Jun Pan
junpan@mit.edu
253-3083
E62-624

MIT Sloan School of Management
15.433-15.4331 Financial Markets, Fall 2016

Classes 9 & 10: Equity in the Time Series, Part 2

Time-Varying Volatility

This Version: September 30, 2016

Just when Prof. Fama and his PhD/MBA students were busy working on the cross section of expected stock returns, another area of Finance was taking shape. In this area, tools developed in Econometrics and Statistics are applied to financial time series such as the time-series of stock returns. Given how difficult it is to estimate the first moment (expected returns), much of the attention was devoted to estimating the second moment, stock return volatility. The most visible figure in this area is Prof. Rob Engle, who was awarded a Nobel Prize in 2003 for "methods of analyzing economic time series with time-varying volatility (ARCH)." The ARCH paper was published in 1982 when Prof. Engle was an Economics professor at UCSD. The more famous GARCH extension was later published in 1986 by his PhD student Prof. Bollerslev.

1 Volatility models and market risk measurement

- **The need for better risk management tools:** In the early 1990s, there were two developments that made volatility models attractive and relevant. First, the need of a better option pricing model becomes quite obvious after the 1987 stock market crash. The Black-Scholes model builds a very good foundation, but it lacks flexibility in handling the richer reality. In the Black-Scholes model, stock returns are normally distributed with a constant volatility σ. Any casual inspection of the data would inform us that volatility is not a constant. So having a better volatility model would be a first step toward a better option pricing model.

  Second, and even more pressing was the need for better risk management tools. The mortgage-backed security was developed in the 1980s, and the over-the-counter derivatives market started to take off by the late 1980s. By the early 1990s, the increasing activity in securitization and the increasing complexity in the fixed-income products have made the trading books of many investment banks too complex and diverse for the chief executives to understand the overall risk of their firms.

It was an industry wide phenomenon. For example, in "Money and Power," Cohan wrote about the difficult year of 1994 at Goldman after the two amazingly profitable years in fixed-income in 1992 and 1993. In the book, he quoted Henry Paulson, "What came out of the 1994 debacle was best practices in terms of risk management, The quality of the people, and the processes that were put in place – anything from the liquidity management to the way we evaluated risk and really the independence of that function – changed the direction of the firm."

- **JPMorgan's RiskMetrics:** Among all Wall Street firms, JPMorgan's effort was by far the most visible and influential. This 2009 New York Times article titled "Risk Mismanagement" gave a very good account of the events.

> JPMorgan's chairman at the time VaR took off was a man named Dennis Weatherstone. Weatherstone, who died in 2008 at the age of 77, was a working-class Englishman who acquired the bearing of a patrician during his long career at the bank. He was soft-spoken, polite, self-effacing. At the point at which he took over JPMorgan, it had moved from being purely a commercial bank into one of these new hybrids. Within the bank, Weatherstone had long been known as an expert on risk, especially when he was running the foreign-exchange trading desk. But as chairman, he quickly realized that he understood far less about the firms overall risk than he needed to. Did the risk in JPMorgans stock portfolio cancel out the risk being taken by its bond portfolio – or did it heighten those risks? How could you compare different kinds of derivative risks? What happened to the portfolio when volatility increased or interest rates rose? How did currency fluctuations affect the fixed-income instruments? Weatherstone had no idea what the answers were. He needed a way to compare the risks of those various assets and to understand what his companywide risk was.

What later became RiskMetrics was an internal effort developed within JPMorgan in 1992 in response to the CEO's question. Quoting the New York Times article again,

> By the early 1990s, VaR had become such a fixture at JPMorgan that Weatherstone instituted what became known as the 415 report because it was handed out every day at 4:15, just after the market closed. It allowed him to see what every desk's estimated profit and loss was, as compared to its risk,

and how it all added up for the entire firm. True, it didn't take into account Taleb's fat tails, but nobody really expected it to do that. Weatherstone had been a trader himself; he understood both the limits and the value of VaR. It told him things he hadn't known before. He could use it to help him make judgments about whether the firm should take on additional risk or pull back. And that's what he did.

- **Global risk factors:** In 1994, JPMorgan started to make RiskMetrics publicly available. It published its technical document outlining its risk measurement methodologies. It also made available two sets of volatility and correlation data used in the computation of market risk. In Spring 1996, I was hired as a research assistant to Prof. Darrell Duffie to work on risk management and Value-at-Risk. I spent a lot of time reading this technical document of RiskMetrics and I also downloaded the volatility and correlation dataset everyday just to play with them and track the movements. It was through having to deal with the datasets that the immensity of the global financial markets became real to me.

  For example, the 1996 RiskMetrics data files covered over 480 financial time series that were important for the trading books of most investment banks. This includes

  - Equity indices across the world.
  - Foreign exchange rates.
  - The term structure of interest rates across the world:
    - money market rates (1m, 3m, 6m, and 12m) for the short end.
    - government bond zero rates (2y, 3y, 4y, 5y, 7y, 9y, 10y, 15y, 20y, and 30y) for the longer end.
  - The term structure of swap rates across the world (2y, 3y, 4y, 5y, 7y, and 10y).
  - Commodities: spot and futures of varying maturities.

- **The variance-covariance matrix:** In order to measure the firm-wide risk exposure, one need to first map each position to these risk factors, and then calculate the volatility of the overall portfolio or portfolios by asset class: interest rates, equity, currency, and commodities. One of the key building block of this calculation is the variance-covariance matrix of the risk factors. For the 480 risk factors used by JPMorgan in 1996, this involves calculating the volatility for each of the 480 risk factors, and then calculate the pair-wise correlations between the 480 risk factors. In the rest of the class, we will be busying doing these calculations.

Let me quote a few paragraphs from the 2012 annual report of Goldman Sachs so as to give you an update on the risk management effort on Wall Street since 1996.

> We also rely on technology to manage risk effectively. While judgment remains paramount, the speed, comprehensiveness and accuracy of information can materially enhance or hinder effective risk decision making. We mark to market approximately 6 million positions every day. And, we rely on our systems to run stress scenarios across multiple products and regions. In a single day, our systems use roughly 1 million computing hours for risk management calculations.
>
> When calculating VaR, we use historical simulations with full valuation of approximately 70,000 market factors. VaR is calculated at a position level based on simultaneously shocking the relevant market risk factors for that position. We sample from 5 years of historical data to generate the scenarios for our VaR calculation. The historical data is weighted so that the relative importance of the data reduces over time. This gives greater importance to more recent observations and reflects current asset volatilities, which improves the accuracy of our estimates of potential loss. As a result, even if our inventory positions were unchanged, our VaR would increase with increasing market volatility and vice versa.

As you can see, roughly 6 million positions are mapped into 70,000 market factors in Goldman's risk management system. If I understand their statement correctly, this implies a variance-covariance matrix of 70,000 by 70,000.

