Ökonometrie der Bitcoinzeitreihen

Marius Lehne
Westfälische Wilhelms-Universität Münster
Institut für Wirtschaftsinformatik
Leonardo Campus 3
48149 Münster (Westf)
m_lehn03@uni-muenster.de

ABSTRACT

Das Bitcoinnetzwerk ist abgesichert durch Rechenleistung, für einen Proof-of-Work, welche von Dritten zur Verfügung gestellt. Die Analyse der Rechenleistung steht im Fokus dieser Arbeit. Insbesondere welche Faktoren auf diese einen Einfluss haben, in welchem Ausmaß sie wirken und ob eine Prognose der zukünftigen Rechenleistung möglich ist. Der Ausmaß der dieser ist ein zentraler Faktor dafür, dass das Bitcoinsystem schwierig zu kompromitieren ist. Eine genauere Untersuchung kann helfen Schwachstellen zu identifizieren. In der Arbeit werden dazu zunächst durch eine Analyse des Protokolls und des Ökosystmes verschiedene Einflussfaktoren identifiziert. Deren Relevanz wird mit Hilfe von Zeitreihenanalysen überprüft und quantisiert. Zudem wird ein autoregressives Prognosemodell vorgestellt. Es wird festgestellt, dass es starke Hinweise darauf gibt, dass der Wechselkurs in Dollar und die Effizienz der verwendeten Mininghardware einen signifikanten Einfluss auf die Rechenleistung haben.

Keywords

Bitcoin, Zeitreihenanalyse, Ökonometrie, Mininganreize

1. EINLEITUNG

Bitcoin ist eine dezentrale kryptographische Währung. Die Bedeutung dieser Währung wird durch ihre Marktkapitalisierung von ca. 1,4 Milliarden USD [4] im Mai 2013 und durch ihre zunehmende Beachtung in den Medien deutlich. Durch die zugrunde liegende Architektur ist es möglich, auf alle jemals getätigten Transaktionen und deren Details zuzugreifen. Somit ist eine Vielzahl von Daten frei verfügbar, die eine Analyse erleichtert. Im Bitcoin-System werden Transaktionen nicht durch eine zentrale Stelle, sondern durch Teilnehmer des Systems verifiziert. Als Absicherung dient die Berechnung eines Proof-of-Work. Diese kostet Rechenzeit, somit Energie und letztendlich auch Geld. In dieser Arbeit sollen die Anreizstrukturen für die Teilnahme am Verifizieren von Transaktionen und somit dem Bereitstellen von Rechenleistung identifiziert werden. Zudem soll festgestellt werden, ob dies anhand historischer Daten möglich ist, diesen Einfluss nachzuweisen und gegebenenfalls zu quantisieren. Zuletzt soll untersucht werden, ob eine Prognose möglich ist.

Die Bereitschaft Rechenleistung zur Verfügung zu stellen, ist im Bitcoinsystem essentiell, um gegen verschiedene Angriffe abgesichert zu sein. Durch eine niedrige Rechenleistung fällt es einem Angreifer leichter die Entwicklung der Blockchain zu seinem Gunsten zu beeinflussen. Eine genaue Untersuchung der Zusammenhänge ist für die Identifikation von möglichen Angriffsstrategien, die auf den Proof-of-Work abzielen zentral. Für Miner ist ein Verständnis über die Einflussfaktoren der Rechenleistung und die daraus resultierende Difficulty wichtig, um eine Wirtschaftlichkeitsberechnung zur Beschaffung neuer Mininghardware zu betreiben. Mit einer theoretischen Untersuchung von Mining werden möglichen Faktoren auf die Rechenleistung des Bitcoinnetzwerkes identifiziert. Mit Hilfe von multivariaten Zeitreihenmodellen soll in dieser Arbeit der Einfluss dieser Faktoren untersucht. Dazu werden verschiedene Modelle vorgestellt, die die entsprechenden Zusammenhänge beschreiben, von denen unter Einschränkungen ein Modell welches die Rechenleistung aus der Effizienz des Miningequipment und dem Wechselkurs beschreibt, als relevant identifiziert werden kann.

