Monograph:
Gascoigne, J. (2004) Estimating threshold vector error-correction models with multiple cointegrating relationships. Working Paper. Department of Economics, University of Sheffield ISSN 1749-8368

Sheffield Economic Research Paper Series 2004013

Sheffield Economic Research Paper Series

SERP Number: 2004013

Jamie Gascoigne

Estimating threshold vector error-correction models with multiple cointegrating relationships.

November 2004

* Corresponding author

Department of Economics
University of Sheffield
9 Mappin Street
Sheffield
S1 4DT
United Kingdom
www.shef.ac.uk/economics

Abstract

Hansen and Seo (2002) outline procedures to test for threshold cointegration, and to estimate a bi-variate model. However, in their conclusion they note that future research will have to find a way of estimating larger systems with multiple cointegrating vectors. This paper proposes a new algorithm that can be used to estimate such models. Simulation experiments are used to compare the algorithm's performance with that of Hansen and Seo, and a practical application to the term structure of UK interest rates is also presented.

Keywords: Nonlinearity, Cointegration, Term Structure.

JEL numbers: C13, C32.

This paper has benefited from helpful comments made by David Chappell, Athina Kanioura, and Paul Turner. Any remaining errors or ambiguities are, of course, the author's responsibility.

1. Introduction

Threshold models are based on the principle that the data generation process for a time series is characterised by separate regimes, each with its own independent behaviour. The simplest class is the univariate Threshold Autoregression or TAR, developed by Tong (1983,1990). One of the key challenges is estimating the values of the thresholds that divide the separate regimes, since the sum of squares function is discontinuous and non-differentiable with respect to these parameters. Tong suggests performing a grid search using the observations of the time series as potential candidates for the threshold, selecting the values that minimise the residual sum of squares or maximise the log likelihood (exactly equivalent under the assumption of Gaussian errors). This is probably the most common estimation procedure used in the applied literature.

Balke and Fomby (1997) introduce threshold cointegration which allows non-stationary variables to be modelled in such a framework. The idea is intuitively appealing because costs of adjustment may prevent the restoration of equilibrium in a variety of economic circumstances. However, this will mean that often we will want the regimes to be defined by the error-correction term, so that the threshold effect is activated depending on the size of the disequilibrium in the system. This creates an additional problem if the cointegrating vector is unknown a priori because the error correction term will not be observable, and hence it becomes unclear how to form a grid search.

Hansen and Seo (2002), hereafter HS, suggest an algorithm that allows estimation of a cointegrating relationship and a single threshold when both are unknown in a bivariate vector error correction model (VECM). They show by simulation that the estimation procedure performs quite well, and apply it successfully to a model of the interest rate term structure in the US. However, this paper suggests there are ways in which their methodology can be improved. Firstly, it is shown that the number of computations necessary can be substantially reduced, without loss of efficiency. Secondly, their procedure is limited to the bi-variate case. A solution is therefore proposed which should be able to deal with these issues.

The rest of the paper is organised as follows. In Section 2 the HS algorithm is described and evaluated. It is argued that improvements can be made, and an alternative procedure is suggested in an attempt to achieve this. Section 3 presents Monte Carlo evidence in order to compare the empirical performance of the two methodologies. Section 4 then illustrates an application to the term structure of interest rates in the UK of the new algorithm – a three dimensional VECM with two cointegrating relationships.

2. The HS approach and a proposed alternative

A linear cointegrated model can be set out as follows:

$$\Delta x_t = A' X_{t-1} + u_t, \tag{1}$$

where

$$X_t' = \begin{bmatrix} 1 & x_{t-1} & \Delta x_{t-1} & \dots & \Delta x_{t-l} \end{bmatrix},$$

$$A' = \begin{bmatrix} a_0 & \alpha\beta' & a_1 & \dots & a_l \end{bmatrix}.$$

Here x_t is a p-dimensional $I(1)$ time series with n observations and l is the maximum lag length. In the parameter matrix A, α and β are p by r matrices, where r is the number of cointegrating relations. The error term, u_t, is assumed to be a vector martingale difference sequence with finite covariance matrix $\Sigma = \mathrm{E}(u_t u_t')$ Using the same notation, a two-regime threshold cointegrated model is written as:

$$\Delta x_t = A_1' X_{t-1} d_{1t}(\beta, \gamma) + A_2' X_{t-1} d_{2t}(\beta, \gamma) + u_t, \tag{2}$$

where

$$d_{1t}(\beta,\gamma) = 1(f(\beta' x_{t-1}) \le \gamma),$$
$$d_{2t}(\beta,\gamma) = 1(f(\beta' x_{t-1}) > \gamma),$$

with $1(\cdot)$ denoting the indicator function, and γ being the threshold parameter.

