SANTOS, Carlos
OLIVEIRA, Maria Alberta

Abstract. In this paper, we build a model for the yield curve in Germany, from 1975 to 2001, and use it to test the Lucas Critique. We provide a first application of the new general-to-specific automatic model selection algorithm embodied in PcGets to term structure modelling, and use new super exogeneity tests. Super exogeneity is rejected, so the model is vulnerable to the Lucas Critique. Inflation was not retained in the model selected for the spread, suggesting that inflation expectations are not relevant (for the short maturities considered) to forward interest rate movements.
JEL: E43; C12; C22; C51
Keywords: Super exogeneity; Lucas Critique; Term Structure; General-to-Specific

1. Introduction

The motivation for this paper is two-fold: we wish to construct a general-to-specific (GETS) model for the German yield, taking advantage of the noticeable properties of PcGets; and we wish to apply the recently developed automatic super exogeneity tests (see Hendry and Santos, 2006a) to such a model.

Our choice to investigate yield curve models derived from the operational use that practitioners have been giving to the notion of super exogeneity (Engle, Hendry and Richard, 1983), and to super exogeneity testing (Engle and Hendry, 1993): trying to assess the empirical relevance of the Lucas (1976) critique. Lindé (2001) argues that it is not feasible for an econometric model to condition on all possible shocks hitting the economy – hence, parameter invariance, as required by super exogeneity, might, in practice, be simply the outcome of some shocks offsetting others (e.g. monetary policy shocks offsetting supply or technological shocks). Notwithstanding, as argued by Ericsson and Irons (1994), super exogeneity has become the standard empirical tool to assess the Lucas (1976) critique (see also Stanley (2000) and Favero (2001)). Therefore, our choice of a term structure model to build an example of how the new super

* Carlos Santos, Faculdade de Economia e Gestão, Universidade Católica Portuguesa, e-mail: csantos@porto.ucp.pt and Maria Alberta Oliveira (ISMAI).

Acknowledgement: This research was conducted when the first author was a doctoral student at the University of Oxford and a Graduate Research Visitor at the ECB. Comments by David Hendry, Robert Engle, Andreas Bayer, Ignazio Angeloni, Adrian Pagan and Kristoffer Nimark were very useful. Financial support from the Fundação para a Ciência e a Tecnologia, Lisboa, is gratefully acknowledged.


2 In this study we will follow a well established tradition in the literature of using the terms "yield curve" and "term structure" interchangeably. This is not, however, completely accurate: the term structure is a particular yield curve. For a discussion of this see, inter alia, Shiller (1990) and Svensson (1994).
exogeneity tests can be applied bears in mind the main empirical use that is given to this type of tests. We take into account Blanchard’s (1984) remark that the Lucas (1976) critique could only be assessed in the context of models that would deal explicitly with expectations, of which the term structure of interest rates model is a good example (another one being the expectations augmented Philips curve – see, inter alia, Alogoskoufis and Smith (1991)). Nonetheless, as stressed by Lindé (2001), most of the empirical work that has been conducted on the Lucas (1976) critique has used money demand functions and consumption functions. Among the previous studies that have used the term structure model are Blanchard (1984), Favero (1989) and Psaradakis and Sola (1996).

Having defined the aim of this paper, our objective does nonetheless also entail a direct contradiction with some of the results in Psaradakis and Sola (1996). We shall build the marginal and the conditional models using the PcGets algorithm, whilst Psaradakis and Sola (1996) followed a GETS strategy that did not entail congruency checking at each stage nor multiple path searches, being in fact closer to stepwise regression (see Miller, 1990). This paper is organized as follows: the next section will discuss aspects of term structure modelling, and the implications of such models for policy making; section 3 will briefly describe the new super exogeneity tests; section 4 describes the data to be used in our study; section 5 discusses our estimation results; section 6 implements the three new super exogeneity tests; section 7 concludes.