Für die Konstruktion der Modelles, wird als abhängige Variable die Hashrate ausgewählt. Zudem werden mögliche Einflussfaktoren, die diese beeinflussen, identifiziert. Dafür wird zunächst in Abschnitt 3.1 grundlegende Terminologie und Zusammenhänge innerhalb des Bitcoinökosystems erläutert. Aus diesen werden dann vier zu untersuchende exogene Variablen ermittelt. Diese sind der aktuelle Wechselkurs von Bitcoin nach US-Dollar, die Menge der pro Block generierten Bitcoin, die Effizienz der benutzten Mininghardware, die Transaktionskosten, die Risikoaffinität der Miner und die aktuellen Energiekosten. In Abschnitt 5.1 werden die zu den entsprechenden Variablen gehörenden Datensätze erläutert, ihre Herkunft und die daran vorgenommenen Transformationen beschrieben. Zudem werden Besonderheiten in den Datensätzen aufgezeigt. Außerdem werden bestimmte Zeiträume identifiziert, die sich für die Analysen eignen. Im Anschluss werden dann spezielle Modelle vorgestellt, auf denen dann mit Hilfe der Daten Regressionsanalysen durchgeführt werden. Hierzu diese schrittweise um weitere Regressoren erweitert. Insbesondere wird auch untersucht, ob Verzögerungseffekte einzelner Variablen auftreten. Es wird

Münster Bitcoin Conference (MBC), 17–18 July '13, Münster, Germany

zudem ein Prognosemodell vorgestellt Zuletzt erfolgt Bewertung und Vergleich der vorgestellten Modelle.

2. RELATED WORK

Bitcoin und sein Ökosystem sind Gegenstand aktueller Forschung. Die Ursprünglichen Ideen für die Entwicklung von Bitcoin wurden von Santoshi Nakamato [19] veröffentlicht. Speziell im Bereich des Bitcoinminig wurden verschiedene Erkenntnisse gewonnen. Hier wurde insbesondere die Zusammensetzung von Miningpools im Zeitverlauf untersucht. In der Arbeit von Kroll et. al. [17] stellen die Autoren fest, dass der Miningalgorithmus zu einem Gleichgewicht führt, in dem sich alle Beteiligten an die Regeln halten. Es wurde zudem untersucht, in wie weit alte Mininghardware, die durch zunehmenden Schwierigkeitsgrad ineffizient geworden ist, durch neue ersetzt wird [18]. Es gibt Versuche die im Bitcoinsystem zur Verfügung gestellte Rechenleistung mit Hilfe von autoregressiven Zeitreihenmodellen vorherzusagen [7]. Eine umfassende Untersuchung zur Identifikation der Einflussfaktoren die über den Einflusses des Wechselkurses hinausgehen ist dem Autor zum Zeitpunkt des Verfassens dieser Arbeit jedoch nicht bekannt.

3. GRUNDLAGEN

3.1 Bitcoinprotokoll

Zentral für das Konzept von Bitcoin ist die Transaktion. In ihr wird der Transfer von Bitcoin von einer oder mehreren Adressen zu einer oder mehreren Adressen dokumentiert. [12] Die Menge der Bitcoin die einer Adresse zugeordnet sind, ergibt sich aus der gesamten Transaktionshistorie. Eine Transaktion innerhalb des Bitcoinnetzwerkes wird dann als gültig akzeptiert, wenn diese in einen Block eingebaut wird, und dieser Block ein gültiger Bestandteil der Blockchain wird. Das Überprüfen und Zusammenfassen von Transaktionen kann von jedem Teilnehmer des Bitcoinnetzwerkes durchgeführt werden. Hierzu muss ein Proof-of-Work gleistet werden. Das Bitcoinprotokoll sieht hierfür das Berechnen einer kryptographischen Hashfunktion vor, dessen Ergebnis bestimmten Anforderungen genügen muss. Ist dies korrekt geschehen, wird dieser Block veröffentlicht. Er gilt dann als gefunden. Ein neu gefundener Block referenziert den letzten zuvor gefundenen gültigen Block. Hieraus bildet sich eine Kette von Blöcken, die sogenannte Blockchain. Es kann passieren, dass mehrere Blöcke auf den gleichen Block referenzieren. Neue Blöcke verweisen immer auf die längste Kette an Blöcken. Alle Blöcke und die dort enthaltenen Transaktionen, die sich nicht in der längsten Kette befinden, werden als ungültig erachtet.