Assuming that the errors are iid Gaussian, the likelihood function is

$$\mathcal{L}_n(A_1, A_2, \Sigma, \beta, \gamma) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}u_t(A_1, A_2, \Sigma, \beta, \gamma)'\Sigma^{-1}u_t(A_1, A_2, \Sigma, \beta, \gamma). \quad (3)$$

If (β,γ) is held fixed, then model (2) becomes

$$\Delta x_t = A_1(\beta)' X_{t-1}(\beta) d_{1t}(\beta,\gamma) + A_2(\beta)' X_{t-1}(\beta) d_{2t}(\beta,\gamma) + u_t, \quad (4)$$

where

$$X_t(\beta)' = \begin{bmatrix} 1 & w_{t-1}(\beta) & \Delta x_{t-1} & \dots & \Delta x_{t-l} \end{bmatrix},$$

$$A_j(\beta)' = \begin{bmatrix} a_{j0} & \alpha_j & a_{j1} & \dots & a_{jl} \end{bmatrix},$$

and $w_{t-1}(\beta) = \beta' x_{t-1}$ is the error correction term. The concentrated likelihood function for (4) is:

$$\mathcal{L}_n(\beta,\gamma) = -\frac{n}{2}\log\left|\hat{\Sigma}(\beta,\gamma)\right| - \frac{np}{2}. \qquad (5)$$

For given values of (β,γ), all of the parameters in the $A_j(\beta)$ matrices can then be estimated by OLS regression. HS consider the case of model (4) when $p = 2$ and $r = 1$, so there is a single cointegrating vector, and specify that

$$d_{1t}(\beta,\gamma) = 1(w_{t-1}(\beta) \le \gamma),$$
$$d_{2t}(\beta,\gamma) = 1(w_{t-1}(\beta) > \gamma).$$

In order to estimate β and γ, HS suggest using the following algorithm:

1. Use the approach of Johansen (1988) to obtain an estimate $\tilde{\beta}$ from a linear VECM. Given that $\tilde{w}_{t-1} = w_{t-1}(\tilde{\beta})$, let $[\gamma_L, \gamma_U]$ denote the empirical support for $\tilde{w}_{t-1}$, and construct an evenly spaced grid, Γ, on $[\gamma_L, \gamma_U]$. Then construct an evenly spaced grid, B, on $[\beta_L, \beta_U]$ based on a wide confidence interval over $\tilde{\beta}$. The grid search should be constrained to ensure that a certain number of observations remain in each regime (i.e. $\sum_{t=1}^{n} d_{1t}, \sum_{t=1}^{n} d_{2t} \neq 0$), otherwise the specification collapses to the linear model in (1).
2. For all pair-wise combinations of (β,γ) from the respective grids, estimate $\hat{A}_1(\beta,\gamma)$, $\hat{A}_2(\beta,\gamma)$, and $\hat{\Sigma}(\beta,\gamma)$.
3. Define the estimates $(\hat{\beta},\hat{\gamma})$ as the values of (β,γ) that maximise the likelihood function in (5).
4. Set $\hat{\Sigma} = \hat{\Sigma}(\hat{\beta},\hat{\gamma})$, $\hat{A}_1 = \hat{A}_1(\hat{\beta},\hat{\gamma})$, and $\hat{A}_2 = \hat{A}_2(\hat{\beta},\hat{\gamma})$.

