Expectations Hypothesis Tests in the Presence of Model Uncertainty

Erdenebat Bataa

Centre for Growth and Business Cycles Research, Economics, University of Manchester

Dong H. Kim

Dept. of Economics, Korea University

Denise R. Osborn*

Centre for Growth and Business Cycles Research, Economics, University of Manchester

October 2006

* Corresponding author. Denise Osborn, Centre for Growth and Business Cycles Research, Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester, UK, M13 9PL. Email: denise.osborn@manchester.ac.uk, Tel: +44(0)161-275-4861. Details of other authors: Erdenebat Bataa, Email: e.bataa@manchester.ac.uk, Tel: +44(0)161-275-4860, Dong Heon Kim, Email: dongkim@korea.ac.kr, Tel: +82-2-3290-2226.

Acknowledgements: The authors would like to thank Chris Orme, Simon Peters and Econometrics seminar participants at the University of Manchester for their useful comments. Erdenebat Bataa would also like to acknowledge financial assistance from the University of Manchester and the Open Society Institute.

Expectations Hypothesis Tests in the Presence of Model Uncertainty

Abstract

We extend vector autoregressive (VAR) model based expectations hypothesis tests of the term structure by relaxing some specification assumptions in order to reflect model uncertainty. Firstly, the wild bootstrap is used to allow for conditional heteroskedasticity in the VAR residuals without imposing any parameterization on this heteroskedasticity. Secondly, the model selection procedure is endogenized in the bootstrap replications and supplemented with a multivariate autocorrelation test robust to conditional heteroskedasticity to reflect true uncertainty about the VAR order and to ensure the absence of residual serial correlation. Finally, a stationarity correction is introduced to prevent the finite-sample bias adjusted VAR coefficients from becoming explosive. When this new methodology is applied to extensive US structure data ranging from 1 month to 10 years, we find rejection of the theory occurs primarily at the short end of the maturity spectrum in line with the literature, however less often than previously documented.

JEL classification: G10; E43.

Keywords: expectations hypothesis; term structure; wild bootstrap; conditional heteroskedasticity

I. Introduction

The interrelationship between interest rates of various maturities is a fundamental topic in economics and finance. One of the main theories put forward to explain this relationship is the expectations hypothesis (EH). The EH of the term structure implies that, in equilibrium, investing in a succession of short-term bonds gives the same expected return as investing in a long-term bond, when adjustment is made for the (assumed constant) term premium. Various implications of the theory have been tested, e.g. in Sargent (1979), Shiller, Campbell and Schoenholtz (1983), Hall, Anderson and Granger (1992), Evans and Lewis (1994), Lanne (1999) and Bekaert, Hodrick and Marshall (1997, 2001) for the US and in Taylor (1992), Dahlquist and Jonsson (1995), Gerlach and Smets (1997) for other countries, leading to mixed results with perhaps more evidence against the theory in the US data than for the latter group.[1]

Moreover, if one considers multiple maturity pairs and the predictive ability of the spread between long and short rates for future short rate changes, previous studies such as Campbell and Shiller (1991) and Sarno, Thornton and Valente (2006) comfortably reject the EH at the shortest end of the maturity spectrum while Campbell (1995), Rudebusch (1995) and Roberds and Whiteman (1999) note the predictive ability of the spread is better at the short and long ends of the maturity spectrum and less good in the intermediate maturity range for a given short rate, thus creating a "U" shaped pattern. However, Thornton (2006) argues that the slope coefficient used to measure predictive power is not of primary concern, as data generating processes (DGPs) where the EH does not hold are capable of generating the same slope coefficient as that predicted by the EH.

[1] The evidence against the EH is much less severe if high frequency data are used (Longstaff, 2000) and also if expectations are taken from surveys (Froot, 1989).

Thus, empirical results offer mixed support for the EH. However, Mankiw and Miron (1986) argue that the poor performance of the EH over certain periods is related to the monetary policy pursued by the US Fed, with the EH performing better in periods of monetary targeting than in periods of interest rate targeting (and even better before the foundation of the Fed). Rudebusch (1995), Roberds, Runkle and Whiteman (1996), and Balduzzi, Bertola and Foresi (1997) provide models that accommodate Fed behaviour and confirm Mankiw and Miron's finding.

Rather than a failure of the EH, conclusions inconsistent with the theory have sometimes been attributed to the small sample properties of the tests. Early studies use a conventional regression framework (e.g. Mankiw and Miron 1986), volatility tests (Shiller 1979) and VAR-based Likelihood Ratio (Sargent 1979) and Wald (Campbell and Shiller 1987) tests. In their recent seminal paper, Bekaert and Hodrick (2001, B & H thereafter) suggest a Lagrange Multiplier (LM) test and show it has better finite sample properties than Wald and Likelihood Ratio based Distance Metric tests, the former of which had been used almost exclusively in the previous literature. The B & H methodology is fast gaining popularity and is adopted, for example, in Bekaert, Wei and Xing (2006), and Sarno, Thornton and Valente (2006).

The present paper extends the B & H methodology by recognising model uncertainty in the specification of the VAR, and then applies it to re-examine the EH for US term structure data. Finite sample inferences drawn to date from the LM test rely on either an i.i.d. bootstrap or a GARCH model of the VAR residuals. Goncalves and Kilian (2004) argue that i.i.d. re-sampling is inaccurate in the presence of conditional heteroskedasticity, which characterizes many financial time series, while GARCH models can suffer from misspecification problems, see e.g. Wolf (2000) and Belsley (2002). To avoid these problems, we propose the application of a wild bootstrap scheme, which permits heteroskedasticity of unknown form while retaining the contemporaneous error correlation. Further, applications of the i.i.d. bootstrap assume a known VAR order, which does not reflect true uncertainty. We not only endogenize the VAR lag length selection using an information criterion, but also supplement this with an application of a multivariate extension of the Godfrey and Tremayne (2005) autocorrelation test, as the performance of the former may not be reliable in the presence of conditional heteroskedasticity.[2] In addition, we introduce a stationarity correction in the VAR and randomize the initial condition.

Our extended method is applied to US term structure data for January 1952 to December 2003. Following Sarno, *et al.* (2006), who report a regime shift around 1982, we consider not only the whole period but also two sub-samples, before and after the monetary targeting period of 1979-1982. These sub-samples allow us to reassess if Campbell and Shiller's (1991) claim that the EH performs better prior to 1978 than in the whole sample applies only because the latter includes potentially different regimes. While the existence of cheap communications and financial market competition may have reduced transactions costs (whose presence may be a factor against the EH) for the recent period, our empirical results do not indicate any noticeable difference in the performance of the theory across the sub-samples. Moreover, we find that the EH tends to be rejected at the short end of the maturity spectrum, in line with much of the previous literature. Nevertheless, the number of rejections is fewer in our analysis, as a result of the greater recognition of uncertainty in our testing methodology.

The paper has five sections. Section 2 outlines the implications of the EH theory for interest rates and discusses tests of the theory in a VAR framework, focusing on the

[2] See Monte Carlo evidence of Backus and Zaman (1998), Kyriazidou (1998) and Ng and Perron (2005) on the performance of various model selection criteria in this situation.

B & H methodology. Our methodological extensions are discussed in Section 3 and applied in Section 4, while Section 5 concludes.

II. Expectations hypothesis theory and tests

According to the EH, a long term interest rate equals the sum of a constant term premium and an average of current and expected future short term interest rates over the life of the long term interest rate. That is, in a linearized version of the EH (see Shiller, 1979)

$$R_{n,t} = \frac{1}{k}\sum_{i=0}^{k-1} E_t R_{m,t+mi} + \pi_{n,m} \ ; \tag{1}$$

where $R_{n,t}$ and $R_{m,t}$ are long and short rates at time t respectively, $E_t R_{m,t+mi}$, $i = 0$, 1, 2,… k-1, is the expectation formed at time t of short rates at $t+mi$ and $\pi_{n,m}$ is the term premium which can vary across maturities but is assumed constant through time. Since the EH places no restriction on $\pi_{n,m}$, this term can be ignored by working with demeaned series. In (1), $k = n/m$, the maturity multiple, is defined to be an integer, with m the maturity of a shorter rate and n the maturity of a longer rate.

This section first discusses single equation and vector approaches to testing the EH, before considering the B & H methodology.

A. Single equation tests

Equation (1) is rarely tested directly, probably due to the empirical results that imply the series are integrated, in which case conventional statistical theory is not appropriate. Rather, another implication of (1) is usually tested, which is based on the ability of the spread between long and short rates to predict future short rate changes after imposing rationality on the expectations. Rationality requires

$$R_{m,t+mi} = E_t R_{m,t+mi} + v_{t+mi}, \tag{2}$$

where v_{t+mi} has zero mean and is orthogonal to the information available at time t. Subtracting $R_{m,t}$ from both sides of equation (1) and imposing rational expectations as in (2) yields probably the most commonly tested equation of the EH, which, after some rearrangement and parameterization, can be written as

$$\sum_{i=1}^{k-1}(1-\frac{i}{k})\Delta^m R_{m,t+mi} = -\pi_{n,m} + \alpha S_{(n,m),t} + w_{(n,m),t} \tag{3}$$

where $\Delta^m R_{m,t+m} = R_{m,t+m} - R_{m,t}$, $S_{(n,m),t} = R_{n,t} - R_{m,t}$ and $w_{(n,m),t}$ is a moving average process of order (n-m). Equation (3) says that the current spread between long and short term rates predicts a cumulative change in shorter term (m-period) interest rate over n periods, and under the null hypothesis of the EH, α should be unity. [3]

However, as Campbell, Lo and MacKinlay (1997) point out, there are several econometric difficulties with the conventional regression approach in this context. Firstly, we lose n-m observations at the end of the sample period. This may be quite serious, as the data available for analysis may be relatively small and n can, for example, be as large as 10 years. Secondly, the error term $\eta_{(n,m),t}$, is a moving average process, so standard errors have to be corrected, for example using the method described in Hansen and Hodrick (1980), Hansen (1982) or Newey and West (1987). But these adjustments do not work well when n-m is not small relative to the sample size (see e.g. Richardson and

[3] Another implication of (1), which is less empirically supported, is that the yield spread predicts the m-period change in the longer-term yield. This is tested (see e.g. Campbell and Shiller, 1991) using $R_{n-m,t+m} - R_{n,t} = \gamma + \beta \frac{m}{n-m} S_{(n,m),t} + v_t$ under the null β is unity.