2. The term structure of interest rates

In this paper we develop a modelling exercise, not a forecasting one. In this sense, we are not interested in investigating a leading indicators approach to forecasting, based on the pseudo-information content of the yield curve with respect to expectations of future economic activity. Evidence on that domain has been mixed: Berk and Van Bergeijk (2000) have concluded against it, whilst others, like Hardouvelis (1994) and Bernard and Gerlach (1996), have found richer information contents on the yield, with respect to future movements of inflation and output.

This is also not an exercise in testing the expectations hypothesis of the term structure of interest rates. Nonetheless, we are influenced by the expectations theory, since it inspired Psaradakis and Sola’s (1996) conditional model for the yield, which we shall use as a baseline for our work. However, empirical studies tend to exhibit several important departures from the expectations theory that are worth noticing. The most significant of these is the inclusion of a time-varying risk premium.

Evidence on the adequacy of the expectations hypothesis as a theory for the term structure is mixed. Shiller, Campbell and Schoenholtz (1983) provide references to studies that refute the expectations hypothesis. On the other hand, Gerlach and Smets (1997) find evidence in favour of it. Cox, Ingersoll and Ross (1985), Shiller (1990) and Campbell and Shiller (1991) provide systematic discussions of the theory of the term structure of interest rates. In particular, with respect to the expectations theory, a common nutshell representation is given by equation (1):

\[ R(n, t) = \alpha + \beta_0 R(m, t) + \sum_{i=1}^{I} \beta_i R(m, t-i) + \sum_{i=1}^{I} \gamma_i \pi(t-i) + \delta_t V_t + \epsilon_t \]  

(1)
with $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$. \hfill (2)

$R(n,t)$ denotes the long term interest rate at time $t$. $R(m,t)$ is the short term interest rate and $\pi$ is inflation. The risk premium, $V_t$, is proxied by a moving-average of short-term interest rates. The error term is modelled as a stationary AR(1) process, with parameter $\rho$, whilst $u_t$ is white noise.

Underlying equation (1) is the theoretical prior that the long-term interest rate is equal to the expected future spot interest rate plus a risk premium. The expected future spot rate is a linear function of current and lagged values of inflation and short term interest rates. Blanchard (1984) presents this model, with $l = 19$, as the 1979 version of the MPS model, which was first specified and estimated by Modigliani and Schiller (1973). Berk and Van Bergeijk (2001) claim that the model is well known in the profession, since it has often been used as a workhorse for policy debates and has been applied to numerous countries. On the basis of this, and given the model simplicity, the authors use it themselves, in their research on the economic impacts of the creation of the EMU. As argued by Walsh (2003), the inclusion of a risk premium in the model already implies a theoretical concession to empirical modelling. The pure expectations model of the term structure should hold exactly: the term premium implies a stochastic deviation from the exact form of the expectations hypothesis. Hence, Walsh (2003), following closely McCallum (1994), suggests

$$I_t = \alpha i_t + \beta E_t(i_{t+1}) + \zeta_t$$ \hfill (3)

for a two-period model, where $I_t$ is the two-period interest rate, $i_t$ is the one-period interest rate and $\zeta_t$ is the time-varying term premium, where

$$\zeta_t = \rho \zeta_{t-1} + \eta_t$$ \hfill (4)

in which $\eta_t$ is a white noise process and $|\rho| < 1$.

Shiller, Campbell and Schoenholtz (1983) argue that the inclusion of such a risk premium in the expectations hypothesis (even if a slowly time-varying one, or even a constant one) should have shifted interpretations of interest rate phenomena to changes in risk, instead of the still prevailing interpretations based on future interest rate movements. Berk and Van Bergeijk (2000) go so far as to say that a time-varying risk premium reduces the value of the yield curve as far as estimating future inflation is concerned. Psaradakis and Sola (1996) used a univariate conditional model of the yield where the spread was a function of lagged spread and interest rates alone. No risk premium term is included explicitly.

