

Carl, David; Ewerhart, Christian

**Working Paper**

## Ethereum gas price statistics

Working Paper, No. 373

**Provided in Cooperation with:**

Department of Economics, University of Zurich

*Suggested Citation:* Carl, David; Ewerhart, Christian (2020) : Ethereum gas price statistics, Working Paper, No. 373, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-194455>

This Version is available at:

<https://hdl.handle.net/10419/228872>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**University of  
Zurich**<sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series  
ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 373

# **Ethereum Gas Price Statistics**

David Carl and Christian Ewerhart

December 2020

---

# Ethereum Gas Price Statistics

David Carl<sup>†</sup>

Christian Ewerhart<sup>‡</sup>

December 22, 2020

**Abstract.** For users of the Ethereum network, the gas price is a crucial parameter that determines how swiftly the decentralized consensus protocol confirms a transaction. This paper studies the statistics of the Ethereum gas price. We start with some conceptual discussion of the gas price notion in view of the actual transaction-selection strategies used by Ethereum miners. Subsequently, we provide the descriptive statistics of what we call the threshold gas price. Finally, we identify and estimate a seasonal ARIMA (SARIMA) model for predicting the hourly median of the threshold gas price.

**Keywords.** Ethereum, Gas Price, Confirmation time, Forecasting.

**JEL codes.** C57 (Econometrics of Games and Auctions), C22 (Time-Series Models • Dynamic Quantile Regressions • Dynamic Treatment Effect Models • Diffusion Processes), D85 (Network Formation and Analysis: Theory), E37 (Forecasting and Simulation: Models and Applications)

---

<sup>†</sup>D. Carl (dcarl@student.ethz.ch) is a M.Sc. candidate in Statistics at ETH Zurich, Switzerland.

<sup>‡</sup>C. Ewerhart (christian.ewerhart@econ.uzh.ch) is Professor of Economics at University of Zurich, Switzerland.

## 1. Introduction

Among the various cryptocurrency platforms that offer convenient accounting and settlement functionalities, *Ethereum* stands out as the most popular public blockchain platform capable of storing and executing smart contracts.<sup>1</sup> Users interact with the platform for a variety of reasons, such as transferring ether or tokens, calling a function of a smart contract, or deploying a new instance of a smart contract. In any of these interactions, the Ethereum protocol requests the user to set a specific gas price, usually denominated in Gwei.<sup>2</sup> Subsequently, the thereby priced transaction enters a pool of pending transactions. Depending on the transaction-selection strategies implemented by individual miners, the transaction may or may not be included into the next block. However, given that the block reward for the successful miner depends positively on the gas price of included transactions, the gas price is crucial for determining how swiftly the network confirms the transaction. It is, therefore, desirable for users of the Ethereum network to be able to predict the gas price level necessary for obtaining a confirmation within a specified period of time and with a certain probability of confidence.

This paper studies the Ethereum gas price from a statistical perspective. The analysis has three main parts. We start by reviewing the institutional background of the market for Ethereum transactions. In particular, this review entails an examination of the transaction-selection strategies used by Ethereum miners, and the conceptualization of what we call the *threshold gas price* for a block of transactions. Next, we provide the descriptive statistics of the threshold gas price over the considered time interval (January 2018 to November 2019). As the market for Ethereum transactions is far from perfect, we present our findings as a collection of stylized facts. Finally, we develop an empirical model of the Ethereum gas price. Specifically, based on a variety of tests, we identify a seasonal autoregressive integrated moving average (SARIMA) model as a statistical representation for the hourly median of the threshold gas price. Returning to the motivation for the analysis, we evaluate the predictive capability of the model, and discuss how well it serves the purpose of determining a “safe” gas price level for Ethereum transactions.

The literature on blockchain transaction fees and mining strategies is still small. Regarding the Ethereum network, we found three related papers. Pierro and Rocha<sup>3</sup> studied the determinants of transaction fees and evaluated their predictive power using Granger causality. Werner et al.<sup>4</sup> used a recurrent neural network to forecast the minimum gas price. Finally, de Azevedo Sousa et al.<sup>5</sup> estimated Pearson and Spearman coefficients to measure the correlation between gas price and pending time, reporting non-significant values. However, prioritization by miners has not been adjusted for and the gas price itself is potentially a lagged endogenous variable (as longer expected pending time may lead to higher gas prices which a simple correlation cannot account for). Regarding bitcoin, Möser and Böhme<sup>6</sup> concluded that higher transaction fees tend to lead to faster confirmation. However, they were also observant of the heterogeneity and instability of transaction fees. Koops<sup>7</sup> studied the distribution of confirmation times of bitcoin transactions. A game-theoretic model involving both miners and network users has been proposed by Easley et al.<sup>8</sup> As far as we can tell, however, the research pursued in the present paper, viz. the identification of a statistical model for the economically relevant threshold gas price, has not been attempted before.

The remainder of this paper has the following structure. Section 2 provides some institutional background. Section 3 presents the descriptive statistics of the threshold gas price. Section 4 discusses the identification and estimation of our statistical model. Section 5 concludes. An Appendix provides additional background and analysis.

## **2. The market for Ethereum transactions**

This section prepares the main analysis. We first provide some background on Ethereum transactions. We then study miners' transaction-selection strategies. Subsequently, we introduce our definition of the Ethereum gas price, and finally describe the data set.

### ***2.1 Transactions and blocks***

Successful transactions correspond to atomic status changes of the Ethereum network.<sup>9</sup> All miners in the Ethereum network continuously receive authorized transactions into the

*pending pool*. From there, miners consecutively add transactions into their block. By doing so, they take part in the decentralized proof-of-work consensus protocol *Ethash* that allows miners to invest computational power to find a valid hash. The successful miner earns a reward that is equivalent to a sum of block (and uncle) rewards and transaction fees, where the latter corresponds, for any given transaction, to the product of gas price and the amount of gas actually used in the execution of that transaction. Gas is a dimensionless quantity, meaning that there is no particular name for one unit. A new block is mined on average every  $\Delta t = 15$  seconds.

