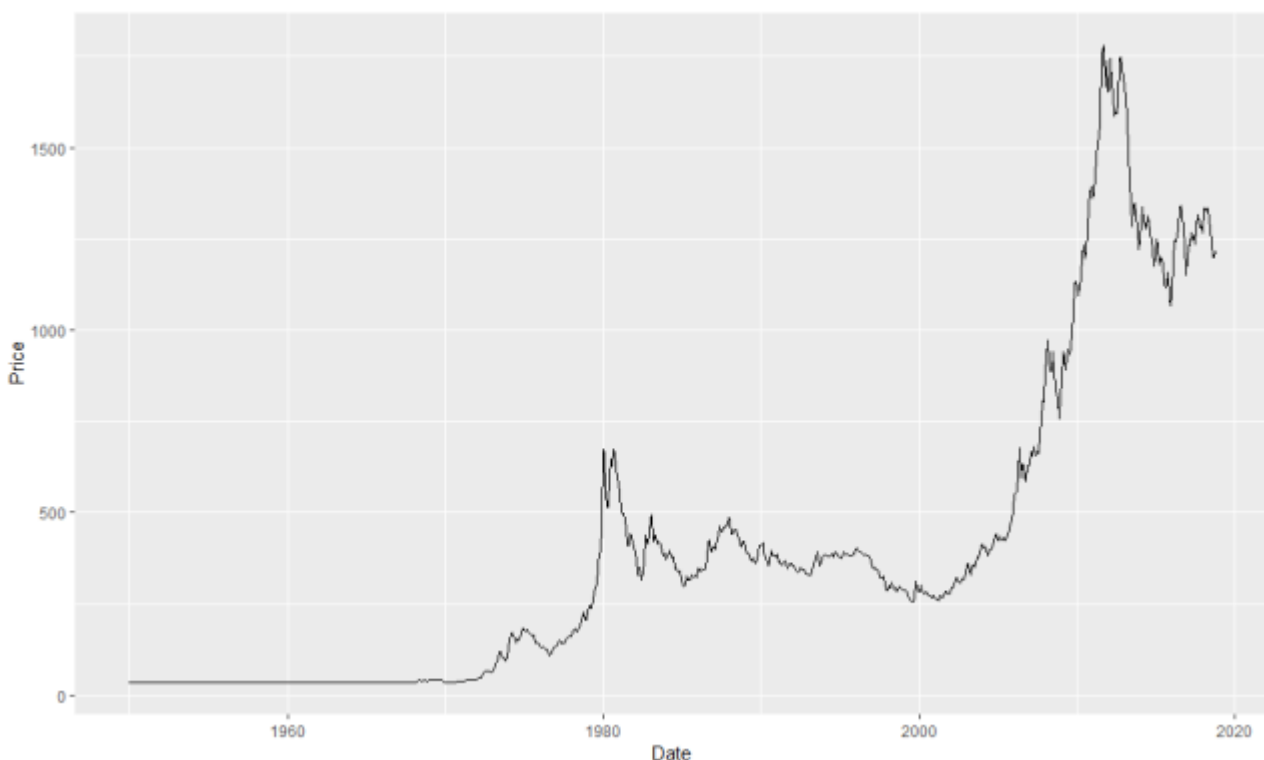


A switching regression model is used to either classify unobservable states or to estimate the transition probabilities for these unobservable states in a time series. It can be considered as a clustering algorithm for time series, which gives you the estimated equation for each cluster and the probability that the time series falls into that cluster at the given point in time. A switching regression can be applied in any business area where you have a time series, and has already been successfully applied by economists to analyze the business cycles, by mutual fund managers in assessing mutual funds and by investment bankers to evaluate stock returns.

I will explain you on the basis of an example what a switching regression can do. A time series is a collection of data where you followed an individual over a longer period of time and recorded specific variables at several points on time. A simple time series is for instance is the price of gold on the stock market. Here you can see the development of the gold price from 1995 until today.

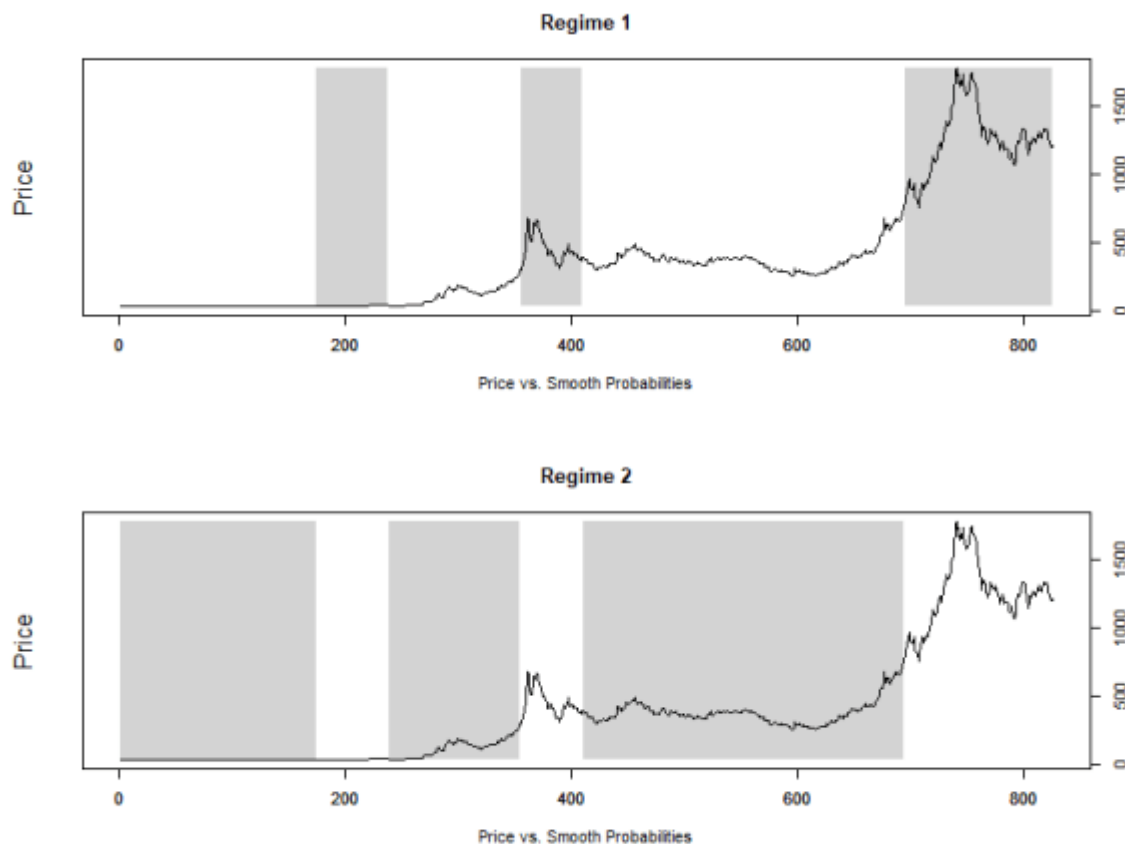


Gold Price Time Series

When you look at the figure, you will realize that fitting a simple linear regression might not be a good idea, because the time series does not grow in a straight line. Ideally, you would hypothesize that the first part until approximately 1970 would fit to a rather very flat regression line, the parts from 1970 until 1983 and from 2000 until 2015 to a steeply

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

increasing regression line, the part from 1983 until 2000 to a mildly decreasing regression line. A switching regression model would help you first to identify, how many different unobservable phases are there, what are their estimated equations, how does the influence of certain variable differ depending on the state, and what is the probability that the time series is in any of the different phases at any point in time. Here is the example, what states a switching regression model would identify for the gold price time series:



Example of different regimes detected by a Switching Regression

Table of Contents

- [Business Area & Impact: Find the unobservable in your time series](#)
- [Procedure](#)
 - [Step 1: Set up Data](#)
 - [Step 2: Decide on States](#)
 - [Step 3: Estimate the Switching Model](#)

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

- [Step 4: Evaluate & Interpret Switching Model](#)
 - [Looking at the coefficients](#)
 - [Looking at the graphs](#)
- [Advantages](#)
- [Disadvantages](#)
- [Further Links](#)
- [References](#)
 - [Share this:](#)

Business Area & Impact: Find the unobservable in your time series

A switching regression analysis can be practically applied any field, where you want to analyze different unobservable states in time series. It has been already successfully applied in the area of finance and economics to understand business cycles, asset allocation, stock returns, interest rates, portfolio management, and exchange rates. However, there are also other possible applications in various areas. Here are a few examples:

- **Human Resources:** What is the driver of employee turnover? If you have followed the turnover rate over a longer period of time, you could easily assemble a great dataset with rich information. Now you can analyze whether there are different phases in your companies' development and how the effects of the variables you select change depending on the phase. You might for instance discover that there are three phases, e.g. expansion when your company is growing, stagnation when your company is stable, and decline when your company is shrinking. The switching regression would tell you, what factors would be relevant to keep employees depending on the state.
- **After Sales repair:** What does influence the incoming volume of after sales order? Even here you might find different underlying phases and depending on a certain phase, other factors are more important. That might help you predict the incoming repair volume.
- **Operations:** What are the main influencers of lead time? What if there is a recurring pattern in lead times, a cycle, and depending where you are in the cycle, a variable might have a positive or negative impact on lead time. However, since you ignored possible underlying phases, the negative and positive effects of this very same variable averaged out, so you wrongfully assume that this variable is irrelevant.
- **Marketing:** Are some marketing tools more effective in certain phases? If you think

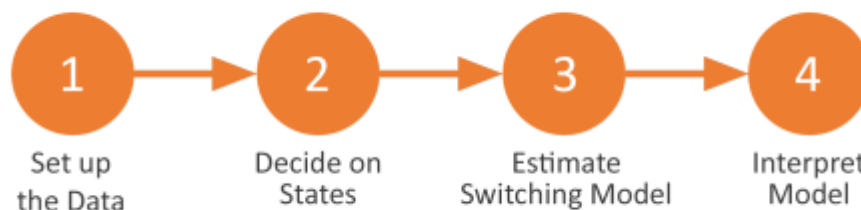
Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

about it, the effects of marketing tools might differ depending on the state of your product. But the question is, what states will you be able to discover and how will these states change the effect size of your variables?

In general, you should consider using a switching regression model for the following five purposes:

1. **Clustering:** If you have experience with clustering algorithms, you might have realized that a switching regression can also be used as a clustering algorithm. If the switching regression assigns certain observations to an underlying states, it can be also interpreted as assigning them to a certain cluster.
2. **State detection:** You want to understand, whether there are different states in your data that you have not observed. Another way to phrase this is: is there any categorical variable that we might have missed out on that interacts with the observed variables?
3. **Estimate differing equations for the different states:** You do not want to find out, whether there are different underlying states that you cannot observe directly, but you also want to understand how the influence of your variables differ depending on the states.
4. **Understand probabilities for states:** You want to understand what is the probability for an observation to be in a certain state and how is that probability influenced?
5. **Switching probabilities between different states:** You want to understand what the probability is for switching from a certain state to another, what are the drivers of state switching and whether certain groups of observations are more or less likely to switch between certain states.

Procedure



Steps for conducting a Switching Regression Analysis

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

You can use a switching regression model when the underlying process is a markov process. This means that your time series is believed to transition over a finite set of unobservable states, where the time of transition from one state to another and the duration of a state is random. It is not difficult to use a switching regression and you can do it in four simple steps. I will show you how to compute and interpret your own switching regression model based on gold data from the introduction.

Step 1: Set up Data

First of all, I need to upload the data and make sure that all the variables have the right data type. In this case, when you upload the data set, you will see that the variable Date is still a character. Therefore, I will convert it to a Date-type using the function `as.Date()`.

```
##### Library
# install.packages("MSwM")
# install.packages("ggplot2")
library(MSwM)
library(ggplot2)
```

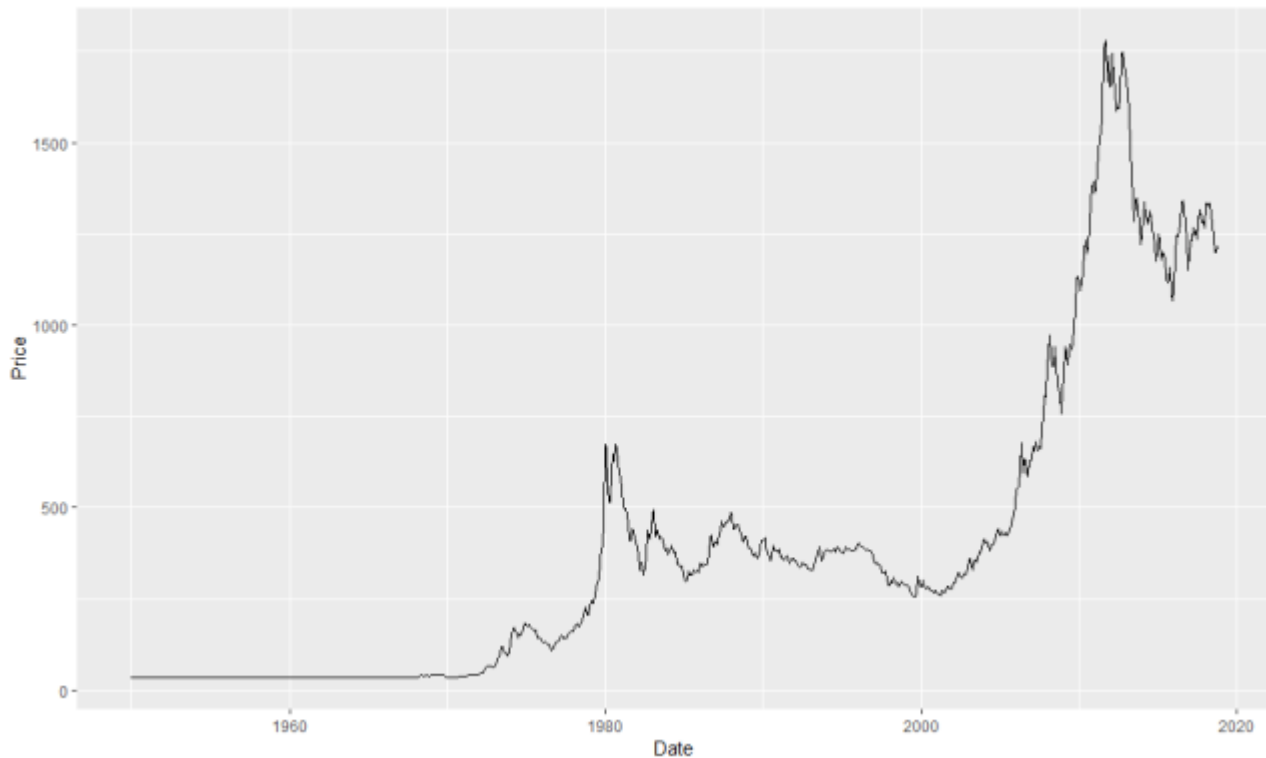
```
##### Step 1: Set up Data
Gold <- read.csv("C:/Users/apivcevic/Desktop/Privat/Switching
Regressions/monthly_csv.csv")
```

```
Gold$Date <- as.Date(paste(Gold$Date,"01",sep="-"), format="%Y-%m-%d")

ggplot(Gold, aes(Date, Price)) + geom_line()
```

Step 2: Decide on States

In the second step, you will need to decide on the number of states that you expect. In the context of switching regressions and Markov processes, you usually say regimes instead of states. However, I will continue using the word states. Your decision on the number of states should be theory-driven. That means that you have a clear theory how many states should be possible and how many states you want to estimate. If you analyze a stock, you might expect only two states: the stock goes up or goes down. Therefore, you would assume only two hidden states. Now let's have look at our example:

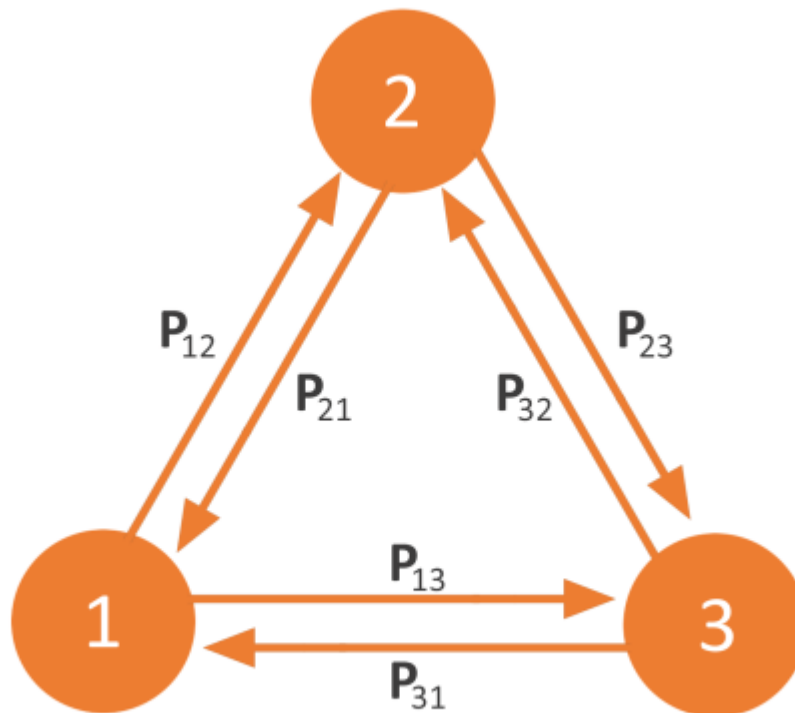


Gold Price Time Series

In our example, I expect three different hidden states. The first one is a stagnating state, the second one is a sharply increasing state that we can observe after 2000, and a volatile stagnating state that we can mostly observe before 2000. Therefore, I assume that there should be three different states. Keep in mind, that you do not want to specify too many states for two reasons. First, the more states you have the more complex the interpretation gets. Second, the estimation of a switching regression model is computationally complex, which means the more data you have and the more states you have, the longer your it will take to compute it.

```
##### Step 2: Decide on States  
nstates <- 6
```

The switching regression will now estimate a different linear equation for each state that we specified. Furthermore, it will calculate the transition probabilities for each state according to the following overview, where p_{ab} stands for the transition probability from state a to state b:



Different states and the transition probabilities

Since I have an economic background, here my small question to you. Why was the price of Gold so stable until 1970 (there is a pretty logical explanation ;))?

Step 3: Estimate the Switching Model

We will use the `msmFit()`-function from the `MSwM`-package to estimate the switching regression. The `msmFit()`-function needs as input a regression model produced by the `lm()`-function.

```
##### Step 3: Estimate Switching Model
olsGold <- lm(Price~Date, Gold)
```

```
msmGold <- msmFit(olsGold, k = nstates, sw = c(FALSE, TRUE, TRUE))
```

At this point, I should mention that there are various types of markov-switching regression models, where each type has its advantages and disadvantages. You can basically apply all

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

statistical tools you know from time series. Here two examples:

- **Univariate or multivariate Markov chains:** First, you can have an univariate Markov chain or multivariate Markov chain. An univariate Markov chain is for instance based on an homogenous Markov chain. That means that it does not have any trend or seasonality. In the context of time series, trend means that the time series is generally going upwards (like in our example) or downwards. Seasonality means that it has a repetitive pattern. For instance ice consumption will show you a seasonality with a peak in summer and a low in winter.
- **Fixed or time-varying transition probabilities:** In switching regression models, it is also possible to let the transition probabilities vary with time. It will make the model more complex, but if there are serious theoretical reasons for letting the probabilities vary over time, it will be useful.
- **First-order or higher-order chain:** A first-order switching regression computes the transition probability based only on the latest state, but not on all previous states.
- **Linear or generalized switching regression:** You can also create switching regressions based on for instance generalized linear models, and not only on linear models. For example, if your variable of interest is a binary variable, you can use a logistic regression analysis.

If you understood the three examples, you will realize that I applied the simplest switching regression model here: a univariate first-order switching regression with fixed transition probabilities. Furthermore, there are two general families of switching regression models:

1. **Markov-switching dynamic regression:** The dynamic models allow states to switch according to a Markov process, but in contrast to the other type, they allow for quick adjustments after a change of state. These types of models are often applied to high frequency data.
2. **Markov-switching AR model:** AR-models allow states to switch according to a Markov process as well, however, they only allow for a gradual adjustment after change. These models are often applied to lower frequency data (quarterly, yearly, etc.).

Step 4: Evaluate & Interpret Switching Model

We can interpret a switching regression model in two ways, first by looking at the coefficients and secondly graphically.

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

Looking at the coefficients

Step 4: Interpret & Evaluate Switching Model
summary(msmGold)

The code will give use the following results:

Markov Switching Model

Call: msmFit(object = olsGold, k = nstates, sw = c(FALSE, TRUE, FALSE))

	AIC	BIC	logLik
	10010.85	10056.59	-5001.427

Coefficients:

Regime 1

```
-----
              Estimate Std. Error    t value  Pr(>|t|)
(Intercept) 121.4889      0.0008 151861.125 < 2.2e-16 ***
Date(S)      0.0209       0.0013   16.077 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 98.68257

Multiple R-squared: 0.8681

Standardized Residuals:

	Min	Q1	Med	Q3	Max
	-9.909234e+01	-1.857703e+01	7.932031e-04	1.963472e+01	1.555179e+02

Regime 2

```
-----
              Estimate Std. Error    t value  Pr(>|t|)
(Intercept) 121.4889      0.0008 151861.125 < 2.2e-16 ***
Date(S)      0.0772       0.0013   59.385 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

Residual standard error: 98.68257

Multiple R-squared: 0.8948

Standardized Residuals:

	Min	Q1	Med	Q3	Max
	-2.803412e+02	-2.134766e+00	-2.736030e-04	3.681998e-04	4.849814e+02

Regime 3

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	121.4889	0.0008	151861.125	< 2.2e-16 ***
Date(S)	0.0502	0.0013	38.615	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 98.68257

Multiple R-squared: 0.92

Standardized Residuals:

	Min	Q1	Med	Q3	Max
	-208.87791580	-4.74304709	-0.07559093	1.23991461	197.72230992

Transition probabilities:

	Regime 1	Regime 2	Regime 3
Regime 1	9.955638e-01	1.712038e-08	0.01485319
Regime 2	5.173007e-09	9.717252e-01	0.02020467
Regime 3	4.436244e-03	2.827476e-02	0.96494213

You will realize, that it will give us a different equation for each state.

There you see now, that none of the regimes has a negative estimate. Apparently, the price of gold has been increasing through all three states it would go through. The effect size is the highest in State 2, therefore this states will probably represent an extreme growth. State 3 has a more moderate effect size, therefore it makes sense to name state 3 moderate growth. Finally, state 1 has the lowest effect size, so I would suggest to name it slow growth. In your further analysis, it might be interesting to include further independent variables to see for instance, how much they were the driver behind the growth in each phase.

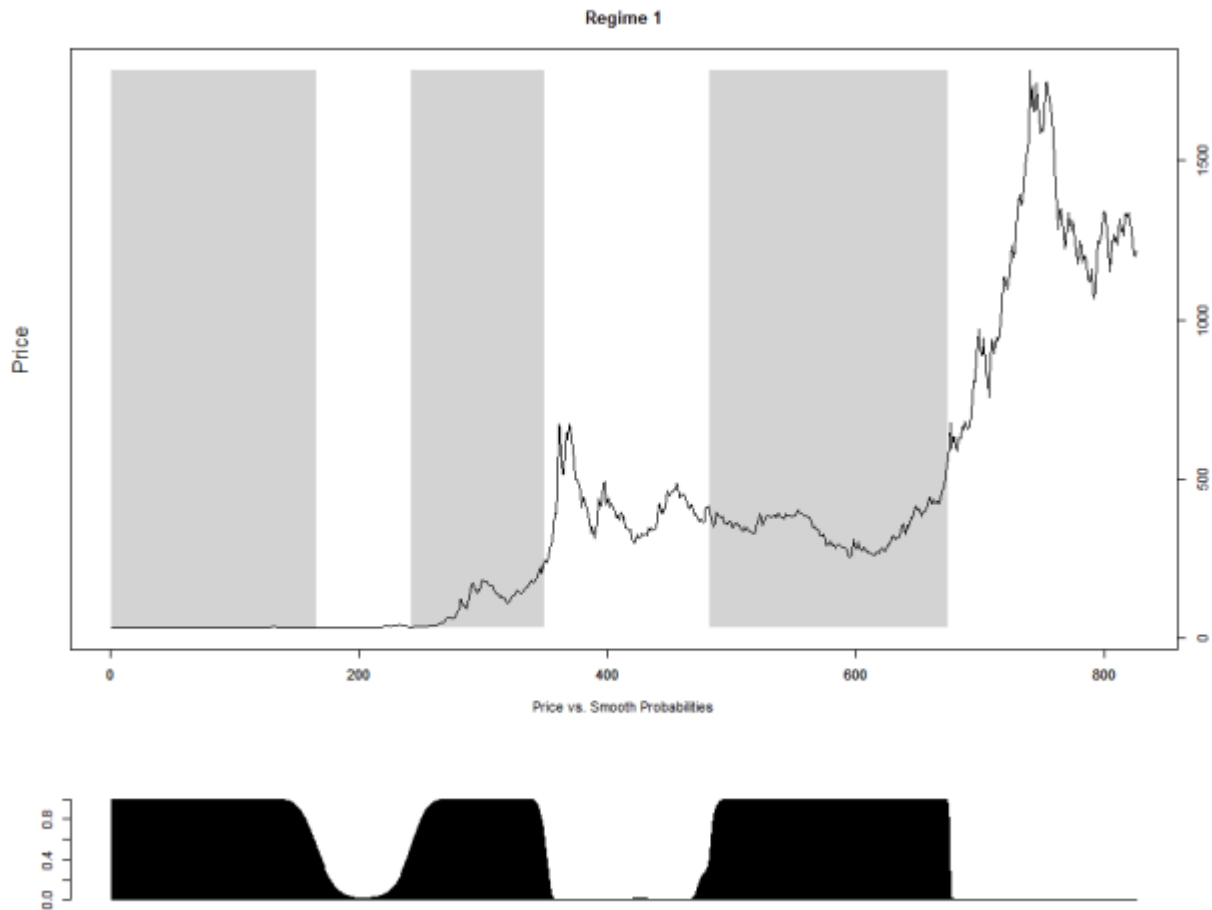
Another thing we can look at are the transition probabilities, which are summarized at the very bottom of the output.

What you will see is that the states are pretty stable, which means that the underlying states change rarely over a period of a month. Furthermore, you will see that the transition probabilities for switching to the first “moderate growth” state are generally higher than for any other. Of course you can go in greater depth in your analysis, but I will leave that to you.

Looking at the graphs

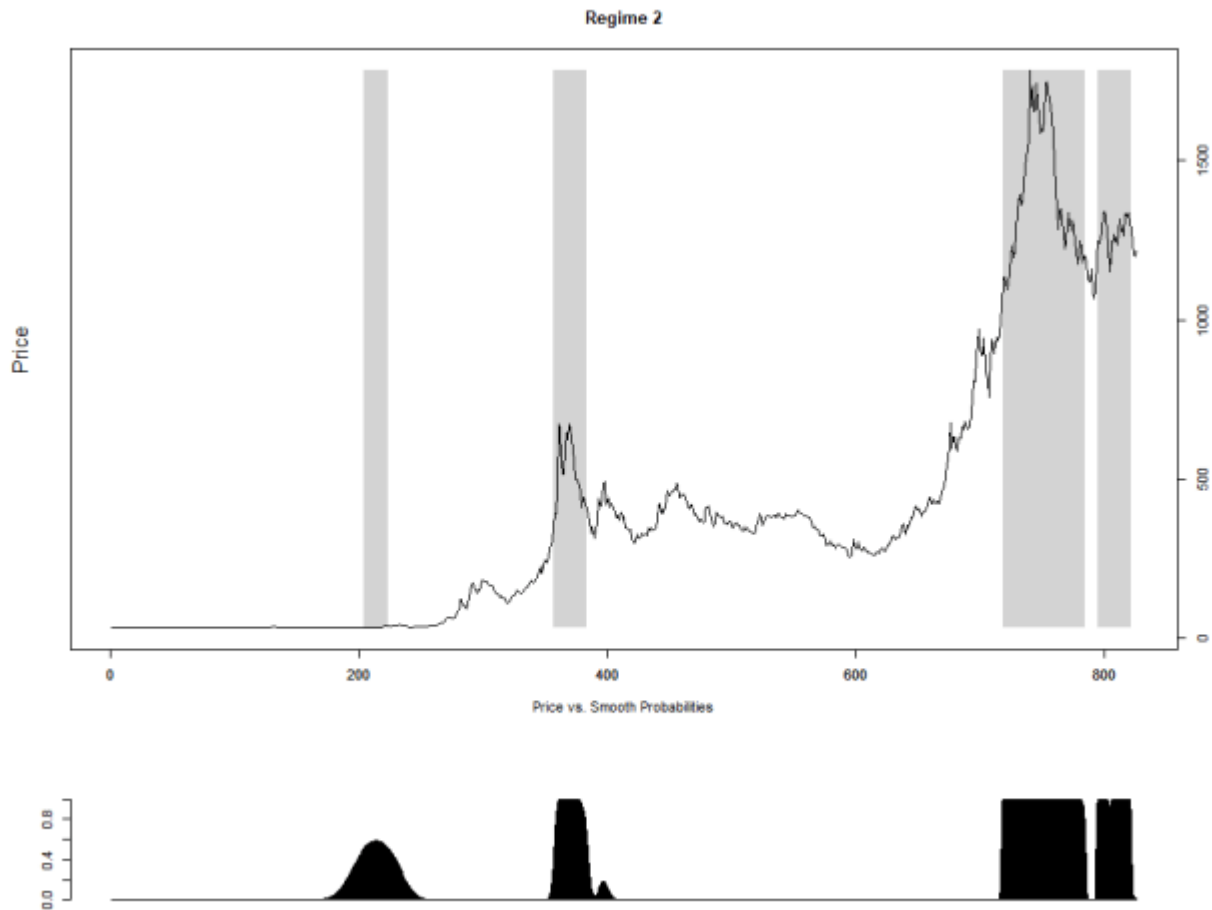
I will use the following code to produce the relevant graphs. I will have one graph for each state. You will see that each graph consists of two figures. The upper one displays the gold time series and grey-highlighted areas. The grey-highlighted areas are where the switching regression model estimated that the time-series was in the respective state. The lower figure displays the probability that the time-series was in the respective state for any point in time.

```
# Graphical Overview of Probability and predictions
plotProb(msmGold, which=2)
plotProb(msmGold, which=3)
plotProb(msmGold, which=4)
```



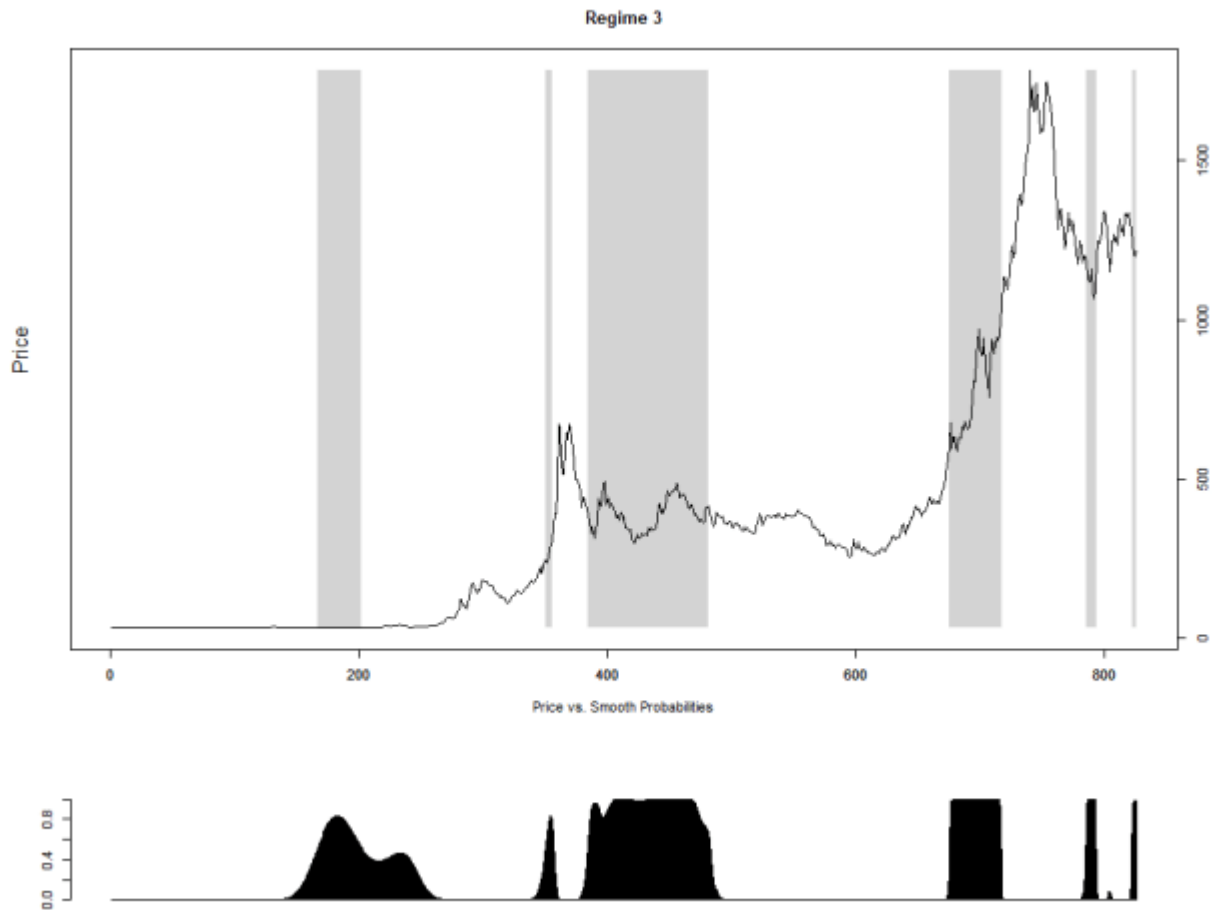
Predicted probabilities and states for regime 1

When we are looking at the upper figure, we can see that this state most likely describes the one of slow growth. The probabilities also seem to be very clear with little chance for misinterpretation.



Predicted probabilities and states for regime 2

The second regime apparently is the high-growth one or the one with the highest volatility, as the gold price increases rapidly, peaks, and then it falls down to a little higher price than it was before it started to soar. Only the increase around 200 has a lower probability as it looks as this part does not necessarily fit that well into this state.



Predicted probabilities and states for regime 3

Finally, the third state seems to be the one of moderate growth. Also here the probabilities are not that clear for the one cluster around 200. Regardless of that, it looks relatively reasonable. I will not dug deeper into the interpretation here as well. I will leave that you.

Advantages

Switching-regression models have a few advantages compared to other regression models. Here is a short overview.

- **Mathematical simplicity:** Switching regressions are generally mathematically tractable. This is especially true for the likelihood, which is computed relatively

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out www.economalytics.com for more!

straightforward.

- **Hidden variables:** Switching regressions are great in discovering hidden variables that might have an impact on your time series, but you have not observed it. One classical example where it does a good job comes from clinical psychology. Imagine you follow a person with a bipolarity disorder and you record basically the mood as dependent variable over a longer period of time. The problem however, is that you cannot observe when the person is in a maniac state and when the person is in a depressive state, so you do not record it as a variable. If you simply compute a linear regression with time as predictor, you will probably get an insignificant coefficient near zero, because the mood on average will show you neither an increasing nor decreasing trend. If you use a switching-regression, it will ideally detect both states and show you how the mood develops over time, while also giving you the matrix with the transition probabilities.
- **Flexibility:** One straightforward advantage is that it is very flexible, as we can use several different version of the same regression.
- **Rich Statistics:** Another advantage is that it offers us a rich statistics to interpret including: the transition probabilities, a separate equation for each state, and the probabilities for each state at a given point in time. Furthermore, you can interpret it using the numbers or the graphs.

Disadvantages

- **Requirements:** The transition between the states has to follow a Markov process, otherwise the switching regression will not produce unbiased estimators. The problem is that not all transitions necessary fulfill the requirements for a Markov process.
- **Number of states:** The number of states is not calculated by the model, but you have to specify before you run the estimation of the model. It is always difficult to say, what the right number of states is and it sometimes makes interpretation difficult if there are truly less or even more states than you have specified.
- **Computationally complex:** One clear disadvantage of the switching regression is that it is relatively computationally complex. That means with increasing and amount of data and especially with an increasing amount of states, your device's computation capacities will hit the limit fast.

Further Links

If you are still interested into the topic, I can recommend you the following readings to dive deeper into the topic:

- https://rpubs.com/ibn_abdullah/markovs

References

MSwM exmaples – Jose A. Sanchez-Espigares, Alberto Lopez-Moreno, Dept. of Statistics and Operations Research

Hamilton, J. D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357-384.

1993. Estimation, inference and forecasting of time series subject to changes in regime. In *Handbook of Statistics 11: Econometrics*, ed. G. S. Maddala, C. R. Rao, and H. D. Vinod, 231-260. San Diego, CA: Elseiver.

Kim, C.-J. 1994. Dynamic linear models with Markov-switching. *Journal of Econometrics* 60: 1-22.

1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press. (Chapter 22)

Share this:

- [Click to print \(Opens in new window\)](#)
- [Click to share on Facebook \(Opens in new window\)](#)
- [Click to share on LinkedIn \(Opens in new window\)](#)
- [Click to share on Twitter \(Opens in new window\)](#)
- [Click to share on WhatsApp \(Opens in new window\)](#)