

The Vector Floor and Ceiling Model*

GARY KOOP

*Department of Economics, University of Edinburgh,
Edinburgh, EH8 9JY, U.K.
E-mail: G.Koop@ed.ac.uk*

SIMON POTTER

*Federal Reserve Bank of New York,
New York, NY 10045-0001, U.S.A.
E-mail: Simon.Potter@ny.frb.org*

January 2000

ABSTRACT: This paper motivates and develops a nonlinear extension of the Vector Autoregressive model which we call the Vector Floor and Ceiling model. Bayesian and classical methods for estimation and testing are developed and compared in the context of an application involving U.S. macroeconomic data. In terms of statistical significance both classical and Bayesian methods indicate that the (Gaussian) linear model is inadequate. Using impulse response functions we investigate the economic significance of the statistical analysis. We find evidence of strong nonlinearities in the contemporaneous relationships between the variables and milder evidence of nonlinearity in the conditional mean.

Keywords: Nonlinearity, Bayesian, Vector Autoregression.

JEL: C32, C52, E30

1 Introduction

Much work in empirical macroeconomics over the last few decades has involved the use of Vector Autoregressive (VAR) models (see, e.g., Sims, 1972 or 1980). VARs have many advantages in that they are simple to work with and their properties are well-understood. However, they have one major disadvantage: they are linear. A plethora of work involving univariate time series models has provided statistical evidence for nonlinearity in many economic time series (see, among many others, Beaudry and Koop, 1993, Hamilton, 1989, Pesaran and Potter, 1997, Koop and Potter, 1999b, Skalin and Teräsvirta, 1999). Furthermore, many theoretical models of the macroeconomy imply nonlinearity. For these reasons, there is an increasing interest in nonlinear extensions to VAR models. In this paper, we propose one such extension: the Vector Floor and Ceiling (VFC) model (see Altissimo and Violante 1999 for a similar extension and Potter 1995b for the general case). The VFC model is a parsimonious extension of the VAR model which, we argue, should be able to capture the types of nonlinearity suggested by economic theory. Estimation and testing in this model raise some interesting issues for either the Bayesian or non-Bayesian econometrician. Accordingly, we develop and compare both econometric approaches in the context of an application involving the commonly-used RMPY variables.¹

In order to carry out an empirical analysis using a nonlinear VAR, two preliminary hurdles must be passed. First, unlike VARs and their relationship to the Wold Representation, there is no general nonlinear model that can be appealed to. The Vector Floor and Ceiling model is a particular choice which, we argue, is an attractive one. It is a multivariate extension of the model used in Pesaran and Potter (1997) which allows nonlinearity to enter through floor and ceiling effects. The intuition for introducing nonlinearity in this way arises from the long tradition in economics that fluctuations may arise from reflective

¹The RMPY variables are R=the interest rate, M=the money supply, P=the price level and Y=real output. The data is described below.

barriers (e.g. floors and ceilings). Furthermore, the VFC model allows nonlinearity to enter in a particularly parsimonious manner. Parsimony is crucial since even with linear VARs it is relatively easy to run into over-parameterization problems. The VFC model is a tightly restricted version of the Threshold Autoregression (TAR) class introduced by Tong (see Tong 1990 for an overview).

Second, econometric methods must be developed to estimate and test the VFC model. It is particularly important to provide evidence that the nonlinear model is statistically superior to the linear model. In this paper, we use both Bayesian and classical methods. Estimation, using either paradigm, is relatively straightforward using iterated generalized least squares² (for classical analysis) or a Markov Chain Monte Carlo (MCMC) algorithm (for Bayesian analysis). However, testing is plagued by the presence of Davies' problem — nuisance parameters which are unidentified under the null hypothesis of linearity. In this paper, we show how the classical test procedures of Hansen (1996) and Pesaran and Potter (1997) can be applied in the vector time series case. In previous univariate time series work (Koop and Potter, 1999a), we have investigated the properties of Bayes factors in the presence of Davies' problem. We extend this work in the present paper and derive a particularly computationally efficient method for Bayes factor calculation.

The methods described above are applied to a system containing Y =GDP in 1987 dollars, P =the GDP deflator, R =3month T-bill and M =M2 observed at the quarterly frequency in the U.S. in the post-Korean war period (1954Q3 through 1997Q4). In terms of statistical significance both classical and Bayesian methods indicate that the (Gaussian) linear model is inadequate. Using impulse response functions we investigate the economic significance of the statistical analysis. We find evidence of strong nonlinearities in the contemporaneous relationships between the variables and milder evidence of nonlinearity in the conditional mean.

²Iterated generalized least squares is asymptotically equivalent to maximum likelihood estimation in the models we discuss and, hence, we use both terms in this paper.

The structure of the paper is as follows: section 2 develops the vector floor and ceiling model; section 3 describes the results and discusses some of the differences between the classical and Bayesian analysis. Much of the technical detail can be found in 4 appendices. Appendix A outlines the calculation of various least squares quantities. Appendix B gives a detailed development of the Bayesian analysis. Appendix C describes classical estimation and testing. Appendix D provides information on the calculation of the generalized impulse response functions.

2 A NONLINEAR VAR WITH FLOOR AND CEILING EFFECTS

In this section we develop and motivate a parsimonious nonlinear VAR which we call the Vector Floor and Ceiling model. Variants on this model should, we feel, be suitable for use in many macroeconomic applications. The univariate version of this model was first introduced in Pesaran and Potter (1997).³

There are many reasons for thinking that linear models might be too restrictive when working with economic time series (see, among many others, Beaudry and Koop, 1993, Koop, 1996 or Koop, Pesaran and Potter, 1996). For instance, in a macroeconomic context, linear models imply that positive and negative monetary shocks have the same effect on output (in absolute value). Furthermore, a monetary shock of a given magnitude will have the same effect regardless of whether it occurs in a recession or in an expansion. However, it is easy to say linear models are too restrictive, but it is much harder to choose a particular nonlinear specification from the myriad of possibilities.

Economic theory offers only limited guidance in this regard. However, a common thread underlying a great deal of macroeconomic theory is that dynamics should vary

³See also Koop and Potter (1997) for a Bayesian analysis of a generalization of the univariate floor and ceiling model.

over the business cycle (see Kapetanios, 1998, for a general discussion). Nonlinear time series econometricians have typically interpreted this as implying different regimes exist. The two most commonly-used classes of models have this property. Threshold autoregressive (TAR) models divide the data into different regimes based on lagged values of the time series (see Tong, 1990 or Teräsvirta, 1994 for a popular variant called the smooth transition threshold autoregressive model or STAR). Markov switching models assume the time series switches between regimes with transition probability following a Markov chain (Hamilton, 1989). Most empirical work has been univariate (exceptions are Chauvet 1998, Weise 1999 and Altissimo and Violante 1999).

Even in univariate contexts, nonlinear time series models have been criticized for their lack of parsimony. For instance, the simplest TAR allows for a different AR(p) model to exist in each different regime. If the number of regimes and/or lags is at all large, over-parameterization can be a worry. Furthermore, the number of regimes is rarely specified by economic theory. Consider the example of real GDP dynamics. One might suspect different regimes to apply in the case where the economy is in a recession, expansion or normal times. This suggests three regimes. However, dynamics might differ within these regimes (e.g. dynamics might change if the economy starts to overheat, indicating that a different specification might be appropriate early in an expansion than that which applies late in an expansion). All in all, a case can be made that many different regimes should exist, exacerbating parsimony problems.

With multivariate models, these over-parameterization worries are greatly increased. It becomes essential to work with a multivariate nonlinear time series model which both allows for many regimes and is not over-parameterized. The Vector Floor and Ceiling model attempts to do this by using the intuition that, although many different regimes might exist, dynamics probably only vary slightly over similar regimes. For example, if the economy has been in a recession for one period, dynamics are probably only slightly different from the case where the economy has been in a recession for two periods. The

key properties of the VFC model are:

- It is a nonlinear extension of a VAR.
- It contains three major regimes which we label “Floor”, “Ceiling” and “Corridor” which can be thought of as relating to recessions, expansions and normal times.
- Within the floor and ceiling regimes there are various sub-regimes.
- Dynamics in the sub-regimes are modelled through variables which can be thought of as reflecting the current depth of recession and the amount of overheating in the economy.

These general considerations are formalized and expanded on in the context of our empirical application in the remainder of this section.⁴ We begin by defining the notation used throughout the paper. \mathbf{X}_t will be a 4×1 vector of observations, with X_{1t} being the log level of output, X_{2t} be the log of prices, X_{3t} the interest rate, X_{4t} the log level of money. \mathbf{Y}_t will be the vector of first differences of \mathbf{X}_t .⁵ $\mathbf{1}(A)$ is the indicator function equal to 1 if the event A occurs, 0 otherwise. \mathbf{V}_t will be a 4×1 vector of i.i.d. multivariate standard Normal random variables.

The previous VAR literature has not produced any consensus on the important issue of how to treat the trending behavior that appears to exist in several of the variables. Since there is only weak and conflicting evidence of a cointegrating relationship, we estimate the system in differences. This has the advantage of allowing us to ignore the behavior of the test statistics for nonlinearity under non-stationarity. The disadvantage is that the long run dynamics of the system might be distorted.

⁴In this paper, we work with a particular variant of the VFC model which is appropriate for the application at hand. The basic econometric techniques developed in this paper are, of course, applicable to other variants which may be relevant in other empirical problems.

⁵We work with log differences multiplied by 100 for all variables except the interest rate. We take raw first differences for the latter, since it is already a percentage.

We begin by defining the three major regimes which we assume are based on the state of the business cycle as reflected in output growth (i.e. Y_{1t}). The model contains three indices defining the three possible regimes of the economy. F_t is the index representing the floor regime. It is activated if output falls below the previous maximum minus a parameter, $r_F < 0$, to be estimated. C_t is the index representing the ceiling regime. This regime is activated when output growth has been ‘high’ for two consecutive quarters and the floor regime is not active. Here the parameter to be estimated will be $r_C > 0$, the value of ‘high’ growth.