Stock, 1989, and Hodrick, 1992). Thirdly, the regressor is serially correlated and correlated with lags of the dependent variable, and this can cause finite sample problems as well (see e.g. Mankiw and Shapiro 1986). Moreover, as Thornton (2006) argues, α can be very close to one even under the alternative hypothesis where the EH does not hold.

B. The VAR approach

Problems associated with the single equation method can be avoided using a VAR framework, as suggested in e.g. Sargent (1979) and Campbell and Shiller (1987, 1991). Assuming that there exists a stationary vector stochastic process for $\mathbf{y}_t = \left[\Delta R_{m,t}, S_{(n,m),t}\right]'$ in (3), then the demeaned process for $\mathbf{y}_t$ can be represented as a VAR of order p,[4]

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{u}_t . \tag{4}$$

Further, (4) can be written as a first order VAR in companion form such that $\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{v}_t$, where the companion matrix **A,** of dimension $2p\times 2p$, has the form:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_2 & \mathbf{0} \end{bmatrix}.$$

While $\mathbf{z}_t$ has $2p$ elements, $\mathbf{z}_t = \left[\mathbf{y}'_t, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p+1}\right]'$, $\mathbf{v}_t$ is the $2p$ vector $\left[\mathbf{u}'_t, 0, 0, \ldots, 0\right]'$ which is uncorrelated over time. Thus $\mathbf{z}_t$ summarizes the whole history of $\mathbf{y}_t$.

[4] We avoid using interest rates in levels as that leads to a further stationarity restriction on the VAR parameters and/or the computational time was very long.

Now define vectors $\mathbf{e}_i$, $i = 1, 2$, each of dimension $2p$, with unity in the i^{th} position and zeros everywhere else such that $\mathbf{e}_1'\mathbf{z}_t = \Delta R_{m,t}$ and $\mathbf{e}_2'\mathbf{z}_t = S_{(n,m),t}$. Using these definitions, the spread predicted by the EH and its restrictions on VAR parameters can be shown to be, respectively,

$$S^*_{(n,m),t} \equiv \mathbf{e}_1'\mathbf{A}[\mathbf{I}_{2p} - \frac{m}{n}(\mathbf{I}_{2p} - \mathbf{A}^n)(\mathbf{I}_{2p} - \mathbf{A}^m)^{-1}](\mathbf{I}_{2p} - \mathbf{A})^{-1}\mathbf{z}_t \tag{5}$$

$$\mathbf{e}_2' = \mathbf{e}_1'\mathbf{A}[\mathbf{I}_{2p} - \frac{m}{n}(\mathbf{I}_{2p} - \mathbf{A}^n)(\mathbf{I}_{2p} - \mathbf{A}^m)^{-1}](\mathbf{I}_{2p} - \mathbf{A})^{-1}. \tag{6}$$

The restrictions in (6) are highly non-linear and are predominantly tested by asymptotic Wald tests, even though these have some undesirable properties in finite samples (see e.g. Gregory and Veall, 1985, Dagenais and Dufour, 1991). In particular, the Wald statistic is not invariant to how one specifies the null hypothesis and, potentially, to units of measurement; Shea (1992) provides a numerical example of how one can reach different conclusions on testing algebraically equivalent EH restrictions using Wald tests.

The next two sub-sections describe the B & H methodology, which is designed to avoid the problems of asymptotic Wald tests.

C. B & H methodology: asymptotic inference

B & H (2001) suggest a Lagrange Multiplier (LM) test based on the restricted VAR parameters. On the basis of a Monte Carlo study, they find that the LM test has much better small sample properties than the Wald test in terms of size and power. They also consider the Likelihood Ratio based Distance Metric (DM) test, but prefer the LM test. Since this methodology is relatively new, and is an important part of this study, we now summarize the methodology.

The LM statistic of B & H is based on Hansen's (1982) Generalized Method of Moments (GMM) estimator, which uses the orthogonality condition implied by (2). Defining $\dot{\mathbf{A}} = [\mathbf{A}_1,...,\mathbf{A}_p]'$, and assuming that the DGP is represented by (4), the vector of nonlinear orthogonality conditions can be written as

$$E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})] = \mathbf{0}, \text{ where } \mathbf{x}_t \equiv (\mathbf{y}'_t, \mathbf{z}'_{t-1})', \ \boldsymbol{\theta} = vecr(\dot{\mathbf{A}}).$$

Estimation uses the corresponding sample moment conditions for a sample of size T, namely

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}).$$

It proceeds by selecting $\boldsymbol{\theta}$ to minimize the GMM criterion function

$$J_T(\boldsymbol{\theta}) \equiv \mathbf{g}_T(\boldsymbol{\theta})'\boldsymbol{\Omega}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}), \tag{7}$$

where, assuming the VAR of (4) is correctly specified with $\mathbf{u}_t$ uncorrelated, the weighting matrix, $\boldsymbol{\Omega}_T$, is a consistent estimate of[5]

$$\boldsymbol{\Omega} \equiv E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})']. \tag{8}$$

Writing the null hypothesis of (6) as

$$H_0 : \mathbf{c}(\boldsymbol{\theta}_0) = \mathbf{0}, \tag{9}$$

[5] Notice that, in their equation (15), B & H use the more general GMM expression $\boldsymbol{\Omega} \equiv \sum_{h=-\infty}^{\infty} E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})\mathbf{g}(\mathbf{x}_{t-h}, \boldsymbol{\theta})']$. This collapses to (8), due to the assumption of uncorrelated VAR disturbances, which can be consistently estimated using a multivariate version of White's (1980) heteroskedasticity consistent covariance matrix, as given in equation (A4) of Appendix A. Defining $\boldsymbol{\eta}_t$ as the VAR residual vector, then we estimate $\boldsymbol{\Omega}$ as

$\boldsymbol{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})' = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\eta}_t \otimes \mathbf{z}_{t-1})(\boldsymbol{\eta}_t \otimes \mathbf{z}_{t-1})' = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{\eta}'_t \otimes \mathbf{z}_{t-1}\mathbf{z}'_{t-1})$, which is (A4).

where $\mathbf{c}(\boldsymbol{\theta}_0)$ is a $2p$ dimensional vector, the Lagrangian for the constrained GMM maximization problem is

$$L(\boldsymbol{\theta},\boldsymbol{\gamma}) = -\frac{1}{2}\mathbf{g}_T(\boldsymbol{\theta})'\boldsymbol{\Omega}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}) - \mathbf{c}(\boldsymbol{\theta})'\boldsymbol{\gamma} \tag{10}$$

where $\boldsymbol{\gamma}$ is a vector of Lagrange multipliers and $\boldsymbol{\Omega}_T$ is obtained from (8) using the sample mean in place of the expectation and replacing unknown parameters by their estimates. Since direct maximization of (10) is difficult, B & H (2001) adopt the approach of Newey and McFadden (1994) to derive a constrained consistent estimator starting from an initial unconstrained consistent one.

Denoting an estimate of the matrix of restricted parameters satisfying (6) as $\overline{\mathbf{A}}$ and $\overline{\boldsymbol{\theta}} = vecr(\overline{\mathbf{A}})$, a Taylor series expansion to the non-linear first order conditions for (10) yields

$$\sqrt{T}\mathbf{g}_T(\overline{\boldsymbol{\theta}}) \approx \sqrt{T}\mathbf{g}_T(\boldsymbol{\theta}_0) + \mathbf{G}_T\sqrt{T}(\overline{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\,; \tag{11}$$

$$\sqrt{T}\mathbf{c}_T(\overline{\boldsymbol{\theta}}) \approx \sqrt{T}\mathbf{c}_T(\boldsymbol{\theta}_0) + \mathbf{C}_T\sqrt{T}(\overline{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\,, \tag{12}$$

where $\mathbf{c}_T(\boldsymbol{\theta})$ is the sample (mean) counterpart of $\mathbf{c}(\boldsymbol{\theta})$, and $\mathbf{G}_T$ and $\mathbf{C}_T$ are gradients, with respect to $\boldsymbol{\theta}$, of the sample orthogonality conditions and the vector of constraints, respectively. Using $\mathbf{c}_T(\boldsymbol{\theta}_0) = 0$ under the null hypothesis and substituting (11) and (12) into the first-order conditions, B & H obtain

$$\overline{\boldsymbol{\theta}} \approx \boldsymbol{\theta}_0 - \mathbf{D}_T^{-1/2}\mathbf{M}_T\mathbf{D}_T^{-1/2}\mathbf{G}_T'\hat{\boldsymbol{\Omega}}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}_0) - \mathbf{D}_T^{-1}\mathbf{C}_T'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{c}_T(\boldsymbol{\theta}_0)\,, \tag{13}$$

$$\overline{\boldsymbol{\gamma}} \approx -(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{G}_T'\hat{\boldsymbol{\Omega}}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}_0) + (\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{c}_T(\boldsymbol{\theta}_0)\,, \tag{14}$$

where $\mathbf{D}_T \equiv \mathbf{G}_T'\hat{\boldsymbol{\Omega}}_T^{-1}\mathbf{G}_T$ and the idempotent matrix $\mathbf{M}_T$ is defined as $\mathbf{M}_T \equiv \mathbf{I} - \mathbf{D}_T^{-1/2}\mathbf{C}_T'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{C}_T\mathbf{D}_T^{-1/2}$.