Among the data fields that comprise an Ethereum transaction, the following three are of particular relevance for the subsequent analysis:

- The *gas price* specifies the amount of Ether (or Gwei) that the creator is willing to pay for gas.
- The *gas limit* is the maximum amount of gas that the creator is willing to spend in order to have the transaction executed.
- The *nonce* is a counter for the transactions originating from the same externally owned address (EOA).

## **2.2 Miners' transaction selection strategies**

When adding transactions from the pending pool to their block, a miner must comply with two main constraints. First, a miner may add a transaction only if all transactions of the same EAO with lower nonces have already been taken account of (either in an ancestor block or in already included transactions). Second, the sum of the actual gas used by all the transactions in the block must not exceed the block gas limit presently set for the Ethereum network. We define three classes of transaction-selection strategies.

- *Default strategies without prioritization* have the property that any transaction A added to a block must carry a weakly lower gas price than the previously added

transaction, unless there is a transaction B already added to the block initiated by the same EOA as transaction A.

- *Default strategies with prioritization* have the property that transactions initiated by one specific EOA are added to the block before any transaction from different EOAs. After all transactions initiated by the specific EOA have been included in the block, the transactions initiated by other EOAs are added following a transaction-selection strategy from the class default strategies without prioritization. The blocks we assigned to the class default strategies with prioritization were additionally classified into two subclasses:
  - blocks in which less than a quarter of the transactions were initiated by the prioritized EOA, and
  - blocks in which more than a quarter of the transactions were initiated by the prioritized EOA.
- *Other strategies* have the property that they neither classify as default strategies without prioritization nor as default strategies with prioritization.

Default strategies with and without prioritization roughly correspond to the standard transaction-selection strategies suggested by Geth and Parity. Standard strategies first construct a list of pending transactions for each EOA. Then, having accounted for any prioritized EOA(s), the transaction with the highest gas price of the transactions with the lowest nonce in each list is committed to the block. This process is iterated until either the block gas limit is reached or other conditions are met.

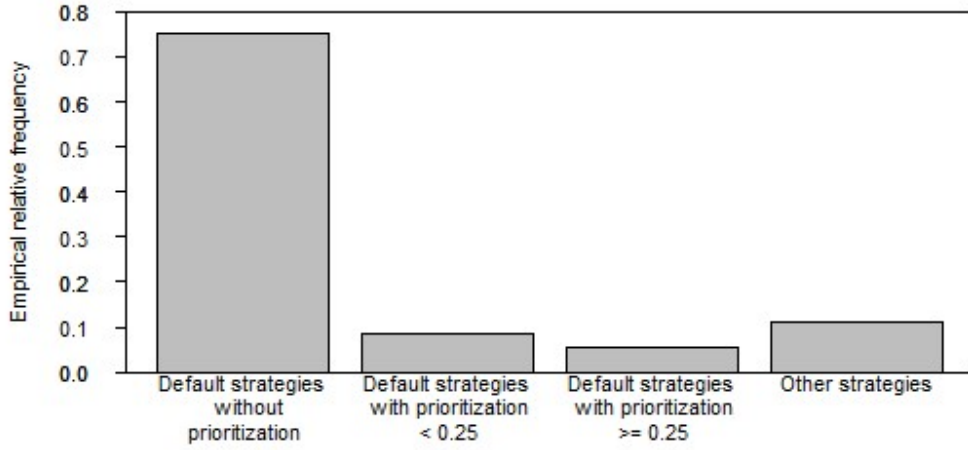


Fig. 1. Empirical distribution of the transaction-selection strategies.

Fig. 1 shows the empirical distribution of these transaction-selection strategies in a random sample of 1'000 blocks mined during the observed period.<sup>10</sup> As can be seen, about 75 percent of the blocks were mined using default strategies without prioritization, 14 percent using default strategies with prioritization, and 11 percent using other transaction-selection strategies. It should be noted, however, that at least a fraction of these strategies might be gas-price based as well. Miners may prioritize transactions from more than one EOA, and if there are not enough valid transactions to fill up the block from the pending pool in an initial round, a second set of transactions may be added if new transactions arrive in the meantime. In fact, during our analysis we rarely encountered a block in which the transactions were not ordered by their gas price at least to some extent. Still, we may conclude that, in only three quarters of the blocks, the minimum gas price properly reflects the economic price for being included in the block.

### ***2.3 Definition of the threshold gas price***

To account for the heterogeneity in miners' transaction-selection strategies, we introduce the following concepts:



- The *minimum gas price* of the block,  $P_{\min}$ , is the gas price of the transaction with the lowest gas price in the block.
- The *first-quartile gas price* of the block,  $P_{\text{gq1}}$ , corresponds to the gas price such that one quarter of the transactions in the block have a gas price weakly below  $P_{\text{gq1}}$ , while three quarters of the transactions in the block have a gas price weakly above  $P_{\text{gq1}}$ .
- The *threshold gas price* of a block,  $P_{\text{gas}}$ , corresponds to the minimum gas price if the minimum gas price is strictly larger than 1 Gwei, and to the first-quartile gas price otherwise, i.e.,

$$P_{\text{gas}} = \begin{cases} P_{\min} & \text{if } P_{\min} > 1 \text{ Gwei} \\ P_{\text{gq1}} & \text{if } P_{\min} \leq 1 \text{ Gwei.} \end{cases}$$

The case distinction in the definition of the threshold gas price reflects that a non-negligible subpopulation of miners uses transaction-selection strategies with prioritization. Thus, if the minimum gas price is too low to be economically plausible, we replace it by the first-quartile gas price.

