On the Expectations Hypothesis in US Term Structure ♣

Erdenebat Bataa*, Dong H. Kim & Denise R. Osborn

University of Manchester

February 2006

Abstract

We extend the vector autoregression (VAR) based expectations hypothesis test of term structure, considered in Bekaert & Hodrick (2001) using recent developments in bootstrap literature. Modifications include the use of wild bootstrap to allow for conditional heteroskedasticity in the VAR residuals without imposing strict parameterization, endogeneous model selection procedure in the bootstrap replications to reflect true uncertainty and the stationarity correction designed to prevent finite-sample bias adjusted VAR parameters from becoming explosive.

JEL classification: G10; E43.

Keywords: expectations hypothesis; term structure of interest rates; vector autoregression

---

♣ Acknowledgements: The authors thank Dick van Dijk, Markus Kraetzig, and especially Daniel Thornton for making their computer codes available on the Internet, and the participants of the econometrics seminar and 1st PhD Conference at Manchester. We also benefited from useful discussions with Simon Peters and Chris Orme. *Corresponding author: N.5.4 Dover Street Building, University of Manchester, Oxford Road, Manchester, UK, M19 9PL. Tel: +44-161-2754860, Email: e.bataa@postgrad.man.ac.uk

1. Introduction

Interrelationship between interest rates of various maturities is a fundamental topic in economics and finance. One of the main theories that explain this relationship is the expectations hypothesis (EH) theory. According to the EH of the term structure, in equilibrium, investing in a succession of short-term bonds gives exactly the same expected return as investing in a long-term bond, when adjustment is made for the term premium. Various tests of this implication have yielded different results over various periods of time. Campbell & Shiller (1991), Campbell (1995), Rudebusch (1995) and Roberds & Whiteman (1999) note the EH works better at the short and long ends of the maturity spectrum and less well in the intermediate maturity range for a given short rate, thus creating a "U" shaped pattern. However, Thornton (2006) argues that analysis should not be based on the slope coefficient of the test equation only, since even under the alternative hypothesis where the EH does not hold, one can have slopes that are numerically close to the theoretical ones.

Mankiw & Miron (1986) argues that the poor performance of the EH over certain periods is related to monetary policy pursued by the Fed, 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 & Whiteman (1996), and Balduzzi, Bertola & Foresi (1997) provide models that accommodate Fed behaviour and confirm Mankiw & Miron's finding.

Inconsistent conclusions about the EH has also been attributed to the small sample properties of various tests. Early studies use conventional regression framework (e.g. Shiller 1979, Mankiw & Miron 1986), volatility test (Startz 1984) and VAR based Wald (Campbell & Shiller 1987, 1991) and Likelihood Ratio (Driffil, Psaradakis & Sola 1997) tests. In their recent seminal paper, Bekaert & Hodrick (2001), B & H thereafter, suggest a LM test and examine its finite sample behaviour by a Monte Carlo study and argue that their test performs better than Likelihood Ratio based Distance Metric and Wald tests, the latter of which had been used almost exclusively in the previous literature. The B & H methodology is fast gaining popularity and is adopted in Thornton (2004), Bekaert, Wei & Xing (2006), and Sarno, Thornton & Valente (2006).

However, recent developments in the bootstrap literature can usefully be employed to improve the B & H methodology. The asymptotic and small sample inferences from the LM test rely on either i.i.d. or factorized GARCH bootstrap of the VAR residuals. Goncalves & Kilian (2004) argue that i.i.d. re-sampling scheme is inaccurate in the presence of the conditional heteroskedasticity, which characterizes many financial time series and the latter type of bootstrap can suffer from misspecification problems, see e.g. Wolf (2000) and Belsley (2002). We therefore apply a wild bootstrap scheme, which does not require strict specification and keeps contemporaneous error correlation, and furthermore introduce stationarity correction, randomize the initial condition and endogenize lag length selection rule in the B & H methodology. As Sarno, *et all* (2006) find a structural break in the assumed data generating process at around 1982 using data that is used in our study and more importantly, to compare our result with theirs' we consider two sub-samples. This will also enable us to assess if Campbell & Shiller's (1991) claim that the EH performed better prior to 1978 is just because they included the Fed's non-borrowed reserve targeting policy period in the latter period. Furthermore, one may also expect the EH performing better in the second sub-sample as the innovation in the communication industry and competition in the financial market must have shrunk the transaction cost, which is one of the main enemies of the theory.

The paper is organised as follows. Section 2 examines the implications of the EH theory of the term structure of interest rates, Section 3 discusses how the theory is tested in VAR framework and outlines the B & H methodology. Suggested extensions are discussed in Section 4 and applied in Section 5. We also discuss our data in the latter section and Section 6 concludes.

2. Expectations hypothesis theory of the term structure

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)

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

where $R_{n,t}$ and $R_{m,t}$ are long and short rates at time t respectively, $E_tR_{m,t+mi}$, $i=0,1,2,\ldots k-1$, is the expectation of the short rates at $t+mi$ formed at time t and $\pi_{n,m}$ is a term premium which can vary across maturities but not through time. Here $k=n/m$ is defined to be an integer, m is the maturity of a shorter rate and n is the maturity of a longer rate. Since the EH places no restriction on $\pi_{n,m}$, this term can be ignored by working with demeaned series.[1]

Equation (1) is rarely tested directly, probably due to the empirical results that conclude 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

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

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

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

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 $e_{(n,m),t}$ is a moving average process of order $(n-m-1)$.

Equation (3) says that the current spread 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.[2]

---

[1] This assumption will simplify the derivation of restrictions on VAR parameters in what follows. See Melino (2001).

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

However, there are several econometric difficulties with the conventional regression approach applied to this equation. Firstly, we lose *n-m* observations at the end of the sample period. This can be quite serious, as the data available for analysis are usually relatively small. Secondly, the error term $e_{(n,m),t}$, is a moving average of order *n-m-1*, so standard errors have to be corrected, for example using the method described in Newey & West (1987). But these adjustments do not work well when *n-m* is not small relative to the sample size (see e.g., Campbell & Shiller 1991). Thirdly, the regressor is serially correlated and correlated with lags of the dependent variable, and this can cause finite sample problems as well (Campbell, Lo & MacKinlay, 1997).

3. VAR Approach to testing the EH

The problems associated with the single equation methods can be avoided using a VAR framework as suggested in Campbell & Shiller (1987, 1991). Let us assume that there exists a stationary vector stochastic process for $\mathbf{y}_t = [\Delta R_{m,t}, S_{(n,m),t}]'$, where $\Delta R_{m,t}$ is a change in short term rate and $S_{(n,m),t}$ is a spread between long and short term rates. Assuming the process for $\mathbf{y}_t$ is represented by a demeaned VAR of order *p*,

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{u}_t, \qquad (4)$$

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

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

$\mathbf{z}_t$ has *2p* elements, $\mathbf{z}_t = [\mathbf{y}'_t, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-p+1}]'$, and $\mathbf{v}_t$ is again a *2p* vector equal to $[u_{1,t}, u_{2,t}, 0, 0, ..., 0]'$, uncorrelated over time. Thus the vector $\mathbf{z}_t$ is assumed to summarise the whole history of $\mathbf{y}_t$.

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, restrictions imposed by the EH on VAR parameters can be shown to be

$$\mathbf{e}_2' = \mathbf{e}_1'\mathbf{A}[\mathbf{I}_2 - \frac{m}{n}(\mathbf{I}_2 - \mathbf{A}^n)(\mathbf{I}_2 - \mathbf{A}^m)^{-1}](\mathbf{I}_2 - \mathbf{A})^{-1}. \qquad (5)$$

The next two sections describe how the B & H methodology is used to test this restriction.

3.1. B&H Methodology: Asymptotic Inference

The restrictions in (5) are highly non-linear and are predominantly being tested by asymptotic Wald test even though it has some undesirable properties in finite samples (see e.g. Gregory & Veall 1985 and Dagenais & Dufour 1991). The Wald statistic is not invariant to how one specifies the null hypothesis and possibly, unit of measurement. Shea (1992) provides a numerical example how one can reach different conclusions on testing algebraically equivalent EH restrictions using Wald tests.

B & H (2001) suggest a Lagrange Multiplier test which employs restricted VAR parameter and argue the LM test has much better small sample properties than Wald test in terms of size and power using Monte Carlo simulations. They also consider Likelihood Ratio based Distance Metric test but they prefer the LM test. Since this methodology is relatively new, and is an important part of this study, it is summarised here.

They derive the LM test statistic based on Hansen's (1982) Generalized Method of Moments (GMM) estimator, which uses the orthogonality condition implied by (2). Let $\overline{\mathbf{A}}$ denote an estimate of the matrix of the restricted parameter satisfying (5) and define $\dot{\mathbf{A}} = [\mathbf{A}_1,...,\mathbf{A}_p]'$. Then the vector of orthogonality condition can be written

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

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

$$J_T(\boldsymbol{\theta}) \equiv \mathbf{g}_t(\boldsymbol{\theta})'\mathbf{W}\mathbf{g}_t(\boldsymbol{\theta}), \tag{6}$$

where, assuming the VAR of (4) is correctly specified with $\mathbf{u}_t$ uncorrelated, the weighting matrix, $\mathbf{W}$, is a consistent estimate of the inverse of

$$\boldsymbol{\Omega} \equiv E\left[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})'\right]. \tag{7}$$

Let the null hypothesis in (5) be expressed as:

$$H_o : \mathbf{c}(\boldsymbol{\theta}_0) = \mathbf{0}, \tag{8}$$

and define a Lagrangian for the constrained GMM maximization problem as

$$L(\boldsymbol{\theta}, \boldsymbol{\gamma}) = -\frac{1}{2}\mathbf{g}_t(\boldsymbol{\theta})'\hat{\boldsymbol{\Omega}}_T^{-1}\mathbf{g}_t(\boldsymbol{\theta}) - \mathbf{c}_t(\boldsymbol{\theta})'\boldsymbol{\gamma}; \tag{9}$$

where $\boldsymbol{\gamma}$ is a vector of Lagrange multipliers and $\hat{\boldsymbol{\Omega}}_T$ is a consistent estimate of $\boldsymbol{\Omega}$ obtained from (7) using the sample mean in place of the expectation. Since direct maximization of (9) is difficult, B & H (2001) suggest extending an approach put forward by Newey & McFadden (1994) who demonstrate how to derive a constrained consistent estimator starting from an initial unconstrained consistent one. Using a Taylor's Series expansion to the non-linear first order conditions for (9) yields

$$\sqrt{T}\mathbf{g}_T(\bar{\boldsymbol{\theta}}) \approx \sqrt{T}\mathbf{g}_T(\boldsymbol{\theta}_0) + \mathbf{G}_T\sqrt{T}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0); \tag{10}$$

$$\sqrt{T}\mathbf{c}_T(\bar{\boldsymbol{\theta}}) \approx \sqrt{T}\mathbf{c}_T(\boldsymbol{\theta}_0) + \mathbf{C}_T\sqrt{T}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{11}$$

where $\mathbf{G}_T$ and $\mathbf{C}_T$ are gradients, with respect to $\boldsymbol{\theta}$, of the sample orthogonality conditions and the vector of constraints, respectively, and under the null hypothesis, $\mathbf{c}_T(\boldsymbol{\theta}) = \mathbf{0}$. Substituting these into the first-order conditions,

$$\bar{\boldsymbol{\theta}} \approx \boldsymbol{\theta}_0 - \mathbf{D}_T^{-1/2}\mathbf{M}_T\mathbf{D}_T^{-1/2}\mathbf{G}_T'\hat{\boldsymbol{\Omega}}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}_0) - \mathbf{D}_T^{-1}\mathbf{C}_T'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{c}_T(\boldsymbol{\theta}_0) \tag{12}$$

$$\bar{\boldsymbol{\gamma}} \approx -(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{G}_T'\hat{\boldsymbol{\Omega}}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}_0) + (\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{c}_T(\boldsymbol{\theta}_0) \tag{13}$$

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

Let $\tilde{\boldsymbol{\theta}}$ represent an initial consistent unconstrained estimate. Then constrained estimates, $\bar{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\gamma}}$, are obtained by iterating on equations (12) and (13), substituting $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$ to derive a second constrained estimate, and so forth until the constraint is satisfied, i.e. $\mathbf{c}_T(\boldsymbol{\theta}) = \mathbf{0}$.[3]

This yields the constrained estimate, together with the Lagrange Multipliers, which under the null hypothesis of EH and assuming i.i.d. disturbances has asymptotic distribution

$$\sqrt{T}\bar{\boldsymbol{\gamma}} \rightarrow \mathbf{N}\left[\mathbf{0}, (\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-\mathbf{1}}\right]. \tag{14}$$

The constrained parameter estimate is not equal to the unconstrained one when the constraints in (8) significantly affect the value of the GMM objective function (6). From (14), LM test statistic is

$$T\bar{\boldsymbol{\gamma}}'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')\bar{\boldsymbol{\gamma}} \rightarrow \chi^2(2p), \tag{15}$$

where p is the lag length of the VAR. Two other test statistics considered in B & H (2001) are DM test, $T\mathbf{g}_T(\bar{\boldsymbol{\theta}})'\boldsymbol{\Omega}_T^{-1}\mathbf{g}_T(\bar{\boldsymbol{\theta}}) \rightarrow \chi^2(2p)$ and Wald test, $T\mathbf{c}_T(\tilde{\boldsymbol{\theta}})'(\mathbf{C}_T\mathbf{D}_T^{-1}\mathbf{C}_T')^{-1}\mathbf{c}_T(\tilde{\boldsymbol{\theta}}) \rightarrow \chi^2(2p)$. Note that the Wald test is based on the unrestricted, while the former two are based on the restricted information.