To use these results for estimation, let $\tilde{\boldsymbol{\theta}}$ represent an initial consistent unconstrained estimate. Then constrained estimates, $\bar{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\gamma}}$, are obtained by iterating (13) and (14), substituting $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$ to derive a second constrained estimate, and so forth until the constraint is satisfied, i.e. $\mathbf{c}_T(\boldsymbol{\theta}) = \mathbf{0}$.[6] This yields the constrained estimate, together with the Lagrange Multipliers, which under the null hypothesis of EH has asymptotic distribution

$$\sqrt{T}\bar{\boldsymbol{\gamma}} \rightarrow \mathbf{N}\left[\mathbf{0}, (\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\right]. \tag{15}$$

The constrained parameter estimate is not equal to the unconstrained one when the constraints in (9) significantly affect the value of the GMM objective function (7). From (15), under the null hypothesis the LM test statistic is

$$T\bar{\boldsymbol{\gamma}}'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')\bar{\boldsymbol{\gamma}} \rightarrow \chi^2(2p), \tag{16}$$

where p is the lag length of the VAR.

B & H (2001) also consider the DM test statistic $T\mathbf{g}_T(\bar{\boldsymbol{\theta}})'\boldsymbol{\Omega}_T^{-1}\mathbf{g}_T(\bar{\boldsymbol{\theta}}) \rightarrow \chi^2(2p)$ and the Wald statistic $T\mathbf{c}_T(\tilde{\boldsymbol{\theta}})'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{c}_T(\tilde{\boldsymbol{\theta}}) \rightarrow \chi^2(2p)$. Note that the Wald test is based on the unrestricted estimates, the LM on the restricted estimates and the DM on both.

D. B & H methodology: finite sample inference

It is well known that estimated VAR parameters, although consistent, are biased in finite samples (see e.g. Tjostheim and Paulsen, 1983, Bekaert, Hodrick and Marshall, 1997). B

[6] Note that we update $\mathbf{C}_T, \mathbf{D}_T, \mathbf{G}_T$ and $\boldsymbol{\Omega}_T$ at each iteration step to speed up the convergence. Preliminary results show doing that does not alter much the final conclusion. In our application the tolerance for convergence is set to 10^{-8}.

& H suggest correcting for this bias using the bootstrap. More specifically, having specified the appropriate VAR order p and obtained the estimated unconstrained VAR parameter matrix, $\hat{\mathbf{A}}$, they use $\hat{\mathbf{A}}$ and an i.i.d bootstrap of the estimated residuals to generate b samples of artificial data. Each of these yields an unconstrained estimate $\hat{\mathbf{A}}_{M,i}$, i = 1, …, b, and the bias, $\mathbf{B}$, is estimated as $\hat{\mathbf{B}} = \hat{\mathbf{A}} - \frac{1}{b}\sum_{i=1}^{b} \hat{\mathbf{A}}_{M,i}$. Finally, bias-corrected estimates are obtained as $\hat{\mathbf{A}}_c = \hat{\mathbf{A}} + \hat{\mathbf{B}}$.

To obtain bias-corrected parameter estimate that satisfies the null hypothesis, they use the bias corrected unconstrained VAR parameter estimate and an i.i.d. bootstrap of the residuals to simulate a very long series (70,000 observations plus 1,000 starting values that are discarded), which is then subjected to the iterative estimation scheme of (13) and (14). This bias corrected and constrained estimate of **A** is then used to calculate the LM test statistic and to conduct inference through the asymptotic result in (16).

B & H also apply finite sample inference directly. Indeed, it is well documented that large sample inference can be misleading for finite samples (see e.g. Mankiw and Miron, 1986, or Horowitz, 2001). The estimate of the bias-corrected constrained parameter matrix, $\overline{\mathbf{A}}_c$, obtained from (13), combined with an i.i.d bootstrap of the unrestricted residuals, is used as the bootstrap DGP to generate artificial null data sets of the actual sample size, plus 1000 observations that are discarded to attenuate the start-up effect. Using the same VAR order p as obtained from the actual data, iterations are performed on (13) and (14) for each bootstrap data set to obtain an LM statistic from (16). This is repeated a large number of times, with the empirical *p-value* then computed as the proportion of bootstrap LM test statistics that are larger than or equal to the sample statistic obtained from the observed data.[7]

B & H recognise the limitations of the i.i.d. bootstrap, and also examine test statistics obtained where conditional heteroskedasticity is allowed in the VAR residuals through a specific parametric factor GARCH model. In this case, the observed LM test statistic is compared to percentiles of the null distribution obtained using the estimated GARCH model rather than i.i.d. bootstrap of the residuals.

III. Extensions to B & H Methodology

In this section we suggest several extensions, motivated by recent developments in the model specification and bootstrap literature, to the B & H methodology in order to encompass more general situations. Our three principal extensions are: i) the use of wild bootstrap to allow for conditional heteroskedasticity in the residuals of the estimated model, which does not require any a priori parameterization, ii) imposition of a stationarity correction and randomizing the initial condition in the bias correction procedure, iii) interactive use of an endogenous lag order selection rule and a vector autocorrelation test, with the restricted residuals used for finite sample inference.

After outlining the wild bootstrap procedure, we discuss how our extensions are employed to improve the bias correction method of B & H and to conduct finite sample inference.

[7] Finite sample inferences from DM and Wald tests follow the same procedure, except the latter does not require iterations of (13) and (14).

A. The wild bootstrap

The EH itself places no restriction on the distribution of the VAR disturbances. It is, however, common that the residuals from estimated models exhibit volatility clustering, especially when financial time series are used (see e.g. Bollerslev, Chou and Kroner, 1992). As discussed above, the bias correction method in B & H (2001) relies on i.i.d. residuals, thus assuming away the presence of heteroskedasticity, whether of the unconditional or conditional form. However, they acknowledge possible conditional heteroskedasticity in the residuals for the purpose of finite sample inference, through the use of a VAR-GARCH model in addition to using i.i.d. bootstrap. But there is no solid reasoning behind why this specific form of volatility clustering model is used (Goncalves and Kilian, 2004), and even if this class of GARCH models is appropriate the precise form of the GARCH model is unknown, leading to the possibility of different results for different specifications (Wolf 2000 and Belsley, 2002). In contrast, we avoid these problems through using the wild bootstrap for both bias correction and to obtain empirical *p-values*.[8]

The wild bootstrap we use was developed in Liu (1988) following recommendations in Wu (1986) and Beran (1986). The particular form we employ is the recursive design wild bootstrap, which has better small sample properties than several other resampling schemes and is comparable with the i.i.d. bootstrap when the errors are indeed i.i.d.; see, e.g., Goncalves and Kilian (2004). Therefore, there appears to be minimal cost in applying the wild bootstrap when the disturbances satisfy the i.i.d. assumption.

---

[8] By this we mean that the wild bootstrap allows the possibility of structural breaks in the disturbance variance-covariance matrix, as well as allowing for conditional heteroscedasticity of the multivariate GARCH form.

For given VAR parameter matrices $\mathbf{A}_i$, $i=1...p$, and corresponding disturbance vector $\mathbf{u}_t$, a recursive design wild bootstrap sample is generated as $\mathbf{y}_t^* = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i}^* + \mathbf{u}_t^*$, $\mathbf{u}_t^* = \omega_t \mathbf{u}_t$, $t = 1,..., T$, in which the scalar random variable ω_t satisfies $E(\omega_t) = 0$ and $E(\omega_t^2) = 1$. Following the evidence of performance in recent Monte Carlo studies of the wild bootstrap (Davidson and Flachaire, 2001, Godfrey and Orme, 2004, Godfrey and Tremayne, 2005), we specify ω_t as having the Rademacher distribution, which takes the possible values of negative and positive unity with equal probabilities.

B. Bias correction

The bias correction of the VAR coefficient estimates is an important part of the B & H methodology. We develop B & H bias correction procedure in three ways.

Firstly, we introduce a stationarity correction, which is important because the asymptotic validity of B & H methodology relies on it. Interest rates are assumed to be $I(0)$ in B & H (2001) and Sarno, Thornton and Valente (2006), the latter of which provides some unit root test results that support this assumption, and they include the level of interest rates in their VAR. Nevertheless, even if they are differenced, as in this study, the random nature of the bias correction does not guarantee that the bias-corrected, companion-form VAR coefficient matrix is stable.

Our procedure ensures $\lambda_{\max}(\hat{\mathbf{A}}_c) < 1$, where $\lambda_{\max}$ is the largest eigenvalue of the estimated companion matrix after bias correction, through an additional step suggested by Kilian (1998b). This step computes $\lambda_{\max}(\hat{\mathbf{A}}_c)$ after bias adjustment. If $\lambda_{\max}(\hat{\mathbf{A}}_c) \geq 1$, we set $\hat{\mathbf{B}}_i = \hat{\mathbf{B}}$, $\delta_i = 1$ and $i = 1$, then define $\hat{\mathbf{B}}_{i+1} = \delta_i \hat{\mathbf{B}}_i$ with $\delta_{i+1} = \delta_i - 0.001$. Finally, we set

$\hat{\mathbf{A}}_c = \hat{\mathbf{A}}_{c,i}$ after iterating on $\hat{\mathbf{A}}_{c,i} = \hat{\mathbf{A}} + \hat{\mathbf{B}}_i$ $i = 1, 2, \ldots$ until $\lambda_{\max}(\hat{\mathbf{A}}_{c,i}) < 1$. The adjustment has no effect asymptotically and does not restrict the parameter space of the OLS estimator, since it does not shrink the OLS estimate $\hat{\mathbf{A}}$ itself, but only its bias estimate.

Secondly, we randomize the initial conditions. B & H discard the first p observations in each of the 100000 bootstrap replications in order to attenuate the start-up effect. However, as p can be one, this may not fully account for the uncertainty associated with the initial conditions. We therefore follow the suggestion of Stine (1987), by splitting the observed data into $T - p + 1$ overlapping blocks of length p and one of these is selected randomly as the starting point.