These two indices are then used to recursively cumulate output growth in the respective regimes. The floor index produces a variable measuring the Current Depth of Recession (CDR_{1t}) as in Beaudry and Koop (1993).⁶ The ceiling index produces a variable measuring the amount of overheating, (OH_{1t}), in the economy. The exact form of the relationships are:

$$F_t = \begin{cases} F_t = \mathbf{1}(Y_{1t} < r_F), & \text{if } F_{t-1} = 0, \\ F_t = \mathbf{1}(CDR_{1t-1} + Y_{1t} < 0), & \text{if } F_{t-1} = 1, \end{cases} \quad (1)$$

$$CDR_{1t} = \begin{cases} (Y_{1t} - r_F) F_t & \text{if } F_{t-1} = 0, \\ (CDR_{1t-1} + Y_{1t}) F_t & \text{if } F_{t-1} = 1. \end{cases} \quad (2)$$

$$C_t = \mathbf{1}(F_t = 0)\mathbf{1}(Y_{1t} > r_C)\mathbf{1}(Y_{1t-1} > r_C), \quad (3)$$

$$OH_{1t} = (OH_{1t-1} + Y_{1t} - r_C) C_t. \quad (4)$$

The third index defines the corridor regime which occurs if neither of the other two regimes are activated (i.e. $COR_t = \mathbf{1}(F_t + C_t = 0)$).

The floor regime and current depth of recession variable are most easily understood if $r_F = 0$. In this case, the floor regime is activated when GDP falls (i.e. a recession begins) and remains activated until GDP has grown back to its pre-recession level. The current

⁶Beaudry and Koop (1993) assumes $r_F = 0$.

depth of recession variable is then a measure of how deep the recession is (i.e. a measure of how much GDP has fallen from its pre-recession level). Note that the ceiling regime is only activated if the floor index is not. This rules out the possibility that recovery from the trough of a recession is labelled overheating. Furthermore, the requirement that the ceiling regime is only activated by two consecutive quarters of fast growth follows from the reasonable notion that a single quarter of fast growth is unlikely to overheat the economy. The overheating variable itself can be interpreted as the reverse of the current depth of recession.

In the model of Pesaran and Potter (1997), the variables CDR_{1t} , OH_{1t} are lagged and entered into a standard univariate nonlinear model for output with the exception that the error variance is allowed to vary across the three regimes. That is, their univariate model is:

$$Y_{1t} = \alpha + \phi_p(L)Y_{1t-1} + \theta_1 CDR_{1t-1} + \theta_2 OH_{1t-1} + \{\sigma_0 COR_{t-1} + \sigma_1 F_{t-1} + \sigma_2 C_{t-1}\} V_t,$$

where $\phi_p(L)$ is a polynomial in the lag operator of order p .

One can extend this model very parsimoniously to the multiple time series case by using the indicator variables F_t and C_t defined from the behavior of output alone. Additional recursively-defined variables similar to CDR_{1t} , OH_{1t} can then be used to allow for over-heating or “under-cooling” effects in the other variables.⁷ In particular, we construct:

$$CDR_{it} = F_t (CDR_{it-1} + Y_{it}),$$

⁷In more general applications it might make sense to consider regimes defined by other variables or by combinations of variables, but for the present case we consider only the simplest extension to the multivariate case. In other words, we use only GDP information to define the state of the business cycle. The case where the growth rates of the other variables are recursively cumulated is considered here but there are many other possibilities. For example one might want to cumulate a linear relationship suggested by economic theory between variables in the VFC model.

$$OH_{it} = C_t \{OH_{it-1} + Y_{it}\}, \text{ for } i = 2, 3, 4.$$

Note that these new variables do not depend on the thresholds r_F and r_C . This makes them easy to interpret. For instance, CDR_{3t} measures the deviation of the interest rate from that which occurred when a recession began. In recessions, interest rates tend to fall; CDR_{3t} is a measure of the magnitude of this fall. Similarly, OH_{2t} is a measure of the deviation of the price level from that which occurred at the beginning of a period of overheating. Hence, it measures how bad inflationary pressures were during expansionary periods. Analogous interpretations hold for the other variables.

The four current depth of recession and overheating variables are entered into a standard VAR framework, with the exception that the error covariance matrix is allowed to vary across the regimes:

$$\mathbf{Y}_t = \alpha + \Phi_p(L)\mathbf{Y}_{t-1} + \Theta_1 \mathbf{CDR}_{t-1} + \Theta_2 \mathbf{OH}_{t-1} + \{\mathbf{H}_0 \mathbf{COR}_{t-1} + \mathbf{H}_1 F_{t-1} + \mathbf{H}_2 C_{t-1}\} \mathbf{V}_t \quad (5)$$

where α is a $k \times 1$ vector, $\Phi_p(L)$ is a p^{th} order matrix polynomial in the lag operator, Θ_1, Θ_2 are $k \times k$ matrices, $\mathbf{H}_0, \mathbf{H}_1, \mathbf{H}_2$ are $k \times k$ matrices with a lower triangular structure (i.e. $\Sigma_i = H_i H_i'$ is the error covariance in regime i , $i=0,1,2$) and

$$\mathbf{CDR}_{t-1} = (CDR_{1t-1}, CDR_{2t-1}, \dots, CDR_{4t-1}), \mathbf{OH}_{t-1} = (OH_{1t-1}, OH_{2t-1}, \dots, OH_{4t-1}),$$

are $K \times 1$ vectors. In our empirical work, $K = 4$.

It is possible to unravel the vector floor and ceiling model in an illuminating manner:

$$\mathbf{Y}_t = \begin{cases} \alpha + \Phi_p(L)\mathbf{Y}_{t-1} + \mathbf{H}_0\mathbf{V}_t & \text{if } COR_{t-1} = 1, \\ \alpha + \Phi_p(L)\mathbf{Y}_{t-1} + \Theta_1(\mathbf{Y}_{t-1} - r_F) + \mathbf{H}_1\mathbf{V}_t & \text{if } F_{t-1} = 1, F_{t-2} = 0 \\ \vdots & \vdots \\ \alpha + \Phi_p(L)\mathbf{Y}_{t-1} + \Theta_1 \sum_{s=1}^j (\mathbf{Y}_{t-s} - r_F) + \mathbf{H}_1\mathbf{V}_t & \text{if } \prod_{s=1}^j F_{t-s} = 1 \text{ and } F_{t-j-1} = 0. \\ \vdots & \vdots \\ \alpha + \Phi_p(L)\mathbf{Y}_{t-1} + \Theta_2(\mathbf{Y}_{t-1} - r_C) + \mathbf{H}_2\mathbf{V}_t & \text{if } C_{t-1} = 1 \text{ and } C_{t-2} = 0, \\ \vdots & \vdots \\ \alpha + \Phi_p(L)\mathbf{Y}_{t-1} + \Theta_2 \sum_{s=1}^j (\mathbf{Y}_{t-s} - r_C) + \mathbf{H}_2\mathbf{V}_t & \text{if } \prod_{s=1}^j C_{t-s} = 1 \text{ and } C_{t-j-1} = 0. \\ \vdots & \vdots \end{cases} \quad (6)$$

where $\mathbf{r}_m = (r_m, 0, \dots, 0)'$, $m = F, C$ are $k \times 1$ vectors and F_t, C_t are determined by Equations 1 and 3.

Equation 6 shows how the VFC model can be interpreted as having many different sub-regimes within each of the two outer regimes. For instance, a different dynamic specification exists when the economy has been in recession for one period ($F_{t-1} = 1, F_{t-2} = 0$) than when the economy has been in recession j periods ($\prod_{s=1}^j F_{t-s} = 1$ and $F_{t-j-1} = 0$). Furthermore, the specification is quite parsimonious in that differences between major regimes only depend on two $K \times K$ matrices, Θ_1 and Θ_2 . Differences between dynamics in the many subregimes also only depend of these two matrices. The conditional mean of the VFC model depends on $\alpha, \Phi_p, \Theta_1, \Theta_2, r_F$ and r_C , which contain $K^2(p+2) + K + 2$ distinct parameters. In contrast, a three regime multivariate TAR model which allowed for different VAR dynamics in each regime contains $3pK^2 + 3K + 2$ parameters. In the present application, $k = 4$. For the case where $p = 1$, the VFC model has 54 conditional mean parameters, whereas the TAR has 62. However, if $p = 4$ these numbers change to 102 and 206, respectively, indicating the strong parsimony of the VFC model relative to the TAR.

To provide even more intuition, let use write out the conditional mean for a sequence

of time periods where the economy is in the corridor regime in period $t-1$, then enters the floor regime in period t and remains there for several periods. We will assume $p=2$ and let Φ_i indicate the appropriate element of $\Phi_p(L)$.

- $\mathbf{Y}_t = \alpha + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2}$
- $\mathbf{Y}_{t+1} = (\alpha - \Theta_1 r_F) + (\Phi_1 + \Theta_1) \mathbf{Y}_t + \Phi_2 \mathbf{Y}_{t-1}$
- $\mathbf{Y}_{t+2} = (\alpha - \Theta_1 r_F) + (\Phi_1 + \Theta_1) \mathbf{Y}_{t+1} + (\Phi_2 + \Theta_1) \mathbf{Y}_t$
- $\mathbf{Y}_{t+3} = (\alpha - \Theta_1 r_F) + (\Phi_1 + \Theta_1) \mathbf{Y}_{t+2} + (\Phi_2 + \Theta_1) \mathbf{Y}_{t+1} + \Theta_1 \mathbf{Y}_t$
- $\mathbf{Y}_{t+4} = (\alpha - \Theta_1 r_F) + (\Phi_1 + \Theta_1) \mathbf{Y}_{t+3} + (\Phi_2 + \Theta_1) \mathbf{Y}_{t+2} + \Theta_1 \mathbf{Y}_{t+1} + \Theta_1 \mathbf{Y}_t$.

For this example, for the common case where Θ_1 is small, one can see how the original VAR(2) specification gradually changes to a VAR(2) with different intercept and first order coefficient matrix, then a VAR(2) with different second order coefficient matrix, then a VAR(3), then a VAR(4), etc. This illustrates how the model allows for many different subregimes, but the coefficients change only gradually across regimes. A similar illustration could be done when the economy starts to overheat.

