Missing Data and Interpolation in Dynamic Term Structure Models

V. Pavlov

Queensland University of Technology (v.pavlov@qut.edu.au)

Abstract: Term structure is often reconstructed from cross-sectional data where observations corresponding to a considerable proportion of maturities are missing. Usually nonparametric methods, such as various interpolation techniques, are employed to fill in these missing observations prior to the estimation of the model of interest. In this paper, we examine the impact of spline interpolation on the properties of expectations theory tests.

Keywords: interest rates; term structure; expectations hypothesis; spline interpolation; vector auto regression.

1. INTRODUCTION

Factors that drive the evolution of the term structure of interest rates have attracted considerable attention in the empirical economic and finance literatures. Term structure is a crucial input into pricing of interest sensitive derivatives. The relationship between short and long interest rates is important from the perspective of understanding the mechanism of transmission of monetary policy. Last, term structure is believed to reflect the expectations of market participants about future evolution of interest rates and therefore could provide a wealth of information.

Default-free term structure datasets often contain a large number of missing observations, which arise due either to thin trading or simple non existence of bonds with certain maturities. The problem is particularly severe in high-frequency (daily, weekly) datasets.

Most empirical studies rely on various cross-sectional interpolation techniques to reconstruct the missing observations. Cubic spline interpolation is still arguably the most popular method and was used by McCulloch and Kwon in the construction of the popular dataset of monthly US interest rates [McCulloch and Kwon, 1993].

The ability of the interpolation procedure to extract information from the observed cross-sections has been studied by a number of authors. Bliss [1996] reports that many such methods appear to leave omitted factors in the residuals. To the best of our knowledge the effect of interpolation on the properties of estimators in popular term structure models have not been attempted. The objective of this paper is to document some of these effects.

The structure of our simulation study is fairly straightforward. We start by estimating an interest rate model which is then used to simulate observations consistent the expectations hypothesis. Observations are then randomly selected and dropped from each cross-section. Missing observations are reconstructed using exact cubic spline interpolation and the interpolated data are used to compute EH statistics.

2. EXPECTATIONS HYPOTHESIS

2.1 Theory

The expectations hypothesis has a long history in economics. It is based on an intuitively plausible idea that the slope of the current term structure contains information about the future evolution of interest rates. In particular, if long term rates diverge too far from the expectations of future short rates the expectations theory prediction is that investors will trade until rates are brought back in line with these expectations.

Unfortunately, it is impossible to find a unique consistent formalisation of this simple idea. Cox, et al. [1981] noted that different formulations of the expectations hypothesis lead to incompatible theories.

Nevertheless, a large empirical literature developed testing the implications of the expectations hypothesis. Earlier papers, most notably Campbell and Shiller [1991], generally rejected the theory; some recent research however finds some support for it at very short maturities [Longstaff, 2000] and in the international data. A recent survey of the theory; with tests and applications can be found in Bekaert and Hodrick [2000].

Most often ET is interpreted as a hypothesis that forward rates of interest are unbiased predictors of the future short rates and somewhat arbitrarily postulated in terms of yields:

$$f(t,n) = E_t[r(t+n)],\ \forall n.$$

Here $f(t,n)$ - denotes the n-period forward rate of interest, $E_t[X] = E_t[X \mid \sigma_t]$ and $r(t)$ is the one period (short) yield. As an almost immediate implication we can obtain that the yields to longer maturity bonds are equal to the expected average short yields over the life of the bond:

$$r(t,n) = \mu_n + E_t\left(\frac{1}{n}\sum_{i=0}^{n-1} r(t+i)\right).$$

In this expression $r(t,n) = -[LnB(t,n)]/n$ is the yield on an n-period bond and μ_n is the time-invariant risk or liquidity premium on holding the n-period bond. This premium is assumed to be zero in the Pure ET.

In this paper we concentrate on a number of popular ET tests.

The first two are the Fama-Bliss regression tests:

Test 1.

$$r(t+1,n-1) - r(t,n) = \alpha + \beta\frac{r(t,n)-r(t)}{n-1} + \varepsilon_1(t+1)$$

Test 2.

$$\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\Delta r(t+i-1) =$$
$$= \gamma + \delta(r(t,n) - r(t)) + \varepsilon_2(t+n-1)$$

The expectations theory restricts the slope coefficients α, δ in the above regressions to unity.