Finally, to bias correct the constrained VAR coefficient matrix used to generate empirical distribution of the test statistics, B & H (2001) use the i.i.d. bootstrap of the residuals and the bias corrected unconstrained VAR coefficients to generate a large sample of 70000 observations, which is subjected to the iterative process. This procedure seems ad hoc, and it is unclear whether any advantage can be gained by the artificial generation of a long time series using an i.i.d. bootstrap in the presence of conditional heteroskedasticity. Instead, we subject the actual data and the bias corrected unconstrained coefficient estimates to the iterative process directly, since Newey and McFadden (1994) show that consistency of the estimator is sufficient for the validity of their expansion.

C. Finite sample inference

In addition to employing the wild bootstrap rather than an i.i.d. bootstrap, our finite sample inference procedure differs from B & H in a number of respects.

Firstly, the B & H methodology assumes the lag order is unknown when the VAR of (4) is specified, but when obtaining an empirical *p value* the lag order is treated as known to be that specified from the actual data. However, it is often emphasized that the bootstrap world should reflect the actual world (see, e.g., Li and Maddala, 1996), and ignoring the uncertainty involved in determining the true lag order in finite samples might lead to spurious inference. Therefore, our procedure separately estimates the lag order for every bootstrap dataset generated, employing the same lag selection criterion as that used for the actual dataset. Although there is no difference asymptotically between endogenizing lag selection or not, since every consistent model selection criterion will then choose the right lag length almost surely, Kilian (1998a) shows that endogenous lag selection improves finite sample inference for impulse response analysis.

By the same argument, when estimating (4) on the artificial restricted data, the resulting estimated VAR coefficients should be bias corrected.

Secondly, we employ a more flexible approach in choosing the appropriate lag length *p* of the assumed VAR-DGP, which is important not only because the asymptotic distributions of test statistics depend on it, but also because a necessary condition for the validity of the bootstrap is the absence of autocorrelation in the VAR residuals. Although B & H (2001) provide residual-autocorrelation test results after choosing the VAR lag length by SIC to argue that autocorrelation is not a concern, the test they use is univariate, applied to each individual equation of the VAR. Thus, their test potentially leads to the problem of mass significance, as discussed in Edgerton and Shukur (1999), and also omits the possibility of cross-equation residual autocorrelation. Moreover, the SIC model selection rule, like many others, implicitly relies on conditional homoskedasticity and is typically derived under conditional normality. The Monte Carlo studies of Backus and Zaman (1998), Kyriazidou (1998) and Ng and Perron (2005), examining the performance of various model selection criteria, greatly reduces our confidence that we can rely on SIC to choose appropriate lag lengths in all 48 models we examine, representing the full spectrum of the term structure, in contrast to only two in B & H (2001).[9] We therefore employ a multivariate extension of the autocorrelation test robust to conditional heteroskedasticity of Godfrey and Tremayne (2005), some details of which are provided in Appendix A, to the residuals of the model, the first of which is specified by SIC, and increase the lag length by one if there is any evidence of autocorrelation. This is repeated until that lag length which ensures the absence of autocorrelation up to 12th order.

Finally, since studies of Davidson and MacKinnon (1985) and Godfrey and Orme (2004) find that the use of the restricted residuals provide improvements in finite sample properties compare with unrestricted ones, we use the former. That is, the bootstrap employs the residuals $\overline{\mathbf{u}}_t = \mathbf{y}_t - \sum_{i=1}^{p} \overline{\mathbf{A}}_i \mathbf{y}_{t-i}$. Sarno *et al.* (2006) also use the restricted VAR residuals in their study.

As the computational costs of bias correction, model identification and application of autocorrelation tests at each bootstrap iteration is high, the number of iterations is 1000 for the bias correction (that is *b*) and 5000 for the empirical distribution of LM statistics (that is *d*) in this context, not 100000 and 25000 respectively as in B & H (2001), and Sarno *et al.* (2006).

---

[9] Campbell and Shiller (1991) and Hardouvellis (1994) assume that the data is known to be generated by VAR(4), Thornton (2006) use SIC, while Shea (1992), and Sarno *et al* (2006) use AIC. Basci and Zaman (1998), Kyriazidou (1998) recommend using SIC when the VAR residuals are suspected to be contaminated with conditional heteroskedasticity.

IV. Empirical Results[10]

In this section we describe the data we use and provide empirical results from our extended B & H methodology. All inferences are made at a 5% significance level.

A. Data and preliminary results

We use the continuously compounded zero coupon yield curve dataset of Sarno *et al.* (2006). This is an update of the McCulloch and Kwon (1993) dataset, which is used in many studies, including Campbell and Shiller (1991), Campbell (1995) and Thornton (2006). The full coverage is from January 1952 to December 2003. However, as there is evidence that the EH performed better prior to 1978 (Campbell and Shiller, 1991) and of a structural break in the VAR parameters around 1982 (Sarno, *et al.,* 2006), with these dates roughly coinciding with the introduction and abandonment of the Fed's reserve targeting policy, respectively, we consider not only the whole, but also two sub-samples, January 1952 to December 1978 and January 1982 to Dec 2003. This sample splitting can also be motivated by the hypothesis that cheaper communication technology and competition in the financial market implies shrinking transaction costs over time, favouring the EH in the second sub-period. The maturities considered range from one month to a maximum of 20 months (10 years).

Panel A of Table 1 shows the VAR lag orders chosen by SIC, separately selected for the entire sample and for the two sub-samples, for all maturity pairs considered[11].

[10] The authors would like to thank Dick van Dijk, Ruud Koning, Markus Kraetzig and especially Daniel Thornton for making their computer codes available on the Internet, modifications of which are used in this study. We also thank Daniel Thornton for providing the dataset.

Tests were conducted for the presence of first order conditional heteroskedasticity in the VAR residuals using the multivariate ARCH-LM test statistic described in Doornik and Hendry (1997). Strong evidence of conditional heteroskedasticity was found in almost all cases, both in the unrestricted VAR residuals and also in the residuals after imposition of the EH restriction, pointing to the need to take account of conditional heteroskedasticity for valid inference.[12] Since Goncalves and Killian (2004) show that the wild bootstrap has performance that is little inferior to than that of the standard bootstrap even if the errors are indeed i.i.d., all subsequent analyses are based on the wild bootstrap.

Table 1 here

Table 2 provides LM test statistics for the null of no autocorrelation against the first order autocorrelation in the VAR residuals and their asymptotic and empirical *p-values*, when applied with the VAR order indicated by SIC.[13] It is notable that the discrepancy between the asymptotic and empirical *p-values* is relatively small and the rejections of the null tend to occur almost exclusively in the first sub-period, rather than the second sub- or the whole samples. However, when the null is tested against higher order (up to 12th order) autocorrelation, the adequacy of SIC in ensuring that all dynamic

---

[11] The maximum lag length considered for the VAR order is $p_{\max} = 12(T/100)^{1/4}$ which is supported in the Monte Carlo study of Schwert (1989).

[12] The *p-value* for this test was almost always 0.00. The only exceptions to significance at 5% occurred for some maturity pairs at the longest end of the maturity spectrum, namely for maturity pair 12 and 120 months in the first sub-sample and 60 and 120 months in the second sub-sample. Full results are available from the authors on request.

[13] See Appendix A for the description of the test and Godfrey and Tremayne (2005) for the use of wild bootstrap in estimating empirical *p values*.

mean relationships are captured by the VAR becomes more suspect. The result of increasing the lag length of the VAR chosen by SIC when there is any evidence of autocorrelation is reported in Panel B of Table 1, with the cases highlighted where the final lag lengths are different from the initial SIC choice.[14] The fewest discrepancies are found over the whole sample, as one would expect given that SIC is consistent, but it is also notable across all samples that the lag length indicated by SIC often does not account for all autocorrelation when the longer maturity considered is relatively short.

Overall, the results indicate the danger of solely relying on SIC as the single lag selection rule across these models. Therefore, our empirical results in sub-section 4.3 are based on the augmented lag lengths.

Table 2 here

The importance of the stability correction we employ is shown in Table 3, where Panel A reports the number of iterations required to ensure stability for the unrestricted bias-corrected companion-form VAR parameters and Panel B gives the average number of iterations for the restricted bias-corrected parameters.[15] It is clear that, without this correction, many bias-corrected estimates would have produced unstable models, invalidating inference based on the assumption of stability. There is also a marked pattern across the different sample periods. The modification has almost no effect in the second sub-sample, but the stability correction is employed much more often in other periods, especially in the first sub-sample.

[14] This is a conservative strategy with respect to autocorrelation as the overall level of significance is less than 5% we use for individual tests.

[15] It may be noted that the bias corrected restricted VAR parameter in Panel A of Table IV in B & H (2001) is unstable, i.e. the maximum eigenvalue is 1.078.

Table 3 here

B. Effect of lag length selection in the bootstrap

As discussed above, our bootstrap inference is designed to capture the uncertainty faced in the specification of the lag order in a VAR, while ensuring that the disturbances are uncorrelated. To indicate the effects of the various lag length treatments, Figure 1 shows the empirical distributions obtained for the various LM test statistic for the maturity pair of 1 and 3 years using data from January 1952 to December 1978. The asymptotic $\chi^2(6)$ distribution is also included for comparative purpose.[16] When the VAR lag length is exogenous in the bootstrap world (that is, set equal to that estimated from the actual data) the empirical distribution closely matches the asymptotic one, implying there is little gain from the bootstrap in improving finite sample inference. However, as discussed above, Kilian (1998a) and others argue this lag treatment does not reflect the true uncertainty associated with choosing the lag length, which implies that this is not the appropriate finite sample distribution.

As can be seen from the right-hand panel of Figure 1, when the lag length selection is endogenized using SIC, the empirical distribution of the lag lengths in the bootstrap never extends beyond the lag estimated from the observed data. More specifically, although only 50% of the lags are estimated as 3, in contrast to 100% in the exogenous lag length procedure, the remaining half in the former case cluster on either 1 or 2 lags and no case exceeds 3. This is not specific to this maturity pair; indeed, the lag distributions for all maturity pairs either cluster at the one estimated from the actual observed data or extend only to that lag length. This reduction in uncertainty in the

[16] The results for all maturity pairs can be obtained from the authors. The degrees of freedom is 6 as the lag is estimated to be 3, see Panel B, Table 1.

bootstrap has the consequence that the empirical distribution of the resulting LM statistic lies closer to the origin than that of the exogenous lag distribution. Thus, this procedure would reduce the number of rejections of the EH, compared to the use of either the asymptotic distribution or an exogenous lag length bootstrap.

