Estimation of holding periods applied to the case of short and leveraged ETFs

Leo Schubert[1]
Constance University of Applied Sciences

David Schubert
Munich

[1] Schubert@HTWG-Konstanz.de

Resumen

La estimación de los períodos de tenencia de una inversión en instrumentos financieros debe realizarse mediante un proceso dinámico, en el cual el tamaño del intervalo de observación influye en los resultados. Pequeños intervalos de observación producirán en promedio períodos de tenencia más pequeños a los obtenidos con intervalos grandes de observación. El enfoque desarrollado en este artículo ofrece la posibilidad de estimar estos promedios, independientemente del tamaño del intervalo de observación. El método es ilustrado con el ejemplo de dos distribuciones, basadas en las leyes de probabilidad exponencial y geométrica. La estimación óptima se encuentra al maximizar la función de verosimilitud.

Los dos ejemplos examinados fueron aplicados al instrumento financiero *Exchange Traded Fund* (ETF). Estos fondos, negociables en la bolsa de valores, tienen factores de apalancamiento -2, -1, +1, y +2. Aunque más de 30 ETFs fueron estudiados, la mayoría de los datos están orientados al “db x-tracker ShortDAX ETF”, “db x-trackers DAX ETF”, “iShares DAX (DE)” y el “Lyxor ETF LevDAX”. La aplicación del método propuesto a las ETFs aumenta la amplitud del período de tenencia entre un 4% y 29%. El aumento es dependiente del tamaño de la ventana de observación, del factor de apalancamiento y del promedio de los períodos de tenencia.

Abstract

The estimation of the holding periods of financial products has to be done in a dynamic process in which the size of the observation time interval influences the result. Small intervals will produce smaller average holding periods than bigger ones. The approach developed in this paper offers the possibility of estimating this average independently of the size of this time interval. This method is demonstrated on the example of two distributions, based on the exponential and the geometric probability functions. The estimation will be found by maximizing the likelihood function.

The two examples will finally be applied to the financial instrument Exchange Traded Fund (ETF). The analysis contains ETFs with leverage factors of -2, -1, +1 and +2. Although different ETFs are treated, the majority of the data is concerned with the "db x-tracker ShortDAX ETF", "db x-trackers DAX ETF", "iShares DAX (DE)" and the "Lyxor ETF LevDAX". By the application of the proposed estimation approaches, the average holding periods of ETFs increase by 4%-29%. This increase depends on the time interval T of observation, the leverage factor, and the average holding period.

Palabras clave: Períodos de tenencia, duración, distribución exponencial, distribución geométrica, muestreo, estimación de máxima verosimilitud, apalancamiento en los Exchange-Traded Funds (ETFs).

Key Words: Holding periods, Duration, Exponential distribution, Geometric Distribution, Sampling, Maximum Likelihood estimation, Short and Leveraged Exchange-Traded Funds (ETFs).

JEL Classification System: G23, G24, C13, C20, C41, C46.

1. Introduction

The measurement of holding periods is done by more or less rough estimation. The issues with this estimation depend on how the holding periods are determined. The floating process of the buying and selling of financial assets without break makes the precise observation of holding periods as difficult as taking a sharp photo of a running horse. The external or classical way uses the aggregated data of the exchange boards; the internal way uses the individual data of financial institutions. Both methods of calculating the average holding periods have their imperfections, which will be discussed in the following chapter. The approaches developed in this paper are concerned with the imperfection of the internal way. In this case, the size of the observation time interval influences the result. Small intervals will produce smaller average holding periods than bigger ones. The method proposed in this paper offers the possibility of estimating the holding period independently of the size of this time interval. This approach is demonstrated on the example of two distributions, based on the exponential and the geometric probability functions. The estimation will be carried out by maximizing the likelihood function.

2. The classical way to calculate holding periods and the internal way

The classical way to calculate the average holding period uses the ratio “market capitalisation / sales”. In the stock markets, the computation of market capitalisation uses stock prices at the end of the year, which must be multiplied with the number of stocks in this market. The sum of sales of stocks in the whole year is used as the divisor. Stocks which are not dealt on exchange boards are not included in the calculation in Table 2-1. This table shows the market capitalisation, the sales and furthermore the holding period in years for the years 1980 to 2014.

Table 2-1 depicts the increasing activity on the exchange boards of stock markets in the last three decades. In this time interval, the holding period of stocks has been severely decreasing: from nearly 9.7 years in 1980 to 0.6 years in 2014. The estimation of holding periods in the way described above has the disadvantage that the date when the capitalization is computed is at the end of the year, while the sales are registered over the whole year. Therefore, the lowest holding period of 0.3 years can be recognised in the year 2008 when the financial crisis took place. At the end of the year, the value of the stocks[2] was only 60% of that for the beginning of the year, which were also used to

[2] See, e.g. German Stock Index DAX or European Index STOXX50.

calculate the sales. Generally in the case of constant holding periods, strongly rising or falling stock prices within one year would cause higher or lower ratios as estimates for these periods.

The classical way's imprecise estimation of the holding periods is due to the lack of the more detailed information that banks have. Financial institutions know when their clients or investors buy or sell assets. Although the internal way to calculate holding periods uses more information, the exact average holding period cannot be calculated either in most cases.

If the *sample consists of investors* that buy assets within a period, it would be necessary to wait until all these engagements are closed again. This could consume too much time to finish the study. The more convenient way would be to take some investors who sell assets within a period and search in the past for when they were bought. This point of view is not so time consuming, but ignores those clients with long holding periods who will sell sometime in the future. Especially if new financial instruments have recently been introduced into the market, this estimation problem exists.

Another internal way to estimate holding periods regards all investors that buy and sell within a *sample window of years or months*. This approach regards only the transactions which occur completely in the time window between time unit 1 and T. The maximal holding period measurable in this window has T-1 time units. Cases in which assets are only sold or only bought outside the window are not taken into account, and neither are cases with holding periods greater than T-1, which belong proportionately to the used sample of years, but were bought earlier and sold later. Depending on whether the transaction is buying or selling, the first case will be called in the following “semi-outside**”** and the second “completely-outside”. A calculation of the holding period which ignores “semi-outside” and “complete-outside” cases will always underestimate the real holding period. This deficiency depends on the number of years in the sample.

Year	Capitalisation in thousand million US Dollars	Sales in thousand million US Dollars	Holding period in years
1980	2.9	0.3	9.7
…			
1990	8.9	5.7	1.6
…			
2000	31.0	49.8	0.6
2001	26.6	38.1	0.7
2002	22.8	33.1	0.7
2003	30.7	32.2	1.0
2004	36.3	40.5	0.9
2005	40.8	52.3	0.8
2006	50.4	69.4	0.7

2007	60.6	99.3	0.6
2008	32.3	113.2	0.3
2009	47.6	81.7	0.6
2010	54.7	85.5	0.6
2011	47.4	102.0	0.5
2012	54.5	82.9	0.7
2013	64.5	93.0	0.7
2014	67.5	111.4	0.6

Table 2-1: Capitalisation, sales and holding period in stock markets of the "World Federation of Exchanges" and the "Exchange boards of the London SE Group" in the years 1980 to 2014[3]

To get better estimations for holding periods out of a sample of years, an approach will be developed in the following using the information base of the internal way. This approach contains two steps. The first step is concerned with including the semi-outside cases using a correction factor and the second step is founded on the assumption that the holding periods are geometrically or exponentially distributed.

3. A correction factor to include the "semi-outside" engagements

In a sample observed within a window of time, only those cases can be registered that are bought and also sold within this time interval. Investments bought before or sold after this time interval must be ignored. The maximal duration which can be observed is T-1, which is the difference between the last date and the oldest date; this interval touches T days. In the case of continuous time measurement, the difference between T and T-1 does not exist.