Although this is shown to work quite well in practice, there are some ways in which it can be improved. Consider the fact that HS use an evenly spaced grid for γ. Let $w_1(\beta)$ be a vector of stacked observations for $w_t(\beta)$ arranged in ascending numerical order. For a given value of β, there is no information about the likelihood function given in (5) for values of γ between observations of $w_1(\beta)$. To put it another way, if more than one value $\gamma_i \in \Gamma$ lies between the same two consecutive observations in $w_1(\beta)$, they will result in an identical construction of $d_{1t}(\beta,\gamma)$ and $d_{2t}(\beta,\gamma)$, yielding identical estimates and value for the likelihood function. In this way, the HS algorithm is likely to perform computations that provide no new information and are hence unnecessary. Conversely, it may be the case when conducting an evenly spaced grid that there are no values of $\gamma_i \in \Gamma$ that lie between sets of two consecutive observations in $w_1(\beta)$. In this case, some candidates for γ are simply not considered, even though the likelihood function does contain information about them. This may result in inefficiency of the estimates for all the parameters in the model.

Of course, if one uses a suitably large grid, they are more likely to suffer from unnecessary computational expense than a loss of efficiency. However, although this is not too important when carrying out the estimation, it is far from trivial when it comes to hypothesis testing. This is because γ is a nuisance parameter that is not present under the null hypothesis of linear cointegration, and hence it becomes impossible to solve for the distribution of any test statistic applied. This is known in the literature as Davies' (1977) problem. Deriving the null distribution under a residual bootstrap will obviously require repeating the estimation procedure for an absolute minimum of 1,000 replications. It is therefore preferable for the algorithm to be as fast as possible, particularly when it is desirable to compare the performance of several different models.

This issue can, however, be dealt with using only a small modification to the HS algorithm. First, the grid B for the cointegrating vector can be constructed as usual over a wide confidence interval for $\tilde{\beta}$. Now for each $\beta_i \in \mathrm{B}$ construct a different grid for the threshold based on the empirical support for $w_t(\beta_i \in \mathrm{B})$ so that $\Gamma_i = w_1(\beta_i)$. Whilst the

dimensions of the grid search remain the same, the grid values of γ are allowed to change with their accompanying values of β, ensuring that there are no superfluous computations, and that all possible points on the likelihood function are considered.

Another limitation of the HS algorithm is that it is only really feasible to implement in a bi-variate VECM with a single cointegrating vector. For larger systems, the HS grid search quickly becomes unmanageable. This is because it involves a joint grid search over the threshold parameters and the cointegrating vector. Suppose for example, we had a bi-variate model, and considered 100 candidates each for γ and β. This would require us to estimate 10,000 VECMs to determine the parameters that maximise the likelihood function. However, if we had a tri-variate system then there would be two cointegrating parameters to estimate, adding an extra dimension to the grid, requiring 1,000,000 estimations. Clearly, this makes the HS algorithm inappropriate for estimating larger systems, particularly when bootstrapping is required to produce p-values for test statistics. An alternative algorithm, which shall be referred to as a Sequentially Modified Grid-search (SMG), is now outlined, that should be able to cope with multiple cointegrating vectors, and the concerns noted above.

1. Use a linear estimator, such as Johansen's, to obtain $\tilde{\beta}$.

2. Construct a grid for γ, containing all the observations of $\tilde{w}_{t-1} = w_{t-1}(\tilde{\beta})$. Using model (4), estimate the parameters A_1, A_2, Σ over the grid, and select the value of γ that maximises the likelihood function in (5).

3. Using the value of γ acquired in the previous step, construct d_{1t} and d_{2t}, and re-estimate β in model (2). This requires estimation using non-linear least squares, under the specification that the cointegrating vector(s) are constant across regimes.

4. Reconstruct $w_{t-1}(\beta)$ and grid for γ based on the estimate of β acquired in the previous step, again selecting the candidate that produces the highest value for the likelihood function.

5. Repeat steps 3-4 while the likelihood function continues to improve.

Whereas the HS algorithm involves a joint search over the threshold and the cointegrating parameter, the SMG alternative repeats a one dimensional grid search over only the threshold (for a two regime specification). This is likely to be quicker, even for a bi-variate model. If we had, for example, 100 observations, the SMG algorithm would require fewer computations than the HS method, unless it needed to be repeated more than 100 times. Because the grid is always based on the observations of $w_{t-1}(\beta)$ for a given step, no superfluous calculations are performed, and each step records as much information about the likelihood function as possible. The biggest advantage, however, is that the grid search does not increase in dimension or size if extra cointegrating parameters, or even whole vectors, are added to the model. Only the number of regimes and observations determines this. It seems likely therefore that the SMG algorithm proposed above should be able to cope with much larger models, although it may be reasonable to expect steps 3-4 to have to be repeated a greater number of times.

