

Forecasting Future Amazon Web Services Pricing

Andrew Kicinski, Hassan Souiri

Abstract

The National Reconnaissance Office (NRO) Cost and Acquisition Assessment Group (CAAG) produces independent cost estimates to support decision making, budgeting and trade studies. Cloud service costs procured from Amazon Web Services (AWS) are becoming increasingly scrutinized. In order to better inform leaders with accurate cloud cost estimates, a thorough analysis was conducted to collect historical AWS prices and model the downward trend. Autoregressive time series models were fit to Simple Storage Service (S3) and Elastic Compute Cloud (EC2) prices and were used to develop annual price reduction rates. Results include rates of 14.9% and 8.2% which can be applied to the AWS US-East-NoVa region for S3 and EC2, respectively; and rates of 11.6% and 6.4% which can be applied to the AWS GovCloud region for S3 and EC2, respectively. Applying defensible factors to more accurately model cloud service costs over time will aid in producing more realistic cost estimates, allowing better decision making and budgeting.

Introduction

AWS has been a leading provider of cloud services since the initial release of S3 in 2006. Today AWS offers a comprehensive set of cloud services, commands 35% of worldwide market share and has an annual revenue run rate of approximately \$20B. AWS prides itself on continually adapting their services to meet customer's needs and frequently adjusting service prices as favorable economies of scale and market conditions are met. By lowering their prices, AWS passes on significant savings to their consumers.

Many established organizations are migrating their Information Technology (IT) infrastructure to the cloud while new organizations are being stood up with cloud native infrastructure. Cloud advocates boast the many benefits over a traditional on-premises IT infrastructure including lower costs and more flexibility, agility and responsiveness. Since any public or private sector organization has some obligation to estimate, manage, monitor or otherwise control its costs, organizations have an incentive to understand the trends of cloud prices and how this history impacts future business decisions.

The NRO finds itself in exactly this position. Since 2014 the NRO has been utilizing cloud services from AWS in a region exclusively created for the Intelligence Community (IC). As usage grows and cloud costs increase it is becoming increasingly important to develop accurate estimates of future cloud costs. The NRO CAAG is responsible for providing independent cost estimates to allow for trade studies, analysis of alternatives, budget formulation and to support senior decision makers. CAAG estimates often extend a decade or more into the future.

Therefore it has become necessary to conduct a thorough study to collect historical AWS pricing data, analyze the trends and develop annual price reduction factors. The analysis and recommendations are described in the following sections.

Why Reduce Prices?

From 2006 to mid-2018 AWS has provided 66 price reductions. It may initially seem illogical that AWS would consistently reduce their prices. Most goods and services experience positive price inflation (including Amazon Prime Membership which increased 20% in June 2018). Rates of inflation vary with the

economy, but one can almost be certain that over any decade period or longer, labor rates, housing prices and cost for a gallon of milk will all increase. Cloud services behave differently for several reasons.

First, computer technology is a fast paced and ever changing industry. Storage costs have decreased significantly over the past decades. Today's USB and external hard drives offer orders of magnitude greater storage than previous generation's CDs and floppy disks. Additionally, personal computer prices remain relatively stationary even as performance and capability increase. As AWS is able to procure cheaper hardware they often return a portion of these savings on to their consumers.

Second, although AWS commands a significant portion of the cloud market, there are several competitors who could threaten AWS' dominance, including Microsoft, IBM and Google. In strong competition, firms must compete for business by offering competitive prices. AWS may lose customers if they cannot maintain similar prices to their competitors. Therefore, it is not uncommon to see an AWS price reduction shortly after a competitor introduces a cloud service price reduction.

Finally, given the massive market owned by AWS, the cloud provider benefits from economies of scale. As their business grows, fixed costs such as facilities and security can be distributed over a larger customer base thereby decreasing the cost per service unit. These savings occasionally get handed down to the consumer in the form of AWS price decreases.

AWS offers price reductions to their cloud services due to cost efficiencies in technology, pressure from competition and realized economies of scale. In order to understand the magnitude and frequency of AWS price reductions, a thorough data collection effort was undertaken.

Data Collection

Given the dozens of service offerings AWS provides, it would be a massive undertaking to evaluate the historical price behavior of each service in commercial AWS regions. To limit the scope of this analysis to a reasonable level that will yield actionable results, the EC2 and S3 services in the Northern Virginia region were selected for analysis. For many users Amazon's compute service, EC2, will account for a large majority of AWS costs. In addition to compute resources, another major cloud service offering is storage. S3 was selected for inclusion due to its popularity and relevance as well as its simplicity in service options and pricing, relative ease to estimate storage sizes and anticipated growth in many organization's storage requirements. The Northern Virginia region has a long AWS price history and is located near the Washington D.C. metropolitan region, making it a good analogy for many government customers.

Price data through 2015, including current and previous generation services, was collected through Amazon's AWS website. In order to find data prior to 2015 an archive website was utilized, providing EC2 data through 2010 and S3 data through 2006. The AWS blog provided additional insight into when price reductions occurred which helped guide the search of the archived pages. In addition to service pricing and dates of price reductions, EC2 data was compiled by operating system, term type, instance type and size.

The raw pricing data collected from the AWS website provided over 30,000 service prices. EC2 data was consolidated and narrowed down to 450 EC2 instance combinations by limiting the geographical region to Northern Virginia and limiting operating system types and billing types to those available to NRO users. In order to further reduce the dataset, two of the most popular EC2 product families were selected for analysis. General purpose instances (m3, m4, m5) and memory optimized instances (r3, r4) typically account for more usage than other specialized instance types.