Das Protokoll ist so konstruiert, dass im Idealfall alle zehn Minuten ein gültiger Hashwert berechnet wird. Um dies zu gewährleisten, werden die Anforderungen an ein korrekten Hashwert angepasst. Nach 2016 Blöcken wird überprüft, wie viel Zeit für das Finden der vorherigen Blöcke benötigt wurde. Dementsprechend wird die Schwierigkeit erhöht oder gesenkt, sodass die erwartete durchschnittliche Zeit zwischen den Blöcken wieder erreicht wird.[10] Die Schwierigkeit wird auch Difficulty genannt. Der Prozess des Berechnen von Hashes wird auch Mining bezeichnet, die Personen die dies betreiben Miner.

Die Wahrscheinlichkeit ein korrektes Ergebnis berechnet zu haben, steigt mit der Anzahl der berechneten Hashes. Die Anzahl dieser Berechnungen pro Sekunde wird Hashrate genannt. Sie beschreibt, wenn Sie auf das gesamte Bitcoinnetzwerk bezogen wird, wie viel Rechenaufwand in diesem insgesamt geleistet wird. Dieser Wert ist kein direkt beobachtbarer Parameter, sondern kann aus der aktuellen Schwierigkeit und der Zeit, die zwischen dem Finden von Blocks vergeht, berechnet werden. Einen Überblick, wie sich die Hashrate im Zeitverlauf verändert hat, findet sich in Abbildung 1. Die Hashrate ist essentiell für die Sicherheit des gesamten Systems. Je größer die Rechenleistung eines Angreifer relativ gesehen zum Rest des Netzwerkes ist, desto großer sind die Erfolgschancen eines Angriffes. Besitzt dieser mindestens 51% der Rechenleistung im System, so hat er die Möglichkeit dieses anzugreifen. [20] Eine höhere Hashrate erschwert dementsprechend ein Angriff. Für jeden erfolgreich gefundenen Block gibt es eine Belohnung (Reward), in Form von Bitcoin. Dies dient zum einen der initialen Verteilung von Bitcoins und zum anderen als Anreiz zur Sicherung des Netzwerkes. Vom ersten bis zum 209999. Block betrug diese 50 BTC, danach 25 BTC. Auch in Zukunft wird sich diese Belohnung alle 210000 Blocks halbieren.[9] Bei einer Transaktion kann ein Benutzer eine Transaktionsgebühr festlegen. Ein Miner kann sich zusätzlich zu dem Reward für das Finden eines Blocks sich allen in diesem Block enthaltenen Transaktionen die dazugehörigen Transaktionsgebühren gutschreiben.[13]

3.2 Bitcoinökosystem

Für das Finden eines Blockes werden in Mai 2013 ca. $5,21 \cdot 10^{16}$ Hashes benötigt.[5] Für einen Großteil der Miner ist daher die Wahrscheinlichkeit, mit ihrem Equipment einen Block zu finden, sehr gering. Um dennoch mit ihrer Tätigkeit Geld verdienen zu können, schließen sich Miner in Miningpools zusammen. Jeder Teilnehmer stellt diesem seine Rechenkapazität zur Verfügung. Wird von einem Teilnehmer ein Block gefunden, so wird die Belohnung an die Teilnehmer verteilt. Meist wird dazu der Anteil des jeweiligen Teilnehmers an der totalen Rechenkapazität des Miningpools berücksichtigt.[21] Im Bitcoinökosystem existieren ein Vielzahl verschiedener Pools. Neben diesen existieren noch Mining Companies. Diese verkaufen Anteile, mit deren Erlös dann in die Beschaffung weiterer Miningkapazitäten investiert wird. Die Hardware verbleibt in diesem Fall bei der Firma. Der erzielte Gewinn wird abzüglich von Verwaltungsgebühren an die Anteilseigner als Dividende ausgeschüttet. Der bekannteste Vertreter ist ASICMiner, welche zum Betreiben des Minings selbst entwickelte Hardware verwenden.