## 2.4 The data set

We used transaction and block data recorded in the Ethereum mainnet blockchain in the period between January 30, 2018 and October 24, 2019. The corresponding block numbers were 5'000'000 and 8'800'000. The transaction and block data cover 98.5 % of the blocks mined during that period. The data was retrieved via the Python script `ethereum-etl` and the `web3` library.<sup>11</sup> During the considered period, individual blocks consisted of about 50-160 transactions in the majority of cases, with an average of 111 transactions. For the reader's

convenience, both the data set and the R code employed in the analysis has been made available on Github.<sup>12</sup>

### 3. Empirical observations

This section presents the descriptive statistics of the threshold gas price, thereby documenting three main empirical observations on the market for Ethereum transactions. Specifically, we observed a very inelastic short-term price elasticity, behavioral pricing strategies, and a pronounced (European) day-night activity pattern.

#### 3.1 Short-term inelastic demand

Fig. 2 shows the development of the threshold gas price in the considered period. The solid line represents a locally estimated scatterplot smoothing (LOESS). The mean of the threshold gas price over the considered period has been 6.6 Gwei. The standard deviation has been 19.9 Gwei and, thus, more than three times the mean. This remarkable volatility is highly asymmetric in nature, as reflected by the fact that the mean has been more than 50 percent higher than the median, which has been 3.0 Gwei. In fact, the threshold gas price has often dropped to nearly zero. Given that the supply for Ethereum transactions is

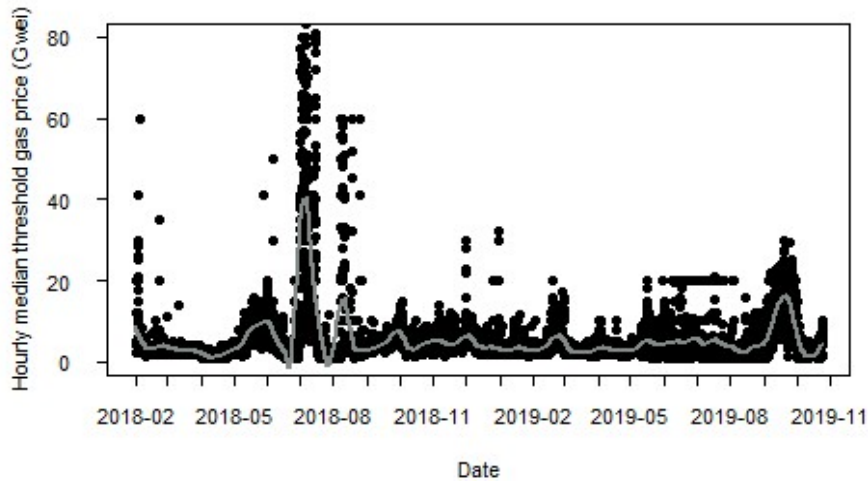


Fig. 2. Time series of the threshold gas price during the considered period. Each dot represents an hourly median. The grey line shows a smoothed development of the gas price.

effectively bounded by the block size limit, this hints towards a very inelastic short-term demand. Ethereum users, therefore, may have been excessively impatient during the considered period.

### 3.2 Behavioral pricing

Fig. 3 shows the relative frequencies of threshold gas prices in Gwei, rounded to integer values. The peaks at 10 Gwei and 20 Gwei, respectively, are close to the median (12.0 Gwei) resp. mean (18.3 Gwei) of the gas prices over all transactions in a block over the observed

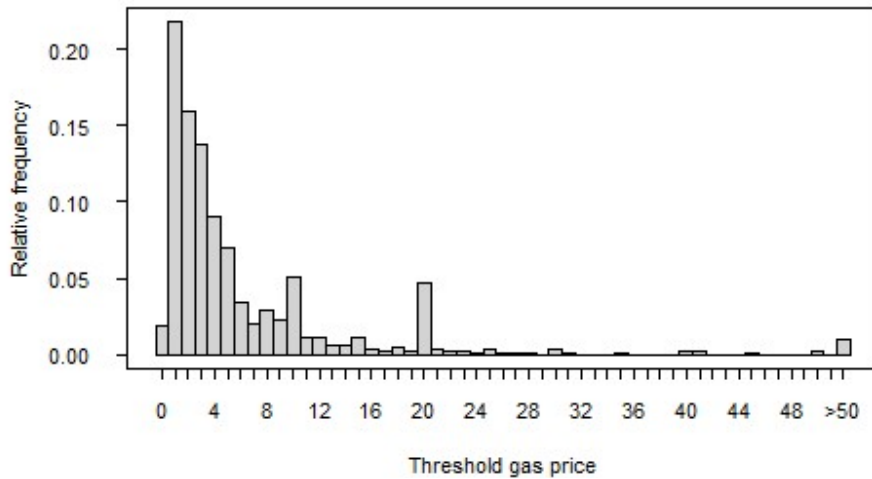


Fig. 3. Relative frequency distribution of the threshold gas price.

period. Moreover, these are prominent numbers and, therefore, might be appealing.<sup>13</sup>

It is also notable that more than 75% of the minimum gas prices and 70% of the median gas prices are integers in Gwei. However, choosing a gas price slightly above integer values—e.g., 4.1 or 4.2 Gwei instead of 4.0 Gwei—may render a transaction considerably more likely to be included into a block while hardly increasing the transaction fee. Apparently, however, higher-level reasoning with respect to this fact is not very common in the Ethereum mainnet, in line with observations made in so-called guessing games.<sup>14</sup>

### 3.3 Day-night pattern

Fig. 4 shows the distribution of the threshold gas price in the observed period. Quite notably, the threshold price typically has been higher during the continental European day (from 7 a.m. UTC to 6 p.m. UTC) than at nighttime. For a better understanding, we plotted the distribution of the *utilization rate* (Fig. 5), defined here as the total gas used in a block divided by the block gas limit. The Ethereum network is seen to be substantially more active during daytime than at night, which is in line with the day-night pattern of the threshold price.