The size t (t = 0, ..., T-1) of a holding period has its lowest value, t=0, in the case of day trades, which touch one day. The maximal holding period t which can be observed is t = T-1, as mentioned above. Bigger holding periods t > T-1 had to be ignored as it was not known whether they exist or not. A solution for this problem will be discussed in a later chapter.

Normally in the calculation of the average holding period, the periods are weighted by a factor, the number n_t of investments with a duration of t days. Depending on the application, the invested volume could be used instead for weighting the holding periods. The following approach is based on the assumption that the probability of buying a product independent of the day.

[3] See "World Federation of Exchanges (WEF): www.world-exchanges.org: Annual Statistics Reports" or "Bundeszentrale für politische Bildung http://www.bpb.de/nachschlagen/zahlen-und-fakten/globalisierung/52590/aktien**".**

Day trades have a holding period of t=0, therefore all n_0 trades are observable. In contrast, most of the investments with a holding period of t=T-1 will not be counted in n_t, although these investments belong partially to the sample of time selected for observing. This incompleteness in the number of observable engagements n_t depends on the size of the holding period t.

Table 3-1 shows three day trades which are completely inside the time interval, four transactions with a duration of t=1 day with only two periods completely inside, and five with a duration of t=2 with only one transaction inside. The maximal length in the example is T-1 = 2; this case touches T = 3 days. The semi-outside cases are registered according to their part inside the time interval. Doing this, the number of investments rises to (n_0=3), n_1=2+1 and n_2=1+2.

t	Mo	Tue	We	Thu	Fr	Sa	Su	inside	semi-outside
0								1	
0								1	
0								1	
sum								**3**	**0**
1									0.5
1								1	
1								1	
1									0.5
sum								**2**	**1**
2									1/3
2									2/3
2								1	
2									2/3
2									1/3
sum								**1**	**2**

Table 3-1: Example of registered numbers of holding periods and their missing part outside

To get a better estimation of the average holding period, the registered number n_t of investments with holding period t has to be adapted. In Table 3-1, the total number of cases (inside plus semi-outside) is T=3 and is shown in the line of the sum. The different lines represent the possible different configurations of buying and selling having a fixed holding period t. Out of these cases there are T– t cases inside the interval and therefore observable within the data of the sample. The sum of parts outside the time interval is t, which belong to the time interval, too. This portion is t / (T - t) and has to be altered by a correction factor ρ_t

$$\rho_t = 1 + \frac{t}{T-t} = \frac{T}{T-t},\ t = 0, \ldots, T\text{-}1. \tag{3-1}$$

The correction factor depends on t and T-1. A large time interval T-1 for the observation of the holding periods reduces the effect of the factor and, depending on the research objective, makes its application needless. Due to the compensation of unobservable semi-outside investments in n_t by the factor ρ_t, the average holding period will not depend on the sizes of the interval T-1. While the observable number $n_t = T - t$ in Table 3-1, the product $n_t \cdot \rho_t = T$ is independent of the size of the holding period t. It must be mentioned that only in the artificial example do the n_t shrink in a discrete linear way. In empirical data n_t will show a different behaviour. The artificial example is based on the assumption that every day the probability of buying a product is equal. In this example, products were bought every day, each one with a holding period of from t = 0 to T-1 days. Using the correction factor, every holding period will have the same weight, $n_t \cdot \rho_t = T$ when calculating the average holding period.

In the small example of Table 3-1 for t= 0, 1, 2, the factors are $\rho_0 = (2+1)/(2+1-0) = 1$, $\rho_1 = (2+1)/(2+1-1) = 1.5$ and $\rho_2 = (2+1)/(2+1-2) = 3$. The factor $\rho_2 = 3$ means that only 1/3 of all cases are known due to the data's being restricted by the selection of a small time interval.

For bigger samples, e.g. with a size of T-1= 1000 in the artificial case, the correction factor ρ_t would be: $\rho_0 = 1$, $\rho_1 = 1.0010$, $\rho_2 = 0.0020$, $\rho_{10} = 1.0101$, $\rho_{50} = 1.0526$, $\rho_{100} = 1.11099$, $\rho_{900} = 9.9109$. In the case of long holding periods, e.g. t=900, the factor is high. For cases with high ρ, it should be taken into account that there will be big gaps between the different observed holding periods t (see the right side of Figure 5-1).

4. Maximum likelihood estimation which includes the complete-outside cases

Investments both bought before and sold after the time interval with length T-1 are not covered by the correction factor developed above. In the example of Table 3-1, an investment which was bought on Monday and sold on Saturday would be such an engagement. To take these cases into account for the calculation of statistical moments like the mean of the holding period, a supposition concerning the distribution of the holding period has to be made. The holding period of financial products is regarded as geometrically distributed. If the time is measured continuously, the difference between T-1 (maximal holding periods) and T (touched days) does not exist and the holding periods t in Equation (3-1) are $0 < t < T$.

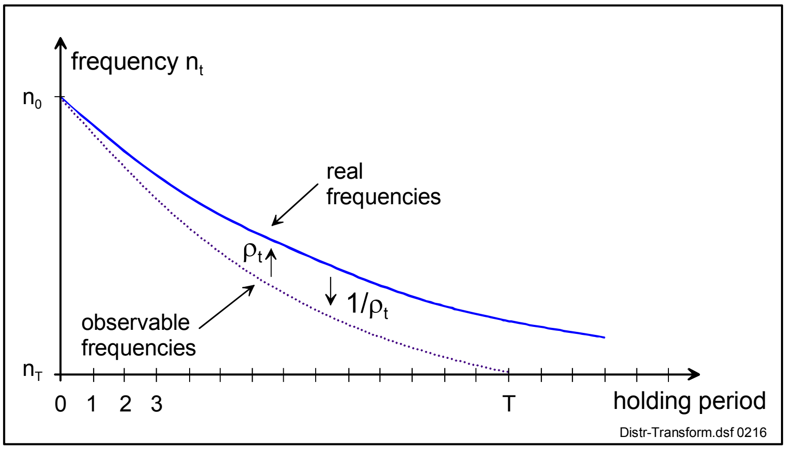

Figure 4-1: Observable and real frequencies

The computation of a maximum likelihood estimator requires independent holding periods. Therefore, the real distribution has to be transformed into the observable distribution[4], multiplying the real distribution with $1/\rho_t = 1 - t / T$ (see Equation (3-1)). For discrete time, the holding period t is the realisation of a geometrically distributed variable[5] with the following probability function with parameter $0 < p < 1$

$$f(t) = \begin{cases} p \cdot (1-p)^t & \text{if} \quad t \in \mathbb{N}_0 \\ 0 & \text{else} \end{cases}, \qquad (4\text{-}1)$$

and for continuous time the holding period is distributed according to an exponentially distributed variable[6] with the density function with parameter $\lambda > 0$

$$f(t)^\circ = \begin{cases} \lambda \cdot e^{-\lambda \cdot t} & \text{if} \quad t \geq 0 \\ 0 & \text{else} \end{cases}. \qquad (4\text{-}2)$$

In Chapter 4.1, the function $f(t) / \rho_t$ with the geometrical probability function f(t) will be used to find a maximum likelihood estimator. In Chapter 4.2, the exponential density function $f(t)^\circ$ will be used.

The process of transformation with $1/\rho_t$ is depicted in Figure 4-1, which shows the observable and real frequencies. Like the absolute frequencies, the probability functions can be transformed. As mentioned above, the transformed probability function which has maximal holding period T will be used to construct the likelihood function.

[4] The transformation of the observed data by the factor ρ_t would destroy the independence of the single observations. Without the independence, the likelihood function cannot be built using this transformed data.
[5] See, e.g. Bosch (1992), p.123.
[6] See, e.g. Bamberg, Baur and Krapp (2012), pp. 100f.