Cooray (2003) takes a different approach, modelling the bivariate system, and hence avoiding the pitfalls of invalid conditioning. De Wachter, Lyrio and Maes (2004) assess the impact of the creation of the EMU in Germany using a continuous time model with three latent factors (the central tendencies of inflation and the output gap and the instantaneous interest rate). Their model fits within the class of affine term structure models.

It is sometimes claimed that even in the absence of the Lucas (1976) critique, the term structure of interest rates is a dubious tool to assess policy effects on its own. Blanchard

\[ ^3 \text{MIT-Pennsylvania-Social Science Research Council.} \]
(1984) and Berk and Van Bergeijk (2000) argue that the mechanism is hardly simple. Key issues are, for instance, price flexibility, labour market expectations and financial markets expectations. In conclusion, the yield curve should be viewed with caution, both when assessing the impact of possible policy changes, and also when trying to interpret its realized movements.

3. New Super Exogeneity Tests

In this section, we will present a very brief overview of the new super exogeneity tests developed in Hendry and Santos (2006a). The interested reader is referred to their paper for details.

A key recent development in testing for parameter non-constancy is doing so by adding a complete set of impulse indicators to a marginal model: see Hendry, Johansen and Santos (2005). This new technique is known as impulse or indicator saturation. Using GETS procedures, the authors establish the null distribution of the estimator of the mean in a location-scale model, after adding T impulses when the sample size is T. A two-fold process is investigated, where half of the indicators are added and the significant ones recorded. Then, the other half is examined, and finally the two retained sets of indicators are combined. The average retention rate of indicators, under the null hypothesis that no indicator matters, is $\alpha T$, matching the binomial result, where $\alpha$ is the significance level: hence there is no overfitting. Moreover, Hendry et al. (2005) show that other splits, namely $T/3$, do not affect the retention rate under the null. Hendry and Santos (2006b) extend the analysis to dynamic models.

This procedure can be applied to marginal models of putative superexogenous conditioning variables. First the significant dummies in the marginal are recorded. Secondly, the retained set is added to the conditional in one of three ways:
- as m indicator variables, matching the m dummies retained from the marginal model after impulse saturation; their joint significance is tested via a joint F test or a likelihood ratio (LR) test;
- as an index (for the theory of indices or linear combinations of indicators see Hendry and Santos (2005), and Santos (2003)), where each indicator carries a weight equal to its estimated coefficient in the marginal model; testing super exogeneity is now testing the individual significance of the index in the conditional model;
- as two indices (the previous one and another where the weights are the previous ones multiplied by the values of the marginal variable at the dates for which dummies were retained); the super exogeneity test is now a test on the joint significance of the two indices in the conditional model.

In all three cases, rejection of the null is equivalent to rejecting the super exogeneity hypothesis. The tests are based in those suggested in Engle and Hendry (1993), but clearly overcome the criticism of ad-hoc selection of the dates for the dummy variables to be included in the conditional model (see Lindé, 2001), since the procedure tests a dummy at each possible date, and can now be fully automated without any user intervention.

Hendry and Santos (2006a) show that all new tests have the correct size, and good power properties against several types of failure of super exogeneity. Furthermore, it is
established that selecting the indicators in the marginal does not affect the critical values for testing in the conditional, a fundamental result for the procedure to be valid.

4. Data

We chose to model a yield curve where the forward interest rate is based on 6 months T-bills and the short-term interest rate is based on 3 months T-bills, as in Psaradakis and Sola (1996). The modelling strategy in this paper and in theirs is GETS. However, our methodologies differ in that GETS essentially meant sequential reduction by deletion of insignificant variables for them, whilst we benefit from the PcGets algorithm, allowing among other things the benefits of multiple path searches.

We use German monthly data, comprising the period between January 1975 and December 2001. Interest rates are annualized and expressed in decimals. The spread at time $t$ is thus defined as

$$\text{spread}_t = T_6 - T_3$$

where $T_6$ refers to the yield on bonds with a six months maturity and $T_3$ is the yield on bonds with a three months maturity.

