Autoregressive conditional tail behavior and results on Government bond yield spreads

Niklas Wagner*

Center for Entrepreneurial and Financial Studies CEFS, Munich University of Technology,
Munich 80290, Germany

Abstract

Previous evidence in empirical finance indicates the potential usefulness of modeling time variation particularly in the tails of speculative return distributions. Based on results from extreme value theory, the present paper proposes a fixed changepoint Pareto-type autoregressive conditional tail (ARCT) model. Regression-based parameter estimation of the unobservable time-varying tail index is carried out via classical Kalman filtering. A model application highlights the tail index dynamics for daily changes in Government bond yield spreads between the U.S. Dollar and the Euro zone.

© 2004 Elsevier Inc. All rights reserved.

JEL classification: C16; C22; C53; G10
Keywords: Tail index; Pareto model; Autoregressive conditional tail; Extreme financial risk; Spread risk

1. Introduction

The forecasting of large price movements in economics and finance is surely a difficult, while at the same time, central issue in financial management. This explains why the prediction of risk—frequently associated with the prediction of return volatility—has been the subject of a vast number of papers within the empirical finance literature. Despite huge advances in the area during the last two decades—including the modeling of generalized autoregressive conditional heteroskedasticity (GARCH)—several authors recently put some doubt on the applicability of available
volatility models for risk management and for extreme price movements in particular. As such, Shephard (1996, p. 21) notes that “GARCH models cannot deal with the extremely large movements in financial markets, even though they are good models of changing variance.” Danielsson and Morimoto (2000, p. 15) suggest that “the wild swings observed in the GARCH VaR [Value-at-Risk] predictions are more of an artifact of the GARCH model rather than the underlying data.” Engle (2000, p. 2) notes “it is not clear whether the tails have the same dynamic behavior as the rest of the distribution as would be assumed by GARCH style models.” Considering higher conditional moments, Rockinger and Jondeau (2002, p. 140) conclude that it seems that “there is little evidence that skewness and kurtosis are dependent on past returns. One possible reason for this finding is that these moments are driven by extreme realizations that occur only infrequently.” At the same time, empirical findings suggest the modeling of time variation in the conditional tails of return distributions. Quintos, Fan, and Phillips (2001, p. 634) summarize earlier work by noting that “there is a consensus from past empirical research that the tail behavior of certain financial series are time varying.” Along with the authors mentioned above, Christoffersen and Diebold (2000, p. 21) point out that “it seems […] that all models miss the really big movements […], and ultimately the really big movements are the most important for risk management. This suggests the desirability of directly modeling the extreme tails of return densities […].”

Given the above statements among others, the present paper aims at going in the indicated direction by proposing a model based on autoregressive conditional tail (ARCT) behavior. The route taken follows the one directed by extreme value theory, i.e., focus is put on the tail of the conditional return distribution function. Instead of modeling conditional return variance, taking volatility clustering into account or modeling conditional quantiles as was done previously, we suggest modeling a time-varying conditional tail index. The so-called tail index is known to be a key parameter which characterizes tail behavior in extreme value theory. As supported by theoretical arguments and empirical evidence, the assumption is made that the time-varying distributions are fat tailed throughout; that is, they all belong to the maximum domain of attraction of the Fréchet-type extreme value distribution. Consequently, a Pareto-type ARCT model results. Introducing fixed changepoints which are evenly spread through time, model estimation is based on a regression equation which is derived from the empirical distribution function. Although other estimation techniques found wider application in extreme value statistics, the static regression approach is known for long; recently, related statistical results were outlined, for example, in Datta and McCormick (1998). Van den Goorbergh (1999) provides a financial application. We choose the approach for three reasons. First, it follows immediately from the

---

assumption of fat-tailedness; second, it has a simple graphical implication; and lastly, it allows us to introduce a dynamic state space formulation for the unobservable time-varying tail index. Parameter estimation can then be carried out via standard Kalman filtering.

By now, no framework was proposed which includes a conditional time-series model for the dynamics of the tail index. While evidence on time-varying volatility is vast, approaches to time-varying tails were considered in a few financial studies so far. Tsay (1999) applies the time-varying inhomogeneous Poisson model by Smith (1989) in a study of S&P 500 index returns and reports a tendency for increasingly fat tails during the 1962–1997 period. Smith (2000) proposes a random changepoint model with time-varying parameters for S&P 500 index returns. Quintos et al. (2001) derive a statistical test of tail index stationarity based on a rolling series of so-called Hill estimators. Their analysis documents structural breaks in the tail index of emerging stock index returns during the 1997 Asian crisis. A related study is by Galbraith and Zernov (2004); using daily U.S. stock market returns during the period 1960–2002, they document increased tail fatness for the subperiod starting in the mid-1980s. In contrast to the other studies, Chavez-Demoulin, Davison, and McNeil (2003) do not model time-varying tails but use extreme value theory in a setting that allows for time variation in the arrival intensity and the distribution of excesses of a given high threshold.

