

PhD Dissertation

Line Elvstrøm Ekner

Department of Economics, University of Copenhagen

Cointegration and Regime Switching Dynamics in Macroeconomic Applications

Academic advisor: Heino Bohn Nielsen

Submitted: January 24, 2014

Contents

Summary	3
Resumé	7
Acknowledgements	11
Chapter 1	12
Okun's Law with Non-Linear Dynamics - a smooth transition VECM applied to the U.S. economy	13
Chapter 2	44
Smooth or Non-Smooth Regime Switching Models (joint with Emil Nejstgaard)	45
Chapter 3	68
A Structural Analysis of the Convenience Yield on U.S. Treasury Bonds	69

Summary

This dissertation is comprised of three self-contained papers, each of which can be read separately. All three chapters are concerned with the same topic, though, namely the application of regime switching models to time series of macroeconomic or financial data. The overall aim is to narrow the gap between theoretically developed multivariate regime switching time series models and the application to data. It is often documented that relationships between economic variables appear non-linear or characterized by regime switching behavior, e.g., they depend on the state of the business cycle. In addition, many of these variables need to be modeled simultaneously to account for all dynamic effects. Chapter 1 and 3 include empirical applications that account for these two aspects by combining regime switching behavior with the cointegrated vector autoregressive model (CVAR) of Johansen (1996) and Juselius (2006). Estimation of such models is, however, non-standard and often possess serious problems to the researcher. Chapter 2, therefore, concerns theoretical and practical details about estimation and identification of smooth regime switching models. The parameter that determines the speed of transition between the regimes of the popular logistic smooth transition autoregressive (LSTAR) model is particular difficult to identify, see Chan and Tong (1986) and Teräsvirta (1994), and chapter 2 analyzes the consequences hereof and suggests useful tools to improve an econometric analysis.

Chapter 1, “*Okun’s Law with Non-Linear Dynamics - a smooth transition VECM applied to the U.S. economy*”, investigates regime switching behavior in the Okun’s law relationship. Okun’s law is a macroeconomic relationship relating the activity in the goods market to the activity in the labor market. The variables of the unemployment rate and output are defined as gap variables where the individual trends are deterministic and estimated simultaneously with the coefficients of the vector error correction model (VECM). Okun’s law is then defined as a linear cointegration relationship with non-linear short-run dynamics. This specification allows for the well-known fact of asymmetry in the steepness of the unemployment rate, see, e.g., Neftci (1984). The relationship is analyzed by means

of a smooth transition VECM (STVECM). The model is estimated by maximum likelihood but is non-standard due to the transition function and its parameters, and the important test of linearity is carried out using a recent proposed sup-LR test of Kristensen and Rahbek (2013). The regimes of the STVECM overlap with those of the National Bureau for Economic Research business cycle indicator and can be interpreted as a recession regime and an expansion regime. The estimated threshold implies that when output falls by more than 0.26% in the previous quarter, the model enters the recessionary regime. The dynamics of the two regimes differ significantly because of the large increases in the unemployment rate during recessions which the model approximates by an explosive root. Okun's coefficient is estimated to -0.28 and thereby close to the $-1/3$ rule of thumb benchmark originally suggested by Arthur Okun in 1962.

Chapter 2, "*Smooth or Non-Smooth Regime Switching Models*" (joint with Emil Nejsgaard), analyzes the problem of selecting between regime switching models with discrete and smooth regime switching. Regime switching models characterized by smooth transitions only differ from discrete regime switching models by the speed of transition parameter. Thus, estimation and identification of this parameter is essential not only for economic interpretation but also for model selection. Nevertheless, the identification problem and its consequences for estimation have received little attention in the LSTAR literature. We show that the original parametrization of the speed of transition parameter is problematic as the likelihood function is characterized by large flat areas caused by all derivatives approaching zero with faster speed of transition. This implies that the magnitude of the estimator may depend on the arbitrary chosen stopping criteria of the numerical optimizer. To circumvent this problem, we propose a reparametrization of the LSTAR model. The reparametrization maps the parameter space of the original speed of transition parameter into a much smaller interval which facilitates identifying the global maximum of the likelihood function as well as numerical optimization. We then show that the threshold autoregressive model can be the global maximum of a LSTAR likelihood function, while it, by construction, is always at least a local maximum. Instead of relying solely on economic theory when justifying the additional parameter of the LSTAR model, we show that information criteria provide a model selection tool that can be applied if the researcher wishes to comment on the speed of transition. We illustrate the benefits of the new parametrization and the model selection procedure with data from two published applications. The new parametrization provides general insights on the shape of the likelihood function in directions of the two parameters of the transition function that can be generalized to a broad range of other models within the smooth switching literature.

Chapter 3, “*A Structural Analysis of the Convenience Yield on U.S. Treasury Bonds*”, analyzes the dynamics of the convenience yield on U.S. Treasury bonds using annual observations from 1919-2008. The convenience yield arises because of a special clientele demand for Treasury bonds as they offer extreme liquidity and safety, attributes valuable for, e.g., foreign central banks and insurance companies. This demand suggests an inverse relationship between the supply of Treasury bonds and the convenience yield, measured by the spread between yields on corporate bonds and Treasury bonds, such that a lower supply is associated with a larger convenience yield. We model both the long-run and short-run dynamics of the persistent variables measuring Treasury supply and convenience yield in VECM. The long-run relationship is negative reflecting a downward-sloping demand function as expected. Yet, this relationship implies that a fall in Treasury supply has adverse economic effects because the associated increase in the convenience yield reflects a muted transmission mechanism to the private sector. The paper emphasizes, however, that this downward-sloping demand function can only be interpreted as a long-run demand relationship. In the short-run, the demand function has a positive slope or is at best flat. This short-run relationship is identified by constructing impulse responses from a just-identified structural VECM. When extending the model to separate the convenience yield into a liquidity premium and safety premium, the safety premium turns out to be the driving factor of the short-run positive relationship. The safety premium is reduced on impact and the first two years after a negative shock to the supply of Treasury bonds, before turning positive and raised in accordance with the long-run negative relationship. The results are robust to a potential kink in the demand relation estimated by a threshold VECM. Hence, the economic effects following a fall in Treasury supply may not be as adverse as first thought.

Bibliography

CHAN, K. S. AND H. TONG (1986): “On estimating thresholds in autoregressive models.” *Journal of Time Series Analysis*, 7(3):179–190.

JOHANSEN, S. (1996): “Likelihood-Based Inference in Cointegrated Vector Autoregressive

Model.” *Advanced Texts in Econometrics*, Oxford University Press.

JUSELIUS, K. (2006): “The Cointegrated VAR Model: Methodology and Applications.” *Oxford University Press*.

KRISTENSEN, D. AND A. RAHBEK (2013): “Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models.” *Econometric Theory*, 29.

NEFTCI, S. N. (1984): “Are Economic Time Series Asymmetric over the Business Cycle?” *Journal of Political Economy*, 92(2):307–28.

OKUN, A. (1962): “Potential GNP: Its measurement and significance.” *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pages 89–104.

TERÄSVIRTA, T. (1994): “Specification, estimation, and evaluation of smooth transition autoregressive models.” *Journal of American Statistical Association*, 89:208–218.

Resumé

Denne afhandling består af tre selvstændige papirer, som hver især kan læses separat. Alle tre kapitler beskæftiger sig dog med det samme overordnede emne: Anvendelsen af regimeskiftmodeller på tidsrække­data fra makroøkonomiske eller finansielle variable. Det overordnede formål er at indsnævre kløften mellem teoretisk udviklede multivariate regimeskiftmodeller for tidsrække­data og anvendelser på økonomisk data. Det er ofte dokumenteret, at relationer mellem økonomiske variable fremstår som ikke-lineære. Fx ændrer mange relationer sig i takt med konjunkturudviklingen. Desuden skal flere af disse variable modelleres simultant for at tage højde for alle dynamiske effekter. Kapitel 1 og 3 omfatter empiriske applikationer, der kombinerer regimeskift-adfærd og den kointegrerede VAR model af Johansen (1996) og Juselius (2006). Kapitel 2 omhandler teoretiske og praktiske omstændigheder omkring estimation og identifikation af den parameter, der bestemmer hastigheden af transitionen mellem regimerne i den populære logistic smooth transition autoregressive (LSTAR) model, jf. Chan and Tong (1986) og Teräsvirta (1994).

Kapitel 1, "*Okun's Law with Non-Linear Dynamics - a smooth transition VECM applied to the U.S. economy*", undersøger, hvorvidt der er regimeskift-adfærd i Okuns lov, som stammer fra Okun (1962). Okuns lov er et makroøkonomisk forhold, der relaterer aktiviteten i økonomien til aktiviteten på arbejdsmarkedet. Arbejdsløshedsraten og output defineres som gap variable, hvor den underliggende trendudvikling er deterministisk og estimeres simultant med koefficienterne i den kointegrerede VAR model. Okuns lov defineres som en lineær kointegrationsrelation med ikke-lineær kortsigtdynamik. Denne specifikation tillader for det velkendte faktum, at der er asymmetri i ændringerne i arbejdsløshedsraten, se fx Neftci (1984). Forholdet analyseres ved hjælp af en smooth transition vector error correction model (STVECM). Modellen estimeres med maximum likelihood men er ikke-standard på grund af transitionsfunktionen og dens parametre. Det afgørende test for linearitet foretages ved hjælp af et nyligt foreslået sup-LR test af Kristensen and Rahbek (2013). Regimerne i den estimerede STVECM overlapper med National Bureau for Economic Research konjunkturindikatoren og kan tolkes som henholdsvis et

recessionsregime og et ekspansionsregime. Den estimerede grænseværdi for regimeskift indebærer, at når output falder med mere end 0,26% i det foregående kvartal, indtræder modellen i recessionsregimet. Dynamikken i de to regimer adskiller sig væsentligt og skyldes, at modellen beskriver de store stigninger i arbejdsløshedsraten under lavkonjunkturer ved hjælp af en eksplosiv rod. Okuns koefficient estimeres til $-0,28$ og dermed tæt på det $-1/3$ benchmark, som oprindeligt blev foreslået af Arthur Okun i 1962.

Kapitel 2, "*Smooth or Non-Smooth Regime Switching Models*" (skrevet i samarbejde med Emil Nejstgaard), analyserer problemet med at vælge mellem regimeskiftmodeller med diskret og kontinuert regimeskift. Regimeskiftmodeller karakteriseret ved kontinuer-te regimeskift adskiller sig kun fra diskrete regimeskiftmodeller i form af en hastighedspa-rameter, der bestemmer hastigheden af regimeskiftene. Derfor er estimation og identifi-kation af denne parameter afgørende ikke blot for den økonomiske fortolkning men også for valg af model. Ikke desto mindre har identifikationsproblemet og dets konsekvenser for estimation fået meget lidt opmærksomhed i litteraturen. Vi viser, at den oprindelige parametrisering af hastighedsparameteren ikke er hensigtsmæssig, da likelihoodfunktio-nen er karakteriseret ved store flade områder forårsaget af, at alle afledte af funktionen går mod nul med hurtigere hastighed. Dette indebærer, at størrelsen af estimatoren kan af-hænge af de vilkårligt valgte stoppekriterier for den numeriske optimeringsalgoritme. For at omgå dette problem, foreslår vi en reparametrisering af LSTAR modellen. Reparametri-seringen afbilder parameterrummet af den oprindelige hastighedsparameter til et meget mindre interval, som letter identifikationen af det globale maksimum af likelihoodfunktio-nen samt numerisk optimering. Vi viser, at threshold autoregressive (TAR) modellen med diskrete regimeskift kan være det globale maksimum på en LSTAR likelihoodfunktio-n, mens det per konstruktion altid er mindst et lokalt maksimum. I stedet for udeluk-kende at skele til økonomisk teori når den ekstra parameter i LSTAR modellen skal ret-færdiggøres, viser vi, at informationskriterier kan bruges til vælge mellem en LSTAR og en TAR model, hvis analytikeren ønsker at kommentere på hastigheden af regimeskiftene. Vi illustrerer fordelene ved den nye parametrisering og metoden til at vælge mellem de to modeller med data fra to publicerede artikler. Den nye parametrisering giver en generel indsigt i formen af likelihoodfunktionen i retning af de to parametre i transitionsfunktio-nen, som kan generaliseres til en bred vifte af andre modeller inden for litteraturen om smooth transition modeller.

Kapitel 3, "*A Structural Analysis of the Convenience Yield on U.S. Treasury Bonds*", ana-lyserer dynamikken i bekvemmelighedsafkastet på amerikanske statsobligationer ved hjælp af årlige observationer fra 1919 til 2008. Bekvemmelighedsafkastet opstår på grund af en

særlig efterspørgsel efter statsobligationer, forårsaget af den ekstreme likviditet og sikkerhed, som de tilbyder, og som er egenskaber værdifulde for fx udenlandske centralbanker og forsikringselskaber. Denne efterspørgsel er et tegn på et inverst forhold mellem udbuddet af statsobligationer og spændet mellem renterne på virksomhedsobligationer og statsobligationer. Denne sammenhæng betyder dog, at et fald i udbuddet af statsobligationer har negative økonomiske effekter, fordi det er forbundet med en stigning i bekvemmelighedsafkastet, hvilket indikerer en begrænset transmissionsmekanisme til den private sektor. Artiklen modellerer både den langsigtede og kortsigtede dynamik i de persistente variable, der måler udbuddet af statsobligationer og bekvemmelighedsafkastet i en vector error correction model (VECM). Den langsigtede sammenhæng er negativ som forventet og afspejler en nedadgående efterspørgselskurve og dermed utilsigtede negative økonomiske konsekvenser af et fald i den offentlige gæld. Vi viser, at denne negative sammenhæng kun kan tolkes som en langsigtet efterspørgselsrelation. På kort sigt har efterspørgselsfunktionen en positiv hældning eller er i bedste fald flad. Dette kortsigtsforhold identificeres ved at konstruere impulsresponsen fra en eksakt-identificeret strukturel VECM. Ved at udvide modellen med en ekstra variabel kan vi opdele bekvemmelighedsafkastet i en likviditetspræmie og en sikkerhedspræmie. Sikkerhedspræmien viser sig at være den afgørende faktor bag det positive kortsigtsforhold. Den reduceres øjeblikkeligt og er negativ i de første to år efter et negativt stød til udbuddet af statsobligationer, før den bliver positiv og øges i overensstemmelse med det langsigtede negative forhold. Resultaterne er robuste over for et potentielt knæk i efterspørgselskurven, som estimeres ved en threshold VECM. De økonomiske effekter af et fald i udbuddet af statsobligationer er måske derfor ikke så dårlige som først antaget.

Litteratur

- CHAN, K. S. AND H. TONG (1986): "On estimating thresholds in autoregressive models." *Journal of Time Series Analysis*, 7(3):179–190.
- JOHANSEN, S. (1996): "Likelihood-Based Inference in Cointegrated Vector Autoregressive Model." *Advanced Texts in Econometrics*, Oxford University Press.

- JUSELIUS, K. (2006): "The Cointegrated VAR Model: Methodology and Applications." *Oxford University Press*.
- KRISTENSEN, D. AND A. RAHBK (2013): "Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models." *Econometric Theory*, 29.
- NEFTCI, S. N. (1984): "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy*, 92(2):307–28.
- OKUN, A. (1962): "Potential GNP: Its measurement and significance." *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pages 89–104.
- TERÄSVIRTA, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models." *Journal of American Statistical Association*, 89:208–218.

Acknowledgements

This dissertation was written during my enrollment at the PhD programme in Economics at the Department of Economics at the University of Copenhagen, 2009-2013.

During this process, a number of people deserve my thanks and gratitude. First, I would like to thank my supervisor, Heino Bohn Nielsen, whose door was always open for guidance, constructive comments and discussions, which I have greatly benefited from.

I am grateful to fellow PhD students for valuable discussions throughout the years. In particular, I owe Andreas Noack Jensen a debt of gratitude for many helpful talks and for enjoyable company at the office.

During my studies, I spent two semesters at Columbia University in New York, US, as a visiting graduate student. I want to thank Dennis Kristensen for the invitation and for guidance during my stay. Moreover, I am thankful to the Bikuben Foundation, the Denmark-America Foundation, Oticon, Augustinus and Dansk Tennis Fond for financial support making my stay at Columbia University possible.

Last, but not least, I want to thank my parents, sister, grandmother, Kristian Pless and friends for their loving support during the many ups and downs.

Line Elvstrøm Ekner

Copenhagen, January 2014

Chapter 1

Okun's Law with Non-Linear Dynamics - a
smooth transition VECM applied to the U.S.
economy

Okun's Law with Non-Linear Dynamics

- a smooth transition VECM applied to the U.S. economy

Line Elvstrøm Ekner*

Department of Economics, University of Copenhagen

ABSTRACT: The relationship of Okun's law between measures of output and the unemployment rate is defined as a linear cointegration relationship with non-linear short-run dynamics. This specification allows for the well-known fact of asymmetry in the steepness of the unemployment rate. The relationship is analyzed for the U.S. economy in a smooth transition vector error correction model (STVECM). The model is estimated by maximum likelihood but is non-standard due to the transition function and its parameters, and the important test of linearity is carried out using a recent proposed sup-LR test. The two regimes of the fitted STVECM have asymmetric dynamics and can be classified as a recession regime and an expansion regime. The main source of non-linearity stems from economic downturns where the model switches to the explosive recession regime to explain the observed steep increases in the unemployment rate, before it subsequently returns to the stationary expansion regime. Okun's coefficient is estimated to -0.28 and thereby close to the -1/3 rule of thumb benchmark originally suggested by Arthur Okun in 1962.

1. INTRODUCTION

Okun's law is a macroeconomic relationship relating the activity in the goods market to the activity in the labor market. Arthur Okun gave name to the relationship after his seminal contribution, Okun (1962), in which he found that output and the unemployment rate were inversely related with a regression coefficient of about 1/3 using U.S. post-war data. This coefficient has since been referred to as Okun's coefficient.

Most empirical specifications assume a relationship with linear dynamics. This assumption, which implies, e.g., that expansions and contractions in output are associated with the same absolute changes in the unemployment rate, might not be appropriate. Several studies characterize that the U.S. unemployment rate as asymmetric with sharper increases than decreases, see, e.g., Neftci (1984), Montgomery *et al.* (1998), McKay and Reis (2008) and Teräsvirta *et al.* (2011). We allow for this observation by defining Okun's law as a linear cointegration relationship with non-linear short-run dynamics.

*email: line.elvstrom.ekner@econ.ku.dk

This paper contributes to the inconclusive vector framework studies of a non-linear Okun's law relationship. Output and the unemployment rate are measured in terms of deviations from a trend and are thereby gap variables. These trends are solely deterministic and, different from previous studies, estimated simultaneously with the coefficients of the cointegrated vector autoregressive (CVAR) model, which is the baseline linear model for the analysis. Thus, the resulting stochastic trends (the gaps) form the Okun's law cointegration relationship which reflects the interdependence of the two variables over the longer term. The short-run dynamics of the relationship are modeled by means of the smooth transition autoregressive (STAR) methodology in which the dynamics depend non-linearly on a transition variable. This extension facilitates asymmetry of the two resulting regimes and the transition between regimes may be smooth rather than discrete. The combination of a STAR model and a VECM is here dubbed a STVECM. Furthermore, and contrary to previous literature, the variable governing the non-linearity is not prespecified. The STVECM is estimated for various choices of this transition variable, including the successful output variable of the single-equation analyses.

A second contribution of this paper lies in the application of the STVECM. There has been extensive research in the field of univariate and single-equation STAR models, see Dijk *et al.* (2002) for a survey. The STVECM is, however, a fairly new model in empirical applications. The first attempts at extending the STAR methodology to a vector framework are found in Weise (1999) and Rothman *et al.* (2001). Contrary to these studies the cointegration vector and short-run dynamics are estimated simultaneously based on (quasi) maximum likelihood (ML) to avoid issues with the error otherwise present from a first step estimation of the cointegration vector.¹ Such procedure is proposed in theoretical contributions by Nedeljkovic (2008) and Kristensen and Rahbek (2010; 2013). Moreover, the essential test of linear short-run dynamics is carried out by a recent proposed sup-LR test which approximates the unknown distribution of test statistic by bootstrap simulations.

The result shows that the dynamics of the Okun's law relationship are indeed non-linear. The linear model is rejected when tested against the STVECM. The two regimes of the STVECM are interpretable as a recession regime and a normal time or expansion regime, and roughly coincide with business cycles as classified by the National Bureau for Economic Research (NBER). During economic downturns the model temporarily switches to the recession regime and explains the observed large and swift increases in the unemployment rate by a mildly explosive root. The estimate of the long-run Okun's coefficient is -0.28 and thereby close to the $-1/3$ rule of thumb benchmark suggested by Arthur Okun

¹In a non-linear model an estimation error from a first step may affect the limiting distribution of the parameters, cf. De Jong (2001) and Seo (2011).

in 1962.

The literature on empirical applications of Okun's law is reviewed in section 2. Section 3 introduces the baseline bivariate cointegration model and define the gap variables. The framework of the STVECM and the ML estimation are discussed in section 4. Section 5 follows with the empirical analysis and the results. Finally, section 6 concludes.

2. LITERATURE

Since Okun (1962), numerous studies have investigated the symmetric relationship between the unemployment rate and output for different time periods and for different countries, see *inter alia* Gordon (1984), Kaufman (1988), Prachowny (1993), Weber (1995), Moosa (1997), Moosa (1999), Attfield and Silverstone (1998), Lee (2000), Silverstone and Harris (2001), Sogner and Stiassny (2002), Frank (2008) and Ball *et al.* (2013). Frank (2008) includes a table with a comprehensive list of estimated Okun's coefficients from different studies. While there seems to be consensus concerning the empirical validity of Okun's law with a negative Okun's coefficient, discrepancy as to the actual magnitude of this coefficient exists. Less attention has been given to analyze asymmetry in Okun's law relationship. As with the literature considering a symmetric Okun's law relationship, studies of asymmetric relationships also tend to use varying model specifications, estimation and detrending methods, as well as different country samples. Table 2.1 summarizes the results of studies considering asymmetry in the Okun's law relationship for the U.S. economy.

Table 2.1: Results of asymmetric empirical applications of the Okun's law relationship to the U.S. economy.

Study	Sample	Approach	s_t	Expected asymmetry
Lee (2000)	1955-1996	gap/static	u_t	yes, but insign.
Lee (2000)	1955-1996	diff./static	Δu_t	yes
Silverstone and Harris (2001)	1978:1-1999:1	diff./TVECM	ecm_{t-1}	no
Viren (2001)	1960-1997	diff.*	u_t	yes
Cuaresma (2003)	1965:1-1999:1	gap/dynamic	y_t	yes
Silvapulle <i>et al.</i> (2004)	1947:1-1999:4	gap/dynamic	y_t	yes/no(short-run)
Frank (2008)	1980:1-2004:1	gap/dynamic	y_t	yes, but insign.
Mendonca (2008)	1948:1-2007:2	gap/TVECM	ecm_{t-1}	no

Note: s_t is the transition variable determining the prevailing regime. s_t is a gap variable, unless else stated. u_t and y_t are variables of unemployment rate and output, respectively. Approach refers to whether variables are defined as gap variables or in first differences, and whether lags are included in the regression. TVECM is a threshold VECM originating in Balke and Fomby (1997). Expected asymmetry refers, broadly, to faster adjustment in the unemployment rate in periods of low growth than in periods of higher growth.

*This study augments a first difference version of Okun's law by a working age population variable and an error correction term reflecting deviations from an equilibrium relation between number of unemployed persons, working age population and a time trend.

A static approach to Okun's law is employed in Lee (2000) and Viren (2001). They identify asymmetry by means of an indicator function² allowing the effects of the output variable to depend on whether transition variable (s_t in table 2.1) is above and below zero, or above and below an estimated threshold value. Notably, Lee (2000) uses the unemployment rate as the independent variable and, consequently, the change in the unemployment rate serves as an indicator of the state of the economy.³ Asymmetry is found using first difference variables with Okun's coefficient being larger (in absolute value) for increases in the unemployment rate than for decreases.

Silvapulle *et al.* (2004), Cuaresma (2003) and Frank (2008) find some evidence of non-linearity of the expected form in their single-equation versions of Okun's law with constructed gap variables as measurements. However, in such single-equation models, with the unemployment rate as the dependent variable, the implicit assumption of no feedback from the unemployment rate to output may be violated and the risk of violation increases with lower frequency of the data. For instance, if annual data is used, changes in the unemployment rate, initiated by output changes, may feedback and affect output within the same period. For this reason, and among other things, Silverstone and Harris (2001) and Mendonca (2008) relax this assumption by analyzing Okun's law relationship in a vector framework given by a threshold VECM (TVECM). The non-linear short-run dynamics are

²The indicator function is defined as $I[A] = 1$ if A is true and $I[A] = 0$ otherwise.

³Lee (2000) argues that general conclusions based on this specification are qualitatively the same as those reported from studies with output as the independent variable.

determined by deviations from a linear Okun's law cointegration relation estimated in a first step. The two studies differ in their deterministic specification but neither of them find asymmetry of the form expected, and, if anything, the unemployment rate adjusts significantly downward but not upward.

Overall, for the U.S. economy, the evidence for asymmetry of the form expected with the unemployment rate adjusting faster in contractions than in expansions is mixed.

3. THE LINEAR MODEL: VECM

Similar to Silverstone and Harris (2001) and Mendonca (2008), we analyze the Okun's law relationship by interpreting it as a long-run linear relationship between output and the unemployment rate with possible non-linear short-run dynamics. The analysis thus rests on the assumption that Okun's law is a linear equilibrium relation serving as a resting point towards which the stochastic processes of output and the unemployment rate are drawn after they have been pushed away. This assumption is motivated by the numerous studies that since Okun (1962) have found a stable linear Okun's law relationship for the U.S., also over longer horizons.

The statistical framework for the analysis is the CVAR model of Johansen (1996) and Juselius (2006). The CVAR framework accommodates a bivariate specification of the output and unemployment rate dynamics which is consistent with the long-run Okun's law interpretation. The baseline linear model for the analysis is the p -dimensional CVAR in VECM form:

$$\Delta Z_t = \alpha \beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Z_{t-i} + \epsilon_t, \quad t = 1, 2, \dots, T. \quad (3.1)$$

Z_t is a $p \times 1$ vector of variables and \bar{Z}_t is $(Z_t', 1)'$. α and β are matrices of dimension $p \times r$ and $(p+1) \times r$, respectively, with $\beta' \bar{Z}_t$ representing the $r \leq p$ cointegrating relationships and α giving the direction and speed of adjustment towards equilibrium. k is the number of lags in the corresponding VAR presentation and the autoregressive coefficients, $\Gamma_1, \dots, \Gamma_{k-1}$, are of dimension $p \times p$. For estimation, it is assumed that ϵ_t is an i.i.d. Gaussian sequence, $N_p(0, \Omega)$, while conditioning on the initial values, Z_{-k+1}, \dots, Z_0 . Estimation is then carried out using the ML procedure of Johansen (1996), where maximization of the log-likelihood function is reduced to solving an eigenvalue problem.

3.1. SPECIFICATION OF GAP VARIABLES

The gap version of Okun's law is constructed by estimating and extracting the deterministic trends of the variables within the CVAR model. I.e., the trends are estimated simultaneously with the coefficients of the CVAR model such that the stochastic trends left in the variables form linear relationships. If one or more of these relationships are stationary, the deterministic detrended (or gap) variables cointegrate. Hence, a cointegration relationship of this kind can be interpreted as an estimated gap version of the Okun's law relationship. Note, however, that these gap variables are not interpretable and comparable to the notion of gap variables normally used in the literature since they are not stationary.⁴ Employing the approach of Christensen and Nielsen (2009), the CVAR model is used to approximate the evolution of the deterministic trends of the unemployment rate and the level of output. To extract information about these trends, the trend of the unemployment rate is assumed to be smooth and a fourth-order polynomial in time is used as approximation. This trend may be seen as an estimate of the evolution of the natural rate of unemployment. The order of the polynomial is chosen rather arbitrary, but is yet a testable assumption. The trend in output is approximated by a linear trend in time.

In effect, the simultaneous estimation of deterministic trends and the CVAR coefficients implies that the system is reformulated in terms of gap variables. Following Christensen and Nielsen (2009), the vector of variables becomes

$$Z_t = X_t - \varphi D_t \quad (3.2)$$

where $D_t = (t \ t^2 \ t^3 \ t^4)'$, such that φD_t is a fourth-order polynomial in time. X_t is the vector of observed variables, $X_t = (y_t^{obs} \ u_t^{obs})'$. To avoid polynomial terms in the trend of y_t^{obs} , and enable only a linear trend, the following restriction is imposed on φ :

$$Z_t = \begin{pmatrix} y_t \\ u_t \end{pmatrix} = X_t - \varphi D_t = \begin{pmatrix} y_t^{obs} \\ u_t^{obs} \end{pmatrix} - \begin{pmatrix} \varphi_{11} & 0 & 0 & 0 \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{24} \end{pmatrix} \begin{pmatrix} t \\ t^2 \\ t^3 \\ t^4 \end{pmatrix}. \quad (3.3)$$

To solve the well-known problem of multicollinearity present in a regular polynomial, D_t is replaced by a Legendre polynomial of the same order. A Legendre polynomial is characterized by its orthogonal terms, see Courant and Hilbert (2008: ch.2), and greatly facilitates

⁴Often gap variables are constructed by extracting deterministic *and* stochastic trends by means of a filter, for instance, the HP, Kalman, or Harvey filter. As a result, such gap series are stationary by construction.

estimation of the polynomial coefficients. The baseline linear model is thus the VECM in (3.1) with $Z_t = (y_t \ u_t)'$ defined in (3.3). Hence, with $p = 2$ at most one, $r \leq 1$, cointegration relationship exists among the variables of Z_t . Note that if $\varphi_{22} = \varphi_{23} = \varphi_{24} = 0$ the model is equivalent to one with two restricted trends. The present specification, however, allows for a richer specification of trends.

