	0.731**	0.685*	0.993	1.137	1.267	0.985	1.026	0.878**
MSARc(4)	0.976*	1.035	0.986	1.037	1.340	1.375	0.992	1.015	1.127**
MSARc(4)x1	0.904	1.341	0.917	1.023	1.302	1.302	1.015	1.124*	0.995
MSARc(4)x2	0.729**	0.817	0.760*	1.023	1.298	1.453	1.012	1.002	0.983
MSARc(4)x3	0.725**	0.644**	0.602*	0.916	1.306	1.217	0.998	0.930	0.931
MSARc(5)	0.911*	1.040	1.006	1.048	1.224	1.442	0.993	1.117**	1.030
MSARc(5)x1	0.916	1.332	0.902	1.024	1.331	1.302	1.087*	1.124*	1.047
MSARc(5)x2	0.715**	0.799	0.817*	0.998	1.135	1.453	0.945	1.075	1.151
MSARc(5)x3	0.712**	0.600*	0.754*	0.993	1.137	1.320	0.983	0.989	1.057
MLP(1)x1	0.922*	0.958	0.990	1.019	1.258**	2.045*	0.992	1.086*	1.166*
MLP(1)x2	0.863**	0.973	0.985	1.122	1.089	1.557	1.057*	1.464	1.184*
MLP(1)x3	0.940	1.332	1.101	1.144**	1.175**	2.288*	1.047	1.229*	1.258*
MLP(2)x1	0.803**	0.807	0.935	0.997	1.216	2.933*	1.046	1.137	1.540*
MLP(2)x2	0.749**	1.077	1.394**	1.210*	1.421	1.537	1.088	1.185	1.361*
MLP(2)x3	1.049	1.109*	1.092	1.019	1.334	1.408	1.019	1.157*	1.264*
MLP(3)x1	1.841*	0.762	1.036	1.062	1.391**	1.696	1.017	1.265	1.434*
MLP(3)x2	0.774	0.637	1.097	1.067	1.537*	1.550	1.074	1.082	1.244
MLP(3)x3	0.865*	1.279	1.091	1.000	1.507*	2.129*	1.054	1.315*	1.313*
MLP(4)x1	0.826	0.892	1.172	1.184**	1.426**	1.701	1.158*	1.153	1.445*
MLP(5)x1	1.071	0.953	0.949	1.209**	1.218**	1.405*	1.103*	1.097	1.422*
MLP(6)	0.906**	1.166	1.201*	1.029	1.678*	2.741*	1.017	1.114*	1.242*
MLP(7)	0.845*	0.949	1.518**	1.250*	1.297*	1.657	1.022	1.065*	1.147*
Forecast combination schemes									
mean	0.762*	0.761*	0.772*	0.954	1.036	1.161	0.963	0.959	0.986
med	0.781*	0.753*	0.794*	0.946	1.065	1.283	0.956	0.982	0.971
mAR	0.801*	0.770	0.846	1.002	1.037	0.809	0.983	0.983	1.078
mST	0.883**	0.934	1.109	1.001	1.104*	0.968	1.113*	1.071	1.509*
mMS	0.751*	0.813*	0.736*	0.954	1.116	1.284	0.935*	0.991	0.941
mMLP	0.816*	0.752*	0.878	1.024	1.192	1.553	0.982	1.014	1.104
tmean	0.761*	0.755*	0.778*	0.953	1.042	1.177	0.957	0.954	0.983
best	1.080	1.595*	1.243	1.195**	1.729*	2.661*	1.053	0.998	0.981
best5	1.083	1.596*	1.243	1.195**	1.725*	2.661*	1.053	0.996	0.980
MSEs90	0.944	0.689**	0.890	0.973	0.959	0.910	0.973	1.081	0.941
MSEm90	0.818*	0.678*	0.750*	1.158	1.075	0.950	0.913*	1.029	0.890
OLSr	1.098	0.796	1.339	1.042	1.113	0.973	1.008	1.075	1.187
MSEw	0.758*	0.732*	0.764*	0.956	1.030	1.103	0.962	0.960	0.957

Notes: Numerical entries in this table are relative mean square forecast errors (rMSFE) with respect to the AR(p) model (benchmark model). Forecast are monthly for the period 1991.01 to 2014.12. (P = 244). Numerical values less than unity indicate that the alternative model forecasts best. Where "modxi" is the label for model "mod" and "xi" for $i = 1, 2, 3$, indicates the vector x_t^i included in the multivariate version of the model. If no "xi" is attached, then the label refers to the univariate version of the model. The composite forecasts were made using the entire set of models (i.e., forty one models in total).
(*) Denoting rejection at the 5% level.
(**) Denoting rejection at the 10% level.

Results for IP are quite different from the ones reported in the previous section. For $h = 1$, evidence shows that several nonlinear models significantly outperform the benchmark model. In particular the MSARc(3)x3 model decreases the rMSFE by almost thirty two percent (32%). To illustrate the performance of the "best" model, Figures C.2.7 and C.2.8 of Appendix C show the performance of this model and it highlights the period for which was able to forecast best than the linear benchmark model. Similar results are reported for $h = 3$ for the MSARc(1)x3, MSARc(2)x3, MSARc(3)x3, MSARc(4)x3 and MSARc(5)x3 models with an average improvement of the order of thirty five percent (35%) in the rMSFE. At this forecast horizon, the MSARc(2)x3 model clearly outperforms the rest of the models with predictive gains of the order of forty five percent (45%). For $h = 12$, fewer models are able to forecast best than the benchmark model but by no means this imply a loss in the predictive accuracy of the MSAR type of models. In particular, the MSARc(2)x1, MSARc(2)x3, MSARc(3)x3 and MSARc(4)x3 models report a decrease of almost thirty five percent (35%) on average in the rMSFE. At this forecast horizon the MSARc(2)x3 model clearly dominates the rest of the models. Regarding the forecast combination schemes, the time-varying combination schemes are able to detect the superiority of some of these models at each forecast horizon but this is not enough to outperform the best model at each forecast horizon because they also incorporate the predictions from less accurate models. Thus, these results provide evidence to validate the conjecture made in the previous section, the univariate neural network was able to signal the potential gains of nonlinear models for this target variable; hence these type of models may be an effective first filter to detect these cases. We leave this last point open for further research.

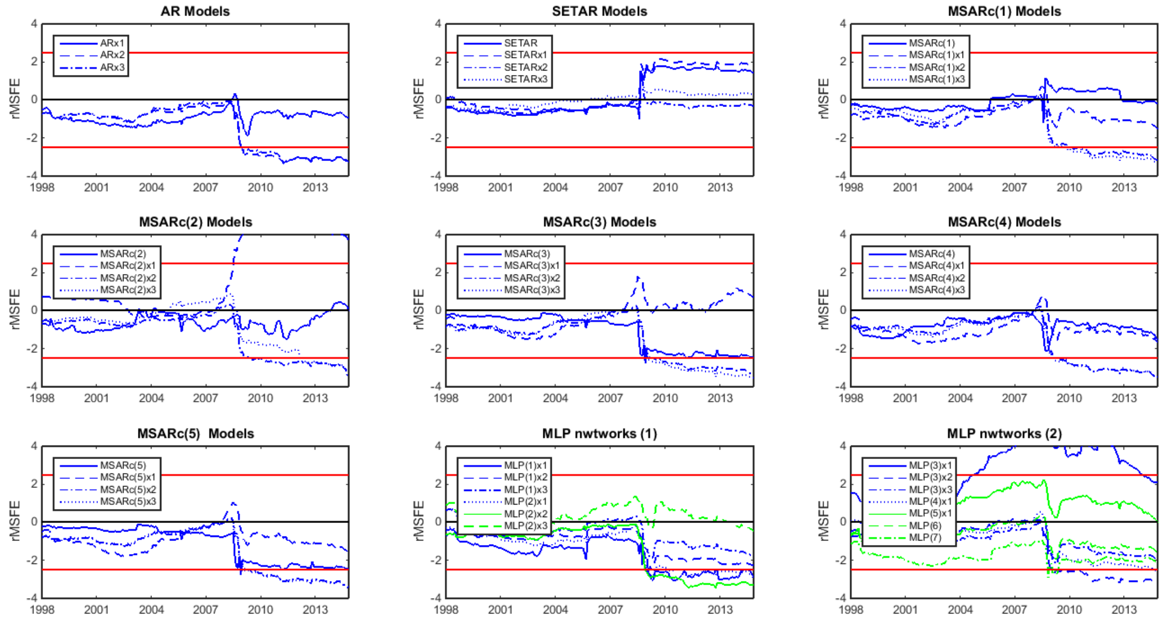
The results for CPI are consistent with previous findings since they show that a simple AR(p) model is able to forecast best than more sophisticated nonlinear models and combination schemes at any forecast horizon. Thus, predictions should be made using autoregressive models or alternative methods not considered in this paper.

The evidence for r10 is not as conclusive as it was for IP, but there are cases on which alternative models outperform the benchmark. In particular at $h = 1$, the MSARc(1)x2, MSARc(1)x3, MSARc(2)x3 and MSARc(3)x2 models report a reduction of eight to ten percent (8% - 10%) in the rMSFE; whereas at $h = 12$, two methods decrease the rMSFE in twelve percent (12%). On the contrary, no significant results are reported at $h = 3$. Finally, forecast combination schemes with fix and time-varying combination weights are unable to forecast best than the benchmark at any forecast horizon; the few cases for which the rMSFE is less than unity are not statistically significant with the exception of MSEm90 which reports a decrease of almost ten percent (10%) in the rMSFE. This mix evidence for this target variable may be due to the zero lower bound policy conducted by the Fed after the financial crisis. Most of the interest rates dropped to low (and constant) unprecedented values; thus in this setting nonlinear models would be unable to forecast this period effectively. To assess the robustness of these results, in the next section I analyze the average performance of the models before and after the financial crisis.

2.4.3 Unveiling the Local performance of the models

Figures 2.1 through 2.3 report the FL test for the 1-step-ahead forecast ($h = 1$) for the three target variables, results for the remaining forecast horizons are reported in Appendix 2.C. Each graph shows the sample path of the statistic ($F_{t,m}$) along with its critical values (cv_{lw}, cv_{up}) for a test level of 10%. If $F_{t,m}$ crosses one of the critical values then it signals that one of the model was outperformed by the other one at some point. In particular, if $F_{t,m} < cv_{lw}$ the alternative model forecasts best, if $F_{t,m} > cv_{up}$ the benchmark model forecasts best and if $F_{t,m} \in [cv_{lw}, cv_{up}]$ then both models have equal out-of-sample performance at each point in time. For ease of exposition, the models are grouped according to their functional form.

Figure 2.1
Fluctuation test for the industrial production index (IP)
1-step ahead forecast

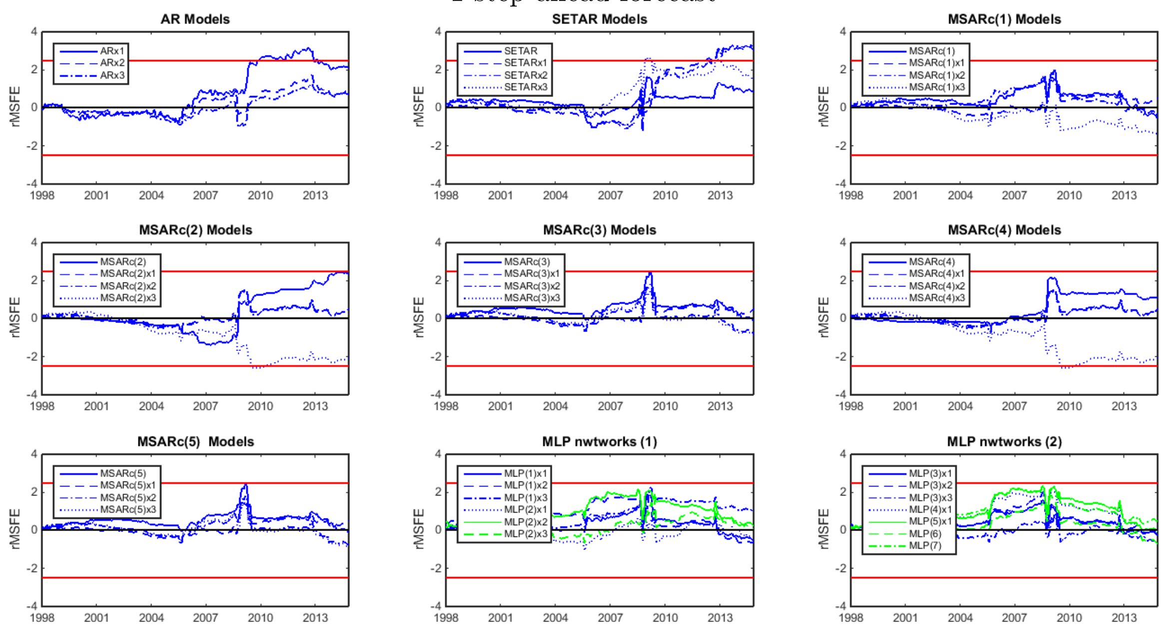


Forecasting IP over time. Each point of the solid blue line reports the relative performance of the alternative model with respect to the benchmark model computed with a rolling windows of five years of data. Red lines show critical values (cv_{lw}, cv_{up}) for a test level of 10%. If the blue line crosses cv_{lw} then the alternative model forecasts best, if it crosses cv_{up} then the opposite holds; otherwise no statistical difference exists between the models.

Figure 2.1 shows that the predictive accuracy between models is quite similar until 2008 and after this point several alternative models are able to forecast best than the benchmark model. In addition, results show a systematic decrease in the performance of threshold models after 2008. Regarding the MSAR type of models, the local evaluation shows that many of these specifications models are able to forecast best than the benchmark model at the end of the evaluation window. But there is a small subset of them on which no significant difference exists in the performance of the competing models; and two cases on which the benchmark model forecasts best. Similar results are reported for neural networks. Thus, the performance for these two type of models (MSAR and MLP) are sensitive to their specification; moreover these results further validate the argument that an exploration of the functional form of nonlinear models may lead to significant predictive gains. The evaluation for $h = 12$

report similar results but for this case the benchmark model is able to outperform the vast majority of neural networks. The FL tests for $h = 3$ show similar results as well, but there is a decrease in the number of models that are able to forecast best than the benchmark model after 2008. It is interesting to note that the local performance of thresholds models increases with the forecast horizon.

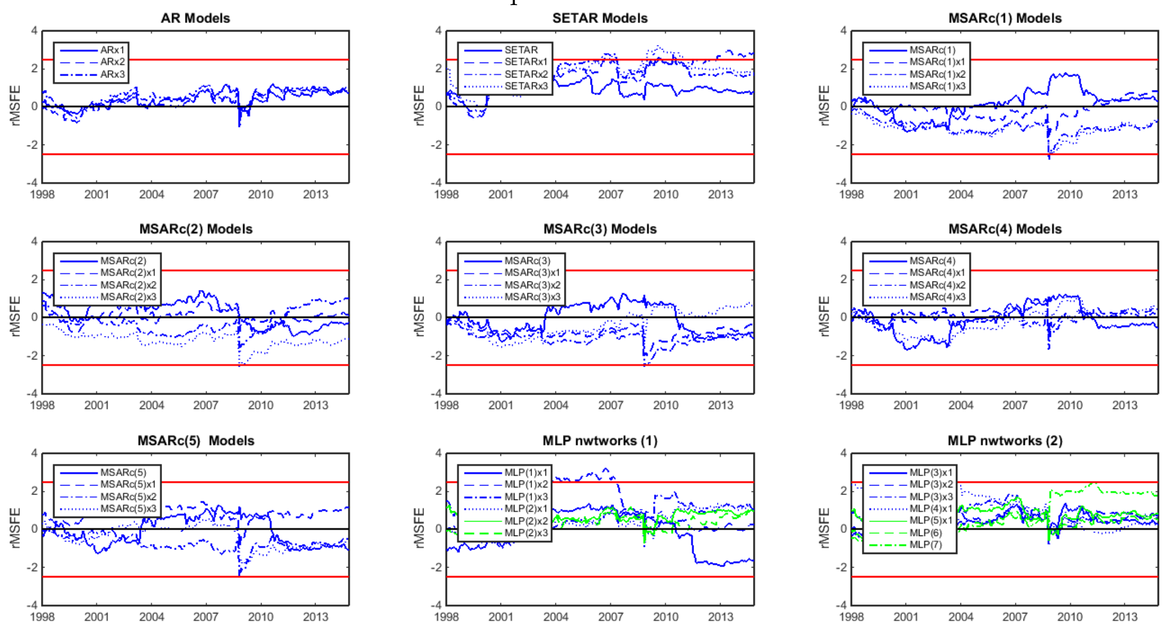
Figure 2.2
Fluctuation test for the consumer price index (CPI)
1-step ahead forecast



Forecasting CPI over time. Each point of the solid blue line reports the relative performance of the alternative model with respect to the benchmark model computed with a rolling windows of five years of data. Red lines show critical values (cv_{lw}, cv_{up}) for a test level of 10%. If the blue line crosses cv_{lw} then the alternative model forecasts best, if it crosses cv_{up} then the opposite holds; otherwise no statistical difference exists between the models.

For CPI, Figure 2.2 shows results that are consistent with the ones reported in the global evaluation. More precisely, alternative models are not able to predict the target variable significantly better than the benchmark model in most of the cases; in addition the evaluation also shows that the benchmark model forecasts best than some of the alternative models. However, the local evaluation shows that at longer forecast horizons the multivariate versions of the threshold model report promising results and further research in this line may reveal significant predictive gains for these type of models.

Figure 2.3
Fluctuation test for the 10 year Treasury-bond yield (r10)
1-step ahead forecast



Forecasting r10 over time. Each point of the solid blue line reports the relative performance of the alternative model with respect to the benchmark model computed with a rolling windows of five years of data. Red lines show critical values (cv_{lw}, cv_{up}) for a test level of 10%. If the blue line crosses cv_{lw} then the alternative model forecasts best, if it crosses cv_{up} then the opposite holds; otherwise no statistical difference exists between the models.

Finally, Figure 2.3 shows the results of the fluctuation test for r10. The FL tests indicate that in most cases no statistical difference exists between the models with the exception of few cases on which the benchmark model forecasts best. Finally, the local evaluation for the remaining forecast horizons show similar results but they also indicate that the local predictive ability of the alternative models decreases with the forecast horizon. However, it is worth noting that the fluctuation tests for this target variable display an erratic path through time and this pattern remain in all the forecast horizons.