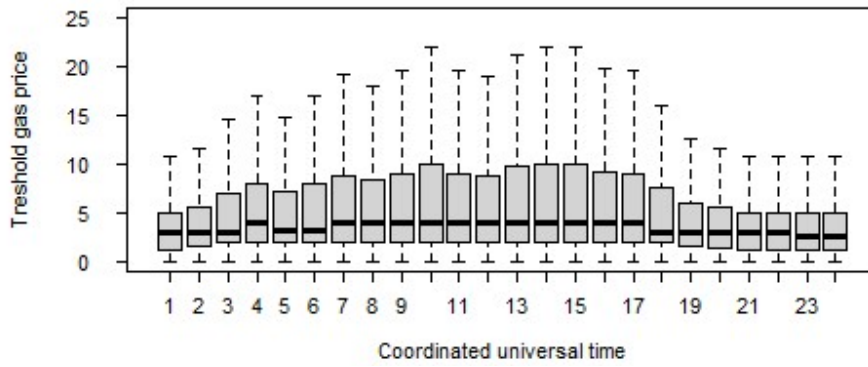


Fig. 4. Threshold gas price distribution for each hour of the day.

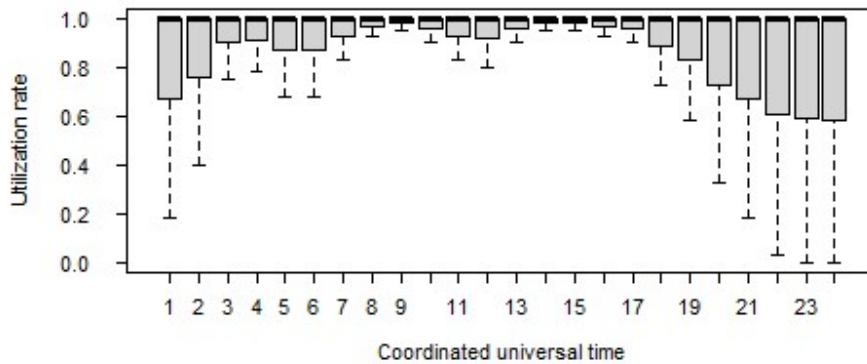


Fig. 5. Utilization rate distribution for each hour of the day.

## 4. Statistical analysis

This section summarizes the main steps and results of our statistical analysis.

### 4.1 Choosing the time series

Given the short-term inelasticity of demand identified in Section 3, the time series of threshold gas prices is too volatile to serve as a basis for statistical modeling. We therefore use instead the *hourly median*, i.e., the median of the threshold gas prices of all blocks mined within the same full hour on the clock.

Fortuitously, the hourly median relates in a natural way to the gas price that guarantees execution of a transaction within a given time span and with a given confidence level:

- The *operational gas price*,  $P_{op} = P_{op}(T, p)$ , is the theoretical gas price ensuring that a transaction will be confirmed by the network within time  $T$  with probability  $p$ .

Indeed, under certain assumptions detailed in Appendix A, the hourly median has the property that a transaction is accepted in the next block with probability 0.5. Therefore, given that in expectation four blocks are mined within  $T = 60$  seconds, the probability that the transaction with the hourly median threshold gas price will be mined within, say, a minute is  $p = 1 - (1 - 0.5)^4 = 93.75$  percent.

Below, we focus on the hourly median as the basis for our statistical analysis. Later in the paper, to account for the user impatience documented in Section 3, we will also allow for other hourly quantiles as an extension.

### 4.2 Identification

In a first step, we applied a Box-Cox transformation to the original time series of the hourly median threshold gas price using the R-function *BoxCox* from the forecast package.<sup>15</sup> This accounts for the asymmetric distribution of the threshold gas price—the threshold gas price

and thus also the hourly median has a lower bound at zero, but theoretically no upper bound. In the following,  $\{Y_t\}_t$  represents the transformed time series of the hourly median threshold gas price. The time series  $\{Y_t\}_t$  was then checked for stationarity – a necessary property for an ARMA model to be suitable. In line with the observations made in Section 3, we identified daily seasonality. Therefore, we applied a 24th order differencing ( $Y_t - Y_{t-24}$ ). An augmented Dickey-Fuller test (ADF) was performed to check whether further differencing was necessary. The test result supported our working hypothesis that the seasonal differenced time series can be considered stationary.

In a second step, we identified a suitable order for the model (meaning the number of non-seasonal and seasonal autoregressive and moving average parameters). We plotted the partial autocorrelation function (PACF) and estimated models for various possible combinations using the R-function *Arima* from the forecast package.<sup>15</sup> Aikake’s Information Criterion (AIC) has been used as a measure of goodness-of-fit.

In a third step, we assessed the prediction accuracy of our model within a forecast horizon of one day (24 hours), via a backtesting approach (an outline of the method used can be found in Subsection 4.4) with the mean absolute percentage error (MAPE) as a measure of fit. The results were compared against a naïve and a seasonal naïve method. Further details regarding the identification may be found in Appendix B.