3. An empirical comparison of the HS and SMG algorithms

There is currently no asymptotic distribution theory for the estimates of Threshold-Vector-Error-Correction-Models. However, it is still possible to explore the finite sample distribution of estimators via Monte-Carlo simulation. To do this, data is generated according to the following model:

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = d_{1t} \begin{bmatrix} -0.75 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + d_{2t} \begin{bmatrix} -0.25 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \qquad (6)$$

where $d_{1t} = 1(x_{1t-1} - \beta x_{2t-1} \le \gamma)$, $d_{2t} = 1(x_{1t-1} - \beta x_{2t-1} > \gamma)$,

$\begin{bmatrix} u_{1t} & u_{2t} \end{bmatrix}' \sim Niid(0,\Sigma)$, and $\Sigma = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$.

For $\beta = 1$ and $\gamma = 0$ the model is estimated using both the HS and SMG algorithms for 1000 replications, using sample sizes of n = 100 and n = 250. To elicit a clear comparison, the HS grid sizes for β and γ are set equal to the number of observations (the same as used by SMG for the threshold [1]). To preserve degrees of freedom, candidates from the grid are only considered when their choice leaves a minimum of five percent of the total observations in each regime. The results are presented in Table 1, reporting the mean, root mean squared error (RMSE), mean absolute error (MAE), and percentiles from the distributions of the estimators.

[Insert Table 1 here]

For 100 observations the performance of both algorithms is quite similar. SMG has a slightly lower RMSE and MAE for β, and marginally higher for γ, although these differences are not statistically significant. The distributions show a roughly equal degree of dispersion; however, HS appears to have a higher bias for both the cointegrating coefficient and the threshold parameter. When n = 250, there are no statistically significant differences in bias or efficiency between the two estimation procedures.

These results generally suggest that both algorithms have a roughly equal degree of efficiency. However, the most striking difference is the number of computations necessary for each replication, which is reported in Table 2.

[Insert Table 2 here]

The grid size for the HS algorithm is the same for each replication, requiring 9,000 models to be estimated when the sample size is 100, and 56,250 when there are 250 observations. Since the SMG algorithm repeats the grid search for the threshold until the likelihood function fails to improve, the number of computations varies for each model. Consequently, the average, minimum, and maximum number of estimations are reported. The average number of estimations is substantially lower for SMG, particularly as the sample size is increased. For 100 observations HS requires nearly thirty times the number of computations, and just over seventy times as many when n is set to 250. In this exercise, the maximum number of grid searches required when applying SMG was eight for both sample sizes. Even at this extreme, SMG is still much quicker than HS. To illustrate the extent of the computational burden, the number of calculations required for the full simulation are also shown. With 250 observations, the HS method required over 56 million, but for SMG it was less than 800,000. This difference is far from trivial to the applied researcher, who will need to run such simulations to acquire accurate p-values to perform hypothesis tests on each individual model they estimate.

These results can be summarised as follows. The newly proposed SMG algorithm is comparable in terms of efficiency to the existing HS method, but appears to have a slightly lower bias in smaller samples. In computational terms, SMG is far quicker. However, the main advantage of SMG is that it can be applied to much larger systems for which HS would be unfeasible. Such an application is now illustrated in the next section.

4. UK term structure of interest rates

Following Campbell and Shiller's (1987) suggestion that interest rates for various bonds should be cointegrated, a number of studies using linear VECMs for the term structure have appeared in the literature. As noted by HS, however, the theory does not necessarily imply a linear relationship. A threshold VECM is now applied to the following UK interest rates:

$R1$ = One month LIBOR rate.

$R3$ = Three month LIBOR rate.

$R6$ = Six month LIBOR rate.