Figure 1 here

However, applying our sequential lag ordering strategy, i.e. first choosing the lag order by SIC and modifying it when the autocorrelation test detects any evidence of autocorrelation up to 12th order, the lag distribution in the right-hand pane of Figure 1 becomes more symmetric around the actual lag in the bootstrap DGP. This increases the uncertainty associated with LM statistic and shifts its distribution to the right compared with the use of SIC alone.

Different lag length treatments in the bootstrap have important empirical consequences. In general, the proposed lag selection method will produce more favourable results for the EH than others we consider. For the specific maturity pair example illustrated in the distributions of Figure 1, the LM statistic estimated from the data is 13.29, with the corresponding 95% quantiles for the four graphed distributions being 12.14, 11.56, 15.08 and 12.59. In this case, the EH would be rejected using all distributions, except when the endogenous lag length selection rule is supplemented with the autocorrelation test. Nevertheless, we believe that our extended procedure is to be preferred because it replicates the procedure applied in analysis of the observed data and hence better captures the uncertainty associated with the value of the test statistic. This is also consistent with Hansen (2005) who emphasizes the importance of model uncertainty in empirical inference.

C. Results for EH tests

Finally, we turn to the results of our extended LM test of the EH for the US term structure, which are provided in Table 4, with cases of rejection of the EH highlighted.[17] As can be seen from the table, the LM test does not always work, sometimes leading to non-convergence of the iterative procedure or instability of the restricted VAR coefficient matrix. The former problem is also reported in Sarno *et al*. (2006) and the latter is found in Bekaert *et al*. (2006). In our case, most non-convergences occur at the shortest end of the term structure and when the VAR lag lengths resulting from the application of the autocorrelation test are particularly large, such as 13 and 9 (see Panel B of Table 1). To obtain convergence in such cases, a restriction of maximum lag length of 5 is imposed.[18]

Table 4 here

The first substantive conclusion emerging from the finite sample LM test is that the EH tends to be rejected at the short end of the term structure, consistent with Campbell and Shiller (1991) and Sarno *et al.* (2006), but the number of rejections are much less often. In particular, the former study rejects the EH whenever the maturity *n* of the longer-term rate is below 3 or 4 years using data from January 1952 to February 1987 and the latter study, which uses exactly the same data and sample periods, provides rejections whenever *n* is less than 2 years in most recent sub-period.

Finally, it should be noted that the use of the asymptotic distribution would result in substantially more rejections of the EH than the finite sample distribution, particularly over the whole sample period. Figure 1 also implies that this would also be true in

[17] Results from DM and Wald tests are not reported to conserve space but available from the authors upon request.

[18] Using a 1% level for the autocorrelation test results in lower lag lengths.

relation to the use of exogenously specified lags, emphasizing again the importance of our extensions to the B & H methodology when testing the expectations hypothesis for the term structure.

V. Conclusion

This paper extends the vector autoregressive model (VAR) based expectations hypothesis tests of term structure considered in Bakaert and Hodrick (2001) by relaxing certain assumptions on the VAR model specification. Firstly, we use the wild bootstrap to allow for conditional heteroskedasticity in the VAR residuals without imposing any strict parameterization. Secondly, when making finite sample inferences, we not only endogenize the model selection procedure but also supplement this with an autocorrelation test, employ the restricted not the unrestricted VAR residuals and randomize the initial condition in the bootstrap replications to reflect the true uncertainty. Finally, a stationarity correction is introduced in order to take account of the possibility of obtaining explosive VAR parameter estimates after adjustment for finite sample bias.

When the modified B & H methodology is applied to an extensive US zero coupon term structure data ranging from 1 month to 10 years, we do not find that the EH performs noticeably differently across the sub-samples, namely before and after the Fed's monetary policy change. Nevertheless, we do find the EH is rejected at the short end of the maturity spectrum, in line with Campbell and Shiller (1991) and Sarno *et al.* (2006), although the rejections are less frequent, and also, interestingly, we reject the EH at the longest end over the January 1952 to December 1978 sub-sample. However, overall, our results indicate that the EH provides a reasonable description of the term structure relationship in the US from the period since 1952, provided that separate pre-1979 and post-1982 sub-samples are employed and also provided that the shorter maturity of interest is at least three months.

References

Balduzzi, P., Bertola, G. and Foresi, S. "A Model of Target Changes and the Term Structure of Interest rates", *Journal of Monetary Economics* 39 (1997), 223- 249.

Basci, S. and Zaman, A. "Effects of Skewness and Kurtosis on Model Selection Criteria", *Economics Letters* 59 (1998), 17-22.

Bekaert, G., Hodrick, R.J. and Marshall, D.A. "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates", *Journal of Financial Economics* 44 (1997), 309-348.

________________________, "Peso Problem Explanations for Term Structure Anomalies", *Journal of Monetary Economics* 48 (2001), 241-270.

Bekaert, G. and Hodrick, R.J. "Expectations Hypotheses Tests", *Journal of Finance* 56 (2001), 1357-1394.

Bekaert, G., Wei, M. and Xing, Y. "Uncovered Interest Rate Parity and the Term Structure", *Journal of International Money and Finance,* forthcoming.

Belsley, D.A. "An Investigation of an Unbiased Correction for Heteroskedasticity and the Effects of Misspecifying the Skedastic Function", *Journal of Economic Dynamics & Control* 26 (2002), 1379–1396.

Beran, R. Discussion of "Jackknife Bootstrap and Other Resampling Methods in Regression Analysis" by Wu, C.F.J., *Annals of Statistics* 14 (1986), 1295- 1298.

Bollerslev, T., Chou, R.Y. & Kroner, K.F. "ARCH Modelling in Finance", *Journal of Econometrics* 52 (1992), 5- 59.

Campbell, J.Y. and Shiller, R.J. "Cointegration and Tests of Present Value Models", *Journal of Political Economy* 95 (1987), 1062-1088.

________________________ "Yield Spreads and Interest Rate Movement: A Bird's Eye View", *Review of Economic Studies* 58 (1991), 495-514.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C.. *The Econometrics of Financial Markets*, Princeton University Press: New Jersey (1997).

Campbell, J.Y. "Some Lessons from the Yield Curve", *Journal of Economic Perspectives* 9 (1995), 129-152.

Dahlquist, M. and Jonsson, G. "The Information in Swedish Short-maturity Forward Rates", *European Economic Review* 39 (1995), 1115-1131.

Dagenais, M.G. and Dufour, J.M. "Invariance, Non-linear Models, and Asymptotic Tests", *Econometrica* 59 (1991), 1601- 1615.

Davidson, R. and MacKinnon, J.G. "Heteroskedasticity Robust Tests in Regression Directions", Ann. De'INSEE 59/60 (1985), 183- 218,

Davidson, R. and Flachaire, E. "The Wild Bootstrap, Tamed at Last", Working Paper, Darp58 (2001), STICERD, LSE.

Doornik, J.A. and Hendry, D.F. *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, International Thomson Business Press: London (1997).

Edgerton, D. and Shukur, G. "Testing Autocorrelation in a System Perspective", *Econometric Reviews* 18(1999), 343- 386.

Evans, M.D.D. and Lewis, K.K. "Do Stationary Risk Premia Explain It All?", *Journal of Monetary Economics* 33 (1994), 285- 318.

Froot, K.A. "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates", *Journal of Finance* 44 (1989), 283-305.

Gerlach, S. and Smets, F. "The Term Structure of Euro-rates: Some Evidence in Support of the Expectations Hypothesis", *Journal of International Money and Finance* 16 (1997), 305-321.

Goncalves, S. and Kilian, L. "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form", *Journal of Econometrics* 123 (2004), 89- 120.

Godfrey, L.G. and Orme, C.D. "Controlling the Finite Sample Singificance levels of Heteroskedasticity-robust Tests of Several Linear Restrictions on Regression Coefficients", *Economics Letters* 82 (2004), 281-287.

Godfrey, L.G., and Tremayne, A.R. "The Wild Bootstrap and Heteroskedsticity- Robust Tests for Serial Correlation in Dynamic Regression Models", *Computational Statistics & Data Ananlysis* 49 (2005), 377- 395.

Gregory, A.W. and Veall, M.R. "Formulating Wald Tests of Nonlinear Restrictions", *Econometrica* 53(1985), 1465- 1468.

Hafner, C.M. and Herwartz, H. "Testing for Vector Autoregressive Dynamics under Heteroskedsticity", Econometric Institute Report EI 2002-36, Erasmus University Rotterdam.

Hall, A. D., Anderson, H. M. and Granger, C. W. J. "A Cointegration Analysis of Treasury Bill Yields", *Review of Economics and Statistics* 74 (1992), 116-126.

Hansen, L.P. "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 50 (1982), 1029-1054.

Hansen, L. P. and Hodrick, R.J. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis", *Journal of Political Economy* 88 (1982), 829- 853.

Hansen, B.E. "Challenges for Econometric Model Selection", *Econometric Theory* 21 (2005), 60-68.

Hardouvelis, G.A. "The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries", *Journal of Monetary Economics* 33 (1994), 255-283.

Hodrick, R. "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement", *Review of Financial Studies* 5 (1992), 357- 386.

Horowitz, J.L. "The Bootstrap", in Heckman, J.S. and Leamer, E. eds. *Handbook of Econometrics* 5 (2001), North-Holland, Amsterdam, The Netherlands, 3159- 3228.

Kilian, L. “Accounting for Lag Order Uncertainty in Autoregressions: The Endogenous Lag Order Bootstrap Algorithm”, *Journal of Time Series Analysis* 19(1998a), 532- 548.