The use of internal risk models according to the Basel Capital Adequacy Directive II requires competitive models which ensure safety of the financial system while at the same time avoid costly over-allocation of capital. Hence, the second part of the paper is devoted to a first application of ARCT behavior to financial time series. In particular, we study changes in Government bond yield spreads. Such changes are relevant to financial risk management as they have a direct impact on the pricing of foreign exchange currency forward rate agreements (FRAs) and currency options (see, e.g., Hull, 2000 for a survey on these derivative instruments). Even mature Government bond markets can be subject to spread changes of substantial magnitude. We study daily changes in yield spreads between U.S. and German Government mid-maturity swap rates from October 1, 1997 to June 30, 2003. The sample period contains several crisis periods, such as the October 1997 Asian crisis, the August–September 1998 Russian crisis, as well as September 11, 2001. In addition, the introduction of the Euro in January 1999 is contained. As Germany—given European Monetary Union convergence—started to represent market conditions for the Euro-Zone well before mid-1997 (see, e.g., Stracca, 1999), the results do in fact represent an analysis of yield spreads for the U.S. Dollar versus the Euro-Zone. We analyze in-sample tail index dynamics and discuss tail index prediction. The findings indicate substantial tail index variation without trend. The dynamics are consistent with stationarity, tail index predictability under mean reversion and provide some suggestive evidence of cyclic behavior.

---

2 The Russian debt crisis and the related case of Long-Term Capital Management (LTCM) may serve as a prominent example where changes in yield spreads played a central role in causing financial distress. The BIS Committee on the Global Financial System is quoted as saying that “only a small number of market participants declined to characterize the 1998 crisis as ‘exceptional’.” For a detailed public perspective crisis review and a view on management implications see for example Jorion (2000).
The paper precedes as follows. Section 2 proposes a model of ARCT behavior and a regression-based approach to model estimation is outlined here. The empirical application for Government bond yield spreads is given in Section 3. Section 4 concludes.

2. The conditional tail framework

Starting with a review on classical extreme value theory, this chapter outlines a model of ARCT behavior. Related useful overviews on extreme value theory and financial applications can be found for example in Coles (2001) and Embrechts, Klüppelberg, and Mikosch (1997). Shiryaev (1999) includes sections on discrete time financial models as well as on extreme value theory.

2.1. Classical extreme value theory

Assume in the following that the random returns \( R_t \) on some asset are observed in discrete time, \( t=1,2,\ldots,T \), and with respect to some probability space \((\Omega, \mathcal{F}, P)\). In the simplest case, the returns are independent and identically distributed (i.i.d.) with common distribution function \( F \). Classical extreme value theory is then concerned with the asymptotic distribution of standardized maxima from the series \( (R_t) \).

The classical result by Fisher/Tippett and Gnedenko states that, if, for given normalizing constants \( a_T \geq 0 \) and \( b_T \in \mathbb{R} \) and \( M_T = \max (R_1, \ldots, R_T) \), some \( H \) exists as the nondegenerate distributional limit of the standardized maximum \( a_T^{-1}(M_T - b_T) \), that is

\[
P(a_T^{-1}(M_T - b_T) \leq r) = F_T(a_T r + b_T) \to H_x(r), \quad \text{as } T \to \infty, \tag{1}
\]

then \( H_x \) is equal to one of three different types of extreme value distributions.

The shape parameter, \( x \in \mathbb{R} \), also denoted as “tail index,” characterizes the extreme behavior of the distribution function. Return distributions typically do not exhibit a finite upper (lower) endpoint which implies thin tails given by the case of the Weibull-type extreme value distribution \((x<0)\). On the other hand, the Gumbel-type extreme value distribution results when the tail index exceeds all limits \((x \to \infty)\) and then such distribution is called thin tailed. As supported by empirical evidence (see, e.g., Longin, 1996; Mandelbrot, 1963), the following treatment will be restricted to the heavy-tailed (also called fat-tailed) case, where—as a prior—we assume \( x>0 \). In this case, \( F \) belongs to the maximum domain of attraction of the Fréchet-type extreme value distribution and, as \( T \to \infty \),

\[
F_T(a_T r + b_T) \to \exp(-r^{-x}), \quad r > 0. \tag{2}
\]

It is known that the above characterization (Eq. (2)) is equivalent to requiring that the tail \( \tilde{F} = 1 - F \) of the distribution function \( F \) is regularly varying at infinity with index \(-x<0\), that is

\[
\tilde{F}(r) = L(r)r^{-x}, \quad r > 0, \tag{3}
\]
where the function $L(r)$ is slowly varying at infinity:

$$\lim_{r \to \infty} \frac{L(vr)}{L(r)} = 1, \quad v > 0.$$ 

The above result demonstrates that the upper tail of a fat-tailed distribution function $F$ behaves roughly like the tail of a Pareto distribution. For a sufficiently high threshold $u > 0$, one may therefore write Eq. (3) as a Pareto-type tail approximation of the form

$$\tilde{F}(r) \approx cr^{-\alpha}, \quad r \geq u. \quad (4)$$

The estimation approach for $\alpha$ may then be based on Eq. (4) and requires the choice of a sufficiently high threshold $u$ or, equivalently, the choice of a sufficiently small fraction $f$ of largest sample observations (cf. DuMouchel, 1983). The tail index estimate can be obtained for example with the estimator of Hill (1975) or with a traditional linear regression approach; recent results are obtained in Datta and McCormick (1998), for example.