In the CVAR, the short-run adjustment is linear. This linearity property imposes at least two strong restrictions on the underlying economic behavior:

- (i) The short-run dynamics are a constant proportion of the equilibrium error in the previous period and of lagged changes in the variables.
- (ii) The short-run dynamics are symmetric. For example, all phases of the business cycle induce identical dynamics.

The VECM in (3.1) can be modified to relax restrictions (i) and (ii). The resulting model is here the STVECM presented next.

4. THE NON-LINEAR MODEL: STVECM

The STAR methodology enables gradual transitions between two or more regimes, such that the regime switches are not necessarily abrupt like in a threshold autoregressive (TAR) model. Moreover, the smoothness of the STAR likelihood function simplifies asymptotic theory significantly compared to a TAR model. The STVECM results from adding the STAR methodology to the VECM framework of section 2.

4.1. THE MODEL

The two-dimensional STVECM with two regimes is defined as:

$$\begin{aligned} \Delta Z_t = & \left(\alpha^1 \beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i^1 \Delta Z_{t-i} \right) (1 - G(s_t; \gamma, c)) \\ & + \left(\alpha^2 \beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i^2 \Delta Z_{t-i} \right) G(s_t; \gamma, c) + \varepsilon_t \end{aligned} \quad (4.1)$$

for $t = 1, 2, \dots, T$. The dimensions of the vectors and matrices are equivalent to those of the VECM in (3.1). Note that only the short-run parameters differ between regimes, while the cointegration vector, β , is assumed constant across regimes. s_t is the transition variable governing the regime shifts, whereas c and γ are parameters of the transition function

$G_t := G(s_t; \gamma, c)$, which is continuous and bounded between zero and unity. Thus, one of the extreme regimes is associated with $G_t = 0$ and adjustment coefficients $\alpha^1, \Gamma_i^1, \dots, \Gamma_{k-1}^1$, and the other extreme regime with $G_t = 1$ and adjustment coefficients $\alpha^2, \Gamma_i^2, \dots, \Gamma_{k-1}^2$. The model is thereby a regime switching model in which the transition between regimes may be smooth.

A frequently applied transition function is the logistic transition function, leading to a logistic STAR (LSTAR) model:

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(s_t - c)/\hat{\sigma}_s\}}, \quad \gamma > 0 \quad (4.2)$$

where $\hat{\sigma}_s$ is the sample standard deviation of s_t . The inclusion of $\hat{\sigma}_s$ normalizes the deviations of s_t from the threshold value c and facilitates interpretation and estimation of the speed of transition parameter, γ . The logistic transition function is characterized by changing monotonically from zero to unity as s_t increases. The threshold parameter c determines when the transition occurs, while γ indicates, as a function of s_t , how rapid the transition from zero to unity is. As γ increases, G_t approaches the indicator function, $I[s_t > c]$, and, consequently, G_t changes from zero to unity almost instantaneous at $s_t = c$. Hence, the TVECM of Silverstone and Harris (2001) and Mendonca (2008) is on the boundary of the STVECM with $\gamma \rightarrow \infty$. The linear model, VECM, is also on the boundary of the STVECM and prevails when $\gamma = 0$. The logistic transition function provides a suitable framework in the present context where the expected asymmetry of the Okun's law relationship is with respect to different states of the economy approximated by the magnitude of s_t (e.g., GDP growth).⁵

The STVECM applied in the empirical analysis consists of (4.1) with the logistic transition function defined in (4.2), and the model is estimated by ML.

4.2. MAXIMUM LIKELIHOOD ESTIMATION

The parameters of the STVECM are estimated based on the Gaussian log-likelihood implying that for the log-likelihood function, the residuals are defined as

$$\varepsilon_t(\theta) = \Delta Z_t - \left(\alpha^1 \beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i^1 \Delta Z_{t-i} \right) (1 - G(s_t; \gamma, c))$$

⁵The related exponential transition function, leading to an ESTAR model, is in contrast appropriate in situations where asymmetry is believed associated with the absolute size of the transition variable.

$$-\left(\alpha^2 \beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i^2 \Delta Z_{t-i}\right) G(s_t; \gamma, c) \quad (4.3)$$

with $\theta = (\alpha^1, \Gamma_1^1, \dots, \Gamma_{k-1}^1, \alpha^2, \Gamma_1^2, \dots, \Gamma_{k-1}^2, \beta', \gamma, c)$. Then, given T observations, Z_1, Z_2, \dots, Z_T , and the initial values Z_{-k+1}, \dots, Z_0 fixed, the concentrated log-likelihood function takes the form (up to the scale of a constant):

$$L_T(\theta) = L_T(\hat{\Omega}(\theta)) = -\frac{T}{2} \log |\hat{\Omega}(\theta)| \quad (4.4)$$

where $\hat{\Omega}(\theta) = \frac{1}{T} \sum_{t=1}^T \varepsilon_t(\theta) \varepsilon_t(\theta)'$ is the ML estimator of the covariance matrix Ω . The ML estimator of θ is thus defined as:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} L_T(\theta) \quad (4.5)$$

$\Theta = \{\alpha^1, \Gamma_1^1, \dots, |\gamma > 0, P(|s_t| < c) \leq 0.8\}$. The maximization problem is constrained to ensure economic interpretability of the estimated model by avoiding sorting too few observations in one regime. In contrast to ML estimation of the CVAR model, no analytical solution exist and the problem is accordingly solved by numerical optimization.

The estimation of the STVECM is non-standard due to well-documented problems with joint estimation of the two parameters of the transition function, cf., Teräsvirta (1994; 1998) and Dijk *et al.* (2002). Several local maxima may exist for different combinations of c and γ , and, hence, the estimation result is sensitive to the choice of starting values for γ and c . The examples of profiled log-likelihood functions in figure 4.1 below illustrate the high degree of curvature present in the current model for some choices of the transition variable s_t .

The search for starting values of γ and c are found by a two-dimensional coarse grid, as recommended by Dijk *et al.* (2002).⁶ Another issue is the difficulty of identifying γ , in particular when being large. To obtain an accurate estimate of γ one needs many observations in the immediate neighborhood of c since even large changes in γ have only small effect on the shape of the transition function. This is, however, rarely the case and a

⁶To reduce the computational time of the grid search, the likelihood function in (4.5) is further concentrated on (β, γ, c) , while the rest of the parameters are estimated by OLS. The two-dimensional grid search over γ and c is performed for $\gamma \in [0.5; 1.5, \dots, 20.5] \cup [30, 40, 60, 80, 100]$ and $c_{min} = s_t|_{t=T*10\%}$ to $c_{max} = s_t|_{t=T*90\%}$ with a step of $(c_{max} - c_{min})/25$, respectively.

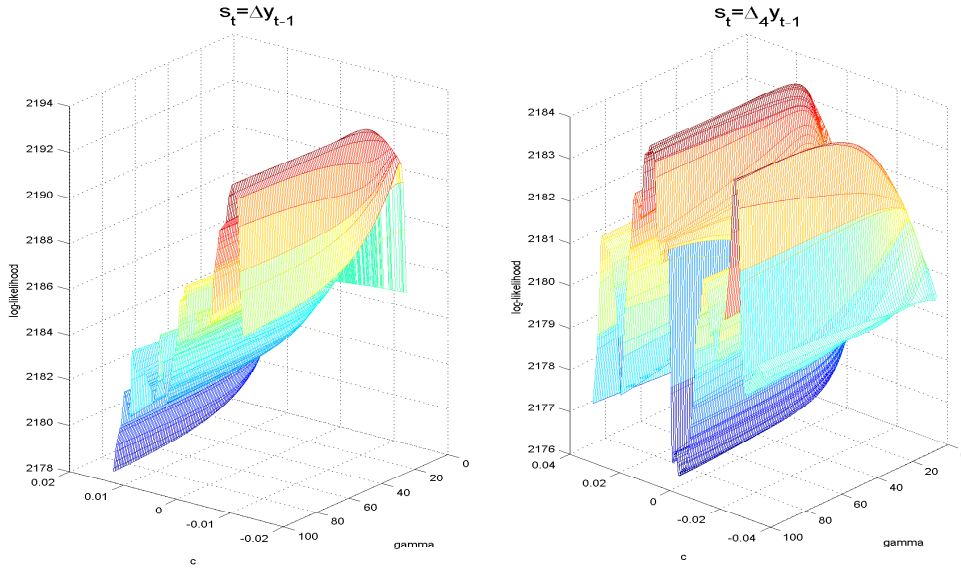


Figure 4.1: Examples of curvature of the profiled log-likelihood functions for combinations of c and γ .

large standard error of $\hat{\gamma}$ is consequently obtained.⁷ It turns out that estimation is greatly facilitated by holding γ fixed in the estimation at the value found in the grid search without significantly altering the result. This is also supported by figure 4.1; the log-likelihood functions are fairly flat in the direction of γ . The procedure implies that the rest of the parameters are estimated conditional on γ . Starting values for the rest of the parameters are obtained from the linear VECM.

4.3. TEST OF LINEARITY

If the observed dynamic relationship between output and the unemployment rate can be modeled adequately by a linear model, then applying a non-linear model is redundant. Testing for linearity is therefore of great importance. The test is, however, non-standard because the parameters of the transition function become nuisance parameters; they are identified under the alternative hypothesis of a STVECM, but not under the null hypothesis of a VECM. This complicates the testing problem with the main consequence that the classical likelihood ratio (LR) and Lagrange multiplier (LM) test statistics do not have the usual asymptotic distributions under the null hypothesis. The unknown distribution of test statistics can, however, be approximated. The most recent contribution within this strand of the literature is by Kristensen and Rahbek (2013), who propose a sup-LR test of

⁷Note that a corresponding small t -value does not suggest redundancy of the non-linear component. The t -value does not have its customary interpretation as a test for the hypothesis $\gamma = 0$, because of the nuisance parameter problem discussed in connection with the test of linearity in section 4.3.

linearity based upon ML estimation and implemented with a wild bootstrap procedure. The test is carried out by the following steps:

1. Estimate the model (3.1) under the null hypothesis of linearity. Save the estimated parameters, $\hat{\theta}_0 = (\hat{\alpha}, \hat{\beta}', \hat{\Gamma}_i, \dots, \hat{\Gamma}_{k-i})$, and the maximized log-likelihood value $L_T(\hat{\theta}_0)$.
2. Estimate the model under the alternative (4.1) by the estimation procedure presented in section 4.2. Save the (recentered) residuals, $\hat{\varepsilon}_t$, and the maximized log-likelihood value $L_T(\hat{\theta})$.
3. Compute the sup-LR test statistic

$$\sup_{c \in C, \gamma \in Y} LR_T = 2 [L_T(\hat{\theta}) - L_T(\hat{\theta}_0)] \quad (4.6)$$

where C and Y are the parameter spaces of c and γ , respectively.

4. Bootstrap the distribution of the test statistic (4.6) by generating $B = 399$ bootstrap samples under the null hypothesis of linearity with the saved estimates from the linear model in step 1, $\hat{\theta}_0$. That is, compute B samples of Z_t^* from

$$\Delta Z_t^* = \hat{\alpha} \hat{\beta}' \bar{Z}_{t-1}^* + \sum_{i=1}^{k-1} \hat{\Gamma}_i \Delta Z_{t-i}^* + \epsilon_t^*, \quad t = 1, 2, \dots, T.$$

where the resampled errors, ϵ_t^* , are generated using a wild bootstrap procedure, originally suggested by Wu (1986): $\epsilon_t^* = \hat{\varepsilon}_t \omega_t$, where ω_t is i.i.d. $N(0, 1)$. Recall that $\hat{\varepsilon}_t$ are the residuals from the non-linear model estimated in step 2. The k initial values equal the k first observations from the original data.

5. For each bootstrap sample, Z_{tb}^* , repeat step 1-3 to obtain a test statistic $\sup LR_{Tb}^*$.
6. Compute critical values of the empirical distribution of $\sup LR_T^*$ and compare with $\sup LR_T$. An approximate p -value is

$$p^*(\sup LR_T) = \frac{1}{B} \sum_{b=1}^B I(\sup LR_{Tb}^* > \sup LR_T)$$

where $I(\cdot)$ denotes the indicator function.

One advantage of such a sup test approach is that linearity is tested against a specified non-linear model and no approximative model is used. In addition, Francq *et al.* (2010) show that for an $AR(1)$ model a sup LR test of linearity is superior in terms of power compared to the frequently applied LM test of Luukkonen *et al.* (1988a;b) which is derived

using a Taylor series expansion of the non-linear part of the model. Another advantage of the sup-LR test is in terms of robustness. Autocorrelation may lead to spurious findings of non-linearity and the same is true for neglected heteroskedasticity. Although heteroskedasticity robust versions of the test by Luukkonen *et al.* (1988a;b) exist, they remove most of the power of the test, cf., Dijk *et al.* (2002). In contrast, the use of the wild bootstrap in the implementation should make the sup-LR test robust to ARCH effects, cf., Kristensen and Rahbek (2013).

5. EMPIRICAL ANALYSIS

The empirical analysis initiates from the VECM in (3.1). It then proceeds with the STVECM in (4.1) with the logistic transition function defined by (4.2). The specification of the STVECM is subsequently discussed before the results are presented and evaluated with respect to the non-linear hypothesis of the Okun's law relationship. The analysis is carried out using Ox and PcGive (see Doornik (2009) and Doornik and Hendry (2009)).

5.1. DATA

The data is quarterly observations of output and the unemployment rate for the U.S. for the period 1948:1-2011:2. Output is measured by the log of GDP in constant 2005-prices and is obtained from Bureau of Economic Analysis (BEA). The unemployment rate is a three month average of monthly observations of the unemployment rate, in which the rate is calculated as the number of unemployed persons as a percentage of the civilian labor force (16 years and over). The resulting series is divided by 100 to match the scaling of the output variable. The monthly series is obtained from the Bureau of Labor Statistics (BLS). The output and unemployment rate series are both seasonally adjusted and illustrated in figure 5.1 (red solid).

The underlying trend in output appears well approximated by a linear deterministic trend in time. A more flexible trend is needed to approximate the trend, or natural rate, in the unemployment rate, and, as discussed in section 3.1, it is approximated by a fourth-order polynomial.

5.2. RESULTS FROM THE LINEAR MODEL - VECM

As a first step in the econometric analysis, the unrestricted VAR model, $H(2)$, is estimated. The model is estimated with the preferred deterministic specification in (3.3). Table A.1

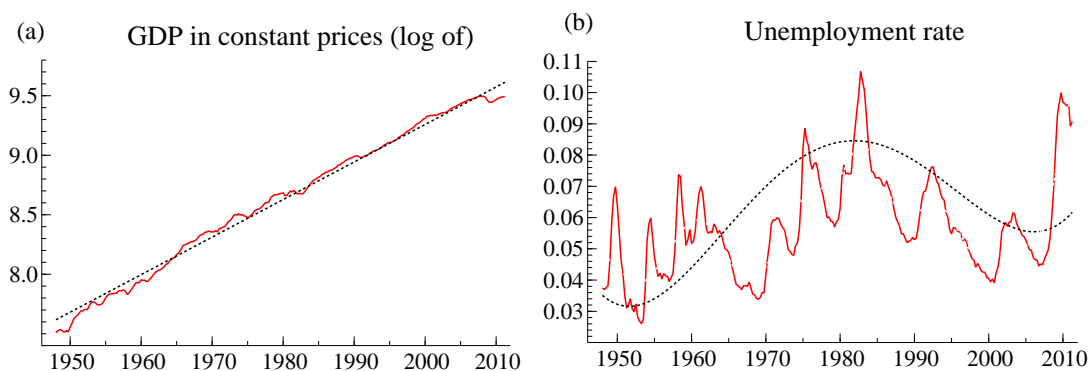


Figure 5.1: (a) Log of GDP in constant prices from BEA (red solid) and the estimated deterministic trend from the reduced rank VECM (black dotted). (b) The unemployment rate from BLS (red solid), the estimated deterministic trend from the reduced rank VECM (black dashed).

and A.2 in appendix A report tests of misspecification and lag-length determination. Information criteria and general-to-specific testing indicate that a lag structure of $k = 3$ is sufficient to capture the time dependence, and, hence, no severe residual autocorrelation is present. The normality assumption of the residuals is rejected because of outliers. Outliers are, however, not removed by including dummy variables, because the non-linear model may be successful in explaining some of these outliers.

The test of cointegration rank, or equivalent the number of unit roots in the data, is based on the VAR model following the Johansen procedure, originating in Johansen (1988). Table 5.1 reports the LR test statistics of the maximum eigenvalue tests and the trace tests for the cointegration rank, $H(r)$. The reduction $H(1)|H(2)$ is easily accepted with a small

Table 5.1: LR tests for rank determination.

	LR_{\max}	5% CV	LR_{trace}	5% CV
$H(0) H(1)$	17.82	26.66		
$H(1) H(2)$	8.36	20.41	8.36	20.41
$H(0) H(2)$			26.17	34.23

Note: The critical values are from simulated asymptotic distributions. The deterministic specification includes a restricted constant in addition to the restricted trends reflecting the linear trend and the fourth-order polynomial in the levels of output and unemployment rate, respectively.

test statistic of 8.36 suggesting that restricting the system to contain one unit root is reasonable. Restricting both variables to be unit root processes, $H(0)|H(2)$, is, however, also accepted. It is not entirely clear how these rank tests perform if the true model includes non-linear error correction of the STAR type. In such case, the linear model reflects the average across the regimes. Moreover, Corradi *et al.* (2000) claims that the trace test in

general has no simple limiting distribution in the presence of neglected non-linearities.⁸ The roots of the companion matrix provide additional information about the unit roots in the data. They are shown in figure B.1 in appendix B together with a plot of the suggested cointegration relation for the system restricted to include one unit root, that is $p - r = 1$. When this (acceptable) restriction is imposed, the second largest root is 0.82. In this area it is hard to discriminate between a unit root and a stationary relation with slow adjustment back to equilibrium. The plot confirms that the mean reversion is quite slow. The combination of the low power of the rank test and the plot, which does not reject the hypothesis of a stationary relation, warrants the choice of $r = 1$.

The reduced rank VECM serves as the benchmark linear model and will be tested against the non-linear model. In the cointegration space, the parameter to the unemployment rate, u_t , is normalized, which is sufficient for identification of the cointegration vector. The results are presented below (t -values in round parenthesis and p -values in square parenthesis)⁹:

$$\begin{aligned}
 \begin{pmatrix} \Delta y_t \\ \Delta u_t \end{pmatrix} &= \begin{pmatrix} 0.124 \\ (1.40) \\ -0.172 \\ (-4.81) \end{pmatrix} \begin{pmatrix} 0.290 & 1 & -2.557 \\ (12.02) & (...) & (-12.30) \end{pmatrix} \begin{pmatrix} y_{t-1} \\ u_{t-1} \\ 1 \end{pmatrix} \\
 &+ \begin{pmatrix} 0.169 & -0.856 \\ (2.03) & (-3.29) \\ -0.096 & 0.483 \\ (-3.90) & (6.28) \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta u_{t-1} \end{pmatrix} \\
 &+ \begin{pmatrix} 0.169 & 0.792 \\ (1.99) & (3.43) \\ -0.088 & -0.220 \\ (-3.51) & (-3.19) \end{pmatrix} \begin{pmatrix} \Delta y_{t-2} \\ \Delta u_{t-2} \end{pmatrix}. \tag{5.1}
 \end{aligned}$$

$$\begin{aligned}
 T = 251 \text{ Loglik} = 2745.72 \quad AR(1) : 5.33[0.25] \quad AR(1-2) : 11.48[0.18] \\
 ARCH(2) : 23.05[0.19] \quad Normality : 26.17[0.00]
 \end{aligned}$$

⁸Nevertheless, the monte carlo study of Corradi *et al.* (2000) reveals that an acceptable level of power of the trace test can be achieved, but only when the linear model includes a constant in the non-stationary direction. Otherwise, the power is low and the test performs poorer with increased complexity of the non-linearity, e.g., when the non-linear component enters more than one of the equations of the system.

⁹The misspecification tests are multivariate LM tests, except the normality test which is the multivariate test of Doornik and Hansen (2008). $AR(1)$ and $AR(1-2)$ are tests of no autocorrelation of order one and of order one and two, respectively. The ARCH test is the test of Lütkepohl and Krätzig (2004) which has power against a broad range of ARCH effects.

As for the unrestricted model, the reduced rank VECM appears well specified. Table C.1 in appendix C shows that a fourth-order polynomial trend is sufficient to capture the deterministic trend in the unemployment rate, interpretable as an estimate of the natural level of the unemployment rate, and pictured in figure 5.1(b) (black dotted). The recent financial crisis has induced a notable deviation from the historical trend growth in output, see figure 5.1(a), and a large and swift increase in the unemployment rate, increasing the natural rate once again after the fall during the 80s and 90s.

5.3. THE RESULTS OF THE NON-LINEAR MODEL - STVECM

The next step in the analysis is the STVECM. As for the VECM, the STVECM is estimated with $k = 3$ VAR lags and $r = 1$. To facilitate estimation, precision of the estimates and reduce the computational time, it turns out to be necessary to reduce the large number of parameters in the STVECM by estimating the deterministic trends in a first step. Consequently, observed data is detrended with the estimated trends from the VECM before estimating the STVECM. Hence, no deterministic trends are estimated in the STVECM.

Before turning to the estimation results, the transition variable, s_t , needs to be determined. Based on the p -values from the linearity test of the STVECM for different candi-

Table 5.2: Log-likelihood value and p -value of the null hypothesis of linearity.

s_t	Δy_{t-1}	Δy_{t-2}	Δy_{t-3}	Δy_{t-4}	$\Delta_4 y_{t-1}$	Δu_{t-1}	Δu_{t-2}	Δu_{t-3}	Δu_{t-4}	$\Delta_4 u_{t-1}$
LogL	2771	2764	2754	2758	2762	2768	2764	2760	2755	2765
p^* -value	0.005	0.113	0.822	0.479	0.105	0.023	0.068	0.607	0.739	0.065

Note: The p^* -value is the wild bootstrap p -value of the null hypothesis of linearity with 399 bootstrap replications. $\Delta_4 x_t = x_t - x_{t-4}$, $x_t = y_{t-1}, u_{t-1}$. The non-standard distribution of the LR test statistic is confirmed by plots of some of the bootstrap distributions, see figure D.1 in appendix D. They are shifted considerable to the right and with a much fatter right tail than the χ^2 -distribution, which under normal circumstances is the asymptotic distribution of the LR test statistic.

dates of s_t , reported in table 5.2, misspecification tests and inspection of the found optimum for each model, $s_t = \Delta y_{t-1}$.

The exogenous given NBER business cycle indicator was considered as transition variable. However, as noted by Camacho and Perez Quiros (2007), such choice of s_t would create a potential endogeneity problem because the NBER indicator has been constructed on the basis of knowing the actual value of output growth. More importantly, this model would fail to capture the fact that the economy can recover on its own since the timing of the regime shifts depends entirely on the NBER indicator, which has been exogenously defined.

The estimation results and misspecification tests for the model with $s_t = \Delta y_{t-1}$ are reported in table 5.3 and 5.4, respectively. The usual LM test for residual autocorrelation is modified because the gradient vector must be augmented with elements related to the non-linear part of the model. The test is derived for the current model in appendix E. In contrast, the usual LM test of ARCH effects is unchanged from the VECM since the information matrix is still block diagonal. The same holds true for the normality test of the residuals, cf., Teräsvirta (1998). The model appears reasonable well specified, although, as for the linear model, the normality assumption of the errors is violated.

Table 5.3: Estimated STVECM with $s_t = \Delta y_{t-1}$.

	Regime 1		Regime 2	
	Δy_t	Δu_t	Δy_t	Δu_t
$\hat{\beta}' \bar{Z}_{t-1}$	–	–0.153 (–1.89)	–	–0.099 (–4.93)
Δy_{t-1}	0.350 (2.44)	–0.181 (–4.06)	0.179 (1.84)	–0.064 (–2.35)
Δu_{t-1}	–1.159 (–2.13)	0.596 (4.04)	–1.020 (–3.03)	0.390 (4.61)
Δy_{t-2}	0.511 (1.71)	–0.250 (–3.04)	–	–0.050 (–2.49)
Δu_{t-2}	3.093 (3.49)	–0.844 (–3.67)	0.463 (2.44)	–0.110 (–1.42)
$\hat{\beta}' \bar{Z}_{t-1}$	$\left(\begin{array}{l} 0.277 y_{t-1} + u_{t-1} - 2.437 \\ (12.95) \quad \quad \quad (-13.27) \end{array} \right)$			
\hat{c}	–0.010 (–99.05)			
$\hat{\gamma}$	100			
<i>EV</i>	1.10 (<i>i</i>)		0.87 (<i>r</i>)	
<i>Obs</i>	31(12%)		220(88%)	
<i>T</i>	251			
<i>LogL</i>	2767.95			

Note: The restrictions are accepted with a p -value of 0.16. Due to signs of ARCH effects, the heteroskedasticity-consistent standard errors of White (1982) are used. γ is fixed at the value found in the grid search for initial values, as explained in section 4.2, and thus no standard error of $\hat{\gamma}$ is computed. *EV* is the modulus of the largest unrestricted eigenvalue present in each extreme regimes. (*i*) and (*r*) refer to imaginary and real root, respectively.

Table 5.4: Misspecification tests of the STVECM.

	Δy_t	Δu_t	Vector
<i>AR</i> (1)	2.49[0.18]	1.01[0.34]	9.07[0.08]
<i>AR</i> (1 – 2)	3.56[0.45]	3.97[0.28]	15.51[0.06]
<i>AR</i> (1 – 4)	1.96[0.98]	2.40[0.88]	21.21[0.27]
<i>ARCH</i> (2)	4.90[0.09]	6.71[0.03]	19.77[0.34]
<i>Normality</i>	11.91[0.00]	11.17[0.00]	20.69[0.00]

Note: The *AR* tests are the robust version of single-equation *F*-tests of autocorrelation in STAR models, originally derived by Eitrheim and Teräsvirta (1996). The vector test is from Camacho (2004) and applied with β fixed at $\hat{\beta}$. The tests of ARCH and normality of the errors are unchanged from VECM.

Comparing the size of the estimated parameters with those from the VECM in (5.1), regime 2 is similar to the VECM. About 88% of the quarterly observations give rise to a larger weight of regime 2 ($y_{t-1} > \hat{c}$) and the other 12% to a larger weight of regime 1 ($y_{t-1} < \hat{c}$). Figure 5.2 presents different charts related to the estimated model. The relative few observations belonging to regime 1 are highlighted by figure 5.2(a). The value of \hat{c} corresponds to an approximative quarterly growth rate in output of -0.26% . Hence, when output shrinks by more than 0.26% in the previous quarter regime 1 dominates. Economically, regime 1 can thus be interpreted as a recession regime whereas regime 2 represents a normal time or expansion regime. Such an interpretation is confirmed by figure 5.2(b) which depicts \hat{G}_t and NBER defined business cycles.¹⁰ The large estimate of $\hat{\gamma}$ indicates a very fast speed of regime switching, which figure 5.2(c) also highlights by the great steepness of \hat{G}_t . As a result, the estimated STVECM approximates a threshold model.

The non-linearity is present in every term of the STVECM. The non-linearity may, however, only be relevant for the error correction mechanism or one of the equations of the system. Testing the extent of the non-linearity is fairly simple since such restricted models can be shown to be nested without the presence of nuisance parameters. The null hypothesis of non-linearity only in the error correction mechanism is rejected with a p -value of 0.00. To obtain an econometric model closer to the original Okun's law equation, the non-linear dynamics is restricted to the unemployment rate equation, while the dynamics of output growth are linear with no error correction. Such a hypothesis is rejected with a p -value of 0.016.

The largest unrestricted eigenvalue of the companion matrix for each regime provides information about the local dynamics of the model and are reported in table 5.3. Regime

¹⁰The time series of NBER indicator is obtained from www.nber.org.