It is well known that estimated VAR parameter is, although consistent, biased in finite samples (see e.g. Tjostheim & Paulsen, 1983, and Bekaert, Hodrick & Marshall, 1997) and the bootstrap bias correction method used in Bekaert & Hodrick (2001) is as follows.

Step 1.1. Estimate VAR parameter $\mathbf{A}$.

---

[3] Note that we update $\mathbf{C}_T, \mathbf{D}_T, \mathbf{G}_T$ and $\hat{\boldsymbol{\Omega}}_T$ at each iteration step. In our application the tolerance level for convergence is 10^{-8}.

Step 1.2. Generate an artificial data set of the actual sample size using first p actual observations, $\hat{\mathbf{A}}$ from previous step and an i.i.d. bootstrap of the estimated residuals, then discard first p observations.

Step 1.3. Estimate "Monte Carlo" or artificial VAR parameter, $\hat{\mathbf{A}}_{M,i}$, using the data set obtained in Step 1.2.

Step 1.4. Repeat Steps 1.2 and 1.3 a large number of times, say b.[4]

Step 1.5. Estimate the bias, $\mathbf{B}$, as the difference between the original VAR parameter estimate and the mean of the Monte Carlo parameters, i.e. $\hat{\mathbf{B}} = \hat{\mathbf{A}} - \frac{1}{b}\sum_{i=1}^{b} \hat{\mathbf{A}}_{M,i}$.

Step 1.6. Estimate bias-corrected parameter by adding the estimated bias to the original VAR parameter estimate, i.e., $\hat{\mathbf{A}}_c = \hat{\mathbf{A}} + \hat{\mathbf{B}}$.

To obtain bias-corrected parameter estimate that satisfies the null hypothesis, they use the bias corrected unconstrained VAR parameter and 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 (12) and (13). This bias corrected, constrained parameter is then used to derive the *LM* test statistics and corresponding asymptotic inference through (15).

3.2. B&H methodology: Finite Sample Inference

B & H also consider their methodology in a finite sample. Indeed, it has been well documented that large sample inference can be misleading for finite samples (see e.g. Mankiw & Miron 1986 and Horowitz 1997). The limiting distribution of the LM test statistic is asymptotically pivotal and it is proven that the bootstrap provides a first order asymptotic refinement, i.e. improved finite sample inference (see e.g. Horowitz 2001, p.3184). The finite sample significance level associated with the null hypothesis of the EH in B & H (2001) is derived as follows:

Step 2.1. Use an estimate of the bias-corrected constrained parameter $\overline{\mathbf{A}}$, after transforming $\overline{\boldsymbol{\theta}}$ from (12), and an i.i.d bootstrap of the unrestricted residuals as a data

[4] As in B & H (2001) and Sarno *et al* (2006) b is 100000 in our study.

generating process (DGP) to generate an artificial data set of the actual sample size plus 1000 observations that are discarded to attenuate the start-up effect.

Step 2.2. Estimate a LM test statistic for the artificial data set, i.e., first estimate model (4), using the same VAR lag length *p* obtained from the actual data, and then iterate on (12) and (13) to obtain a LM statistic from (15).

Step 2.3. Repeat Steps 2.1 and 2.2 a large number of times. [5]

Step 2.4. Calculate an empirical *p-value*, which is the proportion of LM test statistics from Step 2.3 that are larger than or equal to the sample statistic estimated on the actual data.

When the conditional heteroskedasticity in the VAR residuals is allowed for they estimate a specific factor GARCH model for the residuals and derive residuals from that model rather than resampling with replacement in Step 2.1.

4. Extensions to B & H Methodology

In this section we suggest several modifications, which are motivated by recent developments in bootstrap literature, to the B & H methodology in order to encompass more general situations. Suggestions include stationarity correction and randomizing the initial condition in the bias correction procedure, endogenizing the lag order selection and the use of restricted residuals for the finite sample inference and finally, allowing for conditional heteroskedasticity in the residuals of the estimated model without imposing a priori parameterization, all of which are discussed in this section.

4.1. On the bias correction

We consider three modifications in the bias correction part of the B&H methodology. Firstly, we introduce stationarity correction. In contrast to much of the previous literature on term structure Bekaert & Hodrick (2001) assume interest rates are stationary. Sarno, Thornton & Valente (2006) also assume they are *I(0)* and provide some unit root test results that indeed support their assumption. But it is almost a stylized fact that interest rates are highly persistent. Therefore, eigenvalues of the

[5] B & H (2001), Bekaert *et al* (2006), and Sarno *et al* (2006) replicate 25000 times.

estimated companion-form VAR parameter can be very close to unity if not more than that, which means there is no guarantee that bias corrected parameter to be stable, i.e. $\lambda_{\max}(\hat{\mathbf{A}}_c) < 1$, where $\lambda_{\max}$ is the highest eigenvalue. This is the stationary correction issue, which, for example, is addressed in Kilian (1998b). He suggests to add an additional step:

Step 1.7. If $\lambda_{\max}(\hat{\mathbf{A}}_c) \geq 1$, let $\hat{\mathbf{B}}_1 = \hat{\mathbf{B}}$, $\delta_1 = 1$ and define $\hat{\mathbf{B}}_{i+1} = \delta_i \hat{\mathbf{B}}_i$ and $\delta_{i+1} = \delta_i - 0.001$. 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,… until $\lambda_{\max}(\hat{\mathbf{A}}_c) < 1$.

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, to attenuate the start up effect Bekaert & Hodrick (2001) discard first *p* observations in each of the 100000 bootstrap replications. As *p* can be one, for example, it might not fully account for the uncertainty associated with initial condition. We therefore follow an alternative suggested in Stine (1987), where the observed data are split into *T-p+1* overlapping blocks of length *p* and one of them is selected randomly as a starting point.

Finally, to bias correct the constrained VAR parameter, that is eventually used to generate empirical distribution of the test statistics, B & H (2001) use the i.i.d. bootstrap of the residuals and bias corrected unconstrained VAR parameter to generate 71000 observations, which are then subjected to the iterative process. But this procedure seems rather ad hoc, and is certainly not valid in the presence of the conditionally heteroskedasticity. We subject the actual data and the bias corrected unconstrained parameter to the iterative process directly since Newey & McFadden (1994) argue the consistency of the estimator is sufficient for their expansion to work.

4.2. On the finite sample inference

One can observe that there is an inconsistency in treating the lag order in the B & H methodology discussed in Section 3. When estimating (4) it assumes the lag order is unknown and estimates it but when obtaining an empirical *p value* it treats it as known and uses the lag order estimated from the actual data in Step 2.2. But it is often emphasized that the bootstrap world should always reflect the actual world (see e.g. Li & Maddala, 1996). Incorrectly ignoring the uncertainty involved in determining the true lag order in finite samples can lead to spurious inference. Therefore we estimate the lag order for every bootstrap dataset in Step 2.2 employing the same criteria that used for the actual dataset. Asymptotically there is no difference between endogenizing or not since every consistent model selection criterion will be choosing the right lag length almost surely. The idea is formalized in Kilian (1998a) and was shown to improve the finite sample inference in impulse response analysis framework.

Moreover, by the same consistency argument when estimating model (4) on the artificial restricted data in Step 2.2 one has to bias-correct the resulting VAR parameter, following Steps 1.1-1.5. This approach essentially reduces to the usage of double bootstrap and is consistent with Hansen (2005). As computational cost of the bias correction at each bootstrap iteration is high we replicate 10000 times not 100000, i.e. b=10000. Consequently, in the estimation of the finite sample distributions of the test statistics the number of bootstrap simulations is reduced to 399 as opposed to 25000 that used in B & H (2001), Thornton (2004) and Sarne *et al* (2006).[6]

As Davidson & MacKinnon (1985) and Godfrey & Orme (2004) suggest the use of the restricted residuals not the restricted ones provide improvements in the finite sample we use the former, $\overline{\boldsymbol{\eta}}_t = \mathbf{y}_t - \overline{\dot{\mathbf{A}}}\mathbf{z}_{t-1}$. Sarno *et al* (2006) also use restricted VAR residuals.

4.3. Allowing for conditional heteroskedasticity

The EH, itself, places no restriction on the distributions of the VAR disturbances. It is common that the residuals from the estimated models exhibit volatility clustering, especially when financial time series are used (see e.g. Bollerslev, Chou & Kroner 1992). The bias correction method in B & H (2001) relies on i.i.d. residuals, although they do accept possible conditional heteroskedasticity in the residuals when deriving

[6] Estimation of the finite sample distributions of test statistics in this way required minimum of 2 hours, longer than that when AIC is used as a model selection method, for each maturity pair on a Pentium 4 H.T. 2.00 GHz machine.

finite sample inference. In particular, they estimate a VAR-GARCH model and generate the residuals in Section 3.2 through the resulting model. But there is no solid reasoning behind why this specific form of volatility clustering model is being used (Goncalves & Kilian 2004), and even if this class of GARCH models is appropriate the precise from of the GARCH model will be unknown, leading to different results for different specifications (Wolf 2000 and Belsley 2002).

In general we can consider four cases, virtually only one of which is considered in B and H (2001) and can be referred to as the benchmark model or Case I. Three other cases are:

Case II) disturbances are i.i.d. under the null and conditionally heteroskedastic under the alternative;

Case III) disturbances are conditionally heteroskedastic under the null and i.i.d under the alternative;

Case IV) disturbances are conditionally heteroskedastic under both null and alternative hypotheses.

The way the finite sample distribution of the LM test statistic derived is no longer valid for Cases III and IV and the bias adjustment procedure based on an i.i.d. bootstrap of the unconstrained VAR parameter discussed in Section 3.1. may not be justified for Cases II and IV.

To bias correct unconstrained VAR parameters for Cases II and IV and make finite sample inferences for Cases III and IV, we adopt a wild bootstrap developed in Liu (1988) following recommendations in Wu (1986) and Beran (1986). The particular form used is the recursive design wild bootstrap which is shown to be better in small samples 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 & Kilian (2004). A bootstrap sample, in this case, is generated as $\mathbf{y}_t^* = \dot{\mathbf{A}}\mathbf{z}_{t-1} + \boldsymbol{\eta}_t^*$, $\boldsymbol{\eta}_t^* = \omega_t \boldsymbol{\eta}_t$, *t=1,…,T,* where $\dot{\mathbf{A}}$ and $\boldsymbol{\eta}_t$ are rearranged VAR parameter and a vector of residuals at time *t* respectively, $E(\omega_t) = 0$ and $E(\omega_t^2) = 0$. Although Liu (1988) also suggested to have $E(\omega_t^3) = 0$, in our study ω_t is assumed to have the Rademacher distribution, which takes negative and positive ones with equal probabilities, and is preferred in most recent Monte Carlo studies, including Davidson & Flachaire (2001), Godfrey & Orme (2004) and Godfrey & Tremayne (2005).

5. Empirical evidence on the EH

In this section we describe the data used, explain why different model selection methods are employed in our study and provide empirical results from the modified B&H methodology.

5.1. Data

We use continuously compounded zero coupon yield curve data, as used in Sarno, Thornton & Valente (2006), which are update of famous McCulloch (1990) data. Its full coverage is from January 1952 to December 2003, but as there is significant body of evidence that the EH performed better prior to 1978 (e.g. Campbell & Shiller, 1991) and as Sarno, *et al* (2006) find a structural break in the VAR parameters at around 1982, which roughly coincide with the abandonment of the Fed's reserve targeting policy, we consider two sub-samples, Jan 1952- Dec 1978 and Jan 1982- Dec 2003. Such sample splitting can also be motivated by a hypothesis that given the advent of cheap communication means and competition in the financial market, the transaction costs must have shrunk over time favouring the EH in the second sub-period.

5.2. Model selection method

Sections 3 and 4 illustrate the inference on the validity of the EH depends on the lag length of the VAR, which is assumed to have generated the data. But there are various model selection criteria, the finite sample properties of which are unknown in the presence of conditional heteroskedasticity. But we know that a necessary condition for the validity of the residual bootstrap is the absence of autocorrelation in the VAR residuals. As different methods will potentially lead to different residual properties in finite samples, one has to be careful when choosing the model selection method. B & H (2001) use SIC as the model selection criterion, in contrast to Campbell & Shiller (1991) and Hardouvellis (1994) who assume that the data is known to be generated by VAR(4), and provide residual-autocorrelation test results for each equation of the VAR. Although there is some evidence of autocorrelation, they stick to the lag length of 1 that is chosen by the criterion. There is no consensus in the term structure literature over which model selection rule is appropriate to use. For example, Tzavalis & Wickens (1998) and Thornton (2004) also use SIC while Shea (1992), and Sarno *et* *al* (2006) use AIC. However, recent results in econometrics literature seem to favour AIC or its modifications. Ng & Perron (2002, 2005) prefer modified AIC while Ivanov & Killian (2005) suggest using AIC when monthly data are used. It is therefore interesting to see how the final conclusion about the validity of the EH differs depending on a specific model selection method used.

Furthermore, the way B&H reaches their conclusion that there is no autocorrelation in the residuals is problematic. Firstly, the autocorrelation test used in their 5 equation VAR is asymptotic and it is known that the large and finite sample properties depart as the number of equations in a system grows, see e.g. Laitinen (1978). Secondly, the test is used for each individual equation of the VAR, potentially leading to the problem of mass significance, as discussed in Edgerton & Shukur (1999). In order to make sure that there is no autocorrelation in the residuals we employ multivariate autocorrelation test robust to conditional heteroskedasticity, that is studied in Bataa (2006) and some details of which are provided in the Appendix A.