4.1. Geometrically distributed data

The probability function of the observable data, in form of the transformed geometrical probability function (see (4-1)) is

$$f(t) \cong \begin{cases} (1-\frac{t}{T}) \cdot p \cdot (1-p)^t & \text{if} \quad t \in IN_0 \\ 0 & \text{else} \end{cases} . \qquad (4.1\text{-}1)$$

As the cumulative distribution function of (4.1-1) has to be normalized to 1, a constant factor c(p,T) has to be introduced.[7] With this factor, the probability function becomes

$$f(t) = c(p,T) \cdot \left(1-\frac{t}{T}\right) \cdot (1-p)^t \quad \text{for } t = 0, 1, 2, \ldots, T \qquad (4.1\text{-}2)$$

$$\text{with } c(p,T) = \frac{T \cdot p^2}{(T+1) \cdot p - 1 + (1-p)^{T+1}} .$$

The cumulative distribution function[8] of (4.1-2) is

$$F(t) = \frac{(1-p)^{t+1} \cdot (1 + p \cdot t - T \cdot p) + T \cdot p - 1 + p}{(T+1) \cdot p - 1 + (1-p)^{T+1}} \quad \text{for } t = 0, 1, 2, \ldots, T \qquad (4.1\text{-}3)$$

The function F(t) = 0, respectively, F(t) = 1 for t < 0, respectively, t > T. The likelihood function, built from a sample of n independent observed periods t_1, t_2, …, t_n using (4.1-2), is

$$L(t_1,\ldots,t_n \mid p) = \frac{p^{2n}}{\left((T+1) \cdot p - 1 + (1-p)^{T+1}\right)^n} \cdot (1-p)^{\sum t_i} \cdot \prod_{i=1}^{n} (T - t_i) \qquad (4.1\text{-}4)$$

and the logarithm of (4.1-4) is

$$\ln(L(t \mid p)) = 2 \cdot n \cdot \ln(p) - n \cdot \ln\left((1-p)^{T+1} + p \cdot (T+1) - 1\right) + \left(\sum_{i=1}^{n} t_i\right) \cdot \ln(1-p) + \sum_{i=1}^{n} \ln(T - t_i) . \qquad (4.1\text{-}5)$$

To find the maximum likelihood estimator, Equation (4.1-5) has to be differentiated with respect to p:

$$\frac{\partial \ln(L(t \mid p))}{\partial p} = \frac{2 \cdot n}{p} - \frac{n \cdot \left(-(T+1) \cdot (1-p)^T + (T+1)\right)}{(1-p)^{T+1} + p(T+1) - 1} - \frac{\sum t_i}{1-p} . \qquad (4.1\text{-}6)$$

Exploring (4.1-6) shows, that for empirical data this function has one zero point. To characterise “empirical data” in a more general and applicable way, the following inequality[9] (4.1-7) is used. If this inequality is satisfied, the zero point of (4.1-6) will additionally not be in the immediate vicinity of p=0. This inequality will be called the *threshold criterion* in the following:

---

[7] Proof: see Appendix A1.
[8] Proof: see Appendix A2.
[9] See Appendix A5.

$$\frac{\sum t_i}{n \cdot T} < \frac{1}{3} \cdot \frac{T^2 + T}{T^2 - 1}. \tag{4.1-7}$$

As the limit of (4.1-6) will obviously be positive for p → 0 and negative for p → 1, the zero point is a maximum. For large time intervals T, the threshold given by (4.1-7) is 1/3, as in the following chapter in which exponentially distributed data are treated. Appendix A5 illustrates that this threshold will be met in a special case in which the supposition of the geometric distribution is obviously not true. In this special case, the holding periods are equally distributed. The maximum likelihood estimator can be found quickly by standard optimization tools, like, e.g. EXCEL-Solver.

4.2. Exponentially distributed data

The probability function of the observable data, in form of the transformed exponential density function (remember the transformation factor $1/\rho_t$ = 1- t / T) with parameter $\lambda > 0$ (see (4-2)) is

$$f(t)^{\circ} = \begin{Bmatrix} c(\lambda, T) \cdot \left(1 - \frac{t}{T}\right) \cdot e^{-\lambda \cdot t} \text{ if } & 0 \le t \le T \\ 0 & \text{else} \end{Bmatrix}. \tag{4.2-1}$$

The function (4.2-1) has to be normalized. Therefore the constant factor $c(\lambda,T)$ has to be determined. This normalization factor[10] is

$$c(\lambda(T) = \frac{T \cdot \lambda^2}{e^{-\lambda T} + \lambda \cdot T - 1}. \tag{4.2-2}$$

The density function (4.2-1) with the integrated normalization factor can be written as

$$f(t)^{\circ} = \begin{Bmatrix} \frac{\lambda^2}{e^{-\lambda T} + \lambda \cdot T - 1} \cdot (T - t) \cdot e^{-\lambda \cdot t} \text{ if } & 0 \le t \le T \\ 0 & \text{else} \end{Bmatrix}. \tag{4.2-3}$$

The distribution function[11] of the function (4.2-3) is defined for the variable $0 \le t \le T$:

$$\boxed{F(t)^{\circ} = \frac{T \cdot \lambda + \lambda \cdot (t - T) \cdot e^{-\lambda \cdot t} - 1 + e^{-\lambda \cdot t}}{T \cdot \lambda - 1 + e^{-\lambda \cdot T}}}. \tag{4.2-4}$$

The function $F(t)^{\circ} = 0$ and $F(t)^{\circ} = 1$ for t < 0 and t > T respectively. The likelihood function, built from a sample of n independent observed periods $t_1, t_2, \ldots, t_n$ using density function (4.2-3) is

---

[10] Proof: see Appendix A3.

[11] Proof: see Appendix A4.

$$L(t_1,\ldots,t_n \mid \lambda) = \frac{\lambda^{2\cdot n}}{\left(e^{-\lambda T} + \lambda \cdot T - 1\right)^n} \cdot \prod_{i=1}^{n} \left(T - t_i\right) e^{-\lambda \cdot \sum t_i} \qquad (4.2\text{-}5)$$

and the logarithm of (4.2-5) is

$$\ln\left(L(t_1,\ldots,t_n \mid \lambda)^{\circ}\right) = 2 \cdot n \cdot \ln(\lambda) - n \cdot \ln\left(e^{-\lambda \cdot T} + \lambda \cdot T - 1\right) - \lambda \cdot \sum_{i=1}^{n} t_i + \sum_{i=1}^{n} \ln\left(T - t_i\right). \qquad (4.2\text{-}6)$$

To find the maximum likelihood estimator, Equation (4.2-6) has to be differentiated with respect to λ:

$$\frac{\partial \ln(L(t \mid \lambda)^{\circ})}{\partial \lambda} = \frac{2 \cdot n}{\lambda} - \frac{n \cdot \left(-T \cdot e^{-\lambda \cdot T} + T\right)}{e^{-\lambda \cdot T} + \lambda \cdot T - 1} - \sum t_i = \frac{2 \cdot n}{\lambda} - \frac{n \cdot T\left(1 - e^{-\lambda \cdot T}\right)}{e^{-\lambda \cdot T} + \lambda \cdot T - 1} - \sum t_i . \qquad (4.2\text{-}7)$$

Examining (4.2-7) shows that for empirical data, this function has one zero point. To characterise "empirical data" in a more general way, the following inequality[12] (4.2-8) is used. If this inequality is satisfied, the zero point of (4.2-7) will additionally not be in the immediate vicinity of $\lambda = 0$. This inequality will also be called threshold criterion in the following:

$$\frac{\sum t_i}{n \cdot T} < \frac{1}{3} . \qquad (4.2\text{-}8)$$

As the limit of (4.2-7) will obviously be positive for $\lambda \to 0$ and negative for $\lambda \to 1$[13], the zero point is a maximum. Appendix A5 illustrates that (4.2-8) will be met in a special case in which the supposition of the exponential distribution is obviously violated. In that special case, the holding periods are equally distributed.