### ***4.3 Statistical model***

For modelling the underlying process of the time series, we decided to use a SARIMA model. SARIMA models are a well-known and widely used class of prediction models for time series. Compared to ARIMA models, SARIMA models can take account for seasonality in the data (which is reflected in the day-night pattern identified above). While ARMA models are only suitable for modeling stationary processes, ARIMA models (and thus also SARIMA models) can be applied to processes that may not be stationary but can be made stationary by applying differencing. In general, a SARIMA model takes the form

$$Y_t^{d,D} = c + \sum_{h=1}^p AR_{ns,h} Y_{t-h}^{d,D} + \sum_{h=1}^P AR_{s,h} Y_{t-h*S}^{d,D} + \sum_{h=1}^q MA_{ns,h} \varepsilon_{t-h} + \sum_{h=1}^Q MA_{s,h} \varepsilon_{t-h*S} + \varepsilon_t ,$$

where the interpretation of the parameters is as follows:

Parameter	Interpretation
$p$	Order of the nonseasonal autoregressive part
$d$	Order of first order differencing
$q$	Order of the nonseasonal moving-average process
$P$	Order of the seasonal autoregressive part
$D$	Order of seasonal differencing
$Q$	Order of seasonal moving-average process
$S$	Length of seasonal cycle
$Y_t^{d,D}$	Differenced time series
$c$	Constant
$AR_{ns,h}$	Non-seasonal autoregressive parameter
$AR_{s,h}$	Seasonal autoregressive parameter
$MA_{ns,h}$	Non-seasonal moving average parameter
$MA_{s,h}$	Seasonal moving average parameter
$\varepsilon_t$	White noise: sequence of uncorrelated zero mean random variables

The specification of a SARIMA model may be expressed in a more compact form as  $ARIMA(p,d,q)(P,D,Q)[S]$ .

#### 4.4 Estimation results

We identified an  $ARIMA(2,0,1)(0,1,1)[24]$  model to be the most suitable for modelling the underlying process of the (Box-Cox transformed) hourly median threshold gas price time series  $\{Y_t\}_t$ . Table 1 shows the parameter estimates of this model when fitted to our data set.

The differenced process is modelled by a non-seasonal autoregressive part, a non-seasonal moving average part and a seasonal moving average part.

Table 1: Model Parameter Estimates.

Model	Type	Parameter	Value	S.E.	AIC
(2,0,1)(0,1,1)[24]	non-				-
	seasonal	AR1	1.60	0.014	9941
		AR2	-0.60	0.014	
		MA1	-0.89	0.009	
	seasonal	MA1	-0.95	0.004	

The residual analysis shows that there is significant autocorrelation in the squared errors. This is most likely caused by volatility clusters in periods with high network traffic. Thus, expanding the model by integrating a suitable GARCH model might be worthwhile in further studies.

#### 4.5 Backtesting

To measure the accuracy of prediction, we used cross-validation, and calculated the mean absolute percentage error (MAPE) of the back-transformed predictions. The cross-validation was performed in the following way: We randomly selected 500 data points, a data point being the observation of the threshold gas price at this time. At each of these data points, the data set was split into a training and a test set. The training set contained all data points up to and including the randomly selected data point, and the test set contained the data points for the following 24 hours.

We compared the prediction accuracy to a naïve model ( $\hat{Y}_{T+t|T} = Y_T$ ) and a seasonal naïve model ( $\hat{Y}_{T+t|T} = Y_{T+t-24 \times k}$ ), where  $k$  is the smallest integer greater than  $(t-1)/24$  for the same 500 training sets. The respective prediction of the hourly median gas price, say, two hours in the future would be the current price for the naïve model, and the price 22 hours in the past for the seasonal naïve model (since we have a seasonal cycle of length 24h). These approaches may also be represented by ARIMA(0,1,0)(0,0,0)[24] and



ARIMA(0,0,0)(0,1,0)[24] models. In the naïve approach, the time series is assumed to follow a random walk. The results are shown in Fig. 6.

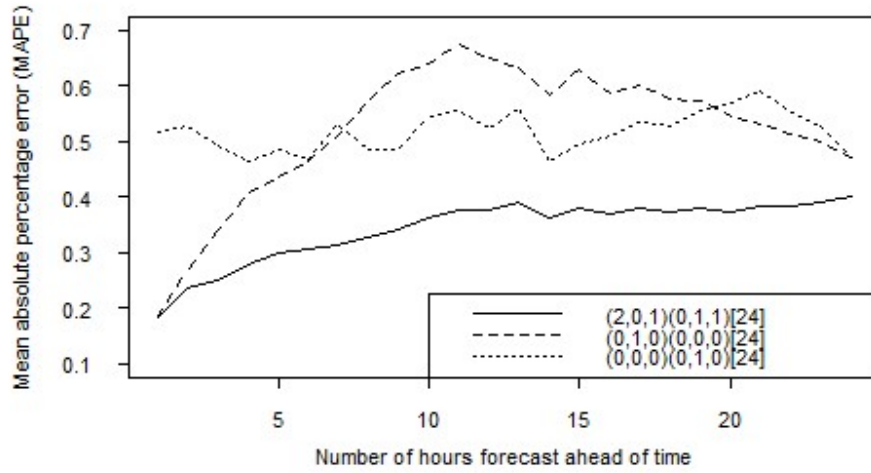


Fig. 6. MAPE of the SARIMA model compared to naïve approaches.

To check whether an ARIMA(2,0,1)(0,1,1)[24] model is also suitable for other hourly quantiles of the threshold gas price time series, we estimated models for the first and third quartiles (corresponding to a confirmation time of less than three minutes and less than

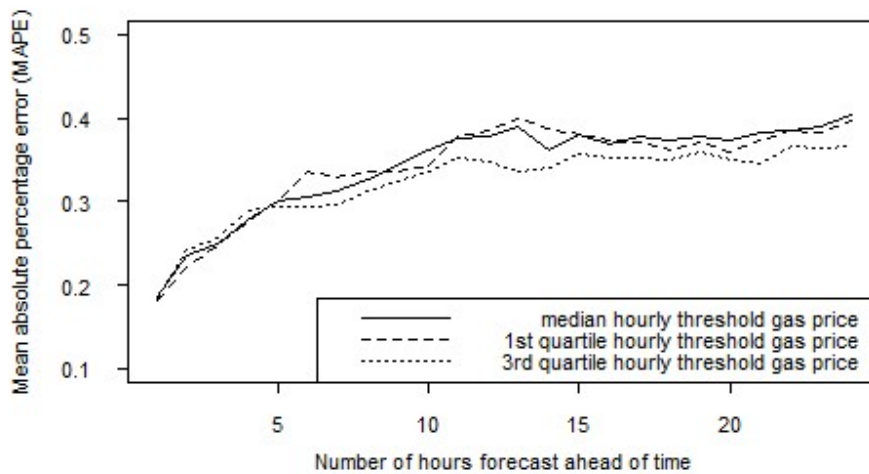


Fig. 7. MAPE of the SARIMA models for the first quartile, median, and third quartile hourly threshold gas price.

thirty seconds, respectively), and used the same cross-validation approach as before. As shown in Fig. 7, the MAPEs for the quartiles are similar to those obtained for the median.