5.3. Empirical results

In addition to the LM test we also use the DM and Wald tests to derive the finite sample inference. As discussed in Section 3.1, the finite sample distribution of the latter is poorly approximated by the first order asymptotic theory, but this does not necessarily mean the bootstrap approximation will be poor as well (see, e.g. Horowitz 1997). One can also notice there is nothing in B & H methodology that controls the restricted VAR parameter to be stable, i.e. there is no guarantee that the finite sample distribution of the LM test statistic can always be estimated, which means the asymptotic Wald test can potentially be the only way to make any inference[7].

Since there is not much information on the relative performance of various model selection criteria in finite samples and in the potential presence of conditional heteroskedasticity, as discussed above, we provide all our empirical results based on AIC and SIC model selection criteria, the most popular ones in the term structure literature. All the inferences are made on the basis of 5% significance level.

[7] In fact, 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 1. Multivariate ARCH test in VAR residuals

Panel A. AIC is used as model selection method

	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	**55.0**									**24.8**								
	0.00									0.00								
	192.6									**159.0**								
	0.00									0.00								
	12									*10*								
3	**102.4**									**47.8**								
	0.00									0.00								
	159.9									**41.5**								
	0.00									0.00								
	9									*9*								
4	**73.9**	**66.2**								**148.2**	**149.6**							
	0.00	0.00								0.00	0.00							
	94.9	**67.8**								**235.3**	**291.3**							
	0.00	0.00								0.00	0.00							
	9	*9*								*3*	*4*							
6	**38.4**	**28.7**	**53.2**							**130.2**	**131.0**	**176.9**						
	0.00	0.00	0.00							0.00	0.00	0.00						
	34.6	**82.6**	**79.7**							**214.9**	**251.9**	**145.7**						
	0.00	0.00	0.00							0.00	0.00	0.00						
	9	*14*	*14*							*3*	*4*	*1*						
9	**30.0**		**75.2**							**96.5**		**166.1**						
	0.00		0.00							0.00		0.00						
	111.9		**N.A.**							**171.4**		**145.6**						
	0.00									0.00		0.00						
	9		*14*							*3*		*1*						
12	**24.4**	**41.7**	**79.9**	**105.0**	**78.9**					**69.0**	**120.9**	**154.3**	**175.8**	**178.8**				
	0.00	0.00	0.00	0.00	0.00					0.00	0.00	0.00	0.00	0.00				
	121.3	**189.9**	**N.A.**	**96.7**	**139.1**					**154.7**	**143.3**	**139.0**	**138.0**	**133.8**				
	0.00	0.00		0.00	0.00					0.00	0.00	0.00	0.00	0.00				
	9	*14*	*14*	*9*	*9*					*4*	*1*	*1*	*1*	*1*				
24	**37.9**	**72.9**	**88.6**	**86.6**	**59.1**		**64.6**			**30.3**	**44.1**	**128.3**	**135.9**	**141.3**		**105.6**		
	0.00	0.00	0.00	0.00	0.00		0.00			0.00	0.00	0.00	0.00	0.00		0.00		
	43.4	**112.2**	**N.A.**	**N.A.**	**17.7**		**34.2**			**78.3**	**99.1**	**117.5**	**114.0**	**113.4**		**99.7**		
	0.00	0.00			0.04		0.00			0.00	0.00	0.00	0.00	0.00		0.00		
	8	*12*	*15*	*15*	*15*		*3*			*4*	*4*	*1*	*2*	*2*		*2*		
36	**57.1**	**75.0**	**122.1**	**118.1**	**87.8**	**70.5**	**37.9**			**19.5**	**38.7**	**104.9**	**112.6**	**129.5**	**126.3**	**115.5**		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00			0.02	0.00	0.00	0.00	0.00	0.00	0.00		
	55.3	**25.1**	**N.A.**	**N.A.**	**59.4**	**39.4**	**32.6**			**62.1**	**65.4**	**107.7**	**99.7**	**101.6**	**118.8**	**111.6**		
	0.00	0.00			0.00	0.00	0.00			0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	8	*14*	*12*	*7*	*4*	*3*	*16*			*4*	*4*	*1*	*2*	*2*	*2*	*2*		
48	**58.4**	**71.4**	**112.7**	**92.1**	**69.1**		**36.2**	**30.9**		**16.1**	**30.0**	**80.8**	**100.1**	**134.7**		**117.6**	**45.0**	
	0.00	0.00	0.00	0.00	0.00		0.00	0.00		0.07	0.00	0.00	0.00	0.00		0.00	0.00	
	51.4	**63.1**	**99.5**	**N.A.**	**N.A.**		**28.1**	**29.9**		**45.8**	**55.8**	**96.9**	**111.8**	**122.0**		**113.7**	**62.9**	
	0.00	0.00	0.00				0.00	0.00		0.00	0.00	0.00	0.00	0.00		0.00	0.00	
	9	*14*	*12*	*16*	*16*		*16*	*3*		*4*	*4*	*1*	*2*	*2*		*2*	*2*	
60	**50.7**	**63.9**	**102.1**	**83.2**	**61.9**		**30.4**			**14.4**	**28.0**	**72.0**	**99.0**	**138.4**		**115.2**		
	0.00	0.00	0.00	0.00	0.00		0.00			0.11	0.00	0.00	0.00	0.00		0.00		
	45.6	**64.5**	**104.7**	**N.A.**	**N.A.**		**22.6**			**34.7**	**52.6**	**88.1**	**115.1**	**134.8**		**112.4**		

	0.00	0.00	0.00				0.01			0.00	0.00	0.00	0.00	0.00		0.00		
	9	*14*	*12*	*16*	*16*		*16*			*4*	*4*	*1*	*2*	*2*		*2*		
120	**37.9**	**64.6**	**87.1**	**78.9**	**32.8**		**15.4**	**25.8**	**38.1**	**17.8**	**12.3**	**72.3**	**92.8**	**137.1**		**79.7**	**30.9**	**11.7**
	0.00	0.00	0.00	0.00	0.00		0.08	0.00	0.00	0.04	0.20	0.00	0.00	0.00		0.00	0.00	0.23
	38.5	**80.1**	**118.0**	**102.2**	**39.9**		**18.4**	**29.2**	**61.5**	**26.2**	**8.2**	**94.8**	**114.0**	**135.6**		**84.4**	**32.8**	**11.6**
	0.00	0.00	0.00	0.00	0.00		0.03	0.00	0.00	0.00	0.52	0.00	0.00	0.00		0.00	0.00	0.24
	6	*6*	*6*	*6*	*6*		*6*	*6*	*6*	*4*	*9*	*1*	*1*	*2*		*1*	*1*	*1*
	Panel B. SIC is used as model selection method																	
	Panel A. Jan 52- Dec 78									*Panel B. Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	**114.0**									**153.4**								
	0.00									0.00								
	200.9									**176.0**								
	0.00									0.00								
	1									*1*								
3	**157.5**									**184.2**								
	0.00									0.00								
	208.2									**193.4**								
	0.00									0.00								
	1									*1*								
4	**105.2**	**67.5**								**195.3**	**199.6**							
	0.00	0.00								0.00	0.00							
	95.5	**67.0**								**207.5**	**201.6**							
	0.00	0.00								0.00	0.00							
	1	*1*								*1*	*1*							
6	**31.3**	**63.1**	**149.4**							**178.6**	**174.6**	**176.8**						
	0.00	0.00	0.00							0.00	0.00	0.00						
	28.9	**173.3**	**291.2**							**201.0**	**176.1**	**145.7**						
	0.00	0.00	0.00							0.00	0.00	0.00						
	1	*1*								*1*	*1*	*1*						
9	**28.4**		**162.0**							**142.7**		**166.2**						
	0.00		0.00							0.00		0.00						
	26.5		**276.7**							**186.0**		**145.6**						
	0.00		0.00							0.00		0.00						
	1		*1*							*1*		*1*						
12	**25.2**	**116.7**	**172.5**	**171.2**	**102.5**					**110.6**	**120.8**	**154.0**	**175.7**	**178.8**				
	0.00	0.00	0.00	0.00	0.00					0.00	0.00	0.00	0.00	0.00				
	27.1	**224.8**	**255.3**	**236.5**	**117.5**					**168.0**	**143.3**	**139.0**	**137.9**	**133.8**				
	0.00	0.00	0.00	0.00	0.00					0.00	0.00	0.00	0.00	0.00				
	1	*1*	*1*	*1*	*1*					*1*	*1*	*1*	*1*	*1*				
24	**50.9**	**95.7**	**108.7**	**124.7**	**90.2**		**64.6**			**59.1**	**89.3**	**128.3**	**156.2**	**162.3**		**122.8**		
	0.00	0.00	0.00	0.00	0.00		0.01			0.00	0.00	0.00	0.00	0.00		0.00		
	45.4	**104.7**	**125.9**	**128.2**	**77.8**		**34.2**			**92.7**	**99.9**	**117.5**	**131.8**	**130.0**		**105.1**		
	0.00	0.00	0.00	0.00	0.00		0.01			0.00	0.00	0.00	0.00	0.00		0.00		
	1	*1*	*2*	*1*	*1*		*3*			*1*	*1*	*1*	*1*	*1*		*1*		
36	**60.2**	**80.4**	**104.0**	**99.6**	**84.5**	**70.5**	**54.4**			**43.6**	**72.9**	**104.9**	**122.5**	**129.6**	**118.1**	**103.3**		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	54.5	80.0	107.3	101.1	76.6	39.4	32.1			75.3	90.2	107.7	116.5	121.9	128.0	118.5		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	2	*2*	*2*	*2*	*1*	*3*	*3*			*1*	*1*	*1*	*1*	*1*	*1*	*1*		
48	**71.0**	**85.6**	**106.9**	**100.4**	**76.3**		**44.9**	**21.1**		**33.9**	**54.1**	**80.9**	**99.6**	**121.2**		**99.9**	**35.8**	
	0.00	0.00	0.00	0.00	0.00		0.00	0.00		0.00	0.00	0.00	0.00	0.00		0.00	0.00	
	65.3	**83.6**	**106.6**	**98.7**	**72.7**		**30.0**	**20.7**		**54.3**	**74.6**	**96.9**	**112.1**	**130.4**		**113.4**	**41.9**	

	0.00	0.00	0.00	0.00	0.00		0.00	0.00		0.00	0.00	0.00	0.00	0.00		0.00	0.00	
	1	*2*	*2*	*2*	*1*		*3*	*1*		*1*	*1*	*1*	*1*	*1*		*1*	*1*	
60	**64.8**	**83.4**	**107.2**	**99.3**	**64.6**		**36.3**			**29.2**	**47.3**	**72.1**	**94.4**	**122.6**		**100.3**		
	0.00	0.00	0.00	0.00	0.00		0.00			0.00	0.00	0.00	0.00	0.00		0.00		
	62.5	83.2	106.9	97.8	64.6		30.1			44.0	64.7	88.1	108.2	132.7		109.4		
	0.00	0.00	0.00	0.00	0.00		0.00			0.00	0.00	0.00	0.00	0.00		0.00		
	1	*2*	*2*	*2*	*1*		*3*			*1*	*1*	*1*	*1*	*1*		*1*		
120	**31.7**	**67.9**	**97.8**	**91.5**	**37.9**		**12.7**	**31.7**	**46.3**	**27.9**	**47.9**	**72.4**	**92.9**	**115.1**		**79.7**	**30.9**	**11.7**
	0.00	0.00	0.00	0.00	0.00		0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.23
	35.5	**76.4**	**107.2**	**98.5**	**41.8**		**13.1**	**31.5**	**59.2**	**35.3**	**65.5**	**94.8**	**114.0**	**129.0**		**84.4**	**32.8**	**11.6**
	0.00	0.00	0.00	0.00	0.00		0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.24
	1	*1*	*1*	*1*	*1*		*2*	*1*	*2*	*1*	*1*	*1*	*1*	*1*		*1*	*1*	*1*

Note: There are five numbers for each maturity pair. First and second numbers are ARCH-LM test statistic and its p value that there is no first order ARCH effect in the residuals of the unrestricted VAR. The third and fourth numbers provide the same information, but for the restricted VAR. Fifth number is the VAR lag order. N.A. indicates was not estimated either because of the instability of the constrained VAR parameter or non-convergence of the iterative process.

Tables 1 provides evidence on the first order conditional heteroskedasticity in the VAR residuals, in particular multivariate ARCH-LM test statistics, described in Doornik & Hendry (1997), corresponding *p values* that there is conditional homoskedasticity and the lag lengths of the VAR in *italics*[8]. Panel A is based on VAR's that are selected by AIC while Panel B relies on SIC model selection method. There are two sets of results for each maturity pair, one for the unrestricted VAR and another for the restricted VAR. As expected the test always detects conditional heteroskedasticity in both restricted and unrestricted residuals, except for maturity pairs 12&120 and 60&120 months, suggesting that we are working with Case IV, as defined in Section 4.3. Therefore all the remaining analyses are based on the wild bootstrap rather than i.i.d. one.

Table 2. Multivariate autocorrelation test

	Panel A. AIC is used as model selection method																	
	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	**4.77**									**3.20**								
	0.31									0.52								
3	**1.80**									**9.45**								
	0.77									0.05								
4	**0.50**	**1.50**								**1.82**	**6.67**							
	0.97	0.83								0.77	0.15							
6	**1.24**	**7.08**	**4.89**							**2.15**	**6.45**	**2.72**						
	0.87	0.13	0.30							0.71	0.17	0.61						

[8] The maximum lag length used to select the VAR order is $p_{\max} = 12\left(\frac{T}{100}\right)^{1/4}$ which is more or less preferred in a Monte Carlo study of Schwert (1989).