5. Application to short and leveraged Exchange Traded Funds

The two approaches, using geometrically and exponentially distributed holding periods, were tested using the financial instrument "Exchange Traded Fund" (ETF) with the leverage factors -2, -1 and +2. ETFs with negative leverage factors can be used for hedging[14]. The data were collected by German brokerage institutions (like comdirect bank AG (Quickborn), Deutsche Bank AG (Frankfurt), DAB bank AG (Munich)) for a research project in the year 2010. Appendix 6 depicts the ISIN, leverage factor, and the frequency of the measured holding periods of these ETFs in the sample. The estimation of the holding period in 2010 was done by the observable holding periods, as usual. The summary of the report was published in different financial magazines (e.g. the *Financial Times*

[12] See Appendix A5.

[13] The sufficient condition $2 \cdot n < \sum t_i$ can be met by selecting an adequate scaling for the holding periods t_i.

[14] See e.g. Alexander and Barbosa (2007); Flood (2010); Hill and Teller (2010); Michalik and Schubert (2009); Schubert (2011).

(London)[15], *L'AGEFI* (Paris), *Börse am Sonntag* (München)) and showed that the majority of private investors reduce their holding periods and invested volume according to the risk, measured by leverage factors. Figure 5-1 illustrates, as an example, the distribution of 685 holding periods of ETFs with leverage factor -2. The shape of the distribution is not exactly an exponential one, but similar.

The upper part of Table 5-1 contains the data of the sample differentiated according to the leverage factors of the ETFs. The cases with leverage factor +1 were not included in the report of the German brokerage institutions. Therefore the small sample size of n = 286 of this ETF is missing in the tables of Appendix 6. The time interval T, in which the sample was observed, is depicted in the first row. This row also shows the dates numerically. The following part shows the sample size n and the sum of the holding periods. Their ratio is the traditional mean of the holding periods. In the lower part of Table 5-1, first the maximum likelihood estimators for the geometrical distribution are computed instead of the traditional mean, and then for the exponential distribution. For the determination of the maximum likelihood estimators, the approaches developed in Chapter 4.1 and 4.2 were applied. Before the estimation started, the threshold conditions (4.1-7) and (4.2-8) were tested with the result that these conditions are satisfied. The zero points of the functions (4.1-6) and (4.2-7) offer the estimators p* and λ* which are used to get the estimated

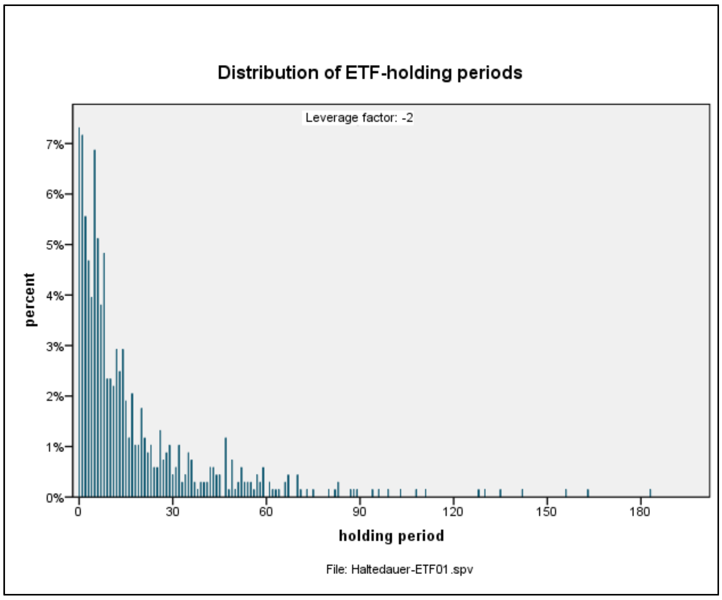

Figure 5-1: Distribution of 685 holding periods with leverage factor -2

[15] See e.g.: http://www.ft.com/cms/s/0/21453158-e670-11df-95f9-00144feab49a,dwp_uuid=d8e9ac2a-30dc-11da-ac1b-00000e2511c8.html

Leverage	**-2**	**-1**	**+1[16]**	**+2**	**Sum**
T *Time interval* *day no.*	224 *20.08.2009-01.04.2010* *40,045 – 40,269*	821 *01.01.2008-01.04.2010* *39,448 – 40,269*	821 *01.01.2008-01.04.2010* *39,448 – 40,269*	821 *01.01.2008-01.04.2010* *39,448 – 40,269*	
Sample size n	685	19,084	286	6,625	26,680
Sum of time $\sum t_i$	12,177	923,014	34,212	311,654	
Mean $\sum t_i / n$ [17]	**17.78**	**48.37**	**119.62**	**47.04**	
Geometric distribution					
$\frac{\sum t_i}{n \cdot T} < \frac{1}{3} \cdot \frac{T^2 + T}{T^2 - 1}$	0.07936 <0.334828	0.058911 <0.337400	0.145703 <0.333740	0.057299 <0.337400	
ML p^*	0.048097	0.018889	0.0064313	0.019453	
$(1 - p^*) / p^*$	**19.79** *(+11.30%)*	**51.94** *(+7.38%)*	**154.49** (+29.15%)	**50.40** *(+7.14%)*	
Exponential distribution					
$\frac{\sum t_i}{n \cdot T} < \frac{1}{3}$	0.07936 < 0.333333	0.058911 < 0.333333	0.145703 < 0.333333	0.057299 < 0.333333	
ML λ^*	0.050839	0.019282	0.00648949	0.019869	
$1 / \lambda^*$	**19.67** *(+10.63)*	**51.86** *(+7.22%)*	**154.10** *(+28.81%)*	**50.33** *(+6.95%)*	

Table 5-1: Data base and maximum likelihood estimators of ETFs 2010 with different leverage factors

"likelihood" mean. The estimated values[18] for the holding period t are $(1 - p^*) / p^*$ and $1 / \lambda^*$. These means are depicted in boldface. As expected, compared with the traditional mean, the estimated means of the holding periods are higher. The differences are shown in italics in line with the estimated average holding periods. The reason for the increase of the holding period lies in the integration of the "semi-outside cases" and "complete-outside cases" described in Chapters 3 and 4. In general, a small time interval T will cause a stronger correction than a bigger time interval. But this time interval has to be regarded in relation to the computed traditional mean, in which the risk of the financial instrument as well as the situation of the economy is included. Although the traditional mean underestimates the mean, it reflects the risk related behaviour of the investor, too. A small traditional mean is a sign that the situation is regarded as risky, and a big mean, as less risky.

Although in the database the time was measured discretely, both approaches produce similar results. Apart from the mean, other parameters, such as the median or the percentiles, can be determined using the developed distribution functions F(t) (see (4.1-3) and (4.2-4)).

---

[16] The ETF with leverage factor +1 was not included in the report for the German broker institutions and refer to only one ETF (ISIN: DE000A0F5UE8: "iShares DJ China Offshore 50 (DE)") which is not listed in Table A6-1.

[17] In the report for the German brokerage institutions 2010 the average holding period is about 0.02-0.04 days bigger. This is due to a higher evaluation of day trading. Instead of 0 days, in that report day trades were counted as 0.5 days.

[18] See, e.g. Bosch (1992), p. 178: $E(t) = 1 / p$ for $f(t)=p\cdot(1-p)^{t-1}$ if $t \in \mathbb{N}$; in the case of $t \in \mathbb{N}_0$ and $f(t)=p\cdot(1-p)^t$ the expected value is $E(t) = 1 / p – 1 = (1 - p) / p$.