Other EC2 price discriminators include operating system, term type and size. Linux and Red Hat Enterprise Linux (RHEL) were the focus of this study. Windows EC2 instances were assumed to be used in only a minority of cases and were therefore excluded from the analysis. The choice of purchasing an On Demand or Reserved instance is determined by the use case of each EC2 user. Given the legitimate reasons for using either billing type, both were included in this analysis. Finally, all EC2 instances are offered in multiple sizes. While price data for all sizes was collected, size was normalized by conducting analysis at a consistent size for all instances. Because instance sizes are comparable (e.g. m5.xlarge and r4.xlarge both have 4 vCPUs), scale (e.g. m5.xlarge costs twice m5.large) and price reductions are applied consistently (e.g. 10% reduction to m4.xlarge and 10% reduction to m4.large), the size selected does not impact the results of the analysis.

Once EC2 and S3 pricing data was collected and organized at the desired level, an initial plot of price vs. time showed significant price reductions over AWS history.

Pricing Trends

Before fitting models to the data, price trends were observed by plotting the data. For EC2, in order to simultaneously analyze On Demand and Reserved prices, all prices were normalized to a unit scale, and therefore began with an initial price of 1 unit. A value of .9 would indicate a 10% price reduction over the associated period of time. For this study On Demand and Reserved instances are assumed to be utilized approximately equally and Linux and RHEL instances are assumed approximately equal as well. Therefore there are four combinations of instances considered – Linux On Demand, Linux Reserved, RHEL On Demand and RHEL Reserved. The average of each normalized price was calculated for each month and trending was observed at the aggregate level.

Finally, rather than take the average of the five EC2 instances considered in the analysis, a phase in profile was applied for the two product families under consideration. For example, m4 was introduced in March 2014. Taking the average price of m3 and m4 instances would suggest that 50% of AWS users immediately use m4 while also indicating that users will always use m3 as much as m4. **Figure 1** shows historical usage of m1 vs. m3 instances for consumers whose data was available to the study. As expected there is a period of time where m3 is phased in and replaces m1 instances. An exponential curve is fit to this data to model the transition between previous and new generation instances. Investigating the transition between m3 and m4 instances produces a similar trend but over fewer months. Therefore the m1 vs. m3 transition was chosen as the representative curve for all new instance types. Applying the phase in profile returns a weighted average price that more accurately reflects the effective price customers may experience during introduction of new instances.

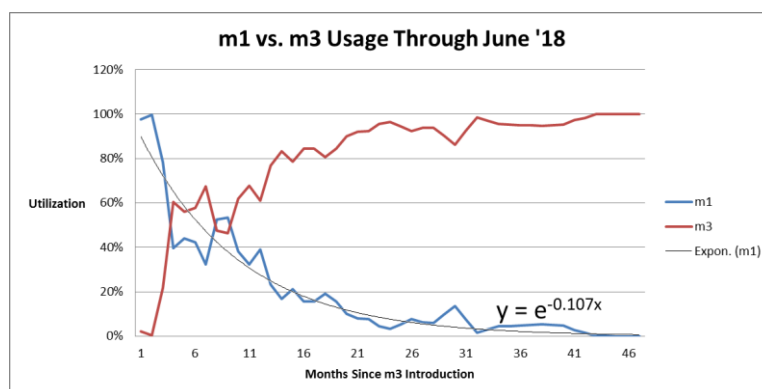


Figure 1

By applying the phase in profile to new generation instances and taking the average of On Demand and Reserved instances and Linux and RHEL instances, the price curve in **Figure 2** is obtained. Over the past five years, the analysis shows a 44% total price reduction, equivalent to an annualized rate of 10.8%. Large drops in price are due to a price reduction of a current instance, while downward sloping curves represent migration from previous generation instances to cheaper, new generation instances. For a period of time there was a small price increase for some instances before being lowered back to the original price. The plot illustrates long term sustained EC2 price reductions.

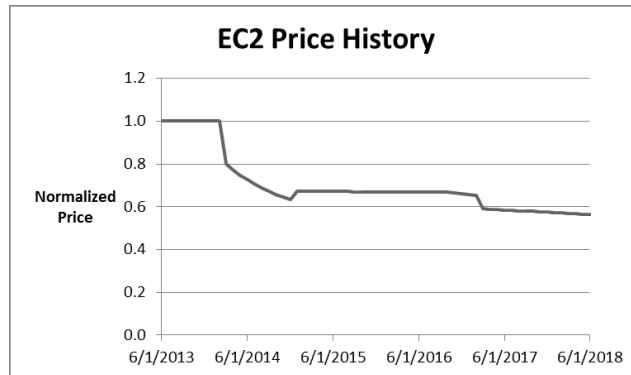


Figure 2

The analysis of S3 pricing data was significantly more straightforward than EC2, but did require making a few assumptions. The pricing structure for S3 is a simple \$/Gigabyte (GB) monthly rate applied to a tiered pricing structure. For example, today a customer’s first 50 Terabytes (TBs) are charged at \$.023/GB per month; their next 450 TBs are charged at \$.022/GB per month; and any storage above 500 TBs is charged at \$.021/GB per month. **Figure 3** shows the S3 price history by effective date and storage tier. The tier structure has changed over time beginning with one tier in 2006, reaching six tiers in 2010 and compressing to three tiers in 2016.