The data is monthly, ranging from June 1996 to September 2001 (a total of 100 observations). An example of an unrestricted linear VECM for the three rates is given by

$$\Delta R_t = \mu + \alpha\beta' R_{t-1} + \Gamma \Delta R_{t-1} + u_t \qquad (7)$$

where $R_t' = \begin{bmatrix} R1_t & R3_t & R6_t \end{bmatrix}$. Applying the Johansen cointegration test to (7), evidence of two cointegrating vectors is found at the five percent significance level. Before proceeding to estimate a corresponding non-linear model, this leaves a number of options available for threshold estimation. For instance, one of the two cointegrating vectors may determine all regime switches, or alternatively each error correction term may respond to separate threshold values (estimating the latter would require a two dimensional grid search over both cointegrating relations). As a compromise, a single threshold value was specified for both error correction terms, with each responding individually. This is reasonable because the spreads $R1$-$R3$ and $R3$-$R6$ have a roughly equal variance. Experimentation with different possibilities also indicated that this specification produced the most significant non-linear model.

The estimated coefficients for the threshold VECM using the SMG algorithm are reported below, with standard errors in parentheses:

$$\begin{aligned} \Delta R1_t = &-0.05(R1_{t-1} - 0.83R3_{t-1} - 0.84) - 0.49d_{1t}(R1_{t-1} - 0.83R3_{t-1} - 0.84) \\ &\quad (0.09) \qquad (0.09) \qquad (0.58) \quad (0.15) \qquad (0.09) \qquad (0.58) \\ &0.29\Delta R1_{t-1} + 0.09d_{1t}\Delta R1_{t-1} + u_{1t} \\ &\quad (0.12) \qquad (0.19) \end{aligned}$$

$$\Delta R3_t = 0.01(R3_{t-1} - 0.78R6_{t-1} - 1.23) - 0.57d_{2t}(R3_{t-1} - 0.78R6_{t-1} - 1.23)$$
$$\quad (0.07) \quad (0.04) \quad (0.26) \quad (0.16) \quad (0.04) \quad (0.26)$$
$$0.20\Delta R3_{t-1} + 0.13d_{2t}\Delta R3_{t-1} + u_{2t}$$
$$(0.11) \quad (0.20)$$

$$\Delta R6_t = -0.08(R3_{t-1} - 0.78R6_{t-1} - 1.23) - 0.68d_{2t}(R3_{t-1} - 0.78R6_{t-1} - 1.23)$$
$$\quad (0.08) \quad (0.04) \quad (0.26) \quad (0.17) \quad (0.04) \quad (0.26)$$
$$0.36\Delta R6_{t-1} - 0.38d_{2t}\Delta R6_{t-1} + u_{3t}$$
$$(0.11) \quad (0.21)$$

$$d_{1t} = 1(|R1_{t-1} - 0.83R3_{t-1} - 0.84| > 0.17) \quad d_{2t} = 1(|R3_{t-1} - 0.78R6_{t-1} - 1.23| > 0.17)$$
$$\hat{\sigma}^2_{u1} = 0.01 \quad \hat{\sigma}^2_{u2} = 0.02 \quad \hat{\sigma}^2_{u3} = 0.02$$
$$\text{SupLM} = 4.09 \qquad \text{p-value} = 0.000$$

The first thing to note is the specification of the regimes. The variables d_{1t} and d_{2t} are activated by the first and second equilibrium relationships respectively, but only when the magnitude of their deviation exceeds the threshold value. This allows the response of the interest rates to change when they are too far from equilibrium, either above or below, so the response is symmetric. The justification for such a specification is based on a simple premise that some agents may have a preference to invest for long periods, whilst others will require shorter commitments. It then follows that a higher rate of return may lead to an aggregate shift towards either long or short term lending, but only if the spread is significantly large. This is very different to the HS application to US term structure, when only large negative deviations induced a change in behaviour.

The findings can be summarised as follows. When the extent of the deviation from equilibrium is less in magnitude than the estimated threshold (0.17 percentage points), response to error-correction terms does not appear to be significant. However, when disequilibrium is greater than this value, the adjustment coefficients change by -0.49, -0.57, and -0.68, which are all significant differences[2]. The short run responses however appear to change less in response to the threshold effect, except in the equation

for the six month LIBOR rate. In this case it appears that there is no short-run response at times when disequilibrium is sufficiently large, thereby increasing the relative importance of the error-correction mechanism in the outer regimes. The equilibrium error for the one and three month rates exceeds the threshold in 27% of the sample observations. The deviation of the three and six month rates from equilibrium is slightly more varied, being greater than the estimated threshold for 34% of the time during this period. The SupLM statistic proposed by HS is also presented. An asymptotic p-value is calculated using the fixed regressor boot-strap proposed by Hansen (1996), with 10,000 replications performed in this case. This tests the null hypothesis of a linear model, where $d_{1i} = d_{2i} = 0$ for all i, against the above threshold specification. The null is clearly rejected in favour of threshold cointegration.