In our empirical work, we also discuss three restricted versions of the VFC model. We will refer to the model with no heteroskedasticity, $\Sigma_0 = \Sigma_1 = \Sigma_2$, as the homoskedastic Vector Floor and Ceiling model (VFC-homo). The model with heteroskedasticity, but linearity in conditional mean, $\Theta_1 = \Theta_2 = \mathbf{0}_{K \times K}$, will be referred to as the heteroskedastic Vector Autoregressive model (VAR-hetero). The standard VAR has restrictions $\Theta_1 = \Theta_2 = \mathbf{0}_{K \times K}$ and $\Sigma_0 = \Sigma_1 = \Sigma_2$. Bayesian and classical estimation and testing in these models is discussed in Appendices A, B and C.

3 Empirical Results

This section contains Bayesian and non-Bayesian results on estimation, testing and impulse response analysis for the data set described above. In order to simplify discussion, we focus on the case where the lag length is equal to one (i.e. $p=1$). The value $p=1$ is chosen using Bayesian posterior odds analysis. Using a standard VAR, the Schwarz (SC) and Hannan-Quinn (HQ) information criteria both also choose $p=1$, although the Akaike information criteria (AIC) chooses $p=4$. If we search over all models (linear and nonlinear) and all lag lengths, we find $p=1$ chosen by SC, $p=2$ chosen by HQ and $p=4$ chosen by AIC. Except for AIC, we are thus finding that virtually all of the evidence indicates $p=1$. Monte Carlo results (e.g. Kapetanios 1998) suggest that the Akaike information criterion tends to overestimate lag length and perform poorly in nonlinear time series models. Furthermore, most of the results presented in this paper are qualitatively similar for $p=1, 2, 3$ and 4 (e.g. P-values for all tests roughly the same). For all these reasons, we feel focussing on the case $p=1$ for simplicity is warranted. We provide Bayesian results for two sets of priors: prior 1 assumes that all 32 coefficients in Θ_1 and Θ_2 are potentially non-zero, prior 2 assumes that only 8 out of 32 coefficients are likely to be non-zero. Further details on the prior are given below.

3.1 Model Comparison Results

Table 1 contains the results of classical tests of linearity. These tests are described in more detail in Appendix C. Suffice it to note here that the tests labelled “SUP WALD”, “EXP WALD” and “AVE WALD” are all based on Wald tests of the null hypothesis of linearity. However, under this hypothesis, the thresholds are not identified. The presence of such nuisance parameters which are not identified under the null leads to a violation of the regularity conditions necessary for deriving the standard Chi-squared asymptotic distribution for Wald statistics. Hence, we use the simulation method of Hansen (1996)

in order to calculate P-values. Following Andrews and Ploberger (1994), we take the average (AVE), exponential average (EXP) and supremum (SUP) of the Wald statistic over all threshold values in order to obtain the three test statistics. We also use a test for nonlinearity in the conditional mean of the series described in Pesaran and Potter (1997). This test, labelled PP below, is based on the observation that if we are only interested in testing $\Theta_1 = \Theta_2 = \mathbf{0}_{K \times K}$, then the local non-identification problem described above does not hold (i.e. the null is the VAR-hetero model and the thresholds still enter the error variance under the null hypothesis of linearity). Hence, a likelihood ratio test comparing the VFC to VAR-hetero models has a standard Chi-squared asymptotic distribution.

As described in Appendix C, classical estimation and testing involves carrying out a grid search over every possible floor and ceiling threshold combination. The floor threshold was allowed to vary from -0.492 to -0.0025 (45 grid points) and the ceiling threshold from 0.599 to 1.284 (61 grid points). The grid was chosen in a data-based fashion to ensure adequate degrees of freedom are available in each regime. These choices imply $45 \times 61 = 2745$ total points in the grid.

Table 1: Classical Linearity Test Results

Test	Statistic	P-value
SUP WALD	63	0.008
EXP WALD	28	0.010
AVE WALD	52	0.012
PP	37.645	0.226

With the exception of the Pesaran and Potter (1997) test, all of the classical tests indicate strongly significant nonlinearities. However, the non-significance of the PP test indicates that these nonlinearities are entering largely through the error variance. This picture is reinforced through an examination of Table 2, which contains various information criteria and Bayesian results for the different models. Kapetanios (1998) provides motivation for the use of information criteria for model selection involving nonlinear time

series models. Bayesian posterior model probabilities can be calculated from the Bayes factors comparing the VFC to VFC-homo models, the VFC to VAR-hetero models and the VAR-hetero to VAR model. These Bayes factors are calculated using the Savage-Dickey density ratio as described in Appendix B.

Table 2: Information Criteria and Bayesian Posterior Model Probabilities⁸

Model	AIC	HQ	SC	Post. Prob:Prior 1	Post. Prob:Prior 2.
VFC	-5.32	-4.93	-4.37	≈ 0	0.2
VFC-homo	-5.19	-4.81	-4.24	≈ 0	≈ 0
VAR-hetero	-5.47	-5.32	-5.10	1	0.8
VAR	-4.80	-4.65	-4.44	≈ 0	≈ 0

As discussed in Appendix C, we also calculated a minimum likelihood (i.e. we find minimum over the gridpoints). In log likelihood ratio units the difference was 72 between the VAR and VAR-hetero (a Chi-squared with 20 degrees of freedom would imply a p-value of approximately zero) and 106 between the VAR and VFC (a Chi-squared with 52 degrees of freedom would imply a p-value of approximately zero).

It can be seen that each of the information criteria select the VAR model with heteroskedasticity (i.e. floor and ceiling effects occur in the error covariance matrix). In each case, however, the second most preferred model is the unrestricted VFC, except for SC which chooses the linear model. The Bayesian posterior model probabilities also indicate strong support for the VAR-hetero model. Under prior 2 the unrestricted VFC also received substantial support, a point we will return to shortly. The extremely conservative minimum-likelihood ratio tests provide further evidence of nonlinearity.

Overall, there seems to be relatively little evidence of nonlinearity in the conditional mean, but overwhelming evidence of nonlinearity in the conditional error variance. Since the error variances control the immediate dynamics, we conclude that there is strong

⁸Prior 1 assumes that the elements of Θ_1, Θ_2 are Normal and independent of one another. Prior 2 assumes that the elements of Θ_1, Θ_2 are a mixture of Normals. Further details on the prior are given below and in Appendix B.

evidence of nonlinearities in the contemporaneous relationships between the variables. That is, the shocks, V_t have a contemporaneous effect on \mathbf{Y}_t and it is the magnitude of Σ_0 , Σ_1 and Σ_2 which control the how the shocks impact on the variables.

There are so many parameters in this model and they are difficult to interpret, so we do not provide a detailed discussion of them here. The dynamic properties of the VFC model are best understood through impulse response analysis, but it is worthwhile to illustrate the importance of the changes in the variance-covariance matrix across regimes by considering the mle estimates of individual variances (i.e. evaluated at the mle of the thresholds). Also presented are the MLE of the variances from the VFC-homo model and the linear VAR.

Table 3:MLE of shock variances

Model/Regime	Y	P	R	M
VFC/Cor	0.709	0.065	0.582	0.210
VFC/Ceil	0.552	0.037	0.088	0.205
VFC/Floor	1.038	0.105	1.230	0.437
VFC-homo	0.716	0.064	0.550	0.245
VAR	0.749	0.067	0.572	0.290

The MLEs of the thresholds associated with these estimates are $r_c = 0.732$, $r_F = -0.479$. At these values there are 85 observation in the corridor regime, 57 observations in the ceiling regime and 30 observations in the floor regime. The point estimates of error variances in the equations for R vary greatly across regimes. Given that the VFC is not a structural model, it is risky to give the error variances a structural interpretation. Keeping this warning in mind, note that interest rate shocks seem to be much bigger in magnitude in the floor regime than in the other two regimes. Interest rate shocks in the ceiling regime seem very small. A similar pattern is found in the money equation, where monetary shocks seem much bigger in the floor regime than in the others. One possible explanation is that the floor regime is picking up the behavior of interest rates in the

1979-1982 period. However, this does not explain the very low variance of interest rates in the ceiling regime.

3.2 A Comparison of Bayesian and Classical Results

With some exceptions the Bayesian and classical analyses yield similar results. That is, MLE's and standard errors are, in most cases, quite similar to posterior means and standard deviations evaluated at the mle estimates of the thresholds. The classical test procedures yield results which point in the same direction as posterior model probabilities. The level of computational difficulty of Bayesian and classical procedures are roughly comparable. There are a few things which the Bayesian approach can do that the classical one cannot (e.g. provide a framework for averaging across many models, yield exact finite sample results for a case where asymptotic approximations are likely to be poor, compare many models simultaneously,⁹ etc.), but so far the differences between sensible Bayesian and classical analyses seem small. However, there are some issues that are worth elaborating on. In order to do this, we must take a detour and present a discussion of Bayesian priors and hypothesis testing procedures.

Our Bayesian methods, including prior elicitation, are described in detail in Appendix B. Here we will provide a heuristic discussion of these methods. Bayesian methods provide the researcher with an intuitively simple measure of model performance: the posterior model probability.¹⁰ As the name suggests, this is just the probability that the model under consideration generated the data. However, one possible drawback in calculating posterior model probabilities is that informative priors are required if the models under consideration are nested ones.¹¹ Some intuition for why this occurs is provided by

⁹This Bayesian advantage is especially important in nonlinear time series models since there are so many model features to be tested (e.g. lag length, nonlinearity in mean, nonlinearity in error variance, etc.). Sequences of classical pairwise hypothesis tests will rapidly run into serious pre-test problems.

¹⁰The Bayes factor is the ratio of posterior model probabilities for two models under the assumption that, *a priori*, each of them is equally likely.