Storage Tier	\$/GB by Effective Date								
	3/14/2006	11/1/2008	1/1/2010	11/1/2010	2/1/2012	12/1/2012	2/1/2014	4/1/2014	12/1/2016
0-1 TB	0.15	0.15	0.15	0.14	0.125	0.095	0.085	0.03	0.023
1-50 TB	0.15	0.15	0.15	0.125	0.11	0.08	0.075	0.0295	0.023
50-100 TB	0.15	0.14	0.14	0.11	0.095	0.07	0.06	0.029	0.022
100-500 TB	0.15	0.13	0.13	0.11	0.095	0.07	0.06	0.029	0.022
500-1000 TB	0.15	0.12	0.105	0.095	0.09	0.065	0.055	0.0285	0.021
1000-5000 TB	0.15	0.12	0.08	0.08	0.08	0.06	0.051	0.028	0.021
5000 TB - Inf	0.15	0.12	0.055	0.055	0.055	0.055	0.043	0.0275	0.021

Storage (TB)	Effective Price (\$ per GB/Month)								
7,110	0.1500	0.1209	0.0781	0.0759	0.0744	0.0596	0.0496	0.0280	0.0211

Figure 3

To conduct meaningful analysis an assumed storage level must be determined. In other words, analysis should be conducted at an effective rate rather than at each tiered rate to allow for simpler analysis and more impactful results. Given this approach, assuming 500 TBs results in a different price trend than assuming 1 Petabyte (PB) of data. Internal analysis determined that an appropriate assumed level of

storage is 7.11 PB. The resulting effective prices throughout AWS history for 7.11 PB of storage are shown at the bottom of Figure 3 with the formula for effective price provided below.

$$Effective\ price = \frac{\sum_{i=1}^n \max\{\min\{g, u_i\} - l_i, 0\} * p_i}{g}$$

n = number of storage tiers with unique pricing
 g = assumed storage in GB
 u_i = storage upper limit in GB for storage tier i
 l_i = storage lower limit in GB for storage tier i
 p_i = price per GB for storage tier i

The effective prices are plotted in **Figure 4**. Over a 12 year history, AWS has reduced their storage service costs eight times totaling 86% cumulatively, equivalent to 14.9% annually - quite a significant reduction over time. Additionally, there has never been a storage price increase.

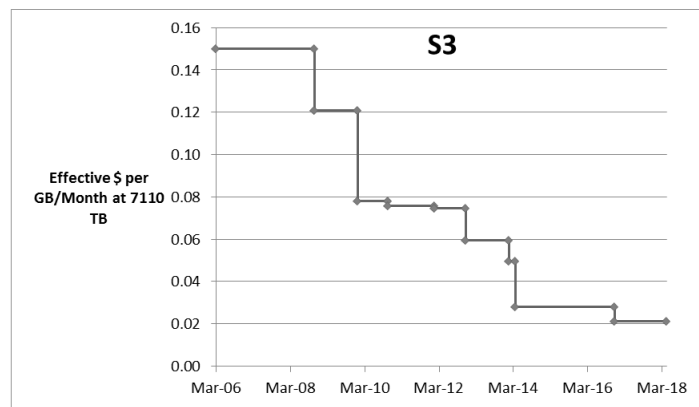


Figure 4

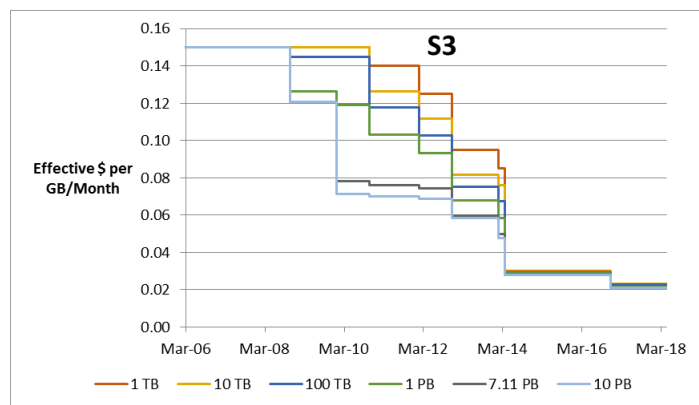


Figure 5

Figure 5 shows the price history for S3 based on varying levels of assumed storage, ranging from 1 TB to 10 PB. As can be seen from either Figure 3 or Figure 5, when initially introduced in 2006 the price for all levels of storage was \$0.15/GB. Similarly, today the price is approximately \$.022/GB with only a little variability depending on storage size. Most notably are the differences in effective prices from 2010 through 2012. The models discussed in the subsequent section are dependent on the behavior of the curve throughout the entire time period under consideration. Therefore, results are ultimately dependent on the assumption to assess storage at 7.11 PB. Organizations with different levels of storage (particularly those with less than

1 PB) should use the results of this study with caution and should instead use this analysis as a reference and framework for conducting their own analysis.

