Comprehensive Statistical Analysis and Modeling of Spot Instances in Public Cloud Environments

Bahman Javadi and Rajkumar Buyya

Cloud Computing and Distributed Systems (CLOUDS) Laboratory
Department of Computer Science and Software Engineering
The University of Melbourne, Australia
Email: {bahmanj, raj}@csse.unimelb.edu.au

Technical Report: CLOUDS-TR-2011-1

Abstract

Due to increase in demand for utilizing public Cloud resources, we are facing with many trade-offs between price, performance and recently reliability. Amazon's Spot Instances (SIs) provide a low price yet less reliable and competitive bidding option for the public Cloud users. Although some works have explored the utilization of SIs to decrease the monetary cost of Cloud computing, the characteristics of SIs have not been investigated yet. In this paper, we provide a comprehensive statistical analysis and modeling of such SIs based on one year price history in four data centers of Amazon's EC2. For this purpose, we analyze all different types of SIs in terms of spot price and the inter-price time (time between price changes). Moreover, we determine the time dynamics for spot price in hour-in-day and day-of-week. The results reveal that we are able to model spot price dynamics as well as the inter-price time of each SI by the mixture of Gaussians distribution with three or four components. The proposed models are validated through extensive simulations, which demonstrate that our models exhibit a good degree of accuracy under realistic working conditions. We believe that this characterization is fundamental in the design of stochastic scheduling algorithms and fault tolerant mechanisms in public Cloud environments for spot market.

1 Introduction

Due to increase in demand for using utility computing systems like public Cloud resources, many trade-offs between price and performance have emerged. For instance, Infrastructure-as-as-Service (IaaS) providers, offer raw computing with various capacity and storage in the form of Virtual Machines (VMs) on a pay-as-you-go basis.

Recently, another aspect, *reliability*, has been added to these trade-offs to make them more challenging than ever. In December 2009, Amazon released a new type of instances called Spot Instance (SI) to sell the idle time of Amazon's EC2 data centers [3]. The price of an SI, *spot price*, depends on the type of instance (see Table 1) as well as VM demand within each data center. The users provide a *bid* which is the maximum price to be paid for an hour of usage. Whenever the current price of an SI is equal or less than the user bid, the instance is made available to the user. If the price of an SI becomes higher than the user's bid, the VM(s) will be terminated by Amazon automatically and user does not pay for any partial hour. However, if the user terminates the running VM(s), she has to pay for the full hour. Amazon charges users per hour by the market price of the SI at the time of VM creation.

Amazon also provides on-demand and reserved VM instances, which are associated with a fixed set price [13]. However, Amazon can increase or decrease these prices based on their own local policy. There are 64 different types of instances with various capacities and prices under two operating systems which are made available by Amazon in four data centers as illustrated in Table 1 (sorted by their prices). In this Table, the prices are given for Linux operating system and the instances labeled as follows:

- m1: standard instances

Table 1. Prices of on-demand instances in different data centers of Amazon (prices given in cents).

Instances	us-west	us-east	eu-west	ap-southeast	EC2 Compute Unit	Memory (GB)	Storage (GB)
m1.small	9.5	8.5	9.5	9.5	1	1.7	160
c1.medium	19	17	19	19	5	1.7	350
m1.large	38	34	38	38	4	7.5	850
m2.xlarge	57	50	57	57	6.5	17.1	420
m1.xlarge	76	68	76	76	8	15	1690
c1.xlarge	76	68	76	76	20	7	1690
m2.2xlarge	114	100	14	114	13	34.2	850
m2.4xlarge	228	200	228	228	26	68.4	1690

- m2: high-memory instances

- c1: high-CPU instances

Spot instances are an alternative to other two classes of instances which offer a low price yet less reliable and competitive bidding option for the public Cloud users. There are a few works on how to utilize SIs to decrease the monetary cost of utility computing [15, 12]. However, thorough statistical analysis and modeling of SIs have not been investigated yet, the focus of our research in this study.

In this paper, we provide a comprehensive statistical analysis and modeling of all SIs in terms of spot price and the inter-price time (time between price changes) in four Amazon's data centers (i.e. us-west, us-east, eu-west, and ap-southeast). In particular, the main contributions of this paper are as follows:

- We provide statistical analysis for all SIs in Amazon's EC2 data centers. We also determine the time correlation in spot price in terms of hour-in-day and day-of-week.

- We model spot price and the inter-price time of each SI with the mixture of Gaussians distribution. A model calibration algorithm is also proposed to deal with an observed artifact in the real price history.

- We validate our proposed models by comparing trace and model simulation to verify the accuracy of our models under realistic working conditions.

We believe that results of this research are essential in the design of stochastic scheduling algorithms and fault tolerant mechanisms (e.g. checkpointing and replication algorithms) in public Cloud environments for spot market.

The paper is structured as follows. In Section 2, we describe the processes that we model in this paper. We discuss related work in Section 3. We examine the pattern in spot price in Section 4. In Section 5, we present the global statistics for all SIs. We then illustrate distribution fitting for spot price and the inter-price time in Section 6. In Section 7, we propose an algorithm for model calibration. We discuss the validation of the proposed models through simulation in Section 8. In Section 9, we summarize our contributions and describe future directions. Moreover, In Section 11 (Appendix) we present the results of some tests for randomness of all SIs as well as distribution fitting by several classic distributions.

2 Modeling Approach

We describe here the variables that we are going to analyze and model. As mentioned in the previous section, SIs have two variables (i.e. spot price and inter-price time) specified by the Cloud provider and, another variable (user's bid) determined by users. In this paper, we focus on the analysis and modeling of two system variables. Thus spot price and the inter-price time of each SI are the processes that we model. These two variables are illustrated in Figure 1 where P_i is the price of an SI at time t_i. So, the inter-price time is defined as $T_i = t_{i+1} - t_i$. Therefore, the time series of spot price (P_i) and the inter-price time (T_i) are analyzed and modeled in the following sections.

The traces that we use in this paper are about one year price history of all SIs from the first of February 2010 to mid-February 2011 where we include the first 10-month (Feb-2010 to Nov-2010) in the modeling process. These 10-month traces along with the last two months are used for the model validation phase. The spot price history is freely provided by Amazon per SI for each data center and also available through other third-parties like [1]. We exclude the data prior to the February
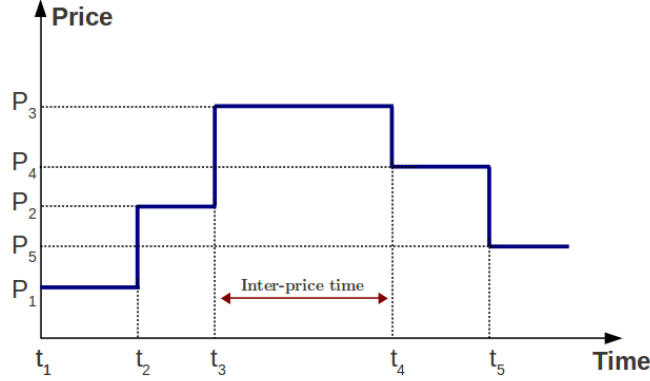
Figure 1. Spot price and the inter-price time of Spot instances.

2010 due to a bug in the pricing algorithm which is reported in [2]. Moreover, we only use the SIs with Linux operating systems from all data centers.

3 Related Work

To the best of our knowledge, this is the first work to analyze and model spot price in public cloud computing environments. However, there are some papers which considered the SIs as an alternative of on-demand and reserved instances and show how we can adopt them to decrease the monetary cost of utility computing.

Yi et. al. in [15] introduced some checkpointing mechanisms for reducing costs of SIs. They used the real price history of EC2 Spot instances, and show how the adaptive checkpointing schemes are able to decrease the monetary cost and improve the job completion times. In [4], a decision model for the optimization of performance, cost and reliability under SLA constrains is proposed. They used the real price history and workload models, to demonstrate how their proposed model can be used to bid optimally on SIs to reach different objective with desired levels of confidences.

Chohan et. al. in [6] proposed a method to utilize the SIs to speed up the MapReduce tasks. They provide a Markov Chain to predict the probability of the SI lifetime. They concluded that having a fault tolerant mechanism is essential to run MapReduce jobs on SIs. Also, in [12], they proposed a hybrid cloud architecture to lease the SIs to manage peak loads of a local cluster. They proposed some provisioning policies and investigate the utilization of SIs compared to on-demand instances in terms of monetary cost saving and number of deadline violations.

Although the existing papers show that SIs are good alternative for on-demand or reserve instances in terms of monetary cost, but the characteristics of the SIs still is not clear for users and researchers in the community. So, we conduct this research to fill this gap and provide a statistical model for SIs in public cloud systems.

4 Patterns of the Spot Price

In this section, we examine hour-in-day and day-of-week time dynamics for the prices of different SIs in all data centers. We use the same approach as [11] to show how the price of one SI changes each hour in the day or day of the week. We have the price information in GMT, so we consider all data sets where the local time is adjusted for time zones. In Figure 2, we create eight 3-hour time slots per day, and determine the average price of each SI in each time slot over all days. We then normalized this average by the maximum average price over all days. In Figure 3, we applied the same procedure except obtained the average price over seven 24-hours slots within the week.

Focusing on the plots in Figure 2, we can see that the y-axis is in the range of [0.98 1.0]. So, the prices varies in a very limited amount in each day. However, we are able to see an increasing price in the first half of each day ([0 12]) and decreasing prices in the second half of each day for all SIs in each data center. Additionally, different SIs in each data centers have the positive correlation where their prices are increasing or decreasing in the same time. This pattern is more pronounced in ap-southeast data center.

In Figure 3, the y-axis has wider range of [0.95 1.0] for all data centers except us-east which is in the range of [0.90 1.0]. As it is observable from these plots, we have more clear pattern in day of the week where in Tuesday we have the maximum

prices for almost all SIs in each data centers. Moreover, the lowest prices are on the first day of weekends and on Sunday we again observe the increasing of SIs' prices.
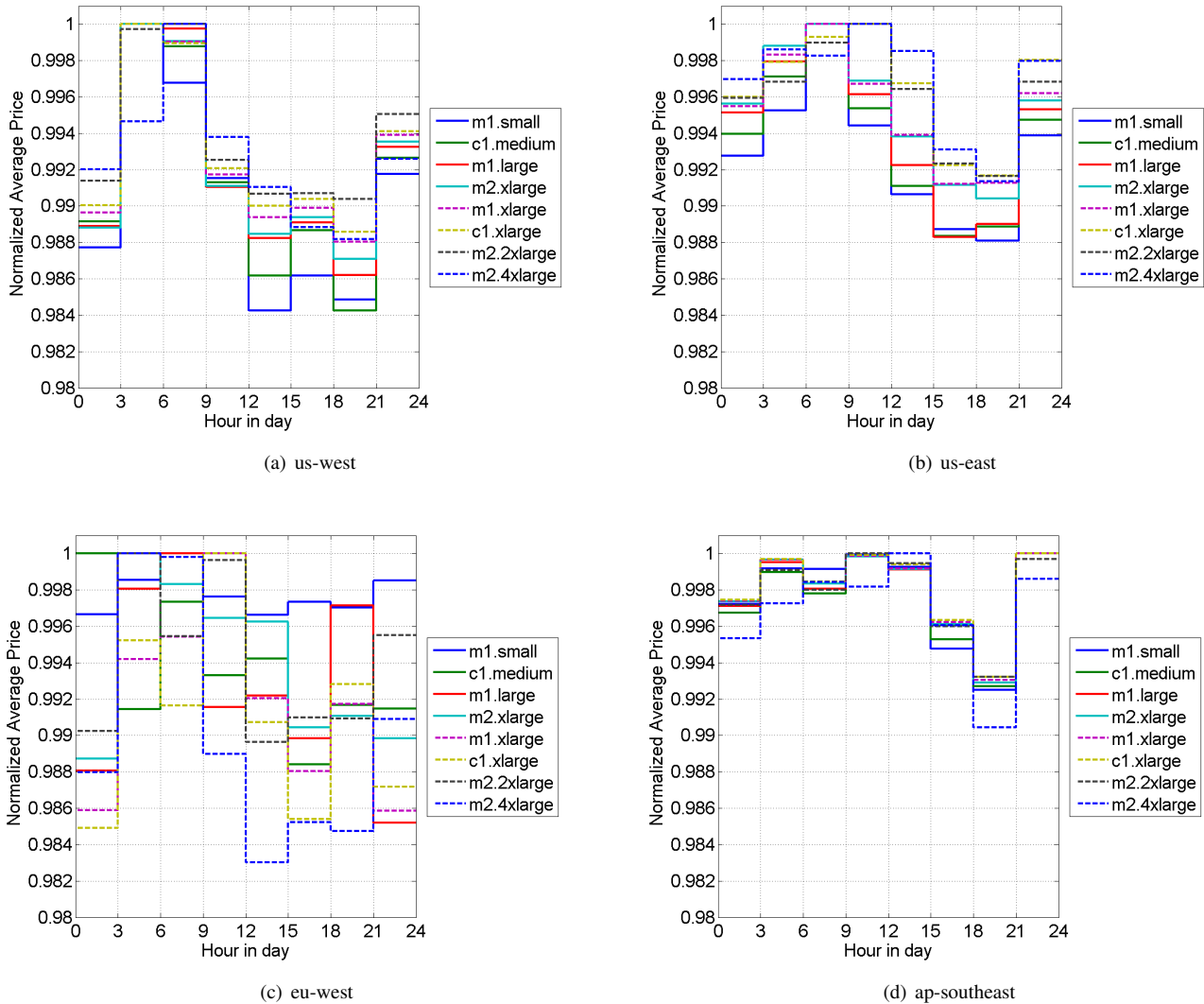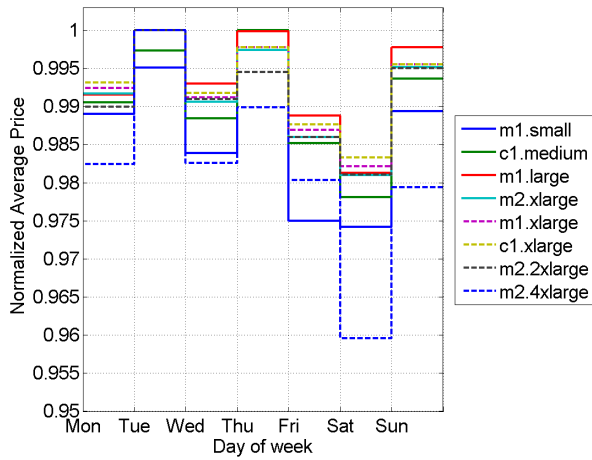

(a) us-west

(b) us-east

(c) eu-west

(d) ap-southeast

**Figure 2. spot price by time in day**

5 Global Statistics

In the following, we analyze data sets of different SIs in all four data centers[1]. It should be noted that we used the trace data from the first of February 2010 up to the end of November 2010 (10 months traces). We used the Spot price history which is freely provides by Amazon. We exclude the data prior to the February due to a bug in the pricing algorithm which is reporting in [2]. Moreover, we only used the SIs with the Linux operating systems from all data centers.

We inspect the basic statistics of the traces in terms of spot price in Table 2, 3, 4 and 5 and in terms of inter-price time in Table 6, 7, 8 and 9. The statistics in the tables are mean, trimmed mean, median, standard deviation (std), coefficient of

---

[1]We conduct all of our statistical analysis using Matlab R2010b on a 32-bit on a Core2Duo 3.00GHz desktop with 3GB of RAM. We use when possible standard tools provided by the Statistical Toolbox. Otherwise, we implement or modify statistical functions ourselves.

(a) us-west
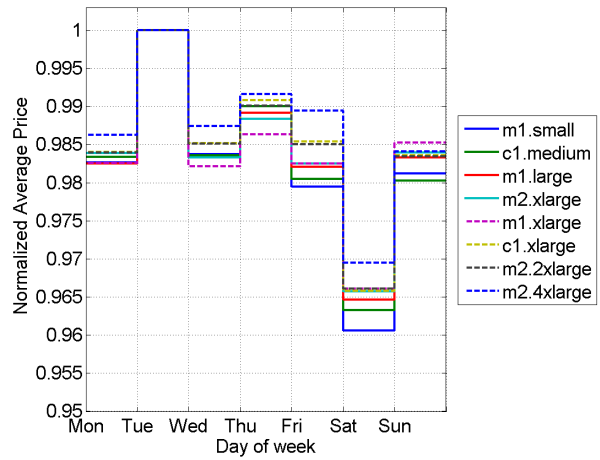
(b) us-east
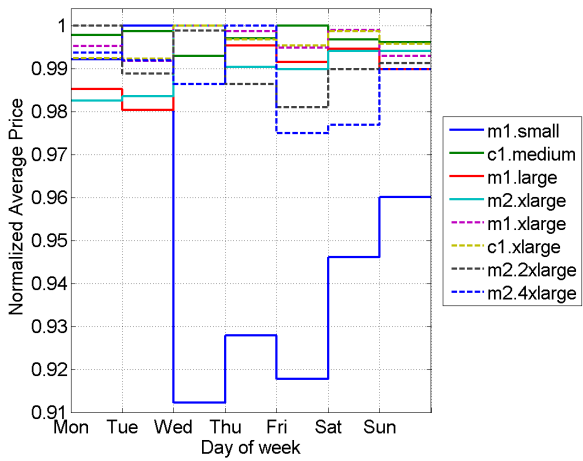
(c) eu-west
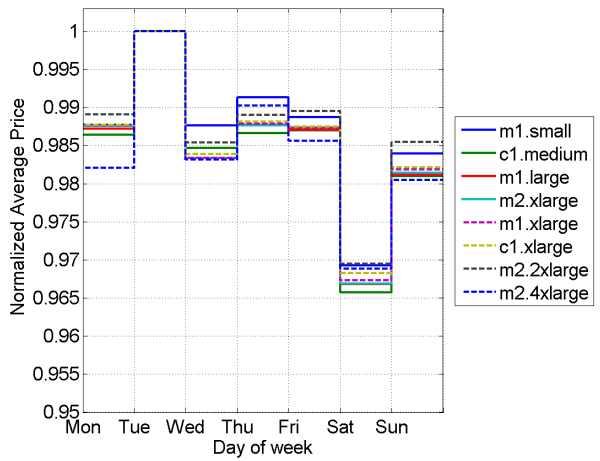
(d) ap-southeast

Figure 3. spot price by day of week

Table 2. Statistics for spot price in us-west data center (Values given in cents).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	4.00	3.99	4.00	0.22	0.06	0.20	8.50	3.80	9.70	177.62	2226
c1.medium	7.99	7.99	8.00	0.26	0.03	0.40	8.40	7.60	0.08	1.62	2376
m1.large	16.09	15.99	16.00	1.53	0.09	1.00	40.00	15.20	12.83	193.20	2489
m2.xlarge	23.98	23.97	23.90	0.79	0.03	1.40	33.00	22.80	0.73	9.05	2285
m1.xlarge	31.99	31.98	32.00	1.02	0.03	1.90	33.60	30.40	0.03	1.56	2480
c1.xlarge	32.03	31.98	32.00	1.56	0.05	1.90	75.00	30.40	12.34	293.12	2439
m2.2xlarge	55.97	55.97	55.90	1.79	0.03	3.40	58.80	53.20	0.02	1.58	2496
m2.4xlarge	111.91	111.90	111.80	3.56	0.03	6.70	125.00	106.40	0.06	1.63	2536

Table 3. Statistics for spot price in us-east data center (Values given in cents).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	3.16	3.02	3.10	0.76	0.24	0.20	15.00	2.90	6.24	50.16	3279
c1.medium	6.07	6.01	6.00	0.53	0.09	0.40	17.00	5.70	7.59	90.49	3643
m1.large	12.98	12.15	12.10	4.47	0.34	0.70	68.00	11.40	6.62	60.29	2034
m2.xlarge	17.78	17.05	17.10	4.87	0.27	1.10	80.00	16.20	7.09	57.62	3524
m1.xlarge	24.18	24.05	24.10	2.56	0.11	1.50	100.00	22.80	22.03	599.91	3704
c1.xlarge	26.01	24.26	24.20	8.68	0.33	1.60	128.00	22.80	4.85	27.78	3600
m2.2xlarge	42.15	42.05	42.20	2.47	0.06	2.50	119.00	39.90	14.91	377.30	3790
m2.4xlarge	84.58	84.04	84.20	8.46	0.10	5.00	240.00	79.80	13.54	218.92	3790

variance (CV), interquartile range (IQR), maximum, minimum, skewness (the third moment), kurtosis (the forth moment) and number of samples.

These tables have three types of descriptive statistics. Statistics of the first type (mean, median, trimmed mean) reflect the central tendency of the distributions. Statistics of the second type (CV, IQR, minimum, maximum) measure the spread of the distribution. Statistics of the third type (kurtosis, skewness) reflect the shape of the distribution.

First of all, we find that on average the price of SIs can be as low as %44 of on-demand instances for us-west, eu-west and ap-southeast, and %38, for us-east data centers. This reveals that there are some opportunities in reducing monetary cost of utility computing in cost of reliability. Moreover, the maximum price of some SIs is bigger than the corresponding on-demand instance price specially for us-east data center. Thus if users bid as high as the on-demand prices, we will still have a probability of out-of-bid (failure) event.

The results reveal that the ratios between the mean and the median for prices and inter-price time of SIs are close to one for each data set. This indicates that single parameter distributions might be a good option for the model. This could be confirmed by the skewness and kurtosis values that show the underlying distributions are right-skewed and short-tailed. However, for few SIs in ap-southeast (see Table 5), the skewness is negative, so spot price is left-skewed.

Additionally, the inter-price times have more variability than prices due to higher values of coefficient of variance. Also, analysis of the trimmed mean (the mean value after discarding 10% of extreme values) confirmed that inter-price times have greater variability. So, we may need distributions with higher degrees of freedom, to model the inter-price time for these data sets. It is worth noting that the minimum inter-price time is almost one hour in all data centers except eu-west which is about a few minutes. Moreover, in all data centers, the set price of SIs are stable on average only for 2-3 hours.

6 Distribution Fitting

Before distribution fitting, we apply some randomness tests for spot price and the inter-price time. Results are presented in Section 11. After randomness testing, we first inspect the distribution using Probability Density Function (PDF) and Cumulative Distribution Function (CDF) for spot price and the inter-price time. Then, we conduct parameter fitting for Mixture

Table 4. Statistics for spot price in eu-west data center (Values given in cents).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	4.00	4.00	4.00	0.19	0.05	0.20	9.50	3.80	9.44	242.97	3702
c1.medium	8.00	8.00	8.00	0.27	0.03	0.40	10.10	7.60	0.28	3.91	3812
m1.large	16.04	16.02	16.10	0.85	0.05	1.00	50.00	15.20	21.55	792.41	3875
m2.xlarge	24.04	24.03	24.10	1.03	0.04	1.40	57.10	22.80	12.91	387.69	3763
m1.xlarge	32.05	32.01	32.10	1.60	0.05	2.00	76.00	30.40	15.34	415.47	3917
c1.xlarge	32.04	32.03	32.10	1.07	0.03	2.00	45.00	30.40	0.54	8.27	3658
m2.2xlarge	56.04	56.04	56.20	1.83	0.03	3.42	76.00	53.20	0.25	4.99	4001
m2.4xlarge	112.08	112.08	112.50	3.62	0.03	6.80	150.00	106.40	0.21	4.55	3912

Table 5. Statistics for spot price in ap-southeast data center (Values given in cents).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	4.01	4.00	4.00	0.25	0.06	0.20	10.00	3.80	15.57	366.17	1641
c1.medium	8.01	8.01	8.00	0.27	0.03	0.60	8.40	7.60	-0.01	1.53	1804
m1.large	16.03	16.00	16.00	0.90	0.06	1.00	38.00	15.20	15.98	392.54	1838
m2.xlarge	24.03	24.03	24.10	0.78	0.03	1.50	25.20	22.80	-0.07	1.51	1823
m1.xlarge	32.04	32.05	32.20	1.04	0.03	2.00	33.60	30.40	-0.09	1.52	1876
c1.xlarge	32.10	32.06	32.20	1.93	0.06	2.00	100.00	30.40	23.81	827.36	1857
m2.2xlarge	56.06	56.06	56.20	1.82	0.03	3.60	58.80	53.20	-0.09	1.52	1848
m2.4xlarge	112.05	112.06	112.50	3.64	0.03	7.10	117.60	106.40	-0.07	1.50	1834

Table 6. Statistics for the inter-price time in us-west data center (Values given in hours).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	3.26	2.88	2.47	2.88	0.88	2.66	34.83	1.00	2.95	18.35	2225
c1.medium	3.06	2.71	2.36	2.65	0.87	2.53	27.73	1.01	2.72	14.76	2375
m1.large	2.92	2.58	2.26	2.55	0.87	2.46	32.44	1.00	3.00	18.61	2488
m2.xlarge	2.95	2.59	2.30	2.67	0.91	2.41	39.16	1.00	3.58	28.02	2284
m1.xlarge	2.93	2.61	2.31	2.43	0.83	2.50	21.08	1.00	2.39	11.05	2479
c1.xlarge	2.98	2.64	2.33	2.55	0.86	2.48	22.49	1.00	2.48	11.72	2438
m2.2xlarge	2.91	2.58	2.28	2.48	0.85	2.47	25.55	1.00	2.75	14.65	2495
m2.4xlarge	2.87	2.55	2.29	2.37	0.83	2.38	19.01	1.00	2.34	10.27	2535

Table 7. Statistics for the inter-price time in us-east data center (Values given in hours).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	2.22	1.66	1.36	3.53	1.59	0.32	76.59	0.78	9.21	130.29	3278
c1.medium	2.00	1.65	1.37	2.09	1.05	0.31	49.91	1.00	6.91	98.48	3642
m1.large	3.58	2.20	1.44	18.60	5.20	1.54	657.29	1.00	26.29	824.35	2033
m2.xlarge	1.91	1.58	1.34	2.02	1.06	0.31	36.26	1.00	6.11	61.19	3523
m1.xlarge	1.96	1.62	1.34	3.05	1.55	0.32	145.98	0.58	30.51	1370.41	3703
c1.xlarge	2.02	1.66	1.35	3.38	1.67	0.33	171.62	1.00	35.74	1758.12	3599
m2.2xlarge	1.92	1.62	1.34	1.94	1.01	0.31	50.40	1.01	8.42	143.99	3789
m2.4xlarge	1.92	1.62	1.35	1.76	0.92	0.32	23.02	1.00	4.50	30.98	3789

Table 8. Statistics for the inter-price time in eu-west data center (Values given in hours).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	1.96	1.61	1.35	2.66	1.35	0.30	109.08	0.02	19.94	727.54	3701
c1.medium	1.91	1.59	1.34	1.86	0.97	0.32	22.81	0.02	4.53	30.63	3811
m1.large	1.88	1.57	1.33	1.79	0.95	0.31	30.94	0.02	5.02	42.02	3874
m2.xlarge	1.79	1.53	1.34	1.56	0.87	0.30	22.83	0.02	4.93	38.54	3762
m1.xlarge	1.86	1.58	1.34	1.78	0.96	0.31	38.20	0.02	7.34	101.43	3916
c1.xlarge	1.99	1.56	1.34	7.22	3.63	0.30	378.19	0.02	44.38	2169.40	3657
m2.2xlarge	1.82	1.55	1.33	1.60	0.88	0.31	29.02	0.02	5.11	45.75	4000
m2.4xlarge	1.86	1.58	1.34	1.71	0.92	0.31	26.51	0.02	5.20	44.28	3911

Table 9. Statistics for the inter-price time in ap-southeast data center (Values given in hours).

Instances	Mean	TrMean	Median	Std	CV	IQR	Max	Min	Skewness	Kurtosis	No.
m1.small	3.34	2.86	2.47	4.27	1.28	2.66	103.60	1.01	13.49	286.88	1640
c1.medium	2.98	2.67	2.34	2.44	0.82	2.48	21.62	1.01	2.27	10.02	1803
m1.large	2.93	2.61	2.34	2.37	0.81	2.43	19.01	1.00	2.38	10.49	1837
m2.xlarge	2.95	2.62	2.37	2.46	0.83	2.45	22.41	1.00	2.73	14.39	1822
m1.xlarge	2.86	2.54	2.28	2.37	0.83	2.36	21.27	1.00	2.52	11.91	1875
c1.xlarge	2.90	2.60	2.37	2.32	0.80	2.46	19.30	1.00	2.38	11.06	1856
m2.2xlarge	2.91	2.57	2.29	2.53	0.87	2.44	29.89	1.00	2.94	17.97	1847
m2.4xlarge	2.93	2.61	2.30	2.44	0.83	2.45	21.03	1.00	2.33	10.24	1833

of Gaussians (MoG) distribution. We considered other distributions, like Weibull, Normal, Log-normal and Gamma distributions as well. However, the mixture of Gaussians distribution shows the better fit with respect to others (see Section 11).

6.1 Spot Price

In the following, we present the distribution fitting for the prices of SIs in all data centers.

6.1.1 PDF and CDF

The PDF and CDF of prices of each SI in all data centers are depicted in Figure 4 and 5. One interesting result from these figures is existing of two modes (peaks) in the probability density functions which imply that we have two components in the distributions. So, looking into some mixture distribution like Gamma and Mixture of Gaussians would be reasonable. However, there are some SIs in the us-east like m1.small which are not follow this type of distribution.

In this part, we conduct parameter fitting for the Mixture of Gaussians distribution with k components which is defined as follows:

$$cdf(x; \vec{\mu}, \vec{\sigma^2}, \vec{p}, k) = \sum_{i=1}^{k} \frac{p_i}{2} \left(1 + erf(\frac{x - \mu_i}{\sigma_i \sqrt{2}}) \right) \quad (1)$$

where $\vec{\mu}$, $\vec{\sigma^2}$, and $\vec{p}$ are the mean, variance and the probability of each component with k items. Also, $erf()$ is the error function and defined as follows:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} dt \quad (2)$$

Data generated by Mixture of Gaussians densities are characterized by clusters centered at mean μ_i with increased density for points closer to the mean.
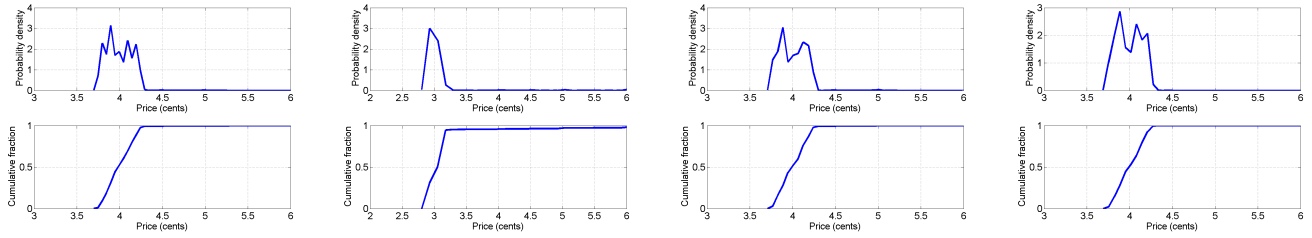
(a) m1.small
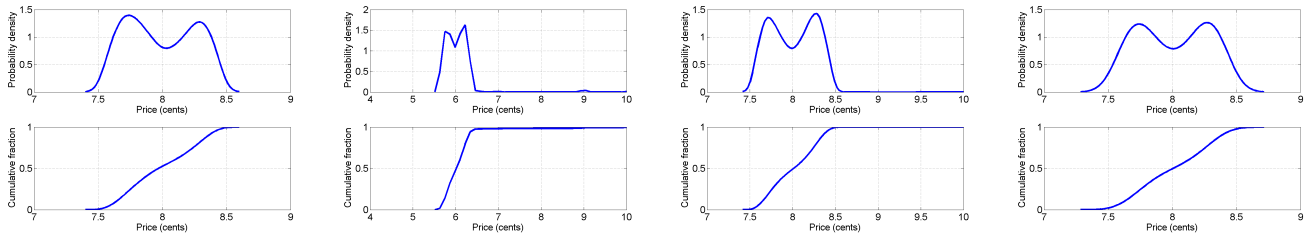
(b) c1.medium
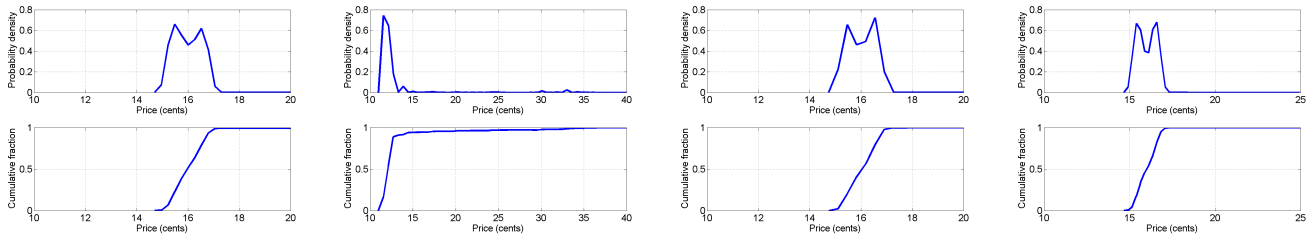
(c) m1.large
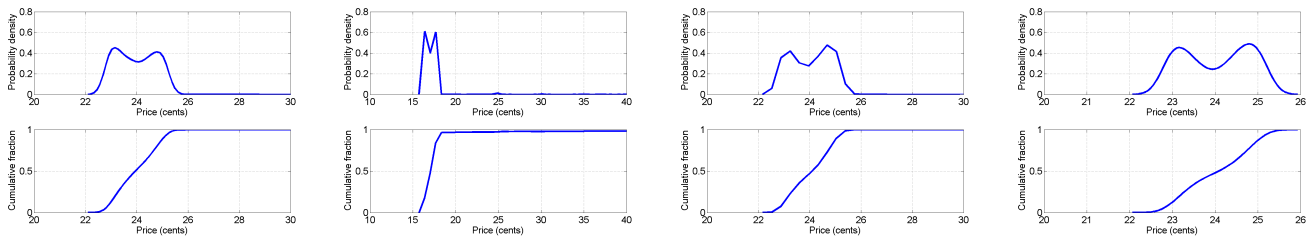
(d) m2.xlarge

Figure 4. PDF and CDF of spot price in the all data centers (us-west, us-east, eu-west, ap-southeast)
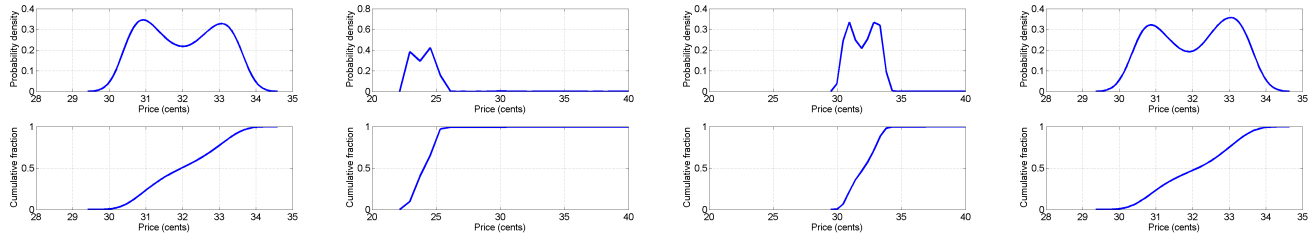
(a) m1.xlarge
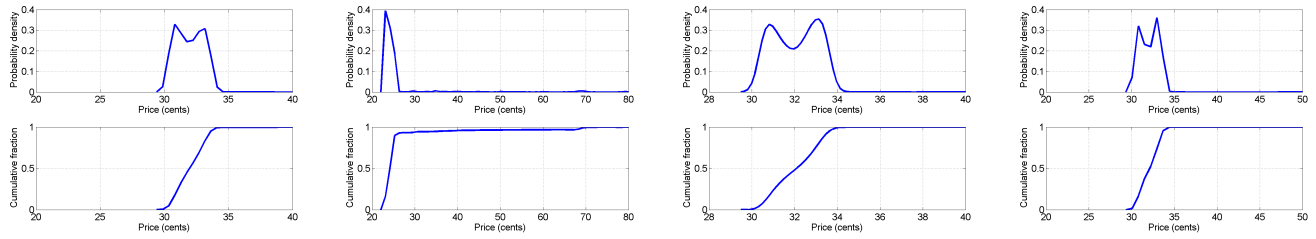
(b) c1.xlarge
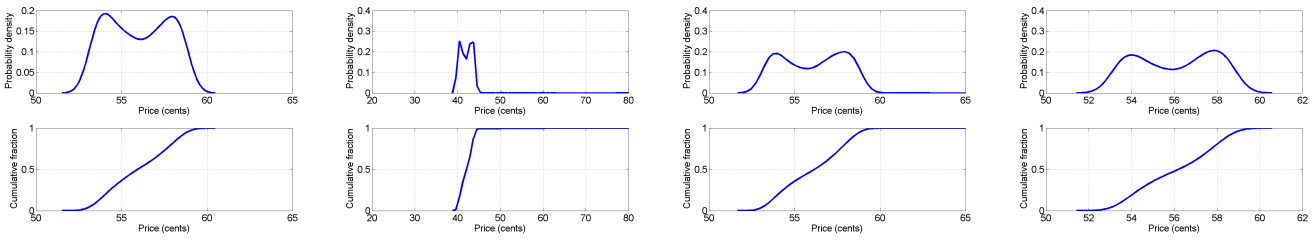
(c) m2.2xlarge
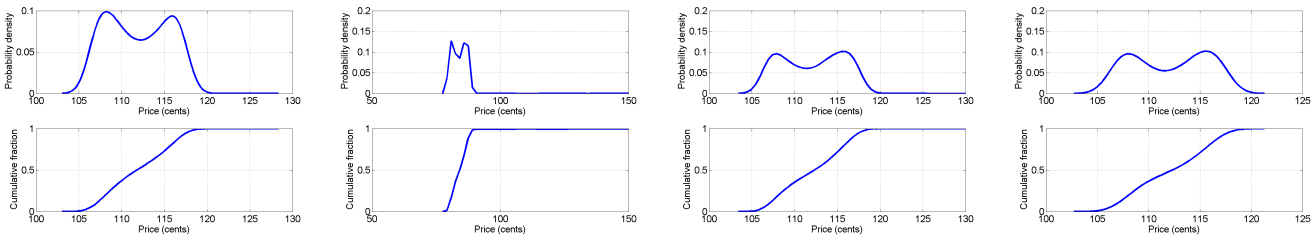
(d) m2.4xlarge

Figure 5. PDF and CDF of spot price in the all data centers (us-west, us-east, eu-west, ap-southeast)

Table 10. p-values resulting from KS and AD tests for spot price in eu-west data center.

Instances	MoG ($k = 2$)	MoG ($k = 3$)	MoG ($k = 4$)
m1.small	0.016 0.791	0.017 0.789	0.053 0.803
c1.medium	0.211 0.779	0.217 0.791	0.224 0.790
m1.large	0.113 0.678	0.319 0.752	0.354 0.754
m2.xlarge	0.139 0.616	0.356 0.721	0.415 0.734
m1.xlarge	0.134 0.570	0.369 0.708	0.431 0.706
c1.xlarge	0.394 0.681	0.444 0.705	0.421 0.707
m2.2xlarge	0.420 0.648	0.469 0.682	0.450 0.672
m2.4xlarge	0.429 0.617	0.463 0.637	0.476 0.653

Table 11. Parameters of some distributions for spot price in eu-west data center.

Instances	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)	MoG($k = 4, p, \mu, \sigma$)
m1.small	0.005 5.216 3.997 1.670 0.020	0.003 0.003 5.216 5.216 3.997 1.670 1.670 0.020	0.431 0.007 0.281 4.139 5.017 3.887 3.887 0.003 1.363 0.006 0.006
c1.medium	0.589 8.194 7.734 0.028 0.010	0.443 0.276 8.018 8.292 7.703 0.045 0.006 0.006	0.138 0.281 0.138 8.292 7.703 8.292 8.018 0.006 0.006 0.006 0.045
m1.large	0.999 16.021 29.486 0.266 143.696	0.492 0.505 15.556 16.470 24.401 0.059 0.048 114.879	0.370 0.002 0.251 15.902 24.429 15.387 16.555 0.092 115.014 0.016 0.026
m2.xlarge	0.999 24.029 53.500 0.593 12.960	0.445 0.001 23.264 53.500 24.643 0.109 12.960 0.135	0.248 0.380 0.001 23.049 24.817 53.500 23.877 0.028 0.062 12.960 0.214
m1.xlarge	0.998 32.006 53.707 1.058 328.521	0.457 0.002 31.010 53.803 32.848 0.184 326.523 0.249	0.002 0.428 0.289 53.873 32.027 33.184 30.762 326.054 0.428 0.069 0.058
c1.xlarge	0.671 32.625 30.836 0.594 0.091	0.261 0.243 33.188 30.756 32.057 0.072 0.058 0.722	0.395 0.353 0.002 31.855 33.134 37.052 30.740 0.400 0.096 17.982 0.054
m2.2xlarge	0.350 53.971 57.149 0.290 1.479	0.492 0.252 56.119 53.784 58.100 1.813 0.157 0.216	0.000 0.269 0.286 76.000 58.117 53.827 56.191 0.000 0.209 0.184 1.308
m2.4xlarge	0.359 108.037 114.338 1.270 5.528	0.263 0.249 116.126 107.609 112.183 0.898 0.660 7.061	0.434 0.233 0.093 111.613 107.564 117.073 115.378 6.820 0.606 0.104 1.023

6.1.2 Goodness of Fit Tests

Parameter fitting was done using Model Based Clustering (MBC) which is introduced by Fraley and Raftery [8]. MBC is a methodological framework that can be used for data clustering as well as (multi)variate density estimation. The assumption is that data has several components where each of which is generated by a probability distribution. The expectation maximization (EM) algorithm, which is a general maximum likelihood estimation is adopted to maximize the data likelihood in terms of parameters $\vec{\mu}$ and $\vec{\sigma^2}$ where k is given as a priori. Model Based Clustering uses Bayesian model selection to choose the best model in terms of number of components. In contrast, we use the goodness of fit (GOF) tests to determine the best model as we have an estimation for the number of components in the model. We choose the number of components between 2 and 4 ($2 \leq k \leq 4$) based on the observation of the density functions. We measured the goodness of fit of the resulting models using a visual method (i.e. standard probability-probability (PP) plots) and Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests as quantitative metrics.

First of all, we presented the graphical results of distribution fitting for price of all SIs in Figure 14, 15, 16, 17, in us-west, us-east, eu-west and ap-southeast data centers, respectively. In these plots, the closer the plots are to the line $y = x$, the better the fit. Based on this figure, Mixture of Gaussians distributions with three or four components are the best fit for the most cases. Also, Log-normal and Gamma distributions provide some good fits for a few cases. Based on the Figure 15, the prices of SIs in us-east data center (specially for m1.small and c1.medium) are hard to fit with any distribution.

To be more quantitative, we also reported the p-values of two goodness-of-fit tests. We randomly select a subsample of 50 of each data set and compute the p-values iteratively for 1000 times and finally obtain the average p-value. This method is similar to the one used by the authors in [10], and was suggested to us by a statistician.

The results of GOF tests are listed in Table 17, 18, 19 and 20 for us-west, us-east, eu-west and ap-southeast data center, respectively. Moreover, in the each row the best fit is highlighted. In some cases, we have two winners as there is one best fit per each GOF test. These quantitative results strongly confirm the graphical results of the PP-plots where the Mixture of Gaussians with three or four components are the best fit for the most cases.

The set of parameters for some fitted distributions are listed in Table 21, 22, 23 and 24 for us-west, us-east, eu-west and ap-southeast data center, respectively. As it can be seen in Equation (1), the number of parameters in the MoG distribution depends on k. So we have a trade-off between accuracy and complexity where the MoG distribution with $k = 3$ which has 10 parameters is the best fit. However, we can utilize other good fit distributions like Log-Normal with only two parameters. It is worth nothing that in the list of parameters for MoG, we just report $k - 1$ items of parameter p_i, as the last item can be computed by others. (i.e. $p_k = 1 - \sum_{i=1}^{k-1} p_i$).
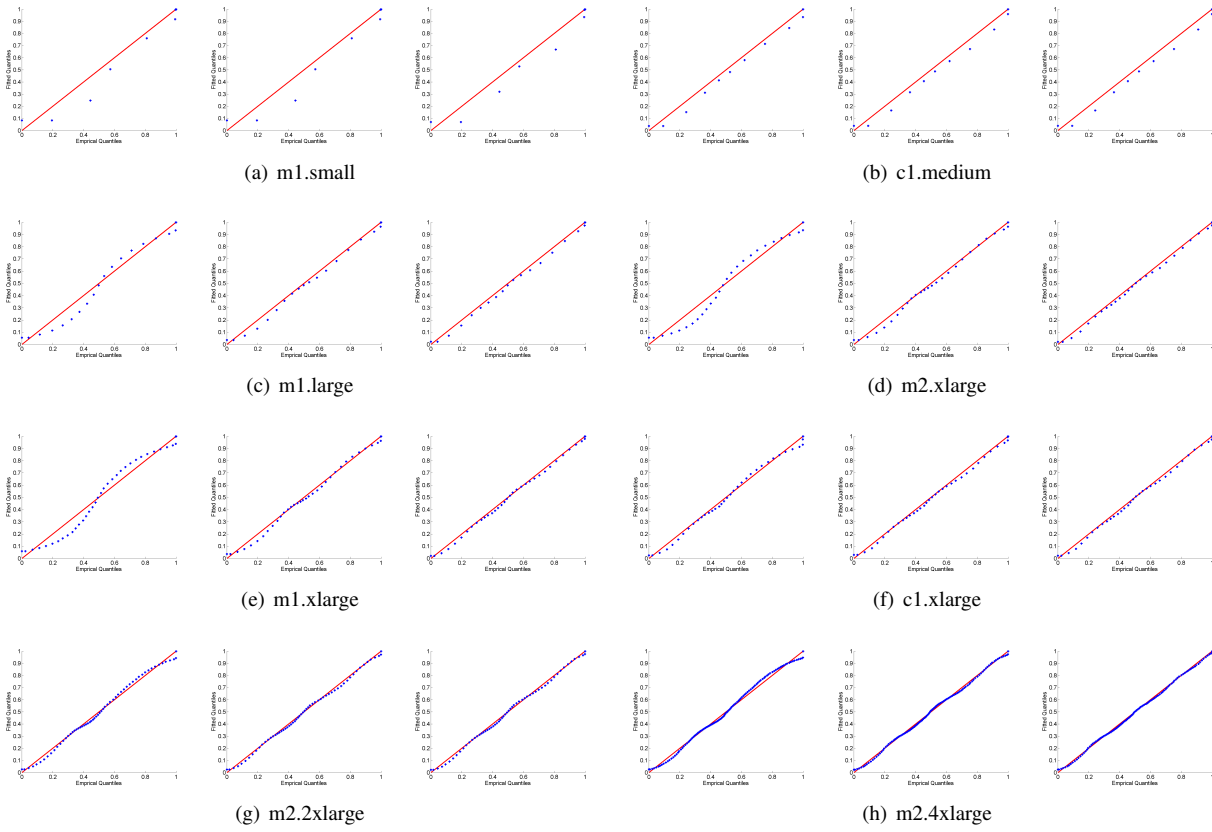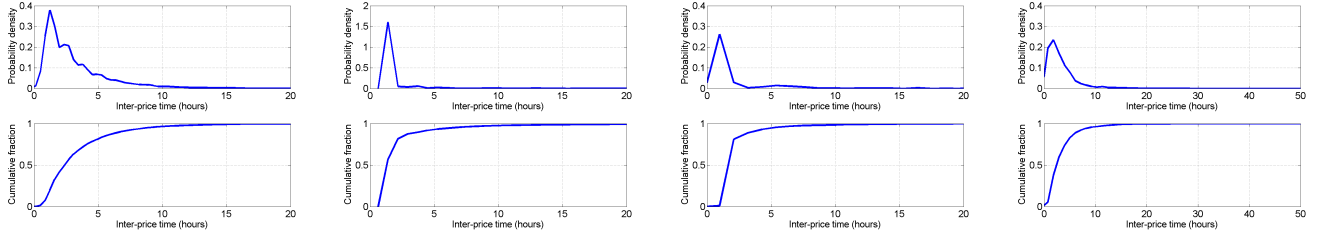
(a) m1.small　　　　　　　　　　　　　　　　　　(b) c1.medium

(c) m1.large　　　　　　　　　　　　　　　　　　(d) m2.xlarge

(e) m1.xlarge　　　　　　　　　　　　　　　　　　(f) c1.xlarge

(g) m2.2xlarge　　　　　　　　　　　　　　　　　　(h) m2.4xlarge

Figure 6. PP-plots of spot price for eu-west data center for Mixture of Gaussians ($k = 2$**,** $k = 3$**,** $k = 4$**)**

6.2 Inter-price Time

In the following, we present the distribution fitting for the inter-price time of different SIs in all data centers.

6.2.1 PDF and CDF

The PDF and CDF of the inter-price time for each SI in all data centers are depicted in Figure 7 and 8. As you can see in these figures, there are one dominant mode (peak) in the probability density functions in compare to two (nearly) equal peaks in the price probability density functions. Moreover, as we expected from the global statistics, the CDFs reveal longer-tail distributions than price distributions.

6.2.2 Goodness of Fit Tests

We presented the PP-plots of distribution fitting for all SIs in Figure 18, 19, 20, 21, for the us-west, us-east, eu-west and ap-southeast data centers, respectively. In these plots, the closer the plots are to the line $y = x$, the better the fit. Based on this figure, Mixture of Gaussians ($k = 4$) distribution is the best fit for the most cases.

To be more quantitative, we also reported the p-values of two goodness-of-fit tests. The results of GOF tests are listed in Table 25, 26, 27 and 28 for us-west, us-east, eu-west and ap-southeast data center, respectively. Moreover, in the each row the best fit is highlighted. These quantitative results strongly confirm the graphical results of the PP-plots where the Mixture of Gaussians ($k = 4$) distribution is the best fit for the inter-price time.

The set of parameters for each fitted distribution are listed in Table 29, 30, 31 and 32 for us-west, us-east, eu-west and ap-southeast data center, respectively.

7 Model Calibration

In this section, we look into the time evolution of the spot price and the inter-price time which may lead us to obtain a more accurate model. As such, we use the scatter plot of spot price and the inter-price time for duration of February 2010 to November 2010. Due to space limitation, we just present the plots for m2.4xlarge instance. The results are consistent for other instances within the data center.

Figure 10(a) depicts the scatter plot of spot price for the duration of the considered history. As it can be seen in this figure, there is no obvious correlation in spot price where they are evenly distributed in a specific rang (the rang depends on the type of instances). However, the congestion of spot price is increased after mid-July and this is the case for all SIs in eu-west data center. To confirm this observation, we depict the scatter plot of the inter-price time for this SI in Figure 10(b). We observe that inter-price time become suddenly shorter after mid-July. That means, the frequency of changing the prices is increased where spot price remain unchanged. The inspections of other SIs within the data center reveal the same result. This is also the reason of very sharp peak in the density function of the inter-price time in Figure **??**.

This artifact is possibly due to some fine tunings in the pricing algorithm which have been made by Amazon. It is worth noting that the same issue has been observed in different dates in other Amazon' EC2 data centers where in us-east happened in August 2010, and in us-west and ap-southeast in January 2011 (see Figure 11).

Table 12. p-values resulting from KS and AD tests for the inter-price time in eu-west data center.

| Instances	MoG ($k = 2$)	MoG ($k = 3$)	MoG ($k = 4$)
m1.small	0.347 0.476	0.415 0.592	0.489 0.627
c1.medium	0.382 0.546	0.390 0.566	0.380 0.566
m1.large	0.390 0.552	0.387 0.573	0.400 0.574
m2.xlarge	0.389 0.556	0.393 0.566	0.405 0.585
m1.xlarge	0.369 0.526	0.391 0.564	0.406 0.581
c1.xlarge	0.221 0.319	0.399 0.561	0.467 0.602
m2.2xlarge	0.376 0.532	0.426 0.570	0.463 0.610
m2.4xlarge	0.368 0.529	0.383 0.569	0.395 0.573

(a) m1.small
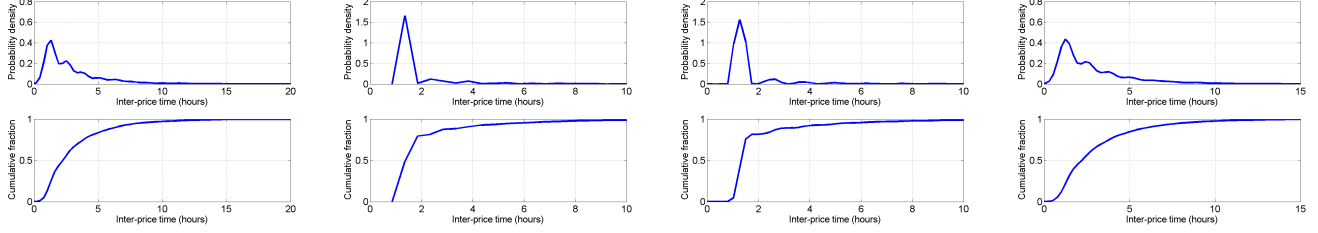
(b) c1.medium
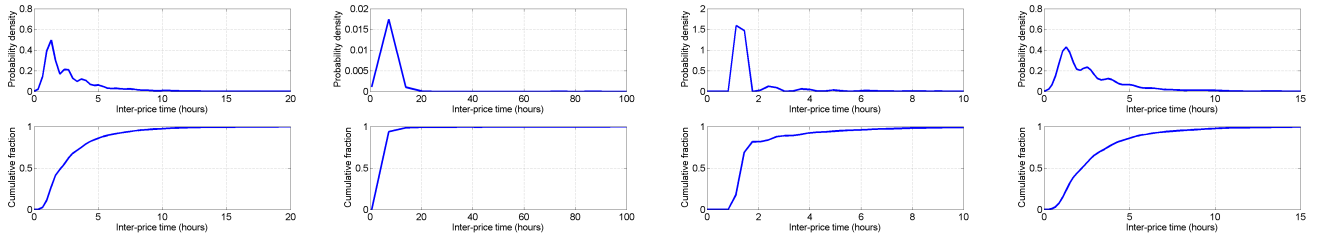
(c) m1.large
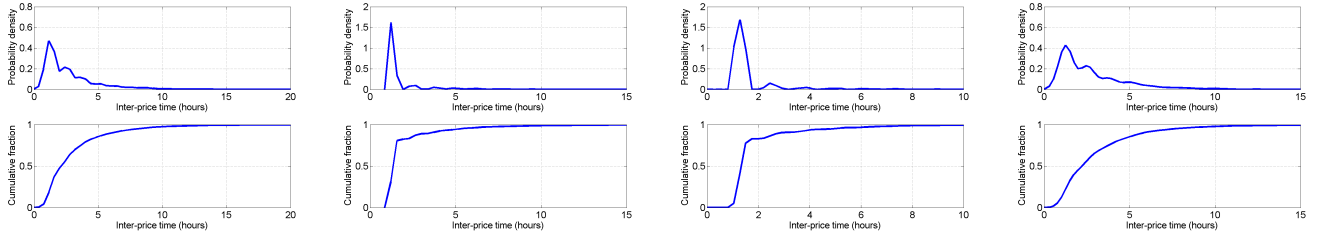
(d) m2.xlarge

Figure 7. PDF and CDF of the inter-price time in the all data centers (us-west, us-east, eu-west, ap-southeast)

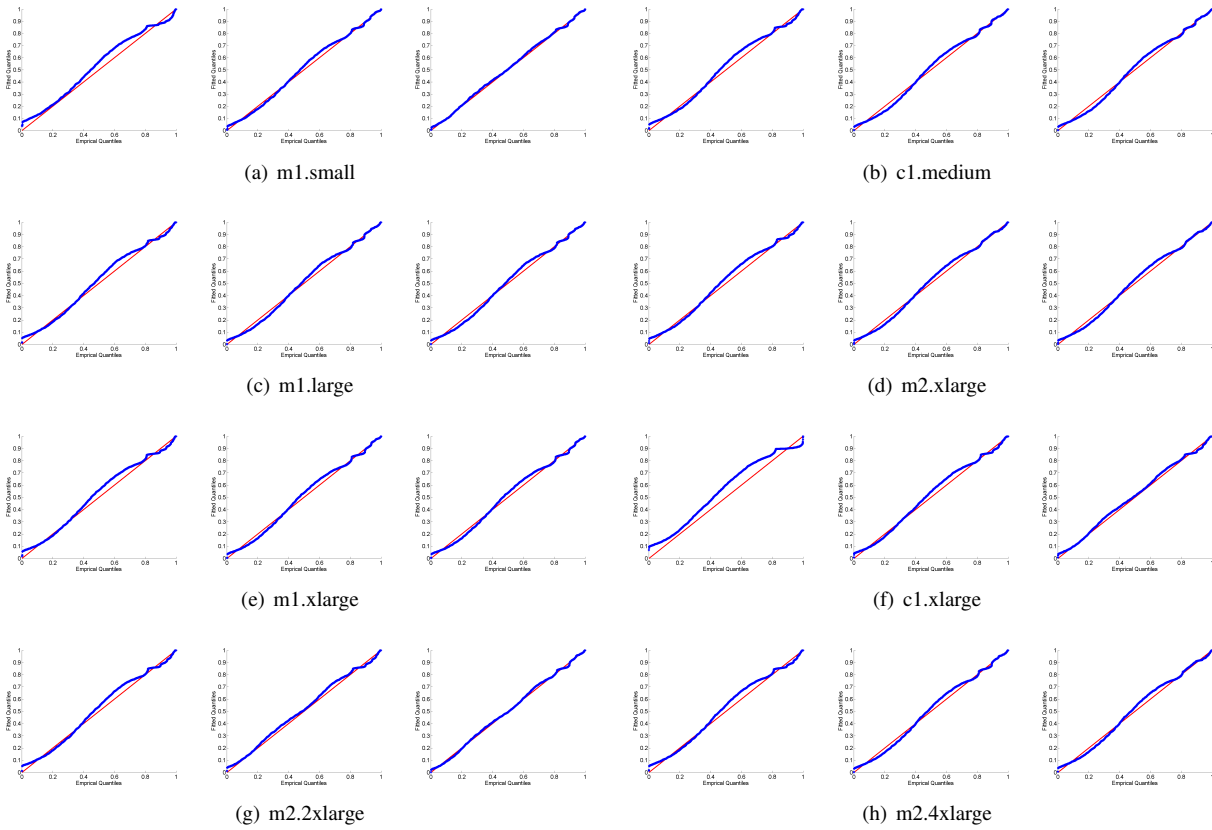**Table 13. Parameters of distributions for the inter-price time in eu-west data center.**

Instances	MoG($k=2, p, \mu, \sigma$)	MoG($k=3, p, \mu, \sigma$)	MoG($k=4, p, \mu, \sigma$)
m1.small	0.798 1.293 4.616 0.022 26.019	0.178 0.028 3.474 11.536 1.292 2.308 120.051 0.022	0.431 0.362 0.028 1.184 1.420 11.546 3.462 0.008 0.007 120.142 2.315
c1.medium	0.803 1.280 4.472 0.022 9.274	0.807 0.090 1.279 6.452 2.876 0.022 12.435 0.528	0.479 0.087 0.107 1.276 6.533 2.902 1.284 0.022 12.618 0.595 0.022
m1.large	0.804 1.277 4.333 0.022 8.707	0.068 0.126 6.793 3.040 1.276 13.803 0.940 0.022	0.510 0.068 0.126 1.271 6.793 3.040 1.285 0.022 13.803 0.940 0.022
m2.xlarge	0.184 4.054 1.284 6.823 0.022	0.824 0.066 1.284 2.506 5.166 0.022 0.035 8.192	0.110 0.410 0.413 5.166 1.284 1.284 2.506 8.192 0.022 0.022 0.035
m1.xlarge	0.207 4.056 1.283 9.182 0.022	0.793 0.177 1.283 3.310 8.356 0.022 1.864 29.730	0.224 0.031 0.569 1.298 8.356 1.277 3.310 0.022 29.731 0.022 1.864
c1.xlarge	0.819 1.287 5.160 0.022 275.687	0.811 0.187 1.286 4.048 84.430 0.022 5.341 15636.817	0.477 0.334 0.002 1.190 1.424 85.037 4.052 0.010 0.005 15724.089 5.339
m2.2xlarge	0.804 1.276 4.036 0.022 6.804	0.405 0.399 1.155 1.398 4.044 0.007 0.008 6.795	0.379 0.145 0.050 1.404 3.121 6.820 1.161 0.007 1.177 12.951 0.007
m2.4xlarge	0.203 4.134 1.280 7.844 0.022	0.063 0.137 6.705 3.001 1.279 13.524 0.863 0.022	0.438 0.367 0.122 1.277 1.283 5.278 2.507 0.022 0.022 9.614 0.036

14
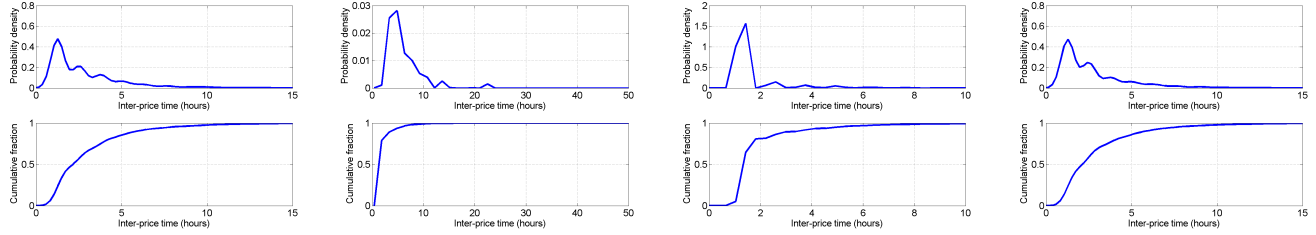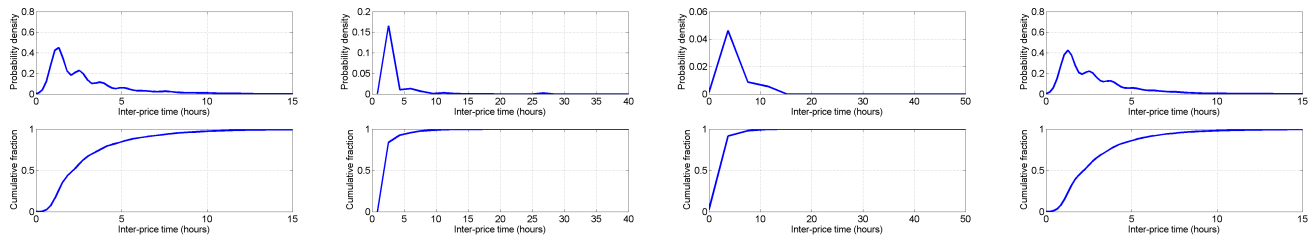
(a) m1.xlarge

(b) c1.xlarge
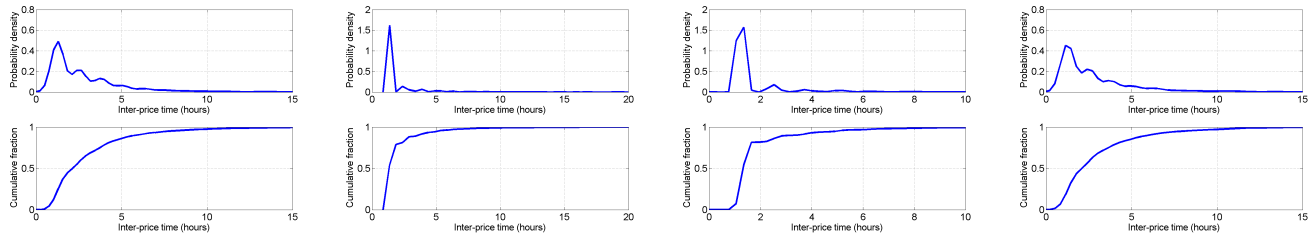
(c) m2.2xlarge
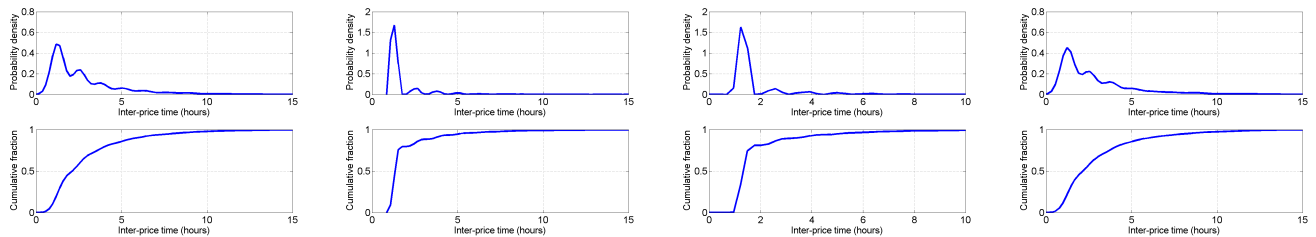
(d) m2.4xlarge

Figure 8. PDF and CDF of the inter-price time in the all data centers (us-west, us-east, eu-west, ap-southeast)

(a) m1.small   (b) c1.medium

(c) m1.large   (d) m2.xlarge

(e) m1.xlarge   (f) c1.xlarge

(g) m2.2xlarge   (h) m2.4xlarge

Figure 9. PP-plots of the inter-price time for eu-west data center for Mixture of Gaussians ($k = 2$, $k = 3$, $k = 4$)

Focusing on the graphical demonstration of the existing components in the inter-price time which is presented in Figure 10(b), we can see that after the aforementioned date only one component remains and other components are almost faded. As this observation is consistent over all SIs, we propose the model calibration algorithm (Algorithm 1) to find the date of changing in the pricing (which is called calibration date) as well as remaining component(s).

The algorithm needs the trace of the inter-price time of an SI ($Trace_{inst}$) and the number of components (k). The result of Mixture of Gaussians fitting with k components is $\overrightarrow{index}$ where $\overrightarrow{date}$ is the vector of correspondence date to each item of $\overrightarrow{index}$. Then, the algorithm computes the probability of each component in each month in the whole trace and after that finds a list ($\overrightarrow{Q_m}$) where the probability of one or more components is less than q_0 (line 5-9). q_0 is a threshold value and we define it as low as 0.01 (i.e. $q_0 = 0.01$). The components that are not in this list are remaining components (line 10,11). The first month in the list of $\overrightarrow{Q_m}$ is the calibration month, called m (line 13). Finally, the last occurrence of the component(s) in month m would be the calibration date ($CalDate$) which is obtained in line 15-22.

The results of applying this algorithm for all SIs in eu-west data center are presented in Table 14. As you can see, all calibration dates are in July. Moreover, for all SIs, except m2.2xlarge, only one component remains after the calibration dates. The remaining components can be examined in the third column of the Table **??**, where the component(s) with higher probability remain(s) beyond calibration date. For instance, the third component of the MoG model for m2.4xlarge with probability of 0.8 remains after 15-July where the mean and variance are 1.279 and 0.022 hours, respectively. The graphical demonstration of Figure 10(b) can confirm the correctness of this algorithm where the component 3 implies a cluster around the mean value of 1.279 hours.

The last step of the model calibration is probability adjustment where the probability of remaining component(s) must be scaled up to one. This can be done by the following formula:

$$p_j = \frac{p_j}{\sum_{\forall i} p_i} \quad i,j \in \overrightarrow{RCmps} \tag{3}$$

In the other words, for the calibrated model for each SI, we just change the probability of remaining component(s) after the calibration date. In the following section, we investigate the accuracy of the calibrated models with respect to the original models.

(a) Price distribution for m2.4xlarge

(b) Inter-price time distribution for m2.4xlarge
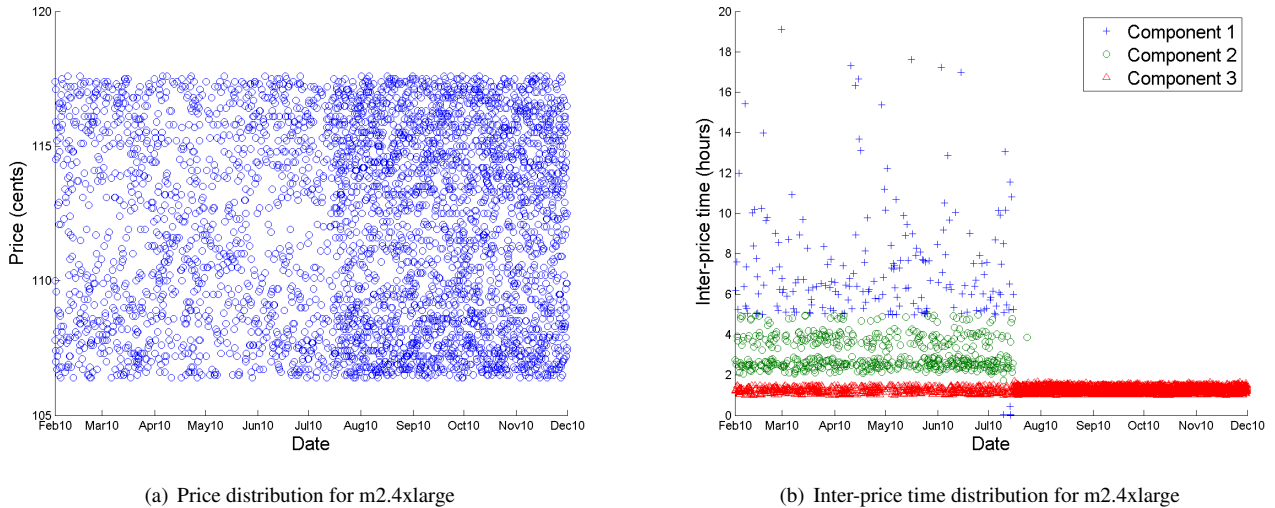
Figure 10. Price and Inter-price time distribution over time for m2.4xlarge in eu-west
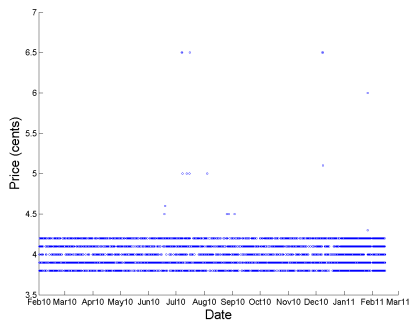
(a) Price distribution for m1.samll (us-west)
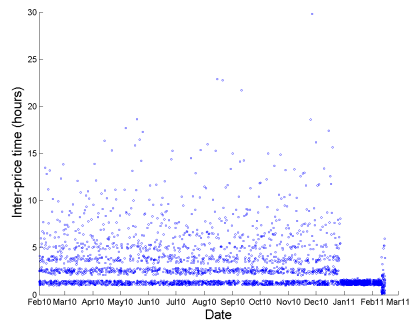
(b) Inter-price time distribution for m1.samll (us-west)
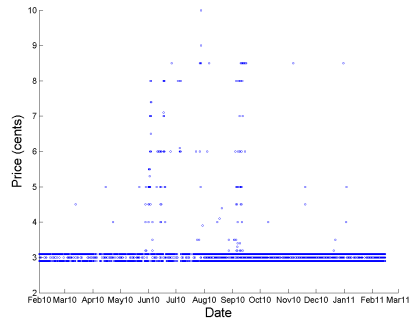
(c) Price distribution for m1.samll (us-east)
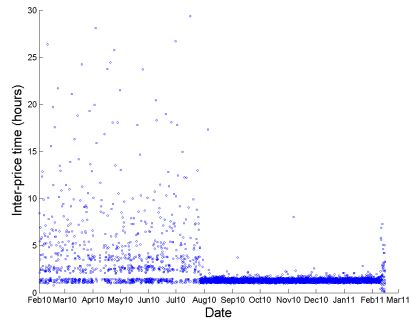
(d) Inter-price time distribution for m1.samll (us-east)
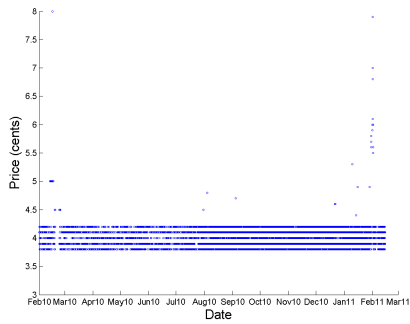
(e) Price distribution for m1.samll (eu-west)
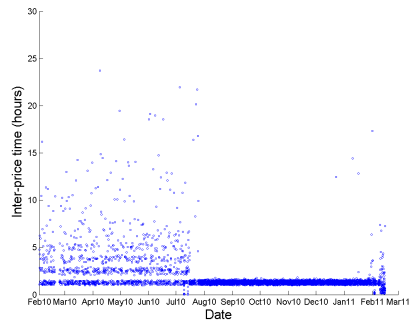
(f) Inter-price time distribution for m1.samll (eu-west)
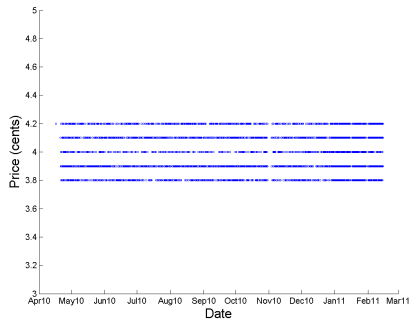
(g) Price distribution for m1.samll (ap-southeast)
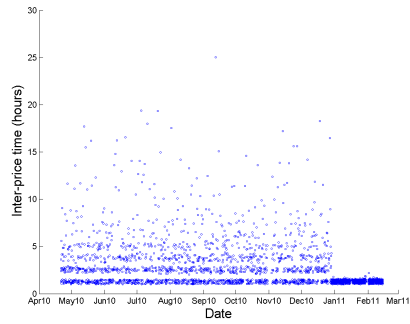
(h) Inter-price time distribution for m1.samll (ap-southeast)

Figure 11. Price and Inter-price time in different data centers

---

Algorithm 1: Model Calibration Algorithm

Input: $Trace_{inst}, k$
Output: $CalDate, \overrightarrow{RCmps}$

1 $T_s \leftarrow Trace_{inst}.start.time$;
2 $T_e \leftarrow Trace_{inst}.end.time$;
3 $n \leftarrow Sizeof(Trace_{inst})$;
4 //$\overrightarrow{index}$ is the result of the MoG model with k components;
5 $\overrightarrow{index} \leftarrow \{c_i | c \in \{1, \ldots, k\}, i \in \{1, \ldots, n\}\}$;
6 $\overrightarrow{date} \leftarrow \{d_i | d \in \{T_s \ldots T_e\}, i \in \{1, \ldots, n\}\}$;
7 $q_{a,b} \leftarrow$ probability of component a in month b;
8 $\overrightarrow{Q} \leftarrow \{q_{a,b} | a \in \{1, \ldots, k\}, b \in \{T_s \ldots T_e\}\}$;
9 $\overrightarrow{Q_m} \leftarrow \{q_{f,e} | q_{f,e} < q_0, q_{f,e} \in \overrightarrow{Q}\}$;
10 $\overrightarrow{Cmps} \leftarrow \{g | q_{g,h} \in \overrightarrow{Q_m}\}$;
11 $\overrightarrow{RCmps} \leftarrow \{1, \ldots, k\} - \overrightarrow{Cmps}$;
12 //find the first month with a low probability;
13 $m \leftarrow min\{h | q_{g,h} \in \overrightarrow{Q_m}\}$;
14 //$Trace_{inst}(m)$ is the trace for month m;
15 $T_{m_s} \leftarrow Trace_{inst}(m).start.time$;
16 $T_{m_e} \leftarrow Trace_{inst}(m).end.time$;
17 $z \leftarrow Sizeof(Trace_{inst}(m))$;
18 $\overrightarrow{Sindex} \leftarrow \{c_j | c \in \{1, \ldots, k\}, j \in \{1, \ldots, z\}\}$;
19 $\overrightarrow{Sdate} \leftarrow \{d_j | d \in \{T_{m_s} \ldots T_{m_e}\}, j \in \{1, \ldots, z\}\}$;
20 //find the last occurrence of component g in month m;
21 $t \leftarrow max\{r_l | \overrightarrow{Sindex}(r_l) == g, l \in \{1, \ldots, z\}\}$;
22 $CalDate \leftarrow \overrightarrow{Sdate}(t)$;

Table 14. Results of model calibration algorithm for all spot instances in eu-west ($k = 3$).

Instances	Calibration Dates	Remaining Components
m1.small	24-July	3
c1.medium	15-July	1
m1.large	15-July	3
m2.xlarge	13-July	1
m1.xlarge	23-July	1
c1.xlarge	23-July	1
m2.2xlarge	23-July	1,2
m2.4xlarge	15-July	3
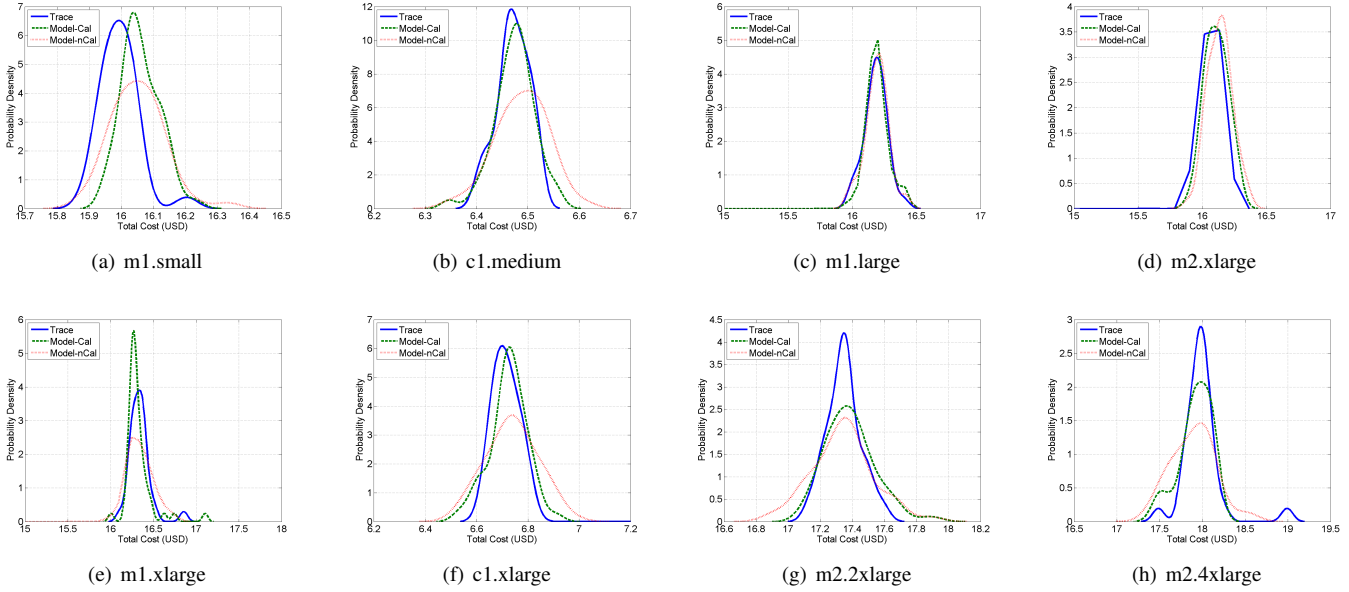
Figure 12. Model validation for all SIs in eu-west for the modeling traces (Feb-2010 to Nov-2010).

8 Model Validation

In order to validate the discovered models, we implemented a discrete event simulator using CloudSim [5][2]. The simulator uses the model or the price history traces to run the input workload. We consider the case where the user requests for one VM from one type of SI and runs whole jobs on that VM. The total monetary cost of running the workload on an SI is the parameter to be considered.

8.1 Simulation Setup

The workload that we used in our experiments is the LCG1 workload traces from LCG Grid which is taken from the Grid Workloads Archive [9]. We used the first 1000 jobs of this trace as the input workload for the experiments which is long enough to reflect the behavior of spot price for different SIs. We assume that one EC2 compute unit is equivalent of a CPU core with capacity of 1000 MIPS[3]. As such, the selected workload needs about two weeks (≈ 400 hours) to finish on a single m1.small instance to complete. For other instance types we consider the linear speedup with the computing capacity in terms of EC2 compute unit which are listed in Table 1.

Moreover, we assume a very high user' bid for each simulation (for example on-demand price) where we do not have any *out-of-bid* event in the execution of the given workload. We use the model for eu-west data center with three components ($k = 3$) for both spot price and inter-price time to show the trade off-between accuracy and complexity. In our experiments, the results of the simulations are accurate with a confidence level of 95%.

8.2 Results and Discussions

In the following, we present the results of two different set of experiments. First, the results of model validation are discussed where we have the price history which was included in the modeling process (i.e, Feb-2010 to Nov-2010). Second, the same results for a new price history which was not included in the modeling process are reported. The new price history is from December 2010 till mid-February 2011.

Figure 12 shows the model validation results where the probability density functions of the total cost of running the given workload for all types of SIs have been plotted. In each plot, Trace, Model-Cal, and Model-nCal refer to the result of using

---

[2]The simulator of Spot Instances will be publicly available on the CloudSim website at: http://www.cloudbus.org/cloudsim/

[3]Amazon mentioned that one EC2 compute unit has equivalent CPU capacity of 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor [13].
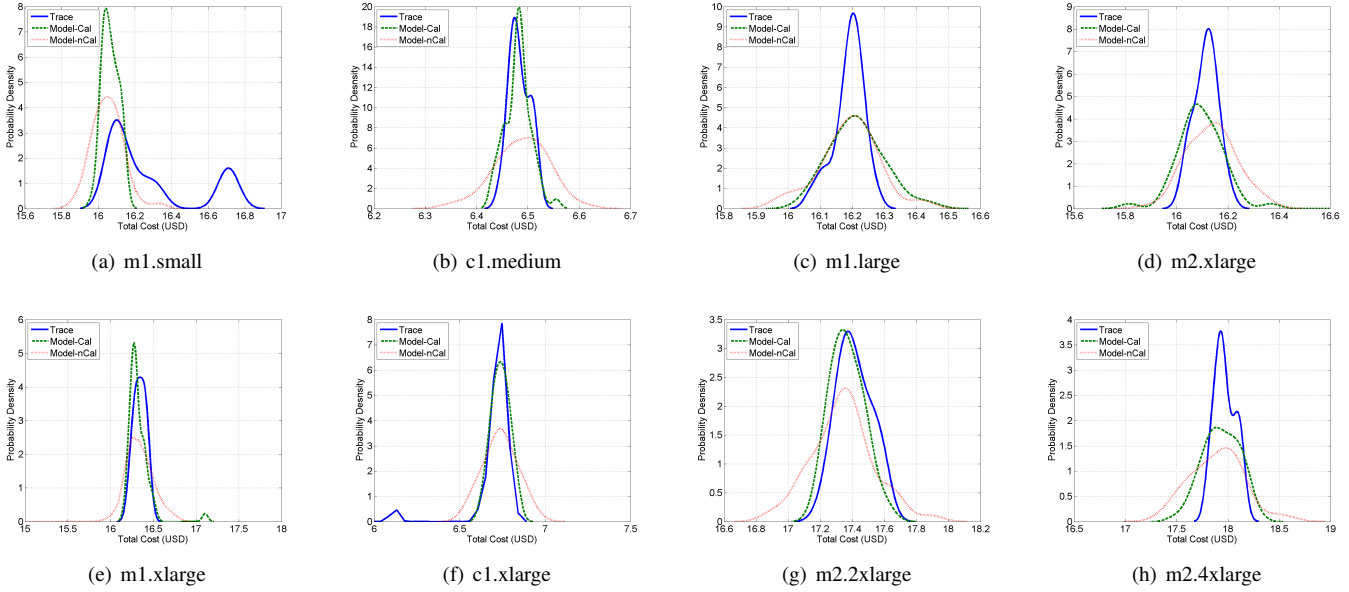
Figure 13. Model validation for all SIs in eu-west for the new traces (Dec-2010 to mid-Feb-2011).

the real price history, the model after calibration and the model without calibration, respectively. Based on these Figures, the discovered models match the real trace simulations with a high degree of accuracy, specially for the calibrated models. As you can see in these plots, in all cases the calibrated models are the better match with the trace simulations. As we expect, there are discrepancies in the results provided by the model and the trace simulation for m1.small instance. However, the mean total cost for running the given workload for all SIs is very accurate where the maximum relative error is less than 3% for both calibrated and non-calibrated model, respectively.

Additionally, we report the same results where we use the new price history from December 2010 to mid-February 2011 to see how good the models are for the future traces. The result of the simulations for the new price history are plotted in Figure 13. The results reveal that the discovers models with three components still conform to the trace simulation results, except for m1.small instance. As it is mentioned before, spot price for m1.small instance is hard to fit and this is the reason of this inaccuracy. That means that for this type of SI, we should use the model with more components (e.g. $k = 4$) to get the better accuracy. As it is expected, the calibrated models again have the better match with respect to the non-calibrated models for all SIs. Besides, the maximum relative error of the mean total cost for all SIs is less than 4% for both calibrated and non-calibrated model, respectively. Therefore, the discovered models are accurate enough for the new price history as well.

9 Conclusions

We considered the problem of discovering models for Spot Instances in Amazon' EC2 data centers in terms of spot price and inter-price time. Based on one year price history given by Amazon, we found the model with Mixture of Gaussians distribution with 3 or 4 components for each type of SI. We also proposed an algorithm to calibrate the discovered models to increase their degree of accuracy. The model is validated through simulations, which have shown that the model predicts the total price of running jobs on spot instances with a good degree of accuracy. We believe that this characterization is fundamental in the design of stochastic scheduling algorithms and fault tolerant mechanisms in public cloud computing environments while using spot market.

In future work, we intend to consider the user' bid as the third parameter and investigate how it can affect the distribution of failures. Moreover, we would like to build a Markov chain as a more sophisticated method for component transition.

10 Acknowledgment

The authors would like to thank William Voorsluys, Sangho Yi and Prof. Ruppa Thulasiram for useful discussions.

References

[1] Cloud exchange website. `http://cloudexchange.org/`.

[2] Amazon Inc. Amazon Discussion Forums. `https://forums.aws.amazon.com`.

[3] Amazon Inc. Amazon Elastic Compute Cloud (Amazon EC2). `http://aws.amazon.com/ec2`.

[4] Artur Andrzejak, Derrick Kondo, and Sangho Yi. Decision model for cloud computing under SLA constraints. In *18th IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, pages 257–266, 2010.

[5] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose, and Rajkumar Buyya. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience*, 41(1):23–50, 2011.

[6] Navraj Chohan, Claris Castillo, Mike Spreitzer, Malgorzata Steinder, Asser Tantawi, and Chandra Krintz. See spot run: using spot instances for MapReduce workflows. In *the 2nd USENIX conference on Hot topics in cloud computing*, HotCloud'10, pages 7–7, 2010.

[7] Feitelson. D. *Workload Modeling for Computer Systems Performance Evaluation*. 2009.

[8] Chris Fraley and Adrian E Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, 2002.

[9] Alexandru Iosup, Hui Li, Mathieu Jan, Shanny Anoep, Catalin Dumitrescu, Lex Wolters, and Dick H. J. Epema. The Grid Workloads Archive. *Future Generation Computer Systems*, 24(7):672–686, 2008.

[10] Bahman Javadi, Derrick Kondo, Jean-Marc Vincent, and David P. Anderson. Mining for statistical availability models in large-scale distributed systems: An empirical study of SETI@home. In *17th IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, pages 1–10, 2009.

[11] Derrick Kondo, Artur Andrzejak, and David P. Anderson. On correlated availability in internet distributed systems. In *9th IEEE/ACM International Conference on Grid Computing (Grid 2008)*, pages 276–283, 2008.

[12] Michael Mattess, Christian Vecchiola, and Rajkumar Buyya. Managing peak loads by leasing cloud infrastructure services from a spot market. In *12th IEEE International Conference on High Performance Computing and Communications*, pages 180–188, 2010.

[13] Jinesh Varia. *Cloud Computing: Principles and Paradigms*, chapter Architecting Applications for the Amazon Cloud, pages 459–490. Wiley Press, 2011.

[14] Ying Wang. Nonparametric tests for randomness. Research report, UIUC, May 2003.

[15] Sangho Yi, Derrick Kondo, and Artur Andrzejak. Reducing costs of spot instances via checkpointing in the amazon elastic compute cloud. In *3rd IEEE International Conference on Cloud Computing*, pages 236 –243, 2010.

11 Appendix

11.1 Randomness Testing

As the preliminary phase before modeling, we apply randomness tests to determine which data sets have truly random spot price and the inter-price times. There are several randomness tests which are divided in two general categories: *parametric* and *nonparametric*. Parametric tests are usually utilized when there is an assumption about the distribution of data. As we cannot make any assumption for the underlying distribution of the data sets, we adopted non-parametric randomness tests. We conducted three well-known non-parametric tests, namely the runs test, runs up/down test, and Mann-Kendall test [14, 7]. For all tests, the data was imported as a given sequence or time series with a significance level of 0.05.

11.1.1 Runs Test

The runs test or Wald-Wolfowitz test, is a non-parametric test in which the number of consecutive values in a data samples that are less or greater than mean will be enumerated as runs. These two values are used to make a hypothesis to check the randomness of the data [14]. Also, there is another runs test (i.e., up/down test) that increasing trend (up) or decreasing trend (down) in a data samples are calculated as runs and the same hypothesis like standard runs test will be examined.

11.1.2 Mann-Kendall Test

The Mann-Kendall test is a non-parametric test for identifying trends in time series data. The test compares the relative magnitudes of sample data rather than the data values themselves In this test the sign of consecutive values in a time series are computed and Kendall's tau coefficient is obtained as follows [14]:

$$T = \sum_{i=2}^{n}(\sum_{j=1}^{i-2} sign(X_i - X_j))$$ (4)

where X is a time series and $sign$ function returns 1,-1, and 0 for positive, negative and equal results. The null hypothesis would be as follows:

$$\mid T \mid \geq (z_{1-\alpha/2})\sigma_3$$ (5)

where $\sigma_3 = n(n-1)(2n-15)/18$.

11.1.3 Results

As there is no perfect test for randomness, we decide to apply all tests and to consider only those that pass at least one of three tests. Table 15 and 16 show the p-values of all three randomness, runs standard (rund std), runs up/down (runs ud) and Kendall (Mann-Kendall) tests, for spot price and the inter-price time, respectively. As it can be seen in Table 15, all prices of SIs in each data center pass the Mann-Kendall test except m1.small and m1.large in the us-east. Moreover, m2.xlarge in the us-east passes the run test as well. So, there are some randomness in the price that we can model by some statistical distributions. Table 16 also shows that inter-price times are more random than spot price as they pass more tests. As it is illustrated by the p-values, all instances in all data centers can pass the runs up/down test. Moreover, all instances in us-west and ap-southeast can pass the runs and Mann-Kendall tests as well.

11.2 Distribution Fitting

After randomness testing, we first inspect the distribution using Probability Density Function (PDF) and Cumulative Distribution Function (CDF) for spot price and the inter-price time. Then, we conduct parameter fitting for various distributions, including the Exponential, Weibull, Normal, Log-normal, Gamma. Parameter fitting was conducted using maximum likelihood estimation (MLE). Intuitively, MLE maximizes the log likelihood function that the samples resulted from a distribution with certain parameters.

Table 15. p-values of Randomness Tests for spot price in different data centers (Runs std, Runs ud, Kendall)

Instances	us-west	us-east	eu-west	ap-southeast
m1.small	0.000 0.000 0.354	0.000 0.000 0.021	0.000 0.000 0.705	0.000 0.000 0.652
c1.medium	0.000 0.000 0.216	0.000 0.000 0.613	0.000 0.000 0.236	0.000 0.000 0.478
m1.large	0.000 0.000 0.577	0.000 0.000 0.000	0.000 0.000 0.279	0.000 0.000 0.955
m2.xlarge	0.000 0.000 0.792	0.760 0.000 0.110	0.000 0.000 0.424	0.000 0.000 0.876
m1.xlarge	0.000 0.000 0.348	0.000 0.000 0.596	0.000 0.000 0.491	0.000 0.000 0.499
c1.xlarge	0.000 0.000 0.064	0.000 0.000 0.062	0.000 0.000 0.703	0.000 0.000 0.524
m2.2xlarge	0.000 0.000 0.818	0.000 0.000 0.480	0.000 0.000 0.774	0.000 0.000 0.122
m2.4xlarge	0.000 0.000 0.945	0.000 0.000 0.435	0.000 0.000 0.951	0.000 0.000 0.563

Table 16. p-values of Randomness Tests for the inter-price times in different data centers (Runs std, Runs ud, Kendall)

Instances	us-west	us-east	eu-west	ap-southeast
m1.small	0.418 0.117 0.391	0.000 0.117 0.000	0.000 0.770 0.000	0.132 0.646 0.865
c1.medium	0.173 0.679 0.871	0.000 0.829 0.000	0.000 0.793 0.000	0.906 0.165 0.822
m1.large	0.594 1.000 0.271	0.002 0.854 0.000	0.000 0.914 0.000	0.597 0.678 0.472
m2.xlarge	0.204 0.301 0.790	0.000 0.063 0.000	0.000 0.356 0.000	0.248 0.034 0.551
m1.xlarge	0.994 0.269 0.567	0.000 0.780 0.000	0.000 0.904 0.000	0.641 0.136 0.742
c1.xlarge	0.341 0.239 0.872	0.000 0.584 0.000	0.000 0.399 0.000	0.676 0.129 0.649
m2.2xlarge	0.593 0.563 0.332	0.000 0.724 0.000	0.000 0.955 0.000	0.768 0.733 0.173
m2.4xlarge	0.109 0.969 0.398	0.000 0.944 0.000	0.000 0.894 0.000	0.339 0.500 0.473

Table 17. p-values resulting from KS and AD tests for spot price in us-west data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k = 2$)	MoG ($k = 3$)
m1.small	0.000 0.371	0.040 0.682	0.075 0.760	0.072 0.734	0.024 0.792	0.023 0.794
c1.medium	0.065 0.702	0.072 0.746	0.073 0.750	0.074 0.746	0.200 0.777	0.221 0.794
m1.large	0.000 0.030	0.000 0.140	0.006 0.320	0.002 0.272	0.150 0.682	0.318 0.752
m2.xlarge	0.002 0.293	0.219 0.634	0.196 0.627	0.188 0.625	0.369 0.707	0.363 0.727
m1.xlarge	0.166 0.546	0.143 0.583	0.144 0.581	0.142 0.580	0.374 0.688	0.418 0.713
c1.xlarge	0.000 0.008	0.114 0.392	0.288 0.527	0.239 0.515	0.187 0.599	0.191 0.600
m2.2xlarge	0.170 0.461	0.181 0.506	0.190 0.509	0.176 0.501	0.187 0.497	0.447 0.679
m2.4xlarge	0.182 0.412	0.190 0.435	0.174 0.426	0.185 0.426	0.416 0.618	0.478 0.651

Table 18. p-values resulting from KS and AD tests for spot price in us-east data center.

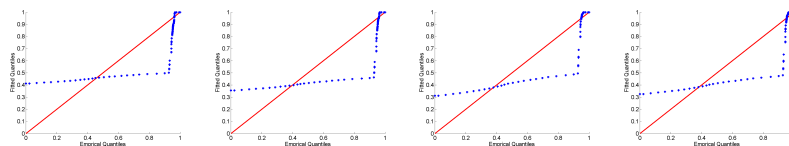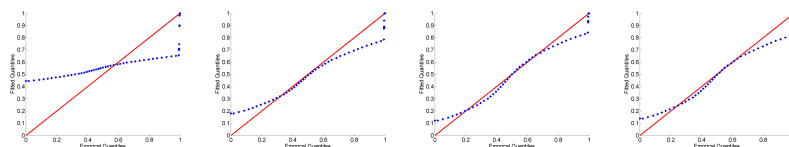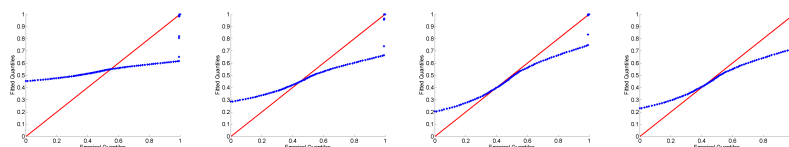| Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k = 2$)	MoG ($k = 3$)
m1.small	0.000 0.476	0.000 0.590	0.000 0.657	0.000 0.632	0.000 0.737	0.000 0.736
c1.medium	0.000 0.281	0.000 0.517	0.003 0.596	0.001 0.571	0.057 0.775	0.056 0.772
m1.large	0.000 0.064	0.000 0.079	0.000 0.155	0.000 0.125	0.167 0.726	0.169 0.727
m2.xlarge	0.000 0.025	0.000 0.042	0.000 0.098	0.000 0.075	0.131 0.691	0.137 0.691
m1.xlarge	0.000 0.006	0.000 0.067	0.007 0.276	0.001 0.203	0.136 0.630	0.131 0.627
c1.xlarge	0.000 0.014	0.000 0.017	0.000 0.041	0.000 0.030	0.185 0.675	0.184 0.674
m2.2xlarge	0.000 0.003	0.019 0.198	0.138 0.410	0.085 0.350	0.166 0.550	0.421 0.681
m2.4xlarge	0.000 0.000	0.000 0.009	0.003 0.060	0.001 0.034	0.182 0.475	0.171 0.460
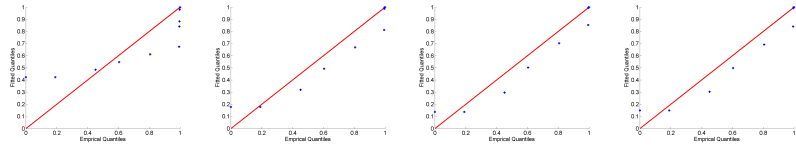
(a) m1.small
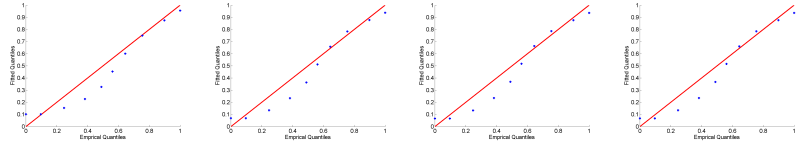
(b) c1.medium
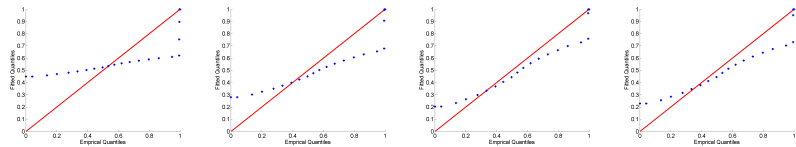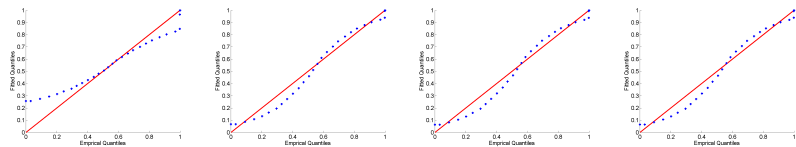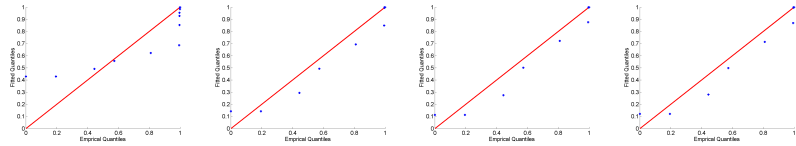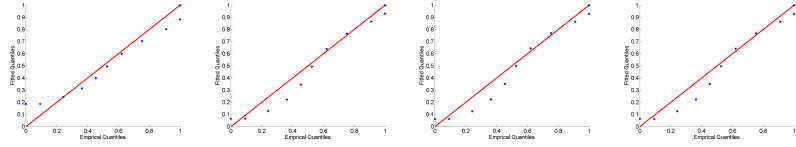
(c) m1.large
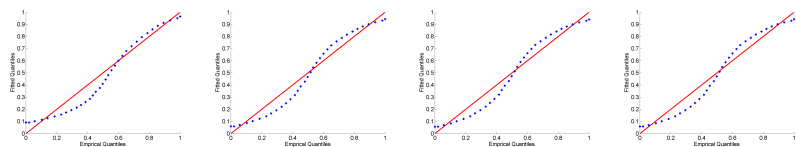
(d) m2.xlarge
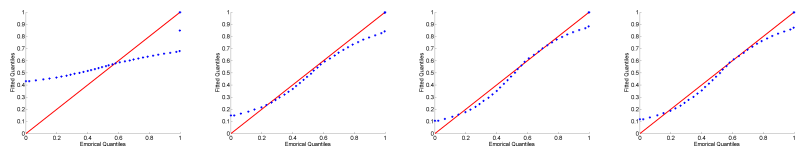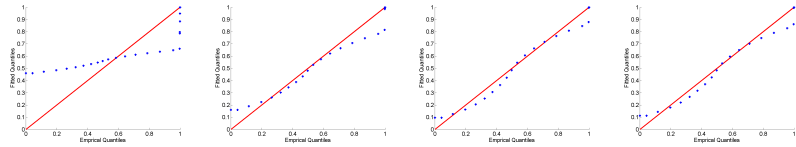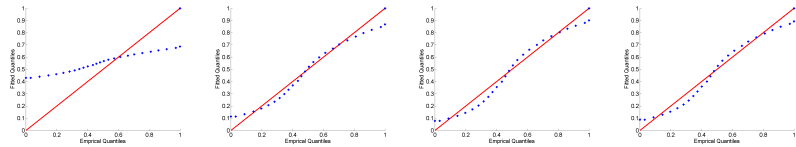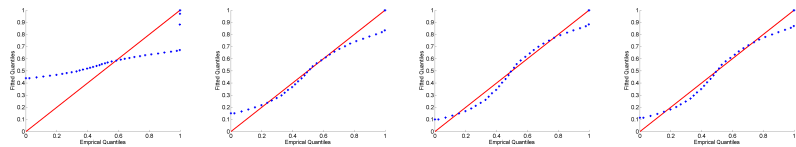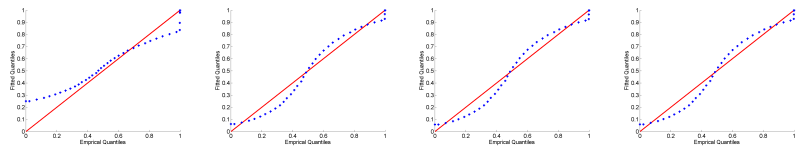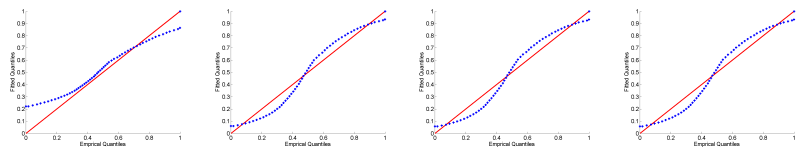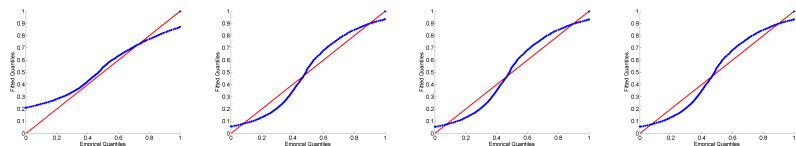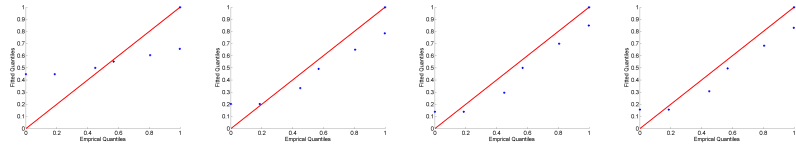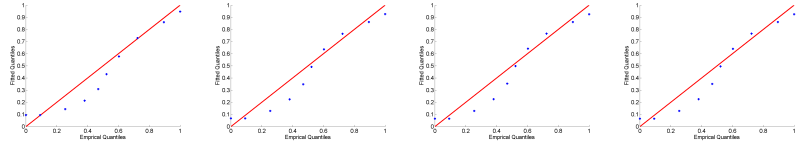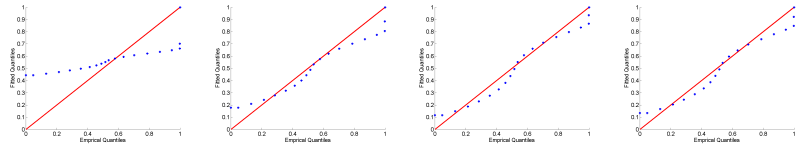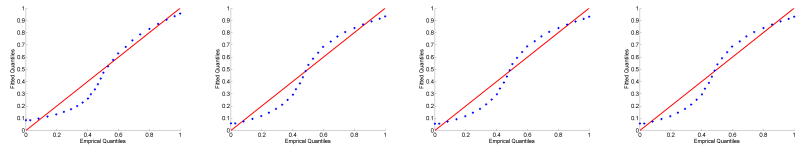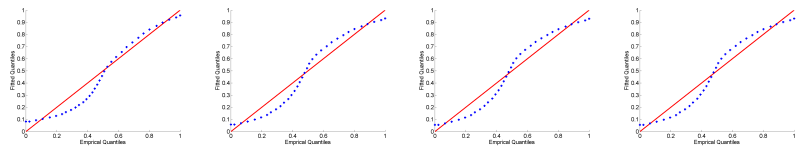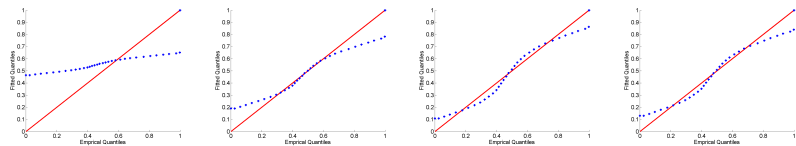
(e) m1.xlarge
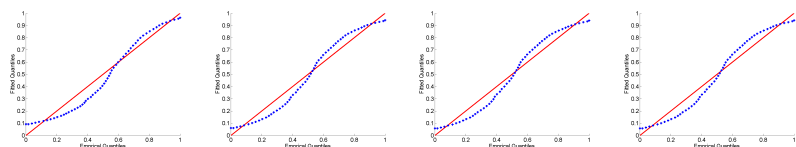
(f) c1.xlarge
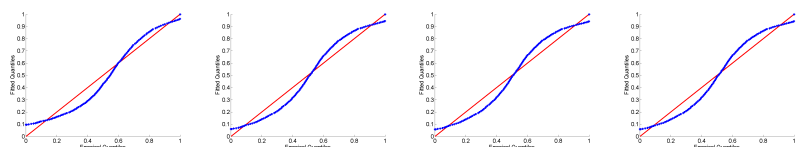
(g) m2.2xlarge

(h) m2.4xlarge

Figure 14. PP-plots of spot price for us-west data center (Weibull, Normal, Log-normal, Gamma)
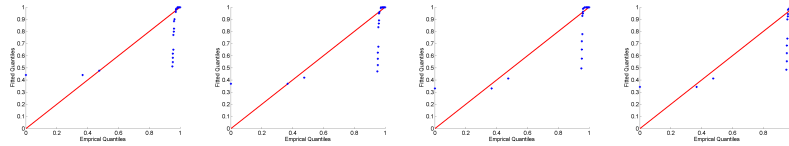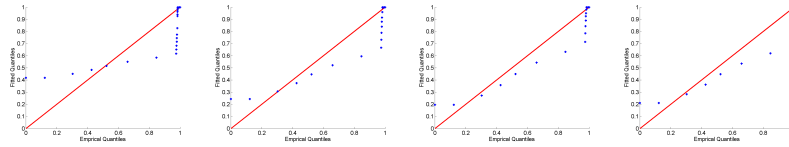
(a) m1.small
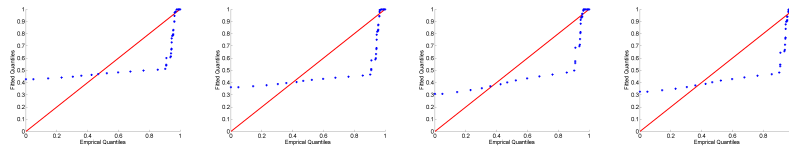
(b) c1.medium
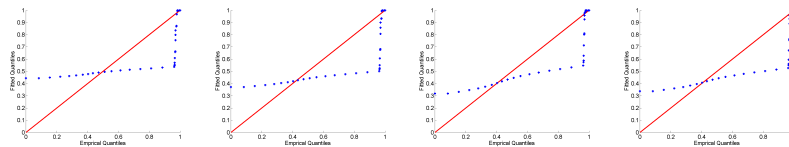
(c) m1.large
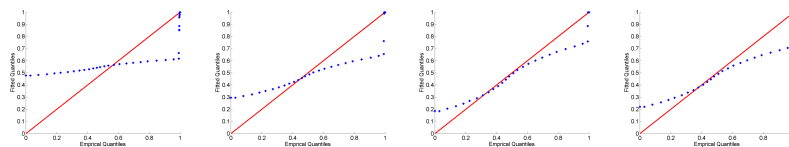
(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge

(g) m2.2xlarge

(h) m2.4xlarge

Figure 15. PP-plots of spot price for us-east data center (Weibull, Normal, Log-normal, Gamma)

26

(a) m1.small

(b) c1.medium

(c) m1.large

(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge

(g) m2.2xlarge

(h) m2.4xlarge

Figure 16. PP-plots of spot price for eu-west data center (Weibull, Normal, Log-normal, Gamma)

(a) m1.small

(b) c1.medium

(c) m1.large

(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge
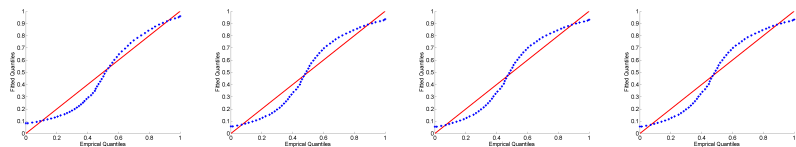
(g) m2.2xlarge
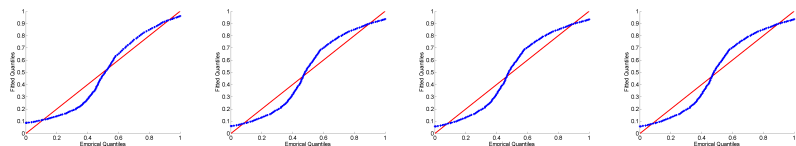
(h) m2.4xlarge

Figure 17. PP-plots of spot price for ap-southeast data center (Weibull, Normal, Log-normal, Gamma)

Table 19. p-values resulting from KS and AD tests for spot price in eu-west data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k = 2$)	MoG ($k = 3$)
m1.small	0.000 0.406	0.069 0.743	0.046 0.773	0.055 0.768	0.016 0.793	0.050 0.800
c1.medium	0.046 0.650	0.078 0.749	0.068 0.745	0.073 0.747	0.210 0.781	0.067 0.740
m1.large	0.000 0.028	0.048 0.480	0.221 0.646	0.186 0.621	0.116 0.679	0.115 0.685
m2.xlarge	0.000 0.015	0.202 0.555	0.231 0.619	0.242 0.593	0.142 0.618	0.132 0.614
m1.xlarge	0.000 0.007	0.089 0.385	0.256 0.540	0.218 0.501	0.143 0.576	0.365 0.702
c1.xlarge	0.002 0.210	0.171 0.576	0.157 0.579	0.163 0.580	0.381 0.675	0.396 0.694
m2.2xlarge	0.010 0.203	0.170 0.490	0.168 0.485	0.164 0.492	0.159 0.476	0.438 0.655
m2.4xlarge	0.015 0.141	0.199 0.441	0.176 0.432	0.180 0.430	0.434 0.613	0.453 0.639

Table 20. p-values resulting from KS and AD tests for spot price in ap-southeast data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k = 2$)	MoG ($k = 3$)
m1.small	0.000 0.341	0.014 0.616	0.058 0.744	0.052 0.714	0.014 0.610	0.012 0.791
c1.medium	0.049 0.706	0.049 0.740	0.048 0.746	0.043 0.738	0.177 0.776	0.201 0.783
m1.large	0.000 0.045	0.033 0.464	0.168 0.604	0.124 0.564	0.078 0.662	0.326 0.750
m2.xlarge	0.115 0.561	0.094 0.588	0.096 0.595	0.101 0.598	0.357 0.719	0.368 0.716
m1.xlarge	0.141 0.533	0.113 0.548	0.116 0.552	0.112 0.549	0.391 0.703	0.418 0.715
c1.xlarge	0.000 0.004	0.012 0.238	0.172 0.488	0.106 0.419	0.101 0.556	0.107 0.556
m2.2xlarge	0.155 0.451	0.128 0.468	0.130 0.466	0.132 0.468	0.410 0.653	0.142 0.478
m2.4xlarge	0.128 0.382	0.108 0.360	0.114 0.379	0.108 0.368	0.441 0.640	0.464 0.644

Table 21. Parameters of some distributions for spot price in us-west data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	1.386 0.047	0.995 3.994 5.903 0.020 2.078	0.003 0.003 5.904 5.904 3.994 2.077 2.077 0.020
c1.medium	2.078 0.033	0.557 7.784 8.254 0.017 0.013	0.374 0.311 7.976 8.304 7.705 0.025 0.006 0.007
m1.large	2.775 0.065	0.006 33.427 15.989 57.477 0.263	0.006 0.545 32.547 15.580 16.483 67.064 0.068 0.049
m2.xlarge	3.177 0.033	0.285 23.089 24.330 0.039 0.418	0.489 0.002 24.651 28.125 23.317 0.143 9.652 0.129
m1.xlarge	3.465 0.032	0.503 32.886 31.075 0.231 0.200	0.396 0.296 31.947 30.814 33.160 0.397 0.070 0.086
c1.xlarge	3.466 0.041	0.002 54.839 31.977 131.688 1.027	0.503 0.002 31.932 54.871 32.022 1.023 131.219 1.028
m2.2xlarge	4.024 0.032	0.554 56.108 55.808 3.179 3.153	0.248 0.474 58.187 56.053 53.860 0.165 1.340 0.192
m2.4xlarge	4.717 0.032	0.504 108.797 115.088 2.577 3.021	0.232 0.309 116.394 107.900 112.355 0.559 0.973 5.524

Table 22. Parameters of some distributions for spot price in us-east data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	1.133 0.156	0.952 3.012 6.009 0.009 3.402	0.476 0.048 3.012 6.009 3.012 0.009 3.402 0.009
c1.medium	1.801 0.070	0.978 6.006 8.916 0.042 2.743	0.978 0.011 6.006 8.916 8.916 0.042 2.743 2.743
m1.large	2.535 0.201	0.094 22.345 12.009 114.856 0.149	0.094 0.472 22.345 12.003 12.015 114.856 0.149 0.149
m2.xlarge	2.860 0.160	0.965 17.020 38.734 0.299 209.065	0.441 0.035 17.021 38.735 17.018 0.299 209.050 0.299
m1.xlarge	3.183 0.063	0.992 24.033 41.511 0.592 425.469	0.486 0.506 24.034 24.032 41.510 0.592 0.592 425.461
c1.xlarge	3.229 0.208	0.929 24.041 51.711 0.594 340.284	0.449 0.071 24.047 51.722 24.035 0.594 340.122 0.593
m2.2xlarge	3.740 0.046	0.994 42.036 61.488 1.807 338.588	0.545 0.449 43.115 40.726 61.084 0.450 0.318 334.571
m2.4xlarge	4.435 0.067	0.993 84.004 166.287 7.064 2455.447	0.496 0.497 83.979 84.029 166.257 7.067 7.059 2457.027

Table 23. Parameters of some distributions for spot price in eu-west data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	1.386 0.042	0.995 3.997 5.216 0.020 1.671	0.007 0.538 5.018 4.111 3.861 1.361 0.006 0.003
c1.medium	2.080 0.033	0.589 8.194 7.734 0.028 0.010	0.500 0.001 8.003 10.000 8.003 0.069 0.007 0.069
m1.large	2.774 0.040	0.999 16.021 29.514 0.266 143.651	0.516 0.482 16.004 16.039 29.487 0.266 0.265 143.694
m2.xlarge	3.179 0.037	0.001 53.500 24.029 12.960 0.593	0.000 0.999 53.500 24.029 53.500 12.960 0.593 12.960
m1.xlarge	3.466 0.040	0.002 53.766 32.006 328.180 1.058	0.539 0.002 32.851 53.802 31.013 0.246 326.527 0.186
c1.xlarge	3.466 0.033	0.329 30.836 32.625 0.091 0.594	0.529 0.469 32.894 31.051 37.283 0.229 0.211 18.992
m2.2xlarge	4.025 0.033	0.598 55.899 56.243 3.233 3.478	0.350 0.325 53.971 57.149 57.149 0.290 1.479 1.479
m2.4xlarge	4.719 0.032	0.641 114.338 108.037 5.528 1.270	0.488 0.250 112.188 107.611 116.127 7.057 0.662 0.896

Table 24. Parameters of some distributions for spot price in ap-southeast data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	1.386 0.047	0.500 4.005 4.005 0.061 0.061	0.999 0.001 3.998 9.750 9.750 0.020 0.063 0.063
c1.medium	2.080 0.034	0.492 8.253 7.765 0.012 0.015	0.309 0.350 8.000 8.300 7.708 0.026 0.006 0.007
m1.large	2.773 0.044	0.999 16.002 38.000 0.278 0.000	0.488 0.001 16.494 38.000 15.532 0.043 0.000 0.052
m2.xlarge	3.179 0.033	0.462 23.264 24.682 0.098 0.120	0.232 0.232 23.266 23.266 24.684 0.099 0.099 0.119
m1.xlarge	3.467 0.033	0.456 31.016 32.904 0.185 0.217	0.195 0.297 30.664 31.376 32.965 0.029 0.195 0.172
c1.xlarge	3.468 0.043	0.001 74.207 32.049 646.887 1.073	0.479 0.520 31.963 32.128 74.166 1.075 1.057 647.976
m2.2xlarge	4.026 0.032	0.525 57.608 54.339 0.616 0.648	0.341 0.354 56.113 56.214 55.812 3.281 3.241 3.287
m2.4xlarge	4.718 0.033	0.548 115.029 108.435 2.838 1.972	0.317 0.362 107.803 115.878 111.917 0.849 1.149 4.709

11.2.1 inter-price time

Table 25. p-values resulting from KS and AD tests for the inter-price time in us-west data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k = 2$)	MoG ($k = 3$)
m1.small	0.057 0.199	0.008 0.062	0.113 0.339	0.100 0.256	0.086 0.271	0.339 0.576
c1.medium	0.045 0.175	0.007 0.058	0.101 0.312	0.083 0.233	0.076 0.267	0.376 0.581
m1.large	0.030 0.146	0.005 0.050	0.075 0.284	0.057 0.192	0.063 0.242	0.361 0.584
m2.xlarge	0.026 0.134	0.004 0.045	0.074 0.276	0.058 0.191	0.054 0.217	0.360 0.579
m1.xlarge	0.040 0.165	0.009 0.068	0.069 0.283	0.064 0.219	0.073 0.251	0.350 0.586
c1.xlarge	0.037 0.157	0.007 0.054	0.091 0.300	0.071 0.208	0.065 0.258	0.349 0.577
m2.2xlarge	0.033 0.152	0.007 0.056	0.077 0.288	0.065 0.210	0.068 0.238	0.361 0.585
m2.4xlarge	0.033 0.146	0.008 0.059	0.081 0.287	0.057 0.195	0.051 0.244	0.380 0.591
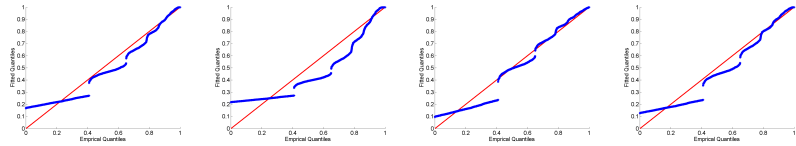
(a) m1.small
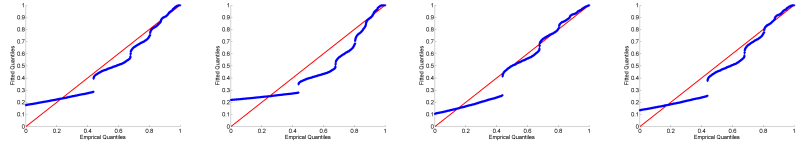
(b) c1.medium
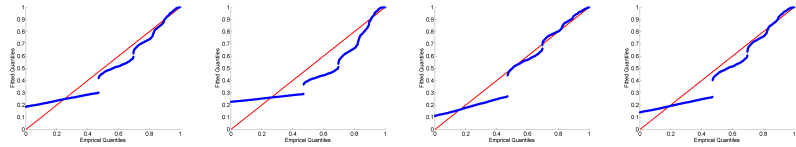
(c) m1.large

(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge

(g) m2.2xlarge

(h) m2.4xlarge

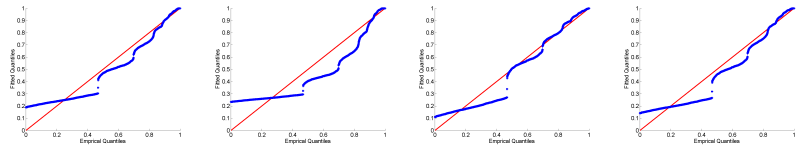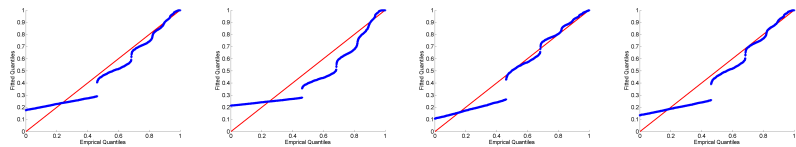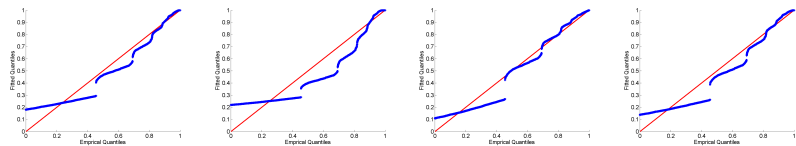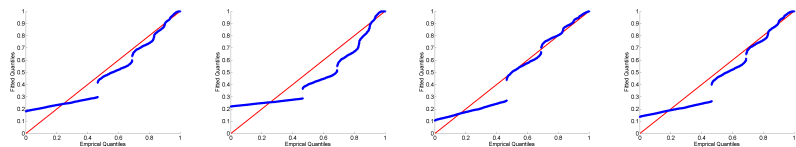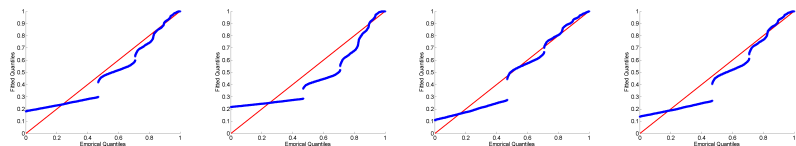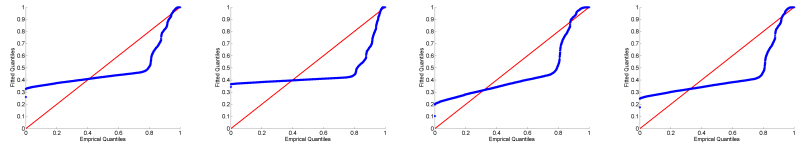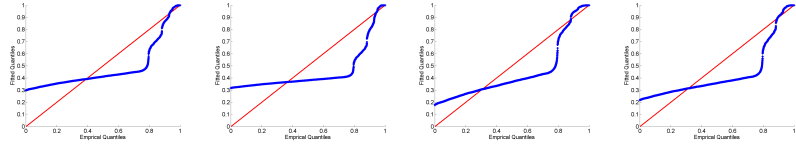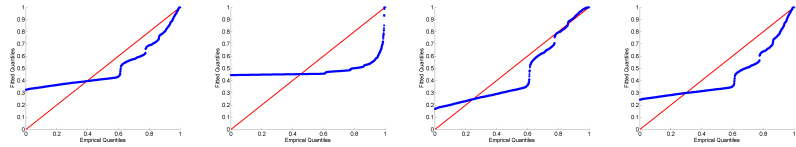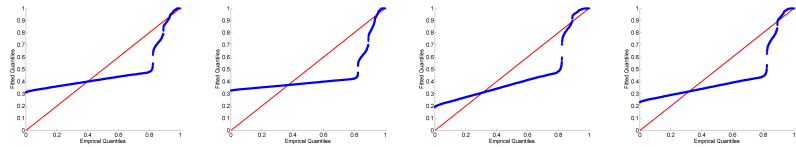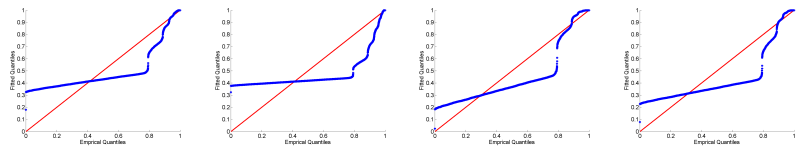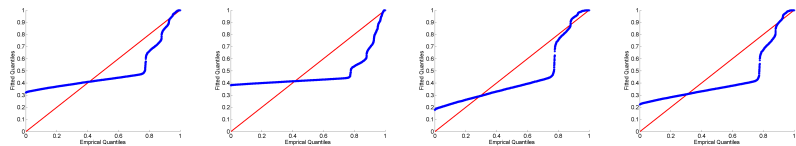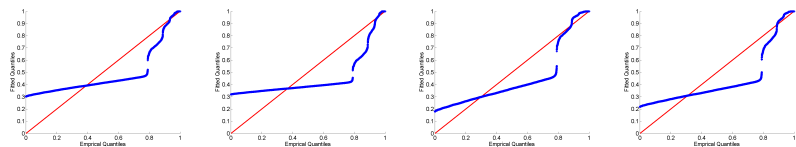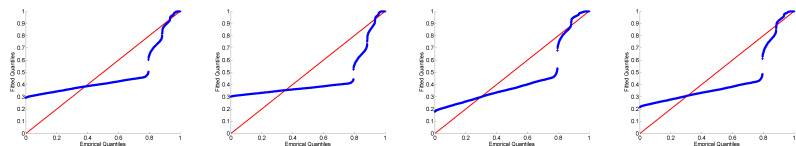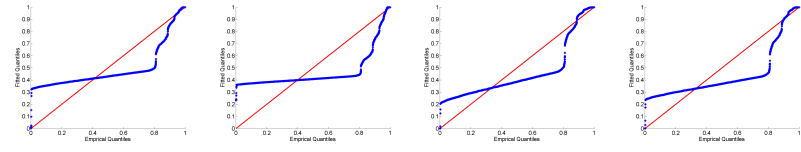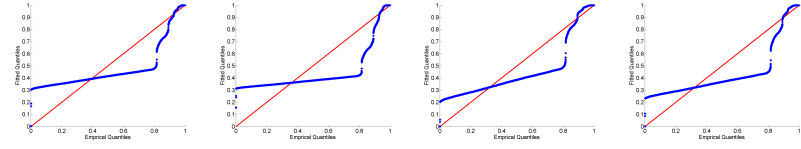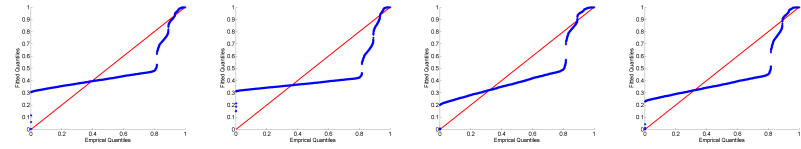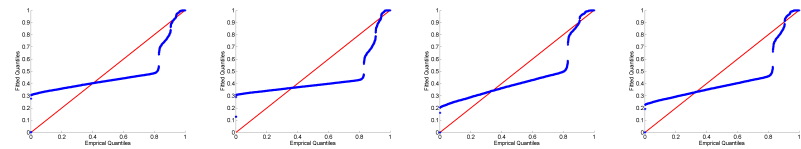**Figure 18. PP-plots of the inter-price time for us-west data center (Weibull, Normal, Log-normal, Gamma)**

(a) m1.small

(b) c1.medium

(c) m1.large

(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge

(g) m2.2xlarge

(h) m2.4xlarge

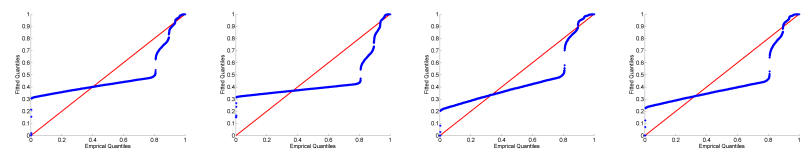**Figure 19. PP-plots of the inter-price time for us-east data center (Weibull, Normal, Log-normal, Gamma)**

(a) m1.small

(b) c1.medium

(c) m1.large

(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge

(g) m2.2xlarge

(h) m2.4xlarge

Figure 20. PP-plots of the inter-price time for eu-west data center (Weibull, Normal, Log-normal, Gamma)
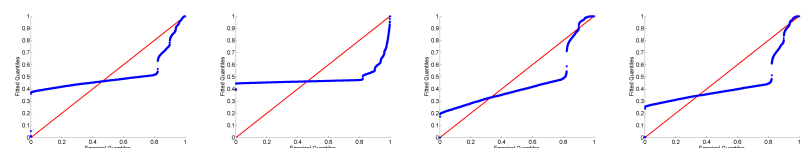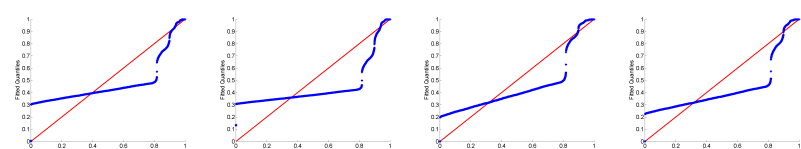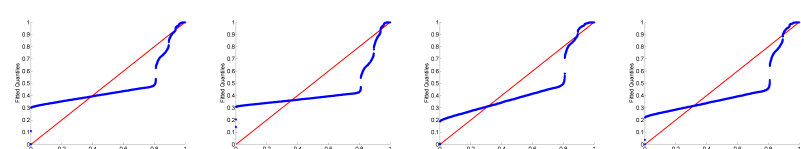
(a) m1.small

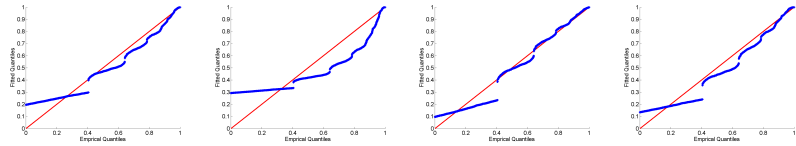(b) c1.medium

(c) m1.large

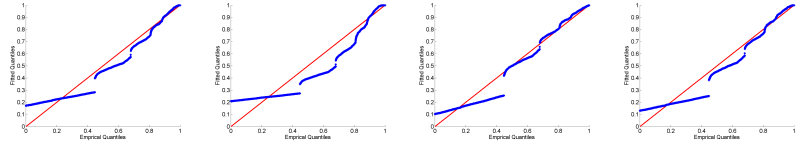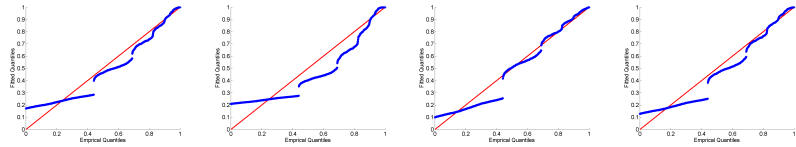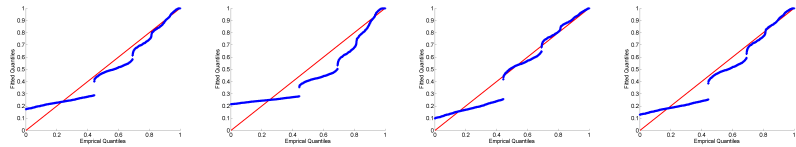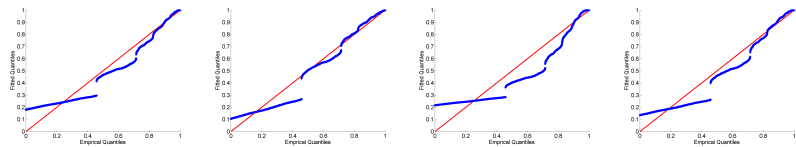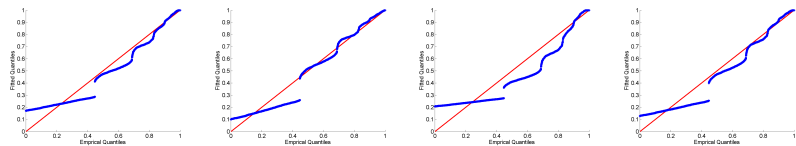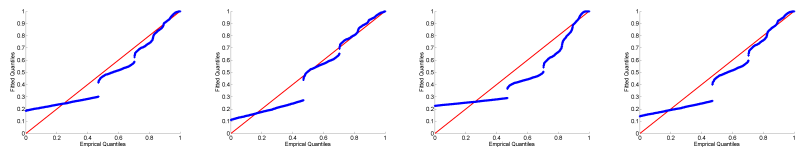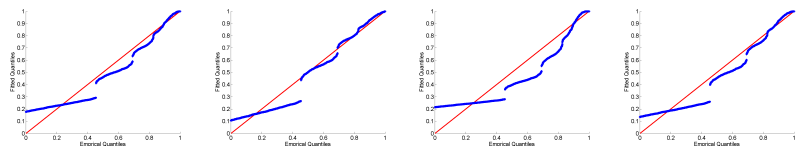(d) m2.xlarge

(e) m1.xlarge

(f) c1.xlarge

(g) m2.2xlarge

(h) m2.4xlarge

Figure 21. PP-plots of the inter-price time for ap-southeast data center (Weibull, Normal, Log-normal, Gamma)

Table 26. p-values resulting from KS and AD tests for the inter-price time in us-east data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k=2$)	MoG ($k=3$)
m1.small	0.000 0.003	0.000 0.000	0.001 0.018	0.000 0.006	0.356 0.465	0.425 0.586
c1.medium	0.000 0.006	0.000 0.002	0.002 0.030	0.001 0.013	0.385 0.525	0.454 0.609
m1.large	0.000 0.010	0.000 0.000	0.015 0.101	0.002 0.017	0.005 0.053	0.341 0.520
m2.xlarge	0.000 0.004	0.000 0.001	0.001 0.020	0.000 0.008	0.372 0.525	0.384 0.571
m1.xlarge	0.000 0.004	0.000 0.000	0.002 0.030	0.000 0.012	0.257 0.377	0.399 0.564
c1.xlarge	0.000 0.005	0.000 0.000	0.003 0.034	0.000 0.013	0.241 0.338	0.390 0.560
m2.2xlarge	0.000 0.006	0.000 0.002	0.002 0.030	0.000 0.015	0.357 0.501	0.410 0.585
m2.4xlarge	0.000 0.007	0.000 0.003	0.001 0.027	0.000 0.014	0.372 0.531	0.399 0.573

Table 27. p-values resulting from KS and AD tests for the inter-price time in eu-west data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k=2$)	MoG ($k=3$)
m1.small	0.000 0.005	0.000 0.001	0.001 0.024	0.000 0.010	0.356 0.492	0.408 0.586
c1.medium	0.000 0.005	0.000 0.002	0.001 0.022	0.000 0.010	0.385 0.549	0.377 0.540
m1.large	0.000 0.005	0.000 0.002	0.001 0.022	0.000 0.011	0.377 0.549	0.385 0.564
m2.xlarge	0.000 0.005	0.000 0.002	0.001 0.022	0.000 0.011	0.393 0.563	0.396 0.570
m1.xlarge	0.000 0.007	0.000 0.002	0.002 0.027	0.000 0.014	0.369 0.524	0.400 0.580
c1.xlarge	0.000 0.002	0.000 0.000	0.001 0.023	0.000 0.006	0.225 0.320	0.393 0.560
m2.2xlarge	0.000 0.006	0.000 0.002	0.001 0.025	0.000 0.012	0.393 0.554	0.395 0.567
m2.4xlarge	0.000 0.006	0.000 0.002	0.002 0.027	0.000 0.012	0.372 0.532	0.391 0.574

Table 28. p-values resulting from KS and AD tests for the inter-price time in ap-southeast data center.

Instances	Weibull	Normal	Log-Normal	Gamma	MoG ($k=2$)	MoG ($k=3$)
m1.small	0.027 0.150	0.000 0.013	0.121 0.357	0.107 0.255	0.075 0.223	0.299 0.543
c1.medium	0.048 0.186	0.011 0.070	0.091 0.310	0.078 0.237	0.067 0.271	0.372 0.593
m1.large	0.050 0.188	0.011 0.068	0.095 0.326	0.084 0.248	0.066 0.272	0.314 0.567
m2.xlarge	0.045 0.181	0.009 0.067	0.091 0.311	0.082 0.244	0.062 0.253	0.337 0.585
m1.xlarge	0.035 0.153	0.007 0.058	0.089 0.297	0.066 0.211	0.055 0.258	0.416 0.569
c1.xlarge	0.048 0.189	0.012 0.075	0.077 0.304	0.080 0.249	0.063 0.257	0.356 0.573
m2.2xlarge	0.029 0.142	0.005 0.050	0.079 0.278	0.058 0.192	0.052 0.229	0.369 0.583
m2.4xlarge	0.041 0.165	0.009 0.062	0.080 0.292	0.070 0.213	0.068 0.255	0.352 0.581

Table 29. Parameters of distributions for the inter-price time in us-west data center.

Instances	LogN(μ, σ)	MoG($k=2, p, \mu, \sigma$)	MoG($k=3, p, \mu, \sigma$)
m1.small	0.914 0.695	0.395 5.542 1.769 11.667 0.495	0.388 0.402 1.258 3.104 7.263 0.020 0.896 14.376
c1.medium	0.859 0.681	0.390 5.222 1.676 9.669 0.409	0.246 0.424 6.439 1.247 2.873 10.758 0.020 0.461
m1.large	0.820 0.665	0.388 4.986 1.613 9.254 0.344	0.386 0.451 3.102 1.260 7.103 0.882 0.022 12.582
m2.xlarge	0.824 0.669	0.366 5.183 1.661 10.963 0.397	0.342 0.454 2.947 1.258 6.720 0.525 0.019 13.498
m1.xlarge	0.831 0.662	0.573 1.570 4.757 0.309 7.533	0.160 0.443 6.983 1.257 3.163 9.983 0.020 0.951
c1.xlarge	0.836 0.674	0.379 5.155 1.651 8.875 0.384	0.330 0.441 2.879 1.249 6.468 0.472 0.021 9.736
m2.2xlarge	0.823 0.660	0.587 1.587 4.803 0.323 8.387	0.147 0.446 7.228 1.258 3.171 12.038 0.020 0.955
m2.4xlarge	0.811 0.655	0.635 1.646 4.985 0.374 7.630	0.314 0.229 2.828 6.129 1.256 0.389 8.098 0.021

Table 30. Parameters of distributions for the inter-price time in us-east data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	0.499 0.586	0.200 5.877 1.303 45.521 0.026	0.793 0.164 1.301 3.582 13.941 0.025 2.641 120.277
c1.medium	0.489 0.523	0.780 1.302 4.452 0.023 11.976	0.780 0.075 1.301 7.381 2.954 0.023 20.605 0.815
m1.large	0.701 0.722	0.079 21.921 2.002 3995.166 1.490	0.390 0.023 3.986 55.662 1.277 4.862 12186.399 0.022
m2.xlarge	0.450 0.511	0.814 1.283 4.676 0.023 12.447	0.814 0.038 1.282 9.249 3.508 0.023 27.356 1.809
m1.xlarge	0.468 0.514	0.783 1.279 4.436 0.022 34.923	0.778 0.206 1.278 3.655 13.907 0.022 3.005 353.820
c1.xlarge	0.488 0.527	0.235 4.432 1.281 41.110 0.022	0.016 0.759 14.300 1.280 3.651 459.734 0.021 3.196
m2.2xlarge	0.467 0.503	0.777 1.279 4.150 0.021 10.468	0.180 0.041 3.230 8.233 1.278 1.355 29.767 0.021
m2.4xlarge	0.470 0.504	0.218 4.190 1.285 7.471 0.022	0.063 0.151 7.043 3.065 1.285 11.974 0.862 0.022

Table 31. Parameters of distributions for the inter-price time in eu-west data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	0.458 0.549	0.798 1.293 4.616 0.022 26.018	0.178 0.028 3.474 11.536 1.292 2.308 120.054 0.022
c1.medium	0.447 0.537	0.197 4.472 1.280 9.274 0.022	0.181 0.623 1.307 1.272 4.472 0.022 0.022 9.274
m1.large	0.441 0.522	0.196 4.333 1.277 8.707 0.022	0.068 0.126 6.793 3.040 1.276 13.803 0.940 0.022
m2.xlarge	0.421 0.498	0.184 4.054 1.284 6.823 0.022	0.110 0.066 5.166 2.506 1.284 8.192 0.035 0.022
m1.xlarge	0.440 0.526	0.207 4.056 1.283 9.183 0.022	0.793 0.177 1.283 3.310 8.354 0.022 1.863 29.717
c1.xlarge	0.440 0.503	0.819 1.287 5.160 0.022 275.687	0.187 0.002 4.048 84.632 1.286 5.344 15666.039 0.022
m2.2xlarge	0.430 0.498	0.804 1.276 4.036 0.022 6.804	0.119 0.070 5.107 2.498 1.276 8.195 0.030 0.022
m2.4xlarge	0.445 0.501	0.203 4.133 1.280 7.844 0.022	0.063 0.137 6.705 3.001 1.279 13.525 0.863 0.022

Table 32. Parameters of distributions for the inter-price time in ap-southeast data center.

Instances	LogN(μ, σ)	MoG($k = 2, p, \mu, \sigma$)	MoG($k = 3, p, \mu, \sigma$)
m1.small	0.917 0.696	0.116 9.558 2.528 97.726 2.113	0.377 0.550 1.263 3.657 11.664 0.021 2.310 139.524
c1.medium	0.848 0.664	0.408 4.935 1.632 7.604 0.349	0.433 0.334 1.263 2.908 6.273 0.021 0.519 8.407
m1.large	0.840 0.648	0.388 4.891 1.683 7.607 0.385	0.419 0.412 1.263 3.034 6.782 0.022 0.816 9.503
m2.xlarge	0.842 0.653	0.383 4.975 1.690 8.459 0.402	0.154 0.420 7.015 1.261 3.142 11.863 0.019 1.034
m1.xlarge	0.816 0.646	0.351 5.044 1.684 7.998 0.383	0.323 0.229 5.344 2.496 1.266 7.589 0.038 0.020
c1.xlarge	0.831 0.648	0.402 4.751 1.652 7.093 0.383	0.211 0.354 6.188 2.949 1.254 8.230 0.502 0.021
m2.2xlarge	0.817 0.664	0.370 5.076 1.638 9.204 0.369	0.333 0.212 2.907 6.466 1.255 0.517 10.745 0.020
m2.4xlarge	0.829 0.664	0.610 1.631 4.970 0.359 7.937	0.439 0.358 1.250 2.936 6.556 0.019 0.593 8.991