2.4.4 Forecasting before and after the financial crisis

This section analyze the robustness of the global average predictive ability tests reported at the beginning of this section to the events of 2008, since several economic variables shown a unique dynamic after the financial crisis. Thus, I divide the evaluation sample in two periods: before and after December 2007 (2007.12). Next, prediction errors are computed for each model at each forecast horizon for each sample. As before, I compute the rMSFE along with the GW test; however this time the comparison excludes the time-varying combination schemes because the fewer number of observations of the second sample (after 2007) make the estimation of the weights for these schemes too noisy. The results are available in Table B.2.3 and B.2.4 of Appendix 2.B.

The results for IP quite similar with those reported using the full evaluation window. The main difference is that some of the GW tests are no longer statistically significant, but this is explained by the loss of efficiency due to the smaller number of observations on each subsample. As with the full sample evaluation, the accuracy of the alternative models decreases with the forecast horizon but at the same time the greatest predictive gains of these models are reported at these horizons.

For CPI, the results are also consistent with the ones reported using the full evaluation window. On each subsample, alternative models and composite forecasts are not able to outperform the benchmark model and in several cases the benchmark model forecasts best. The multivariate versions of the threshold model also show promising results at longer forecast horizons but only for the second subsample (after the financial crisis). In particular, they show a decrease in the rMSFE of almost eighteen percent (18%), but these results are not statistically significant. However, as is was mentioned before, these results may be explained by the loss of efficiency due to the smaller number of observations of this subsample and thus new data may reveal potential significant improvements from these type of models.

Finally, the results for r10 change with those reported using the full evaluation windows depending on the subsample used for the comparison. In the first subsample, there is a small subset of alternative models that is able to forecast best than the benchmark model. In addition, the predictive gains of restricted nonlinear models increase with the forecast horizons. In particular, the model that forecasts best at $h = 12$ decreases the rMSFE by almost twenty percent (20%) whereas at $h = 1$ only by eight percent (8%). These results change considerably in the second subsample (after 2008) since the benchmark model outperforms alternative models and combination schemes in many cases.

2.5 Conclusions

The empirical evidence of this paper yields interesting results about the predictive performance of nonlinear models for the conditional mean because they are able to challenge two of the key results from this literature: (1) nonlinear models are not able to outperform simple linear models and even if they do, the forecast gains are small or not statistically significant; and (2) forecast combination schemes usually outperform single (non)linear models.

First, the forecast comparison (considering only univariate restricted nonlinear models) showed results in line with previous findings in the sense that the benchmark model or the combination schemes were able to forecast best than the competing models in most cases. Second, the global evaluation using the full set of models and combination schemes showed that the predictive performance of the models changed dramatically in some cases. This is especially true for the industrial production index since several alternative models were able to significantly decrease the mean square error with respect to the benchmark model, in some cases by almost fifty percent depending on the forecast horizon. For the interest rate, the predictive gains were not as high as in the previous case and only few models were able to beat the benchmark model. Finally, no significant gains were reported for the consumer price index at any forecast horizon, however threshold models showed promising results. Third, the local evaluation showed results consistent with the global evaluation. For the industrial production index, the fluctuation tests showed an overall improvement with respect to the benchmark model after 2008. For the consumer price index and interest rate, the test showed that the benchmark model forecasted better than alternative models. Fourth, the evaluation before and after the financial crisis showed that the results from the global evaluation for the industrial production index and consumer price index are robust to this unique event. But, this was not the case for the interest rate since they showed that a small subset of alternative models were able to forecast best than the benchmark model before the financial crisis, but after the crisis the benchmark model outperformed most of the alternative models and forecast combination schemes at any forecast horizon.

In conclusions, this paper supports the view that the out-of-sample forecast performance of nonlinear models has being underestimated in previous forecast comparisons due to a deficient specification of their functional that increases parameter estimation error which in turn decreases their predictive ability. The predictive gains from these type models can be elusive, but an exploration of their functional may reveal a dynamic model that vastly outperforms competing models and forecast combination schemes. A natural extension of this work is to include data reduction methods, as in Stock and Watson (2002a,2002b), Bai and Ng (2006a) and Kim and Swanson (2014), to analyze the predictive performance of (un)restricted nonlinear models in this setting.

Appendix 2.A: Nonlinear models

Define y_t as the endogenous or target variable and object $w = (z_t', x_t')'$ as a vector of exogenous variables, where $z_t = (1, y_{t-1}, ..., y_{t-p})'$ and $x_t \in R^q$ as a vector of strongly exogenous variables. Depending on the vector x_t, the univariate or multivariate version of each model can be defined. Four type of such vectors are considered: $x_t^0 = \emptyset$, $x_t^1 \in R^3$, $x_t^2 \in R^6$ and $x_t^3 \in R^9$. Note that x_t^0 defines the univariate version of the model. Finally, define ε_t as a pure stochastic process. A dynamic model with an additive noise component can be defined as follows:

$$y_t = g\left(w(z_t, x_t); \theta\right) + \varepsilon_t \tag{6}$$

Where $g(\cdot)$ is a generic function that maps exogenous variables and parameters to the output variable y_t. I consider three type of models: The Self-Exciting Threshold Autoregressive Model (SETAR), The Markov Switching Autoregressive Model (MSAR) and The Multilayer Perceptron Network (MLP).

2.A.1. Self-Exciting Threshold Autoregressive Model (SETAR)

The model was first introduced by Tong (1978) and developed with more depth in Tong and Lim (1980), Tong (1983) and Tong (1990). Following the notation of Tong (1983) the univariate m-regime SETAR model is denoted as $SETAR(m, p_1, p_2, ..., p_m)$ where p_j denotes the lag order of regime j; and it can be written as:

$$y_t = \begin{cases} \alpha_1 + \sum_{k=1}^{p_1} \beta_{1,k} y_{t-k} + \varepsilon_{1t} & If \;\; q_{t-1} \leq \gamma_1 \\ \alpha_2 + \sum_{k=1}^{p_2} \beta_{2,k} y_{t-k} + \varepsilon_{2t} & If \;\; \gamma_1 < q_{t-1} \leq \gamma_2 \\ \vdots & \\ \alpha_m + \sum_{k=1}^{p_m} \beta_{3,k} y_{t-k} + \varepsilon_{mt} & If \;\; q_{t-1} > \gamma_m \end{cases}$$

Where q_{t-1} is a function of the data and $\varepsilon_{it} \sim N(0, \sigma^2)$. Note that q_{t-1} identifies the regime and therefore it is observable at time t. Define function $q_{t-1} = y_{t-d}$; following Hansen (2000) the $SETAR(m, p)$ model is defined as follows:

$$y_t = \beta_1' w(z_t, x_t) I_{1t}(\gamma, d) + ... + \beta_m' w(z_t, x_t) I_{mt}(\gamma, d) + \varepsilon_t \tag{7}$$

Define object $I_{jt}(\gamma, d) = I(\gamma_{j-i} < y_{t-d} \leq \gamma_j)$ as the indicator function that takes the unit value if the internal condition holds and zero otherwise, with d as the delay parameter ($d < t$) and $\gamma = (\gamma_1, ..., \gamma_m)$ with $\gamma_1 < \gamma_2 < ... < \gamma_m$ as the vector of threshold parameters. Note that if m takes the unit value then the SETAR(1,p) becomes the AR(p) model. Define $\theta = (\beta_1, ..., \beta_m, \gamma, d)$ as the vector of parameters; the estimation can be implemented by minimizing the sum of squares of the residuals (SSR):

$$\hat{\theta} = argmin \sum_{t=1}^{T} (y_t - \beta_1' w(z_t, x_t) I_{1t}(\gamma, d) - ... - \beta_m' w(z_t, x_t) I_{mt}(\gamma, d))^2 \tag{8}$$

The model is restricted in two ways: (1) the number of regimes is set to two ($m = 2$); and (2) each regime has the same $w()$. This setting defines the 2-regime SETAR model:

$$y_t = \begin{cases} \beta_1' w(z_t, x_t) + \varepsilon_{1t} & If \;\; y_{t-d} \leq \gamma \\ \beta_2' w(z_t, x_t) + \varepsilon_{2t} & If \;\; y_{t-d} > \gamma \end{cases}$$

The model has $m(p + k) + (m - 1) + 1$ parameters; where $m(p + k)$ accounts for the number of parameters for each regime, $(m - 1)$ accounts for the number of thresholds parameters and the last term accounts for the delay parameter. Table A.2.1 summarizes the estimation algorithm for this model:

The variable $Z(y_t)$ contains the ordered values from lowest to highest of y_t. Ergo $Z(y_t)$ contains all the relevant values that the threshold may take. To speed up the algorithm a kernel distribution of y_t may be used instead. Note take the algorithm becomes computational more demanding with the sample size and the number of

Table A.2.1
Pseudo estimation code for SETAR model

1. Let $Z(y_t) = sort(y_t)$
2. Define the delay parameter d_j.
3. Define γ_j^i as the $\alpha\%T - 1 + i$ element of $Z(y_t)$.
4. Conditional on γ_j^i and d_j; split the sample and compute the OLS estimators of β_j^i.
5. Define the estimator candidate as $\hat{\theta}_j^i = (\beta_j^i, \gamma_j^i, d_j)$ and construct $SSR(\hat{\theta}_j^i)$.
6. Repeat (2) - (5), for $i = 1, 2, ..., T - 2\alpha\%T$
7. Repeat (2) - (6), for $j = 1, 2, .., w$
8. Choose $\hat{\theta}$ such that minimizes the $SSR(\hat{\theta}_j^i)$.

Where T is the sample size, α is the minimum percentage of observations for each regime and w is the maximum value allowed for the search of the delay parameter ($d \leq p$).

regimes. According to Tong (1990) the information criteria for a 2-regime SETAR model is the sum of the information criteria of each regime.

2.A.2. Markov Switching Autoregressive Model (MSAR)

The model was first introduced in Hamilton (1989) and in its original setting allows for the parameters of an autoregressive specification to change between states or regimes according to a hidden markov process. This regime switching process resembles to the SETAR model, however in this case the states are unobservable and therefore the model uses the data to make an inference about the state of the world at time t.

The m-state Markov Switching Autoregressive Model of order p id defined as follows:

$$y_t = \beta'_{s_t} w(z_t, x_t) + \varepsilon_{t,s_t} \tag{9}$$

Where s_t is the state variable that evolves following an independent Markov Chain and ε_{t,s_t} is distributed as $N(0, \sigma^2_{s_t})$. Following Hamilton (1994), define the conditional likelihood as:

$$f(y_t | s_t = j, w_t; \theta)$$

Where θ is the vector of parameters that characterized the conditional density. Define η_t and P (the transition matrix between states) as:

$$\eta_t = \begin{pmatrix} f(y_t|s_t = 1, X_t; \theta) \\ \vdots \\ f(y_t|s_t = m, X_t; \theta) \end{pmatrix} \quad P = \begin{pmatrix} p_{11} & ... & p_{m1} \\ \vdots & \vdots & \vdots \\ p_{mm} & ... & p_{mm} \end{pmatrix}$$

The element $p_{ij} = Pr\{s_t = j|s_{t-1} = i\}$ defines the transition probability from state i at time $t-1$ to state j at time t. Define $\hat{\xi}_{t|t} = Pr\{s_t = j|Y_t; \theta\}$ for $j = 1, 2, ..., m$; as the conditional probability that the tth observation was generated by regime j. Therefore:

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \bigodot \eta_t}{1'(\hat{\xi}_{t|t-1} \bigodot \eta_t)} \quad \text{and} \quad \hat{\xi}_{t+1|t} = P\hat{\xi}_{t|t}$$

Given an initial value for $\hat{\xi}_{1|0}$ and θ, it is possible to iterate over $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$, to evaluate the log-likelihood function of the model as follows:

$$\ell(\theta) = \sum_{t=1}^{T} log f(y_t|X_t; \theta) \tag{10}$$

Where:

$$f(y_t|X_t; \theta) = 1'(\hat{\xi}_{t|t-1} \bigodot \eta_t) \tag{11}$$

Only two-state models are considered ($m = 2$) since the identification of a higher number of regimens could be difficult in this setting (not enough data to properly identify three or more regimes). Nevertheless, the two-regime model provides a flexible approximation to the DGP process that may lead to significant predictive gains.

The log likelihood function $\ell(\theta)$ is maximized using numerical methods. The initial values for θ are estimated using a SETAR(2,p); for the transition matrix: $p_{ij} = 0.5$ for all i and j. Finally, for: $\hat{\xi}_{1|0}$ the unconditional probability of P is used. Note that $\hat{\xi}_{1|0}$ will change on each iteration of the algorithm.

2.A.3. Neural Networks: The Multilayer Perceptron Network (MLP)

According to Cybenko (1989) neural networks are highly nonlinear models and they are able to approximate almost any function given enough structure; this is known as the Universal Approximation Theorem. However, the main drawback of these models is that the number of parameters increase exponentially when more structure is added to the network[16].

The Feedforward Neural Network with i^* inputs, v^* hidden units (or neurons) and one layer is defined according to McNelis (2005) as:

$$y_t = \beta_0 + \sum_{i=1}^{p+q} \beta_i w_i + \sum_{v=1}^{v^*} \gamma_v N_{v,t}(w(z_t, x_t)) + \varepsilon_t \tag{12}$$

$$N_{v,t}(w(z_t, x_t)) = \frac{1}{1 + e^{-n_{v,t}}}$$

$$n_{v,t} = \omega_{v,0} + \sum_{i=1}^{p+q} \omega_{v,i} w_i$$

Where w_i is the ith component of the vector $w(x_t, z_t) \in R^{p+q}$, $\beta_0 \in R^1$ and each β_i and γ_v are in R^1. In addition, $N_{k,t}$ is the logistic squashing (or activating) function. The Feedforward Neural Network with the logistic activation function defines the Multilayer Perceptron Network (MLP). Note that w_i has a direct and linear impact on the output variable y_t through β_i. The linear function $n_{k,t}$ combines the inputs variables and produces a signal whereas the squashing function transforms this signal into a hidden unit or neuron ($N_{k,t}$). Finally the hidden units are attenuated or amplified by the weights $\{\gamma_k\}_{k=1}^{v^*}$ and summed.

The estimation is done by minimizing the SSR of the model. In particular define Ω as the vector of parameters of the network, then:

$$\hat{\Omega} = argmin \sum_{t=1}^{T} \left(y_t - \gamma_0 - \sum_{i=1}^{p+q} \beta_i w_i - \sum_{k=1}^{v^*} \gamma_k N_{k,t} \right)^2 \tag{13}$$

The direct numerical optimization of (12) can perform poorly due to the dimensionality of the problem. Following McNeil (2005), the parameters are estimated by

[16] This is known as the curse of dimensionality

minimizing the SSR in a two step procedure; in the first stage the function is optimize using a genetic algorithm, whereas in the second stage the function is optimize using a Quasi-Newton algorithm that uses the output from the previous stage as the starting values for the algorithm. A detailed description of the Genetic Algorithm can be found in McNeil (2005), chapter 3 page 72.

Appendix 2.B: Tables

Table B.2.1
Data description (1960.01 - 2014.12)

Variable	Description	Transformation
Industrial production index (IP)	Industrial production total (sa)	$ln\Delta_t(y_{t+h})$
10 year Treasury-bond yield (r10)	Long term interest rate	$\Delta_t(y_{t+h})$
Consumer price index (CPI)	CPI all urban consumers (nsa)	$ln\Delta_t(y_{t+h})/h-$ $ln\Delta_{t-1}(y_t)$
Exogenous variable for each target variable		
IP		
-Federal Funds Rate	Effective Federal Funds Rate	level
-Leading Economic Index	The Conference Board Leading Economic Index	level
-Interest rate spread	With respect to the 10 year treasury rate	level
CPI		
-Federal Funds Rate	Effective Federal Funds Rate	level
-Deliveries	Index of supplier deliveries	level
-Unemployment rate	Unemp. rate: all workers, 16 years and over	level
r10		
-Leading Economic Index	The Conference Board Leading Economic Index	level
-3-Month Treasury Bill	Short term interest rate	level
-Consumer price index	CPI all urban consumers (nsa)	$ln\Delta_t(y_t)$

Source: Federal Reserve Board of Governors and The Conference Board.

Table B.2.2
Relative mean square forecast errors and GW test
Alternative CPI transformation: $y_{t+h}^{cpi} = 1200ln\left(CPI_{t+h}/CPI_t\right)/h$

Method	h = 1	h = 3	h = 12	*Method*	h = 1	h = 3	h = 12
ARx1	0.971	0.940	0.841	MSARc(4)x1	0.919	0.914	1.122
ARx2	0.969	0.941	0.825	MSARc(4)x2	0.895	0.924	0.969
ARx3	0.962	0.944	0.822	MSARc(4)x3	0.907	0.907	1.120
SETAR	0.953	1.213**	1.674*	MSARc(5)	1.032	1.016	1.250*
SETARx1	1.088	1.196*	1.563*	MSARc(5)x1	0.956	0.912	1.170
SETARx2	0.956	1.226**	1.563*	MSARc(5)x2	0.915	0.924	1.052
SETARx3	0.961	1.236	1.745*	MSARc(5)x3	0.934	0.906	1.364**
MSARc(1)	0.990	1.189	0.979	MLP(1)x1	1.035	1.003	1.251*
MSARc(1)x1	0.953	0.898	1.118	MLP(1)x2	0.995	1.107	1.013
MSARc(1)x2	0.903	0.872	1.038	MLP(1)x3	0.950	0.942	1.098
MSARc(1)x3	0.927	0.880	1.105	MLP(2)x1	0.965	0.977	1.346*
MSARc(2)	1.016	1.123	1.266*	MLP(2)x2	1.027	0.949	1.229**
MSARc(2)x1	0.956	0.934	1.166	MLP(2)x3	0.944	0.918	1.151
MSARc(2)x2	0.915	0.932	1.044	MLP(3)x1	0.931	0.900	0.978
MSARc(2)x3	0.907	0.921	1.363**	MLP(3)x2	1.007	0.979	1.385*
MSARc(3)	1.032	1.016	1.250*	MLP(3)x3	0.993	1.062	1.020
MSARc(3)x1	0.956	0.934	1.402*	MLP(4)x1	1.028	0.940	1.088
MSARc(3)x2	0.902	0.934	1.403*	MLP(5)x1	0.982	0.887	1.122
MSARc(3)x3	0.927	0.921	1.363**	MLP(6)	1.121**	0.932	1.116
MSARc(4)	0.973	0.976	0.840*	MLP(7)	0.966	1.006	1.319
Forecast combination schemes							
mean	0.906	0.887	0.962	best	1.674*	1.709*	3.242*
med	0.912	0.885	0.946	best5	1.679*	1.708*	3.247*
mAR	0.967**	0.936**	0.822*	MSEs90	1.093	0.920**	0.809*
mST	0.925	1.111**	1.526*	MSEm90	0.994	0.920	1.067
mMS	0.908	0.903	1.079	OLSr	9.000	9.000	9.000
mMLP	0.953	0.919	0.984	MSEw	0.905**	0.881	0.907
tmean	0.905	0.883	0.943	Mnr	1.760*	1.539*	1.217

Notes: Numerical entries in this table are relative mean square forecast errors (rMSFE) with respect to the benchmark model. Forecast are monthly, for the period 1991.01 to 2007.12 (P = 204). Numerical values less than unity indicate that the alternative model forecasts best. Where "modxi" is the label for model "mod" and "xi" for $i = 1, 2, 3$, indicates the vector x_t^i included in the multivariate version of the model. If no "xi" is attached, then the label refers to the univariate version of the model. The composite forecasts were made using the entire set of models (i.e., forty one models in total).
(*) Denoting rejection at the 5% level.
(**) Denoting rejection at the 10% level.