AWS Pricing Model

Once pricing data was collected, assumptions were addressed and trends were plotted, multiple models were fit to the EC2 and S3 pricing data to determine the best approach for predicating future cloud service prices. Given the simplicity of S3 pricing data, an approach for S3 was determined first, and then model fitting for EC2 followed a similar methodology. First, the annualized factor presented in the previous section (14.9%) was fit against the actual price curve using a step function. While this resulted in a model that produced the correct beginning and ending price, the estimated price for years in between tended to be lower than the actual observed price. **Figure 6** shows that if we were to use this model for S3, storage prices would be underestimated in almost the entire timeframe. Given the assumed 7.11 PB of data, the actual price of storage since 2006 is \$80M. The step function estimate produces \$68M, an estimate error of 15%.

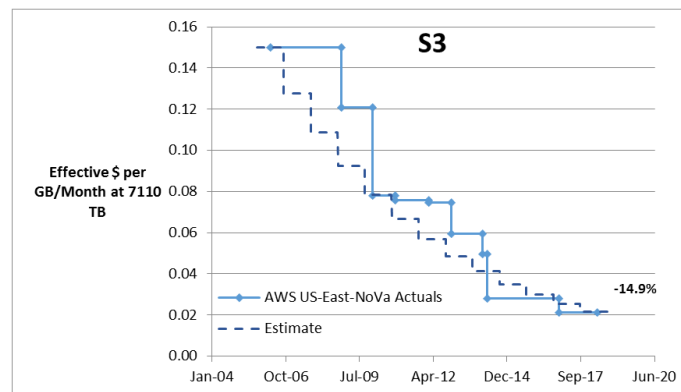


Figure 6

The next model assessed aimed to correct this underestimate bias. The annual rate of price decrease was chosen to minimize the sum squared area between the estimated and actual price functions such that the estimated price was exactly equal to the actual price over the time period. **Figure 7** shows the resulting model. The estimated step function appears much closer to the actual price function than the previous model. While the May 2018 estimate price is very close to the actual price, the estimated beginning price is 25% higher than the actual price when S3 was first introduced in March 2006. Overall the fit is reasonable and produces an annual price reduction factor of 16.1%.

A combination of the two methods represented in Figure 6 and Figure 7 was also examined. This model constrained the beginning price to \$0.15/GB (the actual initial S3 price) and minimized the sum squared area between the estimated and actual price function, constraining the total cumulative estimated and actual prices to be equal. Given the specified constraints, this model suffers from time trending bias by underestimating the beginning years and overestimating the later years. By running this model on varying lengths of S3 history, as shown in **Figure 8**, it is determined this method has a systematic time bias. It was also observed that the lengthier the history, the steeper the estimated price step down function. Furthermore, today's price is estimated to be significantly higher than actual price. Given these effects, there is little confidence this is an appropriate method for determining future S3 prices.

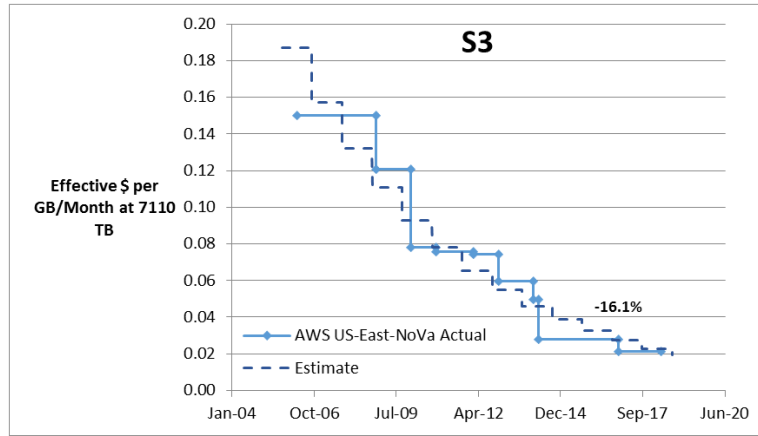


Figure 7

The choice to constrain initial estimated price to the actual price was somewhat arbitrary. The motivation was to simulate results of varying price reduction factors given a known initial condition. Given the variability of results dependent on time period evaluated, constraining the initial price may not be reasonable. Alternatively, the estimated current price could be constrained to the actual current price. Although not held to this constraint, the results in Figure 7 effectively represent this situation. This approach constrains the current price while benefiting from the entire S3 history. Therefore when using this approach to estimate future S3 prices, the current estimated price will reflect the current actual price.

