Author/s:
Hirschberg, J;Lye, J

Title:
Estimating risk premiums for regulated firms when accounting for reference-day variation and high-order moments of return volatility

Date:
2021-04-01

Citation:
Hirschberg, J. & Lye, J. (2021). Estimating risk premiums for regulated firms when accounting for reference-day variation and high-order moments of return volatility. Environment Systems and Decisions, 41 (3), pp.455-467. https://doi.org/10.1007/s10669-021-09812-4.

Persistent Link:
https://hdl.handle.net/11343/295951

Estimating Risk Premiums for Regulated Firms when Accounting for Reference Day Variation and Higher Order Moments of Volatility.

Joe Hirschberg and Jenny Lye [1]

Abstract

In many jurisdictions, the determination of the acceptable rate of return for the assets of a regulated utility is based partially on the Capital Asset Pricing Model (CAPM) to forecast risk premia. However, the traditional estimation of CAPM can be been criticised for not including considerations of reference day risk as well as the higher moments of the rates of return. In this paper we attempt to account for both the potential variation induced by the definition of specific reference days and the higher moments of rates of return in the estimates of *Beta*.

This paper provides a new methodology to simultaneously model both the reference day variation and the non-normality of the errors. To account for reference day variation we construct a set of pseudo-monthly rates of return to identify the influence of reference day choice. These pseudo-monthly asset returns are used to estimate measures of asset volatility for an international panel of regulated firms. To evaluate the influence of the higher moments of the rates of return we examine the errors from the estimation of the CAPM with least squares, least absolute value and a partially adaptive maximum likelihood specification.

Key words: bootstrap; return interval; asset volatility; non-normality, meta-analysis

---

[1] Department of Economics, University of Melbourne, Australia, 3010. Corresponding author (j.hirschberg@unimelb.edu.au).

1. Introduction.

Monthly observations are widely used in the construction of observations of monthly rates of change. Measurements of asset returns are often constructed by the difference of the logs of asset prices based on the first trading day of one month to the log of the asset price on the first trading day of the next month. This is typically done without regard to the timing of the initial trading day in the month nor the length of the month. Monthly asset returns are an important input to the estimation of the Capital Asset Pricing Model (CAPM) estimates of *Beta*, where *Beta* is a measure of an asset's volatility relative to the entire market. In many jurisdictions the determination of the acceptable rate of return for the assets of a regulated utility is partially based on the CAPM to forecast risk premia (e.g. McDonald et al 2010, Buckland et al 2015, Gregory et al 2018). Thus, even slight variations in the estimation of the model can result in significant variations in the findings of utility rate judgements. The choice of reference day used to calculate monthly returns has been shown to produce large variations in estimated asset volatility (see eg., Acker and Duck 2007; Dimitrov and Govindaraj, 2007; Gonzalez et al 2014 and Baker etal 2016). In this paper we introduce a new approach to investigate how the choice of reference days and month length has an impact on the inferences from the CAPM values of Beta by constructing pseudo-monthly rates of return by re-sampling from daily data.

The availability of daily asset price data has been the subject of several studies to compare the estimates of the CAPM with returns computed with data of differing frequencies of observation (e.g. Gilbert et al 2014, Gregory et al 2018) however the monthly returns results for the CAPM continue to be widely employed. Although alternative multivariate approaches such as the Fama and French (1993) inspired models have been applied to asset returns data, the CAPM model continues to be widely used due to the reliance on the single parameter estimate of Beta that has traditionally been employed in the evaluation of rates of return for regulated industries.[2] The availability of higher frequency data such as the daily asset prices considered here, has also led to a growing literature where data measured with varying frequencies have been considered in combination as accounted for in the mixed data sampling (MIDAS) models (see

---

[2] The due to its wide application the appropriateness of the traditional single parameter CAPM model is assumed here.

Ghysels et al 2006, 2007).[3] However, these methods involve the requirement for modelling assumptions that generate additional complications for the interpretation of the resulting multiple parameter estimates.

In this paper we generate sets of pseudo-months to estimate the CAPM using two techniques in addition to the traditional OLS approach. These additional techniques are designed to account for extreme values in the time series of rates of return. The first is the Least Absolute Deviation (LAD) regression model (see eg. Sharpe 1971 and Cornell and Dietrich 1978) although this model only accounts for the leptokurtic characteristics of the errors.

The second approach is a partially adaptive estimation procedure based on maximising a log-likelihood function. This technique assumes a flexible error distribution which can be employed in a quasi-maximum likelihood estimation (QMLE). This estimation uses a log-likelihood function corresponding to an approximating flexible error distribution that is maximized over both regression and distributional parameters. This approach provides more efficient estimates than OLS when the errors are non-normal and produces similar results when the errors are normal, see for example McDonald and Xu (1995).

McDonald and Nelson (1989) show that both leptokurtosis and skewness can affect estimates of asset volatility. The flexible error distribution we assume in the QMLE approach allows the modelling of both possible skewness and kurtosis in the estimation of the parameters of the CAPM. Various distributions have been proposed. A number of these are based on extensions of the Student *t* distribution. McDonald, Michelfelder and Theodossiou (2009) use a five-parameter skewed Generalized *t* distribution that accommodates both skewness and thick tails to model the daily returns of US publicly traded stocks. Nested within this distribution is the Generalized Error distribution (GED) which includes as special cases the Laplace, Normal and uniform distributions. Unlike the Generalised *t* distribution all the moments of the GED exist. In this paper we will used the skewed extension of the GED (SGED) distribution from Theodossiou (2015).

This paper proceeds as follows: First we introduce our innovation of the resampling of the daily asset price, and dividend payments data to generate pseudo- monthly returns for each firm along with the

---

[3] These models have been applied in the finance literature as well as the macroeconomics cases as demonstrated in Andreou (2016).

corresponding risk-free return for the same collection of days. Then we use the simulation of multiple sets of pseudo-monthly returns data to estimate the CAPM using OLS, LAD and partially adaptive MLE of the SGED methods. For each of these sets of estimation results we perform a meta-analysis to establish the degree to which the non-normality of the estimated OLS errors from the estimated equation on the original monthly data, influences the differences in the estimates of Beta found when employing the pseudo-months. As far as we know, such a paradigm has heretofore not been done before. We then report the influence of the choice of reference day on the CAPM estimates from different estimation techniques.

2. The construction of Pseudo-Months constructed by Resampling of Daily Asset Prices.

A series of monthly returns are typically based on the change in the price of an asset from the end of one month to the end of the next month[4]. Thus, they are defined as the differences of daily values based on the calendar for the month. However, one could pick any trading day of a month instead of the end of the month. Choosing different days in the month can result in different series of monthly returns. Several authors (Acker and Duck 2007; Dimitrov and Govindaraj 2007; Gonzalez, Rodriguez and Stein 2014 and Baker, Rajaratnam and Flint 2016) have shown that selecting 20 different trading days to construct a series of 20 monthly returns can produce large variations in asset volatility that has been referred to as "reference day variation" (RDV). This variation in the number of trading days in a month will be dependent on the calendar of the country's holiday schedule where the asset is traded.

In order to account for RDV, we generate series of pseudo-monthly returns by randomly defining the start day of the series and the length of each pseudo-month that follows. The pseudo-monthly returns constitute a nonparametric resampling scheme of the monthly returns that is analogous to a block bootstrap scheme such as proposed by Künsch (1989). Unlike a naïve resampling process where we redraw the rates of return with replacement, the innovation here is that these *pseudo-month*s retain their order and thus retain any implicit times series process and can be used to estimate the corresponding monthly CAPM *Beta*.