9	**1.33**		**3.39**							**2.66**		**3.93**						
	0.86		0.50							0.62		0.42						
12	**1.54**	**2.32**	**3.03**	**4.31**	**2.77**					**3.50**	**5.16**	**5.27**	**4.09**	**1.56**				
	0.82	0.68	0.55	0.37	0.60					0.48	0.27	0.26	0.39	0.82				
24	**6.96**	**4.52**	**1.42**	**1.77**	**3.88**		**2.87**			**3.64**	**3.22**	**11.35**	**3.45**	**1.57**		**6.88**		
	0.14	0.34	0.84	0.78	0.42		0.58			0.46	0.52	0.02	0.49	0.81		0.14		
36	**5.85**	**2.43**	**4.86**	**4.55**	**2.42**	**3.50**	**4.30**			**3.47**	**2.67**	**8.82**	**2.18**	**1.53**	**2.70**	**5.62**		
	0.21	0.66	0.30	0.34	0.66	0.48	0.37			0.48	0.61	0.07	0.70	0.82	0.61	0.23		
48	**5.08**	**2.68**	**4.74**	**4.18**	**3.10**		**4.77**	**1.05**		**3.69**	**1.75**	**6.59**	**2.05**	**1.75**		**6.32**	**5.12**	
	0.28	0.61	0.31	0.38	0.54		0.31	0.90		0.45	0.78	0.16	0.73	0.78		0.18	0.28	
60	**5.93**	**2.88**	**5.19**	**3.97**	**2.53**		**3.74**			**2.89**	**1.00**	**5.37**	**2.91**	**2.53**		**8.47**		
	0.20	0.58	0.27	0.41	0.64		0.44			0.58	0.91	0.25	0.57	0.64		0.08		
120	**3.14**	**2.71**	**2.99**	**3.14**	**3.33**		**3.57**	**3.96**	**4.22**	**6.54**	**1.91**	**1.46**	**1.12**	**3.33**		**1.29**	**3.29**	**3.42**
	0.53	0.61	0.56	0.53	0.50		0.47	0.41	0.38	0.16	0.75	0.83	0.89	0.50		0.86	0.51	0.49

Panel B. SIC is used as model selection method

	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	**11.74**									**1.81**								
	0.02									0.77								
3	**8.83**									**1.45**								
	0.07									0.84								
4	**7.35**	**0.83**								**1.02**	**2.14**							
	0.12	0.93								0.91	0.71							
6	**9.27**	**3.55**	**12.69**							**0.90**	**1.76**	**2.72**						
	0.05	0.47	0.01							0.92	0.78	0.61						
9	**9.61**		**8.41**							**2.40**		**3.93**						
	0.05		0.08							0.66		0.42						
12	**9.89**	**5.77**	**6.87**	**5.35**	**1.19**					**4.50**	**5.16**	**5.27**	**4.09**	**1.56**				
	0.04	0.22	0.14	0.25	0.88					0.34	0.27	0.26	0.39	0.82				
24	**14.26**	**8.92**	**3.95**	**6.98**	**5.82**		**2.87**			**8.15**	**10.25**	**11.34**	**10.67**	**8.69**		**6.93**		
	0.01	0.06	0.41	0.14	0.21		0.58			0.09	0.04	0.02	0.03	0.07		0.14		
36	**5.48**	**4.58**	**4.31**	**4.50**	**8.01**	**3.50**	**0.46**			**7.80**	**8.55**	**8.82**	**8.53**	**6.73**	**5.23**	**6.20**		
	0.24	0.33	0.37	0.34	0.09	0.48	0.98			0.10	0.07	0.07	0.07	0.15	0.26	0.18		
48	**12.92**	**3.69**	**3.89**	**4.57**	**8.77**		**0.84**	**10.71**		**5.80**	**6.46**	**6.59**	**5.90**	**4.42**		**4.27**	**7.63**	
	0.01	0.45	0.42	0.33	0.07		0.93	0.03		0.21	0.17	0.16	0.21	0.35		0.37	0.11	
60	**12.76**	**3.38**	**4.44**	**5.88**	**9.69**		**1.66**			**4.72**	**5.25**	**5.37**	**4.69**	**3.40**		**3.06**		
	0.01	0.50	0.35	0.21	0.05		0.80			0.32	0.26	0.25	0.32	0.49		0.55		
120	**12.61**	**10.66**	**12.18**	**12.89**	**13.43**		**11.53**	**12.01**	**8.91**	**1.93**	**1.65**	**1.46**	**1.12**	**0.90**		**1.29**	**3.29**	**3.42**
	0.01	0.03	0.02	0.01	0.01		0.02	0.02	0.06	0.75	0.80	0.83	0.89	0.92		0.86	0.51	0.49

Note: Table provides first order autocorrelation test results for the VAR residuals. The first number, given in **bold** is the statistic followed by corresponding *p value* that there is no autocorrelation. The VAR lag length is given in Table 1 and the test is described in the Appendix A.

Table 2 provides evidence on the first order autocorrelation in the VAR residuals using a test robust to conditional heteroskedasticity that is explained in the Appendix. Using AIC model selection method seems to be better than SIC in making sure that all the dynamic relationships are captured by the mean of the model, i.e. there is no autocorrelation in the residuals of VAR's that are chosen by AIC but there is some evidence when SIC is used. But this conclusion must be treated with caution since there is still a possibility that a model to be over-specified with AIC. More evidence on higher order autocorrelation and/or use of interactive model selection method and autocorrelation test is desirable but not attempted as this stage, because of the computational cost.

Table 3. Number of iterations required in stationarity correction

Panel A. AIC is used as model selection method

	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	72									68								
3	63									65								
4	65	63								24	45							
6	66	75	75							0	35	0						
9	62		73							0		0						
12	62	73	72	63	63					1	0	0	0	0				
24	61	74	76	73	81		57			8	14	0	0	0		0		
36	62	72	75	72	60	59	85			8	17	0	0	0	0	0		
48	66	73	76	76	85		87	55	.	4	15	0	0	0		0	48	
60	66	75	77	75	85		87			6	19	0	0	0		0		
120	71	68	70	70	69		66	62	50	12	58	0	0	0		0	0	0

Panel B. SIC is used as model selection method

	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	0									0								
3	0									0								
4	0	0								0	0							
6	0	0	0							0	0	0						
9	0		0							0		0						
12	0	0	0	0	0					0	0	0	0	0				
24	0	0	43	0	0		57			0	0	0	0	0		0		
36	48	50	50	50	0	59	60			0	0	0	0	0	0	0		
48	0	54	54	54	0		61	0		0	0	0	0	0		0	0	
60	0	56	56	55	0		61			0	0	0	0	0		0		
120	0	0	0	0	0		53	0	26	0	0	0	0	0		0	0	0

Note: Table reports the number of iterations required to make the bias corrected unrestricted VAR parameter stable. Details of this procedure are given in Section 4.1. Step 1.7.

Table 3 reports the number of iterations required to make the bias-corrected companion-form VAR parameters stable. When SIC is used Step 1.7 introduced in Section 4.1. seems to have no effect in the second sub-sample but we do use this step a lot in the first sub-period, especially for maturity pairs that are not at the short end of the term structure. It is interesting to note that for the VAR's selected by AIC the introduction of this step proves to be more useful, again more often in the first sub-sample. Perhaps this is due to the persistence in the interest rates.

The finite sample distributions of the DM and LM test statistics are almost the same, therefore all the interpretations below are made with respect to the LM and Wald tests only. Figures 1 to 4 in Appendix B illustrate asymptotic and estimates of the finite sample distributions of the test statistics along with the histogram of the endogenized VAR lag lengths. In line with B & H (2001), for most VAR's the finite sample distribution of the LM test is being better approximated by the asymptotic *chi-squared* distribution than that of the Wald test. But working with more maturity pairs than in B & H (2001) reveals that this is not uniformly true, in many cases differences are minor and we do encounter cases where such ordering is difficult to make, for example for maturity pairs 3&6, 3&24 and 6&12 months in the second sub-sample with AIC model selection method. One can also observe extremely bad approximations to both finite sample distributions, such as for the pair 12&48 months in the first sub-period with SIC. Based on this evidence it seems reasonable to reiterate the recommendation given in Sarno *et al* (2006), which is to avoid the use of the asymptotic distribution in making an inference based on either LM or Wald test statistic.

When SIC is used as model selection rule, allowing the lag lengths determined in each bootstrap replication does not seem to alter much the conclusion that would have drawn from the exogenous lag length methodology. In fact, there is no single case in the second sub-sample where the dominant lag length, which is one, is chosen with less than 97% probability.

Given a model selection method it is not obvious how endogenizing the lag selection rule affects the approximation provided by the asymptotic theory. But most extreme departures between the asymptotic and estimates of finite sample distributions seem to occur with AIC model selection method, probably because it is less certain about the assumed true lag order.

Based on the figures one might hypothesize that the difference between finite sample distributions of the LM and Wald test statistics are decreasing functions of the distance between the short and long term interest rates that are included in the VAR. This seems quite plausible, especially in the second sub-sample.

Table 4. LM test of the EH of term structure

Panel A. AIC is used as model selection method

	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	**57.72**									**56.17**								
	0.00									0.00								
	0.00									0.00								
	201.75									**152.91**								
	0.00									0.00								
	0.00									0.00								
3	**47.54**									**29.59**								
	0.00									0.04								
	0.00									U.S.								
	113.21									**84.73**								
	0.00									0.00								
	0.00									U.S.								
4	**45.64**	**33.20**								**18.15**	**16.22**							
	0.00	0.02								0.01	0.04							
	0.00	0.06								0.04	0.16							
	92.81	**65.68**								**31.45**	**41.00**							
	0.00	0.00								0.00	0.00							
	0.01	0.05								0.07	0.23							
6	**46.40**	**44.19**	**38.04**							**14.26**	**12.48**	**7.86**						
	0.00	0.03	0.10							0.03	0.13	0.02						
	0.00	0.01	0.05							0.14	0.24	0.15						
	69.33	**120.07**	**99.56**							**23.73**	**29.79**	**13.52**						
	0.00	0.00	0.00							0.00	0.00	0.00						
	0.02	0.01	0.01							0.22	0.28	0.20						
9	**36.36**		**31.06**							**12.60**		**6.17**						
	0.01		0.31							0.05		0.05						
	0.05		N.S.							0.16		0.24						
	66.48		**83.97**							**22.02**		**9.59**						
	0.00		0.0							0.00		0.01						
	0.05		N.S.							0.22		0.34						
12	**32.62**	**31.80**	**31.2**	**25.82**	**24.95**					**11.45**	**6.15**	**5.88**	**5.66**	**5.83**				
	0.02	0.28	0.3	0.10	0.13					0.18	0.05	0.05	0.06	0.05				
	0.06	N.C.	N.S.	0.33	0.29					0.19	0.20	0.25	0.28	0.22				
	62.46	**69.49**	**69.7**	**56.42**	**53.43**					**21.59**	**8.41**	**8.52**	**8.57**	**7.87**				
	0.00	0.00	0.0	0.00	0.00					0.01	0.01	0.01	0.01	0.02				
	0.08	N.C.	N.S.	0.22	0.17					0.17	0.33	0.33	0.34	0.26				
24	**19.29**	**27.75**	**32.44**	**32.49**	**28.79**		**16.92**			**13.35**	**12.49**	**5.49**	**13.68**	**14.09**		**13.97**		
	0.25	0.27	0.35	0.35	0.53		0.01			0.10	0.13	0.06	0.01	0.01		0.01		
	0.32	0.45	N.S.	N.S.	0.53		0.32			0.10	0.29	0.19	0.11	0.06		0.03		
	38.50	**59.90**	**46.55**	**76.75**	**53.19**		**16.52**			**18.55**	**18.84**	**7.45**	**24.50**	**16.33**		**13.82**		
	0.00	0.00	0.03	0.00	0.01		0.01			0.02	0.02	0.02	0.00	0.00		0.01		
	0.21	0.33	N.S.	N.S.	0.41		0.45			0.13	0.31	0.24	0.13	0.17		0.07		
36	**18.31**	**26.94**	**23.36**	**15.03**	**8.34**	**10.37**	**42.75**			**12.86**	**12.24**	**6.27**	**12.28**	**11.87**	**11.30**	**9.71**		
	0.31	0.52	0.50	0.38	0.40	0.11	0.10			0.12	0.14	0.04	0.02	0.02	0.02	0.05		