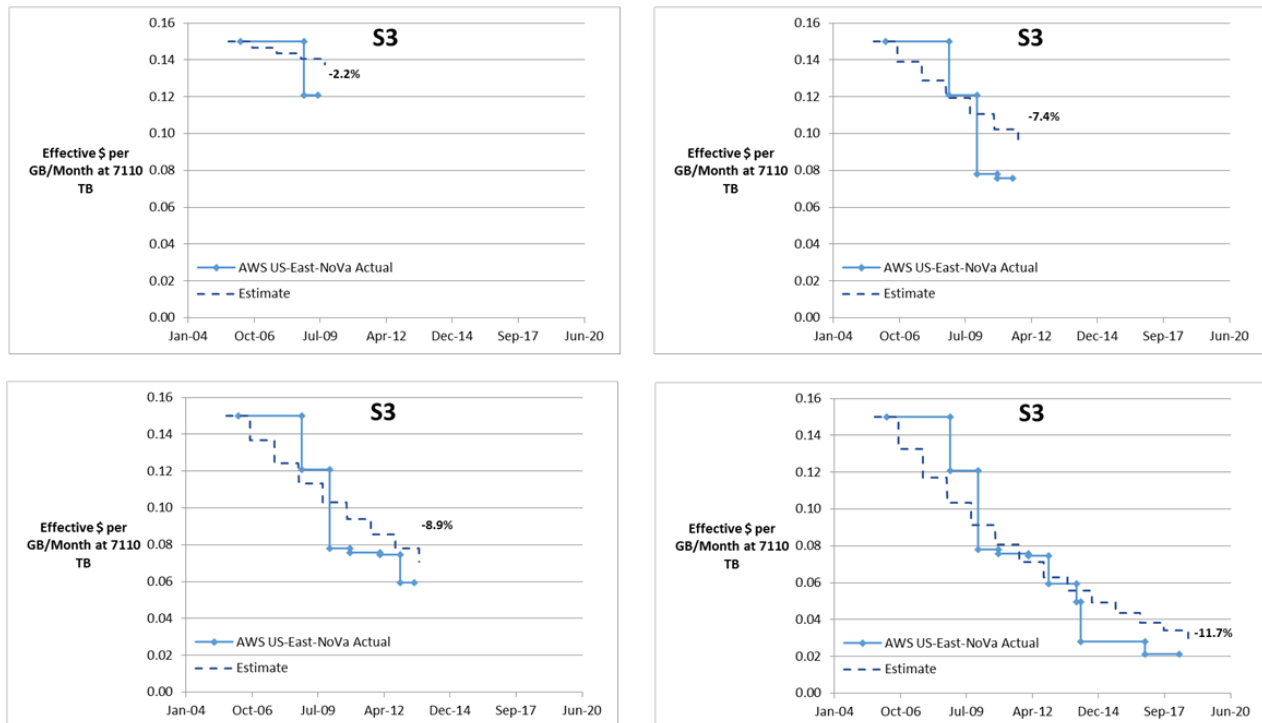


Figure 8

To investigate an alternative method where the current price is effectively constrained to its actual value, an Autoregressive (AR) time series model was considered. The AR model predicts a month's price based on a certain number of previous month's prices (referred to as order p , where p is the number of previous months considered). Multiple AR models were tested with varying orders. Weights, or parameters, for each previous month were selected by Excel Solver to minimize the sum squared area between the model and actual price curve and constrain the area to zero (i.e. the total predicted price for a given amount of storage be equal to the actual total price across the time period). The resulting model coefficients produce an exponential curve that predicts future prices. An annualized factor is then calculated from the continuous exponential curve to produce a step function.

AR(p) Model

$$z_t = c + \sum_{i=1}^p \varphi_i z_{t-i} + \varepsilon_t$$

z_t = price at time t

c = constant, set equal to zero for our models

p = order of model; number of previous months considered

φ = parameters or weights of the model

ε = error term

Note: the authors recognize the use of Excel Solver is not as robust as typical time series packages and Solver may result in solutions dependent on the initial starting solution. The results of this analysis will reasonably answer the proposed problem. Future work may investigate use of more rigorous statistical packages to fit additional time series models.

Time series models are typically fit to data that sporadically increase and decrease, such as stock market indices, home prices or the unemployment rate. Given the monotonic behavior of S3 prices and near-monotonic behavior of EC2 prices, this price analysis differs from other time series models. The AR models considered were set to also be monotonic by constraining all model parameters to be nonnegative. Additionally, the prices under consideration behave as step functions; prices are stable for a period of months before decreasing by a certain amount. The AR model will need to have an order large enough to understand these effects. For example, an AR(1) model only considers the previous month to estimate the current month. Given the price step function, this model would, by definition, not reach back far enough to gather data from more than one unique price. **Figure 9** shows the number of months between price reductions and the number of months between two consecutive price reductions. There was one price reduction that was followed two months later by another price reduction. Otherwise, all successive price reductions were separated by at least ten months. These observations guide the order selection for AR model evaluation. Any AR(p) model where p is less than ten will consider at most two prices, except immediately following the period when there were two price reductions within two months. Therefore we consider AR(p) models with p no less than ten. AR(10), AR(14) and AR(15) seem like logical choices given the distances between price reductions.

Months Between Price Reductions	32	14	10	15	10	14	2	32
Months Between Consecutive Price Reductions	46	24	25	25	24	16	34	

Figure 9

First, consider the AR(10) model. Given the constraints and model objective described above, the model parameters, statistics and results are shown in **Figure 10**. The time series model fits the actual cost function very closely, as expected due to the stable, non-increasing behavior of the cost function. Projecting the model forward yields an exponential curve which can be converted to a step function, producing an annual price reduction factor of 14.3%. Reverse fitting the exponential curve appears to fit the general shape of the data, but generally underestimates actual cost. The model parameters table shows the effective weights given to preceding months. Of the ten months included in the model, only four have nonzero parameters. Therefore, this AR(10) model is defined as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_9 z_{t-9} + \phi_{10} z_{t-10} + \varepsilon_t$$

$$z_t = .9501z_{t-1} + .0092z_{t-2} + .0229z_{t-9} + .0021z_{t-10} + \varepsilon_t$$

It may initially seem that the presence of so many zero parameters indicates a flaw in Excel Solver’s ability to find an optimal solution, but recall the various numbers of months between price reductions presented in Figure 9. It is not a coincidence that the nonzero model parameters are near those which are the number of months between price reductions (i.e. 2 and 10). These have been rightfully identified as the most impactful intervals to consider. To verify Solver has not produced a bias solution, the model was run multiple times with varying starting parameters and results did not significantly differ from those provided in Figure 10.