__________ “Small-Sample Confidence Intervals for Impulse Response Functions”, *Review of Economics and Statistics* 80 (1998b), 218- 230.

Kyriazidou, E. “Testing for Serial Correlation in Multivariate Regression Models”, *Journal of Econometrics* 86 (1998), 193- 220.

Lanne, M. “Near Unit Roots and the Predictive Power of Yield Spreads for Changes in Long-Term Interest Rates”, *Review of Economics and Statistics* 81(1999), 393- 398.

Li, H. and Maddala, G.S. “Bootstrapping Time Series Models”, *Econometric Reviews* 15 (1996), with discussion, 115- 158.

Liu, R.Y. “Bootstrap Procedure under Some Non-I.I.D. Models”, *Annals of Statistics* 16 (1988), 1696- 1708.

Longstaff, F. A. “The Term Structure of Very Short-term Rates: New Evidence for the Expectations Hypothesis”, *Journal of Financial Economics* 58 (2000), 397-415.

Mankiw, N.G. and Miron, J.A. “The Changing Behavior of the Term Structure of Interest Rates”, *Quarterly Journal of Economics* 101(1986), 211-228.

Mankiw, N.G. and Shapiro, M.D. “Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models”, *Economics Letters* 20 (1986), 139-145.

McCulloch, J.H. and Kwon, “U.S. Term Structure Data, 1947 - 1991”, Ohio State University Working Paper 93-6.

Newey, W. K. and McFadden, D. L. “Large Sample Estimation and Hypothesis Testing”, in Engle, R.F. and McFadden, D.L. eds. *Handbook of Econometrics* 4 (1994). Elsevier Science. Amsterdam, The Netherlands, 2111-2245.

Newey, W. K. and West, K. D. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”, *Econometrica* 55 (1987), 703-708.

Ng, S. and Perron, P. “A Note on the Selection of Time Series Models”, *Oxford Bulletin of Economics and Statistics* 67(2005), 115- 134.

Richardson, M. and Stock, J.H. “Drawing Inferences from Statistics Based on Multiyear Asset Returns”, *Journal of Financial Economics* 25 (1989), 323- 348.

Roberds, W., Runkle, D. and Whiteman, C. H. “A Daily View of Yield Spreads and Short-term Interest Rate Movements”, *Journal of Money, Credit and Banking* 28 (1996), 35-53.

Roberds, W. and Whiteman, C.H. “Endogenous Term Premia and Anomalies in the Term Structure of Interest Rates: Explaining the Predictability Smile”, *Journal of Monetary Economics* 44(1999), 555-580.

Rudebusch, G.D. "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," *Journal of Monetary Economics* 35 (1995), 245-274.

Sargent., T. “A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure”, *Journal of Monetary Economics* 5 (1979), 133- 143.

Sarno, L., Thornton, D.L. and Valente, G. “The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields”, *Journal of Financial and Quantitative Analysis,* forthcoming.

Schwert, G. “Tests for Unit Roots: A Monte Carlo Investigation”, *Journal of Business & Economic Statistics* 7 (1989), 147- 159.

Shea, G. “Benchmarking the Expectations Hypothesis of the Interest Rate Term Structure: An Analysis of Cointegrating Vectors”, *Journal of Business & Economic Statistics*, 10 (1992), 347- 66.

Shiller, R.J. “The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure”, *Journal of Political Economy* 87 (1979), 1190- 1219.

Shiller, R. J., Campbell J. Y. and Schoenholtz K. L. "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates", *Brookings Papers on Economic Activity* 1 (1983), 173-217.

Stine, R.A. "Estimating Properties of Autoregressive Forecasts", *Journal of the American Statistical Association* 82 (1987), 1072- 1078.

Taylor, M. P. "Modelling the Yield Curve", *Economic Journal* 102 (1992), 524-537.

Thornton, D.L. "Tests of the Expectations Hypothesis: Resolving Campbell-Shiller Paradox", *Journal of Money, Credit and Banking* 38 (2006), 511-542.

Tjostheim, D., and Paulsen, J. "Bias of Some Commonly-Used Time Series Estimators", *Biometrika* 70 (1983), 389- 399.

White, H. "A Heteroskedasticity-Consistent Covarianace Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica* 48 (1980), 817- 838.

Wolf, M. "Stock Returns and Dividend Yields Revisited: A New Way to Look at an Old Problem", *Journal of Business & Economic Statistics* 18 (2000), 18- 30.

Wu, C.F.J. "Jackknife Bootstrap and Other Resampling Methods in Regression Analysis", *Annals of Statistics* 14 (1986), 1261- 1295.

Table 1. Selected VAR lag orders

	Panel A									Panel B								
									Jan 52- Dec 03									
	1	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*	*1*	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*
2	2									2								
3	2									13								
4	2	2								13	8							
6	2	1	1							8	1	1						
9	2		1							5		1						
12	2	1	1	1	1					3	1	1	1	1				
24	2	2	2	1	1		2			2	2	2	1	1		2		
36	2	2	1	1	1	1	1			2	2	1	1	1	1	1		
48	2	2	1	1	1		1	1		2	2	1	1	1		1	5	
60	2	2	1	1	1		1			2	2	1	1	1		8		
120	2	1	1	1	1		1	1	2	2	1	1	1	1		1	1	2
									Jan 52- Dec 78									
	1	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*	*1*	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*
2	1									2								
3	1									3								
4	1	1								9	3							
6	1	1	1							9	2	2						
9	1		1							9		2						
12	1	1	1	1	1					9	2	1	3	3				
24	1	1	2	1	1		3			2	3	2	1	3		3		
36	2	2	2	2	1	3	3			2	2	2	2	3	3	3		
48	1	2	2	2	1		3	1		2	2	2	2	3		3	3	
60	1	2	2	2	1		3			2	2	2	2	3		3		
120	1	1	1	1	1		2	1	2	3	2	2	3	3		3	3	2
									Jan 82- Dec 03									
	1	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*	*1*	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*
2	1									10								
3	1									6								
4	1	1								5	7							
6	1	1	1							5	4	1						
9	1		1							5		1						
12	1	1	1	1	1					3	8	1	1	1				
24	1	1	1	1	1		1			2	2	2	2	1		2		
36	1	1	1	1	1	1	1			1	1	1	1	1	1	1		
48	1	1	1	1	1		1	1		1	1	1	1	1		1	1	
60	1	1	1	1	1		1			1	1	1	1	1		1		
120	1	1	1	1	1		1	1	1	1	1	1	1	1		1	1	1

Note: Panel A reports the VAR lag orders chosen by SIC for various maturity pairs. Maturities of longer rates are in the first column and those of the shorter rates are in the first rows of the sub-tables corresponding to three samples. The maximum lag length considered is $p_{\max} = 12(T/100)^{1/4}$ which is supported in the Monte Carlo study of Schwert (1989). Panel B provides lag lengths that result from application of the autocorrelation test after SIC, with the highlighting indicating cases where these differ from the SIC results.

Table 2. Multivariate autocorrelation test

	Jan 52- Dec 03								
	1	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*
2	**4.36**								
	0.36								
	0.44								
3	**3.24**								
	0.52								
	0.51								
4	**2.80**	**1.30**							
	0.59	0.86							
	0.59	0.90							
6	**3.29**	**7.03**	**5.05**						
	0.51	0.13	0.28						
	0.53	0.11	0.31						
9	**2.92**		**5.16**						
	0.57		0.27						
	0.57		0.28						
12	**3.37**	**7.79**	**5.98**	**4.86**	**2.77**				
	0.50	0.10	0.20	0.30	0.60				
	0.55	0.09	0.18	0.30	0.64				
24	**2.41**	**4.19**	**6.62**	**5.44**	**3.41**		**7.91**		
	0.66	0.38	0.16	0.24	0.49		0.10		
	0.71	0.42	0.15	0.25	0.53		0.09		
36	**2.70**	**4.42**	**5.53**	**4.42**	**3.16**	**4.30**	**5.76**		
	0.61	0.35	0.24	0.35	0.53	0.37	0.22		
	0.63	0.32	0.28	0.34	0.58	0.37	0.21		
48	**2.76**	**4.63**	**5.72**	**4.53**	**3.04**		**3.83**	**7.79**	
	0.60	0.33	0.22	0.34	0.55		0.43	0.10	
	0.65	0.35	0.23	0.37	0.61		0.46	0.09	
60	**2.88**	**2.61**	**5.49**	**4.32**	**2.90**		**3.17**		
	0.58	0.62	0.24	0.36	0.57		0.53		
	0.64	0.64	0.27	0.36	0.60		0.61		
120	**2.38**	**7.79**	**5.19**	**3.99**	**3.09**		**5.12**	**7.97**	**2.88**
	0.67	0.10	0.27	0.41	0.54		0.27	0.09	0.58
	0.72	0.09	0.31	0.43	0.57		0.28	0.07	0.60

	Jan 52- Dec 78								
	1	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*
2	**11.74**								
	0.02								
	0.02								
3	**8.83**								
	0.07								
	0.06								
4	**7.35**	**0.83**							
	0.12	0.93							
	0.12	0.94							
6	**9.27**	**3.55**	**12.68**						
	0.05	0.47	0.01						
	0.04	0.53	0.01						
9	**9.61**		**8.40**						
	0.05		0.08						
	0.02		0.08						
12	**9.89**	**5.78**	**6.87**	**5.35**	**1.19**				
	0.04	0.22	0.14	0.25	0.88				
	0.04	0.21	0.16	0.28	0.88				
24	**14.26**	**8.93**	**3.97**	**6.97**	**5.83**		**2.87**		
	0.01	0.06	0.41	0.14	0.21		0.58		
	0.01	0.06	0.38	0.16	0.19		0.63		
36	**5.49**	**4.57**	**4.30**	**4.46**	**8.00**	**3.50**	**0.45**		
	0.24	0.33	0.37	0.35	0.09	0.48	0.98		
	0.23	0.37	0.40	0.39	0.09	0.54	0.98		
48	**12.92**	**3.69**	**3.87**	**4.56**	**8.77**		**0.85**	**10.71**	
	0.01	0.45	0.42	0.34	0.07		0.93	0.03	
	0.00	0.44	0.43	0.35	0.08		0.93	0.02	
60	**12.77**	**3.40**	**4.44**	**5.90**	**9.70**		**1.65**		
	0.01	0.49	0.35	0.21	0.05		0.80		
	0.01	0.48	0.42	0.21	0.03		0.81		
120	**12.61**	**10.66**	**12.18**	**12.41**	**13.45**		**11.54**	**12.00**	**8.89**
	0.01	0.03	0.02	0.01	0.01		0.02	0.02	0.06
	0.01	0.03	0.01	0.01	0.00		0.01	0.01	0.08