2.2. Modeling conditional tail behavior

Starting as in Section 2.1, let the random returns $R_t$ be observed in discrete time, $t=1, 2, \ldots, T$. The stationary return series $(R_t)_{t \geq 1}$ is now based on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 1}, P)$. As is standard, $\mathcal{F}_t = \mathcal{F}(R_1, \ldots, R_t)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $(\mathcal{F}_t)_{t \geq 1}$ is the filtration with $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$.

2.2.1. Assumptions

In the model of conditional tail behavior, the time-$t$ conditional distribution functions $F_t$ of the returns $R_t$ are assumed to evolve within a fixed changepoint model. That is, we impose the following model assumptions.

1. Assume constant time intervals $\Delta > 1$. The corresponding changepoints in discrete time are: $0, \Delta, 2\Delta, \ldots, N\Delta$. This forms a grid with points $n\Delta$ where $n = 0, 1, 2, \ldots, N$ and $N = \lceil T/\Delta \rceil$, ($\lceil a \rceil$ denotes the largest integer smaller or equal to $a \in \mathbb{R}$). For $n > 0$, we can write $n = \lceil t/\Delta \rceil$.
2. Any conditional distribution function $F_t$ prevails throughout the time interval $[n\Delta; (n+1)\Delta)$ which includes $\Delta$ time periods.
3. Let the tail behavior of all conditional distributions $F_t$, $t \in [n\Delta; (n+1)\Delta)$, be fully characterized by some time-$n\Delta$ parameter $\varphi_{n\Delta} \in \mathbb{R}^d$, $d \in \mathbb{N}$. The parameter process $(\varphi_{n\Delta})_{n \geq 1}$ has filtered probability space $(\Omega, \mathcal{G}, (\mathcal{G}_{n\Delta})_{n \geq 1}, P)$, is $\mathcal{G}_{(n-1)\Delta}$-measurable and conditionally independent of the return series.
4. The time-$n\Delta$ parameter process $(\varphi_{n\Delta})_{n \geq 1}$ can be embedded in the time-$t$ scale by rewriting $t = m\Delta$ with $1 \leq m \leq N$. Then, set $\tilde{\mathcal{G}}_t = \mathcal{G}_{[m]-\Delta}$ and set $\varphi_t = \varphi_{[m]-\Delta}$ where $[m] = \lceil t/\Delta \rceil$. The embedded parameter process $(\tilde{\varphi}_t)_{t \geq 1}$ has a filtered probability space $(\Omega, \tilde{\mathcal{G}}, (\tilde{\mathcal{G}}_t)_{t \geq 1}, P)$.
Given the assumptions 1 to 4 above, we can write the time-\(t\) conditional distribution function \(F_t^{\ast}\) as
\[
F_t(r) = P(R_t \leq r \mid \mathcal{G}_{([t/\Delta]-1)\Delta}), \quad \text{(5)}
\]
where \([t/\Delta] = n\). Next, we consider the extreme properties of the conditional distribution in Assumption 3. In particular, we assume that—for each point in time \(t\)—the conditional distribution belongs to the maximum domain of attraction of a Fréchet-type extreme value distribution. Formally, given suitable normalizing constants, \(a_t, T > 0\), \(b_t, T \in \mathbb{R}\) and some \(U \in \mathbb{N}\), this is equivalent to imposing the condition that for some \(r > 0\), as \(U \to \infty\), we have
\[
F_t^{\ast}(a_t r + b_t) \to \exp(-r^{-\tilde{z}_t}), \quad \tilde{z}_t, > 0, \quad 1 \leq t \leq T. \quad \text{(6)}
\]
The above condition corresponds to a generalized version of the Pareto model (4) with \(d = 3\) and \(\tilde{\varphi}_t = (\tilde{z}_t, \tilde{c}_t, \tilde{u}_t)\). Let us consider the tail of the distribution function, \(\bar{F}_t(r) = P(R_t > r \mid \mathcal{G}_{(n-1)\Delta})\). Given a sufficiently high time-varying threshold \(\tilde{u}_t > 0\), the Pareto model approximation for \(r > \tilde{u}_t\) is now
\[
\bar{F}_t(r) \approx \tilde{c}_t r^{-\tilde{z}_t}, \quad 1 \leq t \leq T. \quad \text{(7)}
\]

Thereby, the tail is assumed to have bounded support given by \((\tilde{u}_t, \infty)\). As required for a tail function, \(\lim_{r \to \infty} \bar{F}_t(r) = 0\), \(\tilde{c}_t > 0\) denotes a time-varying scaling parameter and \(\tilde{z}_t\) is the time-varying tail index (where all variables are time-\(t\) embedded). In the following, we impose further structure on the three model parameters for the time intervals \([n\Delta; (n+1)\Delta)\), \(n = 1, 2, \ldots, N\).