5. Conclusion

This paper has presented a new algorithm for estimating threshold vector error correction models, based on a sequentially modified grid search. It has been shown by simulation that it produces parameter estimates with an almost identical degree of efficiency to the method proposed by Hansen and Seo, but it has a lower bias, and the computational time saved is considerable. More importantly, the new algorithm can also be used to estimate models containing multiple variables and cointegrating vectors, with the additional computational expense being minimal. This represents a significant step forward from the procedure of Hansen and Seo, which is only applicable in practice to a bivariate system with a single cointegrating vector. To illustrate the potential of the proposed algorithm, an application to the term structure of UK interest rates was then shown. Evidence is found of a threshold effect in the response to two cointegrating relationships, and the null of a linear model is strongly rejected.

References

Balke, N.S., and Fomby, T.B. (1997) 'Threshold cointegration', *International Economic Review*, 38, 626-645.

Campbell, J.Y., and Shiller, R.J, (1987) 'Yield spreads and interest rate movements: a bird's eye view', *Review of Economic Studies*, 58, 495-514.

Davies, R.B. (1977) 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, 74, 33-43.

Hansen, B.E. (1996) 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica*, 64, 413-430.

Hansen, B.E., and Seo, B. (2002) 'Testing for two-regime threshold cointegration in vector error-correction models', *Journal of Econometrics*, 110, 293-318.

Johansen, S. (1988) 'Statistical analysis of cointegration vectors' *Journal of Economic Dynamics and Control*, 12, 231-254.

Tong, H. (1983) *Threshold models in non-linear time series analysis*. Lecture notes in Statistics, No 21, Heidelberg: Springer.

Tong, H. (1990) *Non-linear time series: a dynamical system approach*. Oxford: Clarendon Press.

Table 1
Distribution of estimators

		Mean	RMSE	MAE	Percentiles (%)				
					5	25	50	75	95
Algorithm									
n = 100									
HS	$\hat{\beta} - \beta$	0.0107	0.0747	0.0500	-0.0895	-0.0269	0.0055	0.0441	0.1215
SMG	$\hat{\beta} - \beta$	-0.0000	0.0629	0.0439	-0.0956	-0.0304	0.0002	0.0332	0.0950
HS	$\hat{\gamma} - \gamma$	0.0115	0.1445	0.1243	-0.2152	-0.1068	0.0400	0.1128	0.2405
SMG	$\hat{\gamma} - \gamma$	0.0077	0.1476	0.1290	-0.2207	-0.1126	0.0376	0.1144	0.2427
n = 250									
HS	$\hat{\beta} - \beta$	-0.0004	0.0265	0.0192	-0.0425	-0.0142	-0.0004	0.0142	0.0410
SMG	$\hat{\beta} - \beta$	-0.0009	0.0263	0.0193	-0.0427	-0.0157	-0.0005	0.0128	0.0439
HS	$\hat{\gamma} - \gamma$	-0.0050	0.1074	0.0946	-0.1780	-0.0936	0.0228	0.0779	0.1536
SMG	$\hat{\gamma} - \gamma$	-0.0065	0.1060	0.0923	-0.1769	-0.0892	0.0179	0.0773	0.1527

Table 2
Computational requirements of the HS and SMG algorithms

		Required number per replication				Simulation Total	
		HS	SMG(MIN)	SMG(AVE)	SMG(MAX)	HS	SMG
Estimations	n = 100	9,000	182	305.760	728	9,000,000	305,760
	n = 250	56,250	675	796.725	1800	56,250,000	796,725
Grid searches	n = 100	1	2	3.360	8	1,000	3,360
	n = 250	1	3	3.541	8	1,000	3,541

[1] Note that Hansen and Seo perform a similar simulation exercise. It was, however, felt necessary to produce some new results for their method, not only so that the grid sizes were equivalent, but also so that the HS and SMG algorithms were applied to identical sets of experimental data.

[2] Individual hypothesis tests should be invariant to the nuisance parameter problem since the threshold enters the model through all explanatory variables. This only becomes an issue when testing against the null of the appropriate linear specification.