	0.41	N.S.	0.68	0.40	0.64	0.43	0.02			0.11	0.29	0.18	0.14	0.11	0.07	0.09		
	30.91	**60.45**	**42.61**	**16.56**	**4.11**	**7.06**	**208.12**			**17.47**	**18.14**	**7.40**	**21.60**	**13.03**	**10.25**	**8.14**		
	0.01	0.00	0.01	0.28	0.85	0.32	0.00			0.03	0.02	0.02	0.00	0.01	0.04	0.09		
	0.33	N.S.	0.47	0.51	0.91	0.69	0.01			0.14	0.24	0.24	0.16	0.19	0.17	0.19		
48	**21.80**	**25.26**	**23.24**	**29.78**			**41.31**	**9.37**		**11.61**	**10.71**	**5.62**	**9.81**	**8.93**		**6.67**	**5.19**	
	0.24	0.61	0.51	0.58			0.13	0.15		0.17	0.22	0.06	0.04	0.06		0.15	0.27	
	0.24	0.69	0.62	N.S.			0.03	0.39		0.11	0.26	0.18	0.12	0.14		0.23	0.40	
	32.48	**63.45**	**48.29**	**123.59**			**242.37**	**11.64**		**15.40**	**14.94**	**6.46**	**14.44**	**8.46**		**5.00**	**3.16**	
	0.02	0.00	0.00	0.00			0.00	0.07		0.05	0.06	0.04	0.01	0.08		0.29	0.53	
	0.30	0.34	0.34	N.S.			0.01	0.39		0.14	0.26	0.26	0.16	0.24		0.38	0.64	
60	**21.96**	**25.23**	**22.77**	**31.38**			**41.09**			**10.95**	**9.68**	**5.21**	**8.22**	**7.26**		**5.36**		
	0.23	0.62	0.53	0.50			0.13			0.20	0.29	0.07	0.08	0.12		0.25		
	0.24	0.67	0.63	N.S.			0.01			0.10	0.26	0.23	0.15	0.18		0.27		
	30.72	**68.17**	**55.70**	**147.36**			**273.74**			**13.75**	**12.41**	**5.50**	**10.67**	**6.37**		**3.91**		
	0.03	0.00	0.00	0.00			0.00			0.09	0.13	0.06	0.03	0.17		0.42		
	0.37	0.24	0.30	N.S.			0.00			0.13	0.32	0.29	0.18	0.27		0.39		
120	**17.42**	**14.50**	**15.79**	**17.07**	**16.81**		**21.61**	**23.91**	**30.14**	**7.98**	**11.11**	**3.28**	**2.68**	**2.42**		**0.67**	**0.39**	**0.45**
	0.13	0.27	0.20	0.15	0.16		0.04	0.02	0.00	0.44	0.89	0.19	0.26	0.66		0.71	0.82	0.80
	0.24	0.37	0.36	0.27	0.29		0.14	0.11	0.03	0.24	0.51	0.37	0.42	0.56		0.81	0.91	0.88
	15.80	**13.48**	**14.84**	**16.32**	**15.79**		**22.61**	**26.14**	**82.80**	**11.78**	**14.09**	**2.08**	**1.44**	**2.03**		**0.55**	**0.53**	**0.32**
	0.20	0.33	0.25	0.18	0.20		0.03	0.01	0.00	0.16	0.72	0.35	0.49	0.73		0.76	0.77	0.85
	0.47	0.56	0.54	0.51	0.48		0.35	0.29	0.02	0.17	0.45	0.54	0.64	0.57		0.84	0.87	0.90

Panel B. SIC is used as model selection method

	Jan 52- Dec 78									*Jan 82- Dec 2003*								
	1	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***	***1***	***2***	***3***	***4***	***6***	***9***	***12***	***24***	***60***
2	**30.81**									**0.02**								
	0.00									0.99								
	0.00									0.82								
	91.25									**23.57**								
	0.00									0.00								
	0.00									0.00								
3	**30.70**									**9.54**								
	0.00									0.01								
	0.00									0.00								
	68.00									**16.14**								
	0.00									0.00								
	0.00									0.01								
4	**27.37**	**20.87**								**8.28**	**8.38**							
	0.00	0.00								0.02	0.02							
	0.00	0.00								0.00	0.01							
	45.30	**26.59**								**12.76**	**15.50**							
	0.00	0.00								0.00	0.00							
	0.00	0.00								0.01	0.09							
6	**20.91**	**19.74**	**N.C.**							**7.56**	**7.87**	**7.86**						
	0.00	0.00								0.02	0.02	0.02						
	0.00	0.00								0.01	0.03	0.00						
	23.88	**16.46**								**10.41**	**12.23**	**13.49**						
	0.00	0.00								0.01	0.00	0.00						
	0.00	0.01								0.04	0.07	0.03						
9	**13.10**		**8.09**							**6.70**		**6.17**						

	0.00		0.02							0.04		0.05						
	0.00		0.02							0.00		0.02						
	16.32		**18.17**							**8.59**		**9.59**						
	0.00		0.00							0.01		0.01						
	0.00		0.01							0.07		0.12						
12	**8.89**	**7.29**	**6.17**	**5.58**	**5.43**					**6.36**	**6.15**	**5.88**	**5.66**	**5.83**				
	0.01	0.03	0.05	0.06	0.07					0.04	0.05	0.05	0.06	0.05				
	0.01	0.04	0.02	0.07	0.07					0.02	0.02	0.03	0.04	0.03				
	10.39	**9.98**	**10.51**	**9.86**	**6.25**					**7.90**	**8.39**	**8.53**	**8.57**	**7.87**				
	0.01	0.01	0.01	0.01	0.04					0.02	0.02	0.01	0.01	0.02				
	0.03	0.06	0.03	0.04	0.09					0.08	0.15	0.11	0.11	0.10				
24	**4.52**	**4.11**	**4.69**	**4.09**	**3.62**		**16.92**			**5.85**	**5.73**	**5.49**	**5.25**	**4.99**		**5.02**		
	0.10	0.13	0.32	0.13	0.16		0.01			0.05	0.06	0.06	0.07	0.08		0.08		
	0.11	0.16	0.21	0.15	0.16		0.01			0.06	0.06	0.05	0.06	0.07		0.10		
	3.65	**2.96**	**3.40**	**3.10**	**2.52**		**16.53**			**6.85**	**7.30**	**7.46**	**6.95**	**4.97**		**3.37**		
	0.16	0.23	0.49	0.21	0.28		0.01			0.03	0.03	0.02	0.03	0.08		0.19		
	0.17	0.27	0.34	0.23	0.26		0.05			0.06	0.13	0.09	0.09	0.13		0.21		
36	**3.77**	**3.20**	**3.55**	**4.03**	**2.01**	**10.37**	**13.29**			**6.25**	**6.35**	**6.27**	**6.14**	**5.89**	**5.84**	**5.97**		
	0.44	0.53	0.47	0.40	0.37	0.11	0.04			0.04	0.04	0.04	0.05	0.05	0.05	0.05		
	0.36	0.44	0.34	0.31	0.40	0.07	0.03			0.03	0.03	0.03	0.03	0.04	0.06	0.06		
	2.97	**2.08**	**2.22**	**2.56**	**1.53**	**7.06**	**12.74**			**6.98**	**7.29**	**7.42**	**6.96**	**5.19**	**3.96**	**3.58**		
	0.56	0.72	0.70	0.63	0.46	0.32	0.05			0.03	0.03	0.02	0.03	0.07	0.14	0.17		
	0.46	0.63	0.55	0.49	0.47	0.30	0.09			0.06	0.08	0.08	0.10	0.11	0.18	0.18		
48	**2.89**	**2.73**	**3.03**	**3.50**	**1.31**		**8.74**	**1.90**		**5.67**	**5.81**	**5.62**	**5.31**	**4.73**		**4.33**	**5.26**	
	0.24	0.60	0.55	0.48	0.52		0.19	0.39		0.06	0.05	0.06	0.07	0.09		0.11	0.07	
	0.26	0.50	0.49	0.45	0.51		0.13	0.39		0.05	0.05	0.03	0.07	0.11		0.14	0.11	
	2.36	**1.72**	**1.82**	**2.09**	**1.05**		**8.50**	**1.86**		**6.48**	**6.72**	**6.49**	**5.52**	**3.80**		**2.52**	**3.20**	
	0.31	0.79	0.77	0.72	0.59		0.20	0.39		0.04	0.03	0.04	0.06	0.15		0.28	0.20	
	0.33	0.72	0.68	0.62	0.60		0.25	0.40		0.05	0.09	0.09	0.14	0.18		0.30	0.23	
60	**2.82**	**2.84**	**3.13**	**3.54**	**1.20**		**7.13**			**5.51**	**5.51**	**5.21**	**4.84**	**4.15**		**3.57**		
	0.24	0.59	0.54	0.47	0.55		0.31			0.06	0.06	0.07	0.09	0.13		0.17		
	0.28	0.53	0.45	0.47	0.56		0.31			0.05	0.06	0.07	0.08	0.14		0.21		
	2.29	**1.75**	**1.87**	**2.09**	**0.97**		**7.48**			**6.16**	**6.13**	**5.53**	**4.47**	**2.95**		**2.02**		
	0.32	0.78	0.76	0.72	0.62		0.28			0.05	0.05	0.06	0.11	0.23		0.36		
	0.34	0.69	0.65	0.66	0.62		0.35			0.05	0.11	0.11	0.15	0.28		0.39		
120	**2.99**	**2.58**	**2.67**	**2.59**	**1.94**		**6.93**	**1.27**	**17.77**	**4.29**	**3.87**	**3.28**	**2.68**	**1.68**		**0.67**	**0.39**	**0.45**
	0.22	0.28	0.26	0.27	0.38		0.14	0.53	0.00	0.12	0.14	0.19	0.26	0.43		0.71	0.82	0.80
	0.28	0.28	0.26	0.28	0.39		0.08	0.54	0.00	0.10	0.14	0.24	0.28	0.47		0.75	0.86	0.85
	2.11	**1.49**	**1.53**	**1.53**	**1.30**		**5.49**	**1.13**	**25.48**	**4.12**	**3.13**	**2.10**	**1.45**	**0.88**		**0.55**	**0.53**	**0.33**
	0.35	0.47	0.46	0.46	0.52		0.24	0.57	0.00	0.13	0.21	0.35	0.48	0.64		0.76	0.77	0.85
	0.34	0.44	0.40	0.44	0.52		0.14	0.56	0.00	0.13	0.24	0.40	0.49	0.67		0.77	0.80	0.87

Note: There are two set of results, one from the LM test, the other from the Wald test, for each maturity pair. First number in each set is the test statistic and second and third numbers are asymptotic and finite sample p-values. The cells, that are highlighted indicate the rejection of the EH at 5% empirical significance level. N.S. means instability of the restricted VAR parameter, N.C. means non-convergence of the iterative procedure

Results of the LM and Wald tests of the expectations hypothesis of term structure are provided in Table 4. As one can see the LM and DM tests are not always working, especially with AIC model selection method and/or in the first sub-sample. The reasons are instability of the restricted VAR parameter, and non-convergence.

However, our study is not the first one to point out this problem. The former problem is observed in Bekaert *et al* (2006) and the latter is present in Sarno *et al* (2006). In this case the asymptotic Wald test emerges as the only way to make an inference, which is very likely to be far off than that of the finite sample Wald test. It seems that the theory receives a strong rejection in either at the extreme short end of the term structure or the longest end. As hypothesized earlier the EH seems to be getting more support in the second sub-sample in contrast to Campbell & Shiller (1991)'s claim that it worked better prior to 1978. But notice that the second sub-sample does not include the Fed's reserve targeting policy period.

Another observation one can make is that most rejections occur where k is small, i.e. when the number of roll over investments are small. Together with more rejections at the short end of the maturity spectrum this results might be suggesting because of the transaction costs noise traders specifically concentrate on these portions of the term structure influencing the market to operate away from what the theory would predict.

It is interesting to compare the LM test result of Panel A with its counterpart in Table 8 of Sarno *et al* (2006). If one ignores instability problem, which is non-existent in Sarno *et al* (2006), and marginal inferences, there seems to be pretty good agreement between 2 tables in that the EH is rejected at the short and longest end of the maturity spectrum in the first sub-sample, and only at the short end in the second sub-sample. Thus imposing not only the EH but also conditional homoskedasticity and exogenous lag order selection method to generate data under the null hypothesis does not seem to entail a large inferential difference. But as one would expect the EH receives more support in our study and more so in the second sub-sample. It is also relieving to note that this inference remains almost intact when SIC is used as the model selection method.

Reflecting the poor approximation of the finite sample distributions provided by the first order asymptotic theory, the differences between asymptotic and finite sample inferences are non-negligible. Confirming B&H result, most extreme departures occur with the Wald test. When AIC is used there are 17 inferential differences in the first sub-sample and 27 differences in the second sub-sample at 5% significance level with the Wald test as opposed to 9 and 12 with the LM test.

However, when SIC is used the differences become much fewer, 2 and 19 with the Wald test and 1 and 11 with the LM test statistic.

Interestingly, in terms of the inferential difference between the finite sample Wald and LM tests things are reversed, they meet more often when AIC not the SIC is used as a model selection method. There are 1 and 2 differences with AIC as opposed to 4 and 16, in the first and second sub-sample respectively.

4. Conclusion

We extend the vector autoregression (VAR) based expectations hypothesis test of term structure considered in B & H (2001) using recent developments in the bootstrap literature. Firstly, we use wild bootstrap to allow for conditional heteroskedasticity in the VAR residuals without imposing any strict parameterization. Secondly, when making a finite sample inference, we endogenize the model selection procedure, employ the restricted not the unrestricted VAR residuals and randomize the initial condition in the bootstrap replications to reflect the true uncertainty. Finally, stationarity correction is introduced in order to take account of the possible explosive VAR parameters after they have been adjusted for the 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 find less rejections for the theory in the second sub-sample and when it is rejected it occurs at the very short and/or long end of the maturity spectrum. It is also relieving to note that this inference seems to be robust to both model selection methods used in this study. In terms of the conclusions made about the validity of the EH of term structure, the main difference between this study and its counterpart of Sarno *et al* (2006), which uses the original B&H methodology, is that we reject the theory less often than they do. This is probably as one would expect, since our null hypothesis includes only the EH, what we are interested in, not the conditional homoskedasticity and exogenous VAR lag length hypotheses on top that of the EH.

Appendix A. Multivariate autocorrelation test robust to conditional heteroskedasticity

In this appendix we describe Godfrey & Tremayne (2005) autocorrelation test robust to conditional heteroskedasticity, generalized into multivariate framework in Bataa (2006). Consider a general dynamic system of *n* stochastic equations, the residuals of which are being suspected to have autocorrelation,

$$\mathbf{Y}_0 = \mathbf{Z}_0\mathbf{B}_0 + \mathbf{U}_0 \qquad (1)$$

where $\underset{T\times n}{\mathbf{Y}_i} = \left[\mathbf{y}_{1+i},\ldots,\mathbf{y}_{T+i}\right]'$, $\underset{[np+m]\times n}{\mathbf{B}_0} = \left[\mathbf{A}_1,\ldots,\mathbf{A}_p,\mathbf{\Pi}\right]'$, $\underset{T\times m}{\mathbf{X}} = \left[\mathbf{x}_1,\ldots,\mathbf{x}_T\right]'$,

$\underset{T\times n}{\mathbf{U}_i} = \left[\mathbf{u}_{1+i},\ldots,\mathbf{u}_{T+i}\right]'$, $\underset{T\times[np+m]}{\mathbf{Z}_0} = \left[\mathbf{Y}_{-1},\ldots,\mathbf{Y}_{-p},\mathbf{X}\right]$, $\mathbf{y}_t$ 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 $\left|\mathbf{I} - \mathbf{A}_1 z - \mathbf{A}_2 z^2 \ldots - \mathbf{A}_p z^p\right| = 0$ 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 (1).