### 2.2.1.1. Threshold \(u_{n\Delta}\) and scaling parameter \(c_{n\Delta}\)

Assume that the time-varying tail function (7) is used to model a constant small fraction \(f\), \(0 < f < 1\), of the largest of \(\Delta\) observations of the returns \(R_t\) during an interval \([n\Delta; (n+1)\Delta)\). Given that exactly a fraction of the \(f\) largest observations is above a high threshold \(u_{n\Delta}\), the time-varying threshold is defined as the conditional \((1-f)\)-quantile
\[
P(R_t > u_{n\Delta} \mid \mathcal{G}_{(n-1)\Delta}) = f. \quad \text{(8)}
\]
Given \(\mathcal{G}_{(n-1)\Delta}\), there are various possible ways to model the quantile \(u_{n\Delta}\) in Eq. (8). In the context of financial time series, it seems plausible to assume a simple Markovian structure. In particular, let \(P(R_t > r \mid \mathcal{G}_{(n-1)\Delta}) = P(R_t > r \mid x_{(n-1)\Delta}, u_{(n-1)\Delta})\) and assume a first-order autoregressive process, AR(1), for the logarithmic quantile
\[
\ln u_{n\Delta} = a + \mu \ln u_{(n-1)\Delta} + \zeta_{n\Delta}, \quad n = 1, 2 \ldots N, \quad \text{(9)}
\]
where \(u_0 > 0\), \(0 < |\mu| < 1\), \(a \in \mathbb{R}\) and the innovations \(\zeta_{n\Delta}\) are independent normal with zero expectation and constant variance, \(\zeta_{n\Delta} \overset{\text{d}}{=} \mathcal{N}(0; \sigma_\zeta^2)\).

Note that under the Pareto model \(\bar{F}_t(r) = c_{n\Delta} r^{-\Delta u_{n\Delta}}\), condition (8) holds, if and only if
\[
c_{n\Delta} = f(u_{n\Delta})^{\Delta u_{n\Delta}}. \quad \text{(10)}
\]
Hence, setting the scaling parameter as in Eq. (10) is consistent with modeling a constant small fraction \( f \) of the largest of \( \Delta \) return observations. Their probability distribution is approximated via the Pareto tail model (7).

2.2.1.2. Tail index \( \zeta_{n\Delta} \). There are manifold possibilities for specifying the tail index process. Assuming \( P(R_t > r | G_{(n-1)\Delta}) = P(R_t > r | \zeta_{(n-1)\Delta}, u_{(n-1)\Delta}) \) as above, let the logarithmic tail index follow a first-order autoregressive process

\[
\ln \zeta_{n\Delta} = b + \rho \ln \zeta_{(n-1)\Delta} + \eta_{n\Delta}, \quad n = 1, 2 \ldots, N, \tag{11}
\]

with a suitable start variable \( \zeta_0 > 0 \) and \( a \in \mathbb{R} \). The innovations \( \eta_{n\Delta} \) are assumed to be independent normal with zero expectation and constant variance, \( \eta_{n\Delta} \sim N(0; \sigma_{\eta}^2) \). Process stationarity is given for \( 0 < |\rho| < 1 \).

As above for the quantile process, the logarithmic specification ensures that the tail index is positive any time, \( \zeta_{n\Delta} > 0 \forall n \). Writing \( \beta_{n\Delta} = \ln \zeta_{n\Delta} \) for the logarithmic tail index and using the relation \( b = \bar{\beta}(1 - \rho) \), where \( \bar{\beta} \) denotes the long-run mean logarithmic tail index level, we can rewrite Eq. (11) as

\[
\beta_{n\Delta} - \bar{\beta} = \rho(\beta_{(n-1)\Delta} - \bar{\beta}) + \eta_{n\Delta}. \tag{12}
\]

Note that the unconditional expectation of the tail index then is, \( E \zeta_\infty = \exp \{ \bar{\beta} \} \), which determines the long-run tail index. The conditional tail index expectation is simply given as

\[
E \zeta_{n\Delta} | \zeta_{(n-1)\Delta} = \exp \{ \bar{\beta} + \rho(\ln \zeta_{(n-1)\Delta} - \bar{\beta}) \}. \tag{13}
\]

2.2.2. Model specification

The above section specified a model of ARCT behavior of the Pareto type. As mentioned above, there are various different possible model specifications.