	Jan 82- Dec 03								
	1	*2*	*3*	*4*	*6*	*9*	*12*	*24*	*60*
2	**1.80**								
	0.77								
	0.86								
3	**1.44**								
	0.84								
	0.90								
4	**1.01**	**2.14**							
	0.91	0.71							
	0.97	0.84							
6	**0.89**	**1.75**	**2.72**						
	0.93	0.78	0.61						
	0.97	0.88	0.74						
9	**2.40**		**3.93**						
	0.66		0.42						
	0.79		0.46						
12	**4.51**	**5.16**	**5.27**	**4.08**	**1.56**				
	0.34	0.27	0.26	0.39	0.82				
	0.38	0.27	0.26	0.42	0.84				
24	**8.16**	**10.23**	**11.35**	**10.67**	**8.69**		**6.93**		
	0.09	0.04	0.02	0.03	0.07		0.14		
	0.09	0.02	0.02	0.03	0.07		0.16		
36	**7.79**	**8.55**	**8.84**	**8.53**	**6.73**	**5.23**	**6.20**		
	0.10	0.07	0.07	0.07	0.15	0.26	0.18		
	0.09	0.06	0.06	0.06	0.22	0.33	0.20		
48	**5.80**	**6.46**	**6.59**	**5.90**	**4.42**		**4.27**	**7.63**	
	0.21	0.17	0.16	0.21	0.35		0.37	0.11	
	0.27	0.15	0.13	0.24	0.41		0.45	0.11	
60	**4.72**	**5.25**	**5.37**	**4.69**	**3.41**		**3.06**		
	0.32	0.26	0.25	0.32	0.49		0.55		
	0.33	0.31	0.25	0.33	0.58		0.59		
120	**1.93**	**1.65**	**1.46**	**1.12**	**0.90**		**1.29**	**3.29**	**3.42**
	0.75	0.80	0.83	0.89	0.92		0.86	0.51	0.49
	0.81	0.85	0.89	0.91	0.94		0.89	0.52	0.52

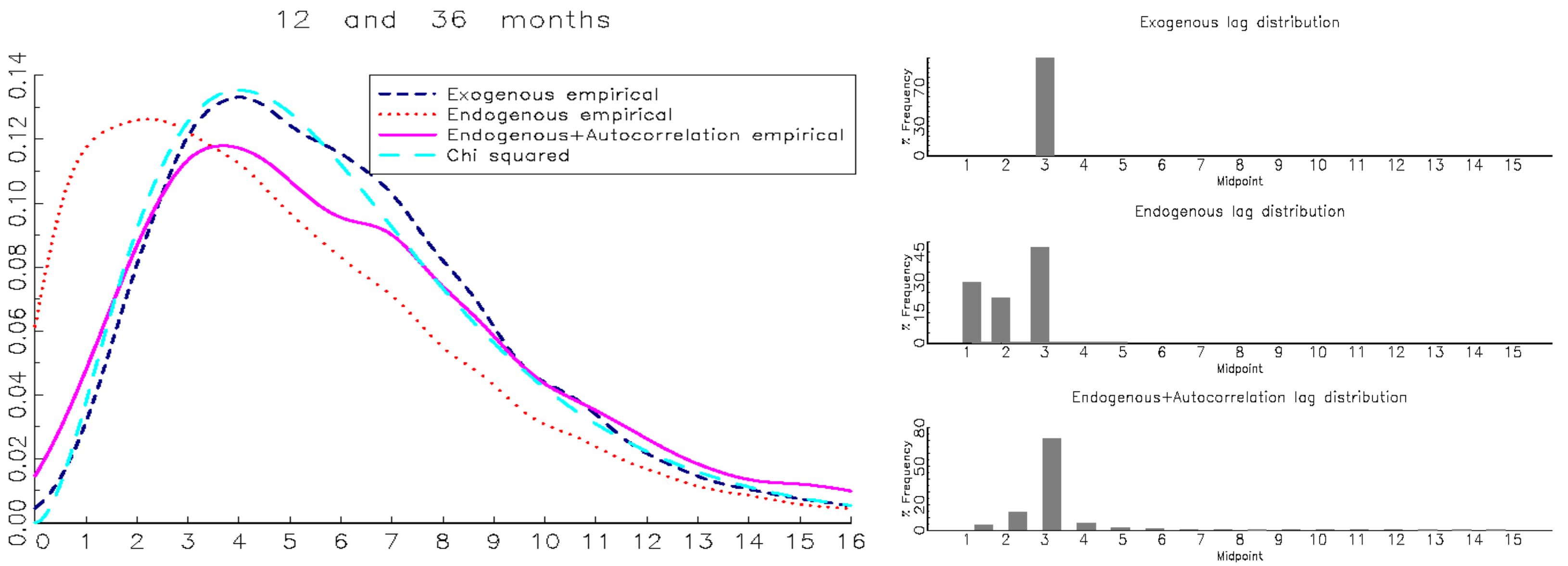
Note: Table provides first order autocorrelation test results for the VAR residuals. The first number, given in **bold** is the test statistic followed by the corresponding asymptotic and wild bootstrapped *p values* (respectively) that there is no autocorrelation. The VAR lag length is given in Panel A of Table 1 and the test is described in the Appendix A. The cases where the first order serial correlation is detected are highlighted.

Table 3. Number of iterations required in stationarity correction

	Jan 52- Dec 03									*Jan 52- Dec 78*									*Jan 82- Dec 03*								
												Panel A. Unrestricted Model															
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	0									0									68								
3	N.C.									0									N.C.								
4	N.C.	55								65	0								N.C.	58							
6	58	0	0							66	0								34	35	0						
9	40		0							N.C.		0							29		0						
12	0	0	0	0	0					62	0	0	13	0					0	0	0	0	0				
24	40	33	27	0	0		42			39	48	42	0	49		57			0	0	0	0	0		0		
36	46	40	0	0	0	0	0			48	50	50	49	56	59	60			0	0	0	0	0	0	0		
48	50	46	0	0	0		0	65		52	54	54	54	59		61	55		0	0	0	0	0		0	0	
60	52	49	0	0	0		N.C.			55	56	56	55	60		61			0	0	0	0	0		N.C.		
120	57	0	0	0	0		0	0	52	62	56	57	62	62		60	56	26	0	0	0	0	0		0	0	0
												Panel B. Restricted Model															
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	2.88									1.16									47.81								
3	N.C.									5.97									N.C.								
4	N.C.	22.79								38.47	3.43								N.C.	37.16							
6	69.78	1.21	1.39							30.59	4.69								57.94	35.13	1.70						
9	32.25		1.25							N.C.		1.99							23.75		2.09						
12	16.44	1.53	1.61	1.60	2.35					56.40	6.99	1.22	10.90	12.07					12.17	7.19	1.91	1.98	1.66				
24	34.58	24.50	15.64	2.47	3.36		25.08			33.84	30.74	22.34	1.76	34.94		42.55			2.17	4.07	10.82	8.65	2.44		2.91		
36	42.46	31.45	2.37	2.34	3.21	2.86	2.54			40.41	35.20	32.84	30.24	43.37	50.86	47.41			2.07	2.09	2.00	2.04	2.51	2.45	2.84		
48	46.74	37.85	2.59	2.53	3.22		2.62	46.78		43.17	39.43	39.37	38.57	48.83		55.87	57.02		2.19	2.19	2.36	2.45	2.69		3.15	3.32	
60	48.12	40.82	2.68	2.89	3.31					43.38	40.43	41.87	40.80	51.53		58.84			2.36	2.11	2.65	2.54	2.68		N.C.		
120	52.38	2.57	2.53	2.64	2.85		3.03	2.92	31.10	45.27	36.12	39.82	52.40	56.27		58.78	55.19	3.26	2.48	2.19	2.74	2.70	3.32		4.47	4.69	2.89

Note: Panel A reports the number of iterations required to make the bias corrected VAR parameters stable for the unrestricted model. Panel B reports the average number of iterations required in the bootstrap simulations that generate empirical *p values*. Details of the procedure are given in Section III.B. NC means non-convergence of the iterative procedure.

Figure 1. Effect of lag uncertainty on empirical distributions of LM statistic



Note: The left-hand panel shows the empirical bootstrap distributions of the LM test statistic for different treatments of the lag uncertainty for the period of January 1952- December 1978, together with the asymptotic distribution $\chi^2(6)$. The distributions of the lag lengths are provided in the right-hand panel. The estimated LM statistic for this maturity pair is 13.29 and corresponding 95% quantiles of the 4 distributions graphed are 12.14, 11.56, 15.08 and 12.59.