Consider a VAR in levels for the yields:

$$y_t = C + \sum_{i=1}^{m} A_i y_{t-i} + \eta_t,$$

here $y_t' = [r(t), r(t,2)..r(t,n)]$ is the vector of yields. Using the estimated VAR we can construct "theoretical spreads" as averages of VAR forecasts:

$$S'(t,n) = \sum_{i=0}^{n-1}\frac{1}{n} r^f(t+i) - r(t),$$

where $r_t^f(T)$ is the time t VAR forecast of one period yield at T.

Under the expectations theory the actual spread is the expectation of the theoretical spread

$$S(t,n) = E_t(S'(t,n)).$$

In addition to the regression tests Campbell and Shiller [1991] suggested that correlations between "theoretical spreads" (Test 3) and actual spreads as well as the ratios of corresponding variances may provide more meaningful tests of economic significance of deviations from the expectations hypothesis.

3. SIMULATIONS

3.1 Basic Model

Interest rate model used in this study was estimated for maturities of 1 to 12 months on the sub-sample from January 1965 to February 1991 of the McCulloch and Kwon dataset.

To obtain simulations consistent with the expectations hypothesis we proceeded in two steps.

Firstly, we reduced the dimensionality of the cross-sections by extracting dynamic factors from the yields. The first factor is assumed to coincide with the shortest maturity (monthly) yield; the remaining two factors are obtained as the principal components of the sample variance-covariance matrix of the spreads. In particular, the two additional factors are the linear combinations of the yields $f_i(t) = w_i' r(t,n)$ with the maximum sample variance normalized to have the norm of unity:

$$w'\tilde{V}w \to \max_w$$
$$s.t.\ w'w = 1$$

Factor loadings are determined by the eigenvectors corresponding to the two largest eigenvalues of the sample variance-covariance matrix of the spreads $\hat{V}$.

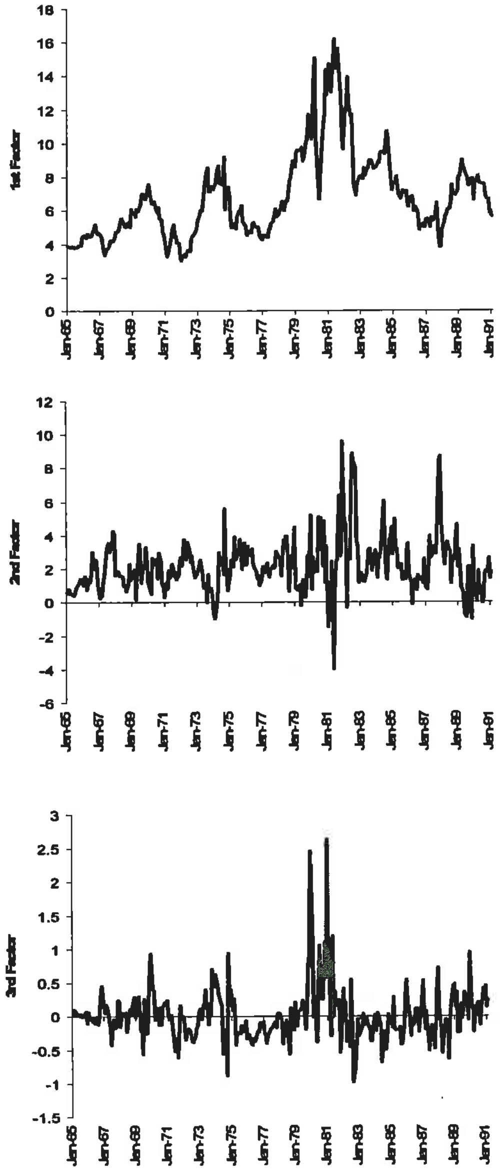

Figure 1. Sample factor realizations.

In the cross-sectional regressions factors capture about 98-99% of the sample variation of the spreads with the only exception of the spread between 2-month and 1-month yields where the amount is somewhat more modest 90%. The rest of the analysis is performed conditional on the estimated factor realizations. Figure 1 illustrates the dynamics of the estimated factors.

The factors are modelled as a 2nd order VAR with GARCH(1,1) residuals. While factors are uncorrelated by construction; we estimate the model under the much stronger assumption that they are uncorrelated conditionally as well as unconditionally. The main benefit of this parameterisation is that GARCH parameters can then be estimated equation by equation. With conditionally uncorrelated factors the model becomes:

$$y_t = C + \sum_{i=1}^{2} A_i y_{t-i} + \eta_t$$

$$\eta_{ti} = \sigma_{ti} \varepsilon_{ti}$$

$$\sigma_{ti}^2 = \omega_i + \lambda_i \varepsilon_{t-1i}^2 + \theta_i \sigma_{t-1i}^2$$

The parameters of the mean equation are estimated by OLS, the residuals are then used to estimate GARCH parameters.[1] The estimates obtained with this two stage procedure are consistent but inefficient [see e.g. Gourieroux, 1997]). To conserve space we do not report parameter estimates.[2]

Table 1 presents some selected diagnostics of the scaled residuals.