In the following model, it is assumed for simplification that the threshold is not relevant as conditioning information. Hence, we do not explicitly model the threshold as a time-varying quantile. Focusing on the tail index only, we outline the econometric specification of one simple model in the following.

2.2.2.1. Model. In short, given a sufficiently high threshold \( u_{(n-1)\Delta} > 0 \) which is assumed to be exceeded with small probability \( f \), the variables \( R_t \) have a time-varying tail

\[
P(R_t > r | G_{(n-1)\Delta}) \approx f(u_{(n-1)\Delta} / r)^{\zeta_{n\Delta}}, \quad r > u_{(n-1)\Delta}, \tag{14}
\]

with \( n = [t/\Delta] \) and \( t = 1, 2 \ldots, T, n = 1, 2 \ldots, N \). The logarithm of the tail index, \( \ln \zeta_{n\Delta} \), is assumed to follow a stationary AR(1) process

\[
\ln \zeta_{n\Delta} = \bar{\beta} + \rho(\ln \zeta_{(n-1)\Delta} - \bar{\beta}) + \eta_{n\Delta}. \tag{15}
\]
2.2.2.2. Estimation. Eq. (14) translates naturally to a regression specification when the model variables are replaced by their empirical counterparts. Let $R_{1,\Delta} \geq \ldots \geq R_{j,\Delta} \geq \ldots \geq R_{k,\Delta}$ denote the $k=[f/\Delta]$ upper order statistics from the $n$th subsample of size $\Delta$ which exceed the threshold $u_{n,\Delta}$ during the time interval $[n\Delta; (n+1)\Delta)$, $n=0, 1, \ldots, N-1$. The tail of the empirical distribution function is $F_{\Delta}(r_{j,\Delta})=(j-1)/\Delta$. It follows that $F_{\Delta}(r_{k+1,\Delta})=[f/\Delta]/\Delta \approx f$ and hence we set $u_{n,\Delta}=r_{k+1,\Delta}$ (see Eq. (8)). Taking logarithms on both sides of Eq. (14) then yields

$$\ln F_{\Delta}(r_{j,\Delta}) \equiv \ln f - \alpha_{n,\Delta}(\ln r_{j,\Delta} - \ln r_{k+1,\Delta}), \quad j = 2, \ldots, k.$$ 

Note that the cross-sectional model is not defined for the largest order statistic $j=1$. The above is equivalent to

$$(\ln r_{j,\Delta} - \ln r_{k+1,\Delta}) \approx -\alpha_{n,\Delta}^{-1} (\ln F_{\Delta}(r_{j,\Delta}) - \ln f), \quad j = 2, \ldots, k. \quad (16)$$

For order statistics $i=j-1$, time intervals $n$, and by introducing additive zero-expectation i.i.d. noise terms $\varepsilon_{i,n,\Delta}$, we may rewrite the model as follows

$$x_{i,n,\Delta} = -\alpha_{n,\Delta}^{-1} z_i + \varepsilon_{i,n,\Delta}, \quad \begin{cases} i = 1, \ldots, k-1, \\ n = 1, 2, \ldots, N-1, \end{cases} \quad (17)$$

where the variables are defined as given in Eq. (16) above. Note that the $z_i$’s do only depend on the order $i$, but not on the chosen time interval $n$. The approximation noise terms are assumed to be uncorrelated with both the explanatory variables and the unobservable noise terms in the tail index process (Eq. (15)).

We may now reformulate the above ARCT model given by Eqs. (14) and (15) in state-space representation. Defining $X_{n,\Delta}=(x_{i,n,\Delta})(k-1) \times 1$, $Z=(z_i)(k-1) \times 1$ and $\varepsilon_{n,\Delta}=(\varepsilon_{i,n,\Delta})(k-1) \times 1$ as variables given in Eq. (17), the model has observation equation

$$X_{n,\Delta} = -Z \exp(\beta_{n,\Delta})^{-1} + \varepsilon_{n,\Delta}. \quad (18)$$

The time-varying coefficient is given by the tail index, where $\beta_{n,\Delta}$ denotes the logarithmic tail index. The latter is assumed to follow the state Eq. (15)

$$\beta_{n,\Delta} - \bar{\beta} = \rho(\beta_{(n-1),\Delta} - \bar{\beta}) + \eta_{n,\Delta}. \quad (19)$$