Table B.2.3

Relative mean square forecast errors (rMSFE) and GW test, before 2008

	IP			CPI			r10		
Method	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12
ARx1	0.897*	1.478**	1.452**	1.003	0.949	1.032	1.030	1.103	1.138**
ARx2	0.858*	0.912	1.091	0.999	0.931	1.013	1.006	0.996	1.213
ARx3	0.872*	0.851	1.037	0.984	0.902**	0.991	1.059	1.008	1.056
SETAR	0.904*	1.553*	1.875*	1.033	1.613*	1.960*	1.047*	1.595*	2.209*
SETARx1	0.928	1.345*	1.483*	0.929	1.122	1.231**	1.127*	1.122**	1.428*
SETARx2	0.806*	0.891	1.273	0.949	1.151	1.213	1.284*	1.154*	1.670*
SETARx3	0.892	0.887	1.142	1.084	1.098	1.258**	1.353*	1.161	1.609*
MSARc(1)	0.971	0.980	0.957	1.110**	1.440	1.035	1.003	1.140*	1.322*
MSARc(1)x1	0.936	0.998	0.975	0.990	1.616	0.989	0.988	1.106	0.925
MSARc(1)x2	0.859**	0.850	0.777	1.176**	1.610	1.468	0.920*	1.072	0.857*
MSARc(1)x3	0.885	0.737**	0.726	0.999	1.591	0.929	0.904*	1.015	0.799*
MSARc(2)	0.987	0.990	1.016	0.924**	1.409	1.428	1.030	1.173*	1.434*
MSARc(2)x1	1.276	0.988	1.013	0.990	1.550	1.366	1.036	2.051*	0.901
MSARc(2)x2	0.926	0.871	0.829	0.987	1.538	1.081	0.987	1.131*	0.795*
MSARc(2)x3	1.000	0.676*	0.616*	0.948	1.516	1.045	0.906*	1.138	0.775*
MSARc(3)	0.946*	0.945**	0.989	1.107**	1.448*	0.984	1.008	1.142	1.017
MSARc(3)x1	0.948	1.048	1.013	0.987	1.477**	0.984	0.955*	1.304*	0.827*
MSARc(3)x2	0.847*	0.863	0.774	1.054	1.465**	1.458	0.912*	1.114**	0.808*
MSARc(3)x3	0.835**	0.795**	0.680**	1.063	1.590	0.862	0.953	1.019	0.788*
MSARc(4)	0.976**	0.964	1.042	1.002	1.419	0.891*	0.997	1.024	1.129
MSARc(4)x1	0.915	0.980	1.114	0.990	1.550	0.954	1.016	1.104**	0.907**
MSARc(4)x2	0.857*	0.868	0.746	0.987	1.515	1.448	1.004	0.876*	0.857*
MSARc(4)x3	0.840*	0.726*	0.611*	0.948	1.517	0.827	0.978	0.935	0.803*
MSARc(5)	0.946*	0.945**	0.989	1.107**	1.448*	0.984	1.008	1.142	1.017
MSARc(5)x1	0.939	1.017	1.076	0.990	1.627	0.954	1.081**	1.092	0.863*
MSARc(5)x2	0.847*	0.853	0.774	1.054	1.607	1.447	0.937**	0.908	0.857*
MSARc(5)x3	0.850*	0.693*	0.715**	1.063	1.590	0.932	1.007	1.000	0.788*
MLP(1)x1	0.932*	1.005	0.983	1.092	1.734*	2.715*	1.000	1.113**	1.132*
MLP(1)x2	0.910*	0.993	0.978	1.340*	1.439*	1.140*	1.077*	1.632	1.149*
MLP(1)x3	1.028	1.089	1.468**	1.162**	1.444*	2.856*	1.023	1.205*	1.137
MLP(2)x1	0.872*	0.850	1.138	1.019	1.553*	4.371*	1.012	1.011	1.327**
MLP(2)x2	0.874*	1.316	2.445*	1.380*	1.238	1.304**	1.076	1.048	1.151
MLP(2)x3	1.135	1.221*	1.433	1.036	1.368*	1.069	1.010	1.119**	1.145*
MLP(3)x1	2.265*	0.952	1.273	1.174*	1.695*	1.548*	1.009	1.217	1.155
MLP(3)x2	0.887**	0.911	1.413	1.174*	2.409*	1.224	1.069	1.006	1.010
MLP(3)x3	0.893*	1.140	1.386	1.029	2.345*	3.078*	1.058	1.253*	1.170**
MLP(4)x1	0.965	1.083	1.289	1.375*	1.524*	1.642*	1.217*	1.042	1.183
MLP(5)x1	1.152	1.173	1.663	1.421*	1.540*	1.688*	1.103	0.993	1.154
MLP(6)	0.921*	1.038	1.480**	1.113**	2.596*	4.305*	1.012	1.074	1.145*
MLP(7)	0.812*	0.955	1.143	1.422*	1.531*	1.267*	0.999	1.062	1.117*
Forecast combination schemes									
mean	0.810*	0.753*	0.735*	1.008	1.271	0.999	0.966	0.950**	0.893**
mAR	0.842*	0.866	0.922	0.990	0.920*	0.870	0.982	0.961	0.977
mST	0.806*	0.946	1.221	0.942	1.129**	1.146	1.110	1.038	1.435*
mMS	0.801*	0.737*	0.674*	0.983	1.412	0.915	0.938*	0.990	0.872*
mMLP	0.879*	0.802*	1.005	1.149*	1.532*	1.583*	0.985	0.992	0.973

Notes: Numerical entries in this table are relative mean square forecast errors (rMSFE) with respect to the benchmark model. Forecast are monthly, for the period 1991.01 to 2007.12 (P = 204). Numerical values less than unity indicate that the alternative model forecasts best. Where "modxi" is the label for model "mod" and "xi" for $i = 1, 2, 3$, indicates the vector x_t^i included in the multivariate version of the model. If no "xi" is attached, then the label refers to the univariate version of the model. The composite forecasts were made using the entire set of models (i.e., forty one models in total).

(*) Denoting rejection at the 5% level.

(**) Denoting rejection at the 10% level.

Table B.2.4
Relative mean square forecast errors (rMSFE) and GW test, after 2008

	IP			CPI			r10		
Method	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12
ARx1	0.944	1.020	0.983	1.027	1.100	0.770	1.041	1.202*	1.646*
ARx2	0.668	0.600	0.765	1.019	1.117	0.769	1.076	1.460	2.133*
ARx3	0.672	0.594	0.770	1.027	1.184	0.802	1.100	1.273	2.252*
SETAR	1.158	1.179	1.152	1.102	1.015	1.328	1.030	1.287*	2.300*
SETARx1	1.216	1.450*	1.543*	1.212**	1.261**	0.835	1.251*	1.372*	1.691*
SETARx2	0.953	0.799	1.007	1.164*	1.202**	0.816	1.308**	1.550*	2.129*
SETARx3	1.063	0.747	0.943	1.135	1.177	0.978	1.307	1.263	1.973*
MSARc(1)	0.981	1.123	1.215	0.952	1.350	1.445	1.020	1.056	1.099
MSARc(1)x1	0.886	1.462	0.870	1.058	1.210	1.441	1.060	1.321*	1.710**
MSARc(1)x2	0.635	0.703	0.849**	0.952	0.914	1.451	0.922	1.168	1.428**
MSARc(1)x3	0.581	0.567	0.711**	0.909	0.941	1.485	0.934	0.996	1.209
MSARc(2)	0.999	1.028	1.013	1.140	0.874	1.315	0.996	1.038	1.087
MSARc(2)x1	1.928	1.618	0.745	1.058	1.194	1.400	1.006	1.392**	1.384
MSARc(2)x2	0.547	0.768	0.837*	1.058	1.204	1.630	1.106	1.456	1.558*
MSARc(2)x3	0.590	0.448	0.608**	0.884	1.214	1.629	0.887	2.758*	1.265
MSARc(3)	0.876**	1.122	1.014	0.989	1.127	1.637	0.962	1.059	1.081
MSARc(3)x1	1.047	1.645	0.745	1.058	1.043	1.450	0.970	1.275**	1.265
MSARc(3)x2	0.588	0.753	0.837*	0.942	1.049	1.455	0.921	1.199	1.316**
MSARc(3)x3	0.525	0.674	0.688**	0.923	0.940	1.439	1.053	1.041	1.235
MSARc(4)	0.978	1.096	0.960*	1.072	1.306	1.581	0.981	0.993	1.122
MSARc(4)x1	0.893	1.656	0.825	1.056	1.194	1.450	1.014	1.171*	1.346*
MSARc(4)x2	0.602	0.773	0.767*	1.058	1.204	1.455	1.028	1.292	1.487
MSARc(4)x3	0.611	0.572	0.597**	0.884	1.214	1.383	1.042	0.918	1.442*
MSARc(5)	0.876**	1.122	1.014	0.989	1.127	1.637	0.962	1.059	1.081
MSARc(5)x1	0.893	1.606	0.821	1.058	1.202	1.450	1.099	1.200*	1.778*
MSARc(5)x2	0.585	0.753	0.837*	0.942	0.930	1.455	0.963	1.462	2.327*
MSARc(5)x3	0.575	0.520	0.772*	0.923	0.940	1.485	0.930	0.963	2.128*
MLP(1)x1	0.913**	0.918	0.992	0.946	1.052	1.760	0.974	1.023	1.302*
MLP(1)x2	0.816	0.955	0.989	0.903	0.937	1.735	1.015	1.071	1.324*
MLP(1)x3	0.853	1.543**	0.930	1.126	1.058	2.046*	1.099	1.285*	1.746*
MLP(2)x1	0.734	0.769	0.839	0.975	1.070	2.321	1.118	1.430	2.393*
MLP(2)x2	0.626	0.869	0.903	1.041	1.500	1.636	1.114	1.505	2.201*
MLP(2)x3	0.964	1.013	0.933**	1.002	1.320	1.552	1.037	1.245*	1.739*
MLP(3)x1	1.420	0.596	0.926	0.951	1.259	1.759	1.034	1.376	2.550*
MLP(3)x2	0.662	0.398	0.950	0.961	1.159	1.688	1.083	1.259	2.180*
MLP(3)x3	0.837	1.399	0.954	0.971	1.144	1.726	1.045	1.461**	1.886*
MLP(4)x1	0.688	0.726	1.118	0.995	1.384	1.725	1.032	1.411	2.494*
MLP(5)x1	0.991	0.761	0.615	0.999	1.079	1.285	1.106	1.340	2.497*
MLP(6)	0.892	1.278	1.071	0.945	1.281	2.076	1.028	1.208*	1.630*
MLP(7)	0.879	0.944	1.693*	1.079	1.197	1.822	1.072	1.072	1.267
Forecast combination schemes									
mean	0.715	0.769	0.789*	0.901	0.935	1.230	0.957	0.979	1.355**
mAR	0.761	0.687	0.810	1.014	1.087	0.784	0.983	1.036	1.483*
mST	0.960	0.923	1.057	1.060	1.094	0.893	1.121**	1.148**	1.806*
mMS	0.702	0.880	0.765*	0.925	0.987	1.441	0.928	0.994	1.213
mMLP	0.754	0.710	0.818*	0.899	1.045	1.540	0.976	1.063	1.626*

Notes: Numerical entries in this table are relative mean square forecast errors (rMSFE) with respect to the benchmark model. Forecast are monthly, for the period 2008.01 to 2014.12 (P = 84). Numerical values less than unity indicate that the alternative model forecasts best. Where "modxi" is the label for model "mod" and "xi" for $i = 1, 2, 3$, indicates the vector x_t^i included in the multivariate version of the model. If no "xi" is attached, then the label refers to the univariate version of the model. The composite forecasts were made using the entire set of models (i.e., forty one models in total).
(*) Denoting rejection at the 5% level.
(**) Denoting rejection at the 10% level.

Appendix 2.C: Figures

Figure C.2.1

Fluctuation test for the industrial production index (IP)

3-step ahead forecast

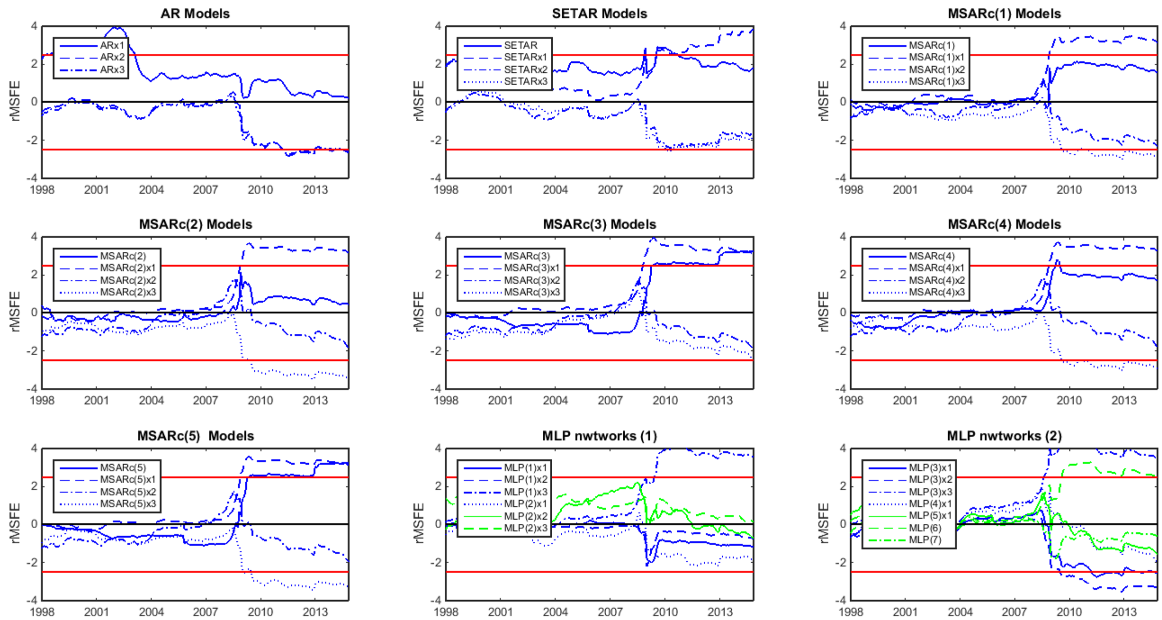
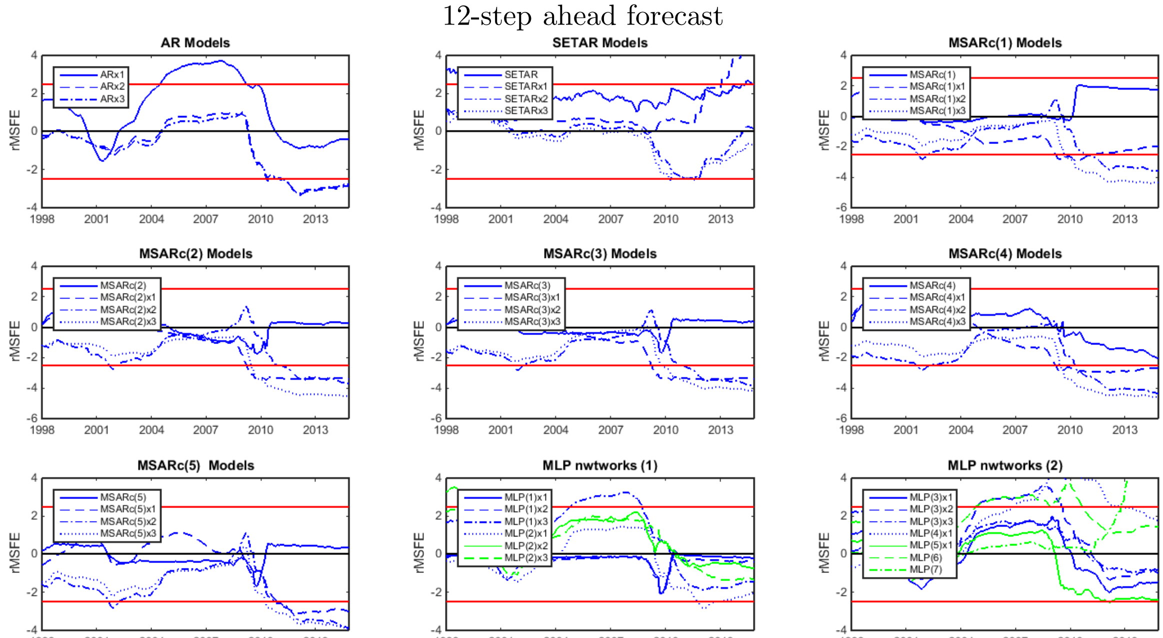


Figure C.2.2

Fluctuation test for the industrial production index (IP)

12-step ahead forecast



See notes in Figure 2.1 for more details.

Figure C.2.3
Fluctuation test for the consumer price index (CPI)
3-step ahead forecast

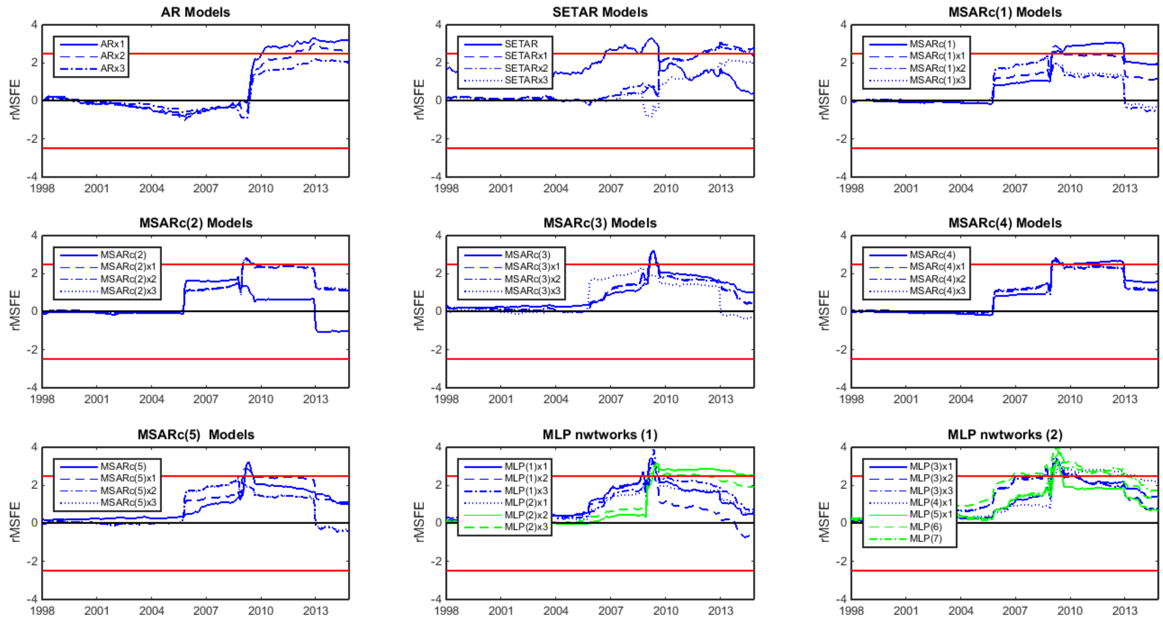
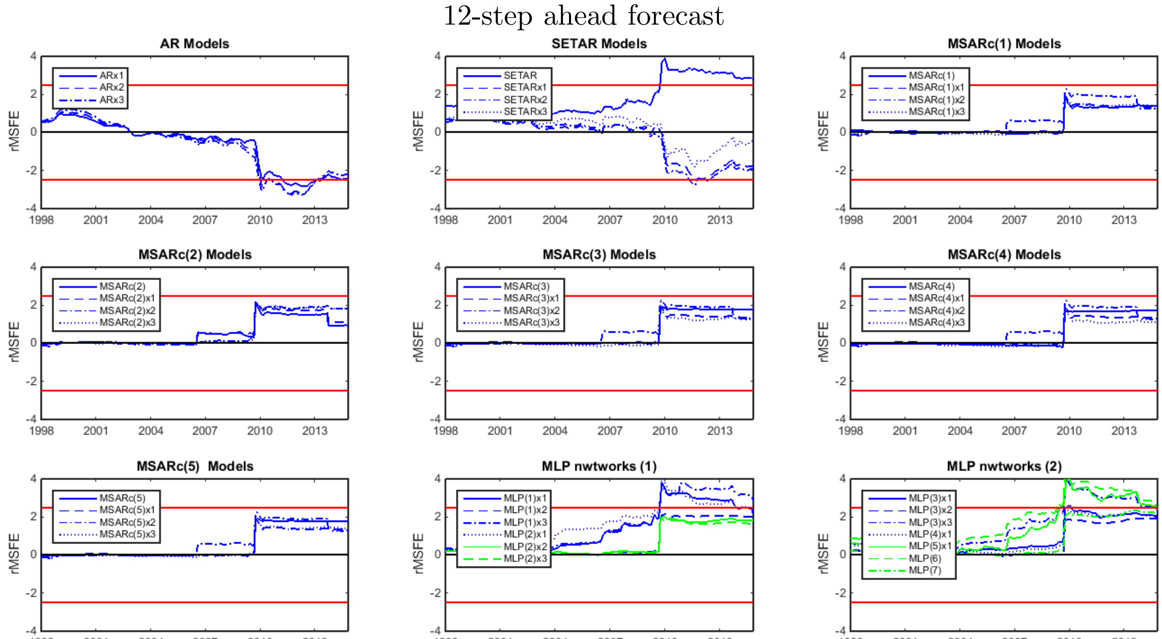


Figure C.2.4
Fluctuation test for the consumer price index (CPI)
12-step ahead forecast



See notes in Figure 2.1 for more details.

Figure C.2.5

Fluctuation test for the 10 year Treasury-bond yield (r10)

3-step ahead forecast

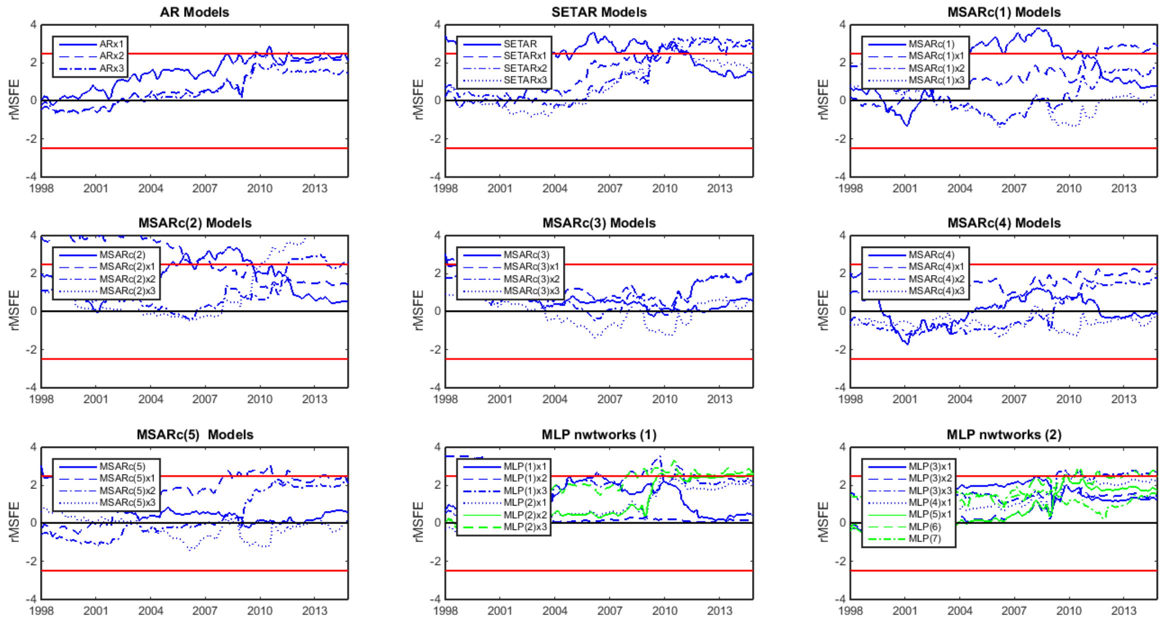
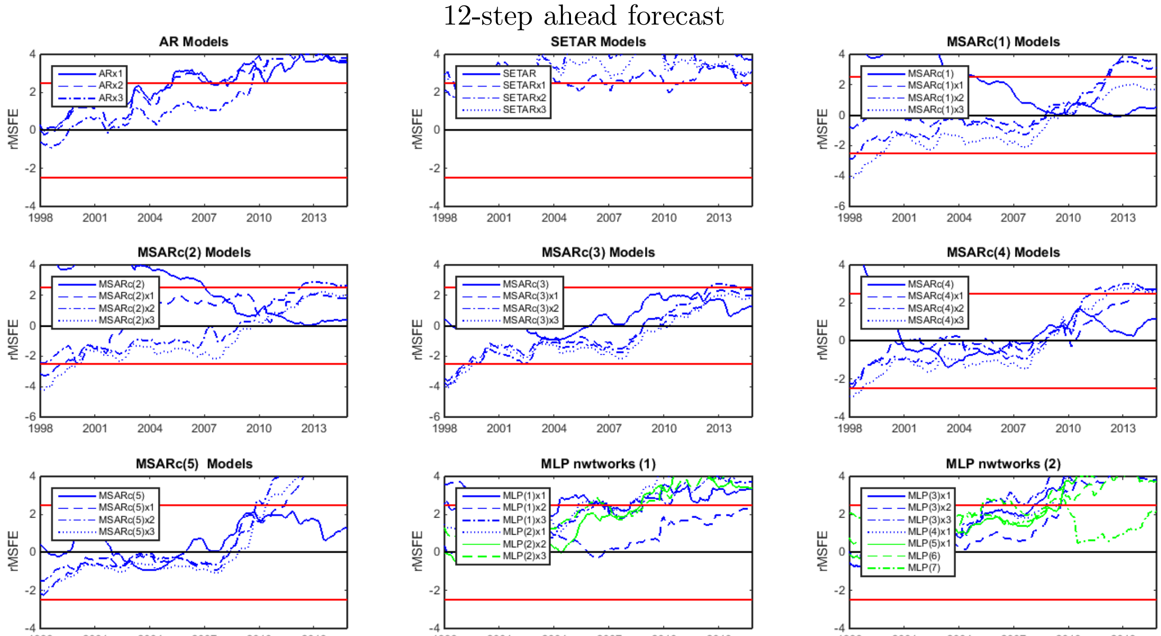


Figure C.2.6

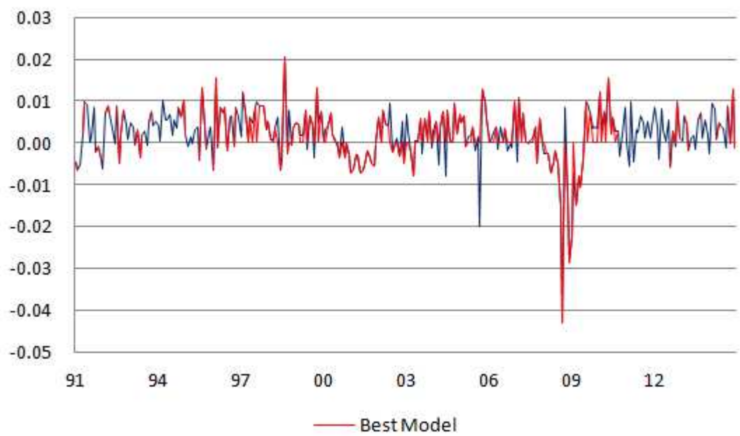
Fluctuation test for the 10 year Treasury-bond yield (r10)

12-step ahead forecast



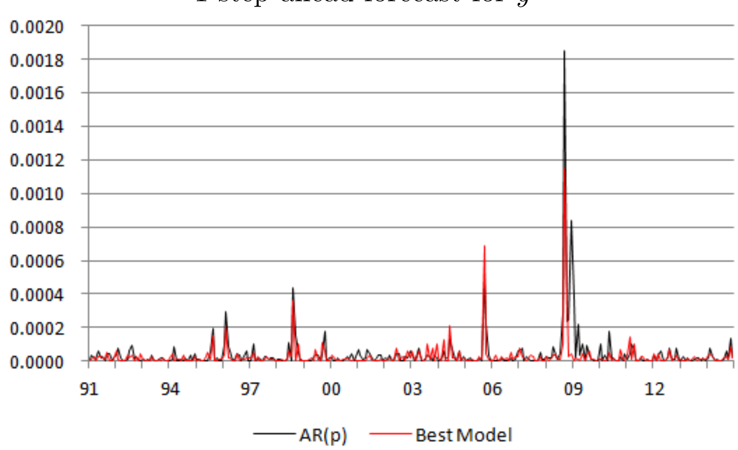
See notes in Figure 2.1 for more details.

Figure C.2.7
1-step ahead forecast for y^{IP}



Notes: Blue line is the evaluation sample for the 1-step ahead forecast for the percentage change of IP; whereas red segments indicate the periods on which the alternative "best" model was more accurate than the linear benchmark model.

Figure C.2.8
Square prediction errors, benchmark model vs "best" model
1-step ahead forecast for y^{IP}



Notes: Comparing the square of prediction error between the benchmark linear model and the alternative "best" model.

Chapter 3
Reassessing the effects of foreign monetary policy on output: new evidence from structural and agnostic identification procedures

3.1 Introduction

In December 2008, the federal funds rate dropped to the zero lower bound and since then unconventional monetary policies dominated the scene[1]. It took almost six years for the Fed to raise its policy rate and finally the zero lower bound was abandoned by the end of 2015. The ongoing period of monetary normalization combines two signals: (i) concrete policy measures and (ii) forward guidance. Currently, several central banks are evaluating the likely effects that the US monetary normalization may have in their economies in order to assess potential risks and improve policy decisions since the propagation of that shock activates different channels (interest rate spread, exchange rate depreciation, problems of excessive debt burden if debt is denominated in dollars, etc.) that affect their economies in different dimension. For example, private debt may have increased significantly due to lower interest rates and thus an increase in foreign rates can generate a domestic depreciation that amplifies the burden of foreign debt in domestic currency. Moreover, the current poor performance in many of these economies could further amplify the impact of the shock over debtors and the economy overall[2].

Thus, this paper investigates the propagation of a foreign monetary policy shock

[1]The Fed had strong reasons to intervene based on historical reasons; fears of a liquidity crisis that could lead the economy to another great depression.

[2]Consider another example to further motivate the discussion. The pass-through of exchange rate to inflation can trigger an increase in domestic interest rates to contain inflation. However, at the same time higher foreign rates can be associated with more adverse foreign conditions that can have a negative impact in output which in turn could help to mitigate the hike in inflation and the response of central bank. Thus, we spot an interesting policy implication from this analysis.

over a small open economy, in particular over the Chilean economy. We use a comprehensive methodological framework that compares the impulse response functions (henceforth IRFs) of three models: two Structural VAR models and a DSGE model tailored for the Chilean economy[3]. We follow this approach because according to Canova (2007), Structural VAR models can be used to judge and validate the responses from DSGE models. Therefore, this comparison provides new lights and insights about the propagation of a foreign monetary policy shock over the Chilean economy and in addition it assess the suitability of the microfounded structure behind the DSGE model (i.e., the theoretical model). To this end, we use the recursive VAR model of Sims (1980a) on which identification of structural shocks is based on a particular order of the variables in the system; along with an extension of the "agnostic" VAR model of Uhlig (2005) and Arias et al. (2014) for small open economies following Koop and Korobilis (2010). In this identification scheme, structural shocks are identified by imposing restrictions directly in the IRF.

Our findings can be summarized as follows. (1) Consistent with several studies such as Bernanke et al. (2005), Mojon (2008) and Castelnuovo (2015) our analysis of IRFs lead us to conclude that identification of foreign monetary shock is not straightforward in recursive VAR models. Therefore, the recursive VAR model fails to provide an informative benchmark to judge the plausibility of results from structural micro-founded models. (2) On the contrary, the "agnostic" VAR model provides IRFs with dynamics that are broadly consistent with macroeconomic theory, hence in our view results provide an informative benchmark to micro-founded models. (3) Beyond the quantitative differences, we find that the IRFs of the "agnostic" VAR model are in line qualitatively with those of the DSGE model except for output. The DSGE model shows an initial hike in activity which is explained by the improving of

[3]A standard Dynamic Stochastic General Equilibrium (DSGE) model for a small open economy with nominal and real rigidities that is closely related to models developed by Christiano et al. (2005) and Smets and Wouters (2003, 2007).

the current account due to the real and nominal exchange rate depreciation; whereas the "agnostic" VAR infers a significant drop in output. (4) The transmission of the shock to the domestic economy in the DSGE model is limited but persistent. At least two reasons may explain this. First, by construction, there are many micro-founded restrictions in the model that increase the persistency of the shock (habit formation in consumption, quadratic adjustment cost for investment, etc.). Second, there is an excessive simplification in the definition of exogenous processes for foreign variables (e.g. foreign interest rates follow an AR(1) process). (5) Finally, we spot different policy implication arising from both models. According to the "agnostic" VAR model, the central bank do not need to rise its policy rate because the drop in activity offsets any burst of inflation; whereas in the DSGE model the rise in prices is partially accommodated by an increase in the policy rate. Thus, this comparison enriches the discussion for the policy maker.

The results for the recursive VAR model are not new and they have been documented many times before in the literature. The identification of monetary policy shocks in this setting has always been a subject of debate and different specifications and models may lead to different responses. Bernanke et al. (2005) provided several reasons to understand this result: (1) the policy shock is not properly identified in the VAR system; (2) variables of the VAR do not represent the real state of the economy; and (3) the impulse response functions are biased because only a subset of the state variables of the economy are used to identify the shocks. Similarly, Weber et al. (2009) argue that structural breaks may be crucial to understand the monetary transmission process. Using data for the Euro area they found two structural breaks in their sample. They report evidence in favor of an "atypical" interim period 1996-1999, but for the rest of the sample, the monetary transmission process remains adequate.

The "agnostic" VAR model of Uhlig (2005) imposes sign restrictions to a subset of the IRFs which in turn imply nonlinear constraints in the structural parameters of the model. In this paper, the author studies the impact of a monetary policy shock on output for the US economy by imposing a set of sign restrictions in all of the variables but leaving the response of output unrestricted. Thus, he refers to this method as an "agnostic" identification scheme[4]. Studies that follow this methodology are Canova and Nicolo (2002), Uhlig (2005), Rubio-Ramírez, Waggoner and Zha (2010) and Arias et al. (2014). These papers extended the VAR framework to also accommodate zero restrictions.

More recently, unconventional monetary policies in the US and the Eurozone have encouraged the use of different frameworks to evaluate the impacts of these shocks (including SVARs, Bayesian VARs, DSGE, etc.), such as Carrera et al. (2015), Baumeister and Benati (2012), Castelnuovo (2012), Christensen and Rudebusch (2012), and Kapetanios et al. (2012), among others. Normally, the choice of restrictions is proposed by the researcher after a careful analysis based on economic theory. For example, if the interest rate differentials increase, then exchange rates are expected to rise due to adjustments one can anticipate from the uncovered power parity relationship. This expected response might be questioned from several angles (e.g. UIP does hold). However, our choice is justified with sound economic theory. Other related applications are presented in Baumeister and Benati (2012) which analyze the effects of unconventional policies with a time varying structural VAR, while Castelnuovo (2012, 2015) use a micro-founded DSGE approach to assess the macroeconomic impacts of an increase in interest rates. Finally, Carrera et al. (2015) have studied the impact of quantitative easing policies on small open economies (a subset of Latin American countries). That piece of research is a very close application to our paper because it uses similar identification methodology, while it differs in the details of the calculation

[4]The key result form this paper is that neutrality of monetary policy is not inconsistent with the US data. More recently, Castelnuovo (2015) addresses this point to the Euro area and analyzes the neutrality of monetary policy on inflation. He reports that the neutrality of VAR models may be due to a deficient identification of the policy shock, omitted variables or structural breaks.

of the posterior distribution[5].

The rest of the paper is organized as follows; the next section presents the VAR models. Section three briefly describes the structural DSGE model economy. Section four reports impulse response functions for each model. Finally, section five concludes.

3.2 Structural VAR models and identification schemes

Structural VAR models were introduced in the seminal paper of Sims (1980a) as an alternative methodology to large-scale macroeconomic models of system of dynamic equations. A complete review of this literature is far from the scope of this paper, but the interested reader may refer to Kilian (2013) and Lütkepohl (2011) for a comprehensive review of this literature.

According to Canova (2007) Structural VAR models can be used to judge and validate theoretical models, such as DSGE models. Because VAR models are able to characterize the joint dynamics of several economic variables with only few assumptions whereas theoretical models rely heavily on a microfounded structure to identify the dynamics between the variables of the system. Thus, the comparison of both methodologies enable us to assess the suitability of the microfounded structure behind a theoretical model if and only if the structural VAR model is properly identified.

The structural VAR model for a SOE with block exogeneity (henceforth SVAR-SOE) is defined as:

$$[y_t^{*\prime} \quad y_t{}'] \begin{bmatrix} A_{01} & 0 \\ A_{03} & A_{04} \end{bmatrix} = \sum_{l=1}^{p} [y_{t-l}^{*}{}' \quad y_{t-l}{}'] \begin{bmatrix} A_{l1} & 0 \\ A_{l3} & A_{l4} \end{bmatrix} + c + [\varepsilon_t^{*\prime} \quad \varepsilon_t{}'] \tag{1}$$

[5] The main difference of Carrera et al. (2015) and our approach is that they estimate the parameters of the blocks of the reduced-form VAR model with block exogeneity independently; whereas our approach remains closer to the original framework of Arias et al (2014) since we estimate the parameters jointly.

The zero blocks in the system reflect the block exogeneity assumption of the model in the spirit of Zha (1999). The $(n \times 1)$ vector y_t contains the endogenous variables for the domestic block (i.e., small open economy), whereas the $(n^* \times 1)$ vector y_t^* the endogenous variables for the foreign block. The A_i matrices and the vector of constants c are the structural parameters, whereas p denotes the lag order of the model. The inclusion of exogenous variables is straightforward but they are excluded to simplify the notation. Finally, the vectors ε_t and ε_t^* are Gaussian with mean zero and variance-covariance matrix I_{n+n^*} (the $(n+n^*)$-dimensional identity matrix).

The model can be compactly written as:

$$Y_t{}'A_0 = X_t{}'A_+ + \xi_t{}' \tag{2}$$

Where $Y_t' = [y_t^{*\prime} \;\; y_t{}']$, $X_t{}' = [Y_{t-1}{}'...Y_{t-p}{}' \; 1]$, $A_+{}' = [A_1{}'...A_p{}' \; c']$; and the reduced-form model is defined as:

$$Y_t = X_t{}'B + u_t{}' \tag{3}$$

Where $B = A_+A_0^{-1}$, $u_t' = \varepsilon_t' A_0^{-1}$ and $E\left[u_t u_t'\right] = \Sigma = (A_0 A_0')^{-1}$. The estimation of SVAR models requires the identification of the structural shocks. Several alternative methodologies are available for the estimation and identification of these type of models. In particular, the most used methodologies can be grouped in three categories: recursive identification schemes, nonrecursive identification schemes and sign restriction schemess; in this paper we explore two of these identification schemes. The next two subsections explain the details of each approach.

3.2.1 Recursive identification scheme

The recursive identification scheme (henceforth recursive scheme or recursive VAR) was introduced in the seminal work of Sims (1980a) and it has become the conventional benchmark that is used in applied macroeconomics to validate responses of micro-founded structural models. The structural model is identified in four steps. First, the variables of the system are ordered in a specific way being the first variable the most exogenous and the last one the most endogenous of the system. Second, the reduced-form model is estimated. Third, the structural innovations are recovered using a Cholesky decomposition over the variance-covariance matrix of the residuals of the reduced-form model (i.e., $\Sigma_\epsilon = PP'$). Finally, the structural parameters are estimated using the map of the reduced-form parameters to the structural parameters defined in the previous subsection:

$$B = A_+ A_0^{-1} \quad u_t{}' = \xi_t{}' A_0^{-1} \quad \Sigma_\epsilon = PP' = (A_0 A_0')^{-1}$$

Note that the P matrix depends on the order of variables and hence is not unique, thus the econometrician needs to rely in some theoretical argument to justify his identification scheme. One of the main drawbacks of this approach is that economic theory can not be incorporated directly into the model. Moreover, even in those cases on which the theory is able to suggest a particular order of causality among the variables of the system, the model can still generate IRF that are counterintuitive or to puzzling results[6].

The block exogeneity assumption for the recursive VAR model for SOE implies that the reduced-form model can not be estimated equation by equation using OLS. Instead, the estimation is performed by quasi-maximum likelihood, see Hamilton (1994) for a comprehensive discussion of this methodology.

3.2.2 Identification with sign and zero restrictions

The sign restriction scheme follows a different approach to identify the structural shocks of the model. In this setting, the IRF of the model are restricted directly according to economic theory. For instance, the contemporaneously dynamic response

[6] Sims (1980a) defines a puzzle as a situation in which the impulse response functions from an identification scheme do not match conventional wisdom from theoretical models.

of inflation is set to be less than zero to a positive monetary policy shock as well as for the first periods following the shock. The methodology imposes linear and nonlinear constraints in the structural parameters of the model. In addition, the methodology does not require the complete identification of the full set of structural shocks of the model as in the recursive scheme. However, in this case the identification of the subset of structural shocks can be contaminated with other structural shocks that look alike. Thus, the full identification of the shocks should generate narrower confidence intervals for the IRF of the system. Alternatively, the researcher can increase the number of restrictions to try to minimize the aforementioned problem[7].

There are several ways in which sign restrictions can be introduced in VAR models. For instance, Blanchard and Quah (1989) developed an algorithm to restrict the long-run response of a set of variables after a structural shock. Other authors have restricted the joint dynamics of the variables after a structural shock, as in Canova and Nicolo (2002). A different approach is used in Uhlig (2005) to study the impact of a monetary policy shock on output for the US economy by imposing a set of sign restrictions in all of the variables but leaving the dynamic response of output unrestricted. The author referred to this method as an "agnostic" identification scheme since no assumptions were made with respect to the response of output. In this setting the restrictions are imposed directly over the dynamics of each variable of the system. More recently, extensions to these approaches can be found in Mountford and Uhlig (2009), Rubio-Ramírez, Waggoner and Zha (2010) and Arias et al. (2014) (henceforth ARW). In particular, ARW expands Uhlig's methodology by incorporating zero restrictions; thus the dynamic responses of the variables after a shock can be set to zero, less than zero or greater than zero. In addition, the methodology allows the combination of these type of restrictions simultaneously in the dynamic response of

[7]Unfortunately, there is little guide to asses the potential gains from this approach. But, further research may help to understand the trade-off between these two approaches.

the variables which in turn should improve the identification of the structural shocks[8].

In this paper we extend the methodology of Arias et al. (2014) for SOE; for ease of exposition we borrow Uhlig's definition and refer to this method as "agnostic" scheme or "agnostic" VAR. The block exogeneity assumption implies that the number of independent variables is not the same between the blocks of the model and thus we follow Koop and Korobilis (2010) to use a more general framework to estimate VAR models. The implications of this identification scheme has not been explored comprehensively in the literature for SOE. This approach enable us to specify an alternative VAR model on which the identification of structural shocks is based on a set of restrictions that are driven by theory (or by stylized facts of the data) and not just by a particular order of the variables as in the recursive scheme. Thus, this method could potentially provide an interesting benchmark to evaluate and validate the responses of theoretical models.

In this setting, the identification of the structural shocks relies on Bayesian methods and the algorithm can be summarized as follows:

1. Draw $(B; \Sigma)$ from the posterior of the reduced-form parameters.

2. Generate $(A_0^*; A_+^*)$ by using a mapping between the reduced-form and the structural parameters[9].

3. Draw an orthogonal matrix Q such that $(A_0^* Q; A_+^* Q)$ satisfies the zero restrictions[10].

[8] More precisely, the inclusion of zero restrictions to Uhlig's method was developed in Mountford and Uhlig (2009) using a penalty function approach. However, according to ARW the method imposes additional sign restrictions in unrestricted variables which generate narrower confidence intervals for the responses of the variable. Thus, ARW shows a new framework to correctly combine the two type of restrictions.

[9] The mapping between structural and reduced-form parameters can be implemented by using a function $h()$ such that $h(X)'h(X) = X$, i.e. Cholesky decomposition: $(A_0^*; A_+^*) = (h(\Sigma)^{-1}; Bh(\Sigma)^{-1})$

[10] Using the QR decomposition $(X = QR)$ which holds for any $n \times n$ random matrix on which each element is $i.i.d.$ from a $N(0, 1)$. In addition, ARW describes an algorithm to obtain recursively each column of Q, which improves the efficiency of the algorithm significantly when the researcher is interested in identifying more than one structural shock.

4. Keep the draw if sign restrictions are satisfied.

5. Repeat 1 to 4 until the desired number of simulations is reached.

6. Compute the median and confidence bands for the full set of IRF that satisfy the restrictions.

If no restriction are imposed over the blocks of the SVAR-SOE, then each equation of the model has the same number of variables. In this case the draws from the posterior of the reduced-form parameters can obtained using the Normal-Wishart Prior (conjugate prior) and the posterior of the parameters are given by[11]:

$$b|\Sigma, y \sim N(\bar{B}, \bar{\Sigma} \otimes \bar{V}) \quad \text{and} \quad \Sigma^{-1}|y \sim W(\bar{S}^{-1}, \bar{\nu})$$

and:

$$\bar{S} = S + \underline{\mathrm{S}} + \hat{B}'X'X\hat{B} + \underline{\mathrm{B}}'\underline{\mathrm{V}}^{-1}\underline{\mathrm{B}} - \bar{B}'(\underline{\mathrm{V}}^{-1} + X'X)\bar{B}$$

The Normal-Wishart prior imposes a Kronecker structure in the variance-covariance matrix of b which in turn implies that for each element of b, say b_i the $cov(b_i, b_j) \neq 0$ for all $i \neq j$. Unfortunately, the block exogeneity assumption requires a block of zeros in the reduced-form model which means that these set of parameters must be independent from the rest of the parameters. Therefore, the Normal-Wishart prior is not suitable to estimate the SVAR-SOE model. Instead, we need to specify a prior that breaks the Kronecker structure in the variance-covariance matrix of b.

Following Koop and Korobilis (2010), we use the Independent Normal-Wishart Prior that defines the posterior of the parameters as follow[12]:

[11]Where $\bar{\nu} = T + \underline{\nu}$; $b = vec(\bar{B})$ and $\hat{B}$ is the OLS estimator of B; $\bar{V} = \left[\underline{\mathrm{V}}^{-1} + X'X\right]^{-1}$ and $\bar{B} = \bar{V}\left[\underline{\mathrm{V}}^{-1}\underline{\mathrm{B}} + X'X\hat{B}\right]^{-1}$; the hyperparameters $\underline{\alpha}$, $\underline{\mathrm{V}}$ and $\underline{\mathrm{S}}$ characterized the prior distributions of the parameters:

$$b|\Sigma, y \sim N(\underline{\mathrm{B}}, \underline{\Sigma} \otimes \underline{\mathrm{V}}) \quad \text{and} \quad \Sigma^{-1}|y \sim W(\underline{\mathrm{S}}^{-1}, \underline{\nu})$$

[12]Where: $\bar{\nu} = T + \underline{\nu}$; $\bar{B} = \bar{V}\left[\underline{\mathrm{V}}^{-1}\underline{\mathrm{B}} + \sum_{t=1}^{T} Z_t'\Sigma^{-1}y_t\right]$ and $\bar{V} = \left[\underline{\mathrm{V}}^{-1} + \sum_{t=1}^{T} Z_t'\Sigma^{-1}Z_t\right]^{-1}$; the hyperparameters $\underline{\alpha}$, $\underline{\mathrm{V}}$ and $\underline{\mathrm{S}}$ characterize the prior distribution of the parameters: $b \sim N(\underline{\mathrm{B}}, \underline{\mathrm{V}})$ and $\Sigma^{-1} \sim W(\underline{\mathrm{S}}^{-1}, \underline{\nu})$ with $p(b, \Sigma^{-1}) = p(b)p(\Sigma^{-1})$

$$b|\Sigma, y \sim N(\bar{B}, \bar{V}) \quad \text{and} \quad \Sigma^{-1}|y, b \sim W(\bar{S}^{-1}, \bar{\nu})$$

and:

$$\bar{S} = \underline{S} + \sum_{t=1}^{T}(y_t - Z_t b)(y_t - Z_t b)'$$

Thus, the main methodological contribution of this paper is to combine the methods of Koop and Korobilis (2010) and Arias et al. (2014) to identify the SVAR-SOE model. In this setting, the model needs to be redefined in the following way. First, rewrite (3) as:

$$y_{mt} = z'_{mt} b_m + \varepsilon_{mt}$$

Where t is the time index and m indicates the variable (i.e., equation); y_{mt} specifies the t^{th} observation of the m^{th} variable and z_{mt} is a vector that contains the explanatory variables for the m^{th} equation at time t. Second, define b_m as the vector that contains the parameters of the m^{th} equation and M as the total number of equations. Note that in this case the z_{mt} vector can vary across equations or blocks of the model. Third, stack the b_i vectors and z'_{mt} matrices as:

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix} \qquad Z_t = \begin{pmatrix} z'_{1t} & 0 & \dots & 0 \\ 0 & z'_{2t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & z'_{Mt} \end{pmatrix}$$

Next, define $y_t = (y_{1t}, ..., y_{Mt})'$, $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Mt})'$ and write the model more compactly as:

$$y_t = Z_t b + \varepsilon_t$$

The total number of parameters is given by $k = \sum_{j=1}^{M} k_j$ and $\varepsilon_t \sim N(0, 1)$. Note that b is a $k \times 1$ vector and Z_t is an $M \times k$ matrix. Finally, stack y_t, ε_t and Z_t as column vectors and define $\varepsilon \sim N(0, I \otimes \Sigma)$ to write the model as:

$$y = Zb + \varepsilon \tag{4}$$

The notation in equation (4) is consistent with the notation of Koop and Korobilis (2010) for the Independent Normal-Wishart Prior. Note that the posterior of Σ is not independent from the draw of b and hence direct sampling from the posterior is not feasible. Instead, A sequential algorithm can be used on which sequential draws are taken from the conditional posterior distributions of $p(b|y, \Sigma)$ and $p(\Sigma^{-1}|y, b)$, i.e. a Gibbs sampling algorithm[13].

3.3 A DSGE model for Chile

In this section we briefly describe the DSGE model for Chile. We use the model of Medina and Soto (2007a) to compute the impulse response to a 1% foreign monetary policy shock. The model is a new Keynesian small open economy model which is closely related to the framework of Christiano et al. (2005) and Smets and Wouters (2003, 2007), but has additional and specific features to describe the Chilean economy, such as a representative commodity-exporting firm, a "structural" fiscal policy rule and a monetary policy rule that responds to changes of headline CPI inflation (we refer to Medina and Soto (2007a) for a more detailed description of the model).

This model has been extended in several directions to address specific questions and has also been reestimated to take advantage of recent data. Examples are the learning extension to replicate the Current Account dynamics of Chile as Fornero and Kirchner (2014) and Fornero et al. (2015) conduct several policy experiments simulating a copper price shock. In the current version, we abstract from these additions[14].

A full description of the model is beyond the scope of this paper. Therefore, in the remainder of the section, we briefly describe main features. The domestic economy is

[13] We use a burn-in period to achieve convergence to the posterior distribution. In particular, we made 5500 simulations and burned the first 500 simulations. We also tried with different number of simulations but the results did not change significantly.

[14] Robustness exercises were done using the model of Fornero and Kirchner (2014) and Fornero et al. (2015) and we did not find any relevant advantage of adding an endogenous commodity-exporting sector in order to compute the IRFs to a foreign monetary policy shock.

composed by a continuum of households, where a fraction are non-Ricardians without access to capital market. These non-Ricardian households consume entirely their wage income. The remaining Ricardian households make intertemporal consumption-savings decisions in a forward-looking manner, so as to maximize the present value of utility.

There are three types of sectors in the domestic economy. First, there is a continuum of firms producing differentiated varieties of intermediate tradable goods, with monopoly power and sticky prices à la Calvo (1983). These firms use labor, capital and oil as inputs and sell their goods to competitive assemblers that produce final domestic goods, which are sold in the domestic and foreign market. There is a representative capital goods producer that rents capital goods to the intermediate goods producing firms. The optimal investment composition is determined through cost minimization, where we assume costs of adjusting investment, following Christiano et al. (2005). All firms are owned by Ricardian households. Second, there is an imported goods sector with a continuum of retail firms that repackage a homogenous good from abroad into differentiated imported varieties. There is a large set of firms that use a CES technology to assemble final imported goods from imported varieties. These firms also have monopoly power and set their prices infrequently. All firms are also owned by Ricardian households. Third, there is an exogenous commodity-producing sector composed by a unique representative firm. The entire production is exported abroad and the international price of the commodity is taken as given. A fraction of the assets of that firm is owned by the government and the remaining fraction is owned by foreign investors, where the revenue is shared accordingly.

Monetary policy is conducted through a simple Taylor-type feedback rule for the nominal interest rate, where the central bank responds to headline CPI. The fiscal policy follows a structural balance fiscal rule, where government expenditure (government consumption and transfers to households) depends on cyclical adjustments of commodity price and output gap. Also, the model includes distortional taxes in consumption, income and capital gains.

There is a foreign sector which is composed by 5 exogenous variables (GDP, inflation, interest rate, oil price and commodity price). We assume that the dynamics of these foreign variables are described by independent autoregressive processes of order one, AR(1), as in Medina and Soto (2007a) and Fornero and Kirchner (2014). We choose this framework instead of a foreign SVAR block (as in Fornero et al. (2015)) to avoid selecting a SVAR identification scheme in the DSGE model[15].

Finally, the model is parameterized using estimates from Bayesian estimation techniques with quarterly data covering the period 2001Q3-20107Q4 and 2001Q3-2014Q4 to analyze the robustness of the results. We use their posterior mean to compute the impulse responses to a foreign interest rate shock[16].

3.4 Results

This section is divided in four parts for ease of exposition. The first part describes the data used to estimate the VAR models along with the set of identify assumptions behind the recursive and "agnostic"' schemes. The second part shows the comparison of the IRFs for both identification schemes and it highlights their similarities and differences. The third part shows the IRF from the DSGE model for the Chilean economy. Finally, the last part compares the IRFs of the VAR and DSGE models. Thus, this comparison between models provides new lights and insights about the propagation of a foreign monetary policy shock over the Chilean economy and in addition it assess the suitability of the DSGE model (i.e., the theoretical model).

[15]In this case, the impulse responses computed by the DSGE would be influenced by the identification scheme chosen for the foreign SVAR block

[16]Details of the Bayesian estimation are available on request. In particular, the persistence of the shock is calibrated to 0.87 following Medina and Soto (2007a). This value arises when the AR(1) process is estimated with a sample that ends before the Subprime Crisis.

3.4.1 Data and identification schemes for SVAR-SOE models

The data are monthly observations covering the period from January 1996 to December 2007[17] (1996.01 - 2007.12). The same data set is used in both recursive and "agnostic" identification scheme. Table 3.1 shows the variables for each block of the SVAR-SOE model.

We transform price indexes in nominal US dollar terms (original sources) to real prices by dividing (deflating) by an external price index constructed to reflect the foreign Chilean trade structure. Domestic real GDP, investment and price indexes are seasonally adjusted using the Census X-12 procedure when they are not available in seasonally adjusted form the original source. The interest rates are defined in levels and the rest of the variables in logs. We choose a 2-month lag based on standard information criteria and also following the recommendation of Castelnuovo (2015).

Table 3.1
Set of variables for the SVAR-SOE models

Foreign block US)	Domestic block (Chile)
- Industrial production index (y*)	- Index of economic activity (y)
- Consumer price index (CPI*)	- Real machinery and equipment investment (Ime)
- US Federal Funds rate (r*)	- Real construction investment (Ic)
- (US Shadow Federal Funds rate)	- Core consumer price index (CPIx1)
- (Real price of oil)	- Nominal monetary policy rate (r)
	- Real exchange rate (RER)

We use the Chow Lin procedure to transform quarterly to monthly frequency (e.g. domestic investments). Variables in parentheses in the foreign block are considered only for robustness exercises and not for the baseline model (exercises not reported). For further details concerning variables, sources and transformations see Table A.3.1 in Appendix 3.A.

We do not include cointegration relationships in the SVAR-SOE because we analyze the short-term dynamics and not the long-run behavior of the model. The main drawback of this approach is that we need to rely in simulation methods to make valid inference over the IRFs of the models, see Sims et al. (1990) for a comprehen-

[17]The data after December 2008 is excluded because we want to isolate the propagation of the shock during a "normal" monetary regime and clearly this was not the case after December 2008 since the federal funds rate experienced a unique path compared to its historical behavior (from September 2007 to April 2008, the policy rate decreased from 5.25% to 2%). But we also estimate the models using the implicit foreign interest rate (shadow federal funds rate) covering the period from January 1996 to December 2014 to analyze the robustness of our results since this rate is not bounded below by zero.

sive discussion of this issue. Finally, we control for the real price of copper, linear time trends and add a constant term to each equation of the model.

The recursive VAR model is specified as in Fornero et al. (2015), the variables for each block were ordered according to Table 3.1 (i.e., most exogenous variables from top to bottom). In particular, this setting assumes that the domestic policy rate reacts contemporaneously with the rest of the variables in the system except with exchange rate and that it can not have a contemporaneous impact in the rest of the variables of the domestic block except the exchange rate; whereas the foreign policy rate has a contemporaneous impact over the domestic block but not over the rest of the variables of the foreign block.

Table 3.2 shows the set of restrictions for the "agnostic" VAR model. In addition, the table also describes two alternative "agnostic" models in order to assess the robustness of the base model. The foreign monetary policy shock is assumed to be positive for at least 1 month. The shock does not have a contemporary impact on the foreign block, neither in domestic output and investment (both type of investment). We remain agnostic with respect to the contemporaneous response of the domestic policy rate and CPI but we assume a real depreciation that last for at least one month. Finally, we assume that the variables of the foreign block react to the shock with a lag as well as domestic investment; but we assume a more persistent impact over the later variable based on empirical data[18]. The two alternative "agnostic" VAR models explore the sensitivity of the results to the restrictions imposed over domestic investment which are perhaps the more controversial of the restrictions. In particular they consider two cases, one on which negative sign restrictions only last one period (Mod A) and a second case on which these restrictions last for at least three periods

[18]A different approach would be to rely on a "agnostic" VAR that restricts heavily the foreign block while minimizing the number of restriction in the domestic block or in the extreme case by leaving it completely unrestricted. However, the short sample of the data available for the Chilean economy makes this approach unsuitable since there is not enough information (data) to unveil the propagation of the shock.

(Mod B). Thus, the base model lies between these two alternative cases. We also consider two additional alternative models on which we increase the restrictions over the foreign monetary policy and the real exchange rate for the base model, see Table A.3.2 of Appendix 3.A for further details of these two cases.

Table 3.2
Sign and zero restrictions for "agnostic" VAR models

	Base Model		Mod A	Mod B
	$h = 0$	$h > 0$	$h > 0$	$h > 0$
Foreign block				
-US Federal Funds rate (rus)	1	?	?	?
-Industrial production index (Yus)	0	-1	-1	-1
-Consumer price index (CPIus)	0	-1	-1	-1
Domestic block				
-Interest rate (r)	?	?	?	?
-Monthly production index (Y)	0	?	?	?
-CPI core	?	?	?	?
-Investment (I)	0	-2	-1	-3
-Real Exchange Rate (RER)	1	?	?	?

Restrictions are imposed over the monthly IRFs of the model after a positive foreign monetary policy shock. Positive or negative entries indicate the length of the sign restrictions, whereas zero entries indicate zero restrictions. Finally, question marks (?) indicate that no restrictions were imposed over the IRF of the variable at that horizon. We also consider two additional alternative set of restrictions for the base model, see Table A.3.2 in Appendix 3.A for more details.

The IFRs for the three cases are computed using monthly data but we aggregate the monthly responses to quarterly responses in order to make the results comparable to the IRFs of the DSGE model, Alternatively, the IRFs can be estimated using quarterly data directly but we argue that the identification of the foreign monetary policy shock is more reasonable at monthly frequency, since at quarterly frequency the restrictions constraint the contemporaneously response of the variables which at this time frequency would imply stronger identifying assumptions. The same argument applies to the recursive scheme.

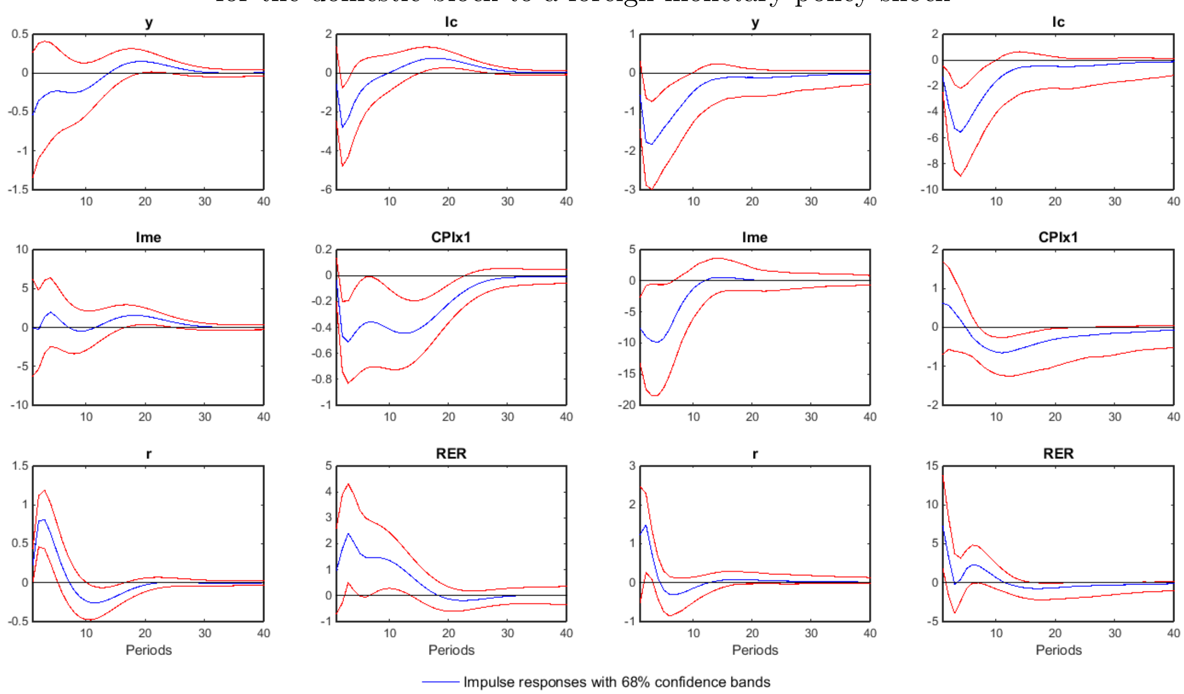
3.4.2 Results for SVAR-SOE models

To begin with, we illustrate in Figure 3.1 the impulse responses of the domestic blocks to a 1% positive shock to the foreign interest rate (100 basis points) for the

SVAR-SOE model according to the recursive (left panel) and "agnostic" (right panel) identification schemes. The responses for the foreign blocks are reported in Figure B.3.1 of Appendix 3.B.

In general, the identification of the recursive VAR model yields puzzling responses. In particular, the monetary policy shock is associated with expansionary conditions in the world economy (a boost in trade partners' activity increases in foreign prices and in real commodity prices). In the domestic economy, the effect on investment are slightly positive and at the same time the impact in local activity is not significant. The fluctuations of RER anc CPIx1 turn out to behave inconsistently because the appreciation of the real exchange rate should be associated with higher inflation, but the CPI drops. The drop on inflation can be associated to the local response of the interest rate.

Figure 3.1
Impulse responses for the recursive and "agnostic" identification schemes
for the domestic block to a foreign monetary policy shock



Recursive VAR in the first two columns; "agnostic" VAR in the last two columns for the base line model. The figure shows the quarterly responses to a 1% positive shock to the foreign monetary policy rate at the monthly frequency. The quarterly responses were computed by aggregating the monthly responses of the model.

Thus, according to these results the foreign shock has a small and limited impact over the domestic economy. In addition, the identification infers that the Central Bank reacts aggressively to contain any burst in inflation due to the pass-through of RER to inflation. However, at the same time the recursive identification scheme infers almost no impact over the local activity and investment[19]. There are at least two problem with this interpretation. First, according to the dynamics of the foreign block the recursive VAR model is not able to properly identify the shock, and thus the previous analysis for the domestic block is not correct. Second, even if we are willing to believe that the model was able to identify the foreign shock, the results suggest that the shock has an extremely limited impact over the domestic economy which seems unrealistic in light to the magnitude of the shock. Thus we conclude that in this case, the recursive VAR model fails to provide an informative benchmark to judge and validate the IRFs of our structural micro-founded model.

The results for the "agnostic" VAR model offer a completely different view of the propagation of the shock. Overall, the impulse responses show results in line with macroeconomic theory. Besides, they are statistically significant at conventional levels (with the exception of inflation and the domestic policy rate). The responses for foreign variables show dynamics that are consistent with those expected after a negative policy shock (i.e., a contractionary effect in foreign prices and activity). It is worth noticing that the responses in the foreign block go further beyond the restrictions that were specified in this identification scheme and thus these results suggest that the shock is properly identified. In the domestic block, the shock has a strong negative impact over output and the two type of investment in the short run (around ten quarters). Moreover, the responses are significant at conventional levels. The fall of investment is mainly due to the large real exchange rate depreciation in line

[19] We explored several alternative specifications to confirm these results. The first exercise consists of changing the order of variables (we assume the interest rate to be the most exogenous variable in the foreign block) and the results are qualitatively very similar.

with tighter monetary conditions abroad (outflow of capitals, etc.). Finally, results show no impact over domestic prices due to the strong drop in the domestic activity that compensates the pass-through of the exchange rate to prices in the short run which would also explain the lack of response for the domestic rate. However, there is a small drop in prices in the median-run due to the normalization of the exchange rate and depressed domestic activity.

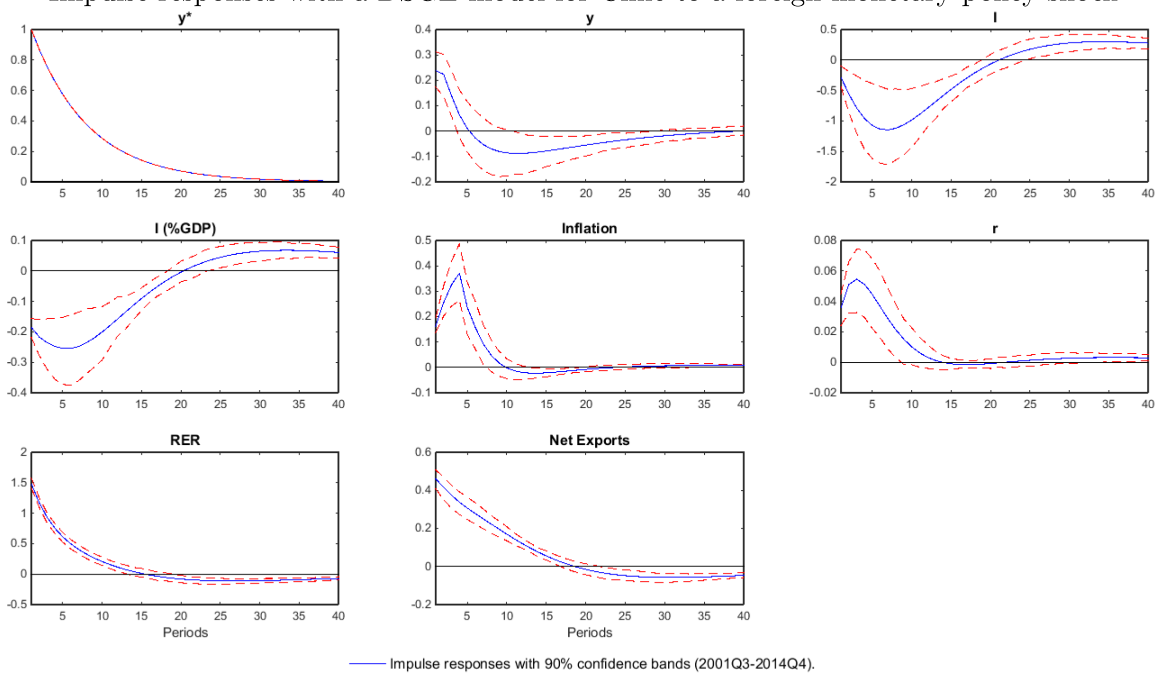
Therefore, we argue that the "agnostic" VAR model is able to proper identify the foreign monetary policy shock and the responses from this identification scheme can be use to validate the responses of our DSGE model. The comparison of these two models will enable us to give new lights and insights about the propagation of the foreign monetary policy shock over the domestic economy. In particular, we can compare and analyze the different policy implications for the domestic Central Bank; as well as the short/long-run dynamics and the convergence toward the equilibrium implied by both models in order to better characterize the propagation of the shock.

To analyze the robustness of the results for this identification scheme we consider four alternatives set of sign restrictions, see Table 3.2 from the previous subsection and Table A.3.2 in Appendix 3.A for more details. The IRFs of these four alternative models are reported in Figure B.3.2 and Figure B.3.3 of Appendix 3.B. In particular, Mod A and B show that restrictions in investment have a significant impact in the real variables but nominal variables show similar dynamics between the alternative cases and base model. Thus, our conclusions hang on the plausibility of these restrictions. Finally, additional restrictions in foreign policy rate and real exchange rate do not change the responses of the variables significantly with respect to those reported for the base model.

3.4.3 Results for the DSGE model

DSGE models are highly parameterized and thus we estimate the model using data covering the period 2001Q3-2014Q4 in order to improve the identification of the parameters of the models. Figure 3.2 illustrates the responses of the DSGE model to a 1% positive shock (100 basis points) to the foreign interest rate.

Figure 3.2
Impulse responses with a DSGE model for Chile to a foreign monetary policy shock



Model is parameterized using estimates from Bayesian estimation techniques with quarterly data covering the period 2001Q3-2014Q4. The figure shows the Bayesian impulse responses to a 1% positive shock (100 basis points) to the foreign interest rate. We assume that the dynamics of the foreign variables are described by independent autoregressive processes of order one, AR(1), as in Medina and Soto (2007a) and Fornero and Kirchner (2014).

The tightening of foreign monetary conditions will lead to capital outflows away from Chile. This will endogenously influence the country risk premium (the debt burden increases if the country is net borrower). Because of this, there will be a depreciation of the local currency both in nominal and in real terms[20]. To fight

[20] Notice that we take a conservative stance regarding the implications of the financial tightening in the U.S.. We can expect additional financial distress triggered by larger volatility in emerging economies such as: (i) an increase of default probabilities of these countries yielding to a boost of country risk premiums; (ii) the appreciation of the U.S. dollar worldwide leading to unfavorable dynamics in commodity prices and in terms of trade of emerging economies. These further effects

against inflationary pressures, the central bank raises the policy rate. The latter causes a large fall in activity, particularly in investment that decreases slightly more than 1% below its steady state value.

The real exchange rate rises persistently and, during the first periods, roughly depreciates by one and a half percent. In consequence, marginal costs increase causing inflationary effects (around 0.2% on impact). As nominal prices are rigid, the inflationary peak is reached at the end of the first year. Also, the results suggest that the immediate pass-through is 0.18 and it increases towards the end of the first year. Moreover, consumption expenses also fall due to the increase in real interest rates (not shown in the figure). Consequently, the model predicts a modest but persistent contraction in output. Notice that the large persistence of the foreign monetary policy shock drives these important fluctuations. Finally, the persistence of the shock contributes to a large improvement of the current account which explains the initial hike in output.

3.4.4 Comparing the results of SVAR-SOE and DSGE models

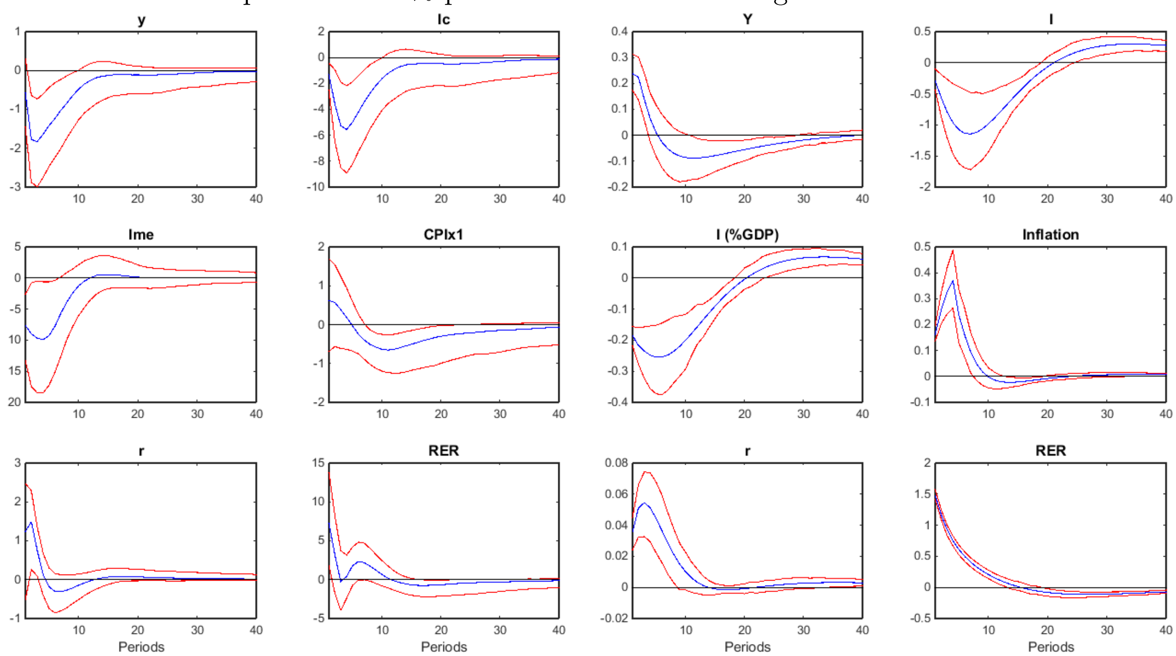
The main results from the IRFs analysis showed that the recursive VAR model was not able to identify the foreign monetary policy shock and thus this identification scheme is excluded from the comparison.

Before jumping into the comparison of the responses between the "agnostic" VAR (right panel Figure 3.1) and DSGE model (Figure 3.2), there are two points that we need to address. First, responses for VAR models were constructed by aggregating monthly responses to quarterly frequency and hence their confidence intervals are wider than they should be because variables are smoother at higher frequencies. Thus the sensitivity of the responses to the restrictions in investment should be re-

can be captured by setting a SVAR for these foreign variables instead of an AR(1) model for each variable. We avoid implementing that SVAR due to strange implications arising from the Cholesky identification discussed above.

considered. Second, the DSG model uses data from the period after 2008 whereas VAR models do not, hence the comparison of the results may not be straightforward. Thus, we also estimated an alternative DSGE model using a more comparable data set but the results did not change significantly[21]. Figure 3.3 summarizes the results for the "agnostic" VAR and DSGE model.

Figure 3.3
Impulse responses for the "agnostic" VAR and DSGE model
Responses to a 1% positive shock to the foreign interest rate



"Agnostic" VAR in the first two columns for the base line model (quarterly responses were computed by aggregating monthly responses). DSGE model in the last two columns (bayesian impulse responses). The figure shows the 68% and 90% confidence bands for the VAR and DSGE model, respectively.

Beyond the quantitative differences, we find that the impulse responses of the "agnostic" VAR model are in line qualitatively with the results of the DSGE model except for output. In the DSGE model the initial hike is explained by the improving of the current account due to the real and nominal exchange rate depreciation; whereas the "agnostic" VAR infers a drop of almost two percent in output.

There are three key issues in the dynamics of the responses inferred by the DSGE

[21] See Figure B.3.4 of Appendix 3.B for the complete set of responses for this alternative DSGE model. The main difference is that the responses are exacerbated in this case.

model that we want to highlight. First, the model infers a limited propagation of the shock to the domestic economy which may seem problematic in light of the size of the shock. Second, the peak of the shock over activity occurs during the second and third year after the shock (impact of the shock accumulates in time slowly). Finally, convergence toward the steady state is reached only in the long-run. The last two issues may be due to the many micro-founded restrictions that are included in the model[22]. Ironically, these mechanisms are added to better-fit the persistence observed in the data. On the contrary, the "agnostic" VAR offers a slightly different view about the propagation of the shock. In particular, it clearly indicates that the shock is much less persistent but at the same time it has a greater impact in the short-run. Finally, policy implications from both models turned out to be different, according to the "agnostic" VAR model, the central bank do not need to rise its policy rate because the drop in activity helps to contain any burst in inflation; whereas in the DSGE model the rise in prices is partially accommodated by the increase in the policy rate.

Of course, both models are approximation and thus we favor the view that the responses will lie between the responses of both models. The main advantage of the DSGE model is that it offers a comprehensive description of the propagation of the shock that enriches policy discussions. However, this comparison enable us to: (1) validate the responses of the theoretical model (i.e., DSGE model) for the Chilean economy; (2) better understand the propagation of the shock over the domestic economy, in terms of duration, length and deep; (3) help us to develop potential improvements to the structure behind the DSGE model in order to address the key three issues outlined in the previous paragraph; (4) offer a richer policy discussion for the policy maker.

[22]One example of these micro-founded restrictions is the delay in domestic consumption because of the assumption of consumption habits.

3.5 Conclusions and further discussion

This paper investigates the propagation of a foreign monetary policy shock over a small open economy, in particular over the Chilean economy. Our motivation is base on the ongoing period of monetary normalization already started by the Fed. We use a comprehensive methodological framework (i.e., two Structural VAR models and a DSGE model tailored for the Chilean economy) in order to give new lights and insights about the propagation of the shock. We use this approach because according to Canova (2007), Structural VAR models can be used to judge and validate the responses from DSGE model. This exercise is important because the main advantage of DSGE models is that they provide a comprehensive description of the economy. Our main methodological contribution is to combine the methods of Arias et al. (2014) and Koop and Korobilis (2010) to develop an "agnostic" VAR model for SOE.

The results suggest that the recursive VAR model is not able to identify the shock since some of the responses are counterintuitive (specially for the foreign block). These results are in line with Bernanke et al. (2005), Mojon (2008) and Castelnuovo (2015). Thus, this identification scheme can not be used to judge the responses of the DSGE model. On the contrary, the "agnostic" VAR model shows results in line with macroeconomic theory. The comparison between the "agnostic" VAR and DSGE model show that both approaches infer similar responses for the economy, except for output. In addition, we identify three points that deserve further attention in the dynamics of the DSGE model: (1) the impact of the shock, (2) peak of the shock and (3) the convergence toward the steady-state. Finally, we spot different policy implication arising from both models. According to the "agnostic" VAR model, the central bank do not need to rise its policy rate because the drop in activity offsets any burst in inflation; whereas in the DSGE model the rise in prices is partially accommodated by the increase in the policy rate. Thus, this comparison enriches the discussion for the policy maker.

Therefore, our results suggest that there is a gap in the interpretation of the propagation of the foreign monetary policy shock in these models. Further research is needed to develop a better propagation mechanism in the DSGE model to solve or improve the short and long-run propagation mechanism of the shock. We leave these issues to further work. However, we recognize and propose two potential improvements for the DSGE model. First, significant gains could be made by improving the time series properties of the foreign shocks in these type of models; the DSGE model combines an AR(1) process to describe the foreign interest rate, which is, admittedly, extremely simple. The lack of a foreign propagation mechanism can help to explain the observed responses in this model. Second, the lack of financial restrictions mitigates the propagation of the shock; the model can be improved by including a financial accelerator as in Bernanke (1999). In brief, these improvements provide an opportunity to investigate the causes of the differences between the "agnostic" VAR and DSGE model.

Finally, we do recognize that our comparison does not have a real benchmark to judge each model independently. A more elegant approach to perform the comparison would be to specify a more general DSGE model and simulate data from it. Then, we can compute and compare the responses of each model according to a loss function. However, our approach remains valid since it fosters the discussion for the policy maker. In addition, the specification of a true model is always a controversial assumption and in this case it would be similar to the DSGE model and thus the comparison can be biased toward this model.

Appendix 3.A

Table A.3.1
Data used for the estimation of SVAR models

Variable	Description
-Log World real GDP	World real GDP index, US index of industrial production (both SA)
-Log foreign price index	Chilean external price index (IPE) and US consumer price index (both SA)
-Foreign interest rate	Fed Funds rate
-Log real copper price	Real copper price
-Log real oil price	Real WTI oil price
-Log domestic real GDP	Monthly economic activity indicator (IMACEC) (SA)
-Log domestic price index	Consumer price index (IPC, 2013=100) (SA)
-Log real exchange rate	Multilateral real exchange rate
-Domestic interest rate	Monetary policy rate
-Log real investment in machinery and equipment	Real gross fixed capital formation in machinery and equipment (SA)
-Log real investment in construction	Real gross fixed capital formation in construction (SA)

Sources: The Central Bank of Chile and Federal Reserve Economic Data - FRED - St. Louis Fed. The Log World real GDP constructed using the Chow-Lin procedure with monthly world production index for the World real GDP index, the Log real copper price and oil price were deflated with international price index (IPE, 2005=100). Finally, an increase in the exchange rate denotes a depreciation.

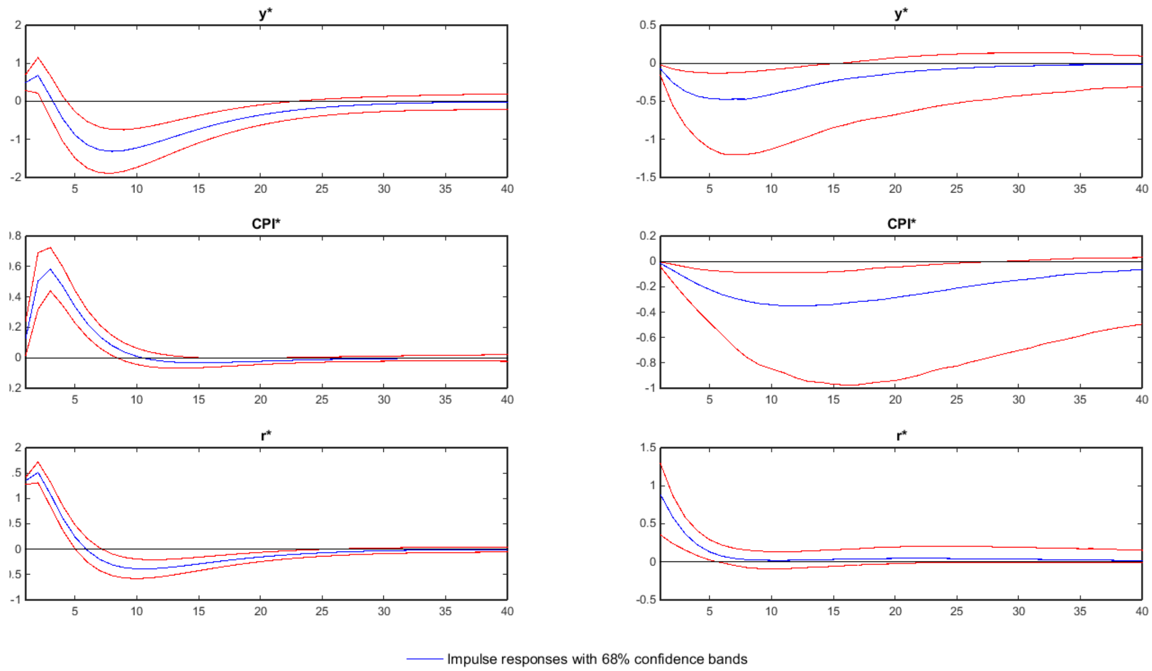
Table A.3.2
Alternative "agnostic" VAR models
Sign and zero restrictions

	Base Model		Mod C	Mod D
	$h = 0$	$h > 0$	$h > 0$	$h > 0$
Foreign block				
-US Federal Funds rate (rus)	1	?	2	2
-Industrial production index (Yus)	0	-1	-1	-1
-Consumer price index (CPIus)	0	-1	-1	-1
Domestic block				
-Interest rate (r)	?	?	?	?
-Monthly production index (Y)	0	?	?	?
-CPI core	?	?	?	?
-Investment (I)	0	-2	-2	-2
-Real Exchange Rate (RER)	1	?	?	2

Restrictions are imposed over the monthly IRFs of the model after a positive foreign monetary policy shock. Positive or negative entries indicate the length of the sign restrictions, whereas zero entries indicate zero restrictions. Finally, question marks (?) indicate that no restrictions were imposed over the IRF of the variable at that horizon. We also consider two additional alternative set of restrictions for the base model; Mod C considers the foreign monetary policy to be positive for at least 3 months. Mod D considers the foreign monetary policy and the real exchange rate to be positive for at least 3 months. Thus, these two alternative "agnostic" schemes are incremental cases of the base model.

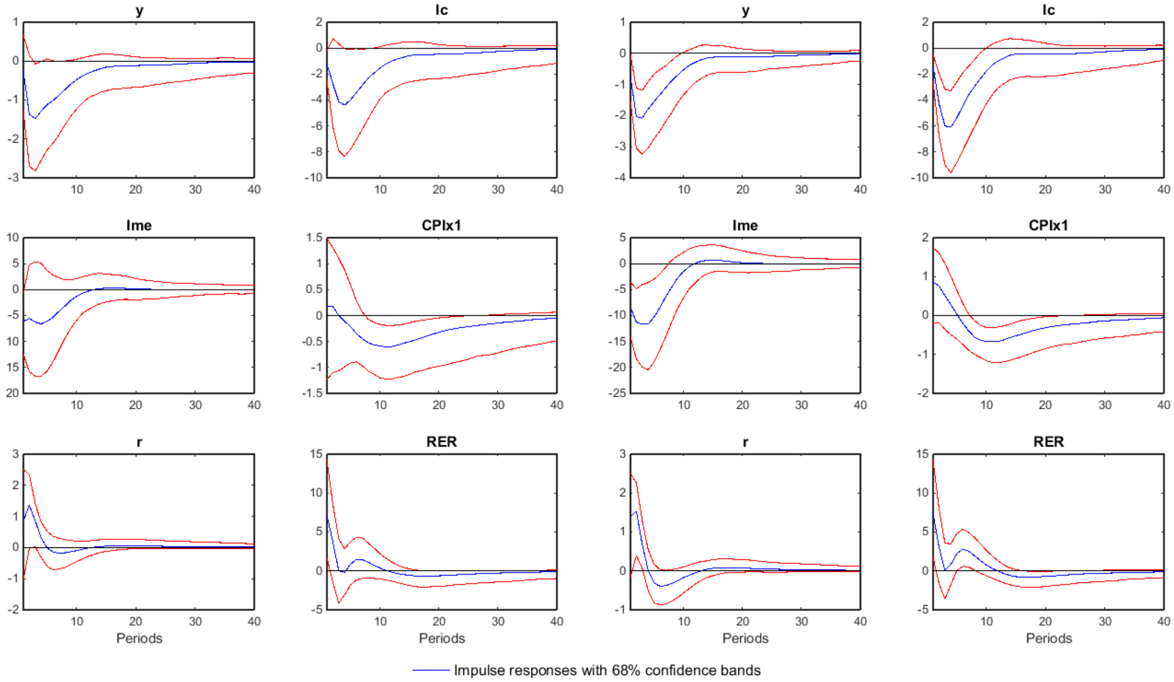
Appendix 3.B

Figure B.3.1
Impulse responses for the recursive and "agnostic" identification schemes for the foreign block to a foreign monetary policy shock



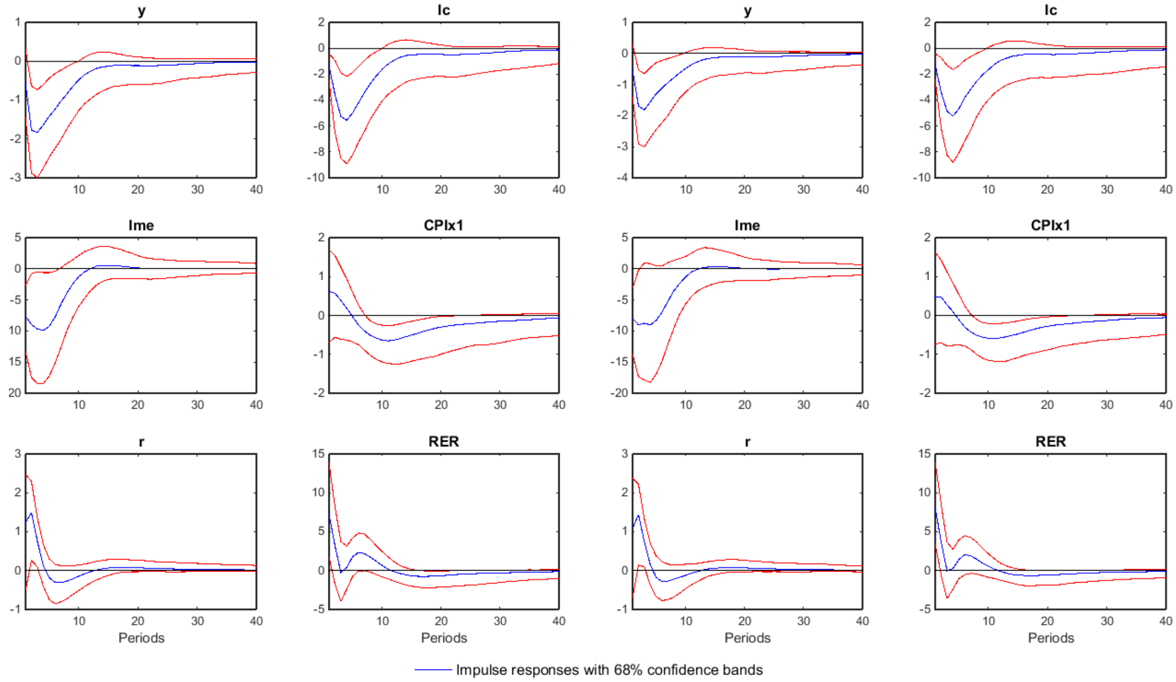
Recursive VAR first two column; "agnostic" VAR last two column for the base line model. The figure shows the quarterly responses to a 1% positive shock to the foreign monetary policy rate at the monthly frequency. The quarterly responses were computed by aggregating the monthly responses of the model.

Figure B.3.2
Impulse responses for alternative "agnostic" VAR models
for the domestic block to a foreign monetary policy shock



Responses for the alternative restrictions over investment for "agnostic" VAR models: (1) Mod A: negative sign restrictions only last one month (left panel); (2) Mod B: negative sign restrictions last for three months (right panel). The figure shows the quarterly responses to a 1% positive shock to the foreign monetary policy rate at the monthly frequency. The quarterly responses were computed by aggregating the monthly responses of the model. The responses for the foreign blocks do not change in these two cases and thus they are not reported.

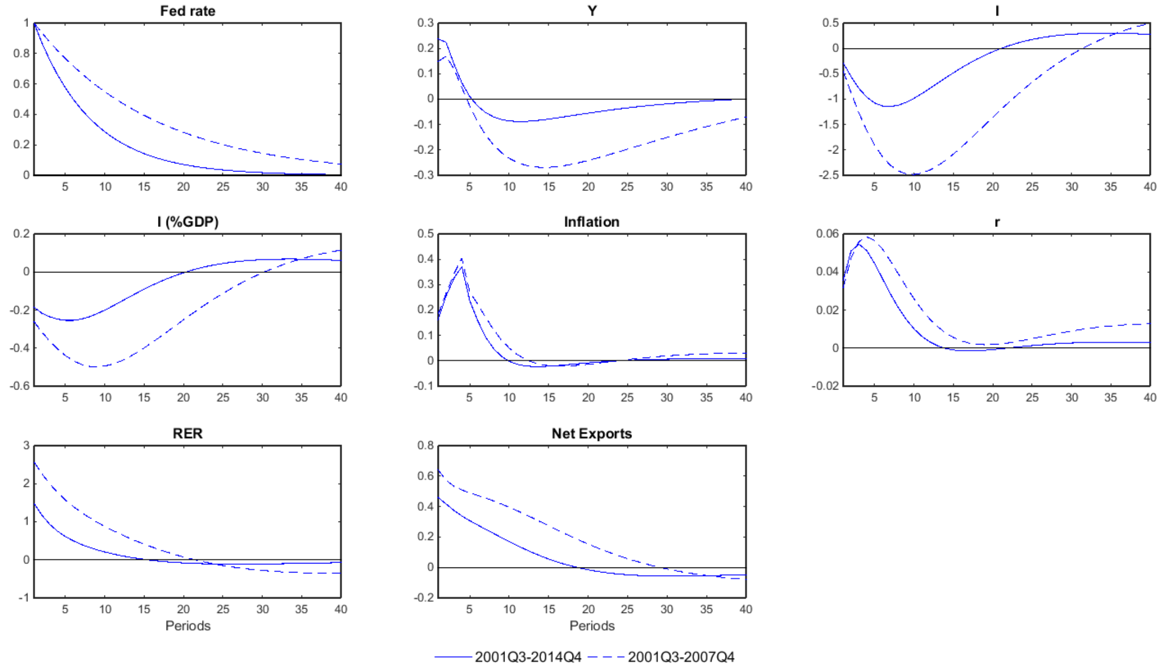
Figure B.3.3
Impulse responses alternative "agnostic" VAR models
for the domestic block to a foreign monetary policy shock



Responses for the alternative "agnostic" VAR models: (1) Mod C: foreign monetary policy is positive for at least 3 months (left panel); Mod D: foreign monetary policy and real exchange rate are positive for at least 3 months (right panel). Thus, these two alternative "agnostic" schemes are incremental cases of the base model. The figure shows the quarterly responses to a 1% positive shock to the foreign monetary policy rate at the monthly frequency. The quarterly responses were computed by aggregating the monthly responses of the model. The responses for the foreign blocks are the same as those in the base model and thus they are not reported.

Figure B.3.4

Impulse responses with a DSGE model for Chile to a foreign monetary policy shock



Model is parameterized using estimates from Bayesian estimation techniques with quarterly data covering the period 2001Q3-2014Q4 and 2001Q3-2007Q4. The figure shows the bayesian impulse responses to a 1% positive shock (100 basis points) to the foreign interest rate. We assume that the dynamics of the foreign variables are described by independent autoregressive processes of order one, AR(1), as in Medina and Soto (2007a) and Fornero and Kirchner (2014).

References

[1] Altig, D., Christiano, L., Eichenbaum, M. and J. Lindé (2003). The Role of Monetary Policy in the Propagation of Technology Shocks. manuscript, Northwestern University.

[2] Arias, J., Rubio-Ramírez, J. and Waggoner, D. (2014). Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications. International Finance Discussion Papers 1100, Board of Governors of the Federal Reserve System (U.S.).

[3] Armah, N, Swanson, N. (2011). Some variables are more worthy than others: new diffusion index evidence on the monitoring of key economic indicator. Applied Financial Economics 21, pages 43-60.

[4] Bai, J. and Ng, S. (2006a). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Econometrica 74.

[5] Bates J. and Granger, CW. J. (1969). The Combination of Forecasts. Operational Reserach Quaterly. Vol. 20, No. 4 (Dec., 1969), pages 451-468.

[6] Baumeister, C. and Benati, L. (2013). Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound. International Journal of Central Banking, vol. 9(2), pages 165-212, June.

[7] Bernanke, B., Boivin, J. and Eliasz, P. (2005). Measuring Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach. Quarterly Journal of Economics, 120(1), 387-422.

[8] Bernanke, B., Gertler M. and Gilchrist, S. (1999). The Financial Accelerator in a Quantitative Business Cycle Framework. In Handbook of Macroeconomics, edited by M. Woodford and J. Taylor, North Holland.

[9] Blanchard, O. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. American Economic Review, American Economic Association, vol. 79(4), pages 655-73, September.

[10] Bredahl A and Teräsvirta, T. (2010). Forecasting with nonlinear time series models. CREATES Research Papers 2010-01, School of Economics and Management, University of Aarhus.

[11] Boero, G. and Marrocu, E. (2004). The performance of SETAR models: A regime conditional evaluation of point, interval and density forecasts. International Journal of Forecasting, 20, pages 305-302.

[12] Calvo, G. (1983). Staggered prices in utility-maximizing framework. Journal of Monetary Economics 12: 383-98.

[13] Canova F. (2007). Methods for Applied Macroeconomic Research. Princeton: Princeton University Press.

[14] Canova, F. and Nicolo, G. (2002). Monetary disturbances matter for business fluctuations in the G-7. Journal of Monetary Economics, Elsevier, vol. 49(6), pages 1131-1159, September.

[15] Carlstrom, C. and T. Fuerst (1997). Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis. American Economic Review 87: 893-910.

[16] Carrera, C., Pérez Forero F. and Ramŕez-Rondń, N. (2015). Effects of the U.S. Quantitative Easing on Latin American Economies. Working Papers 2015-35, Peruvian Economic Association.

[17] Castelnuovo, E. (2012). Monetary Policy Neutrality: Sign Restrictions Go to Monte Carlo. "Marco Fanno" Working Papers 0151, Dipartimento di Scienze Economiche "Marco Fanno".

[18] Castelnuovo, E. (2013). What does a Monetary Policy Shock Do? An International Analysis with Multiple Filters. Oxford Bulletin of Economics and Statistics, volume 75(5), pages 759-784. Octuber.

[19] Castelnuovo, E. (2015). Monetary Policy Shocks and Cholesky-VARs: An Assessment for the Euro Area. Empirical Economics (forthcoming).

[20] Christensen, J. and Rudebusch, G. (2012). The Response of Interest Rates to US and UK Quantitative Easing. Economic Journal, Royal Economic Society, vol. 122(564), pages F385-F414, November.

[21] Christiano, L., Eichenbaum, M. and Evans, C. (1999). Monetary policy shocks: What have we learned and to what end? Handbook of Macroeconomics, in: J. B. Taylor & M. Woodford (ed.), Handbook of Macroeconomics, edition 1, volume 1, chapter 2, pages 65-148 Elsevier.

[22] Christiano, L., Eichenbaum, M. and Evans, C. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. Journal of Political Economy 113(1): 1-45.

[23] Christoffersen, P. and Diebold, F. (1997). Optimal Prediction Under Asymmetric Loss. Econometric Theory, Cambridge University Press, vol. 13(06), pages 808-817, December.

[24] Clark, T and McCracken, M. (2001). Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics, Elsevier, vol. 105(1), pages 85-110, November.

[25] Clark, T and McCracken, M. (2012). Reality Checks and Comparisons of Nested Predictive Models. Journal of Business and Economic Statistics Volume 30, Issue 1.

[26] Clark, T. and West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, Elsevier, vol. 138(1), pages 291-311, May.

[27] Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5, pages 559583.

[28] Clements, M., Franses, P. and Swanson, N. (2004). Forecasting economic and financial time-series with non-linear models. International Journal of Forecasting, Elsevier, vol. 20(2), pages 169-183.

[29] Clements, M. and Hendry F. (2004). Pooling of forecasts. Econometrics Journal, Royal Economic Society, vol. 7(1), pages 1-31, 06.

[30] Clements, M. and Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. International Journal of Forecasting Volume 13, Issue 4, December 1997, pages 463-475.

[31] Cybenko, G. (1989). An Approximations by superpositions of sigmoidal functions. Mathematics of Control, Signals, and Systems, 2 (4), pages 303-314

[32] DeJong, D., B. Ingram, and Whiteman C. (2000). A Bayesian Approach to Dynamic Macroeconomics. Journal of Econometrics 98(2): 203-23.

[33] Deutsch, M., Granger, CW. J. and Teräsvirta, T. (1994). The combination of forecasts using changing weights. International Journal of Forecasting, Elsevier, vol. 10(1), pages 47-57, June.

[34] De Gooijer, J., Kumar, K. (1992). Some recent developments in non-linear time series modelling, testing, and forecasting. International Journal of Forecasting, 8, pages 135-156.

[35] Diebold, F. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. International Journal of Forecasting, Elsevier, vol. 5(4), pages 589-592.

[36] Diebold, F. (2013). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. MIMEO, University of Pennsylvania.

[37] Diebold, F. and Lopez, J. (1996). Forecast evaluation and combination. In G.S. Maddala and C.R. Rao, eds., Handbook of Statistics. Amsterdam: North Holland.

[38] Diebold, F. and Mariano, R. (1995). Comparing Predictive Accuracy, Journal of Business and Economic Statistics, 13, pages 253-263.

[39] Diebold, F. and Nason, J. (1990). Nonparametric exchange rate prediction. Journal of International Economics, 28, pages 315-332.

[40] Diebold, F. and Peter, P. (1990). The use of prior information in forecast combination. International Journal of Forecasting, Elsevier, vol. 6(4), pages 503-508, December.

[41] Elliott, G. and Timmermann, A. (2005). Optimal Forecast Combination Under Regime Switching. International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 46(4), pages 1081-1102, November.

[42] Faust, J.(1998). The robustness of identified VAR conclusions about money. International Finance Discussion Papers 610, Board of Governors of the Federal Reserve System (U.S.).

[43] Fransesa, P. (2011). Model selection for forecast combination. Journal of Applied Economics. Volume 43, Issue 14, 2011

[44] Fornero, J. and Kirchner, M. (2014). Learning About Commodity Cycles and Saving-Investment Dynamics in a Commodity-Exporting Economy. Working Paper, Central Bank of Chile. May.

[45] Fornero, J., M. Kirchner, and A. Yany (2015): ”Terms of Trade Shocks and Investment in Commodity-Exporting Economies,” in: Commodity Prices and Macroeconomic Policy, R. Caputo, R. Chang, and D. Saravia (eds.), vol. 21 of Central Banking, Analysis, and Economic Policies, Central Bank of Chile.

[46] Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 25(4), pages 595-620.

[47] Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. Econometrica, 74(6),pages 1545-1578.

[48] Granger, CW. J. and Ramanathan, R. (1989). Improved methods of combining forecasts. Journal of Forecasting Volume 3, Issue 2, pages 197-204, April/June 1984.

[49] Granger, CW. J. and Teräsvirta, T. (1993). Modelling Nonlinear Economic Relationships. Oxford: Oxford University Press.

[50] Geweke, J. (2005). Contemporary Bayesian Econometrics and Statistics. Hoboken, NJ: John Wiley & Sons.

[51] Hamilton, J. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.

[52] Hamilton, J. (1994). Time Series Analysis. Princeton University Press, New Jersey.

[53] Hansen, B. (1999). Testing for Linearity. Journal of Economic Surveys, Wiley Blackwell, vol. 13(5), pages 551-576, December.

[54] Hansen, B. (2000). Sample Splitting and Threshold Estimation. Econometrica, Econometric Society, vol. 68(3), pages 575-604, May.

[55] Hendry D. and Clements, M. (2004). Pooling of forecasts. Econometrics Journal, Royal Economic Society, vol. 7(1), pages 1-31, 06.

[56] Kapetanios, G., Mumtaz, H., Stevens, I. and Theodoridis, K. (2012). Assessing the Economy-wide Effects of Quantitative Easing. The Economic Journal, 122: F316F347.

[57] Kilian, L. (2013). Structural vector autoregressions. Chapters, in: Handbook of Research Methods and Applications in Empirical Macroeconomics, chapter 22, pages 515-554 Edward Elgar Publishing.

[58] Kim, H. and Swanson, N. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. Journal of Econometrics, Elsevier, vol. 178(P2), pages 352-367.

[59] Koop, G. and Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. Manuscript. University of Strathclyde

[60] Kuan, C.and White, H. (1994). Artificial Neural Networks: An Econometric Perspective. Econometric Reviews 13.

[61] Kydland, F. and Prescott, E. (1982). Time to Build and Aggregate Fluctuations. Econometrica, colume 50(6), pages 1345-70. November.

[62] Lee, T.H., White, H. and Granger, CW. J. (1992). Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Models and Standard Tests. Journal of Econometrics 56: pages 269-290.

[63] Lütkepohl, H. (2011). Vector Autoregresssive Models. Handbook of Research Methods and Applications in Empirical Macroeconomics.

[64] Marcellino, M., (2008). A linear benchmark for forecasting GDP growth and inflation? Journal of Forecasting Volume 27, Issue 4, pages 305-340.

[65] Marcellino, M., Stock, J. and Watson, M. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics, Elsevier, vol. 135(1-2), pages 499-526.

[66] McNelis, P. (2005). Neural Networks in Finance: Gaining Predictive Edge in the Market. Academic Press Advanced Finance.

[67] Medina J.P. and Soto, C. (2005). Proyecto MAS: Modelo para Analisis y Simulaciones. Central Bank of Chile.

[68] Medina J.P. and Soto, C. (2007a). Copper Price, Fiscal Policy and Business Cycle in Chile. Working Paper, Central Bank of Chile. December.

[69] Medina J.P. and Soto, C. (2007b). The Chilean Business Cycles Through the Lens of a Stochastic General Equilibrium Model. Working Paper, Central Bank of Chile. December.

[70] Mojon, B. (2008). When did unsystematic monetary policy have an effect on inflation? European Economic Review, volume 52(3), pages 487-497, April

[71] Mountford, A. and Uhlig, H. (2009). What are the effects of fiscal policy shocks? Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 24(6), pages 960-992.

[72] Parrado, E (2001). Foreign Shocks and Monetary Policy Transmission in Chile. Journal Economa Chilena (The Chilean Economy), Central Bank of Chile, vol. 4(3), pages 29-57, December.

[73] Ramsey, J. (1969). Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. Journal of the Royal Statistical Society Series B 31 (2): pages 350-371.

[74] Rapach, D., Strauss, J and Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. The Review of Financial Studies, Vol. 23, No.2, pp. 821-862

[75] Rossi, B. and Sekhposyan, T. (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when?. International Journal of Forecasting, Elsevier, vol. 26(4), pages 808-835, October.

[76] Rubio-Ramŕez, R., Waggoner D. and Zha, T.(2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. Review of Economic Studies, Oxford University Press, vol. 77(2), pages 665-696.

[77] Sims, C. (1980a). Macroeconomics and Reality. Econometrica, Econometric Society, vol. 48(1), pages 1-48, January.

[78] Sims, C. (1986). Are Forecasting Models Usable for Policy Analysis? Federal Reserve Bank of Minneapolis, Quarterly Review, 10.

[79] Sims, C, Stock, J. and Watson, M (1990). Inference in Linear Time Series Models with Some Unit Roots. Econometrica, Econometric Society, vol. 58(1), pages 113-44, January.

[80] Sims, C. and Zha, T. (1999). Error Bands for Impulse Responses. Econometrica, 67 (5), 1113-1155.

[81] Smets, F. and Wouters R. (2003). An Estimated Stochastic Dynamic General Equlibrium Model of the Euro Area. Journal of the European Ecoomic Association 1: 1123-75.

[82] Smets, F. and Wouters R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. CEPR Discussion Papers 6112.

[83] Stock, J. and Watson, M. (1999). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In: Engle R, White H Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger. Oxford: Oxford University Press; 1999. pages 1-44.

[84] Stock, J. and Watson, M. (2002a). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97.

[85] Stock, J. and Watson, M. (2002b). Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics.

[86] Stock, J. and Watson, M. (2003). Forecasting output and inflation: The role of asset prices. Journal of Economic Literature, 41(3), pages 788-829.

[87] Stock, J. and Watson, M. (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set. J. Forecast, 23.

[88] Stock, J. and Watson, M. (2007). Why Has U.S. Inflation Become Harder to Forecast?. Journal of Money, Credit and Banking, Blackwell Publishing, vol. 39(s1), pages 3-33.

[89] Stock, J. and Watson, M. (2010). Modeling Inflation After the Crisis. NBER Working Papers 16488, National Bureau of Economic Research, Inc.

[90] Swanson N. and White, H. (1995). A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks. Journal of Business and Economic Statistics. Vol. 13, No. 3, pages 265-275.

[91] Swanson N. and White, H. (1997a). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. International Journal of Forecasting. Volume 13, Issue 4, pages 439-461.

[92] Swanson N. and White, H. (1997b). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. Review of Economics and Statistics 79.

[93] Swanson, N. and Zeng, T. (2001). Choosing among competing econometric forecasts: Regression-based forecast combination using model selection. J. Forecast, Issue 20.

[94] Teräsvirta, T. (1998). Modeling economic relationships with smooth transition regressions. In: Ullah, A., Giles, D.E. (Eds.), Handbook of Applied Economic Statistics. Dekker, New York, pages 507-552.

[95] Teräsvirta, T., Dijk, D. and Medeiros, C. (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. International Journal of Forecasting, Elsevier, vol. 21(4), pages 755-774.

[96] Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. Handbook of Economic Forecasting. Volume 1, 2006, Pages 413-457.

[97] Timmermann, A. (2006). Forecast Combinations. Handbook of Economic Forecasting, Elsevier.

[98] Tong, H. (1978). On a Threshold Model, in C.H. Chen (ed.), Pattern Recognition and Signal Processing. Amsterdam: Sijhoff & Noordhoff.

[99] Tong, H. (1983). Threshold Models in Non-Linear Time Series Analysis. Springer-Verlag, New York.

[100] Tong, H. (1990). Non-linear Time Series: A Dynamical System Approach. Oxford: Clarendon Press.

[101] Tong, H. and Lim, K. S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion). Journal of the Royal Statistical Society, Series B, 42, pages 245-292.

[102] Tsay, R. (2005). Analysis of Financial Time Series. Second Edition, Wiley series in probability and statistics.

[103] Uhlig, H. (1997). Bayesian Vector Autoregressions with Stochastic Volatility. Econometrica, volume 65(1). January.

[104] Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics, Elsevier, vol. 52(2), pages 381-419, March.

[105] Uribe, M and Yue, V. (2006). Country spreads and emerging countries: Who drives whom? Journal of International Economics, volume 69(1), pages 6-36. June

[106] Weber, A., Gerke, R. and Worms, A. (2009). Has the monetary transmission process in the Euro area changed? Evidence based on VAR estimates,BIS Working Paper No. 276.

[107] White, H. (1980). A heterocedasticity-consistent covariance matrix estimator and a direct test for heterocedasticity. Econometrica, 48, pages 817-838.

[108] Zha, T. (1999). Block Recursion and Structural Vector Autoregressions. Journal of Econometrics 90, 291-316.