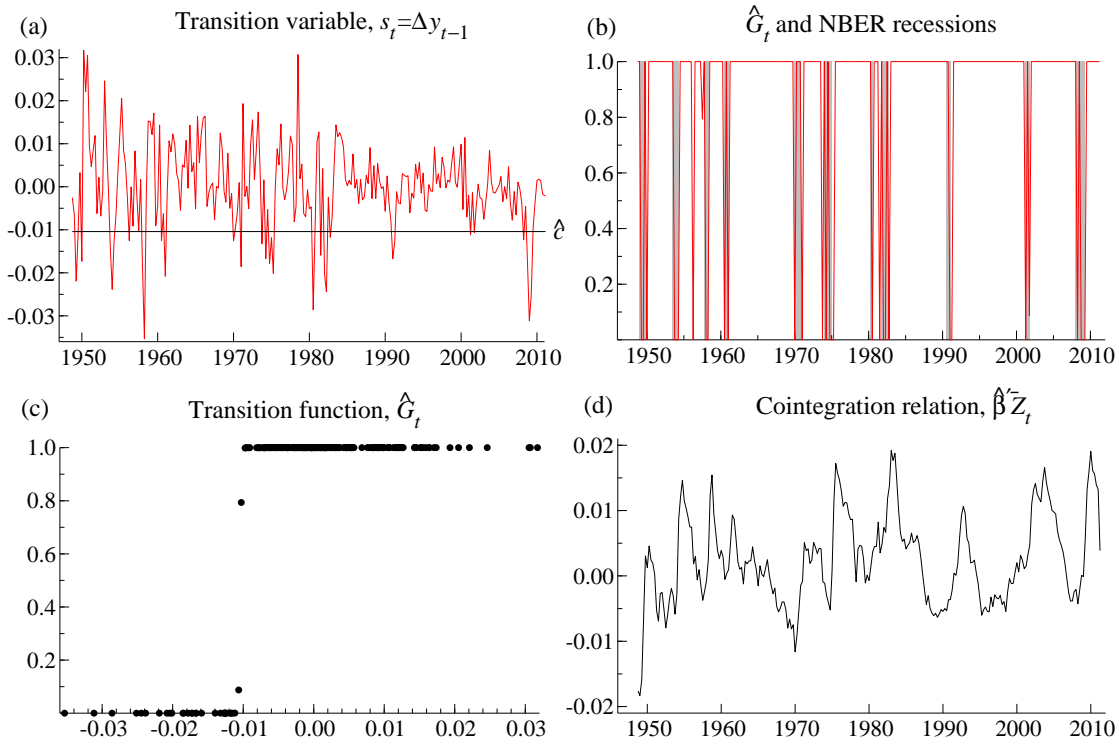


Figure 5.2: All charts are from the restricted STVECM with transition variable $s_t = \Delta y_{t-1}$. (a) The level of the transition variable (red solid) and the estimated threshold parameter \hat{c} (black solid). (b) The estimated transition function, \hat{G}_t , plotted against time (red solid) and NBER recessions (gray shade). (c) The estimated transition function, \hat{G}_t , plotted against s_t . Each dot represents an observation. (d) The estimated cointegration relation $\hat{\beta}' \hat{Z}_{t-1}$.

1 is characterized by an explosive root whereas regime 2 is stationary. The explosiveness of regime 1 is the means by which the model explains the observed large and abrupt increases in the unemployment rate stemming from the flexible U.S. labor market. This is the main source of non-linearity in the model. Importantly, it never degenerates because the model stays only in regime 1 in a maximum of three consecutive quarters, as seen from 5.2(b).

Although such dynamics appear strange, it is not an unusual finding in univariate STAR models, see, e.g., Teräsvirta and Anderson (1992), Teräsvirta (1995) and Teräsvirta *et al.* (2011). The stability of the estimated STVECM can be checked by means of extrapolation of the process.¹¹ Examples of such extrapolation are showed in figure F1 in appendix F. The two processes converge to constant levels, irrespectively of the initial regime, indicating a stable model. For this particular STVECM it is, however, not trivial to show the

¹¹According to Teräsvirta *et al.* (2011), “a necessary condition for stability is that extrapolation of the process after switching of the noise should lead to convergence” (p. 388).

conditions necessary for global stability given the explosive regime 1 and is an area for future research.^{12,13}

The cointegration coefficient to output of -0.277 reflects an estimate of Okun's coefficient. A 1 percentage point output gap is associated with a decrease in the unemployment rate gap of 0.28 percentage point. The estimate is thus close to the $-1/3$ rule of thumb benchmark originally suggested by Okun (1962). Figure 5.2(d) displays the deviations from this long-run relationship over time. While large increases are related to recession periods, the relationship initiates at a all time low in the start of the sample. The estimated cointegration vector of the STVECM is, as expected, similar to that of the VECM, and may be caused by the necessity of prior estimation of the trends. Yet, this exercise shows that in a model with non-linear short-run dynamics, a linear cointegration vector, which is constant across regimes, can be rather precisely estimated simultaneously with the short-run parameters. The constant term has no economic interpretation as it reflects the estimated long run difference between the (unidentified) constant levels of the two observed variables.

The error correction parameters provide information about how the long-run Okun's law relationship is sustained. The observed slow mean reversion in figure 5.2(d) is confirmed by the (numerical) small error correction estimates (coefficients in the top row of table 5.3). Moreover, the output variable is weakly exogenous.¹⁴ This implies that the output variable does not adjust to deviations from the long-run equilibrium. Moreover, the accumulated innovations of the output equation can be considered a common driving trend of the system, and thus shocks to the unemployment rate have no permanent effect on output. Although the dynamics of the regimes differ, the asymmetry of the error correction of the unemployment rate is insignificant.

¹²Seo (2011) develops asymptotic theory for a class of non-linear VECMs, including the TVECM and the STVECM, but with the regime switching depending on the previous period equilibrium error. In the related work of Kristensen and Rahbek (2010; 2013), the previous period equilibrium error is also governing the regime shifts, and in Gonzalo and Pitarakis (2006) an exogenous stationary variable determines the regime switching. The theory for the model of this paper with the lag of an endogenous variables as the transition variable, has, to the best of knowledge of the author, yet to be derived.

¹³This implies that the asymptotic distribution of the parameters is unknown and may be non-standard. Hence, the t -values reported in table 5.3 are not necessarily asymptotically normal distributed and any inference is only indicative.

¹⁴If the hypothesis of weakly exogenous output is tested alone, the null hypothesis is accepted with a p -value of 0.25

6. SUMMARY AND CONCLUSION

This paper estimates a non-linear version of the Okun's law relationship between estimated gaps of the unemployment rate and output by means of a smooth transition cointegration model called a STVECM. Okun's law is defined as a linear cointegration relationship, reflecting the interdependence of the two variables over the longer term, and with any non-linearity prevailing in the short-run dynamics. The result of the sup-LR test of linearity shows that linearity, and hence symmetry, of the short-run dynamics is rejected when tested against the STVECM with the chosen transition variable $s_t = \Delta y_{t-1}$. The regimes of the STVECM overlap with those of the NBER business cycle indicator and can be interpreted as a recession regime and an expansion regime. The estimated threshold implies that when output falls by more than 0.26% in the previous quarter, the model enters the recessionary regime 1. The estimated speed of the regime switching is fast and, thus, the model effectively approximates a threshold model. The dynamics of the two regimes differ significantly and is caused by the large increases in the unemployment rate during recessions which the model approximates by an explosive root. However, the stability of the model appears unaffected since the model returns to the stationary expansion regime shortly after a recession. The error correction mechanism reveals that the output variable does not significantly adjust to deviations from the equilibrium Okun's law relationship, and shocks to the unemployment rate, therefore, have no permanent effect of output. Finally, the estimate of Okun's coefficient is -0.28 . Hence, although five decades have passed, the structure of the economy has changed and a different econometric model is applied, the estimate is still close to the $-1/3$ rule of thumb benchmark originally suggested by Okun (1962).

REFERENCES

- ATTFIELD, C. L. F. AND B. SILVERSTONE (1998): "Okun's Law, Cointegration and Gap Variables." *Journal of Macroeconomics*, 20(3):625–637.
- BALKE, N. S. AND T. B. FOMBY (1997): "Threshold Cointegration." *International Economic Review*, 38(3):627–45.
- BALL, L., D. LEIGH, AND P. LOUNGANI (2013): "Okun's Law: Fit at Fifty?" *National Bureau of Economic Research*.
- CAMACHO, M. (2004): "Vector smooth transition regression models for US GDP and the composite index of leading indicators." *Journal of Forecasting*, 23(3):173–196.
- CAMACHO, M. AND G. PEREZ QUIROS (2007): "Jump-and-Rest Effect of U.S. Business Cycles." *Studies in Nonlinear Dynamics and Econometrics*, 11(4).
- CHRISTENSEN, A. M. AND H. B. NIELSEN (2009): "Monetary Policy in the Greenspan Era: A Time Series Analysis of Rules vs. Discretion." *Oxford Bulletin of Economics and Statistics*, 71:69–89.
- CORRADI, V., N. R. SWANSON, AND H. WHITE (2000): "Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes." *Journal of Econometrics*, 96(1):39–73.
- COURANT, R. AND D. HILBERT (2008): *Methods of mathematical physics*, volume 1. Wiley.com.
- CUARESMA, J. C. (2003): "Okun's Law Revisited." *Oxford Bulletin of Economics and Statistics*, 65(4):439–451.
- DE JONG, R. (2001): "Nonlinear estimation using estimated cointegrating relations." *Journal of econometrics*, 101(1):109–122.
- DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth transition autoregressive models - a survey of recent developments." *Economic Reviews*, 21:1–47.
- DOORNIK, J. (2009): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, London, 6th edition.
- DOORNIK, J. AND D. HENDRY (2009): *PcGive 13 Empirical Econometric Modelling*, volume I-III. Timberlake Consultants Press, London.

- DOORNIK, J. A. AND H. HANSEN (2008): "An Omnibus Test for Univariate and Multivariate Normality." *Working Paper, Nuffield College, Oxford*.
- EITRHEIM, O. AND T. TERÄSVIRTA (1996): "Testing the adequacy of smooth transition autoregressive models." *Journal of Econometrics*, 74(1):59–75.
- FRANCQ, C., L. HORVATH, AND J.-M. ZAKOÏAN (2010): "Sup-test for linearity in a general nonlinear AR(1) model." *Econometric Theory*, 26(04):965–993.
- FRANK, A. (2008): "Asymmetry in Okun's law." *Diplomarbeit, University of Wien*.
- GONZALO, J. AND J. PITARAKIS (2006): "Threshold Effects in Cointegrating Relationships*." *Oxford Bulletin of Economics and Statistics*, 68:813–833.
- GORDON, R. J. (1984): "Unemployment and Potential Output in the 1980s." *Brookings Papers on Economic Activity*, 2:537–564.
- JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control*, 12(2-3):231–254.
- (1996): "Likelihood-Based Inference in Cointegrated Vector Autoregressive Model." *Advanced Texts in Econometrics, Oxford University Press*.
- JUSELIUS, K. (2006): "The Cointegrated VAR Model: Methodology and Applications." *Oxford University Press*.
- KAUFMAN, R. T. (1988): "An international comparison of Okun's laws." *Journal of Comparative Economics*, 12(2):182–203.
- KRISTENSEN, D. AND A. RAHBK (2010): "Likelihood-based inference for cointegration with nonlinear error-correction." *Journal of Econometrics*, 158(1):78–94.
- (2013): "Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models." *Econometric Theory*, 29:1238–1288.
- LEE, J. (2000): "The Robustness of Okun's Law: Evidence from OECD Countries." *Journal of Macroeconomics*, 22(2):331–356.
- LÜTKEPOHL, H. AND M. KRÄTZIG (2004): *Applied Time Series Econometrics*. Cambridge University Press.

- LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988a): “Testing linearity against smooth transition autoregressive models.” *Biometrika*, 75:491–499.
- (1988b): “Testing linearity in univariate time series models.” *Scandinavian Journal of Statistics*, 16:161–175.
- MCKAY, A. AND R. REIS (2008): “The brevity and violence of contractions and expansions.” *Journal of Monetary Economics*, 55(4):738–751.
- MENDONCA, G. P. (2008): “Structural Breaks, Regime Change and Asymmetric Adjustment: A Short and Long Run Global Approach to the Output/Unemployment Dynamics.” *University Library of Munich, Germany, MPRA Paper*.
- MONTGOMERY, A. L., V. ZARNOWITZ, R. S. TSAY, AND G. C. TIAO (1998): “Forecasting the US unemployment rate.” *Journal of the American Statistical Association*, 93(442):478–493.
- MOOSA, I. A. (1997): “A Cross-Country Comparison of Okun’s Coefficient.” *Journal of Comparative Economics*, 24:335–356.
- (1999): “Cyclical output, cyclical unemployment, and Okun’s coefficient. A structural time series approach.” *International Review of Economics and Finance*, 8.
- NEDELJKOVIC, M. (2008): “Testing for smooth transition nonlinearity in adjustments of cointegrating systems.” *Working paper, University of Warwick, Department of Economics*.
- NEFTCI, S. N. (1984): “Are Economic Time Series Asymmetric over the Business Cycle?” *Journal of Political Economy*, 92(2):307–28.
- OKUN, A. (1962): “Potential GNP: Its measurement and significance.” *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pages 89–104.
- PRACHOWNY, M. F. J. (1993): “Okun’s Law: Theoretical Foundations and Revised Estimates.” *The Review of Economics and Statistics*, 75:331–36.
- ROTHMAN, P., D. VAN DIJK, AND P. H. FRANSES (2001): “Multivariate STAR Analysis of Money–Output Relationship.” *Macroeconomic Dynamics*, 5(04):506–532.
- SEO, M. (2011): “Estimation of nonlinear error correction models.” *Econometric Theory*, 27(02):201–234.

- SILVAPULLE, P., I. MOOSA, AND M. J. SILVAPULLE (2004): "Asymmetry in Okun's Law." *Canadian Journal of Economics*, 4:353–374.
- SILVERSTONE, B. AND R. HARRIS (2001): "Testing for asymmetry in Okun's law: A cross-country comparison." *Economics Bulletin*, 5(2):1–13.
- SOGNER, L. AND A. STIASSNY (2002): "An Analysis on the Structural Stability of Okun's Law, A Cross-Country Study." *Applied Economics*, 34(14):1775–87.
- TERÄSVIRTA, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models." *Journal of American Statistical Association*, 89:208–218.
- (1995): "Modelling nonlinearity in U.S. Gross national product 1889-1987." *Empirical Economics*, 20:577–597.
- (1998): "Modelling economic relationships with smooth transition regressions." *Handbook of Applied Economic Statistics*, Marcel Dekker: New York, pages 507–552.
- TERÄSVIRTA, T. AND H. M. ANDERSON (1992): "Characterizing nonlinearities in business cycles using smooth transition autoregressive models." *Journal of Applied Econometrics*, 7(S1):S119–S136.
- TERÄSVIRTA, T., D. TJØSTHEIM, AND C. GRANGER (2011): *Modelling Nonlinear Economic Time Series*. Advanced Texts in Econometrics Series. Oxford University Press.
- VIREN, M. (2001): "The Okun curve is non-linear." *Economics Letters*, 70(2):253–257.
- WEBER, C. E. (1995): "Cyclical Output, Cyclical Unemployment, and Okun's Coefficient: A New Approach." *Journal of Applied Econometrics*, 10:433–445.
- WEISE, C. (1999): "The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach." *Journal of Money, Credit & Banking*, 31(1):85–86.
- WHITE, H. (1982): "Maximum likelihood estimation of misspecified models." *Econometrica: Journal of the Econometric Society*, pages 1–25.
- WU, C. (1986): "Jackknife, bootstrap and other resampling methods in regression analysis." *The Annals of Statistics*, 14(4):1261–1295.

A. MISSPECIFICATION TESTS OF THE UNRESTRICTED VAR

Table A.1: Misspecification tests of the unrestricted VAR.

	Δy_t	Δu_t	Vector
<i>AR</i> (1)	0.55[0.46]	1.35[0.25]	5.95[0.20]
<i>AR</i> (1 – 2)	0.73[0.69]	1.53[0.47]	10.87[0.21]
<i>ARCH</i> (2)	4.87[0.09]	4.01[0.13]	22.36[0.22]
<i>Normality</i>	21.86[0.00]	13.26[0.00]	28.08[0.00]

Note: The misspecification tests are LM tests and asymptotically χ^2 -distributed. The normality test is the test of Doornik and Hansen (2008). *AR*(1) and *AR*(1 – 2) are tests of autocorrelation of order one and order one through two, respectively. The ARCH test is the test of Lütkepohl and Krätzig (2004) which has power against a broad range of ARCH effects.

Table A.2: Information criteria and LR tests of lag-length.

	$k = 2$	$k = 3$	$k = 4$
<i>SC</i>	–21.634	–21.613	–21.547
<i>H – Q</i>	–21.718	–21.733	–21.699
<i>LR</i> ($k = 4 k = 3$)		5.13[0.28]	
<i>LR</i> ($k = 3 k = 2$)		17.52[0.00]	

Note: SC and H-Q are the Schwartz and the Hannan-Quinn information criterion, respectively. The p -value of the LR tests are from the $\chi^2(4)$ -distribution. The models are estimated with the same number of observations.

B. RANK DETERMINATION

Figure B.1 shows the eigenvalues of the companion matrix and the suggested cointegration relation for the reduced rank VECM.

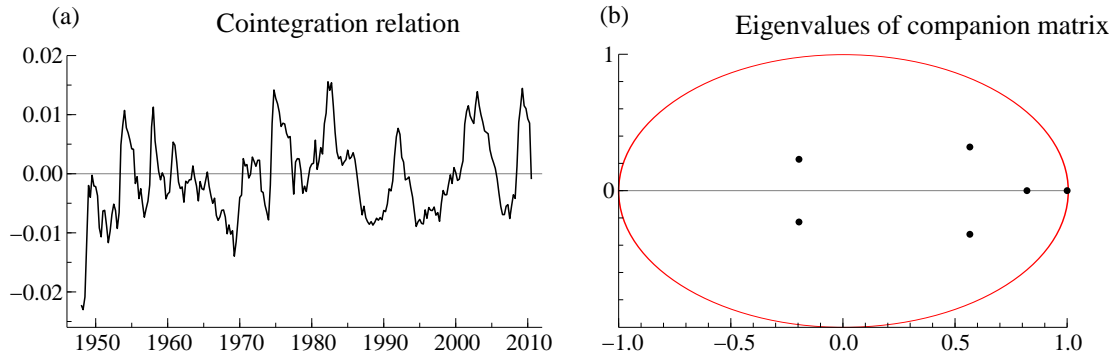


Figure B.1: Cointegration relation and eigenvalues of the VECM with a rank of 1.

C. SHAPE OF POLYNOMIAL TREND OF THE UNEMPLOYMENT RATE

It is not possible to tell, *ex ante*, whether, for instance, a fourth-order polynomial is sufficient to capture the deterministic trend of the unemployment rate or perhaps a fifth-order polynomial is required. Such hypotheses are, however, easily tested by, respectively, imposing the restriction $\varphi_{24} = 0$ and extent φD_t to a fifth-order polynomial in (3.3). The result of LR tests of these hypotheses for the reduced rank VECM are reported in table C.1.

Table C.1: LR test of hypotheses on the shape of the deterministic trend of the unemployment rate, reduced rank VECM.

	LR_T	p -value
$\theta_{24} = 0$	8.21	0.00
$\theta_{25} = 0$	2.53	0.11

Note: The p -value is from a χ^2 -distribution with one degree of freedom. Alternatively, the hypothesis can be tested by t -tests.

D. DISTRIBUTION OF THE SUP-LR TEST STATISTIC

Figure D.1 depicts the wild bootstrap distributions of sup-LR test statistics for four different choices of transition variable. For comparison, the χ^2 -distribution with 12 degrees of freedom is plotted as well. 12 reflects the number of additional parameters in the STVECM compared to the VECM (8 additional lag parameters, 2 additional error correction parameters and 2 transition parameters).

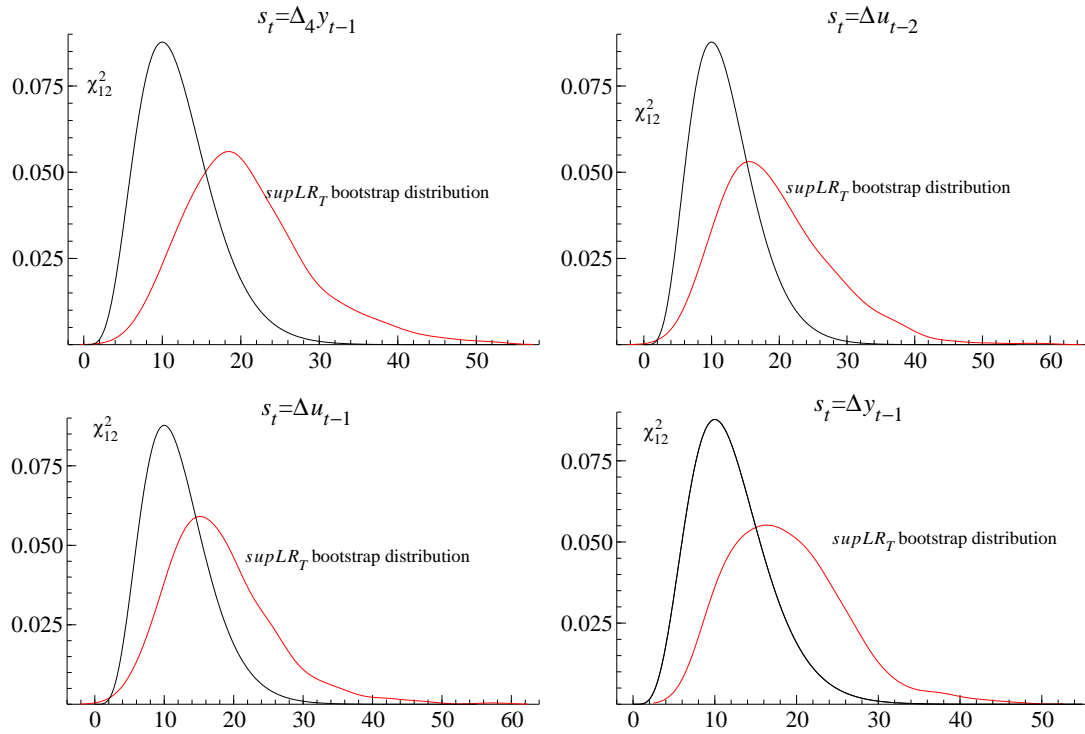


Figure D.1: Distribution of sup-LR test statistics of a null hypothesis of linearity.

E. TEST FOR RESIDUAL AUTOCORRELATION

Eitrheim and Teräsvirta (1996) propose a modified LM test of residual autocorrelation in a univariate STAR model, claiming that the customary Ljung-Box test for residual autocorrelation is inapplicable when applied to STAR residuals. In the framework of the STVECM, the univariate tests are derived below and applied assuming the estimated cointegration vector, $\hat{\beta}$, is a fixed parameter, and thus $ecm_t = \hat{\beta}' \bar{Z}_t$ is perceived as a regressor. The applied vector test is derived in Camacho (2004), also assuming $\hat{\beta}$ fixed.

First, the estimated STVECM in (4.1) is slightly reformulated in compact form as

$$\Delta Z_t := H(W_t, \Phi) + \varepsilon_t = \vartheta' W_t + \delta' W_t G(s_t; \gamma, c) + \varepsilon_t,$$

where $W_t = (ecm_{t-1}, \Delta z_{t-1}, \dots, \Delta z_{t-k-1})'$ for $z_t = y_t, u_t$. $\Phi = (\vartheta, \delta, \gamma, c)'$. The test of no autocorrelation of order q in the STVECM residuals, $\hat{\varepsilon}_{j,t}$, with $j = \Delta y_t, \Delta u_t$ and referring to an equation in the system, is then carried out by running the auxiliary regression:

$$\hat{\varepsilon}_{j,t} = \kappa + \pi_1 \hat{\varepsilon}_{j,t-1} + \pi_2 \hat{\varepsilon}_{j,t-2} + \dots + \pi_q \hat{\varepsilon}_{j,t-q} + \lambda' \hat{V}_t + u_{j,t},$$

where κ is a constant and λ is a $((2k+2) \times 1)$ vector of coefficients to the gradient \hat{V}_t given by

$$\begin{aligned} \hat{V}_t &= \left. \frac{\partial H(W_t, \Phi)}{\partial \Phi} \right|_{\Phi=\hat{\Phi}} \\ &= \left(\frac{\partial H}{\partial \vartheta'}, \frac{\partial H}{\partial \delta'}, \frac{\partial H}{\partial \gamma}, \frac{\partial H}{\partial c} \right)' \Big|_{\Phi=\hat{\Phi}} = (W_t, W_t G(s_t; \hat{\gamma}, \hat{c}), \hat{h}_t^\gamma, \hat{h}_t^c)'. \end{aligned} \quad (\text{E.1})$$

The derivatives \hat{h}_t^γ and \hat{h}_t^c with respect to γ and c , respectively, are evaluated in the ML estimates:

$$\begin{aligned} \hat{h}_t^\gamma &= \frac{\partial G_t}{\partial \gamma} \hat{\delta}' W_t = (1 + \exp\{-\hat{\gamma}(s_t - \hat{c})/\hat{\sigma}_s\})^{-2} \exp\{-\hat{\gamma}(s_t - \hat{c})/\hat{\sigma}_s\} ((s_t - \hat{c})/\hat{\sigma}_s) \hat{\delta}' W_t \\ &= \hat{G}_t (1 - \hat{G}_t) ((s_t - \hat{c})/\hat{\sigma}_s) \hat{\delta}' W_t \\ \hat{h}_t^c &= \frac{\partial G_t}{\partial c} \hat{\delta}' W_t = -(1 + \exp\{-\hat{\gamma}(s_t - \hat{c})/\hat{\sigma}_s\})^{-2} \exp\{-\hat{\gamma}(s_t - \hat{c})/\hat{\sigma}_s\} \hat{\gamma} \hat{\delta}' W_t \\ &= -\hat{G}_t (1 - \hat{G}_t) \hat{\gamma} \hat{\delta}' W_t \end{aligned}$$

The test static is then computed as

$$F_{j,LM} = \frac{(SSR_0 - SSR)/q}{SSR/(T - n - q)}, \quad SSR_0 = \sum_{t=1}^T \hat{\varepsilon}_{j,t}^2 \quad \text{and} \quad SSR = \sum_{t=1}^T \hat{u}_{j,t}^2$$

which approximately follows an F -distribution with $(q, T - n - q)$ degrees of freedom under the null hypothesis of no autocorrelation in $\varepsilon_{j,t}$. The dimension of the gradient vector (E.1) is $n = ((2k + 2) \times 1)$. According to Eitrheim and Teräsvirta (1996), the F -version of the test is preferred to the χ^2 -version which suffers from size problems in small samples. Note that if $\hat{\gamma}$ is large, the derivatives of the transition function are close to zero and \hat{V}_t becomes singular. In such cases, \hat{h}_t^γ and \hat{h}_t^c are dropped from \hat{V}_t .

F. EXTRAPOLATION EXERCISE

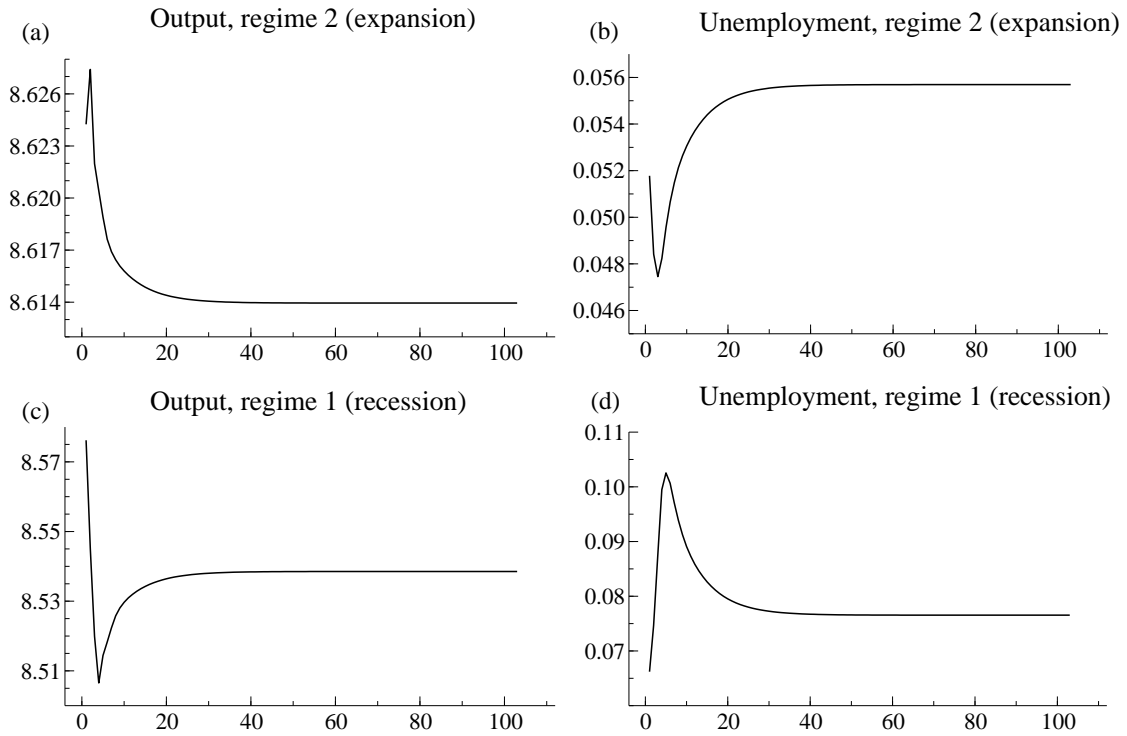


Figure F.1: Chart (a)-(b) display extrapolation of the joint process initiating from observations in the expansionary regime 2. Chart (c)-(d) display extrapolation of the joint process initiating from observations in the recessionary regime 1.