Leverage	**-2**	**-1**	**+1**	**+2**	**Sum**
T *Time interval* *day no.*	863 *20.08.2009-31.12.2011* *40,045 – 40,908*	1460 *01.01.2008-31.12.2011* *39,448 – 40,908*	1460 *01.01.2008-31.12.2011* *39,448 – 40,908*	1460 *01.01.2008-31.12.2011* *39,448 – 40,908*	
Sample size n	5,293	39,851	38,710	18,532	102,386
Sum of time $\sum t_i$	177,845	3,140,259	5,369,077	1,154,544	
$\sum t_i /n$	**33.6**	**78.8**	**138.7**	**62.3**	
Geometric distribution					
$\frac{\sum t_i}{n \cdot T} < \frac{1}{3} \cdot \frac{T^2 + T}{T^2 - 1}$	0.03893 < 0.333720	0.05397 < 0.333562	0.09500 < 0.333562	0.04267 < 0.333562	
ML p*	0.027675	0.011766	0.006291	0.015057	
(1 - p*) / p*	**35.13** *(+4.56%)*	**83.99** *(+6.58%)*	**157.97** *(+13.89%)*	**65.42** *(+5.00%)*	
Exponential distribution					
$\frac{\sum t_i}{n \cdot T} < \frac{1}{3}$	0.03893 < 0.333333	0.05397 < 0.333333	0.09500 < 0.333333	0.04267 < 0.333333	
ML λ*	0.028501	0.011916	0.006337	0.015300	
1 / λ*	**35.09** *(+4.43%)*	**83.92** *(+6.49%)*	**157.80** *(+13.77%)*	**65.36** *(+4.91%)*	

Table 5-2: Data base and maximum likelihood estimators of ETFs 2012 with different leverage factors

In 2012 the measurement was repeated with a bigger time interval (01.01.2008 – 31.12.2011) in collaboration with another set of German brokerage institutions and a wider set of ETFs[19]. The total sample size was n = 102,386. This research project contained n = 38,710 holding periods of ETFs with leverage factor +1, too. Due to the greater time interval for the observation of the holding periods (+639 days) and due to the changed behaviour of investors after the financial crisis of 2008, longer holding periods were observed compared to the report of the year 2010 described above. Table 5-2 shows the data and the results of the study of 2012 in the same design as Table 5-1. Due to the bigger time interval T, in that research project the increase of the estimated holding period (in %) is smaller than in the project of 2010. As in Table 5-1, the leverage factor -2 has the smallest holding period, followed by the factor +2. ETFs with leverage factor -1 have the longest holding period among the leveraged ETFs. For the longest time, nearly 158 days, private investors hold unleveraged ETFs (leverage factor +1). As the financial instrument ETF is a replicated stock index, it is comparable to stocks, as shown in Table 2-1 for the years 2008 to 2011. The interval of 0.3 – 0.6 years means about 110 – 220 days.

[19] See Funke, Gebken, and Johanning (2012), pp. 10-11. In the project of the year 2012 day trades were counted as 0.5 days when the average holding periods were calculated. This augmentation of the average holding period could not be corrected in this paper. For the data of the project of 2010, the holding periods of day trades were counted in this paper as 0.0.

The threshold criterion (4.1-7), respectively, (4.2-8), is the traditional mean of the holding periods divided by T. A small value of this threshold signifies that the investor perceives the riskiness of the leverage factor and the economic situation. If additionally the time interval T is big, this value gets smaller again. The smaller the mean is in relation to T, the smaller will be the correction in % of the traditional mean. Tables 5-1 and 5-2 together illustrate this relationship. The threshold values and the %-augmentation have the same rank order. The biggest threshold value (see Table 5-1 leverage factor +1) is 0.145703 with an increase of 29.15% and the smallest is 0.03893 with 4.56% (see Table 5-2 leverage factor -2).

6. Summary

Two approaches to estimate average holding periods by the maximum likelihood method were applied to two data sets, one with 26,680 and the other with 102,386 holding periods. The bigger data set has 639 more days for the observation of the holding periods than the smaller one. According to the bigger time interval, the difference between the traditionally computed mean of the holding period and the maximum likelihood estimators is obviously smaller than in the first sample.

Like the mean of the holding periods, the maximum likelihood estimators indicate that investors take risk into consideration. Although the augmentation of the traditional mean by the estimation procedure was very different (4.43% to 29.15%), the leverage factor -2 has the shortest holding period, followed by the leverage factor +2. The holding period of ETFs with inverse leverage factor -1 is one-half as long as the holding period of ETFs with leverage factor +1.

Holding periods of financial instruments depend on the economy and will differ in bearish and bullish markets. In the case of stocks, this up and down of the average holding periods is depicted in Table 2-1. The mean value of 0.3 to 0.6 years (i.e. 110 - 220 days) shows that the average holding period of stocks is as high as that of unleveraged ETFs.

To estimate actual average holding periods, the estimation must be done in short time intervals. Therefore the estimation of the holding periods should be independent of the time interval T out of which holding periods are taken as a sample. The approaches developed in this paper offer the possibility of estimating the average holding periods independently of T.

The assumed distribution of holding periods seems to be similar to the exponential or geometrical distribution. To get a more flexible fit of the assumed distribution to empirical data, the developed approach can be applied to distributions with more than one parameter.

7. References

[1] Alexander, C., Barbosa, A. (2007): Hedging and Cross Hedging ETFs, International Capital Markets Association (ICMA), Discussion Paper in Finance DP2007-01, January 2007.

[2] Arnold, G. (1965): Formeln der Mathematik, Westermann.

[3] Bamberg, G., Baur, F., Krapp, M. (2012): Statistik, Oldenbourg.

[4] Bosch, K. (1992): Statistik-Taschenbuch, Oldenbourg.

[5] Flood, C. (2010): Risk betting on short and leveraged ETFs, Financial Times (online version), 2nd November 2010.

[6] Funke, C., Gebken, T., Johanning, L. (2012): Investitionsverhalten bei Short- und Leveraged-ETFs - Eine empirische Analyse des Handels deutscher Privatanleger, WHU – Otto Beisheim School of Management (http://www.fondsprofessionell.de/upload/attach/1355396407.pdf).

[7] Hill, J., Teller, S. (2010): Hedging With Inverse ETFs, Journal of Indexes, November/December 2010, pp. 18–24.

[8] Michalik, T., Schubert, L. (2009): Hedging with Short ETFs, Economics Analysis Working Papers, Vol. 8 No. 9.

[9] Opitz, O., Klein, R. (2014): Mathematik, Oldenbourg.

[10] Schubert, L. (2010: Investitionsverhalten bei Short und Leveraged ETFs, HTWG-Konstanz (www.fondsprofessionell.de/upload/attach/8911.pdf).

[11] Schubert, L. (2011): Hedge ratios for short and leveraged ETFs, Atlantic Review of Economics, Vol. 1, 2011, pp. 1-33.