If there is autocorrelation of order *g* in $\mathbf{U}_0$, it follows a VAR(*g*) without an intercept,

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

A model dependent autocorrelation test is then performed by calculating the least squares residuals from (1) and evaluating the following auxiliary system,

$$\mathbf{Y}_0 = \mathbf{Z}\mathbf{B} + \mathbf{E} \qquad (2)$$

where $\underset{T\times[n(g+k)+m]}{\mathbf{Z}} = \left[\mathbf{Y}_{-1},\ldots,\mathbf{Y}_{-p},\mathbf{X},\hat{\mathbf{U}}_{-1},\ldots,\hat{\mathbf{U}}_{-g}\right]$,

$\underset{[n(p+g)+m]\times n}{\mathbf{B}} = \left[\mathbf{A}_1,\ldots,\mathbf{A}_p,\mathbf{\Pi},\mathbf{C}_1,\ldots,\mathbf{C}_g\right]'$ then the least squares estimator of (2) is $\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ as in familiar univariate case. If we let $\hat{\boldsymbol{\beta}} = vec(\hat{\mathbf{B}})$ then it can be shown,

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}), \qquad 3)$$

with $\mathbf{V} = \mathbf{I}_n \otimes \mathbf{\Gamma}$ and $\mathbf{\Gamma} = p\lim \mathbf{Z}'\mathbf{Z}/T$ under suitable regularity conditions, see Bataa (2006).

Multivariate extension of Godfrey and Tremayne (2005) test is now performed by testing the null hypothesis H$_0$: $\mathbf{C}_1 = \ldots = \mathbf{C}_g = 0$. This test for residual autocorrelation in (1) is equavalent to the Wald test of the hypothesis testing

H$_0$: $\mathbf{R\beta} = \mathbf{0}$ against $\mathbf{R\beta} \neq \mathbf{0}$,

where **R** is a $n^2k\times n^2(p+k)$ 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.

When there is heteroskedsticity in the resuduals, $\boldsymbol{\Sigma}_e$ is no longer the same for all t, and White's (1980) heteroskedsticity consistent covariance matrix estimator in the univarite framework is extended into the current setup by consistently estimating **W** as

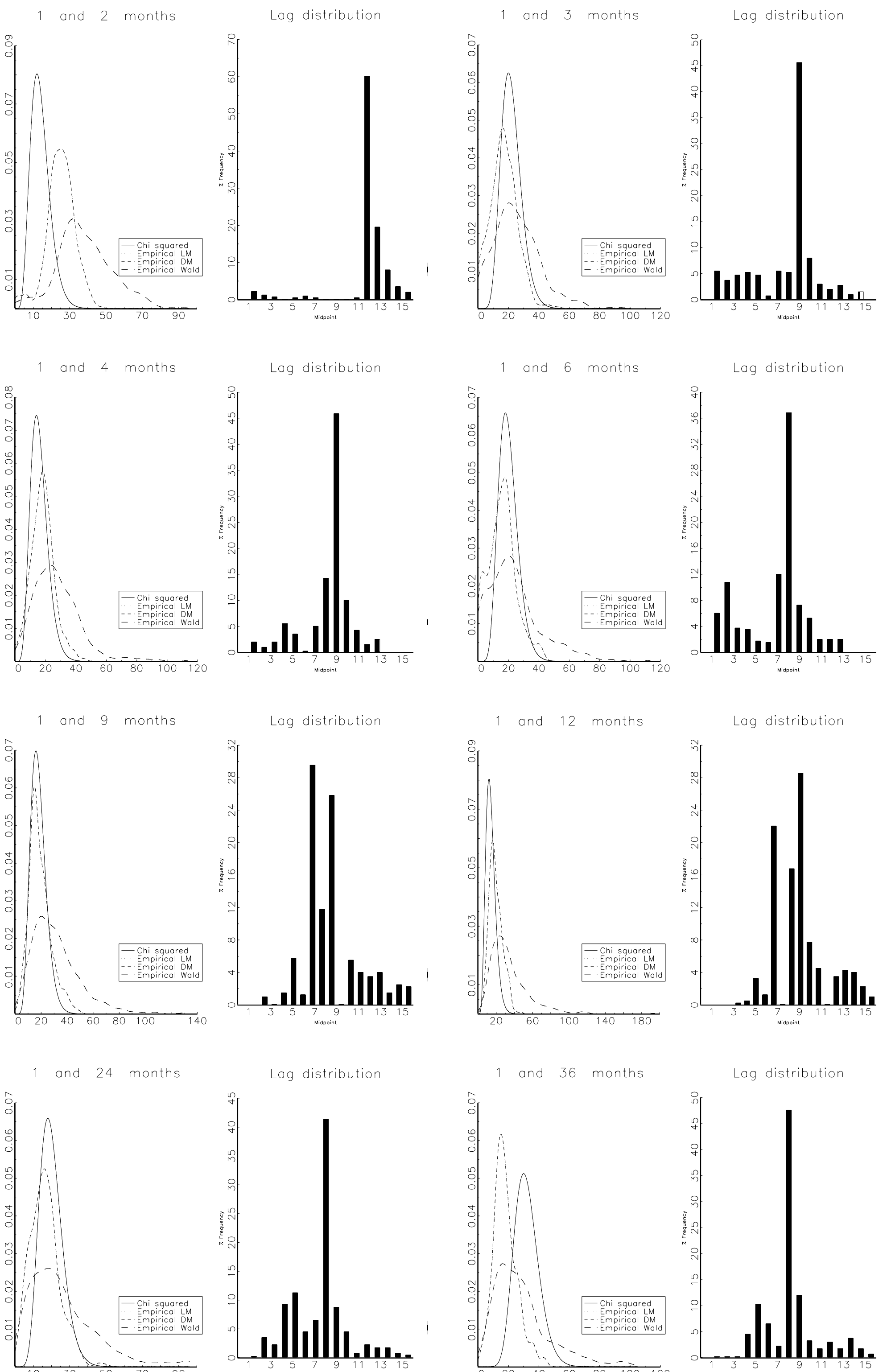
$$\hat{\mathbf{W}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{e}}_t\hat{\mathbf{e}}_t' \otimes \mathbf{z}_t'\mathbf{z}_t \ , \tag{4}$$

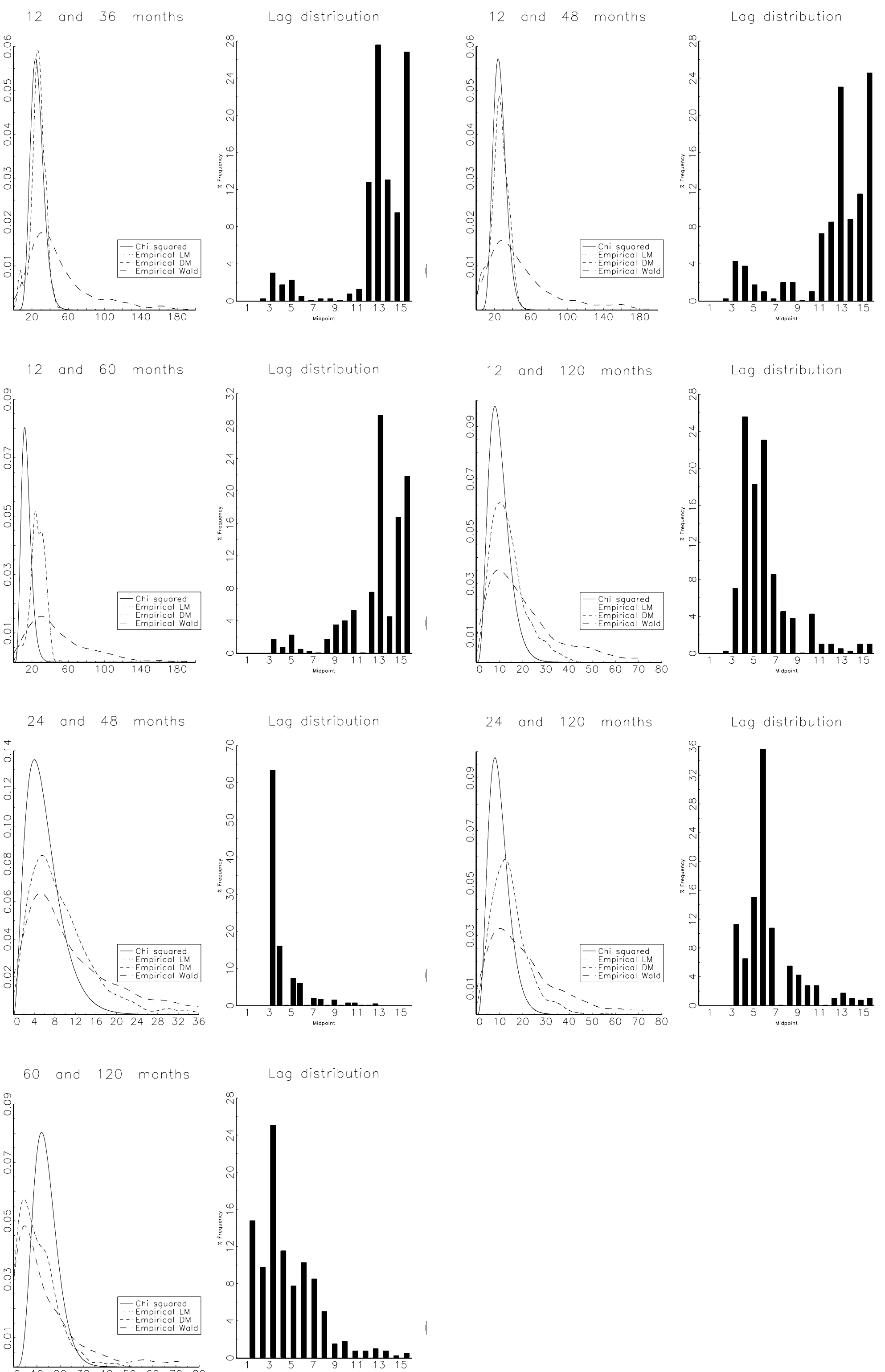
where $\mathbf{z}_t$ is t^{th} row of **Z**. Note that the estimator reduces to the one suggested in White (1980, p.820) when $n=1$ and to the estimator in the previous paragraph under homoskedasticty. We also replace $\hat{\mathbf{e}}_t$ by $\hat{\mathbf{u}}_t$, which is found to improve finite sample inference in the univariate framework as discussed in Davidson and MacKinnon (1985) and Godfrey and Orme (2004). From (3) and (4), the multivariate GT test has the following asymptotic distribution

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

Appendix B. Selected empirical distributions of test statistics and lag length based on AIC