Chapter 2

Smooth or Non-Smooth Regime Switching Models

Smooth or Non-Smooth Regime Switching Models*

Line Elvstrøm Ekner[†] and Emil Nejstgaard[‡]

Department of Economics, University of Copenhagen

ABSTRACT: We propose a reparametrization of the logistic smooth transition autoregressive (LSTAR) model which facilitates identification and estimation of the so-called *speed of transition* parameter. We show that all derivatives of the likelihood function are approaching zero as the parameter measuring the speed of transition increases, and, hence, the threshold autoregressive (TAR) model always represents at least a local maximum of the LSTAR likelihood function. We propose to use information criteria for the choice between the two models and show the effectiveness of this procedure by means of simulations. Two empirical applications illustrate the usefulness of our findings.

1. INTRODUCTION

Regime switching models have become increasingly popular in the time series literature over the last decades and applied to data from potential regime switching processes such as, e.g., the business cycle, the unemployment rate, exchange rates, prices, interest rates, etc. The majority of the models initiate from the threshold autoregressive (TAR) model first presented by Tong and Lim (1980). Nevertheless, the idea of smooth regime switching was discussed by Bacon and Watts (1971), but not formalized in terms of a time series model until Chan and Tong (1986) proposed what they called a smoothed threshold autoregressive model as an extension to the TAR model of Tong and Lim (1980). Heavily cited contributions by Luukkonen *et al.* (1988) and Teräsvirta (1994) changed the label from “smoothed threshold” to “smooth transition” resulting in the label smooth transition autoregression (STAR) used today. For an overview of the TAR and STAR literature, see Tong (2011), Teräsvirta *et al.* (2010), and van Dijk *et al.* (2002).

*We thank Anders Rahbek, Heino Bohn Nielsen, Andreas Noack Jensen, Rasmus Søndergaard Pedersen for useful comments and Søren Johansen for valuable discussions.

[†]Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark, line.elvstrom.ekner@econ.ku.dk

[‡]Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark, emil.nejstgaard@econ.ku.dk

The logistic STAR (LSTAR) model differs from the TAR model by having smooth regime switches over time parametrized by the *speed of transition* parameter. The switches in the TAR model are in contrast discontinuous. The primary economic motivation for the LSTAR model is that economic time series are often results of decisions made by a large number of economic agents. Even if agents are assumed to make only dichotomous decisions or change their behavior discretely, it is unlikely that they do so simultaneously. Hence, any regime switching in economic time series may be more accurately described as taking place smoothly over time. Moreover, the speed of the regime switching can be of separate interest to an economist, e.g., to analyze how fast the economy adapts to another regime or state of the economy. In empirical applications it is, however, often very difficult to identify the speed of transition parameter, and, thus, it is important to test whether this additional parameter of the LSTAR model is at all relevant compared to a TAR model.

The first contribution of this paper is a reparametrization of the LSTAR model. In terms of the proposed parametrization we can explicitly illustrate the problem of distinguishing LSTAR and TAR alternatives.

The second contribution is to show that this distinction is complicated both theoretically and in terms of numerical optimization by the fact that all derivatives of the LSTAR likelihood function are approaching zero with faster speed of transition, i.e., when the LSTAR model approaches the TAR model. Using likelihood analysis, we study the consequences of this identification problem for estimation and inference in the LSTAR model. The new parametrization avoids some of the numerical difficulties that arise when applying the original LSTAR parametrization, and, moreover, clarifies that the LSTAR likelihood function can have a maximum corresponding to a TAR model. In the literature of LSTAR models economic theory is used as the only motivation for modeling an LSTAR model instead of a TAR model, see, e.g., Granger and Teräsvirta (1993) and Teräsvirta (1998). However, our new parametrization facilitates a decision based upon the data, possibly in conjunction with economic theory. We show how information criteria provide a neat, but conservative, tool to select an LSTAR model over a TAR model that can be applied if the researcher wishes to comment on the speed of transition.

The related issue of selecting between an LSTAR model and an autoregressive (AR) model is not treated in this paper. Although testing such hypothesis of linearity is non-standard, procedures are available and well-described in the literature of both the (L)STAR and TAR models, see Davies (1987), Luukkonen *et al.* (1988), Hansen (1996), and Kristensen and Rahbek (2013).

Furthermore, we discuss numerical optimization of the LSTAR likelihood function. In particular, we consider the origin of multiple maxima on the likelihood function and the

use of grid search methods. Finally, we illustrate the benefits of the new parametrization and the model selection procedure with data from two published applications. In the first application, the likelihood function for the reparametrized speed of transition parameter reveals that the published result is only a local maximum on the likelihood function, and that the global maximum is a TAR model. In the second example, data contains insufficient information about the speed of transition parameter which then becomes irrelevant, and, as a result, information criteria prefer a TAR model over the published LSTAR model.

The new parametrization can be applied to all kinds of regime switching models where the regime switching is governed by one or more logistic type transition functions. Identification of the speed of transition parameter in the related exponential STAR (ESTAR) model with an exponential transition function has recently been studied by Heinen *et al.* (2012). However, the problem is different in the ESTAR model since this model approaches an AR model when the speed of transition approaches infinity and not a TAR model. Hence, their results do not carry over to the LSTAR model. Nevertheless, the new parametrization can also be beneficial for estimation of the ESTAR model by facilitating numerical optimization as well as identification of the global maximum of the likelihood function.

2. THE MODEL AND THE IDENTIFICATION PROBLEM

To fix ideas, consider a simple LSTAR model of order one for $y_t \in \mathbb{R}$, cf., Teräsvirta (1994),

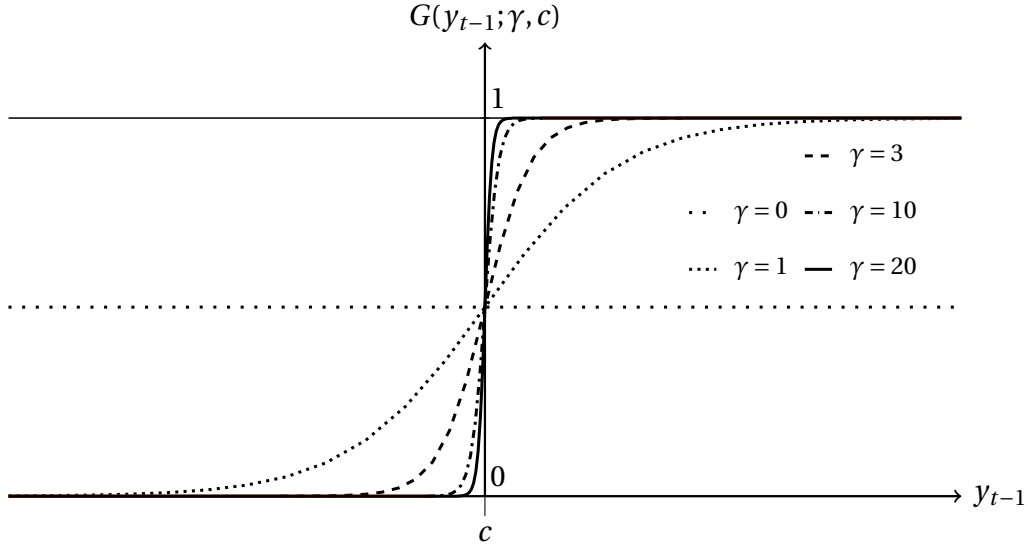
$$y_t = \alpha y_{t-1} G_t + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (2.1)$$

with $\varepsilon_t \sim i.i.N(0, \sigma^2)$ and where G_t is the logistic transition function given by

$$G_t := G(y_{t-1}; \gamma, c) = \left(1 + \exp\{-\gamma(y_{t-1} - c)\}\right)^{-1}. \quad (2.2)$$

The AR parameter is α , γ is the speed of transition parameter and c is the threshold parameter. While $|\alpha| < 1$ and $c \in \mathbb{R}$, we assume that $\gamma \in \bar{\mathbb{R}}_+$ where $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \infty$. Thus, we extend the original definition of the parameter space for γ to include infinity, hereby making it feasible to discuss both the LSTAR model and the TAR model within the same framework. Figure 2.1 shows how the functional form of G_t changes with γ . In particular, note that as $\gamma \rightarrow 0$, $G_t \rightarrow \frac{1}{2}$ and as $\gamma \rightarrow \infty$, $G_t \rightarrow \mathbb{1}_{\{y_{t-1} - c > 0\}}$ where $\mathbb{1}_{\{A\}}$ is the indicator function equal to one when A is true and zero otherwise. Hence, the TAR model is a limiting case of the LSTAR model prevailing when $\gamma = \infty$. This feature of the model is central to the identifica-

Figure 2.1: The logistic transition function $G_t = \{1 + \exp(-\gamma(y_{t-1} - c))\}^{-\gamma}$ for different values of γ .



tion problem discussed in this paper.

The related ESTAR model is given by (2.1) and $G(y_{t-1}; \gamma, c) = 1 - \exp\{-\gamma(y_{t-1} - c)^2\}$. When $\gamma \rightarrow \infty$, $G_t \rightarrow 0$ (with a single blip at $y_{t-1} = c$) and the ESTAR model approaches a white noise process or, in a more general case, an AR model. Hence, poor identification of the speed of transition parameter is, in contrast to the LSTAR model, often anticipated when testing against a linear model, which is standard in the STAR literature.

3. LIKELIHOOD ANALYSIS OF THE SPEED OF TRANSITION PARAMETER

LSTAR models are traditionally estimated by maximum likelihood (ML) or non-linear least squares (NLS). The two approaches are equivalent when the errors are assumed *i.i.d.* Gaussian, and thus the essential insights from the following ML analysis carry over to NLS. Before introducing the new parametrization, we illustrate some of the less attractive consequences of the original parametrization for estimation and inference in the LSTAR model. We are interested in analyzing only the properties of the ML estimator of γ , and, hence, we fix σ^2 , α and c . The (log-)likelihood function is, apart from a constant, given by

$$L_T(\gamma) = \sum_{t=1}^T \ell_t(\gamma) = -\frac{1}{2} \sum_{t=1}^T \varepsilon_t(\gamma)^2, \quad \varepsilon_t(\gamma) = y_t - \alpha y_{t-1} G_t \quad (3.1)$$

where G_t is the logistic transition function given by (2.2). Lemma 1 below provides results on the behavior of the derivatives of the likelihood function as the speed of transition pa-

parameter, γ , tends to infinity. Observe, in particular, that both the score and Hessian tend to zero as $\gamma \rightarrow \infty$, meaning that the likelihood function becomes flat as the LSTAR model approximates the TAR model. Hence, the TAR model always represents at least a local maximum of the likelihood function.

Lemma 1. *With the likelihood function given in (3.1), it holds for $n \geq 1$ that*

$$\lim_{\gamma \rightarrow \infty} \frac{\partial^n \ell_t(\gamma)}{(\partial \gamma)^n} = 0. \quad (3.2)$$

The proof is given in the appendix.

To illustrate the consequences for estimation, we simulate a data set from an LSTAR model with $T = 150$, $\gamma_0 = 2$, $\sigma^2 = 1$, $c = 0$ and $\alpha = 0.5$. The data series and G_t is graphed in figure A.1 in appendix A while the corresponding likelihood function as a function of γ is depicted in figure 3.1(a) below. Observe that the likelihood function gets flatter as the value of γ grows and the maximum is found, roughly, somewhere in the interval $\gamma \in [35; \infty]$ which does not contain $\gamma_0 = 2$. In empirical applications, researchers often estimate a large value of γ with a large standard error. As our parametrization below clarifies, this is in effect identical to estimating a TAR model, and a more satisfactory solution might be to switch to the TAR framework which by now has a well-developed theoretical framework, see Tong and Lim (1980), Hansen (1997), Seo and Linton (2007) among others.

3.1. THE δ -PARAMETRIZATION

To minimize the harmful flat areas of the likelihood function, we propose the following reparametrization. We define a new parameter $\delta \in (0; 1]$, such that

$$\delta = \frac{\gamma}{1 + \gamma} \quad (3.3)$$

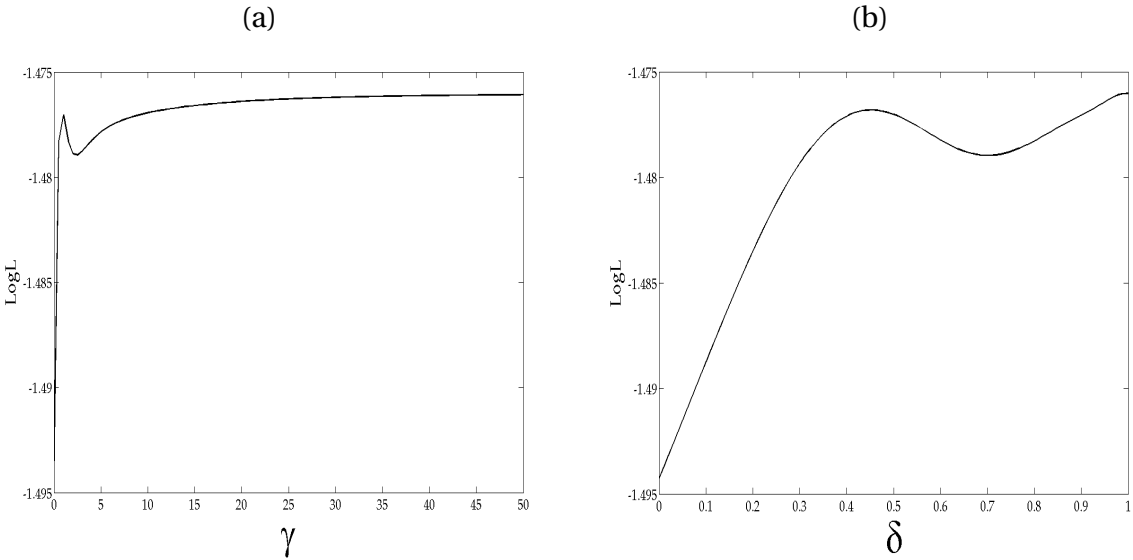
with $\delta \rightarrow 0$ as $\gamma \rightarrow 0$ and $\delta \rightarrow 1$ as $\gamma \rightarrow \infty$. Hence, the transition function in (2.2) is replaced by

$$G(y_{t-1}; \delta, c) = \left(1 + \exp \left\{ -\frac{\delta}{1 - \delta} (y_{t-1} - c) \right\} \right)^{-1} \quad (3.4)$$

Although lemma 1 also applies to an LSTAR likelihood function with (3.4), the reparametrization has advantages compared to the γ -parametrization. The main advantage is that it emphasizes the part of the likelihood function that is of principal interest in an LSTAR model. Essentially, the new parametrization maps $\gamma \in \bar{\mathbb{R}}_+$ into $\delta \in (0; 1]$, where $\delta \in (0; 1)$ is

an LSTAR model, $\delta = 1$ is a TAR model, and $\delta = 0$ is an AR model. Of particular importance is the mapping of $\gamma \in [U; \infty]$ into $\delta \in [u; 1]$, where U is some large value potentially tending to infinity and u is the corresponding value in the δ parametrization. For example, in figure 3.1 set $U \approx 9$ and, hence, $\gamma \in [9; \infty]$ is mapped into $\delta \in [0.9; 1]$. This feature can facilitate numerical optimization of the likelihood function because the large flat part of the likelihood function appearing in figure 3.1(a) is now mapped into a much smaller interval as evident from figure 3.1(b). An example hereof is discussed in the next subsection.

Figure 3.1: The profiled likelihood function as a function of γ (a) and δ (b), respectively. The data set is simulated for $T = 150$, $\gamma_0 = 2$ ($\delta_0 = \frac{2}{3}$), $\sigma^2 = 1$, $c = 0$, $\alpha = 0.5$.



The reparametrization highlights two important aspects that were less clear with the original γ -parametrization. First, the likelihood function is bimodal with a well defined local maximum around $\delta = 0.45$, corresponding to an LSTAR model with $\gamma \approx 0.8$ and, thus, not equal to the true value of $\gamma_0 = 2$ ($\delta_0 = \frac{2}{3}$). Apparently, for this particular realization the local maximum undershoots the true value of the speed of transition. Second, the δ -parametrization stresses that the global maximum of the likelihood function is found close to or at the boundary of the parameter space corresponding to a TAR model. The fact that the likelihood function actually continues to increase until $\delta \approx 1$ is less likely to be seen from the γ -parametrization.

3.1.1. CONSEQUENCES FOR NUMERICAL OPTIMIZATION

Granger and Teräsvirta (1993, p. 123) note that γ tend to be overestimated, and this is also

observed in the literature on LSTAR models where $\hat{\gamma}$ has been reported to have a positive sample bias, see e.g. Areosa *et al.* (2011). This bias may be caused by the estimation of γ without recognizing the behavior of the numerical optimizer when the threshold alternative is the global maximum of a model with a logistic transition function. Table 3.1 shows results from a Monto Carlo study in which the estimated bias in $\hat{\gamma}$ is computed for different values of the stopping criterion of the numerical optimizer used to estimate γ . We focus on the stopping criterion related to the score of the likelihood function. The positive bias in $\hat{\gamma}$ depends heavily on the value of this criterion. This illustrates that the often arbitrary choice of stopping criterion for the numerical optimizer affects the bias in $\hat{\gamma}$. However, for the δ -parametrization, the bias appears (almost) unaffected by the size of the stopping criterion.¹

Table 3.1: Estimated bias in $\hat{\gamma}$ and $\hat{\delta}$ as a function of the stopping criterion for the numerical optimizer.

Number of observations	$T = 150$			$T = 300$		
	$\leq 10^{-2}$	$\leq 10^{-6}$	$\leq 10^{-16}$	$\leq 10^{-2}$	$\leq 10^{-6}$	$\leq 10^{-16}$
$S_T(\hat{x}) = \frac{\partial \ell_T(\hat{x})}{\partial \hat{x}}$						
$\widehat{BIAS}(\hat{\gamma}) = \sum_{m=1}^M (\hat{\gamma}_m - \gamma)$	0.9063	8.2439	49.509	0.6594	5.4652	21.781
$\widehat{BIAS}(\hat{\delta}) = \sum_{m=1}^M (\hat{\delta}_m - \delta)$	0.0533	0.0545	0.0545	0.0146	0.0148	0.0148

Note: The DGP is $\gamma = 1$ ($\delta = \frac{1}{2}$), $\sigma^2 = 1, c = 0$ and $\alpha = 0.5$. $M = 10,000$ and c and α are fixed in estimation.

4. SELECTING BETWEEN LSTAR AND TAR WITH INFORMATION CRITERIA

The bimodality of the likelihood function seen in figure 3.1(b) is a common small sample property of LSTAR models. Typically, there exists one inner maximum corresponding to an LSTAR model and a maximum on the boundary of the parameter space ($\delta = 1$) corresponding to a TAR model. The simulation considered above in figure 3.1 is an extreme example of this, where the global maximum of the likelihood function is at $\delta \approx 1$. A more typical case is one with the inner maximum being the global maximum and a local, smaller maximum is found at the TAR solution, see for example figure 6.2(b). A relevant question is therefore whether the likelihood value of the inner maximum is large enough compared

¹Note that the size of the bias is not comparable across parametrizations due to the different scaling of the parameters and that since the ML estimator is consistent, the bias diminishes as T grows.

to the likelihood value of the boundary TAR maximum to justify estimation of a speed of transition parameter. One way to investigate this question would be to derive a test for the null-hypothesis of $\delta = 1$. However, such a test is highly non-standard since, as Lemma 1 shows, all derivatives are zero and, hence, it is not obvious how to obtain critical values. We propose instead to conduct model selection based on information criteria, where no critical values are needed. Information criteria combine a measure of goodness-of-fit with a penalty for model complexity. Comparing information criteria would therefore indicate whether the additional speed of transition parameter of an estimated LSTAR model leads to a notable improvement of fit compared to a corresponding TAR model. Note that the theoretical foundation for validity of the information criteria when the true model is the TAR model suffers from the same difficulties as a formal test. As a result, we are unable to prove the asymptotic validity of this selection procedure analytically. Rather, we rely on simulation studies which give clear indications that the model selection procedure is indeed consistent. We conjecture that these simulation results are not specific to the selected models, such that information criteria can be used more generally to select between TAR and STAR models.

Psaradakis *et al.* (2009) pursue a comparable idea and consider selecting between several non-linear autoregressive models by means of information criteria. In the following, we conduct a similar simulation study for the choice between a TAR model and an LSTAR model using the proposed reparametrization. With $L_T(\delta)$ being the likelihood function given in (3.1) evaluated with respect to δ , the information criteria are of the form

$$IC_T(\delta, k) = -2L_T(\delta) + kc_T, \quad (4.1)$$

where k is the number of estimated parameters which equals 1 for the LSTAR and 0 for the TAR. The term c_T is a function of T that satisfies $\lim_{T \rightarrow \infty} c_T = \infty$ and $\lim_{T \rightarrow \infty} (T^{-1}c_T) = 0$. We focus on the Bayesian Information Criterion (BIC), Schwarz (1978), with $c_T = \log(T)$, and the Hannan-Quinn Information Criterion (HQIC), Hannan and Quinn (1979), with $c_T = 2\log(\log(T))$. Both criteria fulfill the requirements for c_T . We use the information criteria to estimate the number of parameters, k , and denote this estimate \hat{k} . The selection procedure is consistent if $\hat{k} \rightarrow k_0$ as $T \rightarrow \infty$. We illustrate this selection method using four different models, three LSTAR models with $\delta = \{0.2, 0.5, 0.9\}$, respectively, and a TAR model. We simulate $M = 10,000$ data sets and estimate only the speed of transition parameter δ . The remaining parameters are fixed at the true values: $\sigma^2 = 1$, $\alpha = 0.5$ and $c = 0$. For each replication, we calculate the percentage selected LSTAR models of the two information criteria. The experiment is done for different sample lengths and the selection percentages

are given in table 4.1.

Table 4.1: Percentage selected LSTAR models using information criteria. $c = 0$.

DGP	LSTAR, $\delta = 0.2$		LSTAR, $\delta = 0.5$		LSTAR, $\delta = 0.9$		TAR, $\delta = 1$	
T	BIC	HQIC	BIC	HQIC	BIC	HQIC	BIC	HQIC
100	48	64	13	25	1	4	0	0
250	82	92	26	45	0	3	0	0
500	98	99	47	69	1	3	0	0
1,000	100	100	76	90	1	4	0	0
10,000	100	100	100	100	1	6	0	0
50,000	100	100	100	100	2	21	0	0
100,000	100	100	100	100	6	42	0	0
1,000,000	100	100	100	100	100	100	0	0

Note: Only δ is estimated. Remaining parameters are fixed at true values of $\sigma^2 = 1$, $\alpha = 0.5$ and $c = 0$.

The results show that the slower the speed of transition, the better the performance of the information criteria. Nevertheless, even with a relatively slow transition speed of $\delta = 0.5$ and $T = 1,000$, BIC and HQIC still select a rather large number of incorrect TAR models, 24% and 10%, respectively. For the LSTAR model with $\delta = 0.9$ the information criteria are apparently punishing too severely for the additional parameter and do not choose the LSTAR model in 100% of the cases until $T = 1,000,000$. Again, this shows that while the identification problem for δ is a small sample problem, the label “small sample” is misleading since T needs to be extremely large to get a clear distinction between an LSTAR and a TAR model. This finding is also supported by the results of Castle and Hendry (2013, table 3) which shows that data generated from the two models is highly correlated. Observe that when the TAR model is the DGP, the information criteria perform well.

The results of a repeated Monte Carlo experiment with $c = 1$ in table 4.2 show significant improvements in the selection rates of the information criteria for small samples. It is peculiar that the power of the information criteria depends on the (fixed) value of the threshold parameter c .

Table 4.2: Percentage selected LSTAR models using information criteria. $c = 1$.

DGP	LSTAR, $\delta = 0.2$		LSTAR, $\delta = 0.5$		LSTAR, $\delta = 0.9$		TAR, $\delta = 1$	
	BIC	HQIC	BIC	HQIC	BIC	HQIC	BIC	HQIC
100	52	67	24	42	4	11	0	0
250	84	92	51	71	6	17	0	0
500	98	99	77	89	9	27	0	0
1,000	100	100	96	98	18	45	0	0
10,000	100	100	100	100	98	100	0	0
50,000	100	100	100	100	100	100	0	0
100,000	100	100	100	100	100	100	0	0
1,000,000	100	100	100	100	100	100	0	0

Note: Only δ is estimated. Remaining parameters are fixed at true values of $\sigma^2 = 1$, $\alpha = 0.5$ and $c = 1$.

Overall, if model selection based on information criteria prefer an LSTAR, it is a clear indication that the speed of transition is slow enough to make a difference compared to the TAR model. On the other hand, if the TAR is chosen, there is a risk that one has incorrectly fixed $\delta = 1$ and selected the TAR model. However, this only means that δ is irrelevant for the model. Hence, information criteria provide a conservative tool to select LSTAR models over TAR models that can be used if the researcher wishes to comment on the speed of transition.

5. ESTIMATING LSTAR MODELS

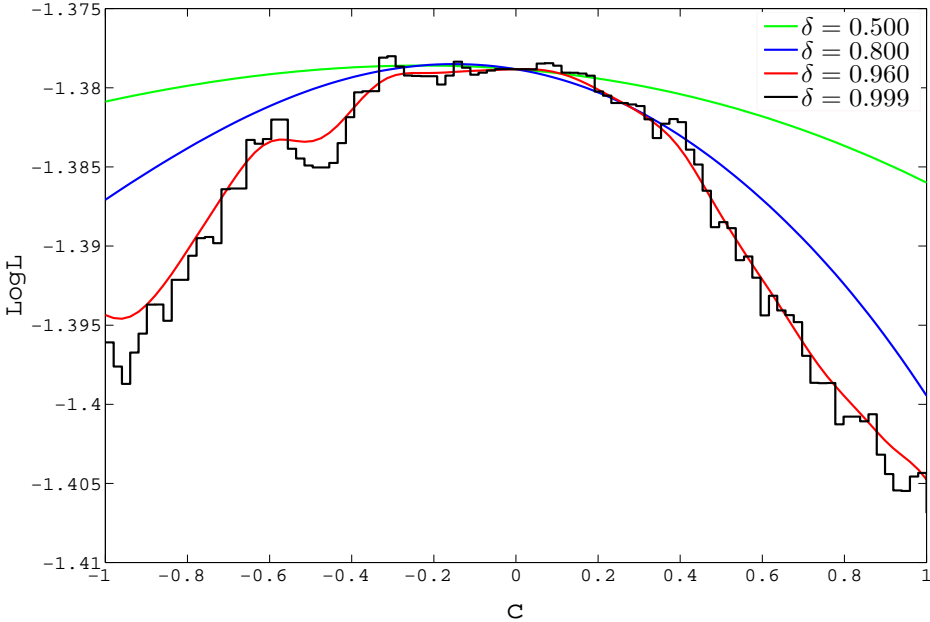
The properties of the likelihood function for LSTAR models discussed so far introduce difficulties for numerical optimization. We observe two separate problems that have to be taken into account. First, the likelihood function might have a multiple maxima in the direction of δ , as described in the previous sections. To handle this, it is useful to estimate δ with a derivative based optimizer and using different initial values from the parameter space $\delta \in (0; 1]$. To ensure that the reached maximum is global, it is important to always calculate the additional likelihood value at the limit, $\delta = 1$.

The second difficulty is that the likelihood function approaches the step-wise likelihood function of a TAR model in the direction of c as $\delta \rightarrow 1$. Consequently, many local

maxima exist in the direction of c and derivative based optimizers will not work well. For an illustration see figure 5.1, which shows the likelihood as a function of c for different values of δ . To circumvent the problem of a step-wise likelihood function, a grid search algorithm over c can be performed with an interval that covers observed values of y_{t-1} spanning from, e.g., the 10th to the 90th percentile of the distribution of y_{t-1} . This grid search technique for c is standard in the TAR literature and ensures that all relevant points for threshold locations are examined. The rest of the parameters are estimated using least squares conditional on the transition function parameters.

When estimating simple models, as the one analyzed in this paper, performing a two dimensional grid search over δ and c and drawing the profiled likelihood function is generally informative. This approach allows the researcher to take into account both problems. Note that this proposal is by no means new and is in fact standard practice in the literature for finding candidates for initial values, see inter alia Bec *et al.* (2008) and Teräsvirta *et al.* (2010, ch. 12). Our contribution is that the δ -parametrization clarifies the reason for doing the grid search, and we emphasize that the main problem of multiple equilibria of the LSTAR model is related to the fact that the likelihood function approaches a step-wise likelihood function as $\delta \rightarrow 1$.

