# Methodology

To further explore our data properties, we used Markov two state autoregressive (lag 1) bivariate multi-asset model. In short, it decomposes descriptive statistics about data into two substates with their unique descriptive statistics. Like in real life every day is not the average day and such assumption is a huge approximation, so now we have two distinct “average days” with known probabilities to change between them during a period. Assumption that there are only two kinds of periods is more precise and useful. Below are implementation of model, implementation of fitting process, results and interpretation.

## Model

We are using the same model as P. Aigner (Modeling and managing portfolios including listed private equity, 2012). Basic bivariate markov-switching model is described by return y at time t function (1) with a mean value μ that depends on state S at time t and a error term ε at time t that is characterized by standard deviation σ that depends on state S at time t. It is called bivariate because both mean and standard deviation depends on state S at time t. States are characterized by probability matrix, where pij is a probability that being in state Si the process in next period will be in state Sj. Error term is assumed to be normally distributed over time and thus E[ε]=0.

Then introduce simple autocorrelation with lag 1 to our model (3). In simple autocorrelation coefficient is state independent, and we will use word simple not to confuse with state-depend autocorrelation. Lag 1 means that only the previous period will affect this period’s returns.

Then introduce multiple assets to our model (4) as column vectors where first element represents Listed Private Equity, second element represents Stocks and third elements represent Bonds. Our model contains 23 unique parameters: **μ**1, **μ**2, **σ**1, **σ**2, **ϕ** as 3x1 vectors, p12, p21 as scalars and Cor1, Cor2, where Cori is 3x3 correlation matric (only 3 elements in it are unique, others are mirrored or 1’s).

Conditional mean and standard deviation is a mean and variance that is specific to a state. Unconditional mean and standard deviation characterizes returns as if there would be only one state. Therefore descriptive statistics as population mean, standard deviation, skewness and kurtosis are equal to unconditional statistics of markov model. Timmermann (Moments of Markov model, 2000) provides formulas to calculate unconditional statistics by moments. First one has to calculate unconditional probabilities **π**i to be in a Si state. It is also called as ergodic probability.

Unconditional mean is calculated by multiplying unconditional probability vector with conditional mean vector for any single asset *i*,

Second moment is used to calculated variance-covariance matrix. For two assets *i* and *j*, variance or covariance is calculated using conditional mean vector , unconditional mean , identity vector filled with 1’s, conditional standard deviation vector , conditional correlation vector and simple autocorrelation coefficient . The sign denotes element-by-element multiplication and everything in bold is a vector.

Third moment is used to calculate unconditional skewness for any asset *i*.

Where bold lowcap letters are vectors and and bold capital letters are matrices. **B** is a backward transition probability matrix that has a direct relationship to **P** as follows:

Fourth moment is used to calculate unconditional kurtosis for any asset *i*.

Timmerman also provides a formula (12) for unconditional autocorrelation. Note that even if we have a simple, state-independent autocorrelation coefficient it does not equal to unconditional autocorrelation.

## Literature review

The basic MS has finite number of states that are characterized by state dependent mean, variance and set of probabilities to change into another state. First time markov switching models were used in economics studying business cycles by (Hamil-ton (1989). There are different variations of markov models. For example basic markov switching, autoregressive markov switching and state-dependent autoregressive markov switching models all used by Allan Timmermann (2000) and adopted others. markov switching poisson multifractal and markov switching multifractal models were introduced by E. CALVET, J. FISHER (2004), but hidden markov-switching model introduced by A. Rossi and G.M. Gallo (2006).

Markov models are favored of its ability to better fit complex data. It has ability to generate non-normal distributions with different coefficients of means, standard deviations, skewness, kurtosis and serial correlation (A. Timmermann, 1999), (A. Taamouti, 2012). Mean reversion, the assumption that after shock the value tends to return a theoretical mean value, is also captured by Markov switching models according to research by Abderrahim Taamouti (2012). It has a better volatility estimate, but the derivation of total volatily depends on variation. And compared to GARCH markov models has advantage in variability of state persistence that describes better exceptional and short-term events. Volatility clustering that characterizes most of real world data is natively captured by model in research by Abderrahim Taamouti (2012).