¹¹The fact that the use of noninformative priors can yield degenerate Bayes factors is often referred to

consideration of the AR(1) model:

$$y_t = \rho y_{t-1} + v_t,$$

where v_t is IIDN(0,1) and y_0 is known. Suppose interest centers on comparing the unrestricted model, M_1 , to the white noise model, M_2 with $\rho = 0$. A Bayesian model involves a prior for all unknown parameters. For M_1 we will assume a Normal prior for ρ (i.e. $\rho|M_1 \sim N(0, c)$). Imagine creating a computer program which artificially generates data from each model. Note that, for M_1 this would involve drawing a value for ρ from the Normal prior, then generating data from the appropriate AR(1) model. Next assume that you have some observed data that is generated from a “true” AR(1) DGP with $\rho = .5$. Consider comparing this observed data with artificial data generated from M_2 , the white noise model and M_1 , the unrestricted model. The observed data will likely be quite different from the data generated from M_2 , the white noise model. In contrast, if the prior for M_1 has $c = 0.5^2$, then some of the artificial data generated from M_1 will likely look quite similar to the observed data (i.e. it will be fairly common for prior draws of ρ to be in the region of .5). Bayesian methods, loosely speaking, will say “It is fairly common for M_1 to generate artificial data sets similar to the observed data, whereas M_2 always generates white noise artificial data which looks quite different from the observed data. Hence, M_1 is supported.” However, if the prior in M_1 becomes “noninformative” (i.e. c gets large) then the program generating artificial data from this model will start generating more and more “bizarre” data sets (e.g. if $c = 16$, then data sets from explosive AR(1)s with $\rho > 1$ will be common) and fewer and fewer “reasonable” data sets (i.e. where ρ is near .5). In this noninformative case Bayesian methods, loosely speaking, will say “In virtually every case, M_1 is generating bizarre data sets which are vastly different from the observed data. M_2 , with its white noise data sets, is not that great, but at least the data sets it is artificially generating are much closer to the observed data than the

as Bartlett’s paradox and is described in detail in Poirier (1995), pages 389-392.

bizarre data sets generated by M_1 . Hence, M_2 is preferred.” In the limit, as $c \rightarrow \infty$, the restricted model will always be preferred to the unrestricted one, regardless of what the observed data looks like and the size of the sample. This type of intuition also motivates why Bayes factors have such a strong reward for parsimony. Adding irrelevant variables to a model will cause it to generate more and more bizarre data sets.

The previous intuition is partly meant to motivate Bartlett’s paradox and why informative priors are crucial for Bayesian model comparison. However, it also illustrates the type of solution that we have used in previous univariate nonlinear time series work (e.g. Koop and Potter 1999a,b). In these papers, we used Normal priors which roughly reflected our subjective prior information about the parameters in the model. Consider, for instance, working with an AR(1) model for real GDP growth. This is highly unlikely to be nonstationary, so choosing c to be 0.25^2 will imply a prior which allocates most of its weight to the ”reasonable” stationary region. In a two regime threshold autoregressive model, we might place such a prior on the AR(1) coefficient in each regime. However, in multivariate cases, models are much more parameter-rich and such an approach does not work well. To be more concrete, in the VFC model, testing for nonlinearity in the mean involves Θ_1 and Θ_2 and, hence, informative priors are required for these parameters. In the univariate floor and ceiling model, Θ_1 and Θ_2 are scalars and using, say, a $N(0, I_2)$ prior seems to work quite well with macroeconomic data. However, in the present case, Θ_1 and Θ_2 each contain 16 parameters, many of which probably are essentially zero. Using, say, a $N(0, I_{32})$ prior for Θ_1 and Θ_2 implies that VFC model can generate plenty of “bizarre” data sets: only $0.68^{32} = 4.4 \times 10^{-6}$ of the prior weight is within the unit circle. In practice, we have found that the combination of including irrelevant explanatory variables and allocating significant prior weight to “bizarre” areas of the parameter space causes the Bayes factors to indicate little support for the VFC model except in cases where the floor and ceiling effects are enormous. The posterior probabilities in Table 2 associated with Prior 1 are indicative of these results. Prior 1 is described in detail in

Appendix B. However, for the present discussion, it suffices to stress that it includes a $N(0, I_{32})$ prior for Θ_1 and Θ_2

In other words, in many cases it is acceptable to use weakly informative priors of convenience, to reflect ideas like “autoregressive coefficients are probably going to be smaller than one”. The present case does not allow for this. In general, we have found that with parameter-rich multivariate nonlinear time series models, care needs to be taken in choosing priors if model comparison is being done. Prior 2 incorporates these considerations and Appendix B describes it in detail. One key aspect builds on an idea discussed in George and McCulloch (1993). Loosely speaking, we combine our prior information that “ Θ_1 and Θ_2 measure deviations from the VAR coefficients in the corridor regime, and such deviations are likely quite small” with additional prior information “ Θ_1 and Θ_2 include 32 parameters in total, many of which are likely zero, however we are not sure *a priori* which ones are likely to be zero.” That is, we use a prior which is a mixture of two mean-zero Normals for each individual element, one with a very small variance (i.e. prior variance is 0.005^2 , saying “this coefficient is effectively zero”), one with a more reasonable variable (i.e. prior variance is 0.1^2). We assume prior independence of the individual coefficients. We allocate a 75% probability to the first component (i.e. we expect 24 of the coefficients in Θ_1 and Θ_2 to be essentially zero, although we do not know which 24) and 25% to the second (i.e. we expect 8 coefficients may be important).

To return to our comparison of Bayesian and classical methods, note that the necessity of careful prior selection adds a burden to Bayesian methods. Many Bayesians would argue that this burden is worth the cost, since adding information, whether data or non-data based, will improve the accuracy of any empirical exercise. Furthermore, by investigating the sensitivity of posterior model probabilities to changes in the prior, we can gain substantial insight into the particular direction nonlinearities are or are not entering. In addition, if the model is to be used for forecasting purposes it is well known that prior information can lead to big out of sample improvements.

The difference in Bayesian and classical model comparison results arises largely due to the fact that different alternative models are being compared. The Bayesian model includes a prior, and this prior can be used to make the alternative model more reasonable. That is, nonlinear multivariate time series models will often have many more parameters than a VAR and only a few of these extra parameters will likely be important.¹² Hence, the completely unrestricted nonlinear model will often be unreasonable. The classical test and Bayes factors using a crude prior of convenience (e.g. a Normal prior) will reflect this and indicate little evidence of nonlinearity in the conditional mean. However, a nonlinear model which reflects the common sense idea that only a few of the extra parameters are important receives much more support. The prior used in this paper, based on George and McCulloch (1993), allows us to develop such reasonable nonlinear models. Similar priors, we feel, should be of great use in many areas of nonlinear multivariate time series modelling.

4 Impulse Response Analysis

Model comparison results and estimates can provide some evidence about the dynamic properties of the VFC model. However, impulse response analysis offers a deeper insight into such dynamics. Our approach will be Bayesian, in that we present impulse response functions that average over both parameter and model uncertainty.

Impulse responses measure the effect of a shock on a dynamic system. In nonlinear time series models, impulse response functions are not unique in that the effect of shocks can vary over the business cycle. In multivariate models which are not structural, the problems of defining impulse responses are complicated by the fact that the errors in the different equations are potentially correlated with one another. In particular, the common

¹²The VFC model is fairly tightly parameterized. Problems of over-parameterization which arise in it will be magnified hugely in models such as vector TARs.

approach of defining structural shocks by a Wold causal ordering applied to a reduced form estimate does not translate easily to the nonlinear case. Firstly, the estimates of the error variance covariance matrix vary across regimes and the same Wold causal ordering might have different effects across regimes. Secondly, the usual experiment is to set all shocks but one equal to zero. In the nonlinear case setting shocks to zero is dangerous since the model can become falsely stuck in a regime. Koop, Pesaran and Potter (1996) discuss these issues in detail and recommend several types of generalized impulse response functions. Here we examine the properties of the following generalized impulse response function:

$$\mathbf{GI}_n = E[\mathbf{Y}_{t+n}|Y_{3t}, Y_{4t}, \mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}] - E[\mathbf{Y}_{t+n}|\mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}].$$

That is, we calculate the difference between two n -period forecasts. The first forecast assumes knowledge all variables last period and the values of interest rates and money today.¹³ The second forecast assumes only knowledge of all variables last period. Note that \mathbf{GI}_n is a 4×1 vector which measures the effects of monetary shocks on each of the four RMPY variables. This generalized impulse response function measures the effect of unexpected changes in the variables most likely to reflect monetary policy, Y_{3t}, Y_{4t} . We stress that, since we use Bayesian methods to calculate generalized impulse responses (see Koop 1996), the expectations operators in the generalized impulse response are taken over V_{t+i} for $i = 0, \dots, n$ ¹⁴ as well as the entire parameter space. Since V_t is IIDN(0, \mathbf{I}_4) and the MCMC algorithm discussed in Appendix B provides random draws from the posterior, the expectations above can be calculated using simulation methods (see Appendix D for details).

In linear time series applications, one would usually consider the effects of monetary policy shocks by making some identification assumptions on the contemporaneous rela-

¹³Note that we have imposed the $p=1$ finding from the estimation.

¹⁴In the case of the first expectation, V_{3t} and V_{4t} are known and are not treated as random variables.

tionships between the variables and perhaps some long run restrictions. Our objectives are less ambitious. To motivate the types of shocks our generalized impulse responses are based on, consider the case where the interest rate unexpectedly increases and the money supply unexpectedly decreases, $\mathbf{GI}_n(v_{3t} > 0, v_{4t} < 0)$:

$$\begin{aligned} E[Y_{3t}|Y_{3t}, Y_{4t}, \mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}] - E[Y_{3t}|\mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}] &< 0, \\ E[Y_{4t}|Y_{3t}, Y_{4t}, \mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}] - E[Y_{4t}|\mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}] &> 0. \end{aligned}$$

One could think of this as an contractionary money supply shock, except we have not restricted the contemporaneous behavior of output or prices. Thus, it is possible the monetary policy is contractionary because of knowledge of some negative shock to output or prices. Further, since our interest rate is the 3 month T-Bill rate and our money stock measure is M2, they will be affected by developments in the financial system outside the control of the Federal Reserve. Our objective is to examine whether the system we have estimated contains any evidence of asymmetric response to negative and positive monetary shocks. There is a debate in the money shock literature concerning this issue (see Rhee and Rich 1996 and Weise 1999).