## 5. Conclusion

The analysis has produced a set of potentially interesting observations related to the gas price in the Ethereum network. Only three quarters of the miners were seen to follow a purely price-based transaction-selection strategy, which necessitated the conceptualization of a threshold gas price. The short-term demand for transactions appears to be very inelastic, and creators of transactions apparently are in general not very price-sensitive regarding transaction fees. In fact, most of the transactions were executed at Gwei integer gas prices. The network traffic on average seems to be higher during the day (7 a.m. to 6 p.m., UTC) than during the night. For a prediction horizon of 24 hours or less, SARIMA models showed superior results in terms of prediction accuracy compared to naïve approaches.

Since the time at which we conducted our econometric analyses, the gas price has undergone a strong increase, leading the Ethereum Foundation to initiate the introduction of a proof-of-stake mining protocol on Ethereum 2.0. However, the data set that we used covers a variety of regimes, even though it did not include a period with prices as exceptionally high as at the time of the completion of this Study. Moreover, the SARIMA model is generally considered robust with respect to level changes provided that the assumptions underlying the statistical model are not validated in parallel. The extent to which the rise in the price levels has been accompanied by structural changes is, again, certainly an issue worthwhile to be explored.

## Appendix A. The relationship between gas price and confirmation time

This Appendix explains our choice of the hourly median threshold price as a proxy for an operational gas price.

We define the *confirmation time* of an individual transaction as the length of the time interval between submission of the transformation to the pool of pending transactions and the later confirmation by the network. We impose three assumptions:

**Assumption 1.** *In each block that is mined during a sufficiently short time interval  $(t_1, t_2]$ , a transaction with gas price  $P_{\text{gas}}$  is accepted with equal probability  $\pi_{P_{\text{gas}},(t_1,t_2]}$ .*

**Assumption 2.** *The acceptance of a transaction with gas price  $P_{\text{gas}}$  in a block mined during the time interval  $(t_1, t_2]$  is independent of the acceptance in all other blocks mined during  $(t_1, t_2]$ .*

**Assumption 3.** *The length of the time interval between two consecutively mined blocks,  $\Delta t$ , is (approximately) constant.*

Under these assumptions, the distribution of the confirmation time  $t^*$  at time point  $t_3$  in the interval  $(t_1, t_2]$ , with  $(t_2 - t_3)$  sufficiently large, may be calculated using the cumulative distribution function of a geometric distribution.

$$\begin{aligned} B_{i,(t_1,t_2]} &\sim \text{Bernoulli}(\pi_{P_{\text{gas}},(t_1,t_2]}) & (i \in \{1,2, \dots, n\}) \\ B_{i,(t_1,t_2]} &\perp B_{j,(t_1,t_2]} & (i, j \in \{1,2, \dots, n\}, i \neq j) \\ \Rightarrow \Pr_{P_{\text{gas}},t_3}(t^* \leq T) &= 1 - \left(1 - \pi_{P_{\text{gas}},(t_1,t_2]}\right)^{\left\lfloor \frac{T}{\Delta t} \right\rfloor} & (t_1 < t_3 < t_2) \end{aligned}$$

The probability  $\pi_{P_{\text{gas}},(t_1,t_2]}$  of acceptance in a block may be estimated as the fraction of blocks with a threshold gas price equal or below the gas price  $P_{\text{gas}}$ . If  $P_{\text{gas}}$  is chosen to be the median threshold price during this time interval, then  $\pi_{P_{\text{gas}},(t_1,t_2]} \approx 0.5$ .

**Proposition 1.** *Suppose that Assumptions 1-3 hold true for the hour between  $t_2 = t_1 + 1$  and  $t_1$ . Then, the hourly median of the threshold gas price guarantees execution within a given time span (1 minute) and with a given confidence level ( $\approx 95\%$ ).*

To apply the proposition to the Ethereum network, suppose that the time between two consecutively mined blocks is  $\Delta t = 15$  seconds. Then, the hourly median threshold price ( $T = 60$  seconds) is almost corresponding to the price that should be chosen if a confirmation time of one minute should not be exceeded  $1 - (1 - 0.5)^4 = 0.9375$ .

Assumptions 1-3 may indeed be reasonable for a short enough time interval since, first, information on gas price developments is only available with a time lag, second, the mining time generally does not vary considerably, and third, changes in gas price levels within this interval, such as due to time-dependent demand or changes in the number of miners, might be negligible. We consider an interval of an hour to be small enough to keep these assumptions plausible. The model may also be used for other quantiles as long as these are not too far away from the median.

The approach underlying Proposition 1 is related to the method implemented in the web3.py library, which implicitly models the confirmation time by a geometric distribution as well. However, instead of the threshold gas price, the 20th percentile of the gas prices of all transactions in a block is employed. Moreover, for a set of sorted gas prices, the number of blocks in which the 20th percentile is below the respective value is counted and divided by the total number of blocks. This ratio then serves as an estimate for the probability that a block accepts a transaction, instead of a quantile (e.g., the median). However, using the 20th percentile might lead to an overestimation of the confirmation time.

## Appendix B. Identification of the statistical model

This Appendix provides some background regarding the reasoning that led us to the choice of our econometric model.