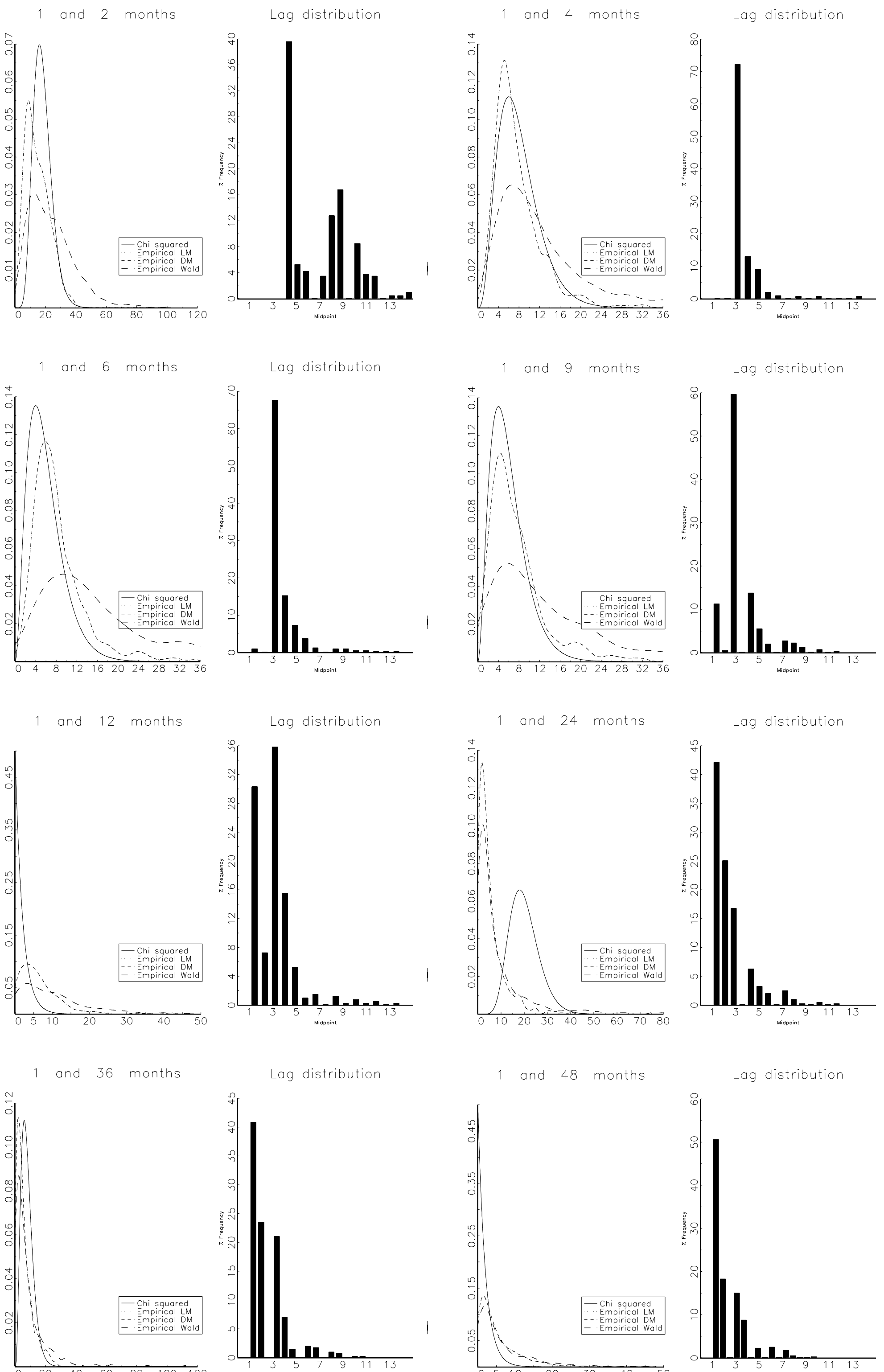
Jan 1952- Dec 1978

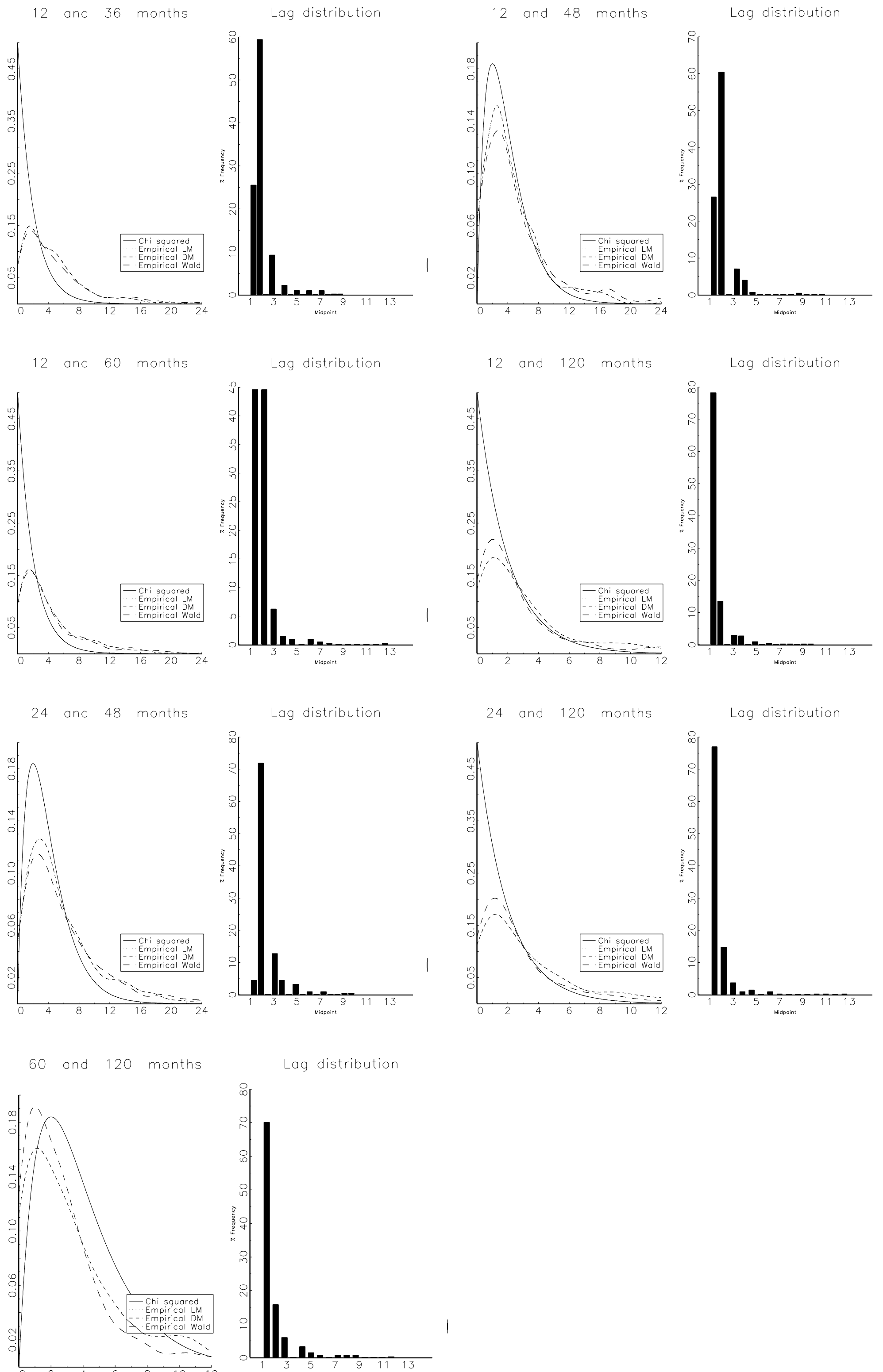




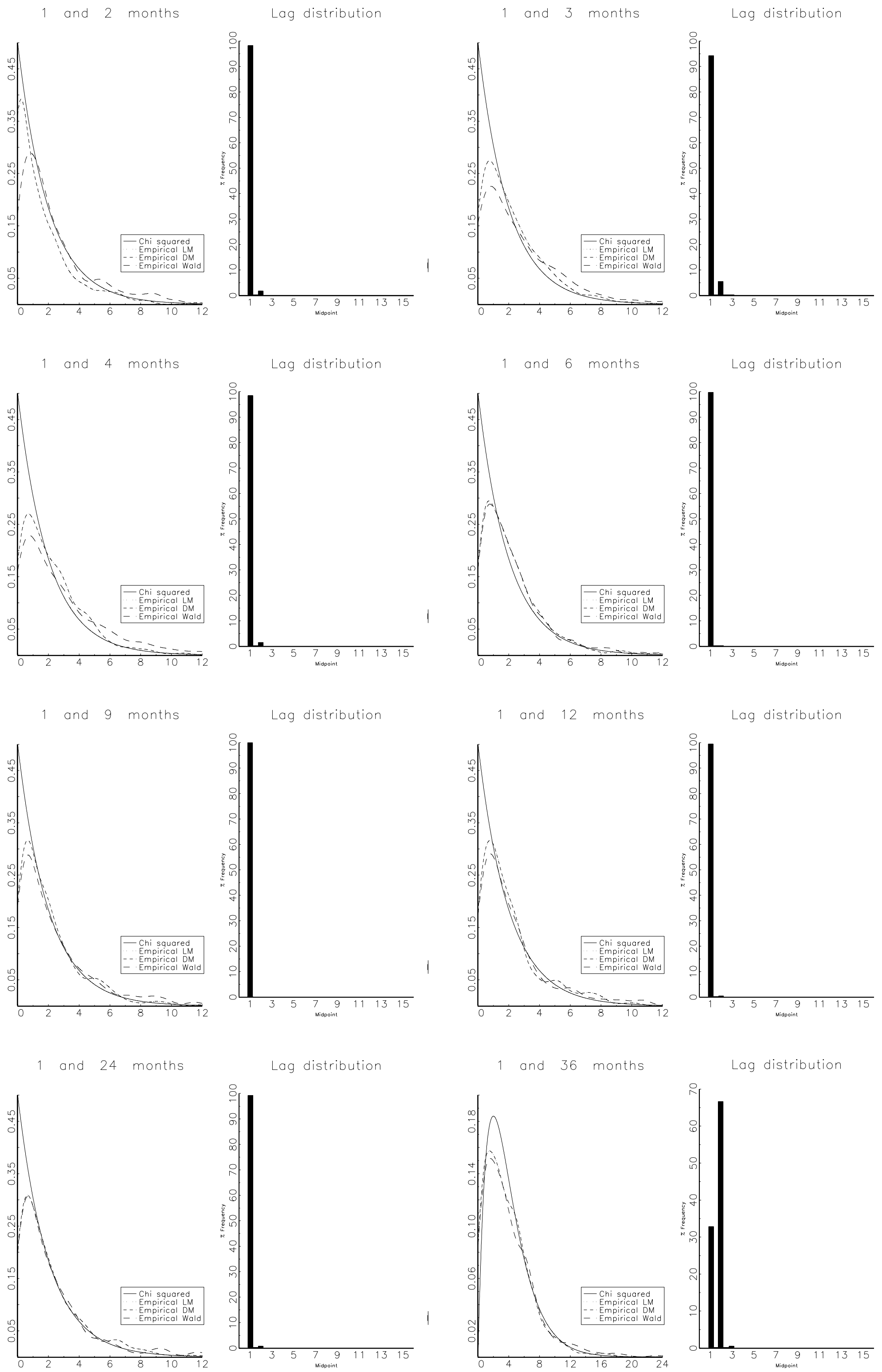
12 and 36 months
Lag distribution
12 and 48 months
Lag distribution
12 and 60 months
Lag distribution
12 and 120 months
Lag distribution
24 and 48 months
Lag distribution
24 and 120 months
Lag distribution
60 and 120 months
Lag distribution
Chi squared
Empirical LM
Empirical DM
Empirical Wald
% Frequency
Midpoint