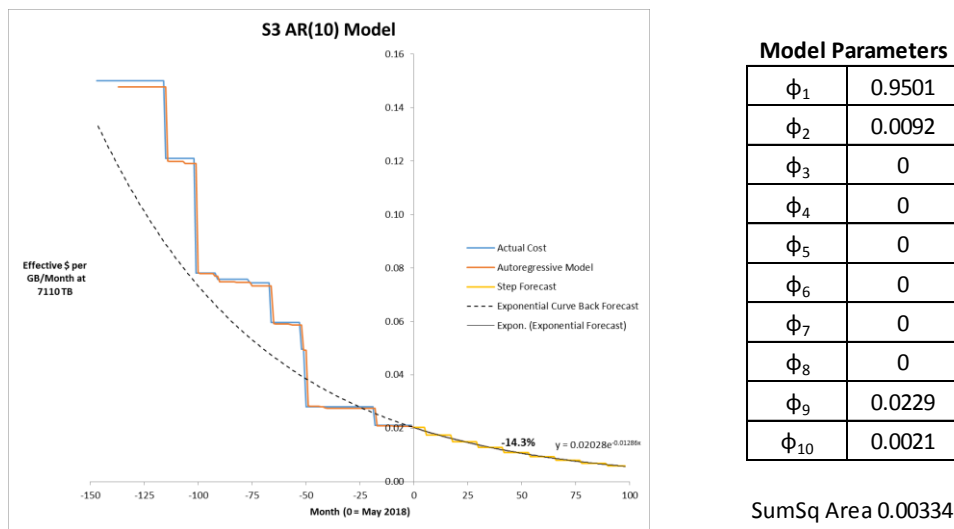
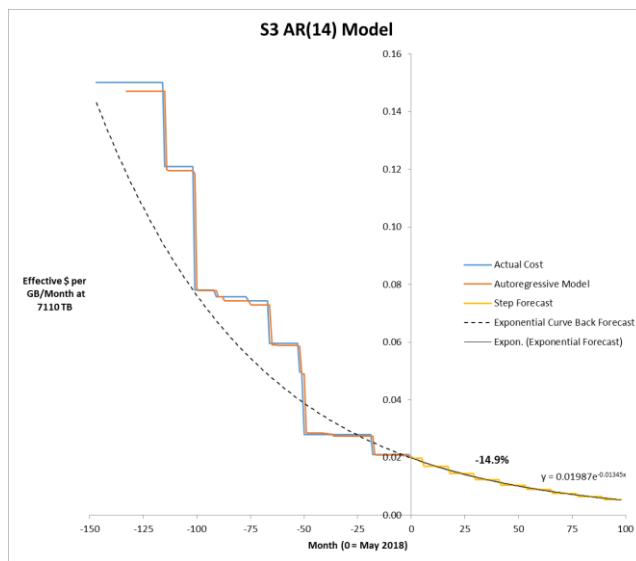


Figure 10

Next, an AR(14) model was considered using the same constraints and objectives as the AR(10) model. The model parameters, statistics and results are shown in **Figure 11**. The model looks very similar to its predecessor. Differences include a slightly steeper predicted exponential cost function, returning a predicted 14.9% annual price reduction in the future. The back forecasted exponential curve is slightly higher than in the AR(10) model and represents a closer fit to the actual historical costs. The AR(14) model parameters does not include a positive ϕ_9 but does include a positive ϕ_{14} . As discussed, this should not be surprising given multiple occurrences of price reductions being spaced 14 months apart. The sum squared area between the estimated cost function and actual cost function is slightly less than that of the AR(10) model. Given this reduced error, the fact that the model includes a longer price history covering more price reductions at some points and the back forecast curve aligning closer to the data, the AR(14) model is preferred over the AR(10) model.



Model Parameters

ϕ_1	0.9373
ϕ_2	0.0092
ϕ_3	0
ϕ_4	0
ϕ_5	0
ϕ_6	0
ϕ_7	0
ϕ_8	0
ϕ_9	0
ϕ_{10}	0.0021
ϕ_{11}	0
ϕ_{12}	0
ϕ_{13}	0
ϕ_{14}	0.0317

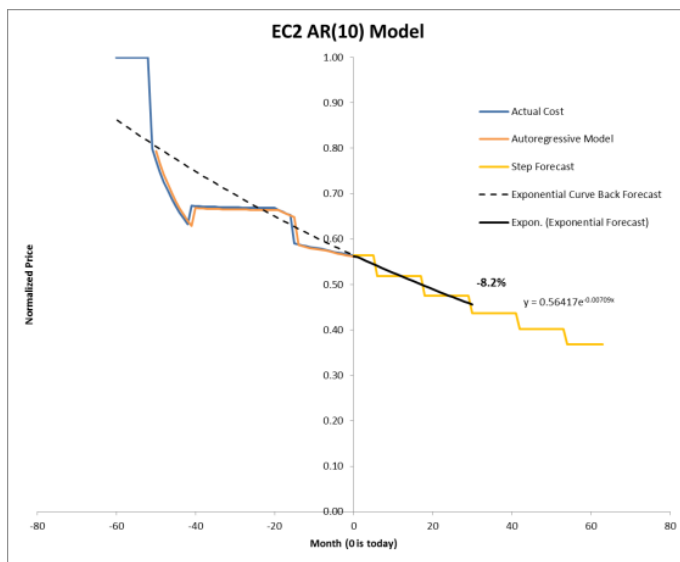
SumSq Area 0.00330

Figure 11

Finally, an AR(15) model was considered, but returned a zero value for the ϕ_{15} parameter. Therefore, the AR(15) model is essentially equivalent to the AR(14) model. Of all the S3 models tested, results are fairly consistent producing annualized reduction factors between 11.7% and 16.1%. The AR(14) model is selected as the best method to estimate future S3 prices at an annual price decrease of 14.9%.

Analysis of EC2 price history followed a similar process. Given the success of the AR model for S3, this approach was also taken for EC2. Various AR(p) models were tested and were assessed using similar logic to evaluation of the S3 AR models. Shown in **Figure 12**, the AR(10) model was the best fit. Given the shorter history of price reductions evaluated for EC2, the single price increase and the integration of EC2 instance transition into the model, AR(p) models varied significantly based on their order. Although less stable than the S3 models, this allowed for easy selection of the best model fit. The AR(10) model has a back forecast curve that goes right through the center of the historical pricing. Interestingly, the model parameters ϕ_2 through ϕ_{10} have practically negligible values while ϕ_1 is nearly 1. This effect is due to the integration of an EC2 new generation instance transition period into the historical price data. Notice that although the actual price function has major steps, there are periods of continuous, downward slope. This feature lends itself to an AR model with a single parameter near but less than one.

Given the results of the AR(10) model it is reasonable to consider an AR(1) model. If the AR(10) model returned nine consecutive model parameters near zero, then maybe it would be best to eliminate these parameters completely. The AR(1) model was tested but produced a very steep curve which did not accurately reflect historical EC2 pricing. Therefore, the AR(10) model remained the recommend model for EC2 historical costs, forecasting future annual price reductions of 8.2%.



Model Parameters

ϕ_1	0.9929
ϕ_2	4.46E-07
ϕ_3	4.52E-07
ϕ_4	4.58E-07
ϕ_5	4.63E-07
ϕ_6	4.69E-07
ϕ_7	4.75E-07
ϕ_8	4.80E-07
ϕ_9	4.86E-07
ϕ_{10}	4.91E-07

SumSq Area 0.00758

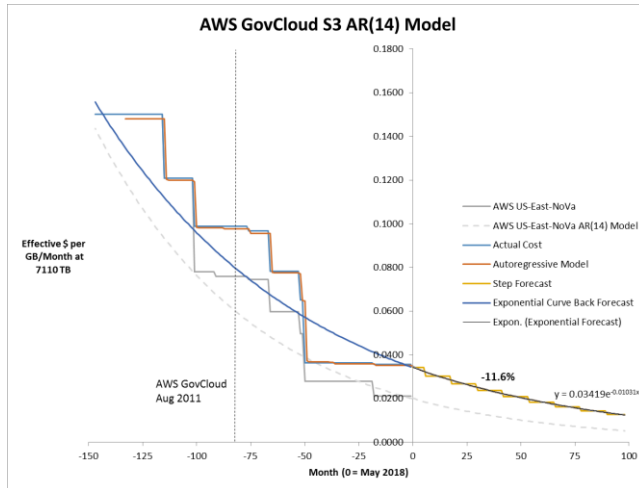
Figure 12

Application of Results

The analysis produced annual price reduction factors of 14.9% and 8.2% for S3 and EC2, respectively. It is important to remember the assumptions behind the underlying pricing data used in these models. For S3, it was assumed that 7.11 PB of data was stored. For EC2, analysis focused on Linux and RHEL general purpose and memory optimized instances. Both S3 and EC2 used pricing from the AWS US-East-NoVa region. These assumptions were made to best align to the conditions experienced by NRO. Because results are sensitive to the underlying assumptions, for cases where the assumptions deviate significantly from those proposed here, this analysis should be redone with updated pricing data.

The AWS US-East-NoVa region was selected due to its long price history and Washington D.C. metropolitan location. The NRO actually utilizes a completely segregated unique AWS region. This region is isolated to allow handling of sensitive material. Given the specialized use of this region, economies of scale are not as extensive as the commercial AWS regions. As the region grows, there will certainly be some economies of scale that will provide savings, but likely not to the extent reached in the commercial market. Furthermore, there is limited competition as AWS has the single current contract to provide cloud services for the IC. AWS must retain competitive pricing considering the contract will be re-competed at some point, but switching cloud providers would be burdensome to the government’s migration efforts. Given this construct, AWS has less incentive than in the commercial market to keep prices low. Therefore, prices should be expected to be higher on the NRO region than for commercial AWS US-East-NoVa. The AWS GovCloud is another region specifically developed for government agencies with higher security standards. The current AWS GovCloud S3 price is approximately 69% higher than the AWS US-East-NoVa price.

Given the differences between these specialized regions for specific customers and the commercial regions, we should not expect the price reductions to be identical. 14.9% annual S3 price reductions for the NRO region is likely too steep given the limited competition and smaller economies of scale. To develop a model that can be applied to the NRO region, we consider a hybrid model that leverages the long history of commercial AWS and the recent directly analogous history of the NRO region. The same can be done for AWS GovCloud. **Figure 13** represents a hybrid AR(14) model for AWS GovCloud.