In order to calculate impulse responses, we must specify the conditions which prevail when the shock hits (i.e. $\mathbf{Y}_{t-1}, \mathbf{OH}_{t-1}, \mathbf{CDR}_{t-1}$) as well as the shock (i.e. Y_{3t}, Y_{4t}). Instead of constructing particular counterfactual choices, we use the observed data to provide histories and shocks. In order to examine the asymmetry issue and gauge the amount of dynamic nonlinearity in the system, we consider the following comparisons of the \mathbf{GI} :

$$\begin{aligned} &\int_0^\infty \int_{-\infty}^0 \mathbf{GI}_n(v_{3t}, v_{4t}, F_{t-1} = 1) f(v_{3t}, v_{4t}) dv_{3t} dv_{4t} \\ &+ \int_{-\infty}^0 \int_0^\infty \mathbf{GI}_n(v_{3t}, v_{4t}, F_{t-1} = 1) f(v_{3t}, v_{4t}) dv_{3t} dv_{4t}, \end{aligned}$$

which equals zero if positive and negative shocks have the same average effect within a regime. That is, the first term in this sum averages over all the responses to expansionary

money supply shocks ($v_{3t} < 0, v_{4t} > 0$) which hit in the floor regime. The second term does the same using all contractionary money supply shocks. If positive and negative shocks have the same effect in the floor regime, then the two terms should sum to zero. Non-zero values for this sum indicate asymmetric responses to positive and negative shocks in the floor regime. A similar calculation can be carried out for the ceiling and corridor regimes.

By construction, for $n = 0$ the measure is zero and for the VAR-hetero it is zero for all n . Thus, this is a measure of the nonlinearity in the conditional mean. Remember that, since our methods are Bayesian, we are integrating out all model parameters using their posteriors and, thus, incorporate parameter uncertainty. Furthermore, the GIs we calculate are averaged over all models which receive non-zero posterior model probability in Table 2 using prior 2. That is, lines labelled "Nonlinear" are actually an average of results for the VFC (with weight of 0.2) and the VAR-hetero (with weight of 0.8).

Alternatively, we can consider asymmetry across regimes:

$$\int_0^{\infty} \int_{-\infty}^0 \mathbf{GI}_n(v_{3t}, v_{4t}, F_{t-1} = 1) f(v_{3t}, v_{4t}) dv_{3t} dv_{4t} \\ - \int_0^{\infty} \int_{-\infty}^0 \mathbf{GI}_n(v_{3t}, v_{4t}, C_{t-1} = 1) f(v_{3t}, v_{4t}) dv_{3t} dv_{4t},$$

For the sake of brevity, we only consider asymmetry across regimes with respect to expansionary monetary shocks ($v_{3t} < 0, v_{4t} > 0$). Note that the first term in this expression measures the average effect of positive monetary shocks in the floor regime and the second term does the same for the ceiling regime. Symmetry across regimes implies these two terms are the same and, hence, the measure above is zero. Note that here we have three possibilities of comparison across regime. This time the measure can be non-zero for $n = 0$ and for the VAR-hetero model it can be different from zero. In this case we present the results only for the VFC model.

The results are contained in Figures 1 to 4, with each figure showing the response for

a RMPY different variable. Each figure contains four graphs. Asymmetry across regimes is presented in the lower right graph and the other graphs show asymmetry within the three regimes. Asymmetry within regime appears to be negligible (i.e. expansionary and contractionary shocks have the same effect in absolute value).¹⁵ There is more evidence of asymmetry across regimes. For example, in Figure 1 we can see that the response of output growth to a positive monetary shock is much greater in the floor regime than the corridor regime. This accords with the idea that expansionary monetary policy is most effective in recessions. From Figure 2 we can see that a positive monetary shock has a bigger effect on inflation in the ceiling regime than the floor regime. This accords with the idea that expansionary monetary policy just increases inflation if the economy is already over-heating. Similarly sensible stories can be told for other variables.

5 Conclusions

In this paper we have introduced a parsimonious nonlinear extension of the VAR model called the Vector Floor and Ceiling model. It is based on the idea that nonlinear fluctuations can arise due to reflective barriers. Classical and Bayesian econometric methods are developed for estimation and testing. An empirical application involving US RMPY data indicates the computational feasibility and usefulness of the VFC model. Our empirical findings indicate that most of the nonlinearity is in the contemporaneous relationship between the variables but there is also some additional, albeit mild, nonlinearity in the conditional mean dynamics.

¹⁵Note that we are integrating over the parameter and shock spaces and thus, our generalized impulse responses are non-random and thus are merely points and no measures of uncertainty (e.g. highest posterior density intervals) are associated with them. If we were not to do this integration, our impulse responses would be random variables.

References

- [1]
- [2] Altissimo and Violante (1999) "Nonlinear VARs: Some theory and an application to US GNP and unemployment," forthcoming, *Journal of Applied Econometrics*
- [3] Andrews, D. and Ploberger, W. (1994). "Optimal Tests when a Nuisance Parameter is Present only under the Alternative," *Econometrica*, 62, 1383-1414.
- [4] Beaudry, P. and Koop, G. (1993). "Do Recessions Permanently Change Output?," *Journal of Monetary Economics*, 31, 149-164.
- [5] Berger, J. and Pericchi, L. (1996). "The Intrinsic Bayes Factor for Model Selection and Prediction," *Journal of the American Statistical Association*, 91, 109-122.
- [6] Chauvet, M. (1998) "An econometric characterization of business cycle dynamics with factor structure and regime switches," *International Economic Review*, 39, 969-996.
- [7] Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327-335.
- [8] George, E. and McCulloch, R. (1993). "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881-889.
- [9] Geweke, G. (1999). "Using Simulation Methods for Bayesian Econometric Modelling: Inference, Development and Communication," *Econometric Reviews*, 18, 1-74.
- [10] Hamilton, J. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.
- [11] Hansen, B. (1996). "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 413-430.
- [12] Kapetanios, G. (1998). *Essays on the Econometric Analysis of Threshold Models*, Faculty of Economics and Politics, University of Cambridge, unpublished Ph.D. dissertation.
- [13] Kim, M. and Yoo, J. (1996). "New index of coincident indicators: A multivariate Markov switching factor model approach," *Journal of Monetary Economics*, 36, 607-630.
- [14] Koop, G. (1996). "Parameter Uncertainty and Impulse Response Analysis," *Journal of Econometrics*, 72, 135-149.

- [15] Koop, G., Pesaran, H., and Potter, S. (1996). "Impulse Response Analysis in Non-linear Multivariate Models," *Journal of Econometrics*, 74, 119-148.
- [16] Koop, G. and Potter, S. (1997). "Bayesian Analysis of Endogenous Delay Threshold Models," manuscript available at <http://www.ed.ac.uk/~gkoop/>.
- [17] Koop, G. and Potter, S. (1999a). "Bayes Factors and Nonlinearity: Evidence from Economic Time Series," *Journal of Econometrics*, 88, 251-85.
- [18] Koop, G. and Potter, S. (1999b). "Dynamic Asymmetries in US Unemployment," *Journal of Business and Economic Statistics*, 17, 298-312.
- [19] Pesaran, M.H. and Potter, S. (1997). "A Floor and Ceiling Model of U.S. Output," *Journal of Economic Dynamics and Control*, 21, 661-695.
- [20] Poirier, D. (1995). *Intermediate Statistics and Econometrics*, Cambridge: The MIT Press.
- [21] Potter, S. (1995a). "A Nonlinear approach to US GNP," *Journal of Applied Econometrics*, 10, 109-125.
- [22] Potter, S., (1995b), "Nonlinear models of economic fluctuations," in *Macroeconometrics: Developments, Tensions and Prospects*, edited by K. Hoover, 517-560.
- [23] Rhee, W., and Rich, R. (1995) "Inflation and Asymmetric Effects of Money," *Journal of Macroeconomics*, 683-702.
- [24] Sims, C. (1972). "Money, income and causality," *American Economic Review*, 62, 540-52
- [25] Sims, C. (1980). "Macroeconomics and Reality," *Econometrica*, 48, 1-48.
- [26] Skalin, J. and Terasvirta, T. (1999). "Another Look at Swedish Business Cycles, 1861-1988," *Journal of Applied Econometrics*, 14, 359-378.
- [27] Terasvirta, T. (1994). "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208-218.
- [28] Tong, H. (1990). *Non-linear Time Series Models*, Clarendon Press, Oxford.
- [29] Verdinelli, I. and Wasserman, L. (1995). "Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio," *Journal of the American Statistical Association*, 90, 614-618.
- [30] Weise, C. (1999), "The asymmetric effects of Monetary policy: a nonlinear vector autoregression approach," *Journal of Money, Credit and Banking*, 31, 85-108.

Appendix A: Sample Information

We start by organizing the vector floor and ceiling model into more convenient form. Define a $1 \times (K(p+2) + K)$ vector $\mathbf{X}_t(\gamma)$ by

$$\left[1 \quad \mathbf{Y}'_{t-1} \quad \cdots \quad \mathbf{Y}'_{t-p} \quad \mathbf{CDR}'_{t-1} \quad \mathbf{OH}'_{t-1} \right],$$

and the $(K(p+2) + K) \times K$ matrix \mathbf{A} by

$$\begin{bmatrix} \alpha' \\ \Phi'_1 \\ \vdots \\ \Phi'_p \\ \Theta'_1 \\ \Theta'_2 \end{bmatrix}$$

This gives

$$\mathbf{Y}'_t = \mathbf{X}_t(\gamma)\mathbf{A} + \mathbf{V}'_t\mathbf{H}'_{t-1}(\gamma),$$

where $H_{t-1}(\gamma) = \mathbf{H}_0\mathbf{COR}_{t-1} + \mathbf{H}_1F_{t-1} + \mathbf{H}_2C_{t-1}$. If we stack the observations we have:

$$\mathbf{Y} = \mathbf{X}(\gamma)\mathbf{A} + \mathbf{U},$$

where

$$E \left[\text{vec}(\mathbf{U}') \text{vec}(\mathbf{U}')' \right] = \begin{bmatrix} \Sigma_1^* & \mathbf{0}_4 & \cdots & \mathbf{0}_4 \\ \mathbf{0}_4 & \Sigma_2^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_4 \\ \mathbf{0}_4 & \cdots & \mathbf{0}_4 & \Sigma_T^* \end{bmatrix} = \mathbf{\Sigma}(\gamma),$$

and $\Sigma_t^* = \Sigma_0\mathbf{1}(COR_{t-1} = 1) + \Sigma_1\mathbf{1}(F_{t-1} = 1) + \Sigma_2\mathbf{1}(C_{t-1} = 1)$.