### B.1 Seasonality

We first checked the Box-Cox transformed hourly median threshold gas price time series  $\{Y_t\}_t$  for *seasonality*. Since the time series covers more than 600 days and, therefore, seasonal patterns over 24 hours or less are very difficult to detect in a scatterplot, we plotted the ACF. The wave pattern repeating itself in a 24-hour rhythm clearly suggests a time series with seasonality. Therefore, we applied a 24th order differencing ( $Y_t - Y_{t-24} =: Y_t^{0,1}$ ).

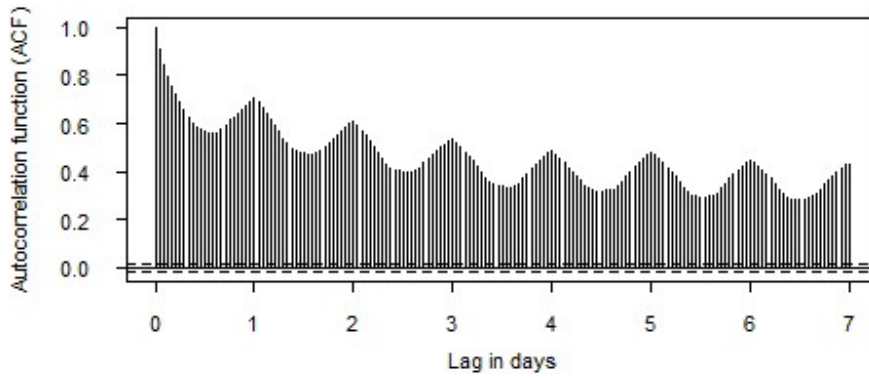


Fig. 8. Plotted ACF of the Box-Cox-transformed hourly median threshold price.

### B.2 Stationarity

Secondly, we tested the seasonally differenced time series  $\{Y_t^{0,1}\}_t$  for *stationarity*. The differenced time series obtained seems to be stationary. To validate this, we used the ADF implemented in the R-package aTSA by the function `adf.test`. The test statistics confirm that no further differentiation is necessary. Therefore, we decided to use an  $ARIMA(p,0,q)(P,1,Q)[24]$  model.

Table 2. Augmented Dickey-Fuller Test Statistic.

Lag	ADF Statistic	P-value	Conclusion
12	-25.1	$\leq 0.01$	Stationarity

Type: no drift, no trend

### B.3 Identification of the order of parameters

To obtain a first indication for the SARIMA parameters, we plotted the PACF for the seasonally differenced time series. The first three spikes in the PACF plot are significant. Subsequent lags do not seem to provide additional information. Therefore, a choice of  $p > 3$  does not seem to be reasonable. The strongly significant values of the seasonal lags in the PACF plot, which seem to be exponentially decreasing, suggest that  $Q \geq 1$  should be chosen.

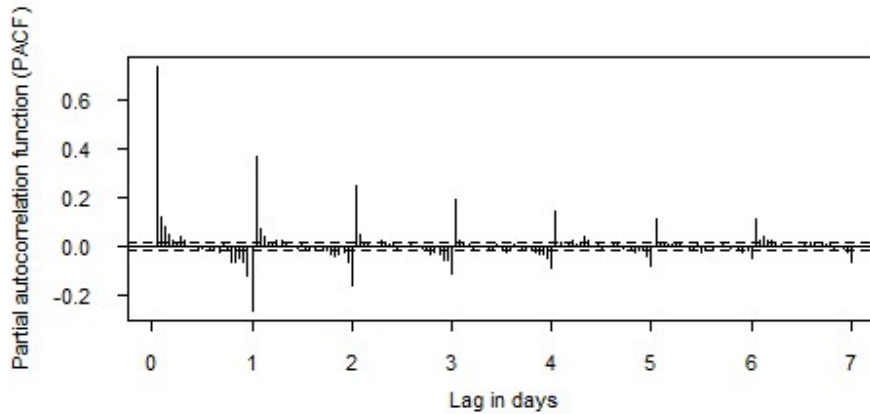


Fig. 9. Plotted PACF of the seasonal differenced Box-Cox transformed time series.

As mentioned in the body of the paper, we estimated models for various possible combinations using the R-function *Arima* from the package *forecast*. Since the distribution of gas prices is clearly skewed to the right due to a lower boundary of zero and a non-existing upper boundary we used a Box-Cox transformation on the time series. The results are summarized in Table 3. As the table shows, the ARIMA(2,0,1)(0,1,1)[24] model clearly

outperforms the two models  $\text{ARIMA}(1,0,0)(0,1,1)[24]$  and  $\text{ARIMA}(2,0,0)(0,1,1)[24]$ . On the other hand, compared to the  $\text{ARIMA}(2,0,1)(0,1,1)[24]$  model,  $\text{ARIMA}(2,0,1)(0,1,2)[24]$ ,  $\text{ARIMA}(2,0,1)(1,1,1)[24]$ , and  $\text{ARIMA}(3,0,1)(0,1,1)[24]$  hardly lead to an improvement when looking at the AIC and the parameter estimates. Therefore, following the KISS principle, we selected  $\text{ARIMA}(2,0,1)(0,1,1)[24]$ .<sup>16</sup>

Table 3: Model Parameter Estimations.