Table 1. Diagnostics of the Standardized Shocks (assymptotic p-values in paranthesis).

	$\frac{\eta_1}{\sigma_1}$	$\frac{\eta_2}{\sigma_2}$	$\frac{\eta_3}{\sigma_3}$
Skewness	-0.313	0.333	0.3893
Kurtosis	5.8209	4.6782	3.9968
JB	105	41.019	20.131
	(0.000)	(0.000)	(0.000)
Serial Correlation	0.276	0.441	0.218
(5 lags)	(0.926)	(0.819)	(0.958)
ARCH LM test	0.129	1.317	0.828
(5 lags)	(0.985)	(0.257)	(0.531)

Our sample period includes the Volcker period from 1979 to 1982 when the FED switched from targeting interest rates to targeting money growth which lead to an episode of extremely volatile nominal rates. This period is usually accommodated by either treating 79-82 as a structural break in the series or by incorporating a

[1] All calculations reported in this paper were performed in Matlab 6. GARCH toolbox was used to obtain estimates of the GARCH parameters and the Econometrics toolbox by J. LaSage was used to compute most VAR statistics.

[2] All unreported calculations are available from the author by request.

	Bias $(\sum_{i=1}^{5000} \frac{\xi}{n} - 1)\%$			**Standard Deviations**		
	$p_m = 0$	$p_m = 10\%$	$p_m = 30\%$	$p_m = 0$	$p_m = 10\%$	$p_m = 30\%$
	Test 1					
Long rate						
2	0.34%	-6.23%	-8.84%	0.13	0.13	0.14
4	1.02%	-2.50%	-12.45%	0.22	0.21	0.21
8	5.67%	3.29%	-5.99%	0.41	0.40	0.39
12	11.40%	9.47%	2.12%	0.52	0.51	0.50
	Test 2					
2	0.39%	-7.42%	-21.72%	0.06	0.07	0.08
4	0.78%	0.33%	0.49%	0.08	0.08	0.08
8	3.55%	3.10%	1.67%	0.12	0.11	0.11
12	6.65%	6.51%	5.73%	0.14	0.14	0.13
	Test 2 (Second order VAR)					
2	-0.78%	-0.92%	-1.31%	0.008	0.009	0.011
4	-2.81%	-3.41%	-3.95%	0.020	0.022	0.023
8	-11.17%	-11.69%	-11.32%	0.050	0.050	0.049
12	-16.03%	-16.50%	-15.74%	0.063	0.062	0.061

Table 2. Monte Carlo distributions of ET statistics.

level effect into the volatility equation. This can be done by augmenting the GARCH equation with a scaling factor equal to the square root of the lagged short yield.

Simulating with the levels related volatility created considerable problems. In particular, in our discrete-time model it is impossible to restrict yields from making trips into the negative territory without additional transformations. With the levels adjustment, volatility is not defined for negative yields. We tried simple specifications for the volatility effect consistent with the possibility of negative yields, such as scaling with $|r(t)|^{0.5}$, but the resulting simulations appear highly implausible.

On the other hand, fitting a simple GARCH process on the selected sub period produced plausible specification (Table 1) and simulations without any need for additional tweaking and was judged sufficient for the purpose of this paper.

3.2 Missing Data and Interpolation

The last step involves bootstrapping standardized VAR shocks, introducing missing data and re estimating ET regressions.

Bootstrap samples were constructed by resampling with replacement from the rows of the standardized GARCH residuals. The simulations were conditional on the sample estimates of VAR-GARCH parameters, so that the estimated VAR-GARCH equation was then used to construct simulated short yields and factors. In order to obtain simulations consistent with the null of the expectations theory we use VAR forecasts to simulate ET consistent spreads as described in section 2.1. The same method was used by Bekaert et al. [1997] to bootstrap small sample properties of the ET statistics.

For each bootstrapped sample we randomly select and drop observations from each cross section by generating a binomial random variable $I(t,k)$, which takes values 1 and 0 with constant probabilities p_m and $1-p_m$ respectively. We report simulations for p_m set to 10% and 30%.