Appendix:

A1: Normalization factor c(p,T)

A2: Distribution function F(t) of the transformed geometrical probability function

A3: Normalization factor c(λ,T)

A4: Distribution function $F(t)^{\circ}$ of the transformed exponential probability function

A5: Effect of equally distributed data

A6: Exchange Traded Funds

A1: Normalization factor c(p,T)

For the normalization of the cumulative distribution function of the transformed geometric probability function (see 4.1-1)

$$f(t) \quad (1-\frac{t}{T})\cdot p\cdot(1-p)^t \text{ for t = 0, 1, …, T} \qquad \text{(A1-1)}$$

the constant factor c(p,T) is used, which includes the parameter p. Therefore the function

$$f(t) = c(p,T)\cdot(1-\frac{t}{T})\cdot(1-p)^t \text{ for t = 0, 1, …, T} \qquad \text{(A1-2)}$$

will be considered. To get the value of the following cumulative distribution function at T, the two factors on the right side of function (A1-2) are transformed and then summarized to

$$F(T) = \sum_{t=0}^{T} f(t) = c(p,T)\cdot\left[\sum_{t=0}^{T}(1-p)^t - \frac{1}{T}\cdot\sum_{t=0}^{T} t\cdot(1-p)^t\right]. \qquad \text{(A1-3)}$$

The two elements in the sum in the bracket will be analysed separately in the following. The <u>first</u> element separates into two terms

$$\sum_{t=0}^{T}(1-p)^t = \sum_{t=0}^{\infty}(1-p)^t - \sum_{t=T+1}^{\infty}(1-p)^t = \sum_{t=0}^{\infty}(1-p)^t - (1-p)^{T+1}\cdot\sum_{t=0}^{\infty}(1-p)^t. \qquad \text{(A1-4)}$$

Now, the sums from t = 0 to t = ∞ in Equation (A1-4) can be joined, to obtain

$$\left[1-(1-p)^{T+1}\right]\cdot\sum_{t=0}^{\infty}(1-p)^t. \qquad \text{(A1-5)}$$

The sum on the right side of (A1-5) is a geometric progression[20] and can be replaced by 1/(1-(1-p)) =1/p with (1-p) < 1. The term with the replaced sum is

$$\left[1-(1-p)^{T+1}\right]\cdot\frac{1}{p}. \qquad \text{(A1-6)}$$

The <u>second</u> sum (without the factor 1/T) in Equation (A1-3) is the differentiation by p of the first element which was transformed above, multiplied by the factor - (1 - p). Therefore the differentiation of the transformed first element (see A1-6) can also be used for the second term:

$$\sum_{t=0}^{T} t\cdot(1-p)^t = -(1-p)\frac{\partial\sum(1-p)^t}{\partial p} = -(1-p)\frac{\partial\left(\left[1-(1-p)^{T+1}\right]\cdot\frac{1}{p}\right)}{\partial p}. \qquad \text{(A1-7)}$$

The right term in Equation (A1-7) can be differentiated, applying the product rule[21] of differentiation. With the factor - (1 – p) this results in

$$-(1-p)\cdot\left(-\frac{1}{p^2}\right)\cdot\left[1-(1-p)^{T+1}\right]-(1-p)\cdot\frac{1}{p}\cdot(T+1)\cdot(1-p)^T. \qquad \text{(A1-8)}$$

---

[20] Opitz and Klein (2014): p. 208.
[21] Opitz and Klein (2014): p. 231.

The rewriting of term (A1-8) compensates the two negative signs in the left product and introduces the factor p/p in the right product, where the factor (1 – p) was included in $(1 – p)^T$:

$$(1-p)\cdot\left(\frac{1}{p^2}\right)\cdot\left[1-(1-p)^{T+1}\right]-\frac{1}{p^2}\cdot p\cdot(T+1)\cdot(1-p)^{T+1}\,. \qquad \text{(A1-9)}$$

Then $1/p^2$ and $(1 - p)^{T+1}$ are shown as separate factors in the following term

$$\left(\frac{1}{p^2}\right)\cdot\left[(1-p)^{T+1}\cdot((-p)\cdot(T+1)-(1-p))+1-p\right] \qquad \text{(A1-10)}$$

which can be formulated in a shorter version:

$$\left(\frac{1}{p^2}\right)\cdot\left[(1-p)^{T+1}\cdot(-1-p\cdot T))+1-p\right]. \qquad \text{(A1-11)}$$

To express the cumulative distribution function (see A1-3) using the results of (A1-6) and (A1-11), the second term must additionally be multiplied by 1/T:

$$F(T)=\sum_{t=0}^{T}f(t)=c(p,T)\cdot\left[\left[1-(1-p)^{T+1}\right]\cdot\frac{1}{p}-\frac{1}{T}\left[\left(\frac{1}{p^2}\right)\cdot\left[(1-p)^{T+1}\cdot(-1-pT))+1-p\right]\right]\right] \qquad \text{(A1-12)}$$

In this function, the factor $1/(T\cdot p^2)$ can be separated, introducing the factor T·p in the first term:

$$c(p,T)\cdot\frac{1}{T\cdot p^2}\cdot\left[T\cdot p\cdot\left(1-(1-p)^{T+1}\right)+(1-p)^{T+1}\cdot(1+p\cdot T)-1+p\right]. \qquad \text{(A1-13)}$$

As the product $T\cdot p\cdot(1 - p)^{T+1}$ has different signs in the first compared with the second term (A1-13) can be designed in a shorter form (see A1-14), which should be 1, to satisfy the necessary condition to be a distribution function:

$$c(p,T)\cdot\frac{1}{T\cdot p^2}\cdot\left[T\cdot p+(1-p)^{T+1}-1+p\right]=1\,. \qquad \text{(A1-14)}$$

To get the value of the normalization factor, Equation (A1-14) must be solved for this factor:

$$\boxed{c(p,T)=\frac{T\cdot p^2}{(T+1)\cdot p-1+(1-p)^{T+1}}} \quad ■ \qquad \text{(A1-15)}$$

A2: Distribution function F(t) of the transformed geometrical probability function

The probability function of the transformed geometrical probability function (see 4.1-2)

$$f(t)=c(p,T)\cdot\left(1-\frac{t}{T}\right)\cdot(1-p)^t \text{ for } t = 0, 1, 2, \ldots, T$$

becomes, with the normalization factor c(p,T) (see A1-15),

$$f(t) = \frac{T \cdot p^2}{(T+1) \cdot p - 1 + (1-p)^{T+1}} \cdot \left(1 - \frac{t}{T}\right) \cdot (1-p)^t \text{ for } t = 0, 1, 2, \ldots, T. \qquad \text{(A2-1)}$$

The cumulative distribution function can be developed as in Appendix A1, where the sum of the two terms in (A1-3) was summarized. In Appendix A1 the sum was calculated over t= 0, 1, …, T. In contrast to the problem of Appendix A1, now in the two sums the variable is $x \in IN_0$ with $x \le t$. Therefore, the variable t has to be replaced in the sum by x and parameter T by t (A1-6) and (A1-11). Doing this, (A1-3) becomes

$$F(t) = \sum_{x=0}^{t} f(t) = \frac{T \cdot p^2}{(T+1) \cdot p - 1 + (1-p)^{T+1}} \cdot \left[\sum_{x=0}^{t} (1-p)^x - \frac{1}{T} \cdot \sum_{x=0}^{t} x \cdot (1-p)^x\right] \qquad \text{(A2-3)}$$

and replacing the variables in (A1-6) and (A1-11) yields

$$\left[1 - (1-p)^{t+1}\right] \cdot \frac{1}{p} \qquad \text{(A2-4)}$$

and

$$\left(\frac{1}{p^2}\right) \cdot \left[(1-p)^{t+1} \cdot (-1 - p \cdot t)) + 1 - p\right]. \qquad \text{(A2-5)}$$

Substituting both elements of the sum in brackets on the right side in the distribution function (A2-3) by (A2-4) and (A2-5) (with the factor 1/T) gives the function F(t) for t = 0, 1, 2, …, T:

$$\frac{T \cdot p^2}{(T+1) \cdot p - 1 + (1-p)^{T+1}} \cdot \left[\left[1 - (1-p)^{t+1}\right] \cdot \frac{1}{p} - \left(\frac{1}{T \cdot p^2}\right) \cdot \left[(1-p)^{t+1} \cdot (-1 - p \cdot t)) + 1 - p\right]\right]. \qquad \text{(A2-6)}$$

Including the factor $T \cdot p^2$ into the left term in brackets, the distribution function of the transformed geometrical distribution can be seen (see 4.1-3)

$$\boxed{F(t) = \frac{(1-p)^{t+1} \cdot (1 + p \cdot t - T \cdot p) + T \cdot p - 1 + p}{(T+1) \cdot p - 1 + (1-p)^{T+1}}} \text{ for } t = 0, 1, 2, \ldots, T \quad \blacksquare \qquad \text{(A2-7)}$$