Back in the mid-1990s, all three of my Chinese classmates at the NYU Physics department went to work at Citibank after graduation. I was the only exception, who went on to get another PhD in Finance. So much for the future of Physics, which was better off without us. During one of my visits back to New York, I visited them at Citibank, thinking how exciting it was for them to be working in the *real* world with a real paycheck. And I was very surprised to see how bored they all looked. One of them worked in the risk management group and his job was to calculate the variance-covariance matrix everyday. He looked miserable. I guess this is not the most exciting job if you have to do it everyday. Many years later, I got an email from Mr. Variance-Covariance, who has become a managing director at Citibank. A happy ending, by Wall Street standard.

2 Estimating Volatility using Financial Time Series

In general, volatility is very easy to estimate. Unlike in the case of expected returns, volatility can be measured with better precision using higher frequency data. The convention in this field is to use daily data. For stock market returns, having one month of daily data could get you a pretty accurate estimate. We will continue to use the time series of aggregate stock returns as an example.

I should mention that in this field, log returns are being used more often than percentage returns:

$$R_t = \ln S_t - \ln S_{t-1}\,,$$

where S_t could be date-t stock price, currency rate, interest rate, or commodity futures price. At the daily frequency, the magnitude of returns are generally very small. Using the handy Taylor expansion, we know that for small x,

$$\ln(1+x) \approx x\,.$$

Repeating this for log-returns, we have

$$R_t = \ln S_t - \ln S_{t-1} = \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln\left(1 + \frac{S_t - S_{t-1}}{S_{t-1}}\right) \approx \frac{S_t - S_{t-1}}{S_{t-1}}\,.$$

In other words, working with log-returns or percentage-returns does not make too much of a difference when returns are small in magnitude.

Also notice that our attention is no longer on the first moment. Getting the volatility right is our main task. We calculate the variance by,

$$\mathrm{var}(R_t) = E\left(R_t - \mu\right)^2 = E(R_t^2) - \mu^2\,.$$

The volatility estimate is

$$\mathrm{std}(R_t) = \sqrt{\mathrm{var}(R_t)} = \sqrt{E(R_t^2) - \mu^2} = \sqrt{E(R_t^2)} \times \sqrt{1 - \frac{\mu^2}{E(R_t^2)}}\,.$$

At the daily frequency, μ is around a few basis points for the US equity market, while the daily volatility is around 100 basis points (i.e., 1%). As a result, $\mu^2/E(R_t^2)$ is a really really small number and $\sqrt{1-\mu^2/E(R_t^2)} \approx 1 - \frac{1}{2}\mu^2/E(R_t^2)$ is very close to one. So it does not make a big difference whether or not we subtract μ from the realized returns such as R_t in estimating the volatility. You will notice that most of the time, people drop μ for simplicity:

$$\text{std}(R_t) \approx \sqrt{E(R_t^2)}\,.$$

Of course, this relationship between daily vol and daily μ exists mostly true for assets in the risky category. For fixed income product, this might not be true. If you would like to be safe, one way to approach the data is to first demean the time-series data: $R_t - \mu$, and then apply the volatility models.

- **SMA:** The simple moving average model fixes a window, say one month, and use the daily returns within this window to calculate the sample standard deviation. The window is then moved forward by one step, say one day, and the whole calculation gets repeated again.

  In Figure 1, I use a moving window of one month and move the window one month at a time to plot a monthly time-series of SMA volatility estimates. In this space, volatility is usually quoted at an annualized level. So I multiply the volatility estimates by $\sqrt{252}$ (assuming 252 business days per year). This annualized volatility corresponds to the volatility coefficient σ in the Black-Scholes model. So we also compare these volatility numbers with the option-implied volatility.

  Just so you are convinced that these volatility estimates can be estimated with a pretty good precision using only one month of daily data, I also plotted the 95% confidence intervals. If you compared Figure 1 against Figure 2, you can see the marked difference in estimation precision. Using one month of daily data to estimate the average return, what you get is very much noise.

  Another observation I would like you to make is the variation of market volatility over time. Its pattern is very different from that of market returns. Volatility is persistent: a day of high volatility is usually followed by another day of high volatility. Volatility also tends to spike up once in a while. If you plot these events against the NBER business cycles, you see that volatility usually spikes up during recessions. But recessions are not the only time when volatility spikes up. Whenever the market is in trouble, volatility goes up. Let me name a few recent events: the 1987 stock market crash, the 1997 Asian Crisis, the 1998 LTCM crisis, the 2000-01 tech bubble/burst, the 9/11, and the 2008 financial crisis. Finally, whenever volatility is at an usually high or low level, it tends to revert back to its historical average. This pattern is called mean reversion. Using a longer sample that includes the Great Depression, the historical average of volatility is around 20%. Using the more recent sample, the average volatility is around 15%.

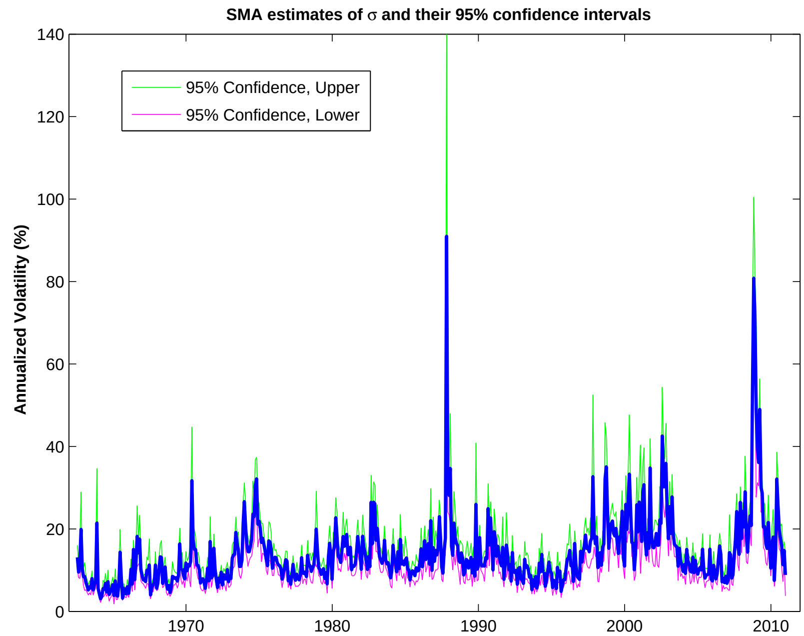


Figure 1: Time-Series of Stock Volatility using SMA.

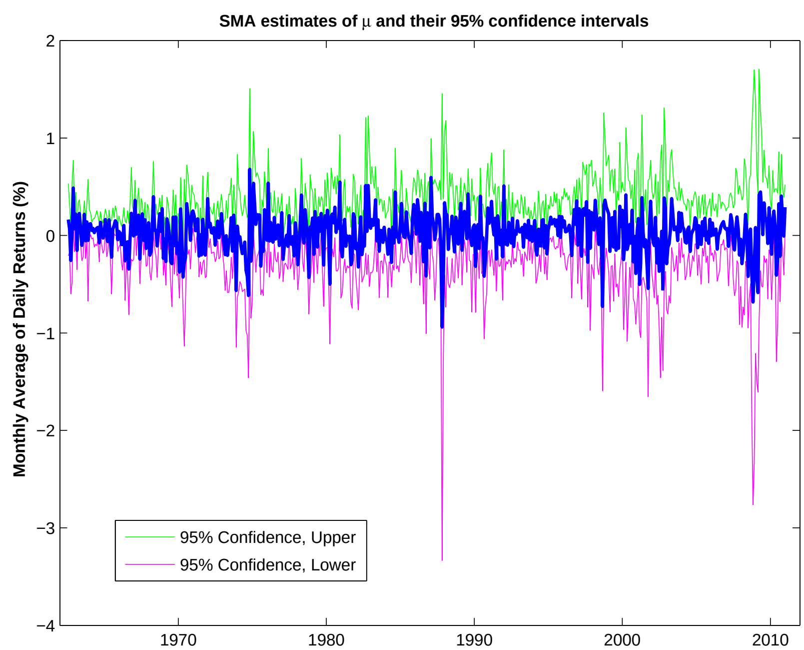


Figure 2: Time-Series of Monthly Average Stock Returns using SMA.

The plot has not been updated for the last few years. Nowadays, the best place to get stock market volatility (without having to do any calculation) is the CBOE VIX index. For example, over the one-week period from August 17 to 24, 2015, the VIX shoots up from 13.02% to 40.74%, because of the concern over the Chinese stock market.

Comparing the volatility estimate with the VIX index, which is effectively the option implied volatility, you notice that the option implied volatility seems to be consistently higher than the volatility estimate. We will visit this issue again when we cover the options market.

- **EWMA:** The exponentially weighted moving average model is an improvement over the SMA model. Instead of applying equal weights to all observations with a fixed window, EWMA applies an exponential weighting schedule. It chooses a decay factor λ, which is a number between zero and one, and performs the volatility estimate by:

$$\sqrt{(1-\lambda)\sum_{n=0}^{N}\lambda^n\left(R_{t-n}\right)^2}\,.$$

Let's first put aside the term $1-\lambda$ and focus on the terms within the summation. We put a weight of 1 for today ($n=0$), λ for yesterday, λ^2 for the day before yesterday, and so on. With this weighting schedule, as we move further back into the history and away from today t, the contribution of $(R_{t-n})^2$ decreases according to the exponential schedule. Hence the name. Figure 3 gives a graphical presentation of this exponential weighting scheme.

Going back to one of the paragraphs I quoted from Goldman's annual report: "We sample from 5 years of historical data to generate the scenarios for our VaR calculation. The historical data is weighted so that the relative importance of the data reduces over time. This gives greater importance to more recent observations and reflects current asset volatilities, which improves the accuracy of our estimates of potential loss." So effectively, the length of the window N is set at five years and a decay factor is selected to put more weight to the more recent events. In Goldman's report, the value of the decay factor was not reported. In RiskMetrics, λ was fixed at 0.94 for all time series. Also, the choice of the window size is not important because the decay factor λ effectively selects the window size for you. Figure 3 gives a nice graphical presentation on how the window size is determined by the decay factor: a strong decay factor ($\lambda=0.8$) implies a smaller window while a mild decay factor ($\lambda=0.97$) implies a larger window. A window of 5 years is definitely not necessary: the return happened

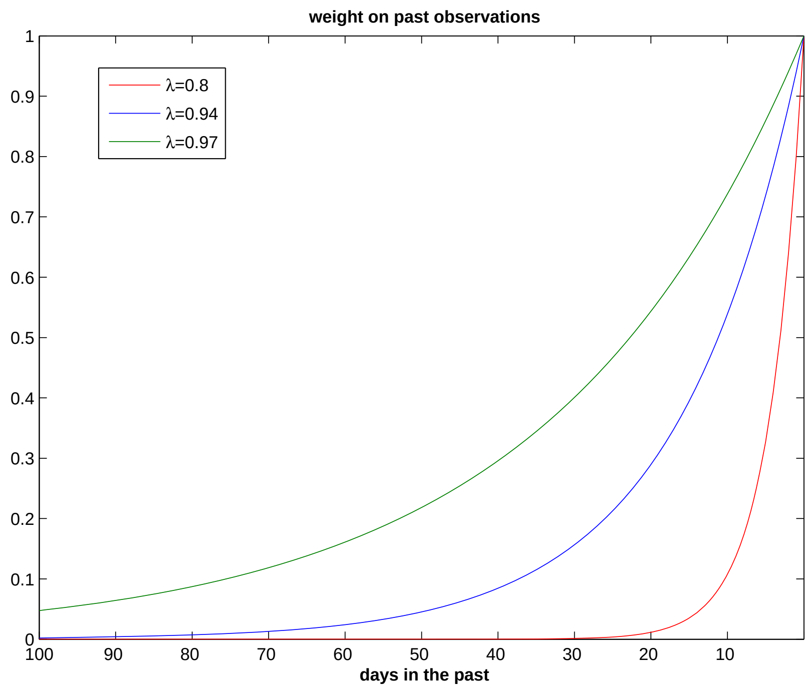


Figure 3: The Exponential Weighting Scheme in EWMA.

5 years ago has a weight of $\lambda^{5\times 252}$, which is too small to matter.

Now let's come back to the term $1-\lambda$. Notice that

$$1+\lambda+\lambda^2+\lambda^3+\ldots = 1/(1-\lambda)\,.$$

So $(1-\lambda)$ is there because of normalization. In the same way, the normalization factor in the SMA model is 1/N. In the SMA model, if I increase the window size N, then each observation carries a smaller weight: a smaller 1/N. Likewise, if I change λ from 0.94 to 0.97 in EWMA, the effective window size increases (see Figure 3). As a result, each observation carries a smaller weight: a smaller $1-\lambda$.

- **SMA and EWMA:** The difference between these two volatility estimates becomes most visible immediately after a large price movement. Figure 4 uses the famous Black Wednesday of 1992 as an example to illustrate this point. This technical document by RiskMetrics was first written around 1994. If it were written today, then the 2008 crisis would be plotted here as an example.

  After a large price movement, up or down, the response of the EWMA estimate is very fast, because it carries a higher weight for the most recent event. If the market calms

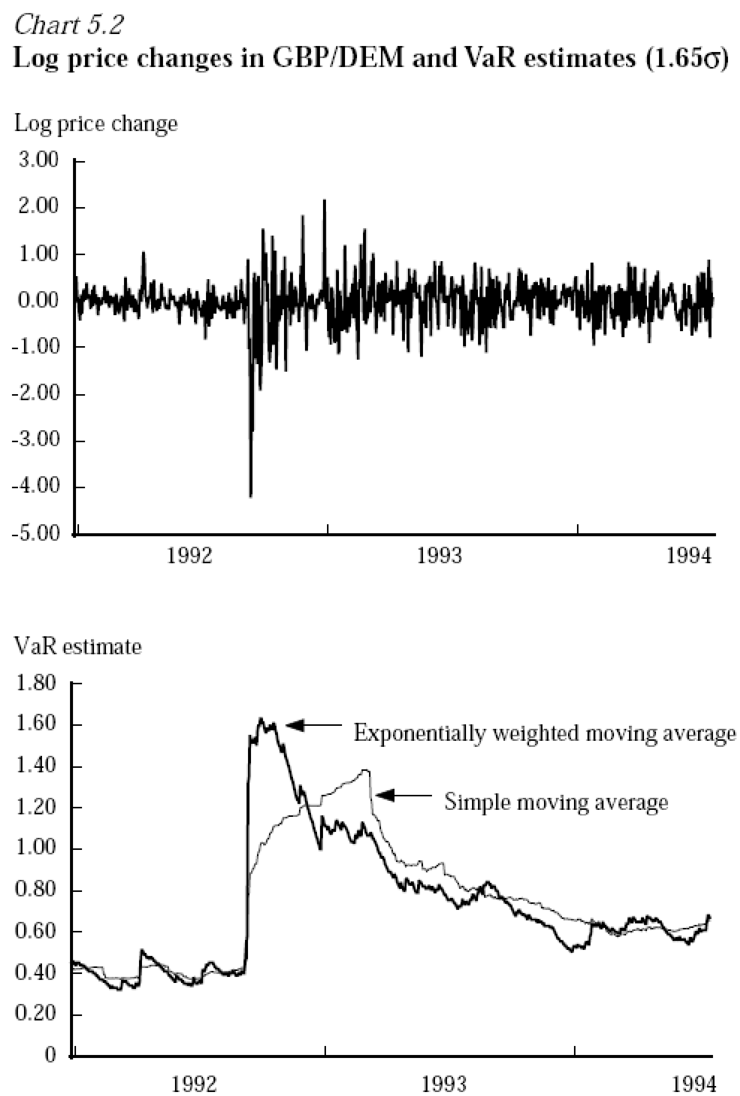


Figure 4: The Black Wednesday 1992 and the Volatility Estimates of SMA and EWMA.

down after the large price movement, then the EWMA estimate will soon come down to a lower level. The behavior of the SMA estimate is the opposite. Its response is typically sluggish and it carries that piece of information for the duration of its window size. For this reason, the EWMA is the preferred volatility estimate when it comes to monitoring market volatility at the daily frequency.

- **Black Wednesday 1992:** As a side, the 1992 sterling crisis was an important event in the global currency market. The followings are excepts from Steven Drobny's book on "Inside the House of Money: Top Hedge Fund Traders on Profiting in the Global Markets."

  The United Kingdom joined the European Exchange Rate Mechanism (ERM) in 1990 at a central parity rate of 2.95 deutsche marks to the pound. To comply with the ERM rule, the UK government was required to keep the pound in a trading band within 6 percent of the parity rate. In September 1992, as the sterling/mark exchange rate approached the lower end of the trading band, traders increasingly sold pounds against deutsche marks, forcing the Bank of England to intervene and buy an unlimited amount of pounds in accordance with ERM rules. Finally, on the evening of September 16, 1992, Great Britain humbly announced that it would no longer defend the trading band and withdrew the pound from the ERM system. The pound fell approximately 15 percent against the deutsche mark over the next few weeks, providing a windfall for speculators and a loss to the UK Treasury (i.e., British taxpayers) estimated to be in excess of £3 billion.

  It was reported at the time that Soros Fund Management made between $1-2 billion by shorting the pound, earning George Soros the moniker the man who broke the Bank of England. But he was certainly not alone in betting against the pound. In fact, the term global macro first entered the general public's vocabulary on Black Wednesday.

  Going back to our class on Predictability and Market Efficiency, there are few things to be learned. First, you predict the market by following the information, which, in this case, includes the ERM rule, the economic condition at UK, the government's ability and political resolve to defend its currency. Second, the "arbitrage" is risky. In order for George Soros to make $1 billion with a 15% drop in sterling, his short position at the time had to be over $6 billion. This is the style of global macro: large and risky directional bets. Of course, they don't always make money and we've seen a few times when Soros lost by the same order of magnitude. Third, most of the global macro opportunities (or losses) in currencies and emerging markets happened because of some frictions outside of the financial markets: currency pegging, government intervention, central bank and policy errors, etc. As the governments and central banks become smarter in their interaction with the markets, such outsize returns may be slowly going away.

- **Computing EWA recursively:** Today is day $t-1$. Let σ_t be the EWMA volatility estimate using all the information available on day $t-1$ for the purpose of forecasting the volatility on day t. Notice the dating convention: the time-t estimate is observed on day $t-1$. In my personal opinion, we should date σ_t by $t-1$, not t. But this is the convention in this area. So let's go with convention.

  Moving one day forward, it's now day t. After the day is over, we observe the realized return R_t. We now need to update our EWMA volatility estimator σ_{t+1} using the newly arrived information (i.e. R_t):

$$\sigma_{t+1}^2 = \lambda\,\sigma_t^2 + (1-\lambda)\,R_t^2\,. \tag{1}$$

  A good exercise for you would be to start right from the beginning,

$$\sigma_2^2 = \lambda\,\sigma_1^2 + (1-\lambda)\,R_1^2$$

  and then apply the recursive formula a few times to convince yourself that this recursive approach does get you the exponential weighting scheme of EWMA:

$$\begin{aligned}
\sigma_3^2 &= \lambda\,\sigma_2^2 + (1-\lambda)\,R_2^2 = \lambda^2\,\sigma_1^2 + (1-\lambda)\left(\lambda R_1^2 + R_2^2\right)\\
\sigma_4^2 &= \lambda\,\sigma_3^2 + (1-\lambda)\,R_3^2 = \lambda^3\,\sigma_1^2 + (1-\lambda)\left(\lambda^2 R_1^2 + \lambda R_2^2 + R_3^2\right)\\
\ldots\\
\sigma_t^2 &= \lambda^{t-1}\,\sigma_1^2 + (1-\lambda)\left(\lambda^{t-2} R_1^2 + \lambda^{t-3} R_2^2 + \ldots + R_{t-1}^2\right)
\end{aligned}$$

  For those of you who like things to be precise: as $t \to \infty$, we are back to the exact formulation of the EWMA. And whatever σ_1 we started with does not make a difference.

  If you are an Excel user, you will appreciate the convenience of this recursive formula. If you care about saving CPU time, you will also appreciate the convenience of this recursive formula. When we update the information on day t to calculate σ_{t+1}, all of the past information has been neatly summarized by σ_t. The new information waiting for us to be included is the realization of R_t. We weight the new information R_t^2 by

$1 - \lambda$ and "decay" the old information σ_t^2 by λ. Adding these two pieces together, we get the updated variance estimate. It would be difficult not to appreciate the elegance of this recursive approach. No?

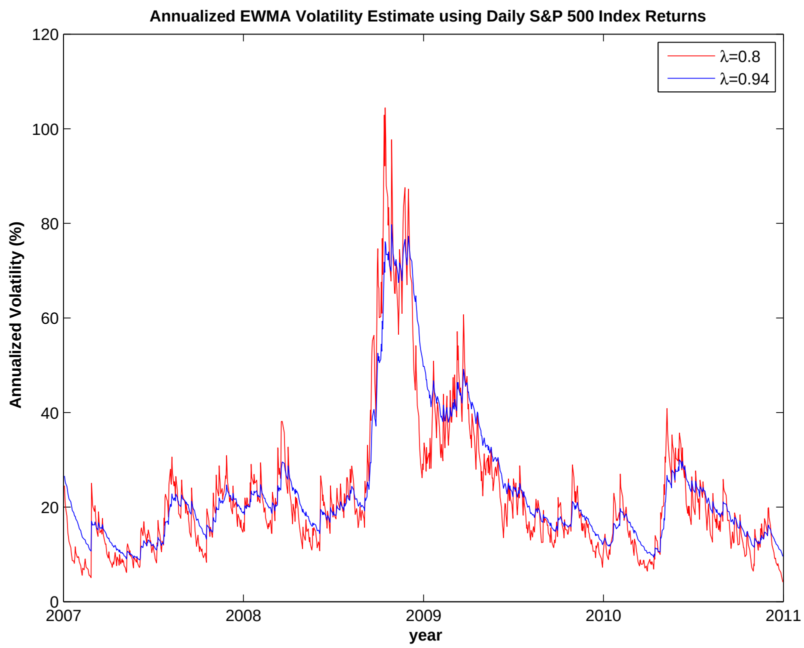
- **The auto-correlation coefficient:** Another way to understand the recursive formula of Equation (1) is that imposes the dynamic structure of σ^2: persistent with an auto-correlation coefficient of λ.

  Recall that we regress stock return R_{t+1} on its own lag R_t to examine the stock return predictability. We find that from 1926 to 2004, the auto-correlation coefficient is positive and statistically significant. But the magnitude of the correlation is very small. Moreover, this predictability is not very robust: over the various subsamples, the auto-correlation coefficients become statistically insignificant. In other words, the random walk model with zero auto-correlation is a reasonable model for the stock returns.

  When it comes to the dynamic structure of volatility, however, the auto-correlation coefficient plays a rather important role. Models such as EWMA and GARCH became popular in practice because they allow volatility to be persistent with a high auto-correlation cofficient. In estimating the auto-correlation in stock returns, we can simply run a regression. In the case of volatility, however, we need to estiamte the volatility along with the coefficient λ. For this, we need a more structured estimation approach than a regression. (If you get into this area called Econometrics, you will realize that the essence is really the same. In particular, a linear regression is really the product of a maximum likelihood estimation. See Appendix A.)

- **Estimating the decay factor:** Figure 3 provides a graphical connection between the decay factor λ and the sample size. A strong decay factor, say $\lambda = 0.8$, pays more attention to the current events and underweights the far-away events more strongly. As a result, the effective sample size is smaller with a stronger decay factor (e.g., smaller λ). As you can see from Figure 5, a strong decay factor improves on the timeliness of the volatility estimate, but the smaller sample size makes the estimate noisier and less precise. On the other hand, a weaker decay factor, say $\lambda = 0.97$, improves on the smoothness and precision, but that estimate could be sluggish and slow in response to changing market conditions, as reflected in Figure 5. So there is a tradeoff.

  - **Minimize RMSE:** Let's consider two ways to pick the optimal decay factor. In the first approach, we would like to minimize the forecast error between the model's prediction and the actual realization. Recall, on day t, we form σ_{t+1} as a



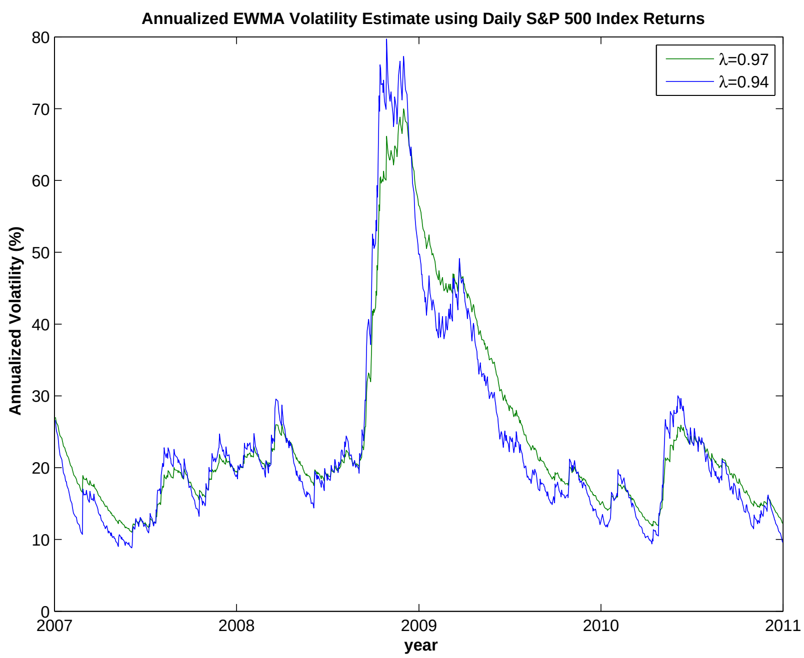


Figure 5: Time Series of EWMA Volatility Estimates with Varying Decay Factors.

forecast for the volatility on day $t+1$. So the model's forecast error is $R_{t+1}^2 - \sigma_{t+1}^2$. Summing these forecast errors over the sample period, we calculate the root mean squared error (RMSE) by,

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2}$$

Note that the only parameter at our disposal is λ. Everything else comes from the data. So let's find the optimal λ^* that minimizes the forecast error:

$$\lambda^* = \arg\min_{\lambda\in(0,1)} \text{RMSE} = \arg\min_{\lambda\in(0,1)} \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2}$$

– **MLE:** In the second approach, let's use the maximum likelihood estimation. To be honest, using the MLE on this problem is really an overkill, but I would like to use this opportunity to introduce you to MLE. Anybody working with data should have done MLE at least once in their life.

Recall that we talk about the pdf of a normal, which is a Gaussian function. In our current setting, the volatility is time-varying. So the stock return R_{t+1} is normally distributed only when conditioning on the volatility estimate σ_{t+1}:

$$f\left(R_{t+1}|\sigma_{t+1}\right) = \frac{1}{\sqrt{2\pi}\sigma_{t+1}} e^{-\frac{R_{t+1}^2}{2\sigma_{t+1}^2}}.$$

Notice that if I wanted to be very precise, I should have replaced R_{t+1}^2 by $(R_{t+1} - \mu)^2$ and use the MLE to estimate both λ and μ. But we talked about this. Setting $\mu = 0$ here is a good compromise to make.

The next step of MLE is to take log of the pdf:

$$\ln f\left(R_{t+1}|\sigma_{t+1}\right) = -\ln\sigma_{t+1} - \frac{R_{t+1}^2}{2\sigma_{t+1}^2},$$

I dropped 2π since it is a constant will not affect anything we will do later. We now add them up to get what econometricians call log-likelihood (llk):

$$\text{llk} = -\sum_{t=1}^{T}\left(\ln\sigma_{t+1} + \frac{R_{t+1}^2}{2\sigma_{t+1}^2}\right)$$

As you can see, the only parameter in llk is our choice of λ. It turns out that the best λ is the one that maximizes llk. In practice, we take -llk and minimize -llk instead of maximizing llk.

What we just did came straight out of Econometrics. A good textbook on this topic is the Time Series Analysis by James Hamilton. Read in particular the chapter on Generalized Method of Moments. Most of the econometrics tasks we encounter in Finance can be understood from the perspective of GMM, which was developed by Prof. Lars Hansen at University of Chicago. Prof. Hansen shared the 2013 Nobel Prize with Prof. Eugene Fama and Prof. Robert Shiller. In the Appendix, I include my old PhD-era code for estimating the standard errors of mean, std, skewness, and kurtosis. As you can see, my approach was very much influenced by the GMM approach. In the Appendix, I wrote a brief note on MLE and linear regression, which could be a nice entry point to motivate you to learn more about Econometrics.

- **ARCH and GARCH:** The ARCH model, autoregressive conditional heteroskedasticity, was proposed by Professor Robert Engle in 1982. The GARCH model is a generalized version of ARCH. ARCH and GARCH are statistical models that capture the time-varying volatility:

$$\sigma_{t+1}^2 = a_0 + a_1 \, R_t^2 + a_2 \, \sigma_t^2$$

As you can see, it is very similar to the EWMA model. In fact, if we set $a_0 = 0$, $a_2 = \lambda$, and $a_1 = 1 - \lambda$, we are doing the EWMA model.

So what's the value added? This model has three parameters while the EWMA has only one. So it offers more flexibility (e.g., allows for mean reversion and better captures volatility clustering). If you are interested in estimating the GARCH model, you can use the MLE method we just discussed. Instead of estimating σ_{t+1} using EWMA, you use the GARCH model. The EWMA has only one parameter λ to estimate. The GARCH model has three parameters to estimate a_0, a_1, and a_2. You will find that, just like λ, a_2 is very close to one. In fact, a_2 captures the auto-correlation of the variance σ_t^2: an autocorrelation coefficient that is close to one indicates a very persistent time series. Moreover, after some calculation, you notice that the long-run mean of the variance in this model is $a_0/(1 - a_1 - a_2)$. You can see how having additional parameters could provide more flexibility to the model.

The GARCH model has a pretty strong influence, and you are encouraged to dig deeper into the model if it interests you. We used to study quite a bit of GARCH at Stanford GSB. But looking back, I feel that I get most of the key intuitions by working with EWMA. Where there are too many moving parts and too many parameters, you tend to focus more on dealing with the formulas and parameters and lose track of the essence of the problem. That's why simplicity is always preferred.

3 EWMA for Covariance

As mentioned at the beginning, our goal is to create the variance-covariance matrix for the key risk factors influencing our portfolio. Suppose that there are two risk factors affecting our portfolios. Let R_t^A and R_t^B be the realized day-t returns of these two risk factors. We estimate the covariance between A and B by

$$\text{cov}_{t+1} = \lambda \, \text{cov}_t + (1-\lambda) \, R_t^A \times R_t^B$$

And their correlation:

$$\text{corr}_{t+1} = \frac{\text{cov}_{t+1}}{\sigma_{t+1}^A \sigma_{t+1}^B} \, ,$$

where σ_{t+1}^A and σ_{t+1}^B are the EWMA volatility estimates.

This calculation of covariance and correlation is pretty straightforward once you master the EWMA recursive formula. But let me use this opportunity to bring in volatility as a risk factor and emphasize on its importance. As recent as the early 2000s, volatility as a risk factor was not widely monitored by market participants. Of course, sophisticated investors pay attention to their exposure to volatility risk. The general intuition is that if you are short on volatility, you are going to lose during crisis. On the other hand, if you are long on volatility, you are partially hedged during these crises. Exposures to volatility risk comes certain non-linearity in one's position. The most straightforward way to be long on volatility is to buy at-the-money S&P 500 index options. As we will cover in our options class, such long positions usually are expensive. That is, you are paying a premium for such positive exposures.

Since the 2008 financial crisis, the volatility risk has got a broader audience. By now, the VIX index is reported daily in a prominent position along with the Dow, the S&P, and Nasdaq. It's often called the fear gauge. Figure 6 plots the historical VIX for the past 15 years. Going over the various events in the past, you can certain appreciate why it is called fear gauge.

Another important observation about the volatility risk factor is its increasing negative

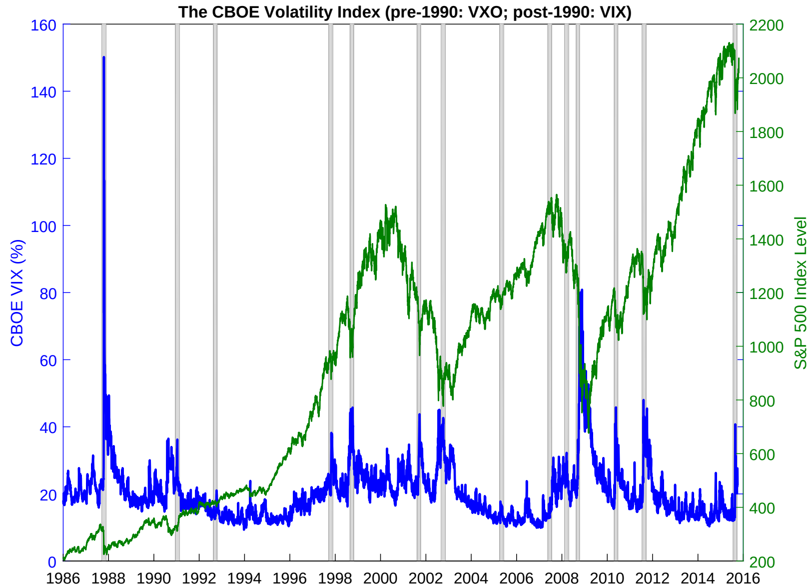


Figure 6: Time Series of CBOE VIX.

correlation between the market risk. Figure 7 plots the EWMA correlation estimate between daily returns of the S&P 500 index and the daily changes in VIX. As you can this, this correlation is always negative. The fact that there is a negative correlation between the stock market return and its volatility has always been documented, well before CBOE published its VIX index. As you can see, during the early sample period, the correlation was hovering around -50%. Was I was a PhD student working on this topic in the late 1990s, a typical number for this correlation would be -60%. There is certainly a trend of this correlation becoming more negative in recent times. After the 2008 financial crisis, this correlation certainly has experienced a regime switch to a more negative territory.

A negative correlation implies that whenever market drops down, the volatility goes up. Using the interpretation of VIX as a fear gauge, this means that a down market is coupled with increasing fear. The more negative correlation in recent years means a higher level of sensitivity to down markets: a market sitting at its edge, more easily spooked. As we move on to the options market, we will look at the "fear" component in VIX more closely.

4 Calculating Volatility and VaR for a Portfolio

- **Portfolio volatility:** Suppose that our portfolio has only two risk factors, whose daily returns are R^A and R^B, respectively. Suppose we've done our risk mapping from individual positions to portfolio weights on these two risk factors: w_A and w_B. If we

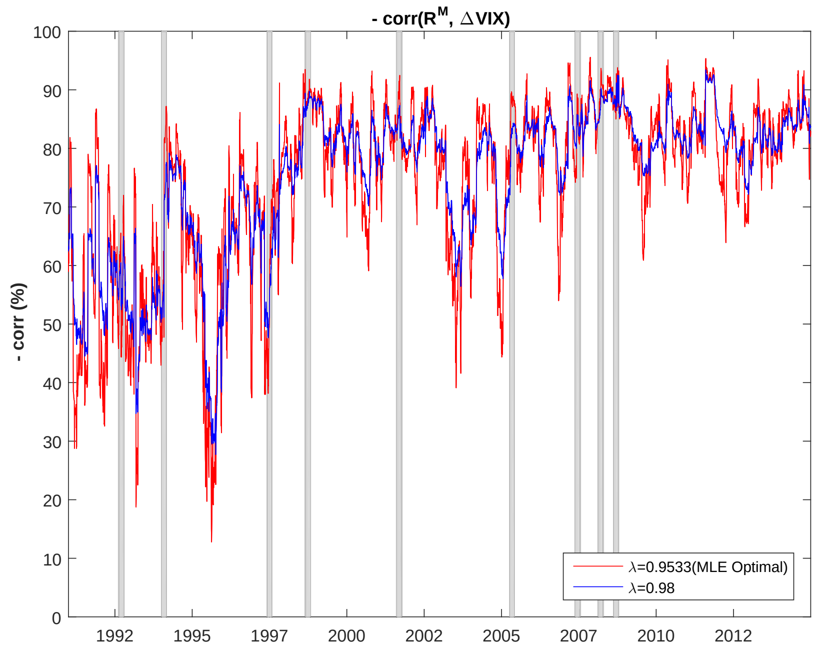


Figure 7: Time Series of CBOE VIX.

focus our attention on the risk part of our portfolio, then we can even normalize the portfolio weights so that $w_A + w_B = 1$. Let's also assume that at the moment, our risk portfolio has a market value of \$100 million.

We first construct a variance-covariance matrix for our risk factors:

$$\Sigma_t = \begin{pmatrix} (\sigma_t^A)^2 & \rho_t^{AB}\sigma_t^A\sigma_t^B \\ \rho_t^{AB}\sigma_t^A\sigma_t^B & \left(\sigma_t^B\right)^2 \end{pmatrix}$$

It is a 2×2 matrix, since we have only two risk factors. If you have 100 risk factors in your portfolio, then you will have a 100×100 matrix. For example, in JPMorgan's RiskMetrics, roughly 480 risk factors were used. In Goldman's annual report, 70,000 risk factors were mentioned. A risk manager deals with this type of matrices everyday and the dimension of the matrix can easily be more than 100, given the institution's portfolio holdings and risk exposures. Notice also the timing here. For σ_t, you are actually using all of the market information on day $t-1$ (e.g., daily returns of assets A and B up to day $t-1$), for the purpose of forecasting volatility for day t.

Let's time-stamp our portfolio weights by the actual time. Suppose today is $t-1$ and let the portfolio weight be written in a vector form:

$$w_{t-1} = \begin{pmatrix} w_{t-1}^A \\ w_{t-1}^B \end{pmatrix}$$

Our portfolio volatility is

$$\sigma_t^2 = \begin{pmatrix} w_{t-1}^A & w_{t-1}^B \end{pmatrix} \times \begin{pmatrix} (\sigma_t^A)^2 & \rho_t^{AB}\sigma_t^A\sigma_t^B \\ \rho_t^{AB}\sigma_t^A\sigma_t^B & \left(\sigma_t^B\right)^2 \end{pmatrix} \times \begin{pmatrix} w_{t-1}^A \\ w_{t-1}^B \end{pmatrix}$$

Using the notation we've developed so far, we can also write

$$\sigma_t^2 = w_{t-1}' \times \Sigma_t \times w_{t-1} \,,$$

which involves using mmult and transpose in Excel.

- **Portfolio VaR:** Suppose that our daily portfolio volatility is σ (daily number, unannualized). The value of our portfolio, marked to the market, is \$100 million. Assuming that the portfolio return is normally distributed, we can estimate how much we stand to lose in market value if a 5% tail event happens to our portfolio over the next day:

$$\text{VaR (95\%)} = \text{portfolio value} \times 1.645 \times \sigma \,,$$

where 1.645 is the critical value for a 5% tail event. Some firms report 99% VaR, which corresponds to the loss in market value if a 1% tail event happens to the portfolio over the next day: our portfolio over the next day:

$$\text{VaR (99\%)} = \text{portfolio value} \times 2.326 \times \sigma \,,$$

where 2.326 is the critical value for a 1% tail event.

As you can see, there are two main drivers for the portfolio VaR: the market value of the portfolio and the portfolio volatility. The market value tells you the dollar exposure of your firm's trading book to risky assets and the portfolio volatility tells you how volatile the risky assets are. For a chief executive of a firm, the VaR number is a useful summary of these two important components of a firm's trading book. Although VaR is framed as a consideration over tail events, it is not really a measure of tail risk since it is driven by volatility. We will come back to this issue again when we spend a class on Market Risk Management.

Appendix

A MLE and Linear Regression

Let's consider the linear regression:

$$Y_t = \alpha + \beta X_t + \epsilon_t \,,$$

where if we replace X with $R^M - r_f$ and Y with $R^i - r_f$, we are back with our favorite CAPM regression.

Thinking in terms of MLE, we focus on the distribution of the regression residual ϵ_t. We assume that ϵ_t is i.i.d. with normal distribution: zero mean and volatility of σ_ϵ. Now let's repeat the MLE steps for this regression:

- We write down the pdf for the residual:

$$f(\epsilon_t) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} e^{-\frac{\epsilon_t^2}{2\sigma_\epsilon^2}}$$

- Take the log of the pdf:

$$\ln f(\epsilon_t) = -\ln \sigma_\epsilon - \frac{\epsilon_t^2}{2\sigma_\epsilon^2} = -\ln \sigma_\epsilon - \frac{(Y_t - \alpha - \beta\, X_t)^2}{2\sigma_\epsilon^2} \,,$$

  where 2π was again dropped.

- Summing up all observations to get

$$\text{llk} = \sum_{t=1}^{T} \left(-\ln \sigma_\epsilon - \frac{(Y_t - \alpha - \beta\, X_t)^2}{2\sigma_\epsilon^2} \right)$$

- Find the parameter values (σ_ϵ, α, and β) that will minimize this,

$$-\text{llk} = T \times \ln \sigma_\epsilon + \frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{T} (Y_t - \alpha - \beta\, X_t)^2$$

In the EWMA case, we use the computer to minimize -llk by varying λ. In the GARCH case, we use the computer to minimize -llk by varying a_0, a_1 and a_2. Here, we can actually do it by hand. Let's forget σ_ϵ for now, and focus on α and β. To miminize $-$llk is the same as finding α and β so that

$$\frac{\partial\,\mathrm{llk}}{\partial\alpha} = \frac{\sum (Y_t - \alpha - \beta\, X_t)}{\sigma_\epsilon^2} = 0$$

and

$$\frac{\partial\,\mathrm{llk}}{\partial\beta} = \frac{\sum [X_t\ (Y_t - \alpha - \beta\, X_t)]}{\sigma_\epsilon^2} = 0\,.$$

Solving for the optimal α and β then reduces to solving the above two equations. The first derivatives of llk with respect to the model parameters (e.g., α and β) are also called score. If there are two parameters, then the score is a vector of two. In our current case, the score should be a vector of three because there are three parameters: α, β, and σ_ϵ. But as agreed, let's focus only on α and β and forget about σ_ϵ.

Solving for the partial derivative (score) with respect to α, we have,

$$\alpha = \frac{1}{T}\sum Y_t - \beta\,\frac{1}{T}\sum X_t$$

Solving for the partial derivative with respect to β, we have

$$\sum X_t\, Y_t - \alpha \sum X_t - \beta \sum X_t^2 = 0$$

Plugging the solution for α into the equation above, we have:

$$\sum X_t Y_t - \frac{1}{T}\sum X_t \sum Y_t = \beta\left(\sum X_t^2 - \frac{1}{T}\left(\sum X_t\right)^2\right)$$

Let me divide both sides of the equation by T so that you can see the result more clearly,

$$\frac{1}{T}\sum X_t Y_t - \left(\frac{1}{T}\sum X_t\right)\left(\frac{1}{T}\sum Y_t\right) = \beta\left(\frac{1}{T}\sum X_t^2 - \left(\frac{1}{T}\sum X_t\right)^2\right)$$

What we have is,

$$\mathrm{cov}(X,Y) = \beta\,\mathrm{var}(X)$$

So, as a by product of our derivation, you get to know why running a regression gets you the CAPM beta.

Also, for those of you who think more carefully, the fact that we assume ϵ is normally distrubted might be bothering you. Don't be. Even if ϵ is not normally distributed, we can still do this procedue, which is called quasi-maximum likelihood estimation. The estimates might not be the most efficient, but they are still consistent. By going through this derivation, my intension is to lead some of you to the door of econometrics. If you are interested, go ahead. If not, turn around. One key calculation left out is how to calculate the standard errors of α and β. For those of you who are interested in learning more, I would recommend the chapters on GMM and MLE of James Hamilton's book on Time Series Analysis.

B Matlab Code

Code 1: Plot SMA Volatility Estimates

```
load SP500_Daily.txt;
Data=SP500_Daily;

yr=Data(:,1);
mn=Data(:,2);
dy=Data(:,3);
Time=datenum(yr,mn,dy);

Ret=Data(:,4)*100;

time_mn=[]; vol_mn=[]; mu_mn=[];
for i_yr=min(yr):1:max(yr),
   for i_mn=1:12,
       i_Ret=Ret(yr==i_yr&mn==i_mn);
       if ~isempty(i_Ret),
          time_mn=[time_mn; datenum(i_yr,i_mn,30)];
          [M,SE]=stat_fun(i_Ret);
          vol_mn=[vol_mn; [std(i_Ret)*sqrt(252) SE(2)*sqrt(252)]]; %monthly
              vol estimate with standard error
          mu_mn=[mu_mn; [mean(i_Ret) std(i_Ret)/sqrt(length(i_Ret))]]; %
              monthly mu estimate with standard error
       end
   end
end

% plot SMA vol estimates
figure(2);
```

```
plot(time_mn,vol_mn(:,1),'b-');
hold on;
datetick('x','yyyy')
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);
BND=axis;
plot([BND(1) BND(2)],std(Ret)*sqrt(252)*[1 1],'k--')
hold off

% plot SMA vol estimates with confidence intervals
figure(10);
plot(time_mn,vol_mn(:,1)+1.96*vol_mn(:,2),'g-',time_mn,vol_mn(:,1)-1.96*
    vol_mn(:,2),'m-');
hold on
plot(time_mn,vol_mn(:,1),'b-','LineWidth',2);
hold off
BND=axis;
datetick('x','yyyy')
legend('95% Confidence, Upper','95% Confidence, Lower')
title('\bf SMA estimates of \sigma and their 95% confidence intervals');
ylabel('\bf Annualized Volatility (%)');
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);

% plot SMA mu estimates with confidence intervals
figure(11);
plot(time_mn,mu_mn(:,1)+1.96*mu_mn(:,2),'g-',time_mn,mu_mn(:,1)-1.96*mu_mn
    (:,2),'m-');
hold on
plot(time_mn,mu_mn(:,1),'b-','LineWidth',2);
hold off
BND=axis;
datetick('x','yyyy')
legend('95% Confidence, Upper','95% Confidence, Lower')
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);
ylabel('\bf Monthly Average of Daily Returns (%)');
title('\bf SMA estimates of \mu and their 95% confidence intervals');
```

```
% plot SMA vol estimates together with NBER recessions
figure(3);
plot(time_mn,vol_mn(:,1),'b-');
BND=axis;
hold on;
datetick('x','yyyy')
FY=[BND(4) BND(3) BND(3) BND(4)];
load NBER_Recession.dat;
hold on
for i=1:size(NBER_Recession,1),
  FX=[datenum([NBER_Recession(i,1:2) 1])*[1 1] ...
      datenum([NBER_Recession(i,3:4) 1])*[1 1]];
  if FX(1)> datenum(1962,1,1),
    fill(FX,FY,[0.75 0.75 0.75]);
    hold on
  end
end
plot(time_mn,vol_mn(:,1),'b-','LineWidth',2);
hold on;
plot([BND(1) BND(2)],std(Ret)*sqrt(252)*[1 1],'k--')
hold off;
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);
```

Code 2: Calculating Standard Errors

```
function [MOMENTS, SE]=my_stat(data)

T=size(data,1);

m1=mean(data);
m2=var(data);
m3=mean((data-m1).^3);
m4=mean((data-m1).^4);

MEAN=m1;
STD=sqrt(m2);
```

```
SKEW=m3/m2^(3/2);
KURT=m4/m2^2;


h1=data-m1;
h2=h1.^2-m2;
h3=h1.^3-m3;
h4=h1.^4-m4;
h=[h1 h2 h3 h4];
T=length(h);
R=h'*h/T;
n_moving=5;
for i=1:n_moving
   R_temp=h(i+1:T,:)'*h(1:T-i,:)/T;
   R=R+(R_temp'+R_temp)*(1-i/(n_moving+1));
end
W=inv(R);
D=[-1 0 0 0; 0 -1 0 0; 3*m2 0 -1 0; 4*m3 0 0 -1];

COV=inv(D'*W*D);
SE=sqrt(diag(COV)/T);

D2=1/2/sqrt(m2);
C2=COV(2,2);
SE(2)=sqrt(D2*C2*D2/T);
D23=[-1.5*m3/m2^(5/2) 1/m2^(3/2)];
C23=COV(2:3,2:3);
SE(3)=sqrt(D23*C23*D23'/T);
D24=[-2*m4/m2^3 1/m2^2];
C24=COV([2 4],[2 4]);
SE(4)=sqrt(D24*C24*D24'/T);

MOMENTS=[MEAN STD SKEW KURT]';
```