We define the *n*th month's asset return r_n y_t where $t = 1 \cdots T$

, by $r_n = y_{l_n} - y_{l_{n-1}}$ r_n where $n = 1 \cdots N$. Since the daily data is

[4] The reference day could be the middle of the month such as the 15th as well.

consecutively numbered, the end of each month is defined as day l_n where $l_n = s + \sum_{j=1}^{n} m_j$, s is the number of days from the beginning of the daily series to the start of the first month and m_j is the length of month j.

The usual definition of monthly returns based on the calendar will have the end of a month (i.e. January 31) as the value for s and the corresponding values of m_j defined by the trading day calendar that varies depending on the holidays. We can construct a matrix $\mathbf{P}_{N\times T}$ to define the vector of monthly rates of return $\mathbf{r}_{N\times 1}$ from a vector of the log of daily asset values given as $\mathbf{y}_{T\times 1}$. Using **P**, we can define the monthly vector of returns from the daily vector of the log of asset values as:

$$\mathbf{r}_{N\times 1} = \mathbf{P}_{N\times T}\mathbf{y}_{T\times 1} \tag{1}$$

where **P** is a matrix of the concatenation of two submatrices of the form:

$$\mathbf{P} = \left[\mathbf{S}_{N\times s-1} \;\vdots\; \mathbf{Q}_{N\times T+1-s}\right] \tag{2}$$

where **S** is an N by (s - 1) matrix of zeros for the start of the series: $\mathbf{S} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$ and **Q** is an N by $(T+1-s)$ matrix of pseudo-month lengths: $\mathbf{Q} = \begin{bmatrix} -1 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & -1 & \cdots & 1 & 0 & \cdots & 0 \\ & & & & \vdots & & & \\ 0 & \cdots & \cdots & \cdots & \cdots & -1 & \cdots & 1 \end{bmatrix}$.

The elements of **P** are defined by the vector of month lengths $L = \{l_1, l_2, \cdots, l_N\}$ such that:

$$p_{t,n} = \begin{cases} 0 \text{ if } t \neq l_n \text{ or } t \neq l_{n+1} \\ -1 \text{ if } t = l_n \\ 1 \text{ if } t = l_{n+1} \end{cases}.$$

Consider a definition of $\mathbf{P}_b$ based on a set of random choices for L_b with elements defined by $l_{bt} = s_b + \sum_{j=1}^{t} m_{bj}$ we can then consider the specification of a typical linear model fit to data for the pseudo-months specified as:

$$\mathbf{P}_b\mathbf{y} = \mathbf{P}_b\mathbf{x}\boldsymbol{\beta}_b + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{y}$ and $\mathbf{x}$ are measured daily. The corresponding OLS estimates based on the pseudo-month definition b would be:

$$\hat{\boldsymbol{\beta}}_b = \left(\mathbf{x}'\mathbf{P}_b{}'\mathbf{P}_b\mathbf{x}\right)^{-1} \mathbf{x}'\mathbf{P}_b{}'\mathbf{P}_b\mathbf{y} \tag{4}$$

where the $\mathbf{P}_b{}'\mathbf{P}_b$ matrix acts as a random weighting matrix of the daily data.

The estimates defined in (4) are analogous to the nonparametric bootstrap estimates obtained from sampling with replacement daily values from the original daily series. Because the order of the pseudo-monthly observations is preserved this resampling process is analogous to the non-overlapping block resampling proposed by Künsch (1989), where our blocks are random in size as defined by the distribution of the length of the months.

Table 1 Number of Trading Days in International Data recorded from July 21, 2003 to July 18, 2013.

Trading Days	*%*	*Cum %*
19	0.85	0.85
20	17.80	18.64
21	27.97	46.61
22	37.29	83.90
23	16.10	100.00

To allow for RDV we construct 5,000 sets of 10 years of pseudo-monthly returns by randomly selecting s the starting day for the series and then selecting a series of random month lengths m_j until we have exhausted the daily data available. In this way we have removed the structure that is imposed by the trading calendar to determine the monthly series of returns. Since each country has their own holidays the values recorded are not limited to trading days alone.[5] The asset price from the last open day was copied to the day that may be missing in a country in order to use the same number of days for all firms. The daily asset prices used here were for observations from July 21, 2003 to July 18, 2013 for a potential total of 120 months. Using the daily data, we found that the number of days in the months had the distribution overall the countries trading days shown in Table 1.

[5] Only major holidays such as Christmas and New Year's Day were added to the weekends

To generate the pseudo-months limits l_t the process proceeded in the following manner for simulation *b*:

1. We draw a value for s_b the starting day where $s_b \sim U(1,23)$ and July 21st, 2003 is the first day. (these days exclude weekends)

2. We draw a value for m_{bj} from a distribution based on the percentages of month length given in Table 1 and compute $l_{bt} = s_b + \sum_{j=1}^{t} m_{bj}$.

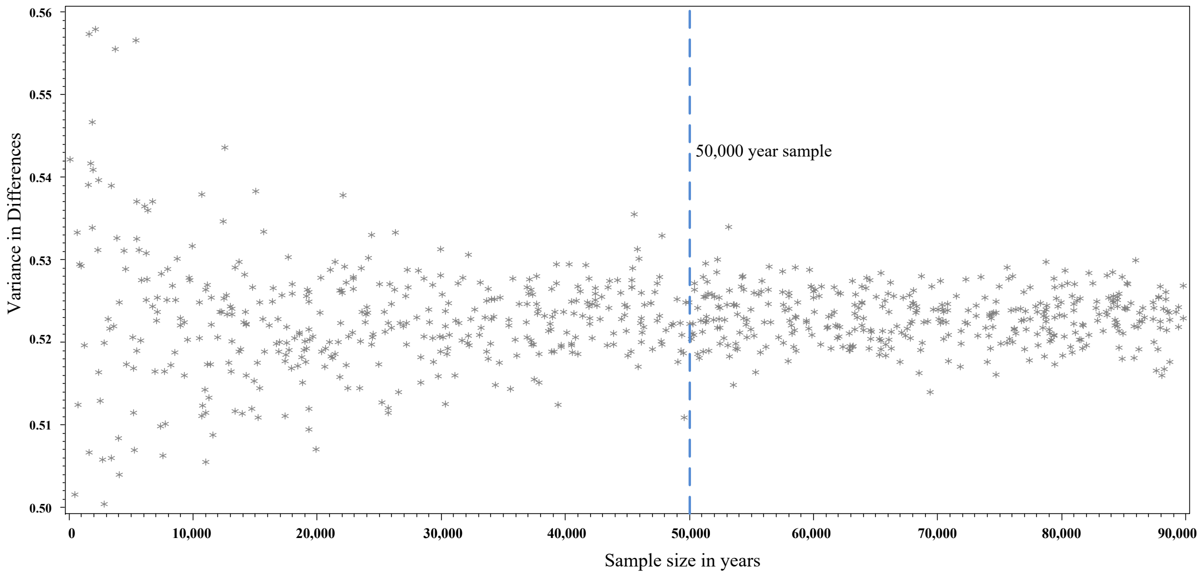
3. We repeat step 2 to draw successive pseudo-month lengths until we have exhausted all the daily values up to July 18, 2013.

The algorithm used for the construction of pseudo-months assigns a pseudo-month number to each day's data for each set of random draws of *s* and *L*. In this way each day in the series has the original calendar month number and a set of pseudo-month numbers. The estimates of the rates of return are then determined from the differences in the log of the values for the last day in each month. Thus, a day may be assigned different order of days within a pseudo-month (in the middle or at the beginning or end). The day could also be moved from one pseudo-month to another.[6]

In order to establish the degree of replication of this random process, we simulated a set of 10.34 billion days (equivalent to 4 million sets of 10 years of potential trading days) that we assigned to pseudo-months. Then, using a uniform distribution we randomly defined a set of samples of sizes from 2 to 90,000 years. Figure 1 is a plot of the variances of the daily differences in the month definitions for various size samples as defined by the number of years of pseudo-months. For example, if the actual month assigned for a day was the 30th month, and the pseudo-month number assigned to that day was 29th then the difference would be -1. From this plot it can be noted that the range of the variance of the differences appears to narrow once we consider samples of 50,000 years and greater. 50,000 years is equivalent to the number of replications we use in our analysis where we use 5,000 sets of 10 years to produce a representative sample of possible pseudo-months.

[6] The simulations of the month definitions for each day are not repeated in the 5,000 cases considered here. The analysis of these allocations is available from the authors on request.

Figure 1. The variance of monthly definitions by sample size in number of years.



To simulate monthly returns for this study, we used the difference between the asset price at the start of the pseudo-month and at the end of the pseudo-month. Then, we added any accumulated dividends paid during the pseudo-month. The market portfolio returns were also based on the same pseudo-month definitions. This entire procedure was done up to approximately 5,000 times for each asset in the data set.

3. Measuring reference-day risk: the estimation of CAPM

In this analysis we investigate the impact of the volatility of the returns on the estimates of *Beta* from the Capital Asset Pricing Model (CAPM) for an international set of firms that can be classified as regulated utilities. These include firms in the energy and transportation industries across countries in Europe the United States and Australasia. The distribution of these firms by country and industry are listed in Table 2. We employ the pseudo-months to perform a resampling so that we may determine the degree to which the calendar definitions of the monthly returns can be included in our comparisons.[7]

We estimate our models using daily data that have been divided into monthly rates of excess returns for various assets and for the market portfolios by country. The risk-free asset employed for each country are the 10-year bond rates of return for the corresponding government bond for each firm. As mentioned above, we include the dividend payments in the returns for an asset when they were paid during the month.

[7] All the daily asset price data, government bond rates, market indices and dividend data were obtained from Bloomberg data services.

Table 3 provides the range of the rates of return for the three firms with the greatest range and the three with the smallest range across the set of pseudo-months. From Table 3 we note that the range of returns based on the pseudo-months is wider than the equivalent range we observe in the original calendar defined months. Thus, demonstrating that we have introduced a greater variance in the observations on returns.

Table 2. The distribution of Firms by Industry and Country

Country	*Industry*							
	Airport	*Coal*	*Energy*	*Port*	*Rail*	*Tollroad*	*Water*	*Total*
Australia	1	2	6	0	1	1	0	11
Austria	1	0	0	0	0	0	0	1
Canada	0	0	6	1	2	0	0	9
France	0	0	0	0	0	2	0	2
Germany	1	0	0	3	2	0	0	6
Greece	0	0	0	2	0	0	0	2
Italy	3	0	0	0	0	2	0	5
New Zealand	1	0	1	4	0	0	0	6
Portugal	0	0	0	0	0	1	0	1
Spain	0	0	0	0	0	1	0	1
Switzerland	1	0	0	0	0	0	0	1
United Kingdom	0	0	4	0	3	0	3	10
United States	0	10	69	0	6	0	10	95
Total	8	12	86	10	14	7	13	150

Table 3 A comparison of the range of the rates of return for the actual months and the pseudo-months.

	Original Calendar Months				**Pseudo-months**			
	Mean	*Min*	*Max*	*Range*	*Mean*	*Min*	*Max*	*Range*
The Smallest Range	0.002	-0.107	0.102	0.209	0.002	-0.175	0.153	0.328
	0.006	-0.157	0.101	0.258	0.006	-0.219	0.125	0.344
	0.002	-0.111	0.099	0.210	0.002	-0.205	0.151	0.356
The Largest Range	-0.025	-0.871	0.667	1.538	-0.027	-1.401	0.831	2.232
	-0.031	-0.770	0.716	1.486	-0.031	-1.387	0.935	2.321
	0.006	-1.661	0.667	2.329	0.003	-1.807	1.062	2.868

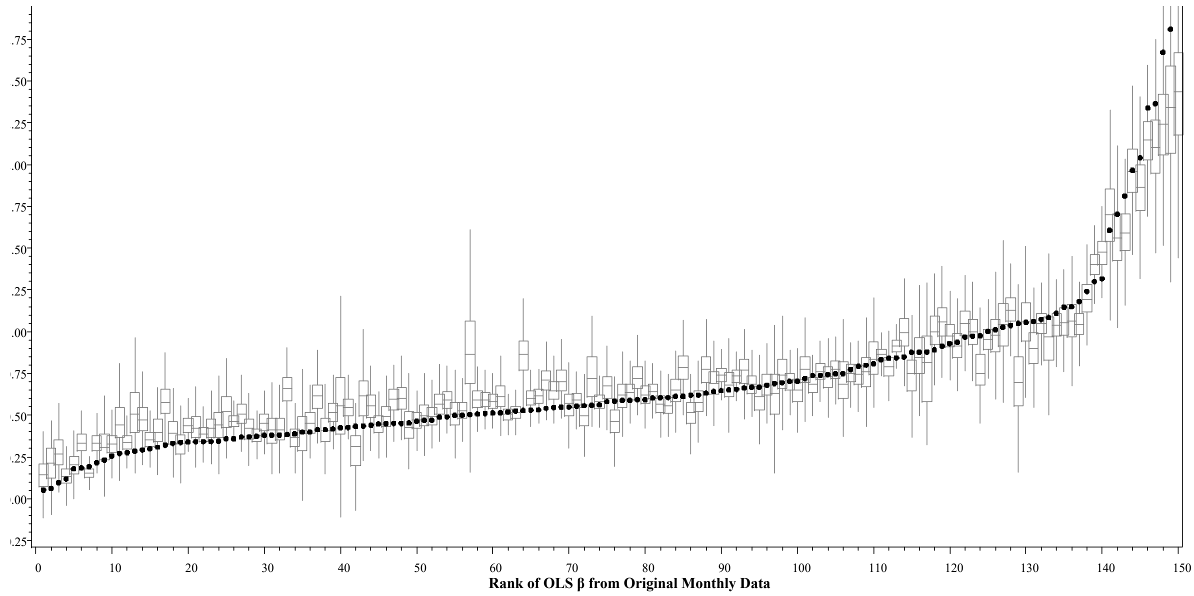
To estimate the CAPM *Beta* we use OLS, LAD and the partially adaptive maximum likelihood estimates under the assumption that the errors are distributed according to the skewed generalized error distribution (MLE-SGED). The estimation via the LAD methods has been proposed as a technique that can mitigate the influence of outliers on the excess returns. The MLE-SGED has been proposed to allow for skewed and leptokurtotic errors. By using the pseudo-month definitions, we can re-estimate estimates of

Beta from the CAPM and to evaluate the conditions when these estimates differ from the estimates obtained from the usual monthly returns.

3.1 OLS estimates of the CAPM

We first estimate the CAPM *Beta* using the original monthly definitions of the returns. Then, to determine the potential for bias in the OLS estimates, we recompute the value of *Beta* using the set of pseudo-months. By comparing the average of the resampled estimates of *Beta* we can measure bias in the original estimates.[8] To examine these, we plot the values of the original OLS estimates with the means of the estimates from the pseudo-months. Figure 2 provides an indication of how the distribution of the OLS estimates of *Beta* based on the original monthly data compare to the distribution of the estimates of *Beta* that we obtain from the pseudo-monthly data. From Figure 2 we have the original estimate as the bold dot with the boxplot of the estimates of *Beta* from the pseudo-months.[9] Note that in some cases the original data estimate is far from the set of *Beta*s estimated from the pseudo-months. These differences appear to be greatest for the lowest and highest values based on the original month definitions.

Figure 2 The OLS estimates of *Beta* based on the original monthly data with the boxplots of the values of *Beta* estimated from the pseudo-monthly data.



[8] Here the bias we refer to can be classed as a form of a bootstrap estimate of bias as described in Efron and Tibshirani (1993).

[9] The boxplots are defined where the upper and lower limits are defined by the 75 and 25 percentiles, the center line is the median and the whiskers extend to the upper and lower bounds.

Figure 3 The relationship between the OLS estimate of *Beta* and the mean of the OLS estimates based on the pseudo-month resampling.

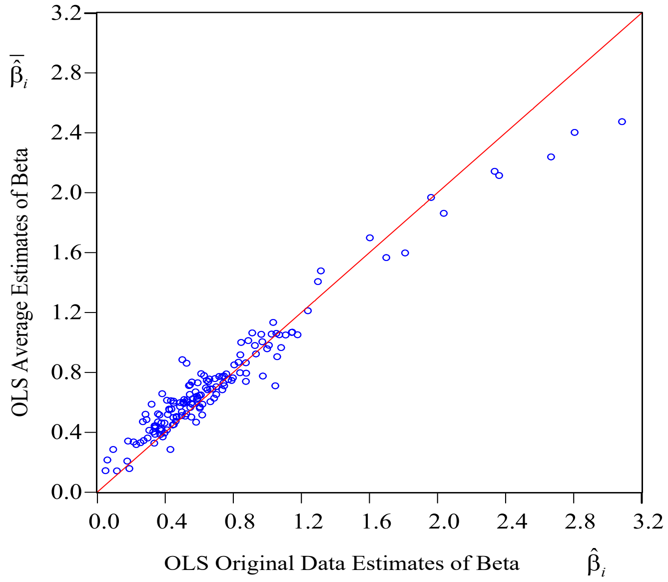


Figure 3 is a plot of the *Beta*s estimated from the original monthly data and the average of the *Beta*s estimated with the pseudo-month data. The diagonal line indicates where the values would fall if these were equal. Those below the line indicate that the OLS estimates are greater than the corresponding average of the OLS estimates based on the pseudo-months. Thus, as in Figure 2 we have a confirmation that there is a tendency for the higher values of *Beta* estimated from the original monthly data to be over-estimated while the lower values are under-estimated.

We also examine the skewness, and kurtosis of the OLS residuals based on the original monthly data as indicators of the influence of non-normality on the OLS results.[10] Table 4 is a cross-tabulation of the tests of significance for excess kurtosis (> 3) and skewness of the OLS residuals.[11] From margins of this table we note that 44% of the firms do not exhibit significant excess kurtosis and 60% do not have skewed residuals. However, of those that exhibit significant skewness over 30% are negatively skewed. From this table we find that (2.67 + 1.33 + 56.00 =) 60.00% of these firms have residuals that either subject to

[10] Several studies have noted that time-varying volatility is not a problem when using low frequency data (e.g. McDonald and Xu 1995, Fong 1997, Mills and Markellos 2008). We also examine the potential influence of a GARCH process in these errors as was investigated by (Mills and Markellos, chapter 8, 2008). Although our resampling method preserves the sequence of the observations and we could include time varying aspects in our models fit to the pseudo-months, we found no indication of the presence of these processes had any significant influence on our results.

[11] We define significance in this case to be 2 standard deviations where the standard deviation for the skewness is given by $\sqrt{6/N}$ and for the excess kurtosis $\sqrt{24/N}$, where N is the number of observations.

significant skewness or excess kurtosis and (28.00 + 8.00 =) 36.00% exhibit both significant skewness and excess kurtosis. When applying the Jarque-Bera test we find the same firms exhibit non-normality.

Table 4 The tests of skewness and excess kurtosis for the OLS residuals.(top is cell %, middle is row % and bottom is column %).

Excess Kurtosis	Skewness			
	Significant Negative	Not Significant	Significant Positive	Total
Not Significant	**2.67**	**40.00**	**1.33**	**44.00**
	6.06	*90.91*	*3.03*	
	8.70	*66.67*	*14.29*	
Significant	**28.00**	**20.00**	**8.00**	**56.00**
	50.00	*35.71*	*14.29*	
	91.30	*33.33*	*85.71*	
Total	**30.67**	**60.00**	**9.33**	**100.00**

To determine the extent to which these indicators of non-normality could be used with the OLS estimate of *Beta* to determine the equivalent bootstrapped value of *Beta* we compute a meta-analysis regression of the form:

$$\begin{aligned}\bar{\hat{\beta}}_i &= \gamma_0 + \gamma_1\hat{\beta}_i + \gamma_{21}\,\mathrm{I}\left(\hat{\theta}_i \le 3\right)\hat{\theta}_i + \gamma_{22}\,\mathrm{I}\left(\hat{\theta}_i > 3\right)\hat{\theta}_i \\ &\quad + \gamma_{31}\,\mathrm{I}\left(\hat{\pi}_i > 0\right)\hat{\pi}_i + \gamma_{32}\,\mathrm{I}\left(\hat{\pi}_i \le 0\right)\hat{\pi}_i + \nu_i, \text{where } \boldsymbol{\nu} \sim (0, \boldsymbol{\Sigma})\end{aligned} \tag{5}$$

The average of the OLS estimates from the pseudo-months $\bar{\hat{\beta}}_i$ is a function of the estimated OLS $\hat{\beta}_i$, the estimated kurtosis of the residuals $\hat{\theta}_i$ is included in two parts – where it is less than 3 and where it is greater than 3 indicating positive excess kurtosis or a leptokurtic case using the indicator function $\mathrm{I}(\bullet)$ that takes a value of 1 when the condition is satisfied and 0 otherwise, the estimated skewness of the residuals $\hat{\pi}_i$ is also included in two parts – for positive skewness and for negative skewness also defined by the indicator function. We assume that the covariance matrix of the error in this model $(\boldsymbol{\Sigma})$ is a diagonal matrix with the estimated variance of the $\bar{\hat{\beta}}_i$ on the diagonal in this way we are able to down-weight those averages that are estimated with less accuracy.[12]

[12] We estimate the model as a GLS that also allows for heteroscedastic errors to also account for the use of computed regressors in the model.

Table 5. The results of the regression for equation (5) (robust standard errors are reported.

Variable	*Coefficient*	*Std. Error*
$\hat{\gamma}_1$	**0.886**	**0.024**
$\hat{\gamma}_{21}$	-0.009	0.008
$\hat{\gamma}_{22}$	**-0.011**	**0.005**
$\hat{\gamma}_{31}$	0.039	0.039
$\hat{\gamma}_{32}$	**-0.070**	**0.028**
$\hat{\gamma}_0$	**0.126**	**0.018**
Adj R^2		.927

From Table 5 we note that the parameter estimates on the OLS value of *Beta* is significantly less than 1 indicating that the slope of the relationship between these two estimates as could be found in Figure 3, would be flatter than the 45-degree line provided for reference. Table 5 also shows a negative estimated significant parameter for excess kurtosis $\hat{\gamma}_{22}$ as well as a significant negative parameter for the influence of negative skewness $\hat{\gamma}_{32}$. This result implies that if the residuals are negatively skewed, the larger the absolute value of the skewness the greater the impact of the skewness on $\bar{\hat{\beta}}_i$. The average of the negative skewness is -.57 which would imply that the impact of the negative skewness would be a .04 increase over the OLS estimate of *Beta*. The impact of the excess kurtosis is to reduce the value of $\bar{\hat{\beta}}_i$. If we assume the mean excess kurtosis of 5.59, we have an average impact of-.06.

By inserting the estimates in Table 5 for γ_{22}, γ_{32} and the data in the formula defined by: $D_i = \hat{\gamma}_{22}\, \mathrm{I}\left(\hat{\theta}_i > 3\right)\hat{\theta}_i + \hat{\gamma}_{32}\, \mathrm{I}\left(\hat{\pi}_i \le 0\right)\hat{\pi}_i$ we find that the average combined impact of these two factors ranges from -.128 to .039 with an average of -.023. This implies that the combination influence of excess kurtosis and negative skewness has an average 5.09% overall inflationary impact on the original monthly OLS estimates of *Beta* on those firms that have either leptokurtic residuals or negatively skewed residuals or both.

3.2 Least Absolute Deviation (LAD) estimates of the CAPM

The LAD method employs the minimization of the sum of the absolute value of the errors in a regression equation (see for example Sharpe 1971 and Cornell and Dietrich 1978). Implicitly, this method reduces the influence of outlier or extreme values of the dependent variable on the resulting parameter estimates. The LAD estimate can be viewed as resistant to error process with significant levels of kurtosis since these errors would be down-weighted in the estimation process. Figure 4 is the equivalent plot to

Figure 2 for the LAD estimates of *Beta*. Note that as in Figure 2, this plot indicates that the pseudo-month estimates tend to be less extreme than those based on the original calendar-based data.

Figure 4 The LAD estimates of *Beta* based on the original monthly data with the boxplots of the values of *Beta* estimated from the pseudo-monthly data.

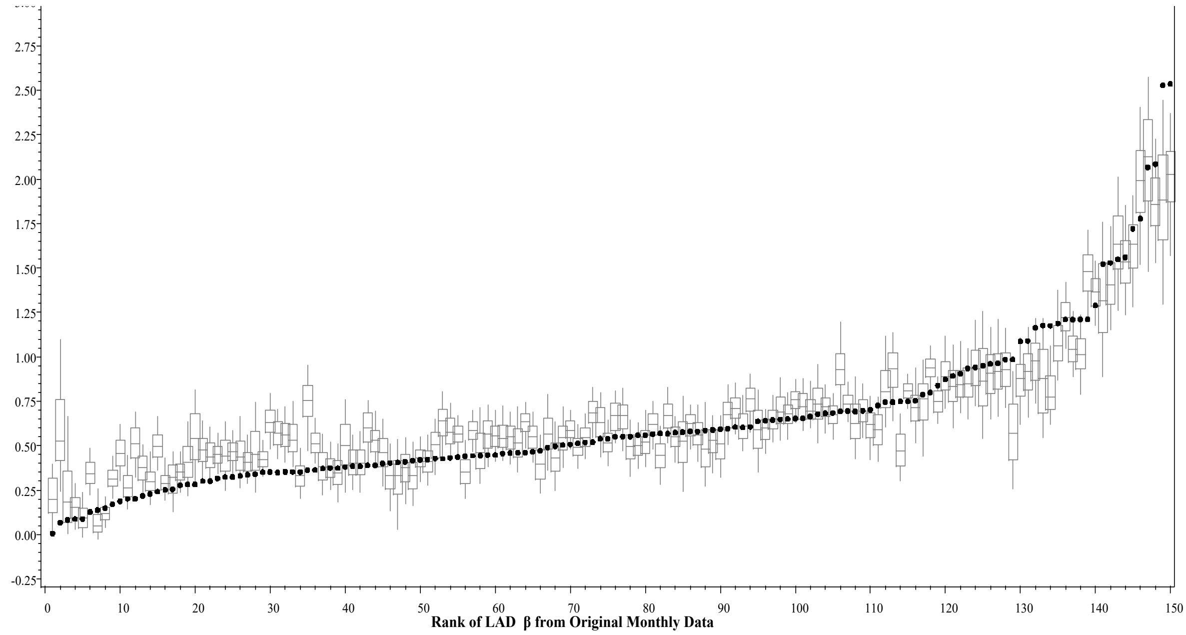


Figure 5. The mean of the LAD estimates of *Beta* based on the Pseudo-months versus the LAD estimates from the monthly observations.

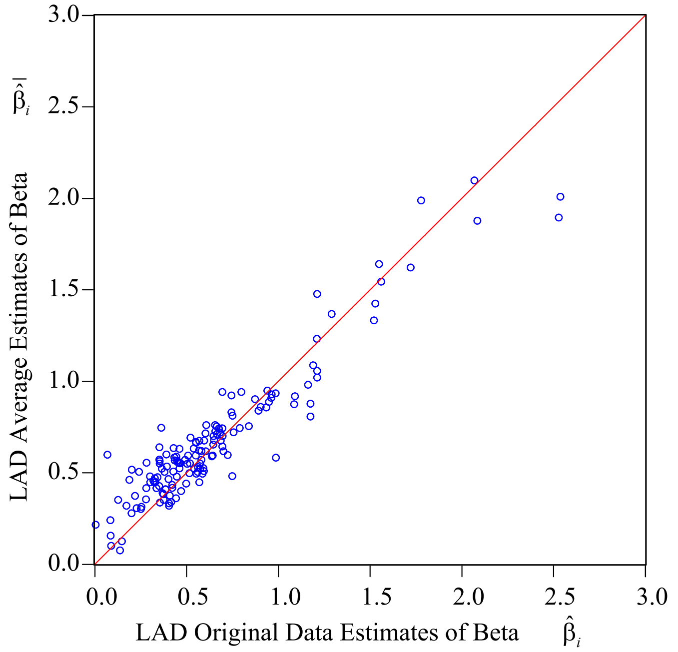


Figure 5 is an equivalent scatter plot to Figure 3 of the LAD estimates of *Beta* versus the average of the estimates from the application of the LAD method to the pseudo-months. From this plot we note that only two assets appear to have much higher estimated values than the mean of the data. We also note that the lower estimates of *Beta* from the actual months are estimated from the pseudo-month data as higher.

Thus, indicating that the LAD method is likely to be subject to potential estimation bias due to calendar influences in much the same manner as the OLS estimates.

We re-estimate (5) using the LAD estimates for $\bar{\hat{\beta}}_i$ and $\hat{\beta}_i$, the result is reported in Table 6. As in the model for OLS results reported in Table 5, the regression errors were assumed to be distributed according to a diagonal covariance matrix with the estimated variance of the average LAD estimates. In this case, we also find that the estimate for the slope parameter on the LAD monthly estimate $\hat{\gamma}_1$ is significantly less than 1.

Table 6. The results of the regression for equation (5) using the LAD estimates of *Beta* (reporting robust standard errors).

Variable	*Coefficient*	*Std. Error*
$\hat{\gamma}_1$	**0.807**	**0.031**
$\hat{\gamma}_{21}$	-0.012	0.009
$\hat{\gamma}_{22}$	-0.009	0.005
$\hat{\gamma}_{31}$	-0.015	0.064
$\hat{\gamma}_{32}$	**-0.079**	**0.031**
$\hat{\gamma}_0$	**0.163**	**0.019**
Adj R^2		0.863

From Table 6 we note that unlike the results for the OLS estimates found in Table 5 we now find that the LAD estimates are not significantly influenced by the leptokurtic residuals. However, they are subject to influence by the negatively skew in the errors as indicated by the negative skew in the residuals. In this case we find that the estimate of γ_{32} is slightly smaller than the equivalent estimate we found in the OLS case as reported in Table 5. The average percent impact on the original monthly LAD estimate is an 8.81% over-estimate based on the negative skewed impact (note this only applies to those firms with negatively skewed residuals).

3.3 The higher order moments in the estimation of the CAPM using the skewed generalized error distribution.

To establish the influence of both skewness and kurtosis of the errors in the estimation of the CAPM when applied to monthly asset prices we consider the partially adaptive maximum likelihood estimation of the SGED (Hansen, McDonald and Theodossiou, 2007). The advantage of the SGED is that it can be used to estimate three parameters for the errors. These three parameters allow the errors to exhibit non-normal levels of skewness and kurtosis as well as the scale of the variance.

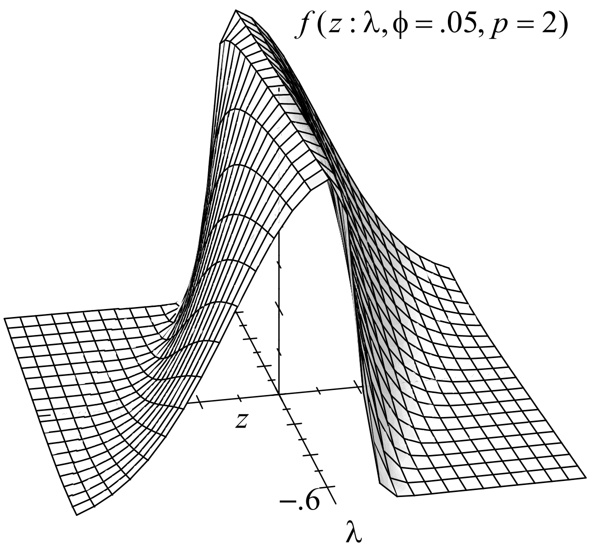
The SGED density function for a random variable *z* can be defined by:

$$f(z:m,\lambda,\phi,p)=p\left(2\phi\Gamma\left(p^{-1}\right)\right)^{-1}\exp\left\{-\left|z-m\right|^{p}\left(\left(1+\tfrac{(z-m)}{|z-m|}\lambda\right)^{-p}\phi^{-p}\right)\right\} \qquad (6)$$

In the case of the estimation of the CAPM we define the model where *z* is the monthly excess returns (*r*) and *m* is the linear combination of the market rate of excess returns (*x*) as: $r=\alpha+\beta x+\varepsilon$. The values of the parameters (λ,ϕ,p) determine the shape of the distribution.

For symmetric errors, the value of $\lambda=0$ and for non-symmetric error $\lambda\neq 0$ where the sign indicates the sign of the skewness in the errors. This distribution can also indicate the degree of kurtosis in the error. The value of *p* is negatively related to the degree of kurtosis present in the error distribution. For symmetric distributions, a value of $p = 1$ indicates a Laplace distribution as estimated via the LAD procedure and $p = 2$ indicates a normal distribution. The scale of the variance is related to the value of ϕ.[13] Figure 6 plots the pdf shape with various values of λ to allow for skewness and Figure 7 plots the pdf shape for various values of *p* to demonstrate how the distribution shape ranges from peaked to normal by changing *p*.

Figure 6. The SGED pdf for different values of λ when $p = 2$ and ϕ=.05 demonstrating how the sign of λ determines the nature of the skewness.



[13] The MLE were obtained using the OLS estimates for α and β as starting values with restrictions on the values of .05 to 10.0 for *p,* -1 to 1 for λ and 0 to 1 for ϕ. The optimization was done with a Newton-Raphson method as implemented in the *nlpnra* routine in SAS *Proc IML*.

Figure 7 The SGED pdf for values of *p* from .5 to 2.5 when $\lambda = 0$ and ϕ=.05.

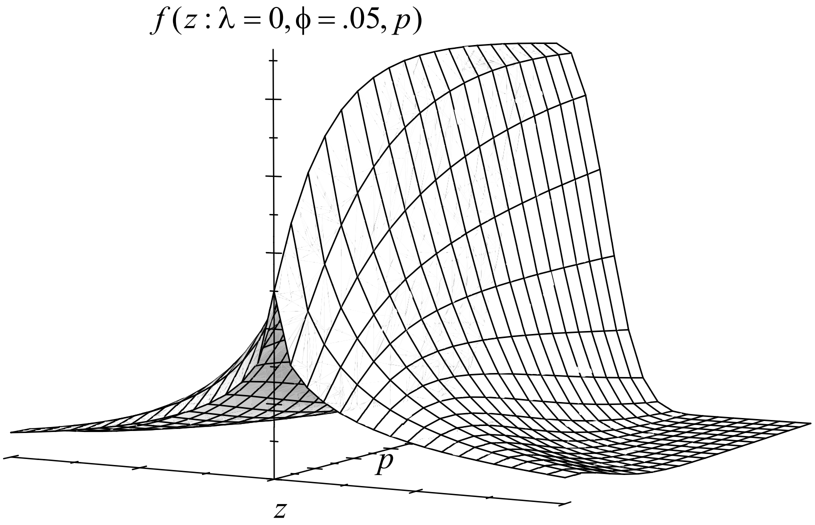


Table 7 lists the cross tabulation of the tests for skewness of the OLS residuals based on the original monthly data and the estimates of the λ parameter from the SGED distribution. Note that the sign of the estimated parameter mirrors the sign of the skewness. The diagonal values in this table indicate that (14.67 + 51.33 + 4.67 =) 70.67% of the skewness in the OLS residuals is mirrored in the estimates of λ in the SGED distribution.

Table 7 The cross tabulation of the estimated values of the parameter λ versus the skewness of the OLS residuals.

	Skewness of OLS Residuals			
λ	***Significantly Negative***	***Not Significant***	***Significantly Positive***	***Total***
Significantly Negative	**14.67**	6.00	0.00	20.67
	70.97	*29.03*	*0.00*	
	47.83	*10.00*	*0.00*	
Not Significant	14.00	**51.33**	4.67	70.00
	20.00	*73.33*	*6.67*	
	45.65	*85.56*	*50.00*	
Significantly Positive	2.00	2.67	**4.67**	9.33
	21.43	*28.57*	*50.00*	
	6.52	*4.44*	*50.00*	
Total	30.67	60.00	9.33	100.00

Figure 8 shows the boxplots for the pseudo-month estimates of the partially adaptive MLE estimates of the SGED for the *Beta* along with the MLE estimates using the original monthly data definitions as we have in Figures 2 and 4 for the OLS and LAD estimates.

Figure 8 The MLE estimates using the SGED of *Beta* based on the original monthly data with the boxplots of the values of *Beta* estimated from the pseudo-monthly data.

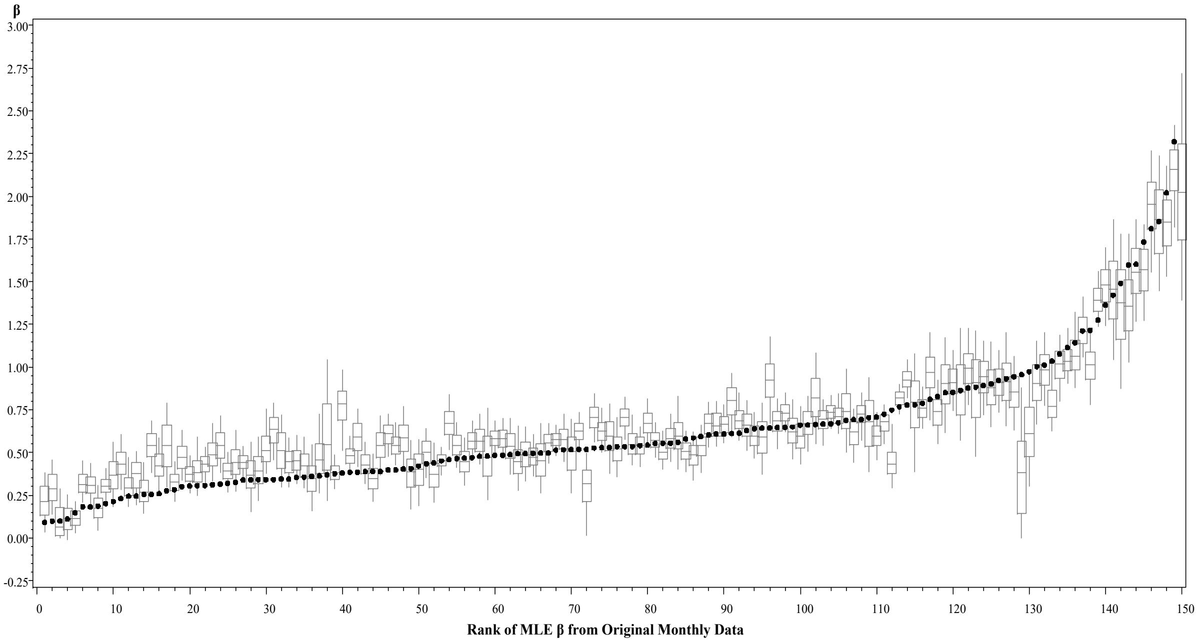


Figure 9 is a plot of the partially adaptive MLE estimates of the SGED for the *Beta* from the CAPM using the original monthly data versus the average of these estimates from the application of this model to the pseudo-monthly data. The deviation of the monthly parameter estimates from the average of the pseudo-month estimates appears to be similar to our findings for both the OLS or the LAD estimates in Figures 2 and 3.

Figure 9. The mean of the MLE SGED estimates of *Beta* based on the Pseudo-months versus the MLE SGED estimates from the monthly observations.

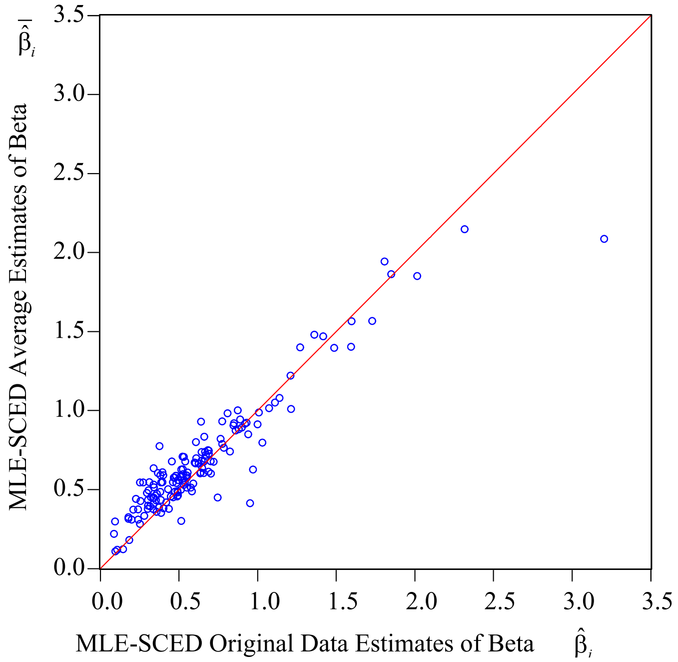


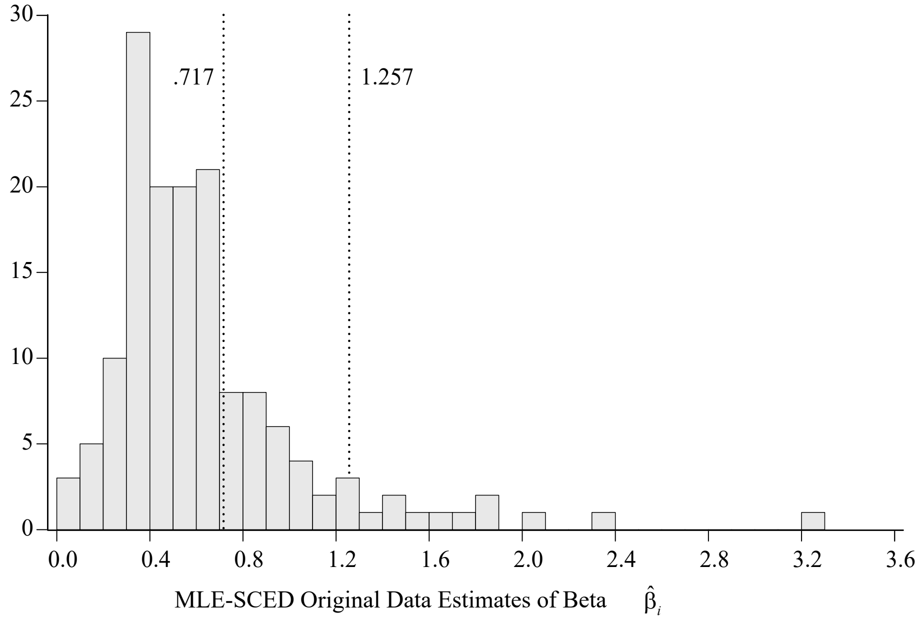
Table 8 lists the parameter estimates for the application of (5) to the average value of the estimated *Beta* from the MLE-SGED model as applied to the pseudo-samples and the MLE-SGED model estimates obtained from the original monthly observations data. In this case we find that the influence of the skewness and kurtosis of the OLS residuals have no significant influence on the difference between the average *Beta* estimate and the original monthly estimate. This would indicate that the estimates obtained for the MLE-SGED are not prone to influences of the skewness and kurtosis of the errors as modelled by the OLS residuals and that the estimates of the extra parameters to account for skewness and kurtosis in this model are able to model the influence of these higher order moments of the data.

Table 8. The results of the regression for equation (5) using the MLE of the SGED estimates of *Beta*.

Variable	*Coefficient*	*Std. Error*
$\hat{\gamma}_1$	**0.849**	**0.029**
$\hat{\gamma}_{21}$	-0.008	0.008
$\hat{\gamma}_{22}$	-0.007	0.005
$\hat{\gamma}_{31}$	-0.016	0.065
$\hat{\gamma}_{31}$	-0.039	0.029
$\hat{\gamma}_0$	**0.149**	**0.020**
Adj R^2		0.873

In Table 8, the estimates for γ_1 and γ_0 are the only ones for which we can reject the null hypothesis that they are zero. From this result we can conclude that the differentials between the two estimates are not influenced by the skewness and kurtosis of the original OLS residuals, and that all the differences between the bootstrapped estimate of *Beta* and the *Beta* estimated on the original monthly data is due to the RDV that is modelled by the pseudo-monthly data. Given this, one could use the monthly data estimates of the *Beta* from the MLE-SGED to predict the equivalent pseudo-month estimate of *Beta*.

Figure 10 The distribution of the MLE-SGED estimates of Beta based on the original monthly data.



Using these parameter estimates from Table 8, we can determine how the estimates of *Beta* from the usual monthly observations differ from those we can estimate using the pseudo-monthly data to establish the bias - when the monthly estimate is too low and when it is too high. From Figure 9 we note that once the monthly *Beta*s reach a little over 1 they over-estimate the average value we obtain using the pseudo-monthly data. We can determine this turning point as the value of $\hat{\gamma}_0 / (1-\hat{\gamma}_1)$, which in this case equal to .987 with an approximate standard error of .135. This would imply that estimates of *Beta* based on the original monthly data that fall within a two standard deviation interval of .717 to 1.257 would be equivalent to those *Beta*s that take the reference day variation into account. However, outside this range an analysis that accounts for reference day variation would lower the higher values and raise the lower ones. From Figure 10 we note approximately 81% of the estimated *Beta*s using the MLE-SGED are outside the range of values where the RDV has no impact.

4. Conclusions

In this paper we applied a new measure of pseudo-monthly rates of return that are constructed by bootstrapping a sampling from daily data to construct monthly-equivalent rates of return. These were used to establish the degree to which the calendar-based sampling structure influence the estimated parameters from the models used to determine asset volatility. Using the pseudo-monthly rates of return, we estimated measures of asset volatility for an international panel of regulated firms using a series of estimation techniques, including least squares, least absolute deviation and partially adaptive maximum likelihood. This approach allowed us to isolate the influence of reference day variation from the non-normality of the errors on the potential bias of the measures of asset volatility based on monthly data.

Our analysis showed that the reference day variation created by calendar effects influenced the results of the application of the CAPM to monthly observations by generating multiple versions of the definition of the months. Using three different models for the estimation of the CAPM *Beta* we found that the calendar influence appears to result in more extreme values than those that are estimated when the days are divided in a random definition of the length and beginning of the month.

In addition, we also noted that both the presence of negative skewness and leptokurtic errors have an influence on the degree to which the OLS CAPM estimates are subject to positive bias. We found that by estimating the CAPM model with the LAD model we could account for the leptokurtic nature of the errors, but results are still influenced by the negative skewness. The partially adaptive MLE estimate of the SGED of the CAPM model accounted for both skewness and kurtosis appears to have removed the influence of both the negative skewness and leptokurtic nature of the errors on the resulting parameter estimates so that we were able to isolate the influence of the reference day variation created by the calendar effects.

Our results indicate that future monthly estimates of the CAPM using models such as the SGED which can account for the presence of both skewed and leptokurtic errors should be used for the estimation of *Beta*. From the meta-analysis reported in Table 8 we isolated the influence of the reference day variations on the estimates of *Beta*. We showed that if the typically defined monthly returns data are used, while there will be a range of estimates over which reference day variations will be negligible, outside this range not taking reference day variations into account will result in over- estimates and under-estimates of *Beta*.

Additionally, we recommend that models of asset return that are based on monthly observations can be evaluated in a similar manner as we have shown here. In addition, the MLE-SGED method can also be employed to account for higher order moments in asset returns in models that include additional covariates.

References

Acker D. and N. Duck (2007), “Reference-day risk and the use of monthly returns data”, *Journal of Accounting Auditing and Finance*, 22, 527–557.

Andreou, E., (2016), "On the use of high frequency measures of volatility in MIDAS regressions", *Journal of Econometrics*, 193, 367-389.

Baker, C., K. Rajaratnam and E. Flint (2016), “*Beta* Estimates of shares on the JSE Top 40 in the context of Reference-Day Risk”, *Environment Systems and Decisions*, 36, 126-141.

Buckland, R., J. Williams, J. and Beecher (2015), “Risk and regulation in water utilities: a cross-country comparison of evidence from the CAPM”, *Journal of Regulatory Economics,* 47, 117-145.

Cornell, B. and J. Dietrich (1978), “Mean-Absolute-Deviation Versus Least-Squares Regression Estimation of *Beta* Coefficients,” *Journal of Financial and Quantitative Analysis*, 13, 123-131.

Dimitrov V and S. Govindaraj (2007), “Reference-day risk: observations and extensions”, *Journal of Accounting Auditing Finance*, 22, 559–572.

Efron, B. and R. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York, NY.

Fama, E. F., and K. R. French (1993), "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics,* 33, 3–56.

Fong, W. (1997), “Robust *Beta* estimation: Some empirical evidence”, *Review of Financial Economics*, 6, 167-186.

Ghysels, E., P. Santa-Clara, and R. Valkanov, (2006), "Predicting volitity: getting the most out of returned data sampled at different frequencies", *Journal of Econometrics*, 131, 59-95.

Ghysels, E., A. Sinko, R. Valkanov, (2007), "MIDAS Regressions: Further Results and New Directions", *Econometric Reviews*, 26, 53-90.

Gilbert, T, C Hrdlicka, J. Kalodimos, S. Siegel, (2014), "Daily Data is Bad for Beta: Opacity and Frequency-Dependent Betas", *The Review of Asset Pricing Studies*, 4, 78-117.

Gonzalez, M., A. Rodriguez and R. Stein (2014), “Adjusted *Beta*s under reference-day risk”, *The Engineering Economist*, 59, 79–88.

Gregory, A., S. Hua, and R. Tharyan, (2018), "In search of beta", *The British Accounting Review*, 50, 425-441.

Hansen, C., J. McDonald, and P. Theodossiou (2007), “Some Flexible Parametric Models for Partially Adaptive Estimators of Econometric Models”, *economics,* No 2007-7, www.economics-ejournal.org/economics/journalarticles.

Künsch, H. (1989), “The Jackknife and the Bootstrap for General Stationary Observations”, *The Annals of Statistics*, 17, 1217-1241.

McDonald, J. and R. Nelson (1989), “Alternative *Beta* estimation for the market model using partially adaptive techniques”, *Communications in Statistics*, 18, 4039- 4058.

McDonald, J., and Y. Xu (1995), “A generalization of the beta distribution with Applications”, *Journal of Econometrics*, 66, 133-152.

McDonald, J., R. Michelfelder and P. Theodossiou (2009), “Robust Regression Estimation Methods and Intercept Bias: A Capital Asset Pricing Model Application”, *Multinational Finance Journal*, 13, 293–321.

McDonald J., R. Michelfelder and P. Theodossiou (2010), “Robust Estimation with flexible parametric distributions: estimation of utility stock betas”, *Quantitative Finance*, 10, 375-387.

Mills, T. and R. Markellos (2008), *The Econometric Modelling of Financial Time Series*, 3rd ed., Cambridge University Press, Cambridge, UK.

Sharpe, W. (1971), "Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios", *Management Science*, 18, B1-B13.

Theodossiou, P. (2015), "Skewed generalized error distribution of financial assets and option pricing", *Multinational Finance Journal*,19, 223-266.