The ordinary least squares estimator of \mathbf{A} for γ is direct as:

$$\mathbf{A}_{OLS}(\gamma) = \left[\mathbf{X}'(\gamma)\mathbf{X}(\gamma) \right]^{-1} \mathbf{X}'(\gamma)\mathbf{Y}.$$

The generalized least squares estimator requires more work. We have:

$$\mathbf{Y}_t = (\mathbf{I}_K \quad \mathbf{X}_t(\gamma)) \text{vec}(\mathbf{A}) + \mathbf{H}_{t-1}(\gamma)\mathbf{V}_t,$$

and once again stacking observations we have

$$\mathbf{y} = \mathbf{x}(\gamma) \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{U}'),$$

where \mathbf{y} is a $(T \times K) \times 1$ vector and \mathbf{x} is a $(T \times K) \times (K^2(p+2) + K)$ matrix. Thus the GLS estimator of \mathbf{A} for γ is:

$$\text{vec}(\mathbf{A}_{GLS}(\gamma)) = \mathbf{a}_{GLS}(\gamma) = [\mathbf{x}'(\gamma) \Sigma^{-1}(\gamma) \mathbf{x}(\gamma)]^{-1} \mathbf{x}'(\gamma) \Sigma^{-1}(\gamma) \mathbf{y}.$$

The estimator is implemented by constructing $\mathbf{x}'(\gamma) \Sigma^{-1}(\gamma)$ from $(\mathbf{I}_K \quad \mathbf{X}_t(\gamma))' \Sigma_t^{*-1}$ (this utilizes the block diagonality of $\Sigma^{-1}(\gamma)$) and the relevant Bayesian or classical estimator of the 3 possible variance covariance matrices).

Define the $2K^2 \times (K^2(p+2) + K)$ selection matrix Υ such that

$$\Upsilon \text{vec}(\mathbf{A}) = \text{vec} \begin{pmatrix} \Theta_1' \\ \Theta_2' \end{pmatrix}.$$

To represent the sample information in the multiple time series we will use the symbol \mathbb{Y} .

Appendix B: Bayesian Analysis of the VFC Model

Geweke (1999) provides a survey of simulation-based Bayesian methods and the reader is referred there for a description of the theory underlying the methods used in this paper.

The VFC model is given in equation (5). Note that the parameter vector $\gamma = (r_F, r_C)'$ enters \mathbf{CDR}_{t-1} , \mathbf{OH}_{t-1} , COR_{t-1} , F_{t-1} and C_{t-1} . The heteroskedasticity and nonlinearity of this model mean that analytical posterior results are not available. However, a Markov Chain Monte Carlo (MCMC) algorithm can be developed which provides pseudo-random draws from the posterior. Given such draws from the posterior, items such as posterior means and standard deviations of all parameters or features of interest (e.g. impulse response functions) can be calculated. To motivate our MCMC algorithm, note that,

conditional on γ , the sample information can be summarized by least squares type estimators and known forms for the conditional posteriors exist under conjugate forms for the prior distributions. The posterior for γ , conditional on all the other parameters, is of non-standard form but, since γ contains only two elements, draws can be provided using a simple Metropolis-Hastings algorithm (see Chib and Greenberg, 1995). To calculate Bayes factors comparing nonlinear to linear and homoskedastic to heteroskedastic models, we use the Savage-Dickey density ratio (see Verdinelli and Wasserman, 1995). Precise details of all our Bayesian computational methods are given in the remainder of this appendix.

5.1 The Prior

A well-known issue in Bayesian statistics is that the use of non-informative priors is typically acceptable for estimation purposes, but that informative priors are required for the purpose of calculating Bayes factors (see Koop and Potter, 1999a, for a detailed discussion of this issue). Given the reluctance of some researchers to accept results obtained using subjectively elicited priors, in this paper we use priors based on the idea of a training sample (see Berger and Pericchi, 1996 for more detailed development of such ideas). The general idea is to begin with a noninformative prior, then divide the sample period into $t = 1, \dots, trs, trs + 1, \dots, T$. The data from period $t = 1, \dots, trs$ is called a training sample. It is combined with the noninformative prior to yield a “posterior”. This “posterior” is then used as “training sample prior” for the data from $t = trs + 1, \dots, T$. Note that strategy yields a posterior which is identical to one based on combining a noninformative prior with the full sample in the classical linear regression model. However, in our case because of the nonlinearity this would not be precisely the case. Typically, trs is chosen to be the smallest value which yields a proper training sample prior and in non-time series applications one can consider averaging over different training samples to obtain an “intrinsic” Bayes factor. Here we set $trs = 31$, a value

which yields a proper prior for values of p up to 6 in the present data set and allows us to start our estimation with a typical post-1954/Korean war dataset. In practice we will find evidence in favor of a lag length of 1, thus the ratio of parameters to observations is approximately 1 : 4.

Formally, in this paper we construct the prior γ , \mathbf{A} and Σ_j^{-1} for $j = 0, 1, 2$ in three main steps. First, we use a flat prior, independent of all other parameters, for the elements of γ . In particular, we allow $p(r_f)$ to be uniform over the interval $[-0.492, -0.0025]$ and $p(r_c)$ to be uniform over the interval $[0.599, 1.284]$. These values are chosen to ensure comparability with the maximum likelihood results and ensure that an adequate number of observations lie in each regime (see the section containing empirical results for more detail).

Second, for the elements of \mathbf{A} shared in common with the linear model we use a training sample. We also use a training sample for the precision matrices. This is implemented by finding the posterior distributions of a linear VAR using a standard noninformative prior plus the training sample. This formally leads to a Normal-Wishart "posterior" for the parameters. In a conventional training sample approach, this "training sample posterior" is used as a prior which is then combined with the remaining observations to produce a posterior using all the data. Rather than using this "training sample posterior" directly, we use the conditional Normality of the VAR coefficients to imply a Normal prior for the elements of A shared in common with the linear model. We then take the implied Wishart form for the VAR error precision matrix and use it as a prior for Σ_j^{-1} for $j = 0, 1, 2$. This suffices to fully specify a prior for all model parameters except Θ_1, Θ_2 .

The third step in our prior elicitation procedure is to use a mixture prior for the nonlinear part of the conditional mean, Θ_1, Θ_2 , as suggested by George and McCulloch (1993). We have informally motivated this prior in Section 3 of the paper, here we provide precise details. We assume that the individual elements of Θ_1, Θ_2 are *a priori* independent and identically distributed. Thus without loss of generality consider the

prior for the generic element, θ :

$$f(\theta) = \begin{cases} \frac{1}{\sqrt{2\pi\tau}} \exp[-0.5(\frac{\theta}{\tau})^2] & \text{if } \xi = 1 \\ \frac{1}{\sqrt{2\pi\varsigma}} \exp[-0.5(\frac{\theta}{\varsigma})^2] & \text{if } \xi = 0 \end{cases},$$

where $\varsigma \approx 0$. Stacking the individual Bernoulli random variables ξ , corresponding to each element of Θ_1, Θ_2 into the vector $\boldsymbol{\xi}$ we have:

$$\text{vec} \begin{pmatrix} \Theta_1' \\ \Theta_2' \end{pmatrix} \sim N(\mathbf{0}, \mathbf{G}_\xi),$$

where $\mathbf{G}_\xi = \tau \boldsymbol{\xi} \mathbf{I}_{k^2} + \varsigma (e_{k^2} - \boldsymbol{\xi}) \mathbf{I}_{k^2}$.

Combining the two elements of prior for \mathbf{A} and Σ_j^{-1} for $j = 0, 1, 2$ we have

$$p(A, \Sigma_0^{-1}, \Sigma_1^{-1}, \Sigma_2^{-1} | \boldsymbol{\xi}) = p(\text{vec}(\mathbf{A}) | \boldsymbol{\xi}) p(\Sigma_0^{-1}) p(\Sigma_1^{-1}) p(\Sigma_2^{-1}), \quad (\text{A.3})$$

where

$$p(\text{vec}(\mathbf{A}) | \boldsymbol{\xi}) \propto f_N^{(K^2(p+2)+K)}(\mathbf{a}_0, C_\xi^{-1}) \quad (\text{A.4})$$

and

$$p(\Sigma_j^{-1}) = f_W^K(\nu_0, D_0^{-1}), \quad (\text{A.5})$$

for $j = 0, 1, 2$. In equations (A.4) and (A.5), $f_N^M(b, B)$ is the M -variate Normal p.d.f. with mean b and covariance matrix B , and $f_W^K(b, B)$ is the K -dimensional Wishart p.d.f. with b degrees of freedom and mean bB . Furthermore, ν_0, D_0^{-1} are obtained from the training sample prior and \mathbf{a}_0, C_ξ^{-1} are training sample prior results augmented to include the prior for Θ_1, Θ_2 (e.g. C_ξ includes \mathbf{G}_ξ).

MCMC Algorithm

The MCMC algorithm used in this paper involves sequentially drawing from the posterior conditional distributions described below.

The joint posterior for \mathbf{A} conditional on γ, ξ and the error covariances involves only the Normal distribution. Combining the prior with the likelihood function, we find the posterior for $\text{vec}(\mathbf{A})$ conditional on $\Sigma_0^{-1}, \Sigma_1^{-1}, \Sigma_2^{-1}, \gamma, \xi$ is multivariate normal with mean

$$\tilde{\mathbf{a}} = \tilde{\mathbf{C}}^{-1}[\mathbf{C}_\xi \mathbf{a}_0 + [\mathbf{x}'(\gamma) \Sigma^{-1}(\gamma) \mathbf{x}(\gamma)] \mathbf{a}_{GLS}(\gamma)],$$

and variance covariance matrix:

$$\tilde{\mathbf{C}}^{-1} = \{\mathbf{C}_\xi + [\mathbf{x}'(\gamma) \Sigma^{-1}(\gamma) \mathbf{x}(\gamma)]\}^{-1}.$$

In the case of the restricted model VFC-homo the conditional posterior is multivariate normal form with mean

$$\tilde{\mathbf{a}} = \tilde{\mathbf{C}}^{-1}[\mathbf{C}_\xi \mathbf{a}_0 + [\mathbf{x}'(\gamma)(\Sigma^{-1} \quad \mathbf{I}_T) \mathbf{x}(\gamma)] \mathbf{a}_{OLS}(\gamma)],$$

and variance covariance matrix

$$\tilde{\mathbf{C}} = \{\mathbf{C}_\xi + [\mathbf{x}'(\gamma)(\Sigma^{-1} \quad \mathbf{I}_T) \mathbf{x}(\gamma)]\}^{-1}.$$

The posteriors for $\Sigma_0^{-1}, \Sigma_1^{-1}, \Sigma_2^{-1}$, conditional on $\text{vec}(\mathbf{A}), \gamma, \xi$, are Wishart distributions with degrees of freedom $\tilde{\nu} = T_j(\gamma) + \nu_0$, and scale matrices $\tilde{\mathbf{D}} = \mathbf{S}_j + D_0$, where

$$T_0(\gamma) = \sum_{t=trs+1}^T 1(COR_{t-1} = 1), \quad T_1(\gamma) = \sum_{t=trs+1}^T 1(C_{t-1} = 1), \quad T_2(\gamma) = \sum_{t=trs+1}^T 1(F_{t-1} = 1),$$

and

$$\begin{aligned} \mathbf{S}_0 &= \sum_{t=trs+1}^T 1(COR_{t-1} = 1)(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A})'(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}), \\ \mathbf{S}_1 &= \sum_{t=trs+1}^T 1(C_{t-1} = 1)(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A})'(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}), \\ \mathbf{S}_2 &= \sum_{t=trs+1}^T 1(F_{t-1} = 1)(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A})'(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}). \end{aligned}$$

We use a Gibbs sampling algorithm conditional on $\gamma, \boldsymbol{\xi}$ to simulate from these two sets of conditional posterior distributions.

So far, the discussion has proceeded conditionally on $\gamma, \boldsymbol{\xi}$. Following George and McCulloch we assume *a priori* that each element of $\boldsymbol{\xi}$ is equal to 1 with independent and identical probability q_0 . The posterior of a specific element of $\boldsymbol{\xi}$ conditional on the draw of respective element of \mathbf{A} is given by:

$$P[\xi = 1|\theta] = \frac{\tau^{-0.5} \exp[-0.5(\frac{\theta}{\tau})^2]q_0}{\tau^{-0.5} \exp[-0.5(\frac{\theta}{\tau})^2]q_0 + \varsigma^{-0.5} \exp[-0.5(\frac{\theta}{\varsigma})^2](1 - q_0)}.$$

To complete the MCMC algorithm, we need to draw from $p(\gamma|\mathbb{Y}, \mathbf{A}, \Sigma_0^{-1}, \Sigma_1^{-1}, \Sigma_2^{-1}, \boldsymbol{\xi})$ which does not have a standard form. Hence, we use an independence chain Metropolis-Hastings algorithm for drawing from $p(\gamma|\mathbb{Y}, \mathbf{A}, \Sigma_0^{-1}, \Sigma_1^{-1}, \Sigma_2^{-1}, \boldsymbol{\xi})$. This involves taking random draws from a candidate generating density, which we choose to be a density constructed from the maximum likelihood concentrated likelihood functions for the thresholds (see Appendix C). Candidate draws are either accepted or rejected. If they are rejected, the MCMC chain retains the previous draw for γ . Since the candidate density is constructed from the concentrated likelihood function then the acceptance probability has a simple form. We found this method to be quite efficient with 16% of the draws being accepted.

Selection of Prior Hyperparameters

It remains to specify the prior hyperparameters: q_0, ς, τ . We start by assuming that about two of the nonlinear variables will be significant in each equation. This leads to $q_0 = 0.25$. Next we need to consider what it means for the nonlinear coefficients to be different from zero. This issue is discussed in the main text and is implemented with the assumption that $\tau = 0.1^2$ and $\varsigma = 0.005^2$. This is prior 2 in the main text. We also include a more traditional prior where the elements of Θ_1, Θ_2 are *IIDN*(0, 1). This latter prior is labelled

”Prior 1” in the text. Prior 2 leads to a posterior where 20% of the elements of ξ were non-zero on average.

Bayes Factor Calculation Using Savage-Dickey Density Ratio

As discussed in the main text we have four possible models to consider. We use marginal likelihoods to assess the relative support from the observed sample for each type of model. Instead of calculating the marginal likelihoods directly we use the fact that the models are nested to calculate Bayes factors by averaging a conditional form of Savage-Dickey density ratio.. There are two main Bayes factors to find:

1. VAR-hetero vs VFC.
2. VAR vs. VAR-hetero

Bayes Factor for Testing Linearity in Mean

The Bayes factor for testing $\Theta_1 = \Theta_2 = \mathbf{0}$ is straightforward to calculate by drawing on the Normality of the prior and conditional posterior. That is, the Savage-Dickey density ratio implies:

$$B_{VAR-hetero,VFC} = \frac{p(\Upsilon \text{vec}(\mathbf{A}) = \mathbf{0} | \mathbb{Y})}{p(\Upsilon \text{vec}(\mathbf{A}) = \mathbf{0})}, \quad (\text{A.10})$$

where the selection matrix, Υ is defined in Appendix A. The height of the prior density at zero is given by:

$$[\tau^{-0.5} \exp[0]q_0 + \varsigma^{-0.5} \exp[0](1 - q_0)]^{2K^2}.$$

The numerator cannot be directly calculated. Instead at each iteration of the MCMC one calculates the logarithm of:

$$\det(\Upsilon \tilde{\mathbf{C}}^{-1} \Upsilon')^{-0.5} \exp\left(-0.5 \tilde{\mathbf{a}}' \Upsilon' [\Upsilon \tilde{\mathbf{C}}^{-1} \Upsilon']^{-1} \Upsilon \tilde{\mathbf{a}}\right)$$

and subtracts from it the logarithm of the prior density evaluated at zero. The anti-logarithm is then taken and the collection of values are averaged across the MCMC draws. Calculating Bayes factors involving any of the coefficients in the conditional mean can be done by changing the selection matrix in the above formulae. Such a strategy is used to determine lag order selection (in this case the prior density is also normal).

Figure B.1 shows the recursive mean of the Bayes factor for the VAR-hetero vs. VFC calculated in this manner for a run of 80,000 draws from the posterior. It is clear that there is substantial variation in the estimate. The large upward movements were caused by the draw of ξ producing a zero vector. Convergence was checked by a number of long independent runs which confirmed a Bayes factor of 4.

Bayes Factor for Testing Homoskedasticity

The calculation of Bayes factors for testing for $\Sigma_0^{-1} = \Sigma_1^{-1} = \Sigma_2^{-1}$ is more complicated. To ease the derivations we work with the following transformation of the precisions: $R_0 = \Sigma_0^{-1}$, $R_1 = \Sigma_1^{-1} - \Sigma_0^{-1}$ and $R_2 = \Sigma_2^{-1} - \Sigma_0^{-1}$.¹⁶ Homoskedasticity occurs if $R_1 = R_2 = \mathbf{0}$. Using the Savage-Dickey density ratio, the Bayes factor comparing homoskedasticity to heteroskedasticity is thus:

$$B_{Hom,Het} = \frac{p(R_1=\mathbf{0}, R_2=\mathbf{0}|\mathbb{Y})}{p(R_1=\mathbf{0}, R_2=\mathbf{0})}. \quad (\text{A.12})$$

Once again we calculate the numerator by averaging posterior draws from the MCMC. The generic problem of calculating $p(R_1=\mathbf{0}, R_2=\mathbf{0}|\mathbb{Y}, \mathbf{A}, \gamma)$ and $p(R_1 = \mathbf{0}, R_2 = \mathbf{0})$ is essentially the same. It reduces to the following:

Let Σ_i^{-1} be independent $W_K(D_i, \nu_i)$ for $i=0,1,2$ where $W_K(.,.)$ denotes the M -dimensional Wishart distribution (see Poirier, 1995, page 136). Consider the transformation: $R_0 = \Sigma_0^{-1}$,

¹⁶There are many transformations that could be used here. For instance, we could use the fact that $P_0 P_1^{-1} = P_0 P_2^{-1} = I$ implies $P_0 = P_1 = P_2$ in order to derive a Bayes factor for testing homoskedasticity. The choice made in this paper was largely made for simplicity. The Bayes factor (like the Wald test statistic) will depend the exact transformation used.

$R_1 = \Sigma_1^{-1} - \Sigma_0^{-1}$ and $R_2 = \Sigma_2^{-1} - \Sigma_0^{-1}$. What is $p(R_1 = 0, R_2 = 0)$?