Autocorrelation can be implemented in the model, but empirical tests shows that it has low significance. Ding et al. (1993) wrote that stock market returns contain little serial correlation. Also several model variations themselves implicit long memory and thus accounts for autocorrelation (E. CALVET, J. FISHER, 2004). All models use maximum likelihood estimator to fit the best parameters to presented data, and others also use filtering method by Hamilton (1994).

Markov models differ greatly. Some of them deals with linear innovations, other with logarithmic. Number of states usually is small, but for example hidden markov model effectively uses dozens of states. Some of them has ergodic states (one active at one period) while others allow for multiple states to be active at the same time. Some deals with discrete times, but poisson multifractal deals with continuous time. Some excludes auto-regression, some includes it independently or state-dependently, others embody characteristics that already accounts for auto-regression. Some of them use independent parameters, others has parameterized less or more of its parameters.

## Fitting

We are using customized ordinary least squaeres (COLS) to fit markov model to data. For fitting target we are minimizing markov model unconditional properties to data descriptive statistics.

Because of our parameters lies in different scales (means in 10-3, kurtosis above 1, others around 0.1) use of ordinary least squares is not feasible. So we turned to RLS but it had also a drawback. For example, mean return of Stocks and LPE are around 0.0001, but mean return of a bond is 0.000001. All three mean returns need equal weighting relative to other measurements, but in RLS case Bond mean return would get 100 higher weight for precision. Therefore we used OLS in a different way – based to underlying data we used different set of constants for estimation to ensure equal representation of all 6 elements. For dataset #2 we used xmean=100000, xstdev=1000, xcor=0.5, xskew=1, xkurt=0.1.

We used the following constraints: , , , and probability matrix constraint in (2).

For the process itself we use Monte Carlo methodology iteratively. Let us define three variables q, d and k where q is drop rate, d is allowed distance with initial value half-length between constraints, and k is a rate at which d decreases. We uniformly draw 1000 samples of 1000 parameter sets, choose the best fitted parameter set for every sample, drop worst samples at rate of q and decrease d at rate of k. In next step every 1000/q sample is individually considered as a starting point for next 1000 uniformly drawn parameter sets within distance d and inside constraints, and this process iterates until best COLS between samples ceases to go down every iteration under a specified threshold. Then all process is restarted with this sample as a starting point.

Such strategy was chosen by various unexpected reasons and all arguments mentioned onwards were practically tested or gained through practical experimentation. First, size of allowed space seems to matter the most. If starting constraints for mean were 3 standard deviations above statistical mean or if allowed space for conditional standard variance was multitude of 5 or more no reasonable results were gained at all. Sometimes in 0.5mil random parameter sets none is superior to original sample, but after little iteration when d has considerably lowered it starts to find up to 100 times better parameter sets. Therefore starting constraints for mean and standard variance were introduced, and variable d was invented to limit and focus the searching space. In practice (after experimenting with constants) those two improvements were crucial to gain reasonable results and improved accuracy by 2-3 magnitudes. Secondly, q theoretically should bias the results by arbitrary filtering, therefore it was introduced only lately after witnessing in every test that after 2rd up to 4th iteration (depending on constants) in practice lower tail results do not get better, but on other hand q noticeably improves performance. Third, balancing weights for COLS is not only important to get a significant result, but serves also other needs. For example by overpraising mean weight all process starts to work faster and more precise. I suspect it is because all other properties are very sensitive to mean therefore getting it right first avoids huge amounts of local minimum pitfalls. Fourth, correct parameter choice is also important. For example at the start process was built to work with variance-covariance matrix instead of standard deviation and correlation, but that was unable to produce anything significant. Most probably it was because random searching is done linearly over space and standard deviation with correlation is more linear than variance. Fifth, mostly of the time all good fittings has a unconditional probability of state 1 above 99%. That effectively cancels out the markov model as a whole because markov model with 100% weight in one state is the same one-state world descriptive statistics. As fitting become more effective these 99% fits become extinct. It could be explained by a limited precision single state model can describe data and as fitting were more optimized finding better precision after a threshold only two-state models could be found, or on other hand some unseen local minimum problem was solved by unnoticed change in model .

All statistical software do not provide the model, neither anything close enough that could be usable as a substitute. For the record, most similar tool found was hidden markov model by MATLAB. Also searchable internet is void of ready to use special tools, scripts or instructions, and even no wikipedia articles or discussions about this specific markov model variation were found. Therefore all the processes were implemented by team members themselves in programming language Lua during the course consulting mainly academic papers. You are welcome to consult the code itself and various raw data online[1].