A3: Normalization factor c(λ,T)

The determination of the normalization factor c(λ,T) uses the density function (4.2-1)

$$f^{\circ}(t) = c(\lambda, T) \cdot \left(1 - \frac{t}{T}\right) \cdot e^{-\lambda \cdot t} = c(\lambda, T) \cdot \left(e^{-\lambda \cdot t} - \frac{1}{T} \cdot t \cdot e^{-\lambda \cdot t}\right) \qquad \text{(A3-1)}$$

which is the basis for the development of the distribution function and the value of this function at T:

$$F(T)^{\circ} = \int_0^T f(t)^{\circ} dt = c(\lambda, T) \cdot \left(\int_0^T e^{-\lambda \cdot t} dt - \frac{1}{T} \cdot \int_0^T t \cdot e^{-\lambda \cdot t} dt\right). \qquad \text{(A3-2)}$$

The left integral in (A3-2) is solved directly and the right integral by partial integration, in which the variable t is function g(x) and $e^{-\lambda t}$ is f′(t). Applying these transformations leads to

$$F(T)^{\circ} = c(\lambda,T)\cdot\left(\left[-\frac{1}{\lambda}\cdot e^{-\lambda\cdot t}\right]_0^T - \frac{1}{T}\cdot\left(\left[-\frac{1}{\lambda}\cdot e^{-\lambda\cdot t}\cdot t\right]_0^T - \int_0^T -\frac{1}{\lambda}\cdot e^{-\lambda\cdot t}dt\right)\right). \qquad \text{(A3-3)}$$

The determination of the left elements in the sum simplifies the function (A3-3) to

$$F(T)^{\circ} = c(\lambda,T)\cdot\left(\frac{1}{\lambda}\cdot\left(1-e^{-\lambda\cdot T}\right)-\frac{1}{T}\cdot\left(-\frac{T}{\lambda}\cdot e^{-\lambda\cdot T} - \int_0^T -\frac{1}{\lambda}\cdot e^{-\lambda\cdot t}dt\right)\right). \qquad \text{(A3-4)}$$

The integral in (A3-4) is nearly identical to the integral on the left side in (A3-2) but without the factor (-1/λ). Therefore, the solution of this integral on the left side of (A3-2), which is the term on the left side of (A3-4), can be used for the remaining integral in (A3-4). The solution of this integral is

$$\int_0^T -\frac{1}{\lambda}\cdot e^{-\lambda\cdot t}dt = \left(-\frac{1}{\lambda}\right)\cdot\frac{1}{\lambda}\cdot\left(1-e^{-\lambda\cdot T}\right) = \left(-\frac{1}{\lambda^2}\right)\cdot\left(1-e^{-\lambda\cdot T}\right). \qquad \text{(A3-5)}$$

Inserting the result of (A3-5) into (A3-4) gives the distribution function

$$F(T)^{\circ} = c(\lambda,T)\cdot\left(\frac{1}{\lambda}\cdot\left(1-e^{-\lambda\cdot T}\right)-\frac{1}{T}\cdot\left(-\frac{T}{\lambda}\cdot e^{-\lambda\cdot T} + \frac{1}{\lambda^2}\cdot\left(1-e^{-\lambda\cdot T}\right)\right)\right) \qquad \text{(A3-6)}$$

which can be rewritten, using for all elements in the sum of (A3-6) the divisor $T\cdot\lambda^2$:

$$F(T)^{\circ} = c(\lambda,T)\cdot\left(\frac{T\cdot\lambda\cdot\left(1-e^{-\lambda\cdot T}\right)+\lambda\cdot T\cdot e^{-\lambda\cdot T} - 1 + e^{-\lambda\cdot T}}{T\cdot\lambda^2}\right). \qquad \text{(A3-7)}$$

Function (A3-7) can be expressed by a neater form, as some terms compensate for others. Furthermore the distribution function value at T has to be normalized:

$$F(T)^{\circ} = c(\lambda,T)\cdot\left(\frac{T\cdot\lambda - 1 + e^{-\lambda\cdot T}}{T\cdot\lambda^2}\right) = 1. \qquad \text{(A3-8)}$$

Solving the function (A3-8) for $c(\lambda,T)$ gives the result

$$\boxed{c(\lambda,T) = \left(\frac{T\cdot\lambda^2}{T\cdot\lambda - 1 + e^{-\lambda\cdot T}}\right)} \;\blacksquare \qquad \text{(A3-9)}$$

A4: Distribution function $F(t)^{\circ}$ of the transformed exponential probability function

The distribution function $F(t)^{\circ}$ of the transformed exponential probability function $f(t)^{\circ}$ can similarly be developed as depicted in (A3-2) to (A3-7). As the function $F(t)^{\circ}$ has to offer for every $0 \leq t \leq T$ the function value, in (A3-2) the variable t has to be replaced by x and T by t in the two integrals:

$$F(t)^{\circ} = \int_0^t f(x)^{\circ}dx = c(\lambda,T)\cdot\left(\int_0^t e^{-\lambda\cdot x}dx - \frac{1}{T}\cdot\int_0^t t\cdot e^{-\lambda\cdot x}dx\right). \qquad \text{(A4-1)}$$

Solving the integrals in (A4-1) directly or by partial integration as in Appendix A3 leads to the equivalent of (A3-7)

$$F(t)^{\circ} = c(\lambda,T)\cdot\left(\frac{T\cdot\lambda\cdot\left(1-e^{-\lambda\cdot t}\right)+\lambda\cdot t\cdot e^{-\lambda\cdot t}-1+e^{-\lambda\cdot t}}{T\cdot\lambda^{2}}\right) \qquad \text{(A4-2)}$$

in which the normalization factor c(λ,T) determined in (A3-9) has to be inserted:

$$F(t)^{\circ} = \left(\frac{T\cdot\lambda^{2}}{T\cdot\lambda-1+e^{-\lambda\cdot T}}\right)\cdot\left(\frac{T\cdot\lambda\cdot\left(1-e^{-\lambda\cdot t}\right)+\lambda\cdot t\cdot e^{-\lambda\cdot t}-1+e^{-\lambda\cdot t}}{T\cdot\lambda^{2}}\right). \qquad \text{(A4-3)}$$

The term (A4-3) can be simplified to

$$\boxed{F(t)^{\circ} = \frac{T\cdot\lambda+\lambda\cdot(t-T)\cdot e^{-\lambda\cdot t}-1+e^{-\lambda\cdot t}}{T\cdot\lambda-1+e^{-\lambda\cdot T}}}\ \blacksquare \qquad \text{(A4-4)}$$

A5: Effect of equally distributed data

The special case of equally distributed data is shown in Figure A5-1. The time period for the sample is restricted to T. In the case of a discrete variable, the frequency n_t of the observable holding period t is depicted by bold red lines and in the case of a continuous variable by the dotted violet line. By the transformation factor ρ_t = T / (T – t), the real frequency (see red points in Figure A5-1) can be estimated.

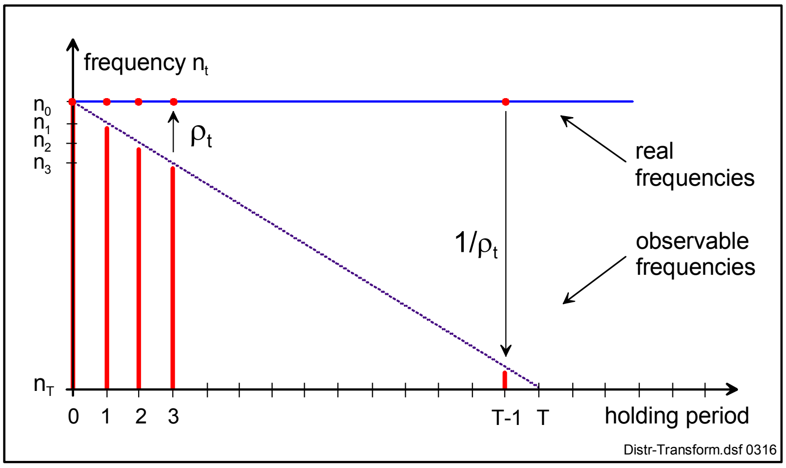

Figure A5-1: Real and observable frequencies in the case of equally distributed data

For the equally distributed data, as shown in Figure A5-1, the value of $\frac{\sum t_i}{n\cdot T}$ will be derived in the following. This will be shown first for (discrete) geometrically distributed holding periods and second for (continuous) exponentially distributed data.

In the <u>discrete case</u>, the total number of observed holding periods is $n=\sum n_t$. As this number is descending with the constant slope n_0/T the number n can be computed by

$$n=\sum_{t=0}^{T}\left(n_0-\frac{n_0}{T}\cdot t\right)=(T+1)\cdot n_0-\frac{n_0}{T}\sum_{t=0}^{T}t=(T+1)\cdot n_0-\frac{n_0}{T}\cdot\frac{(T+1)\cdot T}{2}=\frac{n_0}{2}\cdot(T+1). \qquad \text{(A5-1)}$$

For the computation of the sum of holding periods, the factor t must be multiplied with its frequencies n_t:

$$\sum_{i=1}^{n} t_i = \sum_{t=0}^{T}\left(n_0 - \frac{n_0}{T}\cdot t\right)\cdot t = n_0 \cdot \sum_{t=0}^{T} t - \frac{n_0}{T}\sum_{t=0}^{T} t^2 \,. \qquad \text{(A5-2)}$$

The sums of t and t^2 in (A5-2) are well known series[22] which can be replaced by ratios:

$$n_0 \cdot \frac{(T+1)\cdot T}{2} - \frac{n_0}{T}\cdot\frac{(2\cdot T+1)\cdot(T+1)\cdot T}{6}\,. \qquad \text{(A5-3)}$$

In (A5-3) the first element of the sum can be multiplied by 3/3 and the second fraction can be reduced by T. The multiplication of the terms in brackets leads to

$$\frac{n_0}{6}\cdot\left(3\cdot\left(T^2+T\right)-\left(2\cdot T^2+2\cdot T+T+1\right)\right) \qquad \text{(A5-4)}$$

which can be simplified as

$$\sum_{i=1}^{n} t_i = \frac{n_0\cdot\left(T^2-1\right)}{6}\,. \qquad \text{(A5-5)}$$

With the results of (A5-1) and (A5-5), the term $\frac{\sum t_i}{n\cdot T}$ can be expressed as

$$\frac{n_0\cdot\left(T^2-1\right)/6}{n_0\cdot(T+1)\cdot T/2} \qquad \text{(A5-6)}$$

and finally yields the term

$$\boxed{\frac{\sum t_i}{n\cdot T} = \frac{1}{3}\cdot\frac{T^2-1}{T^2+T}}\ \blacksquare \qquad \text{(A5-7)}$$

In the <u>continuous case</u>, the sample size n is the area of the triangle between the axis and the dotted line of the observable frequencies. The area of this triangle can be computed by

$$n = \frac{n_0\cdot T}{2}\,. \qquad \text{(A5-8)}$$

In the continuous case the $\sum t_i$ has to be expressed by an integral which uses a descending function as in (A5-2)

$$\int_0^T\left(n_0-\frac{n_0}{T}\cdot t\right)\cdot t\cdot dt = \int_0^T\left(n_0\cdot t-\frac{n_0}{T}\cdot t^2\right)\cdot dt = \left[\frac{n_0}{2}\cdot t^2-\frac{n_0}{3\cdot T}\cdot t^3\right]_0^T = \frac{n_0\cdot T^2}{6}\,. \qquad \text{(A5-9)}$$

With the results (A5-8) and (A5-9), the term $\frac{\sum t_i}{n\cdot T}$ can be expressed for the continuous case:

$$\boxed{\frac{\sum t_i}{n\cdot T} = \frac{n_0\cdot T^2/6}{n_0\cdot T^2/2} = \frac{1}{3}}\ \blacksquare \qquad \text{(A5-10)}$$

[22] See, e.g. Opitz and Klein (2014): p. 201 and Arnold (1965): p. 67.

A6: Exchange Traded Funds

Short and leveraged Exchange Traded Funds	ISIN	Lev.	Frequency
ComStage ETF DJ EURO STOXX 50 Leveraged TR	LU0392496930	+2	204
ETFX DAXR 2x Long Fund	DE000A0X8994	+2	91
ETFX Dow Jones EURO STOXX 50 Leveraged (2x) Fund	DE000A0X9AB6	+2	12
Lyxor ETF LevDAX	LU0252634307	+2	5552
Lyxor ETF Leveraged DJ EURO STOXX 50	FR0010468983	+2	766
Sum (leverage factor +2)			**6625**
ComStage ETF DJ EURO STOXX 50 Short TR	LU0392496856	-1	255
db x-trackers CAC 40 Short ETF	LU0322251280	-1	45
db x-trackers DJ EURO STOXX 50 Short ETF	LU0292106753	-1	1808
db x-trackers DJ STOXX 600 Banks Short ETF	LU0322249037	-1	1477
db x-trackers DJ STOXX 600 Basic Resources Short Daily ETF	LU0412624354	-1	18
db x-trackers DJ STOXX 600 Health Care Short ETF	LU0322249466	-1	49
db x-trackers DJ STOXX 600 Insurance Short Daily ETF	LU0412624602	-1	5
db x-trackers DJ STOXX 600 Oil & Gas Short ETF	LU0322249623	-1	116
db x-trackers DJ STOXX 600 Technology Short ETF	LU0322250043	-1	141
db x-trackers DJ STOXX 600 Telecommunications Short ETF	LU0322250126	-1	40
db x-trackers DJ STOXX 600 Utilities Short Daily ETF	LU0412624867	-1	1
db x-trackers FTSE 100 Short ETF	LU0328473581	-1	97
db x-trackers HSI Short Daily Index ETF	LU0429790313	-1	24
db x-trackers II EURO INTEREST RATES VOLATILITY SHORT TOTAL RETURN INDEX ETF	LU0378818727	-1	1
db x-trackers II iTraxx Crossover 5-year Short TRI ETF	LU0321462870	-1	32
db x-trackers II iTraxx Europe 5-year Short TRI ETF	LU0321462102	-1	4
db x-trackers II iTraxx HiVol 5-year Short TRI ETF	LU0321462441	-1	8
db x-trackers II iTraxxEurope Subordinated Financials 5- year Short TRI ETF	LU0378819881	-1	2
db x-trackers II Short iBoxx € Sovereigns Eurozone TR Index ETF	LU0321463258	-1	95
db x-trackers S&P 500 Short ETF	LU0322251520	-1	1191
db x-trackers ShortDAX ETF	LU0292106241	-1	13637
ETFlab DJ EURO STOXX 50 Short	DE000ETFL334	-1	2
Lyxor ETF Short Strategy Europe	FR0010589101	-1	36
Sum (leverage factor -1)			**19084**
EasyETF DJ STOXX 600 Double Short	FR0010689687	-2	8
EasyETF EURO STOXX 50 Double Short	FR0010689695	-2	24
ETFX DAXR 2x Short Fund	DE000A0X9AA8	-2	607
ETFX Dow Jones EURO STOXX Double Short (2x) Fund	DE000A0X9AC4	-2	46
Sum (leverage factor -2)			**685**
Sum			**26394**

Table A6-1: Short and leveraged ETFs with ISIN (International Security Identification Number), leverage factor and frequency of the project of the year 2010[23]

[23] The ETFs of the project of the year 2012 are published in Funke, Gebken, and Johanning (2012), pp. 10-11.