To solve this problem, we begin by deriving $p(R_0, R_1, R_2)$. This can be done using the change-of variable theorem. Noting that the Jacobean of the transformation is the identity matrix and using the formula for the Wishart density (see Poirier, 1995, page 136), an expression for $p(R_0, R_1, R_2)$ can be derived. If we then note that $p(R_0, R_1 = 0, R_2 = 0)$ simplifies to a form involving a Wishart kernel and use the properties of the Wishart to integrate out R_0 , we can obtain:

$$p(R_1=0, R_2 = 0) = \frac{\tilde{c}}{c_0 c_1 c_2 |D_0|^{\frac{\nu_0}{2}} |D_1|^{\frac{\nu_1}{2}} |D_2|^{\frac{\nu_2}{2}} |\tilde{D}|^{\frac{\tilde{\nu}}{2}}},$$

where c_i for $i=0,1,2$ are the integrating constants for the original Wishart distributions given on page 136 of Poirier (1995),

$$\tilde{A} = \tilde{D}_0^{-1} + \tilde{D}_1^{-1} + \tilde{D}_2^{-1}$$

and \tilde{c} is the integrating constant from the $W_M(\tilde{D}^{-1}, \tilde{\nu})$ with $\tilde{\nu} = \tilde{\nu}_0 + \tilde{\nu}_1 + \tilde{\nu}_2 - 2K - 2$. This result can be used to calculate both the numerator and denominator of the Bayes factor.

This strategy can be used to calculate the Bayes factor comparing the homoskedastic VFC model (VFC-homo) to the heteroskedastic VFC model and comparing the traditional VAR to the heteroskedastic VAR model (VAR-hetero).

Appendix C: Classical Analysis of the VFC Model

In this Appendix, classical estimation and testing of the VFC model given in Equation 5 is presented. Some basic results and notation drawn upon here are given in Appendix A.

Estimation

For fixed values of $\gamma = (r_F, r_C)'$ estimation could proceed by Ordinary Least Squares, equation by equation. This suggests a strategy where a grid of possible values for r_F, r_C

is chosen by considering the discontinuities in the sum of squares function produced by the discreteness of the observed data. Unlike standard threshold models the VFC models does not have a flat sum of squares function between grid points but we ignore this extra complication. At each grid point, the sum of squared errors from each equation is calculated:

$$\mathbf{s}^2(\gamma) = \text{trace} \left(\sum_{t=1}^T (\mathbf{Y}'_t - \mathbf{X}_t(\gamma) \mathbf{A}_{OLS}(\gamma))' (\mathbf{Y}'_t - \mathbf{X}_t(\gamma) \mathbf{A}_{OLS}(\gamma)) \right)$$

One then chooses as estimates of r_F, r_C , the values which minimize $\mathbf{s}^2(\gamma)$.

Although the ordinary least squares estimator so defined will be consistent for the conditional mean parameters and thresholds, it will be more useful for our purposes to consider maximum likelihood estimation. There are three reasons for this. First, as discussed below, it allows the construction of a test for nonlinearity along the lines developed in Pesaran and Potter (1997). Second, it will become apparent that there is substantial variation across threshold values in the matrices $\Sigma_0, \Sigma_1, \Sigma_2$ and that these can have important effects of the dynamics of the model. Thus, the estimation of these matrices is important. Finally, in implementing the simulation methodology of Hansen (1996), the use of GLS type estimators greatly simplifies the construction of the test statistics since it incorporates possible conditional heteroskedascity present in the residuals.

For the VFC model maximum likelihood estimates can be found by iterated feasible generalized least squares techniques. Further, while the grid search is proceeding one can calculate various test statistics and “bootstrap” distributions for them. The procedure is as follows:

1. Generate and store \mathbf{v} which is a $KT \times J$ matrix of standard Normal random variates which i^{th} column \mathbf{v}_i .
2. For a grid point γ estimate the VFC model by OLS.

3. Construct estimates of $\Sigma_0, \Sigma_1, \Sigma_2$ from the estimated innovations:

$$\begin{aligned}\widehat{\Sigma}_0 &= \sum_{t=1}^T 1(COR_{t-1} = 1)(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}_{OLS}(\gamma))'(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}_{OLS}(\gamma)), \\ \widehat{\Sigma}_1 &= \sum_{t=1}^T 1(C_{t-1} = 1)(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}_{OLS}(\gamma))'(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}_{OLS}(\gamma)), \\ \widehat{\Sigma}_2 &= \sum_{t=1}^T 1(F_{t-1} = 1)(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}_{OLS}(\gamma))'(\mathbf{Y}'_t - \mathbf{X}_t(\gamma)\mathbf{A}_{OLS}(\gamma)).\end{aligned}$$

4. Use these estimates to calculate the GLS estimate, $\mathbf{A}_{GLS}(\gamma)$.

5. Keep iterating between steps 2 and 3 (replacing $\mathbf{A}_{OLS}(\gamma)$ with the latest $\mathbf{A}_{GLS}(\gamma)$ in step 2), until the parameters converges.

6. Save the value of the log likelihood at this grid point:

$$-0.5 \left\{ \begin{aligned} \sum_{t=1}^T 1(COR_{t-1} = 1) \ln(\det[\Sigma_{0,MLE}]) &+ 1(C_{t-1} = 1) \ln(\det[\Sigma_{1,MLE}]) \\ &+ 1(F_{t-1} = 1) \ln(\det[\Sigma_{2,MLE}]) \end{aligned} \right\}$$

7. Construct and store the quadratic form:

$$\mathbf{a}'_{MLE} \Upsilon' [\Upsilon' [\mathbf{x}'(\gamma) \mathbf{x}(\gamma)]^{-1} \Upsilon]^{-1} \Upsilon \mathbf{a}_{MLE}$$

and construct and store J additional quadratic forms using the Normal random variates generated in Step 1. That is, for $i = 1, \dots, J$ calculate:

$$\begin{aligned} &\left([\mathbf{x}'(\gamma) \mathbf{x}(\gamma)]^{-1} \mathbf{x}'(\gamma) \mathbf{v}_i \right)' \Upsilon' [\Upsilon' [\mathbf{x}'(\gamma) \mathbf{x}(\gamma)]^{-1} \Upsilon]^{-1} \\ &\left(\Upsilon [\mathbf{x}'(\gamma) \mathbf{x}(\gamma)]^{-1} \mathbf{x}'(\gamma) \mathbf{v}_i \right) \end{aligned}$$

8. Repeat procedure from step 2 for all grid points and choose as the MLE estimate of γ that maximizes the log likelihood function.

9. Calculate various information criteria at the MLE of γ .

The computational demands of this estimation procedure are quite high because a reasonable number of grid points can easily be over 2,000. Figures C.1 and C.2 show the concentrated log likelihood across the grid points for the VFC and VAR-hetero models. This concentrated likelihood was also used in the posterior simulator for the thresholds as discussed above.

Testing

Testing in nonlinear time series models such as the VFC model is complicated by Davies' problem. In the present case, the nuisance parameters, r_F and r_C , are not identified under the linear null hypothesis. Hansen (1996) outlines a simulation based method for calculating P-values for tests of the linear null which we follow here in step 7. The J quadratic forms constructed at each grid point will be draws from a Chi-squared distribution with $2K^2$ degrees of freedom. Under the null hypothesis of $\Theta_1 = \Theta_2 = \mathbf{0}$ the quadratic form constructed using the MLE of Θ_1, Θ_2 will have an asymptotic Chi-squared distribution for a grid point chosen at random. However, test statistics constructed from the grid by maximization and averaging will not have Chi-squared distributions. The distribution of such statistics can be found by performing the same operation on the each of the J simulated statistics individually to construct an approximation to the true distribution. Following Andrews and Ploberger (1994), in this paper we not only consider the SUP WALD test statistic, but also the average of the Wald statistics (AVE WALD) and the exponential average of the Wald statistics (EXP WALD).

An alternative test procedure is outlined in Pesaran and Potter (1997). Davies' problem does not occur if we define the null hypothesis as one with a linear conditional mean, but floor and ceiling effects present in the error covariance (i.e. the VAR-hetero model introduced in Section 2). Hence, a standard likelihood ratio test involving the VFC and VAR-hetero models can be done in this case. It is worth stressing that the Hansen approach yields a test for any floor and ceiling effects in the conditional mean but can be

affected by strong floor and ceiling effects in the error variance-covariances whereas the Pesaran-Potter approach yields a test for floor and ceiling effects in the conditional mean only.

The Pesaran-Potter test requires only the maximum values of the likelihood function for the VFC and VAR-hetero models. Thus, it requires a parallel grid search to estimate the VAR-hetero model. The likelihood ratio test based on these two models will, asymptotically, have a Chi-squared distribution with appropriate degrees of freedom. Constructing this test statistic is clearly computationally intensive, since it doubles the time required for estimation. However, it allows one to investigate properties of the two concentrated (with respect to γ) likelihood surfaces that contain useful information. In particular, given that the grid is chosen so that reasonable amounts of data must be present in all 3 regimes, one can examine extremely conservative test statistics such as a minimum likelihood ratio. Asymptotically, the minimum likelihood ratio will be first order stochastically dominated by a Chi-squared distribution with appropriate degrees of freedom. In the empirical section we consider the minimum likelihood ratio for VFC vs the VAR and the VAR-hetero vs the VAR.

Appendix D: Further Details on Impulse Response Analysis

For the VFC model simulation of V_{t+i} for $i = 0, \dots, n$ is required (in addition to MCMC simulation). Since these are IIDN this additional simulation is easy to carry out. For the VAR and VAR-hetero models analytical results (partially sketched below) can be drawn upon so that no simulation beyond the MCMC is required.

We use the posterior draws from the Bayesian analysis for each type of model to construct the generalized impulse response function (see Koop, 1996). The Normality assumption on the errors is crucial for ease of calculation of \mathbf{GI}_0 . Partitioning the error

covariance matrix as (where we suppress subscripts and superscripts in the partition):

$$\Sigma_t^m = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

we have

$$\widehat{\mathbf{G}}\mathbf{I}_0^m \sim N \left(\begin{bmatrix} \Sigma_{12}\Sigma_{22}^{-1}[v_3, v_4]' \\ [v_3, v_4]' \end{bmatrix}, \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

for each draw m from the posterior. This is then averaged across posterior draws. For the linear-in-mean models the draw of the first order coefficient matrix is used in the recursion:

$$\widehat{\mathbf{G}}\mathbf{I}_n^m = (\Phi^m)^n \widehat{\mathbf{G}}\mathbf{I}_{n-1}^m$$

to obtain the impulse response at other horizons, where Φ is the matrix of VAR(1) coefficients.