Um Bitcoin in gesetzlich geregelte Währungen wie z.B. den Euro umzutauschen, müssen Abnehmer gefunden werden, die bereit sind für diese Währung Bitcoin zu verkaufen. Für diesen Zweck existieren Wechselstuben bzw. Börsen [1], die solche Transaktionen vermitteln. Aus den Kauf- und Verkaufsangeboten ergibt sich ein Wechselkurs.

3.3 Mininghardware

Die zum Minen verwendete Hardware lässt sich chronologisch in vier Geräteklassen einteilen: CPUs, Grafikkarten, FPGAs und ASICs. In der Anfangsphase wurden demnach CPUs eingesetzt, wohingegen der aktuellste Stand die ASICs sind. Bei FPGAs handelt es sich um Chips die sich frei programmieren lassen. Erste Berichte über das Auftreten von

[1] Mt. Gox, bitcoin.de

Hardware	Effizienz in MHs/J	Einsatz seit
CPU	0,14	Beginn
Grafikkarte	1,87	Beginn
FPGA	18,05	2011
ASIC	134	2013

Tabelle 1: Effizienz verschiedener Mininggeräte basierend auf [11]

FPGAs zum Minen von Bitcoin lassen sich auf den Frühjahr 2011 zurückdatieren. Die neuste Generation sind Application Specific Intergrated Circuits (ASIC). Solche Chips werden speziell zum Lösen ein bestimmten Aufgabe entwickelt, in diesem Fall dem Berechnen von Hashfunktionen zum Bitcoinmining. Sie zeichenen sich insbesondere durch ihre hohe Effizienz aus. Erste Lieferungen größerer Mengen erfolgten im März 2013 von Avalon, einem größeren Hersteller von Miningequipment [6]. Die ersten ASICs von ASICMiner wurden im Februar 2013 in Betrieb genommen.[3] Für größere Verwendung von ASICs zum Mining in einem Zeitraum davor existieren keine Berichte. Nach teilweise langen Lieferzeiten ist erst im aktuellen Zeitraum eine zunehmende Verfügbarkeit erkennbar.

Wie aus Tabelle 1 ersichtlich wird, hat die Miningeffizienz über den untersuchten Zeitraum dramatisch zugenommen. Durch einen Zuwachs an Effizienz entstehen dem Miner weniger Kosten durch Energieverbrauch. Im Gegensatz zu den bisher betrachteten Daten liegen keine Daten vor, wie effizient das Netzwerk insgesamt rechnet. Um diesen Faktor dennoch einzubeziehen muss die vorhandene Effizienz geschätzt werden. Dazu befindet sich in Tabelle 1 eine vereinfachte Übersicht der zum Mining verwendeten Geräteklassen.

Die Effizienz ist in MHs/J angegeben. Die Daten einzelner Geräte stammen von [11]. Um eine aggregierte Kennzahl für die Geräteklassen zu erhalten, wird ein Mittelwert über die Effizienzen gebildet wurde. Da die vorhandenen Daten keine repräsentative Auswahl darstellen, kann diese nur als eine grobe Schätzung angesehen. Für die Schätzung der Effizienz der CPUs wurde eine Auswahl aktueller AMD-Prozessoren verwendet. Notebookprozessoren können jedoch deutlich effizienter sein.

4. MODELLSPEZIFIKATION

Im Folgenden sollen die wichtigen Einflussfaktoren auf die Rechenleistung des Bitcoinnetzwerkes identifiziert werden, bevor einige von diesen in Modelle umgesetzt und empirisch Überprüft werden. Neben den rein monetären Einflussfaktoren, sollen auch weitere Faktoren untersucht werden, deren Einfluss relevant sein könnte. Hierzu wird zunächst eine Betrachtung auf individueller Ebene durchgeführt.