Model	Type	Parameter	Value	S.E.	AIC
(1,0,0)(0,1,1)[24]	non-seasonal	AR1	0.89	0.004	-8890
	seasonal	MA1	-0.93	0.005	
(2,0,0)(0,1,1)[24]	non-seasonal	AR1	0.74	0.008	-9329
		AR2	0.17	0.008	
	seasonal	MA1	-0.94	0.005	
(2,0,1)(0,1,1)[24]	non-seasonal	AR1	1.60	0.014	-9941
		AR2	-0.60	0.014	
		MA1	-0.89	0.009	
	seasonal	MA1	-0.95	0.004	
(2,0,1)(0,1,2)[24]	non-seasonal	AR1	1.60	0.015	-9974
		AR2	-0.60	0.014	
		MA1	-0.89	0.010	
	seasonal	MA1	-0.91	0.008	
		MA2	-0.05	0.008	
(2,0,1)(1,1,1)[24]	non-seasonal	AR1	1.60	0.015	-9976
		AR2	-0.60	0.014	
		MA1	-0.89	0.010	
	seasonal	AR1	0.05	0.009	
		MA1	-0.96	0.004	
(3,0,1)(0,1,1)[24]	non-seasonal	AR1	1.60	0.011	-9978
		AR2	-0.55	0.015	
		AR3	-0.06	0.009	
		MA1	-0.92	0.007	
	seasonal	MA1	-0.95	0.004	



#### ***B.4 Residual analysis***

The plotted autocorrelation function of the squared residuals exhibits significant spikes, although the estimated correlations are rather small. Modelling the residuals by a GARCH model might further improve model fit.

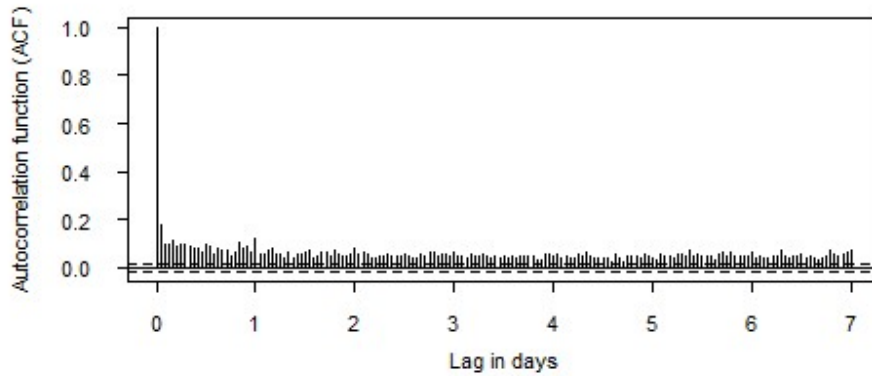


Fig. 10. Plotted ACF of the squared residuals of the SARIMA(2,0,1)(0,1,1)[24] model.

#### **Acknowledgements**

This paper has benefited from comments received during a seminar presentation at the University of Zurich.

#### **Author contributions**

DC conducted the econometric analysis. CE proposed the research question and provided several rounds of detailed feedback. They both contributed equally to manuscript preparation.

---

## Endnotes

<sup>1</sup> Antonopoulos, A.M., Wood, G. *Mastering Ethereum: Building Smart Contracts and DApps*, O'Reilly, Farnham, UK (2018).

<sup>2</sup> 1 GWei =  $1 \times 10^{-9}$  Ether.

<sup>3</sup> Pierro, G.A., Rocha, H. "The influence factors on Ethereum transaction fees." In *2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*. Montreal: IEEE 24-31 (May 2019) doi: 1109/WETSEB.2019.00010

<sup>4</sup> Werner, S.M., Pritz, P.J., Perez, D. "Step on the gas? A better approach for recommending the Ethereum gas price." *eprint arXiv:2003.03479* (March 2020).

<sup>5</sup> de Azevedo Sousa, J.E., Oliveira, V., Valadares, J., Dias Gonçalves, G., Moraes Villela, S., Soares Bernardino, H., Borges Vieira, A. (2020). "An analysis of the fees and pending time correlation in Ethereum" *International Journal of Network Management*;e2113.

<sup>6</sup> Möser, M., Böhme, R. "Trends, Tips, Tolls: A Longitudinal Study of Bitcoin Transaction Fees." In *Financial Cryptography and Data Security*. Berlin, Springer 19-33 (2015).

<sup>7</sup> Kooops, D. "Predicting the confirmation time of bitcoin transactions." *eprint arXiv:1809.10596* (September 2018).

<sup>8</sup> Easley, D., O'Hara, M., Basu, S. "From mining to markets: The evolution of bitcoin transaction fees" *Journal of Financial Economics* **134.1** 91-109 (2019).

<sup>9</sup> Wood, G. "Ethereum: A Secure Decentralised Generalised Transaction Ledger." *Ethereum Project Yellow Paper* **151** 1-32 (2014).

<sup>10</sup> The data were retrieved using `eth.getBlock` and `eth.getTransactionByBlock` from the library `web3.py`. "web3.eth API" *web3.py* (accessed July 11, 2020) <https://web3py.readthedocs.io/en/stable/web3.eth.html>

<sup>11</sup> <https://web3py.readthedocs.io/en/stable/web3.eth.html> (last accessed: Sept. 17, 2020)

<sup>12</sup> <https://github.com/dlcarl/Ethereum-Gas-Price-Statistics> (last accessed: Sept. 17, 2020)

<sup>13</sup> Albers, W., Albers, G. "On the Prominence Structure of the Decimal System." In R. W. Scholz (Ed.), *Decision Making under Uncertainty*. Amsterdam: Elsevier 271-287 (1983).

<sup>14</sup> Nagel, R., Unraveling in Guessing Games: An Experimental Study, *American Economic*

---

*Review* **85**, 1313-1326 (1995).

<sup>15</sup> Hyndman, R. J. et al. “Package ‘forecast’. Forecasting Functions for Time Series and Linear Models” (March 2020) *cran.r-project* <https://cran.r-project.org/web/packages/forecast/forecast.pdf> (last accessed: July 11, 2020).

<sup>16</sup> Kass, R.E., Caffo, B.S., Davidian, M., Meng, X.-L., Yu, B., Reid, N. “Ten Simple Rules for Effective Statistical Practice” *PLoS Computational Biology* **12.6** (2016).