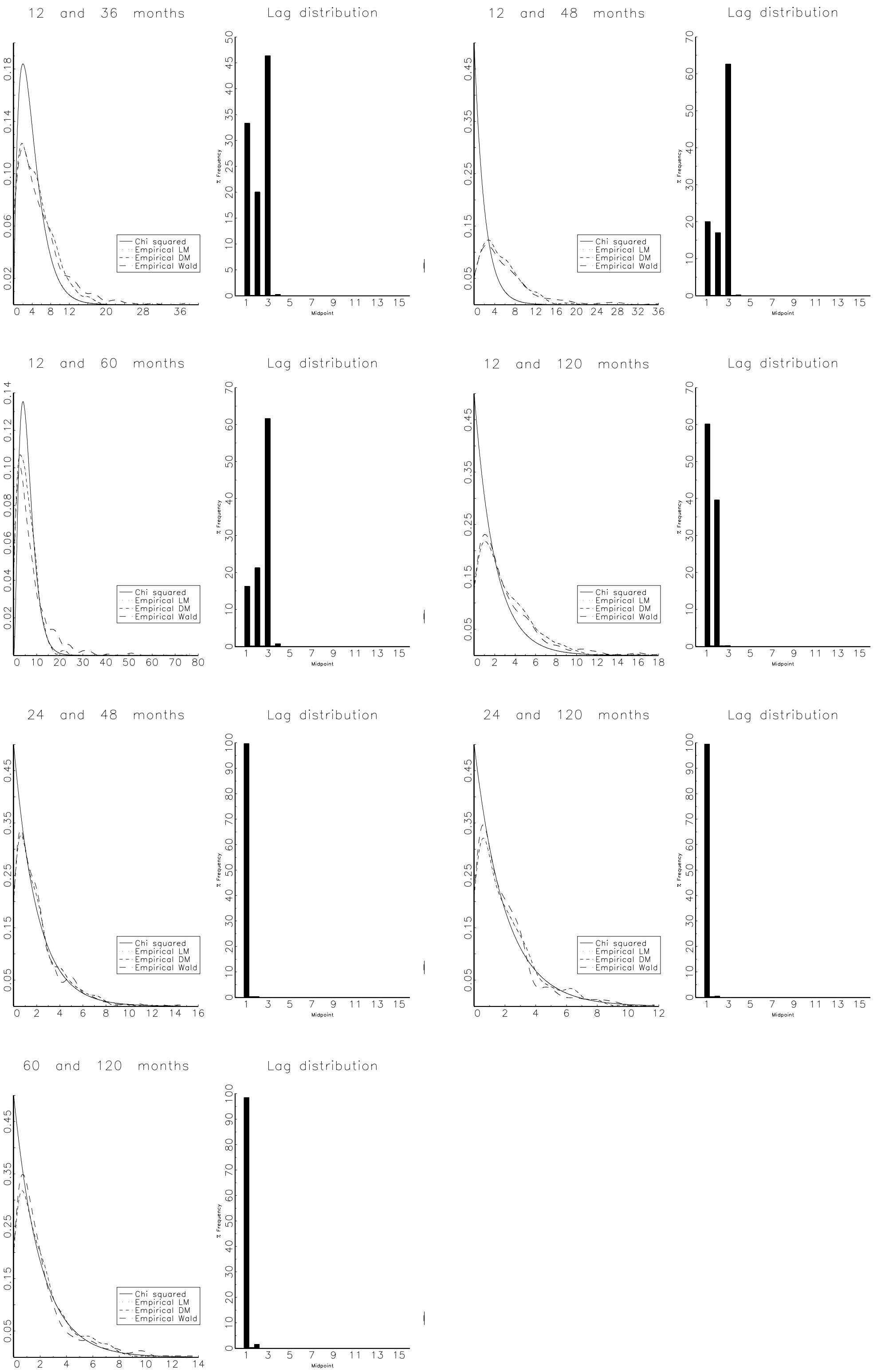
Jan 1982- Dec 2003



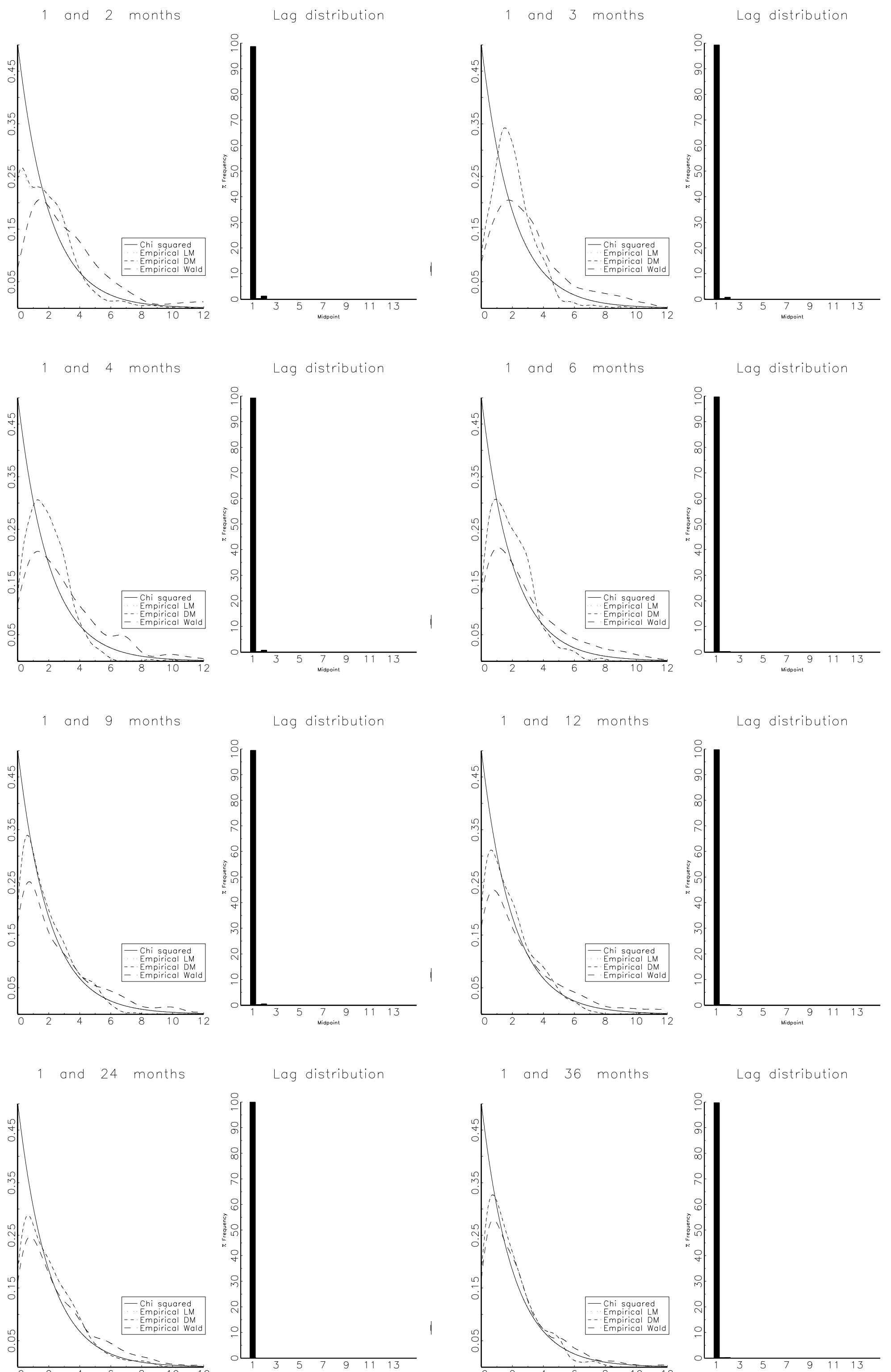

12 and 36 months
Lag distribution
12 and 48 months
Lag distribution
12 and 60 months
Lag distribution
12 and 120 months
Lag distribution
24 and 48 months
Lag distribution
24 and 120 months
Lag distribution
60 and 120 months
Lag distribution
Chi squared
Empirical LM
Empirical DM
Empirical Wald
% Frequency
Midpoint

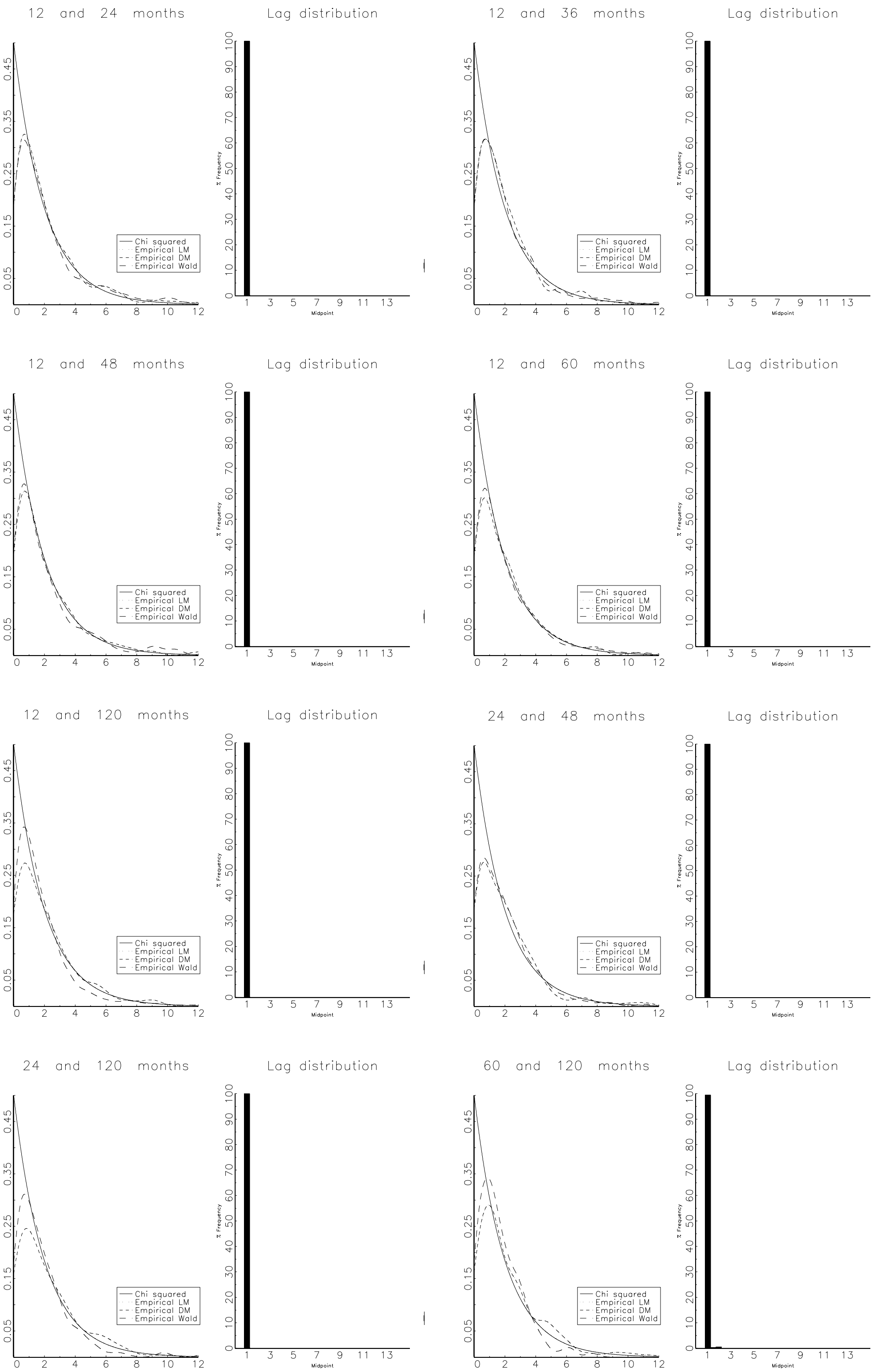
Selected empirical distributions of test statistics and lag length based on SIC Jan 1952- Dec 1978




12 and 36 months
Lag distribution
12 and 48 months
Lag distribution
12 and 60 months
Lag distribution
12 and 120 months
Lag distribution
24 and 48 months
Lag distribution
24 and 120 months
Lag distribution
60 and 120 months
Lag distribution
Chi squared
Empirical LM
Empirical DM
Empirical Wald
% Frequency
Midpoint

Jan 1982- Dec 2003




12 and 24 months
Lag distribution
12 and 36 months
Lag distribution
12 and 48 months
Lag distribution
12 and 60 months
Lag distribution
12 and 120 months
Lag distribution
24 and 48 months
Lag distribution
24 and 120 months
Lag distribution
60 and 120 months
Lag distribution
Chi squared
Empirical LM
Empirical DM
Empirical Wald
% Frequency
Midpoint

References:

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

Bataa, E. 2006. "On the Assessment of the Autocorrelation Tests", Working Paper, University of Manchester,

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

Bekaert, G. & Hodrick, R.J. 2001. "Expectations Hypotheses Tests", *Journal of Finance* 56(4), 1357-1394,

Bekaert, G., Wei, M. & Xing, Y. 2006. "Uncovered Interest Rate Parity and the Term Structure", *International Journal of Money and Finance,* forthcoming,

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

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

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

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

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

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

Campbell, J.Y. 1995. "Some Lessons from the Yield Curve", *Journal of Economic Perspectives* 9(3), 129-152,

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

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

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

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

Driffil, J., Psaradakis, Z. & Sola, M. 1997. "A Reconciliation of Some Paradoxical Empirical Results on the Expectations Model of the Term Structure", *Oxford Bulletin of Economics and Statistics* 59(1), 29- 42,

Edgerton, D., and Shukur, G., 1999. "Testing Autocorrelation in a System Perspective", *Econometric Reviews* 18(4), 343- 386,

Goncalves, S. & Kilian, L. 2004. "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form", *Journal of Econometrics* 123, 89- 120, Godfrey, L.G. & Orme, C.D. 2004. "Controlling the Finite Sample Singificance levels of Heteroskedasticity-robust Tests of Several Linear Restrictions on Regression Coefficients", *Economics Letters* 82, 281-287,

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

Gregory, A.W. & Veall, M.R. 1985. "Formulating Wald Tests of Nonlinear Restrictions", *Econometrica* 53(6), 1465- 1468,

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

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

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

Horowitz, J.L. 1997. "Bootstrap Methods in Econometrics: Theory and Numerical Performance", in Kreps, D.M. & Wallis, K.F. eds. *Advances in Economics and Econometrics: Theory and Applications,* Seventh World Congress, Volume III, 188- 222,

____________. 2001. "The Bootstrap", in Heckman, J.S. & Leamer, E. eds. *Handbook of Econometrics* 5, North-Holland, Amsterdam, The Netherlands, 3159- 3228,

Ivanov, V. & Killian, L. 2005. "A Practitioner's Guide to Lag Order Selection for VAR Impulse Response Analysis", *Studies in Nonlinear Dynamics & Econometrics* 9(1), 1-34,

Kilian, L. 1998a. "Accounting for Lag Order Uncertainty in Autoregressions: The Endogenous Lag Order Bootstrap Algorithm", *Journal of Time Series Analysis* 19(5), 532- 548,

________ 1998b, "Small-Sample Confidence Intervals for Impulse Response Functions", *Review of Economics and Statistics* 80, 218- 230,

Laitinen, K. 1978. "Why is Demand Homogeneity So Often Rejected?", *Economics Letters* 1, 231-233,

Li, H. & Maddala, G.S. 1996. "Bootstrapping Time Series Models", *Econometric Reviews* 15, with discussion, 115- 158,

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

Mankiw, N.G. & Miron, J.A. 1986. "The Changing Behavior of the Term Structure of Interest Rates", *Quarterly Journal of Economics* 101(2), 211-228,

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

Melino, A. 2001. "Estimation of a Rational Expectations Model of the Term Structure", *Journal of Empirical Finance* 8, 639-688,

McCulloch, J.H. 1990. "U.S. Government Term Structure Data, 1947 - 1987." in Freedman, B.M. & Hahn, F. eds. *Handbook of Monetary Economics*, North Holland, 672-715,

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

Newey, W. K. & West, K. D. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708, Ng, S., & Perron, P., 2001. “Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power”, *Econometrica* 69(6), 1519- 1554,

________________, 2005. “A Note on the Selection of Time Series Models”, *Oxford Bulletin of Economics and Statistics* 67(1), 115- 134,

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

Roberds, W., Whiteman, C.H. 1999. “Endogenous Term Premia and Anomalies in the Term Structure of Interest Rates: Explaining the Predictability Smile”, *Journal of Monetary Economics* 44(3), 555-580,

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

Sarno, L., Thornton, D.L. & Valente, G. 2006. “The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields”, *Journal of Financial and Quantitative Analysis,* forthcoming,

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

Shea, G. 1992. “Benchmarking the Expectations Hypothesis of the Interest Rate Term Structure: An Analysis of Cointegrating Vectors”, *Journal of Business & Economic Statistics*, 10(3), 347- 66,

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

Startz, R. 1982. “Do Forecast Errors or Term Premia Really Make the Difference Between Long and Short Rates?”, *Journal of Financial Economics* 10, 323- 329,

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

Thornton, D.L. 2004. “Testing the Expectations Hypothesis: Some New Evidence for Japan”, Bank of Japan *Monetary and Economic Studies* 22(2), 45-69, also published in Federal Reserve Bank of St. Louis *Review* 86(5), 21-39,

___________. 2006. “Tests of the Expectations Hypothesis: Resolving Campbell-Shiller Paradox”, *Journal of Money, Credit and Banking*, forthcoming,

Tjostheim, D., and Paulsen, J., 1983. “Bias of Some Commonly-Used Time Series Estimators”, *Biometrika* 70 (2), 389- 399,

Tzavalis, E. & Wickens, M. 1998. “A Re-examination of the Rational Expectations Hypothesis of the Term Structure: Reconciling the Evidence from Long-run and Short-run tests”, *International Journal of Finance and Economics* 3, 229 – 239,

White, H., 1980. “A Heteroskedasticity-Consistent Covarianace Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica* 48(4), 817- 838,

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

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