Die Grundlegende Annahme ist, dass durch das Betreiben von Bitcoinmining Profite erzielt werden können, die Anreize geben dieser Tätigkeit nachzugehen. Hierfür soll davon ausgegangen werden, dass die Miner rational handeln. Das heißt, dass sie Rechenleistung dann zur Verfügung stellen, wenn es sich für sie finanziell lohnt und diese wieder einstellen, wenn kein positives Ergebnis zu erzielen ist. Des Weiteren wird unterstellt, dass es keine weiteren Besonderheiten gibt, zum Beispiel, dass Mining durch Botnetze betrieben wird.

Die Einnahmen U, die durchschnittlich bei der Berechnung eines Hashes erwartet werden, setzen sich aus der Höhe des Reward R, dem aktuellen Wechselkurs ER und der durchschnittliche Anzahl Hashes die berechnet werden müssen, um einen Block zu finden W, zusammen. Formell lässt sich dies folgendermaßen ausdrücken:

$$U = \frac{ER(R + TC)}{W} \tag{1}$$

Entgegen der Definition von Bitcoin als Währung, muss trotzdem der aktuelle Wechselkurs in einer, von einer Mehrheit akzeptierten Währung, mit in die Betrachtung einfließen. Dies ist notwendig, da Möglichkeiten ausschließlich mit Bitcoin zu bezahlen bisher noch sehr gering sind. Nur wenn der Miner einen tatsächlichen Nutzen hat wirkt der Anreiz.

Den Einnahmen gegenüber stehen die Ausgaben K für die Berechnung eines Hashes für einen individuellen Miner, welche sich aus der Effizienz des verwendeten Mininggerätes E und den aktuellen Stromkosten EC zusammensetzen. Dieser Zusammenhang lässt sich wie folgt darstellen:

$$K = E \cdot EC \tag{2}$$

Ein rationaler Miner wird unter Vernachlässigung von Fixkosten, im Fall ausgeglichener Einnahmen und Ausgaben $U = E$ seine Rechenkapazität auf dem gleichen Niveau belassen. Verändert sich jedoch die Situation, sodass stattdessen $K < U$ gilt, wird der Miner versuchen seine Miningkapazitäten auszubauen. Andersherum wird dieser seine Miningaktivitäten einstellen.

Wird zur Erhöhung der Rechenleistung neue Hardware angeschafft, müssen die Anschaffungskosten mit berücksichtigt werden. Durch eine Investitionsrechnung, können diese Fixkosten verrechnet werden. Ohne die Berücksichtigung von Zinsen ermöglicht dies beispielsweise durch eine Break-Even-Analyse. So kann festgestellt werden , wie viele Hashes X^* berechnet werden müssen, bis sich die Fixkosten K_f amortisiert haben.

$$X^* = \frac{K_f}{U - K} \tag{3}$$

Durch eine Multiplikation der Bandbreite[2] mit X^* kann bestimmt werden, wie viel Zeit aufgewendet werden muss, um die Fixkosten zu decken. Die Einnahmen und die Kosten sind jeweils Prognosen, die eine mögliche Veränderung der Parameter berücksichtigt. Insbesondere bei dem Wechselkurs und der Schwierigkeit kann es zu großen Schwankungen kommen. Abhängig von seiner Risikowahrnehmung wird der Miner die Art und Weise der Parameterschätzung wählen. Zudem ist von seiner Risikoaffinität abhängig, ob er bei der so berechneten Amortisationszeit investiert. An dem Beispiel von ASICMiner lässt sich verdeutlichen, dass eine Investitionsentscheidung nicht zwingend die direkte Beschaffung von neuer Hardware zur Folge haben muss, sondern auch von der indirekte Erweiterung der Rechenleistung bis hin zur Neuentwicklung von Mininghardware beitragen kann.

Sobald gesonderte Beschaffungen für das Mining getätigt werden müssen und nicht einfach vorhandene Hardware ge-

[2]Die pro Zeiteinheit berechneten Hashes

nutzt wird, kann es dazu kommen, dass zwischen der Schwelle Rechenleistung abzuschalten und neue Rechenleistung hinzuzufügen eine Divergenz besteht. Dies kann zum Beispiel dazu führen, dass eine unerwartete Verminderung der Einnahmen kurz nach der Beschaffung, nicht zwingend dazu führen muss, dass dieses Gerät direkt wieder abgeschaltet wird.

Um die Rechenleistung des gesamten Bitcoinsystem zu analysieren, müssen die Gewinnerwartungen aller Miner berücksichtigt werden. Denn sofern durch eine Veränderung eines Faktors das Mining für einen Teilnehmer profitabel wird und sich nach seinen Berechnungen die Investition lohnt, wird dieser seine Bandbreite erweitern. Infolgedessen steigt die Rechenleistung des gesamten Netzwerks. Nach spätestens 14 Tagen wird die Schwierigkeit gemäß des Protokolls angepasst. Dies führt dazu, dass der Anreiz weitere Rechenleistung hinzuzufügen sinkt. Zudem kann es dazu kommen, dass nicht wirtschaftliche Mininggeräte abgestellt werden. Sofern sich dadurch keine weiteren Veränderungen ergeben, stellt sich wieder ein Gleichgewicht ein. Daraus folgt, dass die Hashrate im Netzwerk abhängig von Einflussfaktoren ist, die auf jeden einzelnen Miner wirken. Die Faktoren, die für jeden Miner individuell sind, werden aggregiert. Dies sind die im Netzwerk verwendete Effizienz des Miningequipments $\tilde{E}$, die jeweils verwendeten Stromkosten $\tilde{EC}$, sowie die Risikoaffinität der Teilnehmer $\tilde{RS}$.

Aus diesen Überlegungen kann dann folgender Zusammenhang zur Erklärung der Hashrate formuliert werden:

$$\text{Hashrate} = f(ER\uparrow, R\uparrow, \tilde{E}\uparrow, \tilde{EC}\downarrow, TC\uparrow, RS) \qquad (4)$$

Die aggregierten Faktoren sind nicht direkt beobachtbar und müssen aus diesem Grund geschätzt werden. Dies ist nicht immer so einfach möglich. So gibt es gravierende Unterschiede bei den Strompreisen in unterschiedlichen Regionen. Zum Beispiel liegt im November 2012 der niedrigste Strompreis innerhalb der Europäischen Union bei ca. 11 Cent/kWh und der höchste Strompreis bei 26 Cent/kWh. [15].

Der Einfluss der Transaktionsgebühren ist klein, denn er macht nur einen relativ kleinen Teil des Betrages aus, den sich ein Miner gutschreiben kann. Im Mai 2013 waren dies durchschnittlich $0,18$ BTC[3]. Mit einer Verringerung des Rewards wird dieser Faktor eine größere Gewichtung bekommen.

Nicht alle Einflussfaktoren haben eine sofortige Wirkung auf die Rechenleistung. Durch lange Lieferzeiten und mangelnde Verfügbarkeit entsprechender Hardware ist zu erwarten, dass sich verändernde Einflussfaktoren, einen verzögert einen Einfluss auf die Zielvariable haben.

Damit auf einen potentiellen Anreiz reagiert werden kann, müssen ausreichend Personen über dessen Existenz bescheid wissen. Der Bekanntheitsgrad ist somit auch ein möglicher Faktor, der die Rechenleistung beeinflusst. Zwar kommt eine Studie des IT-Branchenverband Bitkom aus dem Mai 2013 zu dem Schluss, dass nur 15% der Deutschen von Bitcoin gehört habe[14]. Jedoch legt das vermehrte Auftreten von professionellen Firmen, die sich auf Miningequipment spezialisiert haben, nahe, dass dieser Faktor vernachlässigbar ist.

[3]Analyse aus in Kapitel 5 vorgestellten Datensatz

5. EMPIRISCHE ANALYSEN

5.1 Datensatz

Im folgenden sollen die Datensätze erläutert deren Herkunft beschrieben werden, die für die empirischen Analysen benötigt werden.

Die Daten für die Zielvariable Hashrate_t beschreibt eine Schätzung der zum Zeitpunkt t im gesamten Netzwerk berechneten Hashes pro Sekunde. Sