Time-Varying Beta: 
Heterogeneous Autoregressive Beta Model

Kunal Jain
Spring 2010

Economics 201FS
Honors Junior Workshop in Financial Econometrics
1 Introduction

Beta is a commonly defined measure of market risk, which essentially measures the volatility of returns on assets or securities and co-movements of the market portfolio. Conventionally, the Beta coefficient of the market model in security analysis is widely accepted as a relevant valuation of risk in a portfolio. Many members of the investment community have accepted Beta as a practical and convenient methodology to determine the risk of securities.

Traditionally, the Capital Asset Pricing Model (CAPM) uses a constant beta computed over a specified time horizon, typically months. Banz (1981) cites the conventional measure of estimating betas through monthly returns over a 5-year time horizon. But, recent criticism has suggested that a static CAPM is unable to satisfactorily explain the cross-section of average returns on stocks. An essential feature of the beta-metric is the degree of predictability it lends to assess future portfolio risk and returns. Given a lack of predictable power in the beta-metric, portfolio managers would not be able to forecast returns. This inability to forecast returns would surmount the applicability of this market-theory to be restricted (Klemkosky and Martin, 1975).

Given the theoretical relevance of beta predictability in securities analysis, many researchers have attempted to predict betas using different extrapolative methods. Harvey (1989) proposed tests of asset pricing models that allowed time variation in conditional covariance to capture the dynamic behavior of asset returns. Ferson and Harvey (1993) then took this methodology one step further in examining national equity market returns in relation to global economic risks to examine the global risk premia due to time variation. In refining the CAPM to harvest conditional characteristics, Jagannathan and Wang (1996) assumed that betas and the market risk premium vary over time to explain the cross-section of average returns. Andersen, Bollerslev, Diebold and Wu (2004) review this panel of literature to assess the dynamics in realized betas, utilizing the dynamics of realized market variance and individual equity covariances with the market. This recent research suggests that a static beta-model may produce incoherent results.

This paper provides a further look into a time-varying beta analysis that calculates realized betas. Utilizing these realized betas, calculated using high-frequency data, this paper seeks to evaluate the coherency of a Heterogeneous Autoregressive Beta-model that applies daily, weekly and monthly realized betas to find a conditional beta prediction.
First, in Section 2, this paper develops the elements of a time-varying beta model by explaining the concept of a beta and realized betas. Then, in Section 3.1 and 3.2, it provides the motivation for a time-varying beta by citing elements of standard deviation and first order autocorrelations. Next, in Section 4, it builds the theoretical framework of the HAR-Beta model and cites the impetus for important inputs, including sampling frequencies to decrease market microstructure noise. Section 4.4 completes the theoretical framework by explaining how the beta predictions are made within the HAR-Beta model.

Section 5 then provides a brief description of the high frequency data used, including the in sample and out of sample time intervals. To normalize the results, Section 6 provides benchmark comparisons of a constant returns model and a constant beta model. To examine these results, the development of mean squared errors is made in Section 7 to explain the statistical methods. Section 8 provides the in sample and out of sample results that are supplemented by final conclusion in Section 9.

2 The Elements of a Time-Varying Beta

2.1 Basic Variables

In order to understand the elements of stock price and market variation over time, it is necessary to obtain a scale of measurement. The units used within this scale of measurement will be the logarithmic price:

\[ p(s_i) = \log(m(s_i)) \]  (1)

where \( m(s_i) \) is the realized market price of an equity at time \( i \).

Subsequently, the logarithmic of returns at a specified time interval will be the scale:

\[ r_i = p(s_i) - p(s_{i+t}) \]  (2)

where \( p(s_i) \) denotes the realized logarithmic market price of an equity at time \( i \) and \( t \) represents the specified time interval. The impetus for a logarithmic scale is the comparability in percentages of change between two data points that logarithmic scales provide.

2.2 Beta

The Beta Coefficient (\( \beta \)) is a key parameter in the Capital Asset Pricing Model. This can be represented through the Security Characteristic Line (SCL):
\[ r_{i,t} - r_f = \alpha_i + \beta^* (r_{m,t} - r_f) + \varepsilon_{i,t} \]  
\[ (3) \]

where \( r_{i,t} \) is the rate of return on asset \( i \) in time \( t \), \( r_{m,t} \) represents the rate of return on the market in time \( t \) and \( r_f \) is the risk-free rate. For ease of theoretical portrayal, it will be assumed that markets are efficient (\( \alpha_i = 0 \)) and the effective risk-free rate (\( r_f \)) is zero:

\[ r_{i,t} = \beta^* r_{m,t} + \varepsilon_{i,t} \]  
\[ (4) \]

The SCL then represents the relationship between the return of the market (\( r_{m,t} \)) and the return of a given asset \( i \) at time \( t \) with a sensitivity measure Beta (\( \beta \)). Beta is a measure that describes the relationship of an asset’s return in reference to the return of a financial market or index. The formula for the Beta of an asset is:

\[ \beta_a = \frac{\text{Cov}(r_{a,t} - r_{m,t})}{\text{Var}(r_{m,t})} \]  
\[ (5) \]

where \( r_{a,t} \) is a measure of the rate of return on asset \( a \) at time \( t \) and \( r_{m,t} \) is a measure of the rate of return on the market or index being used at time \( t \). Beta is derived from linear regression analysis in which the returns of an individual asset (\( r_a \)) are regressed against the returns of the market (\( r_m \)) in a specific time interval to find the covariance of the asset and the market. Then, the covariance is scaled by the variance of the returns on the market (\( r_m \)) to measure the sensitivity of the asset’s returns to the market’s returns. Theoretically, beta measures the statistical variance or systematic risk, of an asset that can not be mitigated through diversification.

### 2.3 Realized Beta

In order to calculate a way to measure the beta (\( \beta \)) of an asset over a given time interval, the Realized Beta is calculated. Realized beta is a proxy for the convention of realized volatility labeled by Andersen, Bollerslev, Diebold, and Labys (2001). Realized betas are realizations of the underlying ratio between the integrated stock and market return covariance and the integrated market variance\(^1\). If the instantaneous volatility \( \sigma(t) \) were known, then the true variance and covariance, called the integrated variance and integrated covariance, could be found by integrating the spot volatility over the time interval:

\[ \text{Integrated Variance} = \int_{t-1}^{t} \sigma^2(u) du \]  
\[ (6) \]

\[ \text{Integrated Covariance} = \int_{t-1}^{t} \sigma_{a,m}(u) du \]  
\[ (7) \]

---

\(^1\) Underlying theory is developed in Andersen, Bollerslev, Diebold, and Labys (2001, 2003).
Though, in practice, the underlying spot volatility is impossible to observe. Given that it is possible to observe realized prices, discrete measures of variation can be used to numerically approximate integrated variance and integrated covariance. Using an approach similar to that mentioned in Song (2009), sparse volatility estimators typically spaced evenly over some time interval can be defined over a single grid of price data. Considering a set of price data over the interval \([t-1, t]\), where \(t\) is measured in days, one can choose a sampling interval \(0 < \Delta t < 1\).

Then, selecting an initial sampling point, \(m\), \(0 < m < \Delta t\), an equally-spaced data interval is defined by \(M = \{m + k\Delta t \text{ where } k = 0, 1, 2, \ldots, \lfloor 1/\Delta t \rfloor\}\). Then, logarithmic returns can be calculated, \(r(t,k) = p(t, m + k\Delta t) - p(t, m + (k-1)\Delta t)\), with \(p(t, m + k\Delta t)\) being the observed price at time \(s + k\Delta t\).

Hence, using the formula given above for Beta, Realized Beta is calculated by:

\[
R\beta = \frac{\sum_{j=1}^{N_t} \text{Cov}(r_{i,t}, r_{m,t})}{\sum_{j=1}^{N_t} \text{Var}(r_{m,t})}
\]  

(8)

where \(\text{Cov}(r_{i,t}, r_{m,t})\) is the covariance of realized returns on an asset \(i\) and returns on the market on sampling interval \(t\), \(\text{Var}(r_{m,t})\) is the variance of realized returns on asset \(i\) on the sampling interval \(t\), and \(N_t\) is the number of units into which the sampling interval is partitioned into.

3 Predictability of a Time-Varying Beta

3.1 Standard Deviations of Realized Beta

Given that realized betas were constant over time, one would expect the time-series plot of beta over any designated time interval to be a horizontal line fixed at a horizontal index with a theoretical standard deviation of zero. But, Figure 1 displays a time-series plot of realized betas over the in sample interval used within this paper. Additionally, Table 1, summarizes the standard deviations of the equity betas in the aforementioned in sample period.

3.2 Autoregressive (1) Model

The autoregressive (AR) model is a random process which is often used in statistics to model and predict different types of phenomenon. The AR(1) process is generally defined by:

\[
X_t = c + \theta X_{t-1} + \epsilon_t
\]  

(9)

where \(c\) is a constant and \(\epsilon_t\) is a “white noise” process with zero mean and variance of \(\sigma^2\).
Adapting this notation to beta, the observed $\beta^*(t)$ at time is given by:

$$\beta^*(t) = c + \theta \beta(t-1) + \varepsilon_t$$  \hspace{1cm} (10)$$

where $\varepsilon_t$ is once again a “white noise” term, and represents a deviation from the fundamental beta $\beta(t)$. But, given that beta is in fact predictable, one would predict that the first order autocorrelation of this term should be positive. Chosen any random $\varepsilon_t$ in one period that causes the observed price to react, one would expect $\varepsilon_{t+1}$ to be relatively comparable at a similar index. Given that beta was not relatively predictable, one would expect to see a negative first order autocorrelation. The first order autocorrelation for the equities used within this paper are relatively positive and are summarized in Table 2. This suggests the predictability of beta and the impetus behind a time-varying beta analysis.

4 Theoretical Framework

4.1 HAR-Beta Regression Model

This paper emulates the Heterogeneous Autoregressive (HAR) model that was introduced by Corsi (2003) to forecast volatility. An alternative approach of latent volatility or daily squared returns could have been used as proxy, but Andersen and Bollerslev (1998) suggested that these measures of volatility can be noisy. Instead, this paper utilizes a formulated Heterogeneous Autoregressive Beta model that uses a linear combination of historical betas calculated over different time horizons, to capture the persistence of financial data. Numerous studies have suggested additional autoregressive moving averages, such as the Autoregressive Fractionally Integrated Moving Average (ARFIMA) or the Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity (FIGARCH), to supply better persistence in data. These models typically require nonlinear maximum likelihood estimation procedures other than the conventional Ordinary Least Squares. Andersen, Bollerslev and Huang (2007) have empirically tested models such as FIGARCH and ARFIMA, and deduced that simple linear models can oftentimes predict future volatility more accurately. Given this theoretical backing, the HAR model was chosen as an adaptation of time-varying beta analysis.

The HAR-Beta model uses latent integrated variances and integrated covariances to compute realized betas over different time horizons longer than one day. Then, given these realized betas, the model estimates normalized sums of designated time intervals to calculate a simple average of the designated quantity:
where $t$ represent time and $n$ represents the number of units into which time is partitioned into. In this paper, daily returns are predicted using:

$$R\beta_{t+1} = \beta_0 + \alpha_D*R\beta_{t+1} + \alpha_W*R\beta_{t+1 - 5} + \alpha_M*R\beta_{t+1 - 22} + \epsilon_{t+1}$$

(12)

where the dependent variables correspond to lagged daily ($t = 1$), weekly ($t = 5$) and monthly ($t = 22$) regressors. This convention of lagged regressors is taken directly from the formulization of the HAR-RV model in Corsi (2003).

**4.2 Beta Volatility**

Financial and economic literature (Andersen, Bollerslev and Diebold, 2004) have cited evidence of fluctuations and high persistence in asset price conditional covariances with the market and conditional variances. Thus market betas, which are ratios of time-varying conditional covariances and variances, are expected to display statistically persistent fluctuations (Andersen, Bollerslev, Diebold, and Wu, 2004). Tofalis (2008) provides a critique of the standard beta by citing the associated volatility inherent within beta. This can be evidenced by manipulating the standard textbook way of estimating beta using ordinary least squares (OLS) regression to find a resulting slope of:

$$\beta = r * \left( \frac{\sigma_i}{\sigma_m} \right)$$

(13)

where the $\sigma$’s are the standard deviations of the rates of return on asset $i$ and the market, and $r$ is the coefficient of correlation between the rates of return. This formula is equivalent to the the ratio of covariance between market and investment returns divided by variance of the market returns as noted by Tofalis. Hence, this decomposition portrays the combination of volatility and correlation in the calculation of beta. This also leads to an inequality (since $|r|$ is not greater than one):

$$\sigma \geq |\beta| \sigma_m$$

(14)

From this it is noticeable to see that in the presence of market microstructure noise, the realized beta is an inconsistent estimate of the true beta over the period.
4.3 Microstructure Noise

The underlying assumptions of classical economics state that financial markets are homogenous and that short-term prices movements follow a random walk. Given that the data used within this analysis is high-frequency, market microstructure noise, such as bid-ask spread, instantaneous information asymmetry, and other trading anomalies, may affect the short-term market price movements to reflect a value other than the true price of the asset. Hansen and Lunde (2006) define the market microstructure noise:

\[ u(t) = p(t) - p^*(t) \] (15)

where \( p(t) \) is the observable log price in the market at time \( t \) and \( p^*(t) \) is the latent real log price at time \( t \). Although, the latent real log price is not observable and hence cannot be used as an estimator. Using an approach similar to that defined in Amatyakul (2009), the information available on changes in price can be defined as:

\[ p(t + \theta) - p(t) = [p^*(t + \theta) - p^*(t)] + [u(t + \theta) - u(t)] \] (16)

where \( \theta \) is a real number increment , \([p(t + \theta) - p(t)]\) represents the change in price over a time interval and \( u(t) \) is i.i.d. and represents the microstructure noise applicable to the price change over the specified time interval. Now, as \( \theta \to 0 \), or the time interval is decreased, the magnitude of change in the latent real log price, \([p^*(t + \theta) - p^*(t)]\), should decrease. This could be due to the lack of new information or liquidity that forces a smaller price movement at decreased time intervals. Although, since \( u(t) \) is i.i.d., the microstructure noise term, \([u(t + \theta) - u(t)]\), does not decrease. Given this intuition used with high frequency data, at decreasingly small time intervals, as \( \theta \to 0 \), the change in latent real log price is minimal and the observed microstructure noise is relatively large.

Andersen, Bollerslev, Diebold and Labys (2000) recommend a graphical tool, called a volatility signature plot, as an approach to minimize market microstructure noise. This graphical approach displays how average realized variance corresponds to sampling frequency. Given that integrated variance is used to compute realized variance, this graphical tool presents a parallel simulation for the usage of integrated covariance and integrated variance used in the calculation of realizes betas. The theoretical idea as explained by Andersen et al. if variance is independent of the sampling frequency at which prices are observed, then variance should be the same if microstructure noise is not present. Although, since the practicality of real data is not always theoretically ideal, price data at different sampling frequencies will exhibit trends. Andersen et
al. also noted on theoretical grounds that liquid equities will have a downward sloping volatility signature plot since at high sampling frequencies, microstructure noise such as bid-ask bounce and instantaneous pricing asymmetries will be dissolved. The intuition to find the optimal sampling frequency is as follows. As noted in Equation 16, as the time interval, \( \theta \), decreases, realized variance will increase due to microstructure noise \( [u(t + \theta) - u(t)] \). Though, as \( \theta \) is increased by arbitrary amounts, given a liquid equity, the microstructure noises should diminish and subsequently lower the realized variance. Hence, at some point, the variance should stabilize, in which case the most number of data points can be used that are relatively robust to market microstructure noise. Figure 2 shows the volatility signature plots for all the equities used within this analysis. Utilizing the aforementioned technique and visually inspecting Figure 2, a sampling frequency of 10 minutes was chosen as optimal for subsequent analysis.

### 4.4 Beta Predictions

Realized betas of each equity were computed over the whole sampling interval, January 2, 2001 to January 3, 2009, at the 10 minute sampling frequency using the formula stated in Equation 11. Then the HAR-Beta model,

\[
R \beta_{t,t+1} = \beta_0 + \alpha_D \cdot R \beta_{t,t-1} + \alpha_W \cdot R \beta_{t,t-5} + \alpha_M \cdot R \beta_{t,t-22} + \epsilon_{t+1}
\]

or Equation 12, was trained over the whole sample time period of January 2, 2001 to January 3, 2009 to compute normalized average daily (t-1), weekly (t-5) and monthly (t-22) realized betas. Once the realized betas and the HAR-betas were computed over the whole sample, regression analysis was used. The realized betas from the in sample period of January 2, 2001 to January 2, 2006 were regressed on the HAR-Beta daily, weekly and monthly realized betas from the in sample period. The results of this regression analysis were daily, weekly and monthly beta coefficients from the HAR-Beta in-sample regression. The computed coefficients are summarized in Table 3.

Subsequently, these computed coefficients from the in sample time interval were used to compute the beta predictions for the corresponding out of sample time interval. To do this, corresponding trained beta coefficients and out of sample realized betas were multiplied and summed.
5 Data

The price data used within this paper are based on minute-by-minute price quotes from a commercial vendor, price-data.com that includes every minute from 9:35 AM to 4:00 PM on trading days from 1997 to 2009. The Standard & Poor’s Repository Index 500 (SPY) was used as the market index for calculation. Additionally, there were a total of eight equities chosen for this paper, including Coca Cola Company (KO), Pepsico, Inc. (PEP), Microsoft Corporation (MSFT), JPMorgan Chase & Co. (JPM), Bank of America Corporation (BAC), Johnson & Johnson (JNJ), Wal-mart Stores Inc. (WMT), and Exxon Mobil Corporation (XOM). These particular companies were chosen due to their liquidity, market capitalization and representation across industries including Consumer Goods, Technology, Financial, Healthcare, Services and Integrated Oil & Gas. Due to the incongruence across the price data available for the chosen equities, the full time interval used for analysis is January 2, 2001 to January 3, 2009, which includes approximately 1989 trading days.

5.1 In Sample

The in sample time interval chosen for the HAR-Beta model prediction coefficients was January 2, 2001 to January 2, 2006. The impetus for this particular in sample “training” interval was a balance between the congruence of the data available and the incorporation of a wide time interval containing numerous data points. This 5-year time period used to “train” the model is also useful in creating benchmark comparisons that will be explained later in the analysis.

5.2 Out of Sample

The first out of sample period used was a two-year period mapped from January 3, 2006 to January 3, 2008. This two-year sampling period was used as a benchmark length for a corresponding in sample period that would necessitate a significant amount of data points. Additionally, this specific time interval is utilized as a time-scale independent of the 2008 financial crisis.²

5.2B Out of Sample

² http://www.gop.gov/wtas/09/02/09/wsj-how-government-created-the
The second out of sample period used was a three-year period mapped from January 3, 2006 to January 3, 2009. The reasoning for a secondary out of sample data period is to verify the robustness of results with the wake of the 2008 financial crisis. Hence, the reasoning for two out of sample periods is not particularly a congruent comparison between the two samples, but rather a portrayal of the robustness of the beta predictions given a time period characterized by financial uncertainty.

6 Benchmarks of Comparison

6.1 Constant Mean Return

One of the benchmarks of comparison used within this analysis will be the constant mean logarithmic return. This comparison is independent of a beta coefficient that tracks the movement of an asset with the market. Instead, this comparison assumes that returns follow a model such that:

\[
R(t) = \frac{1}{n} \sum_{i=1}^{n} R_{t-1}
\]

where logarithmic returns at time \( t \), \( R(t) \), are observed as a sum of all latent logarithmic returns leading up to time \( t-1 \), \( R(t - 1) \), divided by the number of observations \( n \). Particularly, in this assumption, logarithmic returns at time \( t \) are thought to be estimated by the mean of all latent logarithmic realized returns noted at a corresponding sampling frequency.

6.2 Constant Beta

An alternate benchmark comparison to the time-varying beta used within this analysis is the conventional constant beta calculated for the Capital Asset Pricing Model. The usage of a beta computed from monthly returns over a 5-year time period has been noted by numerous studies including Banz (1981). Hence, given the in sample 5-year period of January 2, 2001 to January 2, 2006, a realized beta was computed for all equities on the basis of realized monthly returns, integrated variances, and integrated covariances over the corresponding time interval. A summary of the computed constant betas can be found in Table 4.

7 Statistical Methods
7.1 Root Mean Square Error

Accuracy measures on the predictability of beta involved the \textit{beta predictions} computed using the simplified Capital Asset Pricing Model:

\[ R_{a,t} = \alpha_a + \beta_a(R_{m,t}) + \varepsilon_{a,t} \] (18)

where \( R_{a,t} \) is the predicted return on Asset a at time t, \( R_{m,t} \) is the temporally corresponding market return, \( \alpha_a \) is a parameter whose value is such that \( \mathbb{E}[\varepsilon_{a,t}] = 0 \), \( \beta_a \) is defined as the corresponding \textit{beta prediction}, and \( \varepsilon_{a,t} \) is a random error term.

The test of beta prediction accuracy that follows will make use of the mean square error as a measure of forecast error. Mean square forecast error (RMSE) is defined as:

\[ MSE = \frac{1}{n} \sum_{t=1}^{n} (R_{a,t} - \hat{R}_{a,t})^2 \] (19)

where \( n \) is the number of predictions contained, \( R_{a,t} \) is out of sample realized logarithmic return on Asset a at time t and \( \hat{R}_{a,t} \) is the predicted return on Asset a at the corresponding time t. In order to standardize the results, the MSE is multiplied by the number of trading days in a year, 252. Additionally, the square root of the MSE*252 is taken:

\[ = \sqrt{MSE \times 252} \] (20)

The result is the root mean squared error (RMSE) which can be interpreted in annualized standard deviation units.

8 Results

8.1 In Sample

\textit{Table 1} summarizes the standard deviations of the in sample betas across all equities using the 10-minute sampling frequency. As described in section 3.1, if beta were constant over time, then theoretically the difference between predicted and observed beta would be zero given no microstructure noise. In account for the microstructure noise, at the 10 minute sampling frequency, the average standard deviation of beta is still 0.3737 which is statistically significant under the null hypothesis of zero deviations. Additionally, \textit{Table 2} summarizes the results of the first order autocorrelations from the in-sample betas. As described in section 3.2, positive first order autocorrelation suggest the predictability of a time-varying element. Given a positive first order autocorrelation, any random \( \varepsilon_t \) in one period that causes the observed price to react would
be relatively comparable to \( \varepsilon_{t+k} \), where \( K \) approaches the number of observations in the sample. The average first order autocorrelation is approximately 0.3484 which suggests the predictability of a time-varying beta. It is notable that Microsoft Corporation (MSFT) and Exxon Mobil Corporation (XOM) display a relatively low first order autocorrelation of 0.0866 and 0.1521.

Literature suggests that it is reasonable to expect that there is some instability in most econometric relationships across time or space. Typically in cross sections with market data, there is likely some degree of heterogeneity among assets (Elliott and Müller, 2006). There are numerous factors that contribute to the heterogeneity of a time series including regulation, economic policy. Giacomini and White (2006) contests that as long as this heterogeneity is not “too strong”, standard regression methods still have reasonable properties with the replacement of “true” values of the coefficients with averages of the individual or inter-temporal true values of coefficients. *Table 3* illustrates the HAR-Beta coefficients of the in sample HAR-Beta estimates regressed on the in sample realized betas. The sum of the beta coefficients across all equities approximate one which alludes to the persistency of the regression coefficients. One exception is Microsoft Corporation (MSFT) which has a sum of coefficients equal to -0.2387. Notably, Exxon Mobil Corporation (XOM) has a sum of regression coefficients equal to 0.7623 which is relatively below one. These results seem to make intuitive sense given the results from *Table 2* which suggested that MSFT and XOM betas were relatively less predictable over the in sample period.

### 8.2 Out of Sample (A)

The out of sample (A) time interval refers to dates January 3, 2006 to January 3, 2008. *Table 5* summarizes the root mean squared errors (RMSE) calculated from the differentials between realized returns and predicted returns based on the HAR-Beta model, constant beta benchmark comparison and the constant return benchmark comparison. Additionally, *Table 3* contains the \( R^2 \) values from the regression model. The direct comparison of RMSE between the hypothesized HAR-Beta model and the benchmark comparisons display a significant reduction. When compared to the constant returns benchmark, the HAR-Beta model gives an approximate 21.94% reduction in RMSE. Additionally, when compared to the constant beta benchmark comparison, the HAR-Beta model gives an approximate reduction of 6.62%.
Notably, Wal-mart Stores (WMT), has a 35.63% increase in RMSE when the HAR-Beta model is compared to constant returns. This seems surprising when referring back to the results from Table 1, Table 2 and Table 3 in which WMT has the highest Standard Deviation of Beta in sample (0.4783), the highest first order autocorrelation (0.5701) and the highest R² value (0.5391). This result could be the outcome of outliers which complicate the extrapolation of Ordinary Least Squares regression models. Poon and Granger (2003) note the common problems of possible sample outliers in volatility estimation and suggest alternate methods of robust regressions.

Returning to the results from Table 2 and Table 3, which suggested that MSFT and XOM were relatively less predictable, the subsequent differences in RMSE seem to concur. When constant returns of MSFT and XOM are compared to the HAR-Beta model, there is a relatively low 0.07% and 5.98% reduction respectively. Although when compared to the constant beta benchmark, MSFT and XOM display the largest RMSE reductions of 13.45% and 7.96% respectively. The R² values for MSFT and XOM are also the lowest, 0.0113 and 0.1106 respectively, concur with beta being less predictable. This point is interestingly quite salient. Given the low predictability of beta measured by the results from Table 2 and Table 3, which is concurred with the low R² values in Table 3, the RMSE of the HAR-Beta model are still reduced in terms of both benchmark comparisons. Although, in both cases of MSFT and XOM, when constant beta is compared to constant returns, there is a significant increase in RMSE. This point confirms the results displayed in Table 1 that beta deviates given a time interval despite measures of predictability.

8.3 Out of Sample (B)

The out of sample (B) time interval refers to dates January 3, 2006 to January 3, 2009. Table 6 summarizes the root mean squared errors (RMSE) calculated from the differentials between realized returns and predicted returns based on the HAR-Beta model, constant beta benchmark comparison and the constant return benchmark comparison. The impetus for this out of sample, as defined earlier, is to check the robustness of the HAR-Beta given the wake of the 2008 financial crisis. It is important to note, the increase in RMSE present when comparing constant returns to the HAR-Beta model of Microsoft Corporation (MSFT). This point once again can be surmised to the presence of an outlier that leads to extrapolation errors. Aside from
this, given this out of sample period, RMSE are reduced by approximately 19.67% when compared to the constant beta benchmark, and approximately 39.28% (including the MSFT increase) when compared to the constant return benchmark.

9 Conclusion

In line with recent literature that suggests alternatives to the constant beta conventionally used within the Capital Asset Pricing model, this paper demonstrates that betas do in fact vary over time. At a determined optimal sampling frequency of 10 minutes, the average standard deviation of betas over the whole sample was approximately .3663. Positive autocorrelation in the in sample beta time series suggest the predictability of these betas. In lieu of beta predictability, the results from benchmark comparisons to the HAR-Beta model showed an overall reduction in RMSE across all equities, within out of sample A and B.

It is important to note the inherent weakness of Ordinary Least Squares regression analysis when dealing with outliers. Given one or two outliers in a large sample, the R$^2$ can inaccurately represent the actual fit of the regression and give imprecise approximates of the coefficient. This misspecification can lead to subsequent analysis which could inaccurately represent the true data.

But, despite this limitation, the majority of the results from this analysis are still fairly conclusive. The usage of logarithmic returns coupled with the optimal sampling frequency gives way to measures of sanity that do not deem these results inconclusive. Overall, the results within this paper suggest the importance of a conventional change from the constant beta used within the CAPM model.
10 Reference


11 Figures

Figure 1: Time Series- Realized Betas (January 2, 2001 – January 3, 2009)
Figure 2: Volatility Signature Plots (January 2, 2001 - January 3, 2009)

KO

PEP

MSFT

JPM

BAC

JNJ

WMT

XOM
## Tables

All tables display statistics that use 10-minute sampling intervals.

### Table 1: Standard Deviation of Beta (In Sample)

<table>
<thead>
<tr>
<th>Company</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coca Cola Company (KO)</td>
<td>0.3925</td>
</tr>
<tr>
<td>Pepsico, Inc. (PEP)</td>
<td>0.2745</td>
</tr>
<tr>
<td>Microsoft Corporation (MSFT)</td>
<td>0.3417</td>
</tr>
<tr>
<td>JPMorgan Chase &amp; Co. (JPM)</td>
<td>0.4485</td>
</tr>
<tr>
<td>Bank of America Corporation (BAC)</td>
<td>0.3721</td>
</tr>
<tr>
<td>Johnson &amp; Johnson (JNJ)</td>
<td>0.3822</td>
</tr>
<tr>
<td>Wal-mart Stores Inc. (WMT)</td>
<td>0.4783</td>
</tr>
<tr>
<td>Exxon Mobil Corporation (XOM)</td>
<td>0.2997</td>
</tr>
</tbody>
</table>

### Table 2: AR(1) - First Order Autocorrelation of Beta

<table>
<thead>
<tr>
<th>Equity</th>
<th>First Order Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coca Cola Company (KO)</td>
<td>0.5345</td>
</tr>
<tr>
<td>Pepsico, Inc. (PEP)</td>
<td>0.2910</td>
</tr>
<tr>
<td>Microsoft Corporation (MSFT)</td>
<td>0.0866</td>
</tr>
<tr>
<td>JPMorgan Chase &amp; Co. (JPM)</td>
<td>0.3296</td>
</tr>
<tr>
<td>Bank of America Corporation (BAC)</td>
<td>0.3550</td>
</tr>
<tr>
<td>Johnson &amp; Johnson (JNJ)</td>
<td>0.4673</td>
</tr>
<tr>
<td>Wal-mart Stores Inc. (WMT)</td>
<td>0.5701</td>
</tr>
<tr>
<td>Exxon Mobil Corporation (XOM)</td>
<td>0.1531</td>
</tr>
</tbody>
</table>
Table 3: HAR-Beta Regression Coefficients

<table>
<thead>
<tr>
<th>Company</th>
<th>$\beta_0$</th>
<th>$R_{\beta_{t-1}}$</th>
<th>$R_{\beta_{t-5}}$</th>
<th>$R_{\beta_{t-22}}$</th>
<th>Sum of Coefficients</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>KO</td>
<td>0.0096**</td>
<td>0.0289*</td>
<td>0.2695*</td>
<td>0.6367**</td>
<td>0.9447</td>
<td>0.4784</td>
</tr>
<tr>
<td>PEP</td>
<td>0.0051**</td>
<td>0.0201*</td>
<td>0.0880*</td>
<td>0.6754**</td>
<td>0.7886</td>
<td>0.0019</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.0107**</td>
<td>0.1023**</td>
<td>-0.0226</td>
<td>-0.3291*</td>
<td>-0.2387</td>
<td>0.0113</td>
</tr>
<tr>
<td>JPM</td>
<td>0.0078*</td>
<td>0.0210*</td>
<td>0.4598**</td>
<td>0.3476**</td>
<td>0.8362</td>
<td>0.2547</td>
</tr>
<tr>
<td>BAC</td>
<td>0.0011*</td>
<td>0.0416*</td>
<td>0.0864*</td>
<td>0.7470**</td>
<td>0.8761</td>
<td>0.2797</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.0066*</td>
<td>-0.0375*</td>
<td>0.3764**</td>
<td>0.5924**</td>
<td>0.9379</td>
<td>0.4361</td>
</tr>
<tr>
<td>WMT</td>
<td>0.0094*</td>
<td>-0.0151</td>
<td>0.4014**</td>
<td>0.5613**</td>
<td>0.9570</td>
<td>0.5391</td>
</tr>
<tr>
<td>XOM</td>
<td>0.0034*</td>
<td>0.0094</td>
<td>-0.0934*</td>
<td>0.8429**</td>
<td>0.7623</td>
<td>0.1106</td>
</tr>
</tbody>
</table>

The significance levels of the coefficients are denoted by the asterisk: * → $p < 0.05$, ** → $p < 0.01$

Table 4: Constant betas computed using monthly data over 5-year period (In-Sample)

<table>
<thead>
<tr>
<th>Company</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>BKO</td>
<td>0.4989</td>
</tr>
<tr>
<td>BPEP</td>
<td>0.4074</td>
</tr>
<tr>
<td>BMSFT</td>
<td>1.1039</td>
</tr>
<tr>
<td>BPM</td>
<td>1.5781</td>
</tr>
<tr>
<td>BBAC</td>
<td>0.4830</td>
</tr>
<tr>
<td>BJNJ</td>
<td>0.2958</td>
</tr>
<tr>
<td>BWMT</td>
<td>0.6255</td>
</tr>
<tr>
<td>BXOM</td>
<td>0.5899</td>
</tr>
</tbody>
</table>
Table 5: Root Mean Squared Error (RMSE)-HAR-Beta, Constant Beta, Constant Return, and Standard Deviation of Beta (*Out of Sample-A*)

<table>
<thead>
<tr>
<th></th>
<th>RMSE HAR-Beta</th>
<th>RMSE Constant Beta</th>
<th>Constant Returns</th>
<th>Standard Deviation of Beta (Out Sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KO</td>
<td>0.1732</td>
<td>0.1867</td>
<td>0.4935</td>
<td>0.2628</td>
</tr>
<tr>
<td>PEP</td>
<td>0.1963</td>
<td>0.2095</td>
<td>0.2511</td>
<td>0.2703</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.2831</td>
<td>0.3271</td>
<td>0.2833</td>
<td>0.3800</td>
</tr>
<tr>
<td>JPM</td>
<td>0.4347</td>
<td>0.4763</td>
<td>1.1433</td>
<td>0.4410</td>
</tr>
<tr>
<td>BAC</td>
<td>0.3340</td>
<td>0.3410</td>
<td>0.5289</td>
<td>0.3479</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.1650</td>
<td>0.1674</td>
<td>0.2716</td>
<td>0.2425</td>
</tr>
<tr>
<td>WMT</td>
<td>0.3441</td>
<td>0.3652</td>
<td>0.2215</td>
<td>0.3614</td>
</tr>
<tr>
<td>XOM</td>
<td>0.2892</td>
<td>0.3142</td>
<td>0.3076</td>
<td>0.3669</td>
</tr>
</tbody>
</table>

*All units are expressed in Annualized Standard Deviation Units.*

Table 6: Root Mean Squared Error (RMSE)-HAR-Beta, Constant Beta, Constant Return, and Standard Deviation of Beta (*Out of Sample-B*)

<table>
<thead>
<tr>
<th></th>
<th>RMSE HAR-Beta</th>
<th>RMSE Constant Beta</th>
<th>Constant Returns</th>
<th>Standard Deviation of Beta (Out Sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KO</td>
<td>0.1953</td>
<td>0.2496</td>
<td>0.5994</td>
<td>0.2442</td>
</tr>
<tr>
<td>PEP</td>
<td>0.1986</td>
<td>0.2195</td>
<td>0.3570</td>
<td>0.2575</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.2291</td>
<td>0.3796</td>
<td>0.1774</td>
<td>0.3613</td>
</tr>
<tr>
<td>JPM</td>
<td>0.4262</td>
<td>0.6329</td>
<td>1.2493</td>
<td>0.4884</td>
</tr>
<tr>
<td>BAC</td>
<td>0.4706</td>
<td>0.4924</td>
<td>0.6348</td>
<td>0.4208</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.2149</td>
<td>0.2230</td>
<td>0.5767</td>
<td>0.2207</td>
</tr>
<tr>
<td>WMT</td>
<td>0.2293</td>
<td>0.3106</td>
<td>0.3274</td>
<td>0.3339</td>
</tr>
<tr>
<td>XOM</td>
<td>0.2188</td>
<td>0.2719</td>
<td>0.4135</td>
<td>0.3476</td>
</tr>
</tbody>
</table>

*All units are expressed in Annualized Standard Deviation Units.*