For the innovation vector $\varepsilon_{n,\Delta}$ in the observation Eq. (18), assume $E(\varepsilon_{n,\Delta})=0$ and $E(\varepsilon_{n,\Delta}^2)=\Omega$, where $\Omega=(\sigma_{i,j})(k-1) \times (k-1)$. This generally allows for a heteroskedastic-dependent time-invariant innovation structure. While differing residual variances across order observations follow necessarily form differing distributional properties of the order statistics in the tail formulation (Eq. (16)), we may impose the model implied condition of uncorrelatedness by setting $\Omega_e=\text{diag}(\sigma_{e,i}^2)(k-1) \times (k-1)$. Orthogonality in the observation Eq. (18) is given as the $z_i$’s are deterministic. The tail index process innovations are
uncorrelated with the cross-sectional innovations, $\text{Cov}(\eta_n; \varepsilon_n) = 0$, and the joint innovation structure is assumed to be normal. Hence, write

$$
\begin{pmatrix}
\varepsilon_t \\
\eta_t
\end{pmatrix} \overset{d}{\sim} \mathcal{N}
\begin{pmatrix}
0 \\
0
\end{pmatrix};
\begin{pmatrix}
\Omega_e & 0 \\
0 & \sigma^2_\eta
\end{pmatrix}
$$

(20)

Estimation of the ARCT model in state space form given in Eqs. (18) and (19) can be carried out via the standard Kalman/Bucy filter under quasi-maximum likelihood assumptions (20). A detailed time-series related treatment of the estimation approach can be found for example in Harvey (1989).

3. An empirical study of yield spreads

This chapter sketches an application of the ARCT model in measuring the tail index dynamics of Government bond yield spread changes. To analyze the dynamics for mature bond markets, we study daily changes in yield spreads between U.S. and German Government bonds. Given European Monetary Union convergence during the second half of the 1990s (see, e.g., Stracca, 1999), the German bond market can be used as a proxy representing Euro-Zone market conditions during the 1997 to 2003 sample period. Hence, we can expect a long-run stationary environment, where intermediate events may cause changes to the tail behavior of spread changes.

Changes in Government yield spreads are of immediate relevance for the management of foreign exchange risks. For example, they have a direct impact on the pricing of currency FRAs and currency options (see, e.g., Hull, 2000). As such, the given model application is directly related to managing the risk of foreign exchange FRAs.³

3.1. The dataset

The dataset consists of daily zero coupon yields $y_{t,t}$ for U.S. and German Government bonds which are based on swap rates as obtained from the Datastream/ICAP zero-curve series. The sample covers the period from October 1, 1997 to June 30, 2003 with $t = 1, \ldots, 1498$. We use annualized yields for Government bonds with mid-maturities of 5, 7, and 10 years. Spreads $s_{t,t}$ are calculated as simple daily differences between U.S. and German yields. Daily spread changes are then given as $\Delta s_{t,t} = s_{t,t} - s_{t-1,t}$; $\Delta s_{t,t} \cdot 100$ denote changes in basis points.

Table 1 reports summary statistics for the yield spread changes $\Delta s_{t,t}$, namely the sample mean, standard deviation, skewness, and excess kurtosis. As the results indicate, all changes have distributions with high excess kurtosis. The variability of the spreads is very

³ The risk implications of changes in spreads for foreign exchange FRAs—assuming continuous compounding—follow from the interest rate parity relation between the time-$t$ fair forward exchange rate $F_{t,t}$ and the spot exchange rate $S_t$. $F_{t,t} = S_t \exp(\tau s_{t,t})$. Hence, for given $S_t$, a spread change $\Delta s_{t+1,t}$ induces a forward rate change of: $\Delta F_{t+1,t}(S_t, \Delta s_{t+1,t}) = F_{t,t}(\exp(\tau \Delta s_{t+1,t}) - 1)$. Note that also FRA-maturity determines position risk.
similar, while skewness is more pronounced for the shorter 5- and 7-year maturities. Negative sample means are due to overall decreasing spreads during the sample period. The first-order sample autocorrelation coefficient acf indicates negative time-series dependence in the spread changes. However, the dependence is not found to be significant at the 95% confidence level given standard asymptotic assumptions.

3.2. Model estimation results

The extreme value approach used here is consistent with asymmetric distributions, in that it is modeling the distribution tails separately from the center as well as separately from each other. Although there is some evidence of skewness in Table 1, it seems questionable as to whether this reflects a substantial and predictable asymmetry. In addition, note that modeling the tails separately comes at a cost which is given by a reduction of the amount of available extreme, that is, tail, observations. This has a potential negative impact on the ability to predict tail fatness. Hence, in this application, we choose to impose symmetry and estimate the tail dynamics of the model from Section 2.2.2 using absolute spread changes $|\Delta s_t|$. 