Model Parameters	
ϕ_1	0.9484
ϕ_2	0.0130
ϕ_3	0
ϕ_4	2.42E-07
ϕ_5	5.27E-07
ϕ_6	8.13E-07
ϕ_7	0.0002
ϕ_8	0
ϕ_9	6.33E-07
ϕ_{10}	9.75E-08
ϕ_{11}	0
ϕ_{12}	0
ϕ_{13}	0.0011
ϕ_{14}	0.0233

SumSq Area 0.00254

Figure 13

The light grey step function in Figure 13 represents the S3 price for the US-East-NoVa region. Notice the actual price step function is higher, representing a higher price for the AWS GovCloud region. Although not introduced until August 2011, the actual price function was set to the initial GovCloud price for 19 months prior to avoid having the curve show an increase in price at the conception of GovCloud. Therefore the first 47 months represent AWS US-East-NoVa prices while the following 100 months represent AWS GovCloud prices. The hybrid model returns an annual price decrease of 11.6% which could be used to model future S3 prices for the AWS GovCloud region. This rate of decrease, while still significant, is considerably more gradual than the US-East-NoVa region rate of 14.9%. This process could be repeated for the NRO region or other standalone region with a shorter history than commercial AWS.

To verify the reasonableness of the model results, the Bureau of Labor Statistics (BLS) Producer Price Index (PPI) was consulted. The BLS releases PPIs with average monthly prices for various industries, goods and services. The BLS PPI for computer storage devices over the last 25 years indicates a -11.4% annual inflation factor, nearly identical to the AR model results for AWS GovCloud. This satisfies a reasonableness test, verifying that the AR model results are realistic.

Using a simple method to develop an annual price reduction rate for AWS GovCloud EC2, we can apply the ratio of S3 GovCloud to US-East-NoVa rates to the US-East-NoVa EC2 rate:

$$\frac{11.6\%}{14.9\%} * 8.2\% = 6.4\%$$

Therefore one could predict annual price reductions of 6.4% for AWS GovCloud EC2.

This analysis was conducted for the NRO region and the resulting rates of price decrease will be applied in future CAAG estimates. The effect of this analysis is essentially an inflationary impact. Consider a cost estimate that does not take into account inflation. Assuming all other estimate inputs are accurate, the estimate will be lower than the incurred cost because costs simply inflate over time. Similarly, if cloud price reductions are not taken into account, a cloud cost estimate will produce an overestimate of actual cost.

Figure 14 shows a simple example of a ten year estimate of \$100 per year without any price reductions,

with a 6.4% annual price reduction and with an 11.6% annual price reduction. The results are deltas of 24% and 39% from the baseline for a 6.4% annual reduction and 11.6% annual reduction, respectively. By simply not modeling cloud service price reductions, a ten year AWS GovCloud S3 estimate could be overestimated by as much as 39% and an EC2 estimate by as much as 24%.

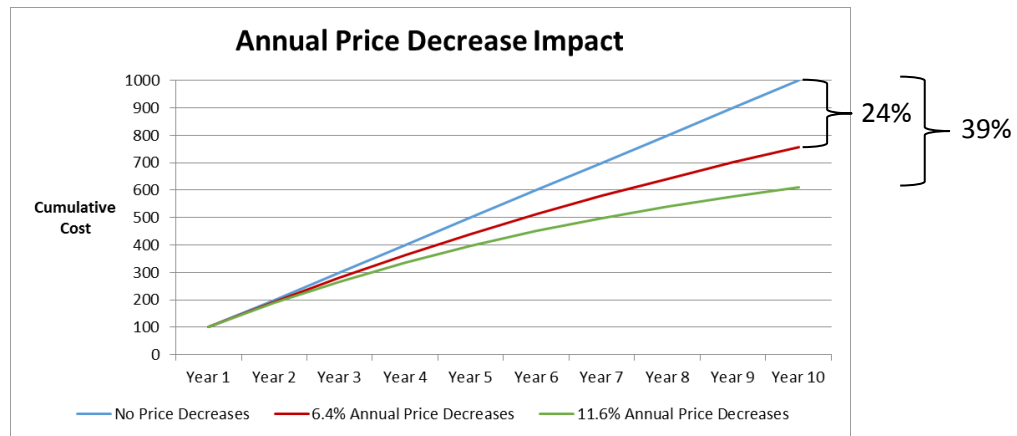


Figure 14

Applying defensible annual price reduction rates will produce more realistic cloud cost estimates and will help improve the budgeting process. By more accurately allocating budget, the NRO can make better decisions and better plan for all IT expenses.

Conclusion

After producing several NRO cloud cost estimates, the CAAG realized the need to conduct a comprehensive study on historical AWS pricing trends. After collecting over a decade of pricing data, narrowing down the dataset and developing multiple models for EC2 and S3, the CAAG has developed annual price reduction rates for compute and storage and recommends their use for all future cloud cost estimates.

Going forward, these models will be refined as more price reductions are realized or as prices stay stagnant, as both outcomes affect the model results.

Incorporating annual price reduction factors is one important step to producing a better AWS cost estimate, but alone does not ensure an accurate estimate. Other measures must be taken as well, including collection of historical cloud and on-premises hardware usage, knowledge of system description and future requirements, and phased modeling of cloud service requirements to include usage growth over time.

References:

1. AWS Pricing, <https://aws.amazon.com/pricing/>
2. AWS News Blog, <https://aws.amazon.com/blogs/aws/category/price-reduction/>
3. Internet Archive, <https://archive.org>
4. Bureau of Labor Statistics, <https://data.bls.gov/cgi-bin/dsrv>