Table 4. LM test of the EH of term structure

Jan 52- Dec 03

	1	2	3	4	6	9	12	24	60
2	28.28								
	0.00								
	0.00								
3	27.14*								
	0.00								
	0.00								
4	23.47*	32.18							
	0.01	0.01							
	0.00	0.01							
6	22.37	16.03	18.00						
	0.13	0.00	0.00						
	0.14	0.01	0.01						
9	22.12		15.66						
	0.01		0.00						
	0.02		0.01						
12	17.12	9.95	11.49	11.51	4.94				
	0.01	0.01	0.00	0.00	0.08				
	0.02	0.05	0.04	0.03	0.21				
24	16.98	16.00	15.67	6.30	3.21		13.03		
	0.00	0.00	0.00	0.04	0.20		0.01		
	0.02	0.02	0.02	0.13	0.35		0.04		
36	13.45	12.34	5.64	5.45	3.18	2.63	3.34		
	0.01	0.02	0.06	0.07	0.20	0.27	0.19		
	0.04	0.05	0.14	0.14	0.31	0.40	0.29		
48	11.39	10.28	5.71	5.41	3.35		2.83	11.33	
	0.02	0.04	0.06	0.07	0.19		0.24	0.33	
	0.06	0.08	0.14	0.15	0.30		0.35	0.20	
60	10.54	9.45	6.05	5.67	3.69		11.91*		
	0.03	0.05	0.05	0.06	0.16		0.29		
	0.08	0.10	0.13	0.14	0.28		0.14		
120	9.88	7.53	7.38	6.90	4.95		3.87	3.62	12.10
	0.04	0.02	0.03	0.03	0.08		0.14	0.16	0.02
	0.10	0.08	0.09	0.10	0.17		0.23	0.26	0.06

Jan 52- Dec 78

	1	2	3	4	6	9	12	24	60
2	36.59								
	0.00								
	0.00								
3	35.56								
	0.00								
	0.00								
4	45.64	21.16							
	0.00	0.00							
	0.00	0.01							
6	46.40	21.16	N.S.						
	0.00	0.00							
	0.00	0.00							
9	25.30*		8.09						
	0.00		0.02						
	0.00		0.08						
12	32.62	8.16	6.17	8.59	9.03				
	0.02	0.09	0.05	0.20	0.17				
	0.02	0.09	0.10	0.14	0.19				
24	5.64	8.09	4.69	4.09	7.63		16.92		
	0.23	0.23	0.32	0.13	0.27		0.01		
	0.28	0.21	0.32	0.22	0.30		0.03		
36	3.77	3.20	3.55	4.03	7.06	10.37	13.29		
	0.44	0.53	0.47	0.40	0.32	0.11	0.04		
	0.50	0.54	0.48	0.41	0.35	0.16	0.07		
48	3.25	2.73	3.03	3.50	5.41		8.74	9.37	
	0.52	0.60	0.55	0.48	0.49		0.19	0.15	
	0.56	0.62	0.55	0.50	0.52		0.25	0.20	
60	3.27	2.84	3.13	3.54	4.68		7.13		
	0.51	0.59	0.54	0.47	0.58		0.31		
	0.54	0.59	0.56	0.49	0.58		0.37		
120	7.49	4.77	5.52	7.69	7.97		9.18	11.39	17.77
	0.28	0.31	0.24	0.26	0.24		0.16	0.08	0.00
	0.21	0.33	0.29	0.27	0.27		0.20	0.12	0.01

Jan 82- Dec 03

	1	2	3	4	6	9	12	24	60
2	56.17								
	0.00								
	0.00								
3	26.45*								
	0.00								
	0.00								
4	23.42*	17.24							
	0.01	0.24							
	0.00	0.08							
6	14.15	12.48	7.86						
	0.17	0.13	0.02						
	0.22	0.10	0.05						
9	11.11		6.17						
	0.35		0.05						
	0.16		0.08						
12	12.39	11.13	5.88	5.66	5.83				
	0.05	0.08	0.05	0.06	0.05				
	0.04	0.05	0.10	0.11	0.11				
24	10.32	11.82	13.07	13.69	4.99		13.97		
	0.04	0.02	0.01	0.01	0.08		0.01		
	0.05	0.05	0.05	0.03	0.16		0.03		
36	6.25	6.35	6.27	6.15	5.89	5.81	5.98		
	0.04	0.04	0.04	0.05	0.05	0.05	0.05		
	0.10	0.09	0.10	0.10	0.12	0.13	0.14		
48	5.67	5.81	5.62	5.32	4.74		4.34	5.42	
	0.06	0.05	0.06	0.07	0.09		0.11	0.07	
	0.11	0.10	0.12	0.15	0.18		0.22	0.18	
60	5.51	5.51	5.21	4.84	4.14		3.46*		
	0.06	0.06	0.07	0.09	0.13		0.18		
	0.12	0.13	0.15	0.17	0.22		0.29		
120	4.28	3.86	3.25	2.67	1.64		0.63	0.41	0.43
	0.12	0.14	0.20	0.26	0.44		0.73	0.81	0.81
	0.19	0.22	0.28	0.36	0.53		0.79	0.86	0.84

Note: First number in each set is the LM test statistic and second and third numbers are asymptotic and finite sample *p-values,* respectively. The cells highlighted indicate the rejection of the EH at the 5% significance level according to the empirical finite sample distribution. N.S. indicates the restricted VAR is unstable.

* indicates cases where the maximum VAR lag length is restricted to be 5 as the iterative procedure did not converge.

Appendix A. Multivariate autocorrelation test robust to conditional heteroskedasticity

In this appendix we describe the vector autocorrelation test of Godfrey and Tremayne (2005) robust to conditional heteroskedasticity. Consider a general dynamic system of n stochastic equations, the residuals of which are suspected to have autocorrelation,

$$\mathbf{Y}_0 = \mathbf{Z}_0\mathbf{B}_0 + \mathbf{U}_0 \tag{A1}$$

where $\underset{T\times n}{\mathbf{Y}_i} = [\mathbf{y}_{1+i},...,\mathbf{y}_{T+i}]'$, $\underset{[np+m]\times n}{\mathbf{B}_0} = [\mathbf{A}_1,...,\mathbf{A}_p,\mathbf{\Pi}]'$, $\underset{T\times m}{\mathbf{X}} = [\mathbf{x}_1,...,\mathbf{x}_T]'$, $\underset{T\times n}{\mathbf{U}_i} = [\mathbf{u}_{1+i},...,\mathbf{u}_{T+i}]'$, $\underset{T\times[np+m]}{\mathbf{Z}_0} = [\mathbf{Y}_{-1},...,\mathbf{Y}_{-p},\mathbf{X}]$, $\mathbf{y}_t$ and $\mathbf{u}_t$ are (n×1), $\mathbf{x}_t$ is (m×1), $\mathbf{A}_i$ is (n×n) and $\mathbf{\Pi}$ is (n×m), and this system reduces to a *VAR(p)* without an intercept when $\mathbf{\Pi}=\mathbf{0}$ and to a static system when $\mathbf{A}_i=\mathbf{0}$, $i=1,..,p$. We assume all values of z satisfying $\left|\mathbf{I} - \mathbf{A}_1 z - \mathbf{A}_2 z^2 ... - \mathbf{A}_p z^p\right| = 0$ lie outside the Argond diagram and that observations $\mathbf{y}_{1-p}$ to $\mathbf{y}_0$ are available for the lagged variables, leaving T number of observations to estimate (A1).

With autocorrelation of order g in $\mathbf{U}_0$,

$$\mathbf{U}_0 = \sum_{j=1}^{g} \mathbf{U}_{-j}\mathbf{C}_j + \mathbf{E},$$

where E has typical rows $\mathbf{e}_t'$, and $\mathbf{e}_t$ and $\mathbf{e}_l$ are uncorrelated for $l \neq t$. A model dependent autocorrelation test is based on the null hypothesis that $\mathbf{C}_1 = ... = \mathbf{C}_g = 0$ in an auxiliary system that includes the lagged least squares residuals from (A1), namely

$$\mathbf{Y}_0 = \mathbf{Z}\mathbf{B} + \mathbf{E} \tag{A2}$$

where $\underset{T\times[n(g+p)+m]}{\mathbf{Z}} = \left[\mathbf{Y}_{-1},...,\mathbf{Y}_{-p},\mathbf{X},\hat{\mathbf{U}}_{-1},...,\hat{\mathbf{U}}_{-g}\right]$, $\underset{[n(p+g)+m]\times n}{\mathbf{B}} = \left[\mathbf{A}_1,...,\mathbf{A}_p,\mathbf{\Pi},\mathbf{C}_1,...,\mathbf{C}_g\right]'$

The least squares estimator of (A2) is $\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ as in the familiar univariate case and if we let $\hat{\boldsymbol{\beta}} = vec(\hat{\mathbf{B}})$ then it can be shown,

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}), \qquad \text{(A3)}$$

with $\mathbf{V} = \mathbf{I}_n \otimes \mathbf{\Gamma}$ and $\mathbf{\Gamma} = p\lim \mathbf{Z}'\mathbf{Z}/T$ under suitable regularity conditions, see Hafner & Herwartz (2002). Under conditional homoskedasticity $\mathbf{W} = \mathbf{\Sigma}_e \otimes \mathbf{\Gamma}$, where $\mathbf{\Sigma}_e = E(\mathbf{e}_t\mathbf{e}_t')$ .The null hypothesis of no autocorrelation can be expressed as

H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ against $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{0}$,

where **R** is a $n^2p\times n^2(p+g)$ nonstochastic selection matrix of zeros except a unity in each row that picks up the parameters of the lagged residuals in $\boldsymbol{\beta}$ one by one.

Under the heteroskedasticity $\mathbf{\Sigma}_e$ is no longer constant, but multivariate extension of White's (1980) heteroskedsticity consistent covariance matrix estimator can be used. It is consistently estimated as

$$\hat{\mathbf{W}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{e}}_t\hat{\mathbf{e}}_t' \otimes \mathbf{z}_t\mathbf{z}_t', \qquad \text{(A4)}$$

where $\mathbf{z}_t$ is t^{th} row of **Z**. In our application we replace $\hat{\mathbf{e}}_t$ by $\hat{\mathbf{u}}_t$, which is found to improve the finite sample inference in the univariate framework as discussed in Davidson and MacKinnon (1985) and Godfrey and Orme (2004). From (A3) and (A4), the multivariate GT test has the asymptotic distribution

$$GT = T(\mathbf{R}\hat{\boldsymbol{\beta}})'\left[\mathbf{R}(\hat{\mathbf{V}}^{-1}\hat{\mathbf{W}}\hat{\mathbf{V}}^{-1})\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}) \xrightarrow{d} \chi^2(n^2k).$$