## Results

### Dataset #1

Our best fit for dataset #1 is in the table below. As it is seen, mean value is precise only 2 digits after comma and for LPE is of by 0.6% or 30 times but from other side it is within 1.5 standard deviations. Standard deviation is rather precise by 2 digits after comma, variance-covariance matrix is very precise by 6 digits after comma. Skewness and autocorrelation is off by factor of 10 and Kurtosis is off by factor of 100. Relative precision can be controlled by coefficients in COLS and for Dataset #2 we raised relative weight for mean value.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Statistical properties, Dataset #1 | | | | Markov fitted properties, Dataset #1 | | |
|  | GE LPE | GE Stocks | GE Bonds | GE LPE | GE Stocks | GE Bonds |
| Mean | 0.000159 | 0.000175 | 0.000012 | -0.006113 | 0.000630 | -0.000277 |
| VCV (1) | 0.000023 | 0.000000 | 0.000000 | 0.000060 | -0.000003 | 0.000001 |
| VCV (2) | 0.000000 | 0.000028 | 0.000000 | -0.000003 | 0.000060 | -0.000001 |
| VCV (3) | 0.000000 | 0.000000 | 0.000000 | 0.000001 | -0.000001 | 0.000000 |
| St.dev | 0.004790 | 0.005273 | 0.000315 | 0.007726 | 0.007734 | 0.000449 |
| Skewness | -1.581739 | -0.533728 | -0.333150 | -0.074609 | -0.030972 | 0.071477 |
| Kurtosis | 24.552038 | 6.345066 | 26.810331 | 0.222705 | -0.058676 | -0.011517 |
| AutoCor | -0.072788 | 0.126724 | -0.366573 | -0.745823 | -0.575322 | -0.035191 |

At first we see that unconditional probabilities to be in states are 38.4% and 61.6% but standard deviations are very similar, as well mean for LPE and Bonds. But mean return for Stocks are inverse, and all correlation changes between states. This suggest a time period where LPE and Bonds are indifferent, but Stocks have significant growing periods or falling periods and in this falling phase both LPE and Bonds move to opposite direction. Conditional probabilities show that being in state two one is more likely to stay in state two than being in state two. To sum up, this don’t look realistic but earned best COLS score.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Markow Autoregressive (lag 1) bivariate joint 3 asset model cond. Params, Dataset#1 | | | | | | |
| uProb. | 0.384234 | 0.615766 |  | Prob (1) | 0.395610 | 0.604390 |
| AutoCor | -0.755549 | -0.688452 | -0.040442 | Prob (2) | 0.377136 | 0.622864 |
|  | State 1 | | | State2 | | |
| Mean | -0.007210 | 0.004547 | -0.000447 | -0.005429 | -0.001814 | -0.000172 |
| Cor (1) | 1.000000 | 0.581946 | 0.828801 | 1.000000 | -0.564372 | -0.181633 |
| Cor (2) | 0.581946 | 1.000000 | -0.129448 | -0.564372 | 1.000000 | -0.397085 |
| Cor (3) | 0.828801 | -0.129448 | 1.000000 | -0.181633 | -0.397085 | 1.000000 |
| VCV (1) | 0.000042 | 0.000018 | 0.000002 | 0.000015 | -0.000012 | 0.000000 |
| VCV (2) | 0.000018 | 0.000023 | 0.000000 | -0.000012 | 0.000029 | -0.000001 |
| VCV (3) | 0.000002 | 0.000000 | 0.000000 | 0.000000 | -0.000001 | 0.000000 |
| St.dev | 0.006460 | 0.004753 | 0.000399 | 0.003878 | 0.005369 | 0.000445 |