The process generating missing observations is oversimplified. In reality the missing data problem is a lot more severe for longer maturities. Also real missing data exhibit deterministic trends as missing data move along the yield curve towards shorter maturities.

Observations corresponding to $I(t,k)=1$ are replaced with their interpolated values. The only difference with McCulloch and Kwon is that the cubic spline is restricted to fit the observed maturities exactly.

4. SIMULATION RESULTS

Simulation results are reported in the table 2. The evidence in the table is based on 5000 simulations of each statistics. The table details the average sample bias of the estimates relative to the ET value of unity for all the statistics.

All small sample distributions display nontrivial biases. In terms of the magnitude and direction of these biases our results are roughly consistent with the simulations detailed in Bekaert et. al. [1997] with the exception of the correlation test (test 3). In our simulations, contrary to what is reported in the above paper, the VAR correlation test seems to fair very poorly. It is however doubtful that either of these simulations based on miss-specified bi-variate VARs enable us to make reliable statements about the magnitude of small sample distortions in VAR tests.

Regression test based on the changes in long yields (test 1) are most sensitive to the presence of interpolated data. The contribution of spline interpolation to the bias disappears in tests #2 and #3 but remains almost constant for the regression test #1.

Somewhat counter intuitively there appears to be very little effect on the standard deviation of slope estimates. In fact standard deviations appear to be smaller for the tests using interpolated simulations, at least for longer maturities. This suggests that interpolating observation possibly over-smooths the data, producing optimistic confidence intervals for regression statistics. Quantitatively however the effect is very small.
Figure 2 shows the effect of interpolation on the distribution of the statistic of the test 1 regression. The distribution was estimated using the Gaussian kernel with the usual bandwidth selection criterion based on the IQR of the empirical distribution. As suggested by the standard deviations there are no noticeable shape distortions.

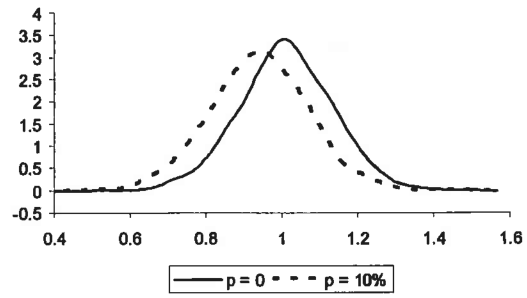

Figure 2. The distribution of the regression coeeficient (Test 1) for S(2,t).

5. CONCLUSION

This study examined the sensitivity of a number of expectations theory tests to cubic interpolation of the missing observations.

The results of our simulations indicate that, even in our highly stylised structure, a small fraction of missing data can introduce significant biases into simple regression and VAR tests. The biases generally shift the distribution of the regression tests to the left, which is consistent with the deviations of the usual empirical estimates from ET values.

Unlike small sample biases documented in the literature, these biases are unlikely to disappear as the sample gets larger. It seems reasonable to hypothesise that these biases can be quite substantial in studies employing interpolation with high frequency observations.

6. REFERENCES

Bekaert, G., R.J. Hodrick, and D.A. Marshall, On the biases of the expectations hypothesis of the term structure of interest rates, *Journal of Financial Economics*, 44, 309-348, 1997.

Bekaert, G., and R.J. Hodrick, Expectations theory tests, NBER working paper, #7609, 2000.

Bliss, R.R., Testing term structure estimation methods, *Federal Reserve Bank of Atlanta working paper*, 96-12a, 1996.

Campbell, J.Y. and R.J. Shiller, Cointegration and tests of present value models, *Journal of Political Economy*, 95, 1062-1088, 1987.

Campbell, J.Y. and R.J. Shiller, Yield spreads and interest rate movements: a bird's eye view, *Review of Economic Studies*, 58, 495-514, 1991.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, A re-examination of traditional hypotheses about the term structure of interest rates", *Journal of Finance*, 4, 769-799, 1981.

Gourieroux, C., ARCH models and financial applications, Springer-Verlag, 1997.

Hall, A.D., H.M. Anderson, and C.W. Granger, A cointegration analysis of Treasury bill yields, *Review of Economics and Statistics*, 74, 116-126, 1992.

Longstaff, F.A., The term structure of very short-term rates: new evidence for the expectations hypothesis, *Journal of Financial Economics*, 58, 397-415, 2000.

McCulloch, H.J., and H.C. Kwon, US term structure data, 1947-1991, *Ohio State Working Paper*, 93-6, 1993.