5. The Marginal and the Conditional models

The first step is to check the order of integration of our candidate variables to be included in the General Unrestricted Model (GUM): we are interested in including current and lagged inflation, current and lagged interest rates and lagged spreads in the GUM for the conditional; current and lagged spreads and lagged interest rates in the GUM for the marginal. This leads us to test for unit roots in the series for inflation, interest rates and spreads. Table (1) reports the results using the Augmented Dickey-Fuller (ADF) test with a constant. $DT_3$ is the first difference of the interest rate. This variable was created since the null of a unit root in $T_3$ was not rejected for the several lag lengths. Inflation and the spread were, on the contrary, found to be $I(0)$. Table (1) also highlights that $DT_3$ is $I(0)$.

<table>
<thead>
<tr>
<th>Table 1: ADF tests with constant</th>
<th>Lag 1</th>
<th>Lag 2</th>
<th>Lag 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>inflation</td>
<td>-10.29**</td>
<td>-8.704**</td>
<td>-8.041**</td>
</tr>
<tr>
<td>spread</td>
<td>-4.452**</td>
<td>-3.359*</td>
<td>-2.937*</td>
</tr>
<tr>
<td>$T_3$</td>
<td>-1.368</td>
<td>-1.597</td>
<td>-1.786</td>
</tr>
<tr>
<td>$DT_3$</td>
<td>-10.69**</td>
<td>-8.342**</td>
<td>7.433**</td>
</tr>
</tbody>
</table>

Figure 1 plots the spread. The movements over time can be related to economic history. The sharp increase in spot interest rates from 1979 to 1983 is related to the inflationary effects of the second oil shock and to the new operating procedures implemented by the Fed in the US (see Baba, Hendry and Starr (1992) for a discussion of these).

---

4 The data used here was collected for Hordhal, Tristani and Vestin (2004).

5 In this paper, a double * following an observed statistic means rejection of the null at a 1% significance level, whilst a single * means rejection at 5%.
German reunification (with financial reunification taking place in June 1990 and political reunification taking place in October) is largely responsible for the upswing in short term interest rates in the early 1990s. Two arguments are usually put forth to justify this: on the one hand, financial reunification meant the enlargement of the monetary base; on the other, it also meant inflationary pressures due to expanded business opportunities for West German firms in the East (although this is said to have been a slower process). Exchange rates play an interesting role here. Whilst the high expected interest rates of the early 1970s seem to be due to the collapse of the Bretton Woods system, the first oil shock and the instability in the snake system, the creation of the EMU does not seem to have caused much disturbance in German interest rates. This is, to a large extent, unsurprising: the process that led to the creation of the EMU has observed a convergence of the other EMU candidate countries’ interest rates to the German one – it was not the German interest rate that converged.

5.1 Selecting the Conditional Model. With respect to the conditional model we have postulated the following GUM:

$$s_t = \alpha + \sum_{i=1}^{6} \beta_i s_{t-i} + \sum_{i=0}^{6} \gamma_i DT3_{t-i} + \sum_{i=0}^{6} \delta_i \text{inflation}_{t-i} + \nu_t$$

where $s_t$ is the spread. The inclusion of inflation on the right hand side of the model follows a certain tradition in the literature on hypothesis testing in the context of the expectations theory of the term structure of interest rates (see, inter alia, Blanchard (1984)).