Figure 5.1: Profiled likelihood functions in the direction of c for different values of δ . Data is one realization from an LSTAR model with $T = 300$, $\sigma^2 = 1$, $c = 0$ and $\alpha = 0.5$.



6. EMPIRICAL APPLICATIONS

We illustrate by two empirical applications from the LSTAR literature the advantages of the δ -parametrization over the γ -parametrization and model selection based on information criteria. The first application illustrates a situation where the δ -parametrization reveals that the reported maximum of the likelihood function is not the global maximum. In the second application, the δ -parametrization confirms that the global maximum is the reported one, but information criteria prefer the TAR model over the LSTAR model because the regime switching is so fast that estimating the additional speed of transition parameter is superfluous.

6.1. WOLF'S ANNUAL SUNSPOT NUMBERS

Teräsvirta *et al.* (2010, p. 390), illustrate a suggested STAR modeling procedure by analyzing Wolf's annual sunspot numbers dating from 1700 to 1979. The data is published at the Belgian web page of Solar Influences Data Analysis Center.² Following Teräsvirta *et al.* (2010) the series is transformed as: $y_t = 2\{(1 + z_t)^{1/2} - 1\}$ where z_t is the original series. The motivation for transformation is that the transformed series is easier to model than the untransformed one. The original estimated LSTAR model is reproduced with both parametrizations and given by (standard errors in parenthesis)

$$\begin{aligned}
 y_t = & \underset{(0.08)}{1.46}y_{t-1} - \underset{(0.13)}{0.76}y_{t-2} + \underset{(0.05)}{0.17}y_{t-7} + \underset{(0.04)}{0.11}y_{t-9} \\
 & + \underset{(0.85)}{(2.65 - 0.54}y_{t-1} + \underset{(0.13)}{0.75}y_{t-2} - \underset{(0.18)}{0.47}y_{t-3} \\
 & + \underset{(0.11)}{0.32}y_{t-4} - \underset{(0.07)}{0.26}y_{t-5} - \underset{(0.05)}{0.24}y_{t-8} + \underset{(0.06)}{0.17}y_{t-10}) \times \widehat{G}_t^x \quad (6.1)
 \end{aligned}$$

$$\begin{aligned}
 x = \gamma: \quad \widehat{G}^\gamma &= 1 + \exp\left\{-\underset{(1.11)}{5.46}(y_{t-2} - \underset{(0.36)}{7.88})/\widehat{\sigma}_{y_{t-2}}\right\}^{-1} \\
 x = \delta: \quad \widehat{G}^\delta &= 1 + \exp\left\{-\frac{\underset{(0.03)}{0.85}}{1 - \underset{(0.03)}{0.85}}(y_{t-2} - \underset{(0.36)}{7.88})/\widehat{\sigma}_{y_{t-2}}\right\}^{-1}
 \end{aligned}$$

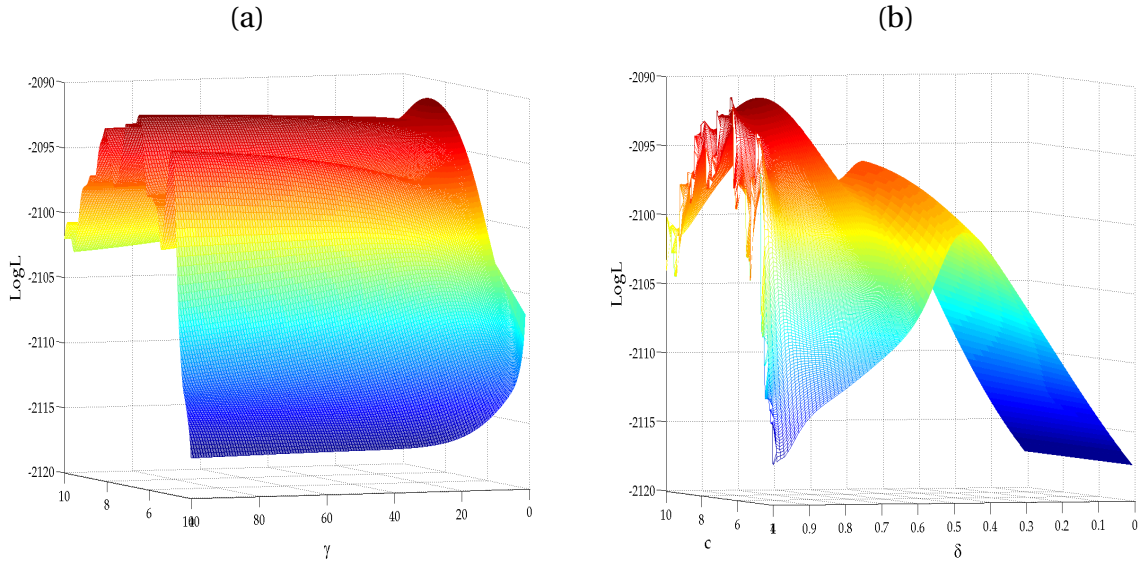
$$T = 270, \quad RSS = 921.84, \quad \text{Log}L = -2,091.2$$

²<http://www.sidc.oma.be/sunspot-data/>

$$BIC = 4,260.8, \quad HQIC = 4,230.7$$

The normalization by $\hat{\sigma}_{y_{t-2}}$ in the transition function is standard in the literature of applied STAR models because it facilitates the choice of grid or initial values for γ , see van Dijk *et al.* (2002). The profiled likelihood function in direction of c and γ for each parametrization is showed in figure 6.1. The characteristically flatness in the direction of γ is pro-

Figure 6.1: Profiled likelihood functions of the LSTAR model for Wolf's sunspot numbers, 1710-1979. (a) is for the γ -parametrization and (b) is the for the δ -parametrization.



nounced in figure 6.1(a), and the reported maximum for $(\hat{c}, \hat{\gamma}) = (5.46, 7.88)$ appears relatively well-defined. However, figure 6.1(b) reveals that the global maximum is actually the TAR model at the boundary $\delta = 1$, whereas the LSTAR model is only a local maximum. The γ -parametrization has effectively blurred the shape of the likelihood function. At the boundary, the TAR likelihood function is characterized by discrete jumps over the range of c . This implies that performing a careful grid search over potential values of c is crucial for the estimation of c , as discussed in section 5 and, more importantly, that inference on c is non-standard, cf., Chan (1993) and Hansen (1997). Estimating the TAR model yields³

$$y_t = \underset{(0.08)}{1.43}y_{t-1} - \underset{(0.14)}{0.77}y_{t-2} + \underset{(0.05)}{0.17}y_{t-7} + \underset{(0.05)}{0.12}y_{t-9}$$

³The grid search of c is performed over values of y_{t-2} , disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of y_{t-2} . No standard error of \hat{c} is reported due to the non-standard inference on the threshold parameter in a TAR model.

$$\begin{aligned}
& + \underset{(0.70)}{2.69} - \underset{(0.11)}{0.45}y_{t-1} + \underset{(0.18)}{0.69}y_{t-2} - \underset{(0.11)}{0.48}y_{t-3} \\
& + \underset{(0.11)}{0.36}y_{t-4} - \underset{(0.07)}{0.27}y_{t-5} - \underset{(0.05)}{0.21}y_{t-8} + \underset{(0.05)}{0.14}y_{t-10} \times \mathbb{I}(y_{t-2} > 6.39). \quad (6.2)
\end{aligned}$$

$$\begin{aligned}
T &= 270, \quad RSS = 920.66, \quad \text{Log}L = -2,090.9 \\
BIC &= 4,254.6, \quad HQIC = 4,226.6
\end{aligned}$$

While the AR parameters are almost identical to those of the LSTAR model in (6.1), the threshold parameter differs between the models. This TAR maximum is preferred by the information criteria to the reported LSTAR model in (6.1) because the TAR model achieves a higher (lower) value of LogL (RSS) in addition to be one parameter short of the LSTAR model.⁴ The TAR maximum (6.2) can easily be reproduced with the δ -parametrization by performing a two-dimensional grid search over c and $\delta \in (0; 1]$. A similar exercise for the γ -parametrization produces, depending on the choice of grid for γ as well as the choice of stopping criterion, either the local LSTAR maximum of (6.1) or an invalid maximum with all observations in one regime. Hence, the model that truly maximizes the likelihood function is impossible to estimate with the γ -parametrization because γ is infinity.

Nevertheless, given that an LSTAR process has a TAR maximum as a small sample property, as found in section 4, and the relatively small sample size of 270, the LSTAR model cannot be discarded as being the DGP of this sunspot data. In addition, the likelihood function in the region of the local LSTAR maximum in (6.1) and appearing in figure 6.1, seems closely approximated by a quadratic form, and is thus a well defined maximum. Based on these considerations, one could also argue that the LSTAR model may be the DGP of the process.

6.2. U.S. UNEMPLOYMENT RATE

The paper by van Dijk *et al.* (2002) illustrates a suggested STAR modeling cycle which includes, among others, impulse response and forecasting analysis. The data series is the monthly seasonally unadjusted unemployment rate for U.S. males aged 20 and over for the period 1968:6-1989:12.⁵

⁴Teräsvirta *et al.* (2010) reach similar conclusion when estimating a TAR model for the same data later in the book, though, without specifying a measurement. Their TAR model is, however, specified differently and non-nested with (6.2) and (6.1) which makes direct comparisons infeasible.

⁵The series is constructed from data on the unemployment level and labor force for the particular sub population. These two series are published together with Gauss programs used to estimate their model at <http://swopec.hhs.se/hastef/abs/hastef0380.htm>.

The LSTAR model is reproduced with both parametrizations and given by (standard errors in parenthesis)

$$\begin{aligned}
\Delta y_t = & \underset{(0.07)}{0.479} + \underset{(0.07)}{0.645}D_{1,t} - \underset{(0.10)}{0.342}D_{2,t} - \underset{(0.09)}{0.680}D_{3,t} - \underset{(0.11)}{0.725}D_{4,t} - \underset{(0.10)}{0.649}D_{5,t} \\
& - \underset{(0.09)}{0.317}D_{6,t} - \underset{(0.09)}{0.410}D_{6,t} - \underset{(0.09)}{0.501}D_{8,t} - \underset{(0.09)}{0.554}D_{9,t} - \underset{(0.07)}{0.306}D_{10,t} \\
& + [-\underset{(0.01)}{0.040}y_{t-1} - \underset{(0.08)}{0.1460}\Delta y_{t-1} - \underset{(0.06)}{0.101}\Delta y_{t-6} + \underset{(0.06)}{0.097}\Delta y_{t-8} - \underset{(0.06)}{0.123}\Delta y_{t-10} \\
& + \underset{(0.07)}{0.129}\Delta y_{t-13} - \underset{(0.06)}{0.103}\Delta y_{t-15}] \times [1 - \widehat{G}_t^x] \\
& + [-\underset{(0.01)}{0.011}y_{t-1} + \underset{(0.08)}{0.225}\Delta y_{t-1} + \underset{(0.08)}{0.307}\Delta y_{t-2} - \underset{(0.07)}{0.119}\Delta y_{t-7} - \underset{(0.09)}{0.155}\Delta y_{t-13} \\
& - \underset{(0.09)}{0.215}\Delta y_{t-14} - \underset{(0.09)}{0.235}\Delta y_{t-15}] \times \widehat{G}_t^x
\end{aligned} \tag{6.3}$$

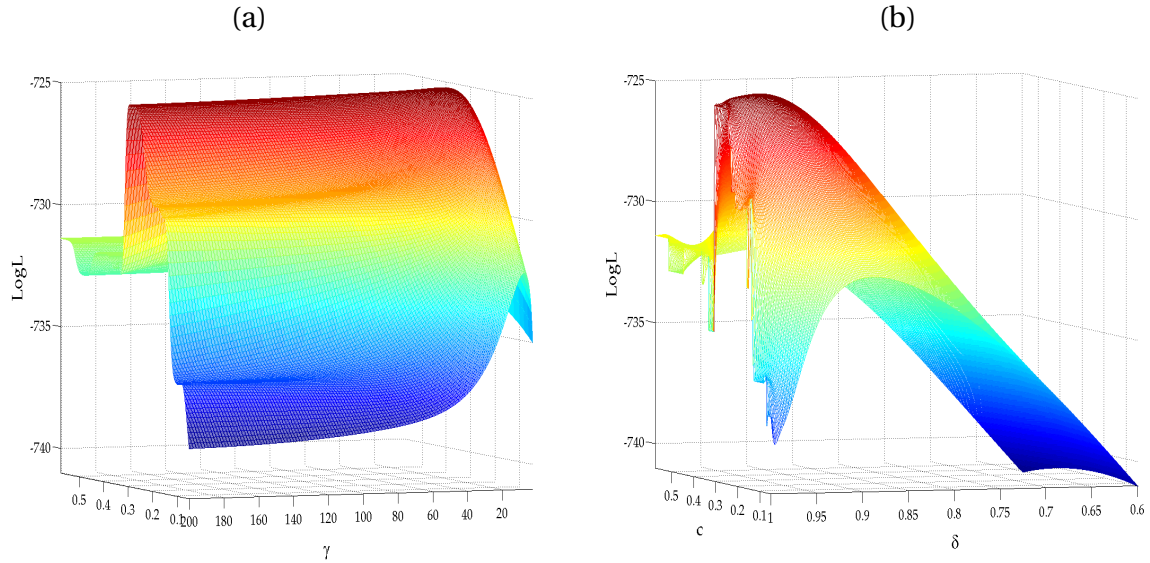
$$\begin{aligned}
x = \gamma: \quad \widehat{G}^\gamma &= 1 + \exp\left\{-\underset{(21.75)}{23.15}(\Delta_{12}y_{t-1} - \underset{(0.04)}{0.274})/\widehat{\sigma}_{\Delta_{12}y_{t-1}}\right\}^{-1} \\
x = \delta: \quad \widehat{G}^\delta &= 1 + \exp\left\{-\frac{\underset{(0.04)}{0.96}}{1 - \underset{(0.04)}{0.96}}(\Delta_{12}y_{t-1} - \underset{(0.04)}{0.274})/\widehat{\sigma}_{\Delta_{12}y_{t-1}}\right\}^{-1}
\end{aligned}$$

$$T = 240, \quad RSS = 8.178, \quad \text{Log}L = -725.0$$

$$BIC = 1,597.9, \quad HQIC = 1,541.8$$

$D_{s,t}$ is monthly dummy variables where $D_{s,t} = 1$ if observation t corresponds to month s and $D_{s,t} = 0$ otherwise. van Dijk *et al.* (2002) have sequentially removed all variables with a t -statistic lower than 1 in absolute value. Observe that γ is rather large and imprecisely estimated indicating that data contains little information about the size of this parameter. The profiled likelihood functions for the two parametrizations are displayed in figure 6.2. Because $\widehat{\gamma}$ is so large, the maximum is visually absorbed by the flatness of the γ -likelihood function in figure 6.2(a). In contrast, the δ -likelihood function in figure 6.2(b) confirms that the reported maximum is in fact the global maximum of the likelihood function. Interestingly, the δ -likelihood function shows that the local TAR maximum at the boundary leads to only a minor drop in likelihood value compared to the LSTAR model. To check whether this TAR model is preferred by information criteria, the TAR model is estimated

Figure 6.2: Profiled likelihood functions of the LSTAR model for U.S. male unemployment rate, 1968:6-1989:12. (a) is for the γ -parametrization and (b) is for the δ -parametrization.



and given by⁶

$$\begin{aligned}
 \Delta y_t = & \underset{(0.07)}{0.473} + \underset{(0.07)}{0.644}D_{1,t} - \underset{(0.10)}{0.343}D_{2,t} - \underset{(0.09)}{0.675}D_{3,t} - \underset{(0.11)}{0.721}D_{4,t} - \underset{(0.10)}{0.641}D_{5,t} \\
 & - \underset{(0.09)}{0.308}D_{6,t} - \underset{(0.09)}{0.410}D_{6,t} - \underset{(0.08)}{0.505}D_{8,t} - \underset{(0.09)}{0.546}D_{9,t} - \underset{(0.07)}{0.295}D_{10,t} \\
 & + [-\underset{(0.01)}{0.040}y_{t-1} - \underset{(0.08)}{0.140}\Delta y_{t-1} - \underset{(0.06)}{0.094}\Delta y_{t-6} + \underset{(0.06)}{0.092}\Delta y_{t-8} - \underset{(0.06)}{0.116}\Delta y_{t-10} \\
 & + \underset{(0.07)}{0.136}\Delta y_{t-13} - \underset{(0.06)}{0.106}\Delta y_{t-15}] \times \mathbb{I}(\Delta_{12}y_{t-1} \leq 0.268) \\
 & [-\underset{(0.01)}{0.012}y_{t-1} + \underset{(0.08)}{0.227}\Delta y_{t-1} + \underset{(0.08)}{0.307}\Delta y_{t-2} - \underset{(0.07)}{0.094}\Delta y_{t-7} - \underset{(0.09)}{0.146}\Delta y_{t-13} \\
 & - \underset{(0.09)}{0.211}\Delta y_{t-14} - \underset{(0.09)}{0.216}\Delta y_{t-15}] \times \mathbb{I}(\Delta_{12}y_{t-1} > 0.268)
 \end{aligned} \tag{6.4}$$

$$T = 240, \quad RSS = 8.191, \quad LogL = -725.3$$

$$BIC = 1,593.2, \quad HQIC = 1,539.1$$

The information criteria prefer this TAR model implying that the speed of transition is too poorly estimated to make a difference.

This application highlights one of the key points of the present paper, namely that a

⁶Similar to the previous TAR estimation, the grid search of c is performed over values of $\Delta_{12}y_{t-1}$, disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of $\Delta_{12}y_{t-1}$. No standard error of \hat{c} is reported due to the non-standard inference on the threshold parameter in a TAR model.

large and imprecisely estimated γ implies that the LSTAR model is effectively a TAR model. Estimation of the LSTAR model is too much to ask of the data. Moreover, we observe the consequences of the flat likelihood function for inference on $\hat{\gamma}$. The estimated standard error of $\hat{\gamma}$ ($\text{s.e.}(\hat{\gamma})$) is large due to the flatness of the likelihood function in direction of γ towards infinity, see figure 6.2(a). However, the large $\text{s.e.}(\hat{\gamma})$ seems less justified towards zero where one observes a large drop in the likelihood. This illustrates how the flatness contaminates the estimation of the variance of $\hat{\gamma}$ for which zero is well within a two $\text{s.e.}(\hat{\gamma})$. Consequently, one might conclude that γ could be zero, which from a look at the function in figure 6.2(a) seems unlikely. For this reason (and because a test of $\gamma = 0$ results in vanishing parameters and, thus, is a non-standard test), it is common practice in the LSTAR literature not to comment on the $\text{s.e.}(\hat{\gamma})$. It is seen from the $\text{s.e.}(\hat{\delta})$ that the δ -parametrization does not suffer from this problem in the present application.

7. CONCLUSION

Regime switching models characterized by smooth transitions only differ from discrete regime switching models by the speed of transition parameter. Thus, estimation and identification of this parameter is essential not only for economic interpretation but also for model selection. Nevertheless, the identification problem and its consequences for estimation have received little attention in the literature. We show that the original parametrization of the speed of transition parameter is problematic as the likelihood function is characterized by large flat areas caused by all derivatives approaching zero with faster speed of transition. This implies that the magnitude of the estimator may depend on the arbitrarily chosen stopping criteria of the numerical optimizer. To circumvent this problem, we propose a reparametrization of the LSTAR model. The reparametrization maps the parameter space of the original speed of transition parameter into a much smaller interval which facilitates identifying the global maximum of the likelihood function as well as numerical optimization. We then show that the TAR model can be the global maximum of a LSTAR likelihood function, while it, by construction, is always at least a local maximum. Instead of relying solely on economic theory when justifying the additional parameter of the LSTAR model, we show that information criteria provide a model selection tool that can be applied if the researcher wishes to comment of the speed of transition. Acknowledging that the LSTAR model considered in this paper is simple and the presented simulation results only apply to this particular framework, the new parametrization provides general insights on the shape of the likelihood function in directions of the two param-

eters of the transition function that can be generalized to a broad range of other models within the smooth switching literature. For example, the double-logistic smooth transition (D-LSTAR), the Multi-Regime Smooth Transition Autoregression (MR-STAR) and the logistic autoregressive conditional root (LACR) model, see, e.g. , Bec *et al.* (2010) and Bec *et al.* (2008).

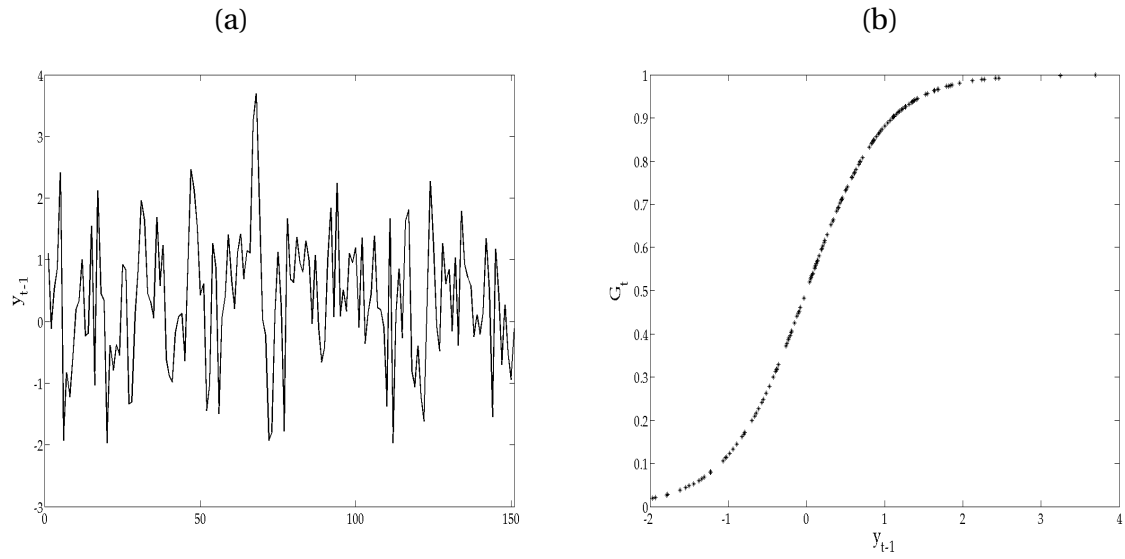
REFERENCES

- AREOSA, W. D., M. MCALEER, AND M. C. MEDEIROS (2011): "Moment-based estimation of smooth transition regression models with endogenous variables." *Journal of Econometrics*, 165(1):100–111.
- BACON, D. W. AND D. G. WATTS (1971): "Estimating the Transition Between Two Intersecting Straight Lines." *Biometrika*, 58:525–534.
- BEC, F., A. RAHBK, AND N. SHEPHARD (2008): "The ACR Model: A Multivariate Dynamic Mixture Autoregression*." *Oxford Bulletin of Economics and Statistics*, 70(5):583–618.
- BEC, F., M. B. SALEM, AND M. CARRASCO (2010): "Detecting Mean Reversion in Real Exchange Rates from a Multiple Regime STAR Model." *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, (99/100):395–427.
- CASTLE, J. AND D. HENDRY (2013): "Semi-automatic Non-linear Model Selection." *Essays in Nonlinear Time Series Econometrics*, Oxford University Press, forthcoming. Edited by Haldrup, N., Meitz M. and Saikkonen, P.
- CHAN, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model." *The Annals of Statistics*, 21(1):520–533.
- CHAN, K. S. AND H. TONG (1986): "On estimating thresholds in autoregressive models." *Journal of Time Series Analysis*, 7(3):179–190.
- DAVIES, R. B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative." *Biometrika*, 74(1):33.
- GRANGER, C. AND T. TERÄSVIRTA (1993): *Modelling nonlinear economic relationships*. Oxford University Press, USA.
- HANNAN, E. J. AND B. G. QUINN (1979): "The determination of the order of an autoregression." *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 190–195.
- HANSEN, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis." *Econometrica: Journal of the Econometric Society*, page 413–430.
- (1997): "Inference in TAR Models." *Studies in Nonlinear Dynamics & Econometrics*, 2(1):1.

- HEINEN, F., S. MICHAEL, AND P. SIBBERTSEN (2012): “Weak identification in the ESTAR model and a new model.” *Journal of Time Series Analysis*.
- KRISTENSEN, D. AND A. RAHBEK (2013): “Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models.” *Econometric Theory*, forthcoming.
- LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): “Testing linearity against smooth transition autoregressive models.” *Biometrika*, 75:491–499.
- PSARADAKIS, Z., M. SOLA, F. SPAGNOLO, AND N. SPAGNOLO (2009): “Selecting nonlinear time series models using information criteria.” *Journal of Time Series Analysis*, 30(4):369–394.
- SCHWARZ, G. (1978): “Estimating the dimension of a model.” *The annals of statistics*, 6(2):461–464.
- SEO, M. AND O. LINTON (2007): “A smoothed least squares estimator for threshold regression models.” *Journal of Econometrics*, 141(2):704–735.
- TERÄSVIRTA, T. (1994): “Specification, estimation, and evaluation of smooth transition autoregressive models.” *Journal of American Statistical Association*, 89:208–218.
- (1998): “Modelling economic relationships with smooth transition regressions.” *Handbook of Applied Economic Statistics*, Marcel Dekker: New York, pages 507–552.
- TERÄSVIRTA, T., D. TJØSTHEIM, AND C. GRANGER (2010): *Modelling Nonlinear Economic Time Series*. Advanced Texts in Econometrics Series. Oxford University Press.
- TONG, H. (2011): “Threshold models in time series analysis—30 years on.” *Statistics and its Interface*, 4(2):107–118.
- TONG, H. AND K. S. LIM (1980): “Threshold autoregression, limit cycles and cyclical data.” *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 245–292.
- VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): “Smooth transition autoregressive models - a survey of recent developments.” *Economic Reviews*, 21:1–47.

A. SIMULATED LSTAR PROCESS AND LOGISTIC TRANSITION FUNCTION

Figure A.1: Simulated data series (a) and transition function (b) for the LSTAR model (2.1) with $\gamma = 2$, $c = 0$, $\alpha = 0.5$ and $T = 150$.



B. PROOF OF LEMMA 1

Observe initially that with G_t defined in (2.2), it holds that

$$\frac{\partial G_t}{\partial \gamma} = G_t(1 - G_t)(y_{t-1} - c) =: \psi_t(y_{t-1} - c)$$

and

$$\frac{\partial^2 G_t}{(\partial \gamma)^2} = \frac{\partial \psi_t}{\partial \gamma}(y_{t-1} - c) = \psi_t(1 - 2G_t)(y_{t-1} - c)^2.$$

Moreover, as $\gamma \rightarrow \infty$, one has $\psi_t \rightarrow 0$ and hence $\partial G_t / \partial \gamma \rightarrow 0$ and $\partial^2 G_t / (\partial \gamma)^2 \rightarrow 0$. In fact, note that all higher order derivatives will have the form

$$\frac{\partial^n G_t}{(\partial \gamma)^n} = \psi_t g(G_t)(y_{t-1} - c)^n$$

where $g(G_t)$ is a function consisting of an integer and of sums and products of G_t . In particular, observe that since $0 < G_t < 1$, it holds for any $n < \infty$ that $g(G_t) = K$ for a constant $K < \infty$. Thus, we have that

$$\frac{\partial^n G_t}{(\partial \gamma)^n} \rightarrow 0 \quad \text{as } \gamma \rightarrow \infty. \quad (\text{B.1})$$

Next, consider the likelihood contribution given by $\ell_t(\gamma)$ in (3.1). Standard calculus gives

$$\frac{\partial \ell_t(\gamma)}{\partial \gamma} = \varepsilon_t(\gamma) \alpha y_{t-1} \frac{\partial G_t}{\partial \gamma} = \alpha y_{t-1} y_t \frac{\partial G_t}{\partial \gamma} - \alpha^2 y_{t-1}^2 G_t \frac{\partial G_t}{\partial \gamma} =: a_t(\gamma) + b_t(\gamma).$$

Observe that the higher order derivatives of the terms $a_t(\gamma)$ and $b_t(\gamma)$ with respect to γ will be of the respective forms

$$\frac{\partial^n a_t(\gamma)}{(\partial \gamma)^n} = \alpha y_{t-1} y_t \frac{\partial^n G_t}{(\partial \gamma)^n} \quad \text{and} \quad \frac{\partial^n b_t(\gamma)}{(\partial \gamma)^n} = -\alpha^2 y_{t-1}^2 \sum_{k=0}^n \binom{n}{k} \frac{\partial^k G_t}{(\partial \gamma)^k} \frac{\partial^{n-k} G_t}{(\partial \gamma)^{n-k}}.$$

Consequently, it holds by (B.1) that $\partial^n \ell_t(\gamma) / (\partial \gamma)^n \rightarrow 0$ as $\gamma \rightarrow \infty$.

Observe that the same result holds for the parametrization with δ since

$$\frac{\partial^n G_t}{(\partial \delta)^n} = \frac{\partial^n G_t}{(\partial \gamma)^n} \frac{\partial^n \gamma}{(\partial \delta)^n} \quad \text{and} \quad \frac{\partial^n \gamma}{(\partial \delta)^n} = \frac{n!}{(\delta - 1)^{(n+1)},}$$

where $\partial^n \gamma / (\partial \delta)^n$ is function that grows as $\delta \rightarrow 1$. However, the grows rate is slower than the

exponential decay of $\partial^n G_t / (\partial \gamma)^n$ and, hence, we still have,

$$\lim_{\delta \rightarrow 1} \frac{\partial^n G_t}{(\partial \delta)^n} = 0.$$

□

Chapter 3

A Structural Analysis of the Convenience Yield on U.S. Treasury Bonds

A Structural Analysis of the Convenience Yield on U.S. Treasury Bonds

Line Elvstrøm Ekner*

ABSTRACT: U.S. Treasury bonds are the most liquid and safe assets on the market. This fact leads to a convenience yield on Treasury bonds relative to high grade corporate bonds which depends on the supply of Treasury bonds. We model the dynamics of the persistent variables measuring Treasury supply and convenience yield in vector error correction model (VECM). The long-run relationship is negative reflecting a downward-sloping demand function for liquidity and safety. However, the short-run relationship is positive and identified by means of a structural VECM. The driving factor of the short-run result is the safety premium of the convenience yield. The safety premium is reduced on impact and the first two years after a negative shock to the supply of Treasury bonds, before turning positive and raised in accordance with the long-run negative relationship. The results are robust to a potential kink in the demand relation estimated by a threshold VECM.

1. INTRODUCTION

An asset is said to carry a convenience yield or is called “special” if it offers benefits from holding it rather than having a contract on the asset. The identifying feature of such an asset is that its price will be inflated relative to other benchmarks, or equivalently, its yield will be lower than benchmarks. It has been argued that U.S. Treasury bonds carry such a convenience yield because they offer extreme liquidity and safety for an investor, see Reinhart *et al.* (2000), Fleming (2000) and Krishnamurthy and Vissing-Jørgensen (2012). The almost certain promise of nominal repayment is valuable because it implies that these assets can be used as guarantee in financial transactions and because investment needs of, e.g., foreign central banks and insurance companies, only can be satisfied by holding safe and liquid assets. Thus, given a limited supply of Treasury bonds, a convenience yield is induced by investors’ demand for liquidity and safety related to this scarce asset. Krishnamurthy and Vissing-Jørgensen (2012) argue that the safety premium on Treasury bonds differs from a standard risk premium because it arises from these investors’ demand and,

*Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark, line.elvstrom.ekner@econ.ku.dk

as a result, the size of the safety premium depends on the supply of Treasury bonds. In contrast, a risk premium is normally unrelated to the supply of the given asset.

This paper analyzes the dynamics of the convenience yield on U.S. Treasuries using annual observations from 1919-2008. We confirm the results found in previous studies, namely, that there is a negative relationship between the supply of U.S. Treasury bonds, measured by U.S. government debt, and the convenience yield, measured by the long-term spread between yields of corporate bonds and Treasury bonds. Corporate bonds offer less liquidity and/or less safety for an investor than Treasury bonds, and, hence, the spread measures the convenience yield. We emphasize, however, that the downward-sloping demand function can only be interpreted as a long-run demand relationship. In the short-run, the results show that the demand function has a positive slope or is at best flat.

In light of the recent emphasis on Quantitative Easing as a monetary policy tool, there is by now a rich literature documenting a positive relationship between the (expected and/or relative) supply of Treasuries and the yield of Treasuries, see, e.g, Kuttner (2006), D'Amico and King (2012), Hamilton and Wu (2012), and D'Amico *et al.* (2012). However, the literature relating the supply of Treasury bonds to yield spreads is still scarce. The earlier literature on a Treasury convenience yield was motivated by declining debt in the late 1990s. Reinhart *et al.* (2000) and Fleming (2000) hypothesized an increasing convenience yield with the debt pay-down and attributed it to higher liquidity and safety of Treasuries compared to other assets. Empirical evidence is presented by Cortes (2003) who observes a positive relation between the expected budget balance and swap spreads using monthly observations for 1994-2003. The swap spread is the difference between swap rates and government bond yields of same maturity. Thus, the more positive expectations, the smaller is the expected government bond issuance, the wider the swap rates. Longstaff (2004) uses monthly observations from 1991-2001 and finds a significant liquidity premium on Treasury bonds and, moreover, that it is negatively related to changes in the supply of Treasuries available to the investors. As a control variable, changes in default credit risk is, among others, included. Longstaff *et al.* (2005) find, using weekly observations for 2001-2002, that a large part of the yield spread between corporate bonds and Treasury bonds is caused by a component unrelated to default risk considerations of the investor. This indicates that the yield spread contains a convenience yield factor. Krishnamurthy and Vissing-Jørgensen (2012) argue that the convenience yield is isolated from the default risk component and present a theoretical model to explain the presence of a convenience yield consisting of a liquidity and a safety premium. Using annual data of different sample lengths they quantify a convenience yield consisting of a liquidity and

a safety premium and disentangle it from default risk. Finally, Sunderam (2013) presents empirical evidence using weekly data for 2001-2007 of a negative relationship between the supply of short-term maturity Treasury notes and the spread between the yield of asset-backed commercial papers and the Treasury notes.

Overall, the previous studies find evidence for a negative demand relation for Treasury bonds induced by a convenience yield. It has the main implication, as also noted in some of the papers, that the Federal Reserve (Fed) purchases of Treasury bonds will increase the yield spreads by disproportionately lowering Treasury yields. Moreover, this may suggest a muted transmission mechanism to the private sector and an adverse economic effect of a reduction in government debt.

Despite the high persistence in the yield spread series and the variable measuring the supply of Treasuries, none of the studies take these properties into account by modeling the dynamics of the relationship directly. More importantly, the nondefault component of the yield spread found in Longstaff *et al.* (2005) is shown to be time-varying. Hence, valuable information about the behavior of the convenience yield may be present in the short-run dynamics, which is the focus of this paper. An important advantage of using yield spreads in the analysis rather than yield levels is that the spreads are unaffected by shocks that equally affect both yields, e.g., expected inflation shocks. This facilitates economic interpretation of the structural shocks in the econometric analysis.

To specify the full dynamics of the convenience yield and the supply of Treasury bonds and, furthermore, model them jointly, we employ a vector autoregression (VAR) model and use a long sample dating from 1919 to 2008. By imposing an acceptable restriction on the cointegration rank, we identify the long-run negative demand relation and the equilibrating force in a vector error correction model (VECM) by Johansen (1996). Similar to Shapiro and Watson (1988), Blanchard and Quah (1989), King *et al.* (1991) we identify a just-identified structural model by constraining the long-run impact as well as the covariance matrix of the residuals. We estimate a structural VECM (SVECM) as a reduced rank VECM and then transform the system by post-multiplying the VECM by a rotation matrix that imposes the necessary restrictions on the covariance matrix. Impulse responses following a convenience yield demand shock and a Treasury supply shock are then computed to shed light on the short-run dynamics of the relationship. Next, we extend the model to include measures of each of the convenience yield components and repeat the structural analysis, albeit the short-run identification becomes more subtle. The structural analysis shows that for shorter horizons, 0-2 years, the supply of Treasury bonds is positively related to the convenience yield which encompasses a positive relation to the safety premium and an insignificant relation to the liquidity premium. This contrasts the long-run

negative relations and two premiums of almost equal size found in previous studies and also in this study. In addition, the transitory demand shocks to liquidity and safety are relatively long-lived and the effects die out after about four years for the spread variables and about six years for the Treasury supply variable. The results are shown to be robust to a potential kink in the demand relations estimated by a threshold VECM (TVECM), an influential observation and choice of identifying restriction. Hence, a reduction in government debt may not have adverse economic effects for the shorter horizons as first thought.

Section 2 introduces the statistical methodology and the structural model used to analyze the short-run effects. The data is then presented in section 3 before the dynamics of the total convenience yield are estimated in section 4. In section 5 the analysis continues by separating the convenience yield into a liquidity premium and safety premium. The robustness of the results are investigated in section 6 while section 7 concludes.

2. METHODOLOGY

The statistical framework for the analysis is the cointegrated VAR (CVAR) model of Johansen (1996) and Juselius (2006). By means of the CVAR we can analyze the short-run relationship between the yield spreads and the supply of Treasuries as well as the long-run relationship within the same framework. The baseline model for the analysis is the p -dimensional CVAR in VECM form, sometimes referred to as the reduced form VECM:

$$\Delta Z_t = \alpha \beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Z_{t-i} + \varepsilon_t, \quad t = 1, 2, \dots, T. \quad (2.1)$$

Z_t is a $p \times 1$ vector of variables, $Z_t = (z_{1t} \ z_{2t} \dots \ z_{pt})'$, and \bar{Z}_t is $(Z_t', 1)'$. α and β are matrices of dimension $p \times r$ and $(p+1) \times r$, respectively, with $\beta' \bar{Z}_t$ representing the $r \leq p$ cointegrating relationships and α giving the direction and speed of adjustment towards equilibrium. k is the number of lags in the corresponding VAR presentation and the autoregressive coefficients, $\Gamma_1, \dots, \Gamma_{k-1}$, are of dimension $p \times p$. For estimation, it is assumed that ε_t is an i.i.d. Gaussian sequence, $N_p(0, \Omega)$, while conditioning on the initial values, Z_{-k+1}, \dots, Z_0 .

2.1. THE GRANGER REPRESENTATION

To gain further insights into the dynamics of the process Z_t , the Granger Representation Theorem of Johansen, cf. Johansen (1995), is applied. It decomposes Z_t into $I(1)$ and $I(0)$ components. If $\alpha'_\perp (I_p - \sum_{i=1}^{k-1} \Gamma_i) \beta_\perp$ has full rank, $p - r$, the solution to (2.1) is

$$Z_t = C \sum_{i=1}^t \varepsilon_t + C^*(L)\varepsilon_t + \tilde{Z}_0 \quad (2.2)$$

where the long-run impact matrix is

$$C = \beta_{\perp} \left(\alpha'_{\perp} (I_p - \sum_{j=1}^{k-1} \Gamma_j) \beta_{\perp} \right)^{-1} \alpha'_{\perp}.$$

The rank of C is $p - r$, and hence when multiplied by the random walks $\sum_{i=1}^t \varepsilon_t$, $C \sum_{i=1}^t \varepsilon_t$ represents the $p - r$ stochastic trends driving the system. The initial values are contained in \tilde{Z}_0 while $C^*(L) = \sum_{i=0}^{\infty} C_i^* L^i$ is a convergent polynomial in the lag operator L , and, thus, $C^*(L)\varepsilon_t$ is an $I(0)$ process.

2.2. IMPULSE RESPONSES

The cointegration vectors give the long-run associations between the variables of the system. These long-run relationships, however, ignore all other relations between the variables which are only summarized in the VECM. The exact form of these relations is usually difficult to see directly from the coefficients. Therefore, impulse response functions are often computed. They represent the marginal responses of the endogenous variables of the system to an impulse in one of the endogenous variables. These may be regarded as conditional forecasts of the endogenous variables given that they have been zero up to the time when an impulse on one of the variables occurs.

The impulse responses can be computed from (2.2) by noticing that the coefficients to the convergent polynomial $C^*(L)$ can be recursively obtained from the equation

$$\Delta C_i^* = \alpha \beta' C_{i-1}^* + \sum_{j=1}^{k-1} \Gamma_j \Delta C_{i-j}^*, \quad i = 1, 2, \dots$$

with $C_0^* = I_p - C$ and $C_{-1}^* = \dots = C_{-k+1}^* = -C$, cf. Hansen (2005). The impulse response function of the reduced form model is then given by the coefficients $C_i = C + C_i^*$ which can be calculated by the recursion, cf. Johansen (2010)

$$C_i = \sum_{j=1}^k \Pi_j C_{i-j}, \quad \text{with initial values } C_0 = I_p, C_i = 0, i < 0 \quad (2.3)$$

where $\Pi_1 = I_p + \alpha \beta' + \Gamma_1$, $\Pi_i = \Gamma_i - \Gamma_{i-1}$, $i = 2, \dots, k-1$, $\Pi_k = -\Gamma_{k-1}$. Note that as the number of impulse responses, i , increases $C_i^* \rightarrow 0$ and $C_i \rightarrow C$.

2.2.1. INFERENCE

We base the statistical inference regarding the impulse responses on bootstrap methods, as first suggested by Runkle (1987). Various bootstrap approaches have been proposed for setting up confidence intervals for the impulse responses. We focus on Hall's percentile interval, cf. Hall (1992), which is pointed out by Benkwitz *et al.* (2001) to be advantageous due to the build-in bias-correction of the bootstrap distribution in case of biased impulse response estimates. The bootstrap samples are generated by resampling the (recentered) residuals with replacement, and the cointegration vectors are reestimated for each bootstrap sample. Alternatively, the estimation uncertainty arising from reestimation of the cointegration vectors could be ignored by fixing them at the superconsistently estimated vectors from the original sample. However, Benkwitz *et al.* (2001) argue that it makes little difference for the impulse response inference and may just cover up actual estimation uncertainty. For the estimated models below, no apparent differences in confidence intervals were observed when fixing the cointegration vector of the bootstrap samples.

2.3. SVECM

A SVECM provides a framework for a structural interpretation of the shocks that drive the system. In contrast to a VAR model, the estimated cointegration relationships impose constraints on the long-run impact matrix, C , in (2.2). This facilitates identification of the structural shocks because these long-run restrictions imply that the number of permanent and transitory components are already given when turning to the SVECM, see, e.g., Shapiro and Watson (1988) and King *et al.* (1991). Here, we focus on the just-identified structural model which is in fact only a rotation of the Granger Representation in (2.2).

The SVECM is given by

$$B\Delta Z_t = B\alpha\beta' \bar{Z}_{t-1} + \sum_{i=1}^{k-1} B\Gamma_i \Delta Z_{t-i} + B\varepsilon_t \quad (2.4)$$

with $\varepsilon \sim i.i.d(0, I_p)$. We introduce the structural shocks or innovations, denoted by μ_t , which are assumed to be related to the reduced-form disturbances, ε_t , by linear relations:

$$\mu_t = B\varepsilon_t \Leftrightarrow \varepsilon_t = B^{-1}\mu_t \quad (2.5)$$

where B is an invertible $p \times p$ matrix. The Granger Representation (2.2) is modified by

inserting (2.5)

$$Z_t = CB^{-1} \sum_{i=1}^t \mu_t + C^*(L)B^{-1}\mu_t + \tilde{Z}_0. \quad (2.6)$$

Let $R_i = R + R_i^*$ where $R = CB^{-1}$ is the structural long-run impact matrix and $R_i^* = C_i^* B^{-1}$. The impulse response function corresponding to the structural shocks becomes

$$R_i = \sum_{j=1}^k \Pi_j R_{i-j}. \quad (2.7)$$

The structural shocks are the central quantities in an SVECM and may be associated with an economic meaning. They are, however, not identified, and B must obey certain requirements to allow for a structural interpretation. How to choose B and achieve identification is the subtle part of the structural analysis.

2.3.1. IDENTIFICATION STRATEGY¹

By adding the rotation matrix B to (2.4) we have introduced $p \times p$ new parameters, and we then need to impose as many restrictions to reach a just-identified model, which is the case considered here. Moreover, a structural interpretation of the shocks requires that the rotation 1) separates the shocks with permanent effects from those with transitory effects; 2) identifies each individual shock; and 3) orthogonalizes the shocks to facilitate impulse response analysis.

To fix ideas, we need to find the rotation matrix B

$$B = H^{-1}EG,$$

where the matrices H , E and G each rotate the system to perform one of the three tasks needed to achieve an identified model with a structural interpretation of the shocks.

1. Separation The shocks are decomposed into two groups with the matrix G

$$G = \begin{pmatrix} \alpha' \Omega^{-1} \\ \alpha'_{\perp} \end{pmatrix}.$$

The first group consists of r shocks with transitory effects and the last group consists of $p - r$ shocks with permanent effects.

¹This section builds on Warne (1993) and Gonzalo and Ng (2001).

2. Identification Additional restrictions are often required because the assumption of orthogonal shocks introduced in step 3 below is not sufficient to achieve identification in many systems. Given a rank of r there can be at most $p - r$ shocks with transitory effects and at least r with permanent effects. Hence, $r(r - 1)/2$ additional restrictions are needed to exactly identify the transitory shocks and $(p - r)(p - r - 1)/2$ restrictions to identify the permanent shocks. These restrictions can be obtained from a “timing scheme” for the shocks. For such a scheme it is assumed that the shocks may affect a subset of variables directly within the current time period, whereas another subset of variables is affected with a time lag only. Given the aim of a just-identified structural model, a rotation which identifies the shocks can be achieved by pre-multiplying G with the inverse of full rank $p \times p$ matrix, E

$$E = \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix}.$$

The shocks with transitory effect are identified with E_1 , which is an $r \times r$ matrix. It consists of the r first elements of the rows in G^{-1} where unit vectors are wanted. Similarly, the shocks with permanent effect are identified with E_2 , which is $(p - r) \times (p - r)$ and it consists of the last $p - r$ elements of the rows of the (until this step) long-run impact matrix CG^{-1} .

Observe that step 1 and 2 imply that the structural covariance matrix $EG\Omega G'E'$ is block-diagonal

$$\begin{aligned} EG\Omega G'E' &= \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix} \begin{pmatrix} \alpha' \Omega^{-1} \\ \alpha'_{\perp} \end{pmatrix} \Omega \begin{pmatrix} \Omega^{-1} \alpha & \alpha_{\perp} \end{pmatrix} \begin{pmatrix} E'_1 & 0 \\ 0 & E'_2 \end{pmatrix} \\ &= \begin{pmatrix} E_1 \alpha' \Omega^{-1} \\ E_2 \alpha'_{\perp} \end{pmatrix} \Omega \begin{pmatrix} \Omega^{-1} \alpha E'_1 & \alpha_{\perp} E'_2 \end{pmatrix} \\ &= \begin{pmatrix} E_1 \alpha' \Omega^{-1} \Omega \Omega^{-1} \alpha E'_1 & E_1 \alpha' \Omega^{-1} \Omega \alpha_{\perp} E'_2 \\ E_2 \alpha'_{\perp} \Omega \Omega^{-1} \alpha E'_1 & E_2 \alpha'_{\perp} \Omega \alpha_{\perp} E'_2 \end{pmatrix} \\ &= \begin{pmatrix} E_1 \alpha' \Omega^{-1} \alpha E'_1 & 0 \\ 0 & E_2 \alpha'_{\perp} \Omega \alpha_{\perp} E'_2 \end{pmatrix}. \end{aligned}$$

3. Orthogonalization The final requirement for the rotation matrix is that the shocks are uncorrelated. This is needed to consider the dynamic impact of an isolated shock. If the shocks were correlated, we should take into account the relationship between the shocks. The shocks $EG\varepsilon_t$ can be orthogonalized by means of a Cholesky decomposition that solves

$$HH' = EG\Omega G'E'.$$

Given the block-diagonal structure of $EG\Omega G'E'$, the Cholesky decomposition results in

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \quad \text{and} \quad H^{-1} = \begin{pmatrix} H_1^{-1} & 0 \\ 0 & H_2^{-1} \end{pmatrix}$$

where H_1 and H_2 are lower triangular matrices resulting from the Cholesky decomposition of $E_1\alpha'\Omega^{-1}\alpha E_1'$ and $E_2\alpha'_\perp\Omega\alpha_\perp E_2'$, respectively.

The final rotation matrix is then $B = H^{-1}EG$. Observe that it also normalizes the variances and standard deviations of the structural shocks to unity

$$B\Omega B' = H^{-1}EG\Omega G'E'(H^{-1})' = I_p.$$

Moreover, note that the Cholesky decomposition is not the single tool used for identification as shocks are also identified by step 2. Consequently, the impulse responses will not depend on the ordering of the variables in the system.

To sum up, a total of $p \times p$ restrictions are needed to just-identify the structural model. Given the cointegration rank, only r shocks have transitory effects and $r(p-r)$ restrictions are thus already imposed on the rotation matrix. Hence, $p^2 - r(p-r)$ restrictions are left. Step 2 imposes $r(r-1)/2 + (p-r)(p-r-1)/2 = \frac{1}{2}p(p-1) - r(p-r)$ restrictions, whereas the final orthogonalization takes up $p(p+1)/2$ restrictions. This adds up to the $p^2 - r(p-r)$ restrictions.

Since the just-identified structural model is only a rotation of the Granger Representation, efficient estimates of the model can be calculated in two steps: first, the reduced form VECM is estimated imposing the cointegration restriction and possibly also short-run restrictions as well. Second, the estimated reduced rank VECM is transformed into the structural model using the three steps given above.

3. DATA

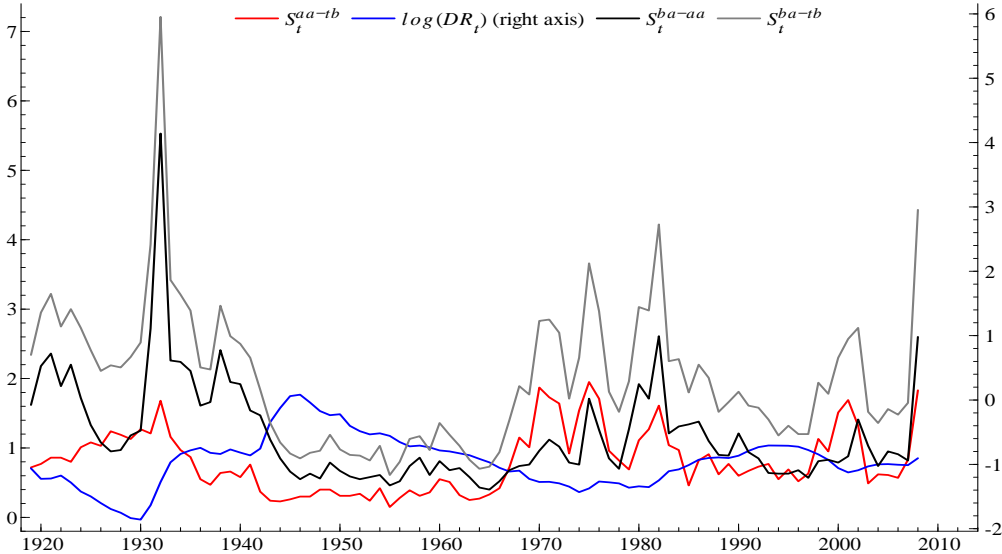
We consider a total of four variables, all sampled at an annual frequency from 1919-2008. The three yield spreads are: the spread between the yield of Moody's AAA-rated corporate bonds and the yield of U.S. Treasury bonds, s_t^{aa-tb} ; the spread between the yield of Moody's BAA-rated corporate bonds and the yield of U.S. Treasury bonds, s_t^{ba-tb} ; and the spread between the yield of Moody's BAA-rated corporate bonds and the yield of Moody's AAA-rated corporate bonds, s_t^{ba-aa} . All bonds are long-maturity bonds. The series are obtained from the Federal Reserve's FRED database², except for the yields on Treasury bonds

²Series AAA, BAA and LTGOVTBD and GS20

for the years 1919-1924 which are obtained from Banking and Monetary Statistics, 1914-1941, table 128.³ The yield spreads are sampled to match the fiscal year; that is in July of the year up to and including 1976 and in October of each year after that. The fourth variable is the (log of) U.S. government debt-to-GDP, ℓDR_t , measuring the supply of Treasury bonds. Debt is ultimo government's fiscal year and GDP is for the same fiscal year. The series is from Henning Bohn's web page and constructed for Bohn (2008).⁴ The data set is identical to the one used in Krishnamurthy and Vissing-Jørgensen (2012), except for debt-to-GDP, here kept in face value.⁵

Figure 3.1 shows the development of the variables over the period. The variables ap-

Figure 3.1: Yield spreads and log of debt-to-GDP, 1919-2008.



pear quite persistent, in particular ℓDR_t , and it is not obvious that any of them are characterized by stationary processes. The rank tests of the VECMs will perform formal tests of the order of integration of the variables. The influence of the large spike in the spread series in 1932 during the crises of the 1930s is investigated as a robustness check of the results.

The basic unrestricted estimations and tests below are carried out with CATS in RATS, RATS (2009). All additional estimations and tests are computed using MATLAB, MATLAB

³<http://fraser.stlouisfed.org/publication/?pid=38>
⁴<http://www.econ.ucsb.edu/~bohn/data.html>
⁵Krishnamurthy and Vissing-Jørgensen (2012) report that they get similar results whether debt-to-GDP is in face value or in market value.

(2010).

4. THE TOTAL CONVENIENCE YIELD

To analyze the relationship between Treasury supply and the convenience yield, we estimate a two-dimensional VECM for $Z_t = (S_t^{ba-tb} \quad \ell DR_t)'$. The yield spread between BAA-rated corporate bonds and Treasury bonds includes both liquidity and safety considerations of the investor, thus, measuring the full effect of a convenience yield. Part of the yield spread is, however, also due to a default risk on the corporate bonds. Nevertheless, we do not include a measure hereof in the analysis because the relationship appears robust in shorter samples to different measures of default risk as well as to the state of the business cycle, see Longstaff (2004) and Krishnamurthy and Vissing-Jørgensen (2012). In addition, no data is available for the entire sample that can match the specification of the other series.

The unrestricted VECM appears well-specified with one lag, see table A.1 and A.2 in appendix. None of the variables are trending, as seen in figure 3.1, and, thus, only a restricted constant is included in the deterministic specification. The result of the trace test by Johansen (1995) is shown in table 4.1 and suggests one cointegration relation in the system, $r = 1$. Hence, both variables are non-stationary, but a linear combination of them is stationary. I.e., the two variables co-move over the period and are driven by the same underlying stochastic trend. The final model is presented in (4.1) (t -values in parenthesis)

Table 4.1: Trace test for rank determination with a restricted constant.

	Trace test	Bartlett corrected
$H(0) H(2)$	25.19[0.01]	23.81[0.01]
$H(1) H(2)$	6.14[0.19]	5.93[0.20]

Note: The asymptotic distribution of the test statistic is non-standard and approximated by the Gamma distribution, see Doornik (1998). The Bartlett corrected tests are computed using Bartlett small sample corrections of Johansen (2002).

and misspecification tests are reported in (4.2).

$$\begin{pmatrix} \Delta S_t^{ba-tb} \\ \Delta \ell DR_t \end{pmatrix} = \begin{pmatrix} -0.50 \\ (-6.82) \\ 0 \end{pmatrix} \begin{pmatrix} S_{t-1}^{ba-tb} & +1.60 \ell DR_{t-1} & -0.45 \\ & (4.53) & (-1.26) \end{pmatrix}$$

$$+ \begin{pmatrix} 0 & 0 \\ 0.04 & 0.66 \\ (3.42) & (8.89) \end{pmatrix} \begin{pmatrix} \Delta S_{t-1}^{ba-tb} \\ \Delta \ell DR_{t-1} \end{pmatrix} \quad (4.1)$$

$$\hat{\Omega} = \begin{pmatrix} 0.455 & \\ 0.016 & 0.006 \end{pmatrix}, \text{corr}(\hat{\varepsilon}_t) = \begin{pmatrix} 1 & \\ 0.320 & 1 \end{pmatrix} \quad (4.2)$$

The zero restrictions on the short-run dynamics are simultaneously accepted with a p -

Table 4.2: Misspecification tests of the two-dimensional VECM.

	ΔS_t^{ba-tb}	$\Delta \ell DR_t$
<i>AR</i> (1 – 2)	0.74[0.69]	3.47[0.18]
<i>ARCH</i> (1 – 2)	1.86[0.40]	1.72[0.42]
<i>Normality</i>	52.35[0.00]	25.34[0.00]

Note: The misspecification tests are LM tests and χ^2 -distributed with two degrees of freedom. The normality test is that of Doornik and Hansen (2008).

value of 0.99. The rejection of normality of the residuals is caused by a few large observations during the crises of the 1930's and the World War 2 period. These observations are not corrected for by dummy variables since the number of observations is limited and the lack of normality has no crucial effect on the further analysis.

The estimated cointegration coefficient to ℓDR_t implies that a decrease of one standard deviation in debt-to-GDP from its mean value of 0.419 to 0.234 is, in the long run, associated with an increase in the convenience yield as measured by S_t^{ba-tb} of 93 basis points.⁶ Note that the total supply of Treasury bonds, as measured by the variable ℓDR_t , is only approximating the supply of long-maturity Treasury bonds and, hence, the effect related to the long-maturity yield spread may be imprecisely estimated. Nevertheless, the result suggests that when the supply of Treasury bonds falls the yields on both BAA-rated corporate bonds and Treasury bonds fall. However, over the longer run, the Treasury yield falls by more than the corporate bond yield, implying an increase in the spread, and, thus, a larger convenience yield. The effect found here is somewhat larger than the one found in a cross-sectional analysis of Krishnamurthy and Vissing-Jørgensen (2012). Although insignificant, the constant term reflects that at zero supply, a demand for the attributes of Treasury bonds is still present and induces a positive spread.

⁶The effect is computed as $1.598(\log(0.419) - \log(0.234)) = 0.933$.

Deviations from the long-run relationship are corrected by changes in S_t^{ba-tb} and the adjustment is quite fast as 50% of the deviation in the previous period is corrected in the current period. ℓDR_t does not error-correct and is, thus, weakly exogenous. This implies that shocks to ℓDR_t have permanent affect on both variables of the system, while shocks to S_t^{ba-tb} have only transitory effect on the system. The instantaneous residual correlation in (4.2) is moderate in size, and, hence, to perform the structural analysis the orthogonalization of shocks becomes crucial for interpretation.

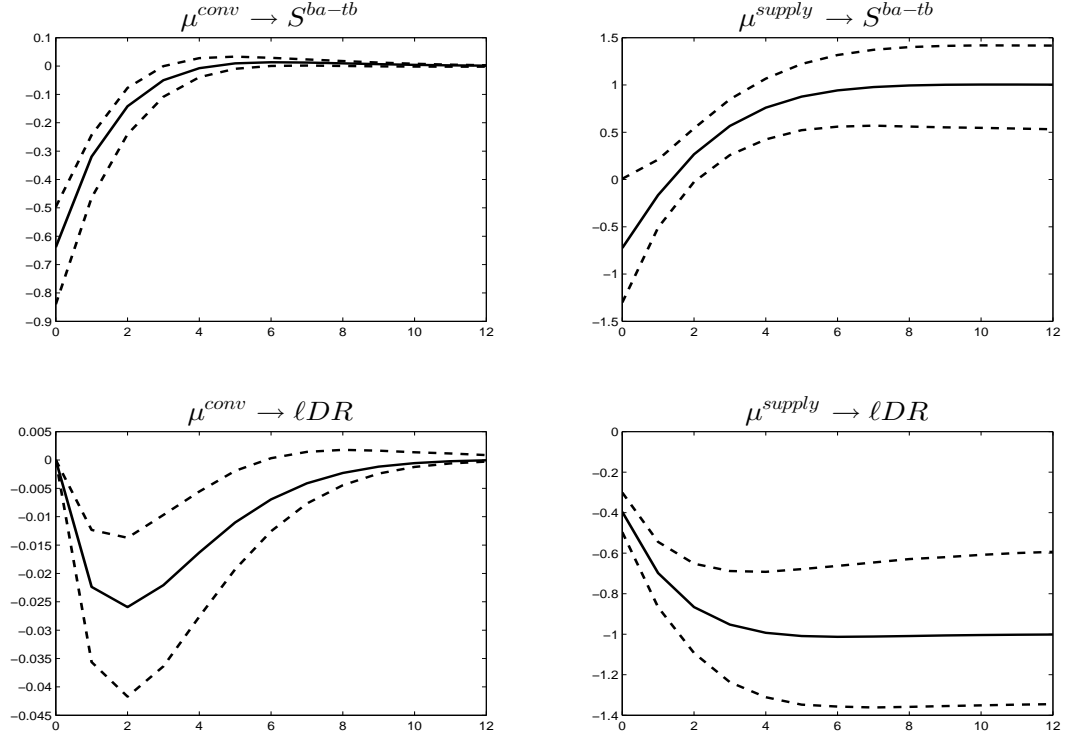
4.1. STRUCTURAL IMPULSE RESPONSE ANALYSIS

The above analysis concerned the long-run relationship between the variables $Z_t = (S_t^{ba-tb} \ \ell DR_t)'$. To analyze the behavior of the variables in the short-run, we construct structural impulse responses. Identification of the structural shocks is in this two-dimensional system given by the cointegration rank, $r = 1$, and the weak exogeneity of ℓDR_t : the system is hit by a temporary shock when the shock is to S_t^{ba-tb} and a permanent shock when the shock is to ℓDR_t . Besides the orthogonalization of two shocks, no further restrictions are required to identify the shocks because $r(r-1)/2 = (p-r)(p-r-1)/2 = 0$. Moreover, we label the shocks as a convenience yield demand shock (or a flight-to-liquidity-and-safety shock) and a Treasury supply shock, $\mu_t = (\mu_t^{conv}, \mu_t^{supply})'$. The long-run impact matrix, $R = CB^{-1}$, and rotation matrix, B , are given by:

$$R = \begin{bmatrix} 0 & * \\ 0 & * \end{bmatrix}, B^{-1} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix}. \quad (4.3)$$

Asterisks denote unrestricted elements. The zero column in R reflects the transitory effects of a convenience yield demand shock. The zero in the first column of B^{-1} is induced by the weak exogeneity of ℓDR_t and implies that a shock to the convenience yield has no instantaneous impact on the supply of Treasury bonds. The estimated structural impulse responses from a negative shock to the system are displayed in figure 4.1. A negative shock is defined as a fall in μ_t . The impulse responses are normalized such that the permanent effects are unity. The confidence bands (dashed lines) are 90% Hall's percentile intervals, Hall (1992), as discussed in section 2.2.1, and thereby corresponding to a one-sided 5% test. The bands are based on 2000 bootstrap replications. The impulse response of S_t^{ba-tb} following a permanent Treasury supply shock, $\mu_t^{supply} \rightarrow S_t^{ba-tb}$, is interesting because the short-run effect has opposite sign of the long-run effect. On impact S_t^{ba-tb} decreases by almost three quarters of the total long-run effect. Two years after the negative shock, the

Figure 4.1: Impulse responses from negative shocks to the system $Z_t = (S_t^{ba-tb} \quad \ell DR_t)'$.



effect turns positive. The confidence interval, however, includes zero the first two years after the initial effect. The estimated long-run impact and rotation matrix are given in (4.4).⁷ The instantaneous effects in the second column of \hat{B}^{-1} are normalized on the permanent effect similar to the impulse responses in figure (4.1).

$$\hat{R} = \begin{bmatrix} 0 & -1.298 \\ & (-3.69) \\ 0 & 0.187 \\ & (-3.55) \end{bmatrix}, \hat{B}^{-1} = \begin{bmatrix} 0.639 & 0.725 \\ (6.20) & (1.78) \\ 0 & 0.396 \\ & (6.86) \end{bmatrix}. \quad (4.4)$$

Given that \hat{B}^{-1} is constructed from a Cholesky decomposition of $\hat{\Omega}$ it is invariant to column-wise sign changes. Hence, the sign of the impulse responses arising from μ_t^{conv} are not identified, and we can only infer that a convenience yield demand shock affects the two variables in the system in the same direction. More importantly, these transitory effects die out relatively slowly. Zero is not included in the confidence interval until about four years after for S_t^{ba-tb} and about six years after for ℓDR_t . The impact on ℓDR_t is rather

⁷The t -values in the brackets are derived using bootstrapped standard deviations of the estimates as suggested by Lütkepohl and Krätzig (2004).

small and reaching a maximum effect of 0.025% after two years. The slower reversion of ℓDR_t is nevertheless noteworthy and may be a result of the high persistence characterizing the process.

Combining the results of the VECM with the impulse response analysis suggests two things. First, the estimated cointegration vector conforms that the long-run demand curve for Treasury bonds is indeed downward sloping reflecting a convenience yield. This suggests an isolated adverse effect on the economy of an otherwise expansionary decrease in the supply of Treasury bonds. The entire fall in the yield of Treasury bonds is not transmitted to the private sector, such that the price of raising capital for the sector does not decrease with the same effect as for the government. The second thing to observe is that this adverse effect is not present at shorter horizons. In the short-run, the demand curve for Treasury bonds is upward sloping or at best flat. When a negative shock hits Treasury supply, the yield spread reduces on impact and enhances the expansionary effect on the economy. The effect on the spread is not significantly positive until three years after the shock.

One possible explanation for the observed opposite short-run effect may be that a negative supply shock is followed by a negative demand shock that is unrelated to the convenience yield. This moves the entire demand function down such that the net effect is a fall in the spread. Over time the demand function moves up again and the lower supply then raises the premium on Treasury bonds. The theoretical model of Krishnamurthy and Vissing-Jørgensen (2012) offers another explanation. If the fall in Treasury supply is accompanied by an increase in the supply of private substitutes resulting in a larger total holdings of assets providing convenience yield services, the convenience yield will fall because the marginal utility of the representative investor is decreasing in total holdings.

To find out whether it is the demand for liquidity, safety or both that is driving the dynamics of the total convenience yield, a three-dimensional VECM is estimated next.

5. LIQUIDITY AND SAFETY PREMIUMS

To split the convenience yield into a liquidity premium and a safety premium, two simplifying assumptions are needed. First, we assume that AAA-rated corporate bonds and Treasury bonds provide similar safety for an investor, and the spread between these bond yields therefore measures the premium investors are willing to pay (in terms of receiving a lower yield) for the excess liquidity of Treasury bonds. Second, we assume that the liquidity of BAA-rated and AAA-rated corporate bonds is alike and the spread between the yield of these bonds includes a safety premium. The later assumption is empirical valid

based on the results of Chen *et al.* (2007) who find similar liquidity for the two types of corporate bonds. We estimate the relationship between each of the two components of the convenience yield and Treasury supply measured by debt-to-GDP by estimating a three-dimensional VECM for $Z_t = (S_t^{aa-tb} \ S_t^{ba-aa} \ \ell DR_t)'$.

As for the two-dimensional VECM, the unrestricted model appears dynamically well-specified with one lag, see table B.1 and B.2 in appendix. Table 5.1 reports the results of the trace test. The test suggests a rank of two, $r = 2$. The estimated reduced rank VECM

Table 5.1: Trace test for rank determination with a restricted constant.

	Trace test	Bartlett corrected
$H(0) H(3)$	40.63[0.01]	37.81[0.02]
$H(1) H(3)$	20.40[0.05]	19.29[0.06]
$H(2) H(3)$	6.30[0.17]	6.01[0.20]

Note: The asymptotic distribution of the test statistic is non-standard and approximated by the Gamma distribution, see Doornik (1998). The Bartlett corrected tests are computed using Bartlett small sample corrections of Johansen (2002).

is given by (5.1) (t -values in parenthesis). We specify the two cointegration relations as reflecting a demand relation for each of the two premiums in the convenience yield. Note that the cointegration matrix, β , is just-identified such that the imposed restrictions on β are non-testable.

$$\begin{pmatrix} \Delta S_t^{aa-tb} \\ \Delta S_t^{ba-aa} \\ \Delta \ell DR_t \end{pmatrix} = \begin{pmatrix} -0.35 & 0 \\ (-4.11) & \\ 0 & -0.42 \\ & (-5.69) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} S_{t-1}^{aa-tb} & +0.76\ell DR_{t-1} & -0.11 \\ (3.65) & & (-0.52) \\ S_{t-1}^{ba-aa} & +0.83\ell DR_{t-1} & -0.36 \\ (2.59) & & (-1.08) \end{pmatrix} + \begin{pmatrix} 0 & 0 & -0.59 \\ & & (-2.23) \\ 0 & 0 & 0 \\ 0 & 0.05 & 0.64 \\ & (4.08) & (8.92) \end{pmatrix} \begin{pmatrix} \Delta S_{t-1}^{aa-tb} \\ \Delta S_{t-1}^{ba-aa} \\ \Delta \ell DR_{t-1} \end{pmatrix} \quad (5.1)$$

$$\hat{\Omega} = \begin{pmatrix} 0.069 & & \\ 0.056 & 0.280 & \\ 0.001 & 0.015 & 0.005 \end{pmatrix}, \text{corr}(\hat{\varepsilon}_t) = \begin{pmatrix} 1 & & \\ 0.405 & 1 & \\ 0.078 & 0.394 & 1 \end{pmatrix} \quad (5.2)$$

The zero-coefficient restrictions on the short-run dynamics are simultaneously accepted

Table 5.2: Misspecification tests of the three-dimensional VECM.

	ΔS_t^{aa-tb}	ΔS_t^{ba-aa}	$\Delta \ell DR_t$
<i>AR</i> (1 – 2)	5.72[0.06]	4.29[0.12]	3.62[0.16]
<i>ARCH</i> (1 – 2)	0.34[0.85]	5.98[0.05]	1.71[0.43]
<i>Normality</i>	10.69[0.01]	73.08[0.00]	23.21[0.00]

Note: The misspecification tests are LM tests and χ^2 -distributed with two degrees of freedom. The normality test is that of Doornik and Hansen (2008).

with a p -value of 0.93. The implications of the estimated cointegration relationships are that a one standard deviation in debt-to-GDP from its mean value of 0.419 to 0.234 is in the long run associated with an increase in S_t^{aa-tb} and S_t^{ba-aa} of 44 and 49 basis points, respectively. Thus, the total convenience yield is in the long run close to being equally divided between the two premiums. For comparison, Krishnamurthy and Vissing-Jørgensen (2012) find a liquidity premium of about twice the size of the safety premium. To see how the two premiums behave for shorter horizons, we perform a structural analysis below.

The two spreads error-correct deviations from their own cointegration relationship and the mechanism is again relatively fast. Observe that, as for the two-dimensional VECM, ℓDR_t does not error-correct deviations from the two cointegration relations such that shocks to this variable have permanent effect on the levels of all the variables. Again, the residuals are correlated, see (5.2), and orthogonalization of the shocks is vital for the structural analysis.

5.1. STRUCTURAL IMPULSE RESPONSE ANALYSIS

To investigate the short-run behavior of the liquidity and safety premium, we perform a structural impulse response analysis similar to the one above for the two-dimensional VECM. However, identification of the structural shocks is now not implicitly given by the rank of the VECM because the system contains an additional variable. Consequently, identification and the choice of rotation matrix follows from the steps outlined in section 2.3.1. Given a rank of $r = 2$ and the weak exogeneity of ℓDR_t , the shocks to S_t^{aa-tb} and S_t^{ba-aa} have only transitory effect whereas the shock to ℓDR_t has permanent effect. To identify the transitory shocks, we need additional $r(r - 1)/1 = 1$ restriction while the permanent shock is already identified, $(p - r)(p - r - 1)/2 = 0$. The chosen rotation matrix B and the

long-run impact matrix $R = CB^{-1}$ are given by (5.3)

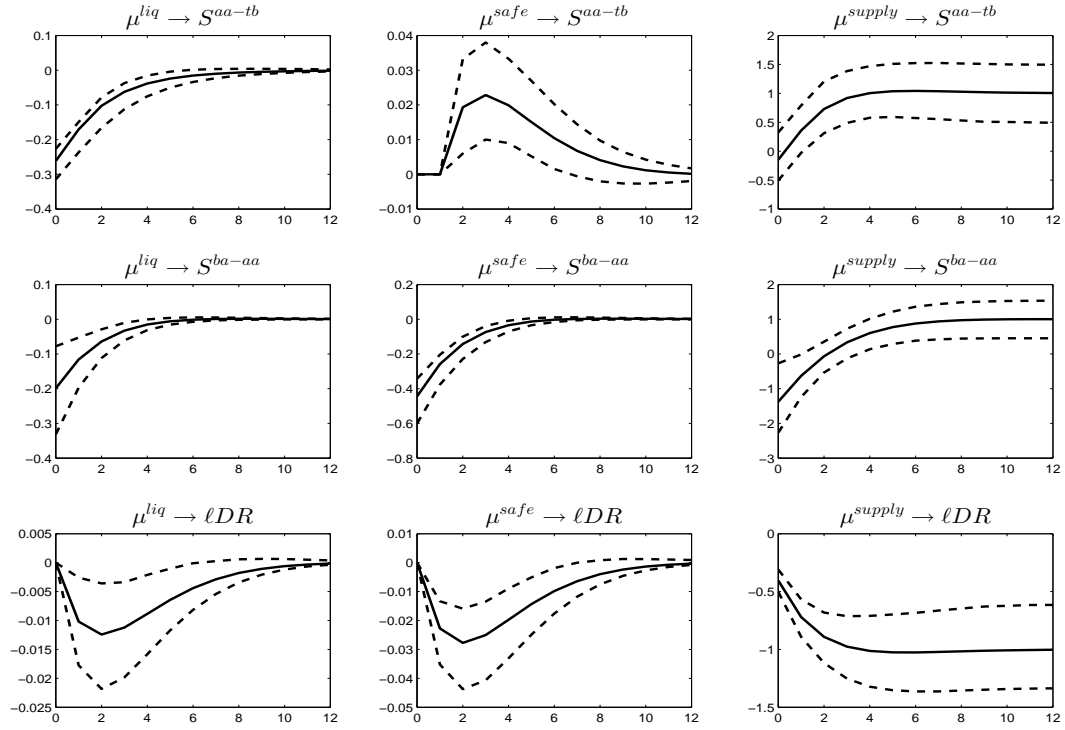
$$R = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ 0 & 0 & * \end{bmatrix}, B^{-1} = \begin{bmatrix} * & 0 & * \\ * & * & * \\ 0 & 0 & * \end{bmatrix} \quad (5.3)$$

Hence, B , separates, identifies and orthogonalizes the structural shocks which we label as demand shocks to liquidity and safety and a Treasury supply shock, $\mu_t = (\mu_t^{liq}, \mu_t^{safe}, \mu_t^{supply})'$. The two zero columns in R reflect the transitory effect of the two shocks, μ_t^{liq} and μ_t^{safe} . The two zeroes in the final row of B^{-1} show that a shock to either of the spreads has no instantaneous effect on the weakly exogenous variable ℓDR_t . The additional restriction needed is imposed via the zero in the first row of B^{-1} in (5.3). It implies that a safety demand shock has no instantaneous effect on S_t^{aa-tb} . This restriction relies on the initial assumption that the spread between yields on corporate AAA-rated bonds and Treasury bonds is due to a liquidity premium on Treasury bonds as they provide similar safety for an investor.

The estimated structural impulse responses following a negative shock are given in figure 5.1. The negative shock is again defined as a fall in μ_t . The impulse responses are again normalized such that the permanent effects are unity and the confidence bands are computed as 90% Hall's percentile intervals with bootstrap replications. The impulse responses of the permanent Treasury supply shock to the two spreads, $\mu_t^{supply} \rightarrow S_t^{aa-tb}$ and $\mu_t^{supply} \rightarrow S_t^{ba-aa}$, respectively, reveal instantaneous effects of opposite sign to the long-run effect. While the initial effect in period 0 on S_t^{aa-tb} is close to zero, as also reflected by the low t -value in (5.4) below, the effect on S_t^{ba-aa} is large and zero is only borderline included in the confidence interval. Moreover, the initial fall in S_t^{ba-aa} is larger in absolute value than the long-run effect. This suggests that the safety premium on Treasury bonds is the most sensitive component of the convenience yield. From period 1 – 5, zero is included in the confidence interval for the impulse responses of $\mu_t^{supply} \rightarrow S_t^{ba-aa}$, before turning positive thereafter.

As for the two-dimensional VECM, the transitory effects in figure 5.1 are only identified up to a sign. The impulse responses for $\mu_t^{liq} \rightarrow S_t^{aa-tb}$ show a moderate effect the first two years after the shock. The effects of the two transitory shocks on ℓDR_t persist for about ten years, although zero is included in the confidence interval after about six years. The effects are, however, small in size. The consequence of the additional restriction needed for identification of the two transitory shocks is seen in the graph for $\mu_t^{safe} \rightarrow S_t^{aa-tb}$ in figure 5.1. The initial zero effect in period 0 is transferred to period 1 because ΔS_{t-1}^{ba-aa} has

Figure 5.1: Impulse responses from negative shocks to the system $Z_t = (S_t^{aa-tb} \ S_t^{ba-aa} \ \ell DR_t)'$.



no significant effect on ΔS_t^{aa-tb} , as seen in (5.1). In section 6.3, we investigate the consequences of the placement of the identifying restriction. The estimated long-run impact matrix and rotation matrix, with the latter normalized on the long-run effect, are given by (t -values in parenthesis)

$$\hat{R} = \begin{bmatrix} 0 & 0 & -0.138 \\ & & (-3.29) \\ 0 & 0 & -0.152 \\ & & (-3.02) \\ 0 & 0 & 0.182 \\ & & (3.11) \end{bmatrix}, \hat{B}^{-1} = \begin{bmatrix} 0.263 & 0 & 0.148 \\ (9.65) & & (0.59) \\ 0.200 & 0.447 & 1.375 \\ (2.59) & (5.58) & (2.21) \\ 0 & 0 & 0.404 \\ & & (7.10) \end{bmatrix}. \quad (5.4)$$

The t -values in (5.4) are computed like those of (4.4) with bootstrap methods. Only the instantaneous effect of the shock $\mu_t^{supply} \rightarrow S_t^{aa-tb}$ in the upper right of \hat{B}^{-1} turns out to be insignificant with a small t -value.

The results of the three-dimensional VECM add several aspects to the relationships. First, a fall in the Treasury supply is in the long run associated with an increase in both spreads of almost equal size. Second, the same does not hold for shorter horizons. A negative Treasury supply shock reduces both spreads on impact. Third, the safety spread,

S^{ba-aa} , is more sensitive in the short run than the liquidity spread, S^{aa-tb} , relative to their individual long-run effects. This stands in stark contrast to the almost equal sized long-run effects. Hence, it is primarily the safety premium that drives the result for the total convenience yield found in section 4.1. The safety premium falls the first two years after a negative Treasury supply shock. This suggests that an unexpected fall in the supply of Treasury bonds provides an expansionary effect for the economy by lowering the safety premium and it lasts for about two years before the long-run adverse effect takes over and raises the premium compared to the pre-shock level.

A small liquidity effect and a dominant safety effect is in line with the findings from an event-study analysis of the effects of Fed policy carried out by Krishnamurthy and Vissing-Jørgensen (2011), but, importantly, with opposite sign of the effects. In addition, Fleming (2000) suggests that the liquidity premium may decrease following a fall in the supply of Treasury bonds, as found here, because the lower supply reduces the actual liquidity in the market, and, thus, also the willingness to pay for the liquidity attribute of Treasury bonds. The reduction in the safety premium may be caused by increasing expectations of better economic outlook following the initial reduction in Treasury supply.

6. SENSITIVITY ANALYSIS

In this section we investigate the robustness of the results from the three-dimensional VECM towards changes in the specification and identification.

6.1. A THRESHOLD ALTERNATIVE

Reinhart *et al.* (2000) and Krishnamurthy and Vissing-Jørgensen (2012) argue that there may be a kink in the demand for Treasury bonds because some investors have an inelastic demand, e.g., foreign official institutions that place great importance on liquidity and safety. This implies that increasing Treasury supply can not drive the two spreads all the way to zero and the relationships will be different at the lower bound of the spreads. To investigate how this potential kink in the demand affects the dynamics of the relationships, a two-regime TVECM, originating in Balke and Fomby (1997), is estimated. This model allows the error-correction coefficients to switch between two regimes depending on whether the (lagged) level of Treasury supply is above or below an estimated threshold value. The cointegration matrix is kept constant across regimes such that any regime differences show up in the error-correction coefficients only. Thereby, we do not estimate a demand relation for each regime, which is deemed too ambitious given the relatively few

observations for estimating a threshold model.

The TVECM is given by

$$\begin{aligned}\Delta Z_t &= \alpha^{low} \beta' \bar{Z}_{t-1} I(\ell DR_{t-1} \leq c) + \alpha^{high} \beta' \bar{Z}_{t-1} I(\ell DR_{t-1} > c) \\ &\quad + \sum_{i=1}^{k-1} \Gamma_i \Delta Z_{t-i} + \eta_t, \quad t = 1, 2, \dots, T\end{aligned}\tag{6.1}$$

where $I(A)$ is an indicator function such that $I(A) = 1$ if A is true and otherwise $I(A) = 0$, and η_t is an i.i.d. Gaussian sequence, $N_p(0, \Omega)$. The definitions of β , \bar{Z}_t and Z_t are unchanged from the VECM in (2.1). The threshold parameter, c , determines the level of Treasury supply where the model switches to the other regime. The likelihood function, L_T , to be maximized is (up to the scale of a constant)

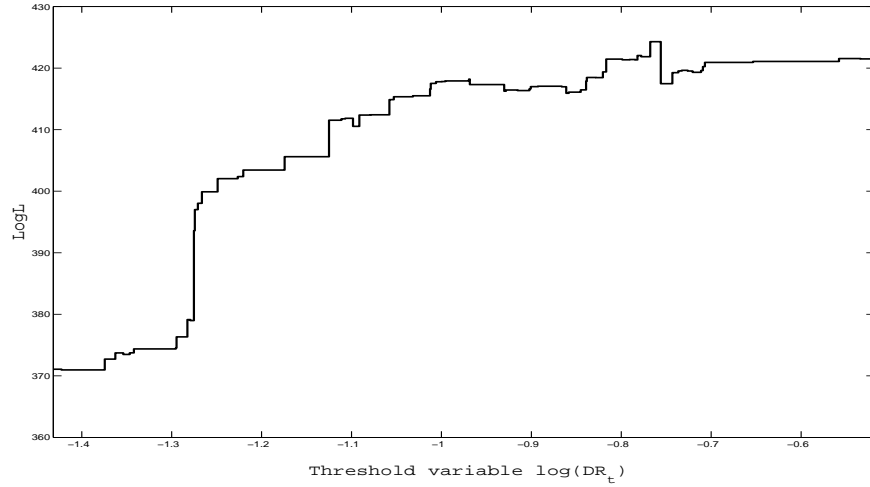
$$L_T(\theta) = -\frac{T}{2} \log |\hat{\Omega}(\theta)|\tag{6.2}$$

where $\hat{\Omega}(\theta) = \frac{1}{T} \sum_{t=1}^T \eta_t(\theta) \eta_t(\theta)'$ is the maximum likelihood estimator of the covariance matrix Ω and

$$\begin{aligned}\eta_t(\theta) &= \Delta Z_t - \alpha^{low} \beta' \bar{Z}_{t-1} I(\ell DR_{t-1} \leq c) - \alpha^{high} \beta' \bar{Z}_{t-1} I(\ell DR_{t-1} > c) \\ &\quad - \sum_{i=1}^{k-1} \Gamma_i \Delta Z_{t-i}.\end{aligned}$$

The parameter vector to be estimated is given by $\theta = (\alpha^{low}, \alpha^{high}, \beta', \Gamma_1, \dots, \Gamma_{k-1}, c)$. Maximum likelihood estimation of the TVECM is non-standard because the likelihood function in (6.2) is not smooth due to the threshold specification. This implies that conventional derivative based optimization algorithms are not suitable for its maximization. The non-differentiability of the likelihood function is illustrated in figure 6.1 which shows the profiled likelihood function for the threshold parameter, c , over values of the threshold variable, ℓDR_t . However, conditional on the threshold parameter, the model can be estimated by maximum likelihood. As a result, it is standard in the literature on threshold models to estimate the threshold parameter using a grid, see, e.g., Hansen and Seo (2002). In the case considered here with a single regime indicator, namely, the level of Treasury supply, the grid search for a value of the threshold parameter is performed over values of ℓDR_t , disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of ℓDR_t . The observations are in effect split into two subsets and the assumption

Figure 6.1: Profiled likelihood function for the threshold parameter, c , over values of the threshold variable ℓDR_t .



about the threshold value is located between the two percentiles ensures that a sufficient number of observations are present in each regime.

The TVECM is estimated using initial values from the VECM and with $k = 2$ and $r = 2$. It results in (t -values in parenthesis)

$$\begin{aligned}
 \begin{pmatrix} \Delta S_t^{aa-tb} \\ \Delta S_t^{ba-aa} \\ \Delta \ell DR_t \end{pmatrix} &= \begin{pmatrix} -0.36 & 0 \\ (-3.8) & \\ 0 & -0.43 \\ & (-5.8) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} S_{t-1}^{aa-tb} & +1.09\ell DR_{t-1} & +0.18 \\ (3.69) & (0.57) \\ S_{t-1}^{ba-aa} & +0.97\ell DR_{t-1} & -0.20 \\ (2.59) & (-0.35) \end{pmatrix} I(\ell DR_{t-1} \leq -0.77) \\
 &+ \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -0.21 & 0.12 \\ (-3.2) & (2.35) \end{pmatrix} \begin{pmatrix} S_{t-1}^{aa-tb} & +1.09\ell DR_{t-1} & +0.18 \\ (3.69) & (0.57) \\ S_{t-1}^{ba-aa} & +0.97\ell DR_{t-1} & -0.20 \\ (2.59) & (-0.35) \end{pmatrix} I(\ell DR_{t-1} > -0.77) \\
 &+ \begin{pmatrix} 0 & 0 & -0.53 \\ & & (-2.0) \\ 0 & 0 & 0 \\ 0 & 0.05 & 0.60 \\ & (4.4) & (8.4) \end{pmatrix} \begin{pmatrix} \Delta S_{t-1}^{aa-tb} \\ \Delta S_{t-1}^{ba-tb} \\ \Delta \ell DR_{t-1} \end{pmatrix} \tag{6.3} \\
 \hat{\Omega} &= \begin{pmatrix} 0.068 & & \\ 0.053 & 0.279 & \\ 0.001 & 0.015 & 0.005 \end{pmatrix}, \text{corr}(\hat{\varepsilon}_t) = \begin{pmatrix} 1 & & \\ 0.383 & 1 & \\ 0.071 & 0.416 & 1 \end{pmatrix}
 \end{aligned}$$

The zero-coefficient restrictions on the short-run dynamics are simultaneously accepted

with a p -value of 0.24. The misspecification tests are reported in table 6.1. The test for no

Table 6.1: Misspecification tests of TVECM.

	ΔS_t^{aa-tb}	ΔS_t^{ba-aa}	$\Delta \ell DR_t$
<i>AR</i> (1 – 2)	2.71[0.26]	6.65[0.04]	3.94[0.14]
<i>ARCH</i> (1 – 2)	0.50[0.78]	7.89[0.02]	3.53[0.02]
<i>Normality</i>	12.02[0.00]	65.23[0.00]	25.61[0.00]

Note: The misspecification tests are LM tests and χ^2 -distributed with two degrees of freedom. The normality test is that of Doornik and Hansen (2008). The test for no autocorrelation is augmented with regressors from both regimes to take the regime structure into account. The tests for no-ARCH and normality are unchanged from the VECM.

autocorrelation is slightly modified to take the regime structure into account. The t -values in parenthesis in (6.3) are computed using standard errors obtained from the square root of the diagonal of the inverted Hessian matrix. They should, however, be interpreted with caution as no formal distributional theory exists for the parameter estimates and standard errors of this model. In addition, no standard error of the threshold estimate, \hat{c} , is reported due to the non-standard inference on the threshold parameter, cf. Chan (1993) and Hansen (1997). Moreover, given that the value of the likelihood function is constant between the values in the grid, as observed in figure 6.1, the threshold estimate could be reported as an interval.

The estimated cointegration relationships from the TVECM enhance the liquidity and safety premiums. A fall of one standard deviation in debt-to-GDP from its mean value of 0.419 to 0.234 is in the long run associated with an increase in S_t^{aa-tb} and S_t^{ba-aa} of 61 and 57 basis points, respectively. The estimated threshold value of -0.77 corresponds to a debt-to-GDP ratio of 0.46 and can be interpreted as the location of the kink in the demand functions. Krishnamurthy and Vissing-Jørgensen (2012) find a slightly higher ratio for the kink in their piece-wise linear demand function. The threshold level is graphed in figure 6.2 together with the threshold variable determining the regime switches. 64 out of 88 effective observations are located in the lower regime with $\text{debt-to-GDP} \leq 0.46$ which characterizes the more normal times and is similar to the estimated VECM in (5.1). As evident from figure 6.2, the TVECM mainly isolates the observations/years with high debt-to-GDP ratios during the 1940s and 1950s in the upper regime. These observations are the most extreme sets of observations in the data set and reflect high debt levels associated

Figure 6.2: Estimated threshold value and threshold variable ℓDR_t .



with low spreads (see also figure 3.1).⁸ Consequently, any economic interpretation of the error-correction mechanism of this regime may be spurious.

Nevertheless, a way to perceive the lower regime of the estimated TVECM is as an outlier corrected version of the VECM in (5.1). The TVECM excludes the extreme sets of observations in the lower end of the demand relations by placing them in a separate regime with a very distinct error-correction mechanism. As a result, the two long-run demand functions are estimated with steeper slopes than in the VECM.

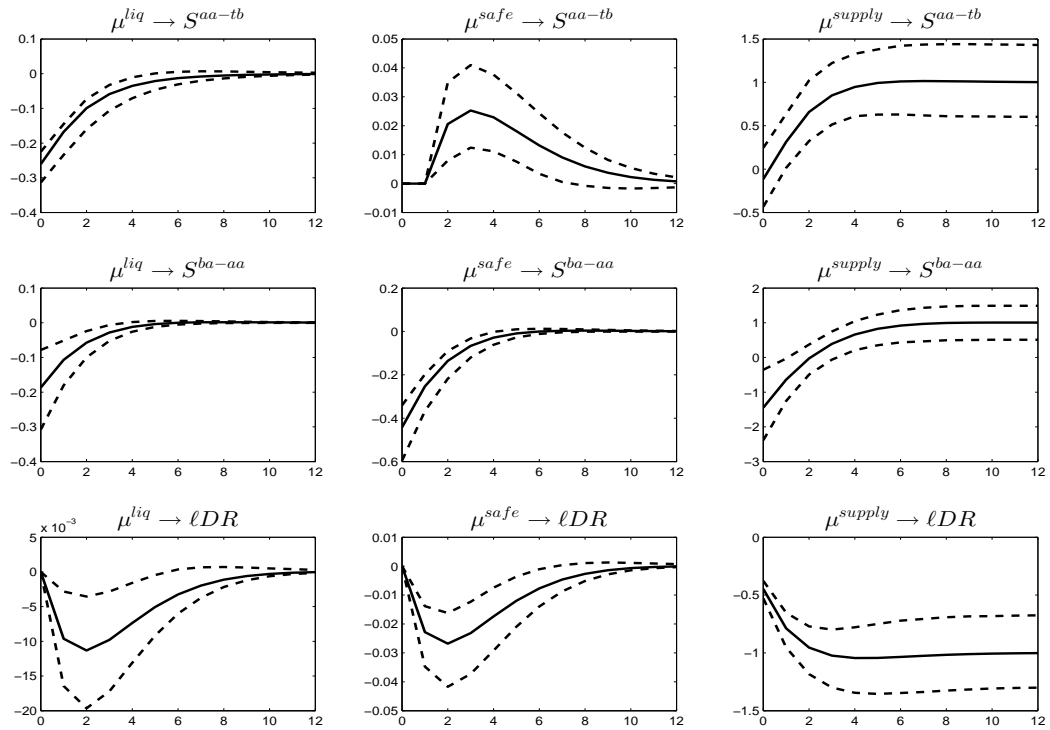
We continue the analysis by inspecting the structural impulse responses of the lower regime of the TVECM for comparability with those of the VECM in figure (5.1). Figure 6.3 shows the impulse responses to the lower regime following a negative shock to each of the variables in the system, and (6.4) displays the estimated long-run impact and rotation matrix B , which gives the instantaneous effects:

$$\hat{R} = \begin{bmatrix} 0 & 0 & -0.162 \\ 0 & 0 & -0.152 \\ 0 & 0 & 0.156 \end{bmatrix}, \hat{B}^{-1} = \begin{bmatrix} 0.261 & 0 & 0.115 \\ 0.187 & 0.442 & 1.448 \\ 0 & 0 & 0.447 \end{bmatrix}. \quad (6.4)$$

As seen from figure 6.3 and (6.4) the effects are close to the effects in the VECM. Relative to their long-run effects, the effect of a negative supply shock reduces the safety premium slightly more than in the VECM while the liquidity premium is less reduced.

⁸During this period the Fed was committed to keep the long-term interest rate low to help finance World War 2 by purchasing long-term Treasury bonds. Nonetheless, the Fed ended up accumulating only a small portion of long-term Treasury bonds, because of an announced commitment to keep yields low, and large purchases of short-term Treasuries, cf. Bernanke (2002).

Figure 6.3: Impulse responses from negative shocks to the lower regime of TVECM.



The results of the TVECM suggest that a steeper demand function for safety does not alter the short-run positive relationship. The previous result of a decreasing safety premium following a fall in the supply of Treasury bonds is, thus, robust to the extreme observations during and after the World War 2 period.

6.2. INFLUENTIAL OBSERVATIONS

The large spike in 1932 observed in the levels of the spread series in figure 3.1 may affect the estimates and the dynamics. We investigate whether this is the case by including a dummy variable to account for the extraordinary large shock in 1932 during the recession. The dummy is included as an additive outlier because no level shifts are observed in the spread series implying no transmission of the shock through the autoregressive structure of the model, cf. Juselius (2006) and Nielsen (2004). This implies that the dummy variable is included with the full lag structure of the endogenous variables.

The model is estimated with $k = 2$ and $r = 2$ and is given by (6.5) and (6.6) (t -values in

parenthesis)

$$\begin{pmatrix} \Delta \widetilde{S}_t^{aa-tb} \\ \Delta \widetilde{S}_t^{ba-aa} \\ \Delta \ell \widetilde{DR}_t \end{pmatrix} = \begin{pmatrix} -0.36 & 0 \\ (-4.30) & \\ 0 & -0.28 \\ & (-4.44) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \widetilde{S}_{t-1}^{aa-tb} & +0.75\ell \widetilde{DR}_{t-1} & +0.11 \\ (3.81) & & (0.56) \\ \widetilde{S}_{t-1}^{ba-aa} & +0.80\ell \widetilde{DR}_{t-1} & +0.36 \\ (2.39) & & (1.07) \end{pmatrix} \\ + \begin{pmatrix} 0 & 0 & -0.47 \\ & & (-1.83) \\ 0 & 0 & 0 \\ & & \\ 0 & 0.10 & 0.69 \\ & (4.84) & (9.68) \end{pmatrix} \begin{pmatrix} \Delta \widetilde{S}_{t-1}^{aa-tb} \\ \Delta \widetilde{S}_{t-1}^{ba-aa} \\ \Delta \ell \widetilde{DR}_{t-1} \end{pmatrix} \quad (6.5)$$

$$\begin{pmatrix} \widetilde{S}_t^{aa-tb} \\ \widetilde{S}_t^{ba-aa} \\ \ell \widetilde{DR}_t \end{pmatrix} = \begin{pmatrix} S_t^{aa-tb} \\ S_t^{ba-aa} \\ \ell DR_t \end{pmatrix} - \begin{pmatrix} 0.38 \\ (1.77) \\ 2.84 \\ (10.77) \\ 0.02 \\ (-0.74) \end{pmatrix} D_t^{1932} \quad (6.6)$$

$$\hat{\Omega} = \begin{pmatrix} 0.066 & & \\ 0.037 & 0.133 & \\ 0.001 & 0.001 & 0.005 \end{pmatrix}, \text{corr}(\hat{\varepsilon}_t) = \begin{pmatrix} 1 & & \\ 0.393 & 1 & \\ 0.041 & 0.357 & 1 \end{pmatrix}$$

The dummy variable is defined as $D_t^{1932} = 1$ for $t = 1932$ and 0 otherwise. Similar to the TVECM the t -values are computed using standard errors obtained from the square root of the diagonal of the inverted Hessian matrix. Note that the estimators related to the dummy variable are not consistent as no additional information about the outlier is gained when increasing the number of observations. Moreover, the reported t -values in (6.6) are not valid for testing. The restrictions implied by (6.6) minimize the degrees of freedom used to approximate the outlier to $p = 3$ parameters. Without the restrictions, the dynamics of the outlier would be approximated by $p(k - 1) = 6$ parameters for the lag structure and with $r = 2$ cointegration parameters, a total of 8 parameters. For larger samples, the difference between the estimated models with and without the restrictions (6.6) may be negligible, but given the small sample of this application, we prefer the restricted version. The misspecification tests are reported in table 6.2. Observe that the test for no-ARCH effects in the S_t^{ba-aa} equation is now accepted with much larger p -value. The slopes of the two demand relations imply that a fall in debt-to-GDP of one standard deviation is associated with increases in S_t^{aa-tb} and S_t^{ba-aa} of 44 and 47 basis points, respectively. Hence, this single observation accounts for a reduction in the associated safety premium change of 2 basis points compared to the model without a dummy variable.

The short-run relationships are by and large unaffected by the dummy variable. The

Table 6.2: Misspecification tests of the three-dimensional VECM with a dummy variable.

	ΔS_t^{aa-tb}	ΔS_t^{ba-aa}	$\Delta \ell DR_t$
<i>AR</i> (1 – 2)	3.08[0.21]	7.81[0.02]	2.82[0.26]
<i>ARCH</i> (1 – 2)	0.73[0.70]	0.27[0.87]	1.33[0.51]
<i>Normality</i>	11.40[0.00]	43.40[0.00]	26.74[0.00]

Note: The misspecification tests are LM tests and χ^2 -distributed with two degrees of freedom. The normality test is that of Doornik and Hansen (2008).

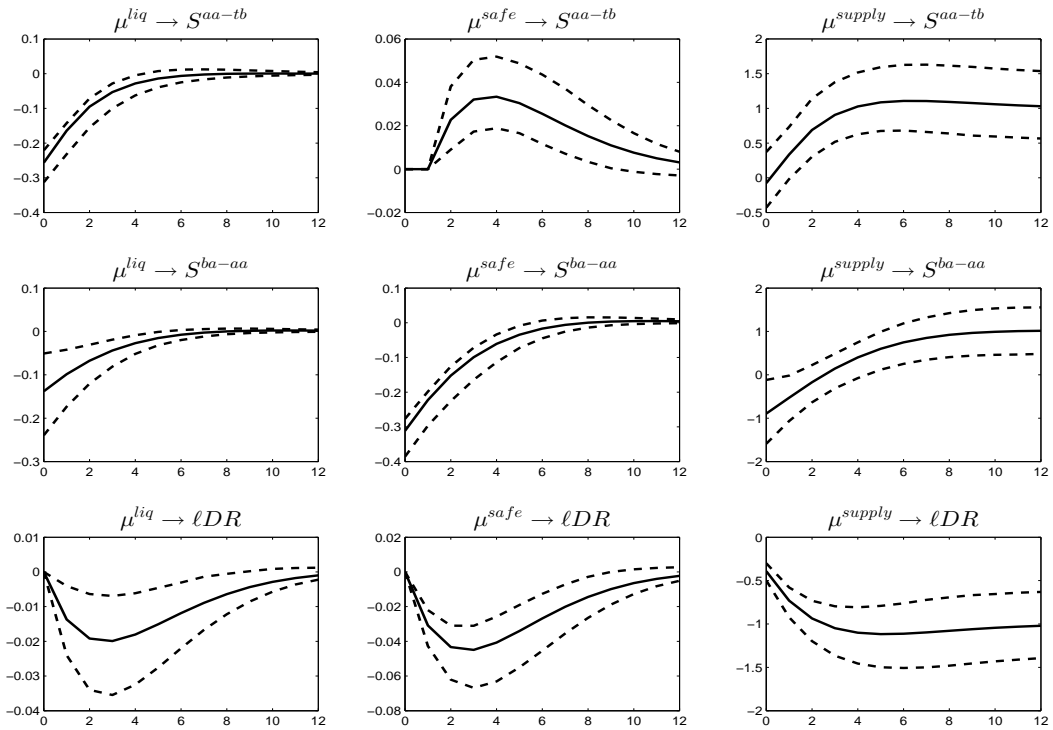
estimated structural impulses and rotation matrix are showed in figure 6.4 and (6.7). The instantaneous effects on the spreads are still of opposite sign to the long-run effect, but somewhat smaller in magnitude than for the model without the dummy variable.

$$\hat{R} = \begin{bmatrix} 0 & 0 & -0.137 \\ & & (-3.47) \\ 0 & 0 & -0.146 \\ & & (-3.07) \\ 0 & 0 & 0.183 \\ & & (4.38) \end{bmatrix}, \hat{B}^{-1} = \begin{bmatrix} 0.256 & 0 & 0.077 \\ (9.10) & & (0.32) \\ 0.138 & 0.311 & 0.895 \\ (2.41) & (9.44) & (2.02) \\ 0 & 0 & 0.392 \\ & & (6.70) \end{bmatrix} \quad (6.7)$$

6.3. IDENTIFICATION STRATEGY

Recall that the impulse responses following the two transitory demand shocks, μ_t^{liq} and μ_t^{safe} , are identified by imposing a zero restriction in the rotation matrix such that μ_t^{safe} has no instantaneous impact on S_t^{aa-tb} . Alternatively, we can restrict μ_t^{liq} to have no instantaneous impact on S_t^{ba-aa} . Imposing both restrictions simultaneously results in an over-identified model which is not pursued here. Changing the identifying restriction to the alternative gives the impulse responses shown in figure C.1 and instantaneous effects in (C.1), both appearing in the appendix. Note, that the impacts of the permanent supply shock are unchanged from the VECM in section 5 because of the orthogonalization of permanent and transitory shocks. The liquidity shock, μ_t^{liq} , now only affects S_t^{aa-tb} because the combination of the identifying restriction and the zero restrictions in the short-run dynamics imply that the shock is not transmitted to the other two variables. The effects are indeed similar to those with the initial identifying restriction, except for the now unrestricted instantaneous effect of $\mu_t^{safe} \rightarrow S_t^{aa-tb}$, which induces changes in S_t^{aa-tb} in the same direction as for the other two variables.

Figure 6.4: Impulse responses from negative shocks to the VECM with dummy variable.



7. CONCLUSION

The presence of a demand driven convenience yield on U.S. Treasury bonds suggests that decreases in the supply of Treasury bonds are associated with increases in the spread between yields on corporate bonds and Treasury bonds. We show that this adverse economic effect of, e.g., an unexpected debt pay-down or Quantitative Easing is only a long-run effect. The dynamics of the yield spread and the supply of Treasury bonds, measured as (log of) debt-to-GDP, are modeled jointly in a reduced rank VECM, which gives a long-run negative relationship. This relationship is interpretable as a downward-sloping demand relation for Treasury bonds. By means of a just-identified SVECM, we present evidence that for shorter horizons, the relationship is positive, and, thus, with opposite sign than in the long-run. An unexpected fall in Treasury supply is on a horizon of 0-2 years associated with a reduction in the convenience yield. When extending the model to separate the convenience yield into a liquidity and safety premium, the safety premium turns out to be the driving factor of the short-run positive relationship. The results are robust to a potential kink in the demand relation for Treasury bonds estimated by a TVECM and also

an influential observation during the recession of the 1930s. Hence, the economic effects following a fall in Treasury supply may not be as adverse as first thought.

REFERENCES

- BALKE, N. S. AND T. B. FOMBY (1997): "Threshold Cointegration." *International Economic Review*, 38(3):627–45.
- BENKWITZ, A., H. LÜTKEPOHL, AND J. WOLTERS (2001): "Comparison of bootstrap confidence intervals for impulse responses of German monetary systems." *Macroeconomic Dynamics*, 5(1):81–100.
- BERNANKE, B. S. (2002): "Deflation: making sure 'it' doesn't happen here." *Remarks before the National Economists Club, Washington, DC*, 21.
- BLANCHARD, O. J. AND D. QUAH (1989): "The dynamic effects of Aggregate Demand and Supply Disturbances." *American Economic Review*, 79:655–673.
- BOHN, H. (2008): "The Sustainability of Fiscal Policy in the United States, in: R. Neck and J. Sturm, "Sustainability of Public Debt." *MIT Press*, pages 15–49.
- CHAN, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model." *The Annals of Statistics*, 21(1):520–533.
- CHEN, L., D. A. LESMOND, AND J. WEI (2007): "Corporate yield spreads and bond liquidity." *The Journal of Finance*, 62(1):119–149.
- CORTES, F. (2003): "Understanding and modelling swap spreads." *Bank of England Quarterly Bulletin*, Winter.
- D'AMICO, S., W. ENGLISH, D. LÓPEZ-SALIDO, AND E. NELSON (2012): "The Federal Reserve's Large-scale Asset Purchase Programmes: Rationale and Effects." *The Economic Journal*, 122(564):F415–F446.
- D'AMICO, S. AND T. B. KING (2012): "Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply." *Journal of Financial Economics*.
- DOORNIK, J. A. (1998): "Approximations to the asymptotic distributions of cointegration tests." *Journal of Economic Surveys*, 12(5):573–593.
- DOORNIK, J. A. AND H. HANSEN (2008): "An Omnibus Test for Univariate and Multivariate Normality." *Working Paper, Nuffield College, Oxford*.
- FLEMING, M. J. (2000): "Financial Market Implications of the Federal Debt Paydown." *Brookings Papers on Economic Activity 2: 2000*, page 221.

- GONZALO, J. AND S. NG (2001): “A systematic framework for analyzing the dynamic effects of permanent and transitory shocks.” *Journal of Economic Dynamics and Control*, 25(10):1527–1546.
- HALL, P. (1992): *The bootstrap and Edgeworth expansion*. Springer.
- HAMILTON, J. D. AND J. C. WU (2012): “The effectiveness of alternative monetary policy tools in a zero lower bound environment.” *Journal of Money, Credit and Banking*, 44(s1):3–46.
- HANSEN, B. (1997): “Inference in TAR models.” *Studies in nonlinear dynamics and econometrics*, 2(1):1–14.
- HANSEN, B. AND B. SEO (2002): “Testing for two-regime threshold cointegration in vector error-correction models.” *Journal of Econometrics*, 110(2):293–318.
- HANSEN, P. R. (2005): “Granger’s representation theorem: A closed-form expression for I(1) processes.” *The Econometrics Journal*, 8(1):23–38.
- JOHANSEN, S. (1995): “Likelihood-based inference in cointegrated vector autoregressive models.” *New York*.
- (1996): “Likelihood-Based Inference in Cointegrated Vector Autoregressive Model.” *Advanced Texts in Econometrics, Oxford University Press*.
- (2002): “A small sample correction for the test of cointegrating rank in the vector autoregressive model.” *Econometrica*, 70(5):1929–1961.
- (2010): “Some identification problems in the cointegrated vector autoregressive model.” *Journal of Econometrics*, 158(2):262–273.
- JUSELIUS, K. (2006): “The Cointegrated VAR Model: Methodology and Applications.” *Oxford University Press*.
- KING, R. G., C. I. PLOSSER, J. H. STOCK, AND M. W. WATSON (1991): “Stochastic trends and economic fluctuations.” *The American Economic Review*, pages 819–840.
- KRISHNAMURTHY, A. AND A. VISSING-JØRGENSEN (2011): “The effects of quantitative easing on interest rates: channels and implications for policy.” *National Bureau of Economic Research*.

- (2012): “The aggregate demand for treasury debt.” *Journal of Political Economy*, 120(2):233–267.
- KUTTNER, K. (2006): “Can central banks target bond prices?” *National Bureau of Economic Research*.
- LONGSTAFF, F. A. (2004): “The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices.” *The Journal of Business*, 77(3):511–526.
- LONGSTAFF, F. A., S. MITHAL, AND E. NEIS (2005): “Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market.” *The Journal of Finance*, 60(5):2213–2253.
- LÜTKEPOHL, H. AND M. KRÄTZIG (2004): *Applied Time Series Econometrics*. Cambridge University Press.
- MATLAB (2010): *version 7.10.0 (R2013a)*. Natick, Massachusetts, USA: The MathWorks Inc.
- NIELSEN, H. B. (2004): “Cointegration analysis in the presence of outliers.” *Econometrics Journal*, 7:249–271.
- RATS (2009): *version 7.20*. Evanston, Illinois, USA: Estima.
- REINHART, V., B. SACK, AND J. HEATON (2000): “The economic consequences of disappearing government debt.” *Brookings Papers on Economic Activity*, 2000(2):163–220.
- RUNKLE, D. E. (1987): “Vector autoregressions and reality.” *Journal of Business & Economic Statistics*, 5(4):437–442.
- SHAPIRO, M. AND M. WATSON (1988): “Sources of business cycles fluctuations.” In: *NBER Macroeconomics Annual 1988, Volume 3*, pages 111–156. MIT Press.
- SUNDERAM, A. (2013): “Money Creation and the Shadow Banking System.” *Harvard Business, Mimeo*.
- WARNE, A. (1993): “A Common Trends Model : Identification, Estimation and Inference.” *Working Paper, Institute for International Economic Studies, Stockholm University*, (555):43.

A. MISSPECIFICATION TESTS AND LAG-LENGTH DETERMINATION OF THE UNRESTRICTED TWO-DIMENSIONAL VECM

Table A.1: Misspecification tests of the two-dimensional VECM, unrestricted.

	ΔS_t^{ba-tb}	$\Delta \ell DR_t$
<i>AR</i> (1 – 2)	0.70[0.70]	3.31[0.19]
<i>ARCH</i> (1 – 2)	2.02[0.36]	1.78[0.41]
<i>Normality</i>	41.75[0.00]	27.04[0.00]

Note: The misspecification tests are LM tests and χ^2 -distributed with two degrees of freedom. The normality test is that of Doornik and Hansen (2008).

Table A.2: Information criteria and LR tests of lag-length, two-dimensional VECM.

	$k = 1$	$k = 2$	$k = 3$
<i>SC</i>	-5.359	-5.668	-5.494
<i>H – Q</i>	-5.462	-5.840	-5.735
<i>LR</i> ($k = 3 k = 2$)		2.99[0.56]	
<i>LR</i> ($k = 2 k = 1$)		44.03[0.00]	

Note: *SC* and *H-Q* are the Schwartz and the Hannan-Quinn information criterion, respectively. The p -value of the LR tests are from the $\chi^2(4)$ -distribution. The models are estimated with the same number of observations.

B. MISSPECIFICATION TESTS AND LAG-LENGTH DETERMINATION OF THE UNRESTRICTED THREE-DIMENSIONAL VECM

Table B.1: Misspecification tests of the three-dimensional VECM, unrestricted.

	ΔS_t^{aa-tb}	ΔS_t^{ba-aa}	$\Delta \ell DR_t$
<i>AR</i> (1 – 2)	3.63[0.16]	1.11[0.57]	2.95[0.23]
<i>ARCH</i> (1 – 2)	0.19[0.91]	6.02[0.05]	1.54[0.46]
<i>Normality</i>	10.56[0.01]	52.08[0.00]	24.68[0.00]

Note: The misspecification tests are LM tests and χ^2 -distributed with two degrees of freedom. The normality test is that of Doornik and Hansen (2008).

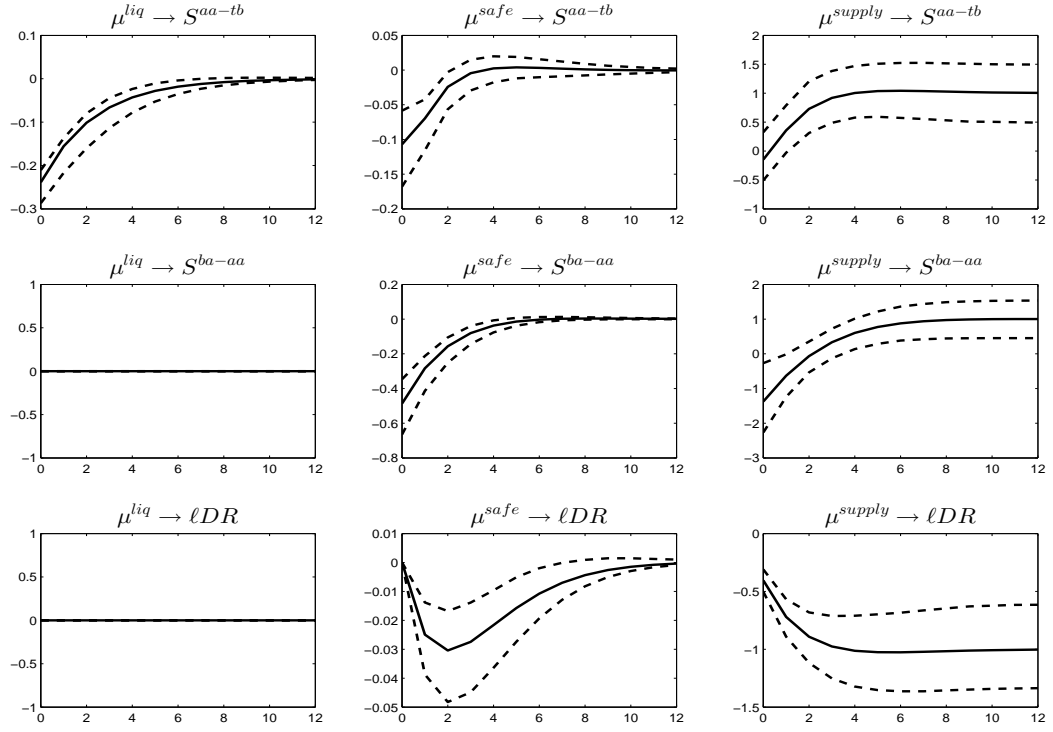
Table B.2: Information criteria and LR tests of lag-length, three-dimensional VECM.

	$k = 1$	$k = 2$	$k = 3$
<i>SC</i>	–8.502	–8.562	–8.228
<i>H – Q</i>	–8.708	–8.923	–8.743
<i>LR</i> ($k = 3 k = 2$)		11.55[0.24]	
<i>LR</i> ($k = 2 k = 1$)		45.10[0.00]	

Note: SC and H-Q are the Schwartz and the Hannan-Quinn information criterion, respectively. The p -value of the LR tests are from the $\chi^2(9)$ -distribution. The models are estimated with the same number of observations.

C. ALTERNATIVE IDENTIFICATION OF TRANSITORY SHOCKS

Figure C.1: Impulse responses from negative shocks to the system $Z_t = (S_t^{aa-tb} \ S_t^{ba-aa} \ \ell DR_t)'$.



$$\hat{B}^{-1} = \begin{bmatrix} 0.239 & 0.107 & 0.148 \\ (10.33) & (3.26) & (0.59) \\ 0 & 0.487 & 1.375 \\ & (4.99) & (2.21) \\ 0 & 0 & 0.404 \\ & & (7.10) \end{bmatrix}. \quad (C.1)$$