3.2.1. Procedure

The results for two ARCT model applications are presented in the following. First, Section 3.2.2 outlines the estimation results for the model of Section 2.2.2. Then, Section 3.2.3 discusses the results obtained from a rolling window regression for day $t > \Delta$ tail and quantile forecasts. In this latter case, observations are dependent by construction. However, this procedure may demonstrate the results of a practitioner who aims at using the model and all available time-$t$ return information for predictions of next period’s (embedded) tail index $\tilde{s}_{t+1}$. In the present context, this would apply to daily tail index predictions.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Mean($\Delta s$)</th>
<th>S.D.($\Delta s$)</th>
<th>Skew($\Delta s$)</th>
<th>Kurt($\Delta s$)</th>
<th>acf($\Delta s$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-0.001</td>
<td>0.050</td>
<td>0.143</td>
<td>1.85</td>
<td>-0.023</td>
</tr>
<tr>
<td>7</td>
<td>-0.001</td>
<td>0.049</td>
<td>0.131</td>
<td>1.56</td>
<td>-0.045</td>
</tr>
<tr>
<td>10</td>
<td>-0.001</td>
<td>0.050</td>
<td>-0.013</td>
<td>1.89</td>
<td>-0.032</td>
</tr>
</tbody>
</table>

Table 1
Summary statistics for the daily yield spread changes

Sample period October 1997 to June 2003.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\hat{\beta}$</th>
<th>$\rho$</th>
<th>$\sigma^2_a$</th>
<th>$R^2 \mid t \mid$</th>
<th>$R^2_{t+1}$</th>
<th>$R^2_{t+2}$</th>
<th>lnL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.56* (50.0)</td>
<td>0.38* (9.22)</td>
<td>.37</td>
<td>.746</td>
<td>.500</td>
<td>.103</td>
<td>-29.18</td>
</tr>
<tr>
<td>7</td>
<td>1.71* (20.9)</td>
<td>0.30* (40.3)</td>
<td>2.18</td>
<td>.422</td>
<td>.655</td>
<td>.650</td>
<td>3.625</td>
</tr>
<tr>
<td>10</td>
<td>2.04 (0.55)</td>
<td>0.18 (0.32)</td>
<td>1.00</td>
<td>.599</td>
<td>.401</td>
<td>.122</td>
<td>-39.81</td>
</tr>
</tbody>
</table>

Table 2
Model estimation results for absolute yield spread changes with $t$ values in parenthesis

Sample period October 1997 to June 2003. Subsamples with $\Delta = 60$, $1 \leq n \leq 24$, and $f=0.1$.

* Denotes 95% confidence level significance for a double-sided test under standard asymptotic assumptions.
In both cases, we choose subsamples of length $\Delta$ and a fraction $f$ of the largest subsample spread change observations yielding $k-1$ cross-sectional observations in the observation Eq. (18). We do not use all of these available observations as—given the small sample sizes—single higher order statistics may exhibit substantial variation and may not always conform well with a theoretical linear fit. Instead, we choose $l \in \{1, \ldots, k-3\}$ as the first of three consecutive pairs of order statistics $(x_i, z_i)_{i=1, \ldots, k-1}$ which yield convergence under application of the Marquardt algorithm.

3.2.2. Tail index dynamics

Parameter estimation results for the fixed changepoint ARCT model are presented in Table 2. The results for choosing the six largest absolute spread changes from a window of

![Graph](image-url)

Table 2
Model estimation results for absolute yield spread changes with $t$ values in parenthesis

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$l$</th>
<th>$\hat{\beta}$</th>
<th>$\rho$</th>
<th>$\sigma_{\hat{\alpha}}^2$</th>
<th>$R_l^2$</th>
<th>$R_{l+1}^2$</th>
<th>$R_{l+2}^2$</th>
<th>$\ln L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
<td>1.60 (0.99)</td>
<td>0.26 (0.73)</td>
<td>1.01</td>
<td>.686</td>
<td>.756</td>
<td>.338</td>
<td>$-2572.93$</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1.61 (4.39)</td>
<td>0.62 (4.25)</td>
<td>.91</td>
<td>.708</td>
<td>.990</td>
<td>.802</td>
<td>$-2287.82$</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2.12 (1.68)</td>
<td>0.36 (2.78)</td>
<td>1.00</td>
<td>.768</td>
<td>.461</td>
<td>.082</td>
<td>$-2986.09$</td>
</tr>
</tbody>
</table>

Rolling sample with $\Delta = 100$, $\Delta \leq t \leq 1498$ and $f = 0.1$. Sample period October 1997 to June 2003.

Fig. 1. Top: 10-year maturity absolute spread changes. Bottom: 10-year maturity subsample tail index estimates; unconditional regression alpha estimated for the overall sample, $\hat{\alpha} = 3.78$ (‘u-alpha’), rolling regression subsample estimates (‘r-alpha’) and Kalman filter smoothed alpha estimates (‘f-alpha’) with estimated unconditional mean, $\exp (2.04) = 7.69$ (‘m-alpha’). All subsamples with $\Delta = 60$, $1 \leq n \leq 24$ and $f = 0.1$. 
approximately 3 months ($\Delta = 60\%$ and $f = 0.1$) may serve as an example of the parameter fit of the model.

The logarithmic mean reversion level is estimated to be positive for all three series. Interestingly, the reversion levels are all estimated at relatively high levels. This allows for the interpretation that a high tail index level (i.e., thin tails) rather than a low index level (i.e., fat tails) dominates the long-run tail index dynamics. For the 5-year spread, for example, the estimate of the beta mean reversion level of 1.56 relates to a tail index mean reversion level 4.76. Two out of the three autoregressive coefficients are significantly positive; all of the estimates lie within the positive and stationarity region. The cross-sectional regressions contribute to the explanation of the tail as indicated by the $R^2$ statistics. The log-likelihood statistics in Table 2 (as well as in Table 3) suggest that the model fit is better for the 7-year spread change series than for the 5- and the 10-year series (see Fig. 1).

3.2.3. Tail index and quantile prediction

Table 3 summarizes the parameter estimation results for the second application. The results are based on a rolling window regression with window length equal to 100 trading days or approximately 5 months. Choosing the 10 largest absolute spread changes ($\Delta = 100$ and $f = 0.1$) yields a series of tail observations. Performing a model estimation for this series, the model parameter estimates reconfirm the basic properties of the

Fig. 2. Top: 7-year maturity absolute spread changes. Bottom: 7-year maturity predicted tail index estimates; rolling regression subsample estimates (‘r-alpha’) and Kalman filter predicted alpha (‘f-alpha’). All subsamples with $\Delta = 100$, $t \leq 1498$ and $f = 0.1$. 
estimation results in Table 2 above. The logarithmic mean reversion level estimates indicate alpha reversion levels between about 8.3 (for the 10-year spread) and about 5 (for the 5- and 7-year spreads). Again, the three autoregressive coefficients are estimated to be positive and smaller than one.

The following two figures illustrate the estimation and forecasting results for the 7-year spread changes in more detail. We derive plots of the forecasted dynamic tail index as well as of one day ahead predicted quantiles. The $p$-quantiles $q_{p,t+1}$ are implicitly given by the estimated tail as follows

$$P(|R_{t+1}| > q_{p,t+1} | F_t) = 1 - p.$$

The tail predictions are conditional on the tail index forecasts. Given the parameter estimates and Eq. (13), tail index forecasting this is done according to the equation

$$\hat{\xi}_{t+1} | F_t = \exp\{\hat{\beta} + \hat{q}(\ln\hat{\xi}_t - \hat{\beta})\},$$

where the Kalman filter prediction is based on the filtration $F_t$, that is, on all available return information up to time $t$.

Fig. 2 plots the 1-day-ahead forecasts of the tail index (bottom) jointly with the absolute daily 7-year spread changes (top). The plots demonstrate that increased levels in spread change volatility tend to be accompanied by lower levels of the tail index. Whereas the

![Figure 3](image-url)

Fig. 3. Seven-year maturity spread changes together with the predicted double-sided 99% quantiles (full line) and 95% quantiles (dotted line).
crosses in Fig. 2 (bottom) indicate the regression estimates of the tail index, the line represents the smoothed Kalman filter forecast of the tail index. As can be seen from the plot, the model is able to reduce the time variability of the tail index estimates. The graphical results also give suggestive evidence of a cyclic behavior of the time-varying tail index parameter.

Based on the tail index predictions in Fig. 2, Fig. 3 shows the symmetric 1-day-ahead predicted 95% and 99% quantiles together with the 7-year spread changes. The model generates relatively stable quantile forecasts; still no-table shortcomings are a somewhat slow reaction to increases in tail fatness as well as some tendency for overshooting (see observation 500). In addition, there is a higher than expected number of violations of the 99% quantiles around observation 250 to 300. This was during fall 1998 when bond markets exhibited increased stress as a result of the preceding August–September Russian debt crisis. After the introduction of the Euro, a period of unexpected and clustered quantile violations is during July and August 1999, which is around observation 450 in Fig. 3. Quantile exceedances during later periods better confirm with expectations.

4. Conclusion

The modeling of potential extreme market events can have an important impact on the calculation of minimum capital risk requirements (see, e.g., Clare et al., 2002). It appears that approaches to time-varying extreme financial risk may provide a potentially useful tool to extend our understanding of market stress and may also contribute to the set of available risk prediction methods. This paper outlines a straightforward model of ARCT behavior. Estimation and application to financial risk modeling yields results on the tail index dynamics of yield spreads. Future work in the area may be devoted to generalize and robustify the model and its estimation as well as to investigate its applicability and predictive performance.

Acknowledgements

The author would like to thank participants for their remarks at the 2003 VHB Zurich meeting, the 2003 LMU-TUM Risk Management Day, the CEFS-ODEON Colloquium and the 2004 CEPR-ESSFM Gerzensee Meeting as well as Bernhard Brück, Claudia Klüppelberg and Christoph Kühn for helpful comments and discussions during earlier stages of the paper. Part of the paper was written while the author enjoyed the hospitality of the Faculty of Economics and Politics, University of Cambridge.

References