We have used PcGets to select a congruent final model, having made two choices: we have opted for a liberal strategy (that is, one that would minimize deletion of relevant variables); and we chose not to check for outliers. This last option is fundamental in the exercise we are performing. The aim of our study is to implement the three new super exogeneity tests suggested in Hendry and Santos (2006a). Hence, we chose not to have any indicators in the conditional to account for outliers, as some of these might be redundant with those that would be added after indicator saturation of the marginal. PcGets selected the following representation for the final model:
\[
\hat{s}_t = 0.6102 s_{t-1} + 0.1694 s_{t-3} + 0.1279 s_{t-4} + 0.042 DT3_{-1} \quad (6)
\]
where the estimated residual standard error is \( \hat{\sigma} = 0.084 \). All variables are significant at 2.5%. Table (2) reports the results for mis-specification tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>observed statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR 1-7</td>
<td>1.0877</td>
<td>0.371</td>
</tr>
<tr>
<td>ARCH 1-7</td>
<td>0.37548</td>
<td>0.916</td>
</tr>
<tr>
<td>Normality</td>
<td>394.33**</td>
<td>0.000</td>
</tr>
<tr>
<td>hetero</td>
<td>0.28087</td>
<td>0.972</td>
</tr>
<tr>
<td>hetero-X</td>
<td>0.29726</td>
<td>0.994</td>
</tr>
<tr>
<td>RESET</td>
<td>2.9747</td>
<td>0.09</td>
</tr>
</tbody>
</table>

With the exception of normality, all mis-specification tests indicate the model describes the data well. The rejection of the null hypothesis of normality was to be expected since no outlier correction was defined from the outset. Therefore, that was a cost we were willing to bear. Figure 2 justifies the claim that the poor behaviour of the normality statistic is due to neglected outliers. The QQ plot suggests the existence of excess kurtosis, as does the residual histogram. The fat tails of the residual density are a signal of neglected outliers. Finally, in figure 2, the comparison of the actual spread with the fitted values is indicative of a very good quality of fit.

![Graphical analysis: residuals from conditional model](image)
Figure 3 plots the residual autocorrelation and partial autocorrelation functions (ACF and PACF), allowing one to conclude there are no neglected serial correlation issues nor nonstationarities. Indeed, based on the ACF and the PACF, the residuals seem to behave as a white noise process.

In conclusion, we will proceed with (5) as the estimated conditional model. The failure of normality due to some neglected outliers is not a major concern, and setting this aside the model is congruent.

5.2 Selecting the Marginal Model. The marginal model’s dependent variable is $DT_3$ for the reasons discussed above. We specify the following GUM:

$$DT_3 = \alpha + \sum_{i=0}^{6} \beta_i s_{t-i} + \sum_{i=1}^{6} \gamma_i DT_3_{t-i} + \omega_t$$

We follow Psaradakis and Sola (1996) in this option. Again, we refrained from specifying an outlier correction algorithm, as we wished to avoid the introduction of indicators in the marginal prior to indicator saturation. PcGets final selected marginal model was

$$DT_3 = 0.1257 DT_3_{t-2}$$
with estimated residual standard error of $\hat{\sigma} = 0.35$. Results for mis-specification tests are reported in Table (3).

<table>
<thead>
<tr>
<th>Test</th>
<th>observed statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR 1-7</td>
<td>1.1939</td>
<td>0.306</td>
</tr>
<tr>
<td>ARCH 1-7</td>
<td>0.034189</td>
<td>1.000</td>
</tr>
<tr>
<td>Normality</td>
<td>148.65**</td>
<td>0.000</td>
</tr>
<tr>
<td>Hetero</td>
<td>1.1832</td>
<td>0.3077</td>
</tr>
<tr>
<td>hetero-X</td>
<td>1.1832</td>
<td>0.3077</td>
</tr>
<tr>
<td>RESET</td>
<td>0.0015276</td>
<td>0.9688</td>
</tr>
</tbody>
</table>

Once more, the only problem arises from the normality test. This was expected as the policy rule (marginal model) is often subject to shocks that might induce outliers. Such odd events are unaccounted for in our modelling strategy, as we have pointed out. In fact, it is not at all uncommon for policy models to be highly unstable without the inclusion of indicators (see Hendry, 1995, for examples and discussion).

6. Automatic Tests for Super Exogeneity

As discussed in Hendry and Santos (2006a), the first step for the implementation of any the three new super exogeneity tests is the indicator saturation of the marginal model. Hence, we proceed by considering a partition of $T/3$ and impulse saturate the model accordingly. The 27 relevant dummies from all 3 partitions entered the conditional model, in each of the three ways described in Section 3: as 27 individual indicators, as a linear combination of such indicators, or as two indices.

6.1 Joint Significance Test. The indicators retained as significant in the marginal model were added to the conditional, where a joint significance test on these 27 new variables was conducted. Table (4) reports the results of this testing method, both when the procedure used is an F-test and when a LR test is used. $c_{\alpha=0.01}$ is the 1% critical value. From Table (4), both procedures lead, even at this stringent significance level, to the rejection of the null hypothesis that none of the 27 indicators retained in the marginal is significant in the conditional. Hence, the null of super exogeneity is rejected, when the joint significance test is used.

<table>
<thead>
<tr>
<th>Test</th>
<th>observed statistic</th>
<th>$c_{\alpha=0.01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>joint F</td>
<td>$F_{obs} = 10.996**$</td>
<td>$F_{(27,320)} = 1.805$</td>
</tr>
</tbody>
</table>

6 The F-test was not conceived for dynamic models, since serial correlation changes the distribution of the test statistic. Nonetheless, for I(0) variables, many empirical studies continue to make use of the F-test in the conventional way. One possible reason for this is that the alternative testing procedure, a likelihood ratio test, is said to behave badly in finite samples when the number of restrictions under the null is large (27 in our case). Reade (2006) provides Monte Carlo evidence of the convergence to nominal size of empirical rejection frequencies, under the null, of both tests when applied in a dynamic setting.
6.2 Index Test. The index was built from a column vector of zeros, replacing the relevant entries by their estimated coefficients. The following augmented conditional model was then estimated:

$$\hat{s}_t = 0.52 s_{t-1} + 0.16 s_{t-3} + 0.14 s_{t-4} + 0.05 DT3_{t-1} - 0.0007 \text{Index}_t$$  

(9)

The t-ratio for the index is greater than the relevant quantile of the standard normal. Hence, the null hypothesis that the index coefficient is zero is strongly rejected. This result matches the outcome from the previous subsection: again we find evidence against super exogeneity.

6.3 Double Index Test. The second index was built introducing in the relevant entries of the column vector of zeroes the product of the weights in the first index by the observed values of the spread at those moments in time. The augmented conditional model is now:

$$\hat{s}_t = 0.52 s_{t-1} + 0.17 s_{t-3} + 0.1 s_{t-4} + 0.04 DT3_{t-1} - 0.00002 \text{Index}_t + 0.27 \text{Index2}_t$$  

(10)

The double index test for super exogeneity postulates as the null hypothesis that both indices’ coefficients are zero. Again, we could follow a LR or an F-test approach. Results for both cases are given in table (5).

<table>
<thead>
<tr>
<th>Test</th>
<th>observed statistic</th>
<th>$c_{\alpha=0.01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>joint F</td>
<td>$F_{obs} = 57.59**$</td>
<td>$F_{(2,318)} = 1.779$</td>
</tr>
<tr>
<td>LR</td>
<td>$LR_{obs} = 113.74**$</td>
<td>$\chi^2_{(27)} = 9.21034$</td>
</tr>
</tbody>
</table>

Evidence of both tests points in the same direction: rejection of the null of super exogeneity at the 1% significance level. Given that the single index variable is insignificant in (9), whilst the second index is significant, it is possible, following the discussion in Hendry and Santos (2006a), to trace rejection of the null to a failure of invariance, rather than to a failure of weak exogeneity.

7. Conclusion

In this article, three new super exogeneity tests are used in a model of the German yield curve. The empirical models for the marginal and the conditional were selected through PcGets. Setting aside some neglected outliers, both models are congruent. Two important outcomes of our study are worth noticing:
- inflation was dropped by the PcGets algorithm from the final model for the spread. This casts some doubts as to economic relevance of inflation in determining expectations about future interest rates, at such short periods (less than 1 year);
- the three new tests led to the same conclusion: rejection of the null of super exogeneity at any relevant significance level. One would therefore be tempted to claim that the conditional model for the German spread is not immune to Lucas (1976) critique, and hence should not be used for policy-making purposes.
References