### Dataset #2

Rasmuss is going to say what does these columns mean in real life language :p

Below is output of best fit for 2nd dataset. We can see that mean and standard deviation matches four digits after comma, skewness and auto-correlation is less than ten times off but kurtosis doesn’t match at all. As we discussed above weights for COLS is going to after this relative precision but it is worthy to note that if other properties were overweighted then only bad overall matches were gained.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Statistical properties | | | | Markov fitted properties | | |
|  | GE LPE | GE Stocks | GE Bonds | GE LPE | GE Stocks | GE Bonds |
| Mean | 0.000104 | 0.000075 | 0.000042 | 0.000158 | -0.000046 | 0.000086 |
| VCV (1) | 0.033534 | -0.000018 | -0.000001 | 0.033545 | -0.000054 | -0.000024 |
| VCV (2) | -0.000018 | 0.000044 | 0.000000 | -0.000054 | 0.000048 | 0.000000 |
| VCV (3) | -0.000001 | 0.000000 | 0.000000 | -0.000024 | 0.000000 | 0.000000 |
| St.dev | 0.183122 | 0.006658 | 0.000211 | 0.183152 | 0.006932 | 0.000221 |
| Skewness | -0.359429 | -0.013262 | 0.151860 | -0.232509 | -0.002668 | 0.327011 |
| Kurtosis | 20.092276 | 7.412959 | 34.388661 | -0.335328 | 0.000595 | 0.193181 |
| AutoCor | -0.271805 | -0.019537 | -0.152000 | 0.087494 | 0.105883 | 0.225368 |

Below we see conditional properties of markov model. At first glance we see that unconditional probabilities are 53.45% and 46.55% and both states has different properties. In conditional probability matrix we see that these states has a great chance to change and first state is more stable than second state. The greatest difference is mean return for LPE – in first state it is growing at 11.47% but in second falling at -13.14%. Other assets has small but inverse situation. Second greatest inversion is correlation – this fitted model shows high inverse relantionships between LPE and Stocks that flips in other state. State two has significantly less correlation between assets, and also has a bit greater volatility overall but with no significant difference. Author [#REF] once used markov model to manually sort timeseries data into growing market and falling market to find out that falling markets has greater volatility, but here a fitted model doesn’t show that.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Markow Autoregressive (lag 1) bivariate joint 3 asset model cond. Params | | | | | | |
| uProb. | 0.5344707 | 0.4655293 |  | Prob (1) | 0.2998600 | 0.7001400 |
| AutoCor | 0.5702866 | 0.1081146 | 0.3669755 | Prob (2) | 0.8038254 | 0.1961746 |
|  | State 1 | | | State2 | | |
|  | GE LPE | GE Stocks | GE Bonds | GE LPE | GE Stocks | GE Bonds |
| Mean | 0.1147493 | -0.0004368 | 0.0000026 | -0.1314034 | 0.0004023 | 0.0001812 |
| Cor (1) | 1.0000000 | -0.9946458 | -0.2369431 | 1.0000000 | 0.8515266 | -0.6357739 |
| Cor (2) | -0.9946458 | 1.0000000 | 0.6927961 | 0.8515266 | 1.0000000 | -0.0612679 |
| Cor (3) | -0.2369431 | 0.6927961 | 1.0000000 | -0.6357739 | -0.0612679 | 1.0000000 |
| VCV (1) | 0.0090043 | -0.0006538 | -0.0000034 | 0.0164326 | 0.0007447 | -0.0000182 |
| VCV (2) | -0.0006538 | 0.0000480 | 0.0000007 | 0.0007447 | 0.0000465 | -0.0000001 |
| VCV (3) | -0.0000034 | 0.0000007 | 0.0000000 | -0.0000182 | -0.0000001 | 0.0000001 |
| St.dev | 0.0948910 | 0.0069275 | 0.0001503 | 0.1281898 | 0.0068225 | 0.0002237 |

## Limitations

By building and experimenting with fitting process we learned that one should have a good experience and intuition, because most of things we found out were out of coincidence like testing a bad idea to find it a great one, as well other unintuitive actions and processes.

From first dataset we learned that good estimate score does not mean a good result. Estimating different properties that has different mathematical characteristics is complicated, one need to account for linear means, nonlinear st.deviations, different space constraints, etc to what a quick and dirty solution is subjective weighting by trial and error.

From second dataset we read an interesting story that introduced to interpretation capacity this model has. Sadly because of bad fitting none of this is trustful, and it is extremely hard to get a good fit with Monte Carlo methodology. Also, both datasets have roughly the same total score (between 5 and 10 points) but their contents and visual precision are very different.

Sharpe ratio

New Refs

[1] https://github.com/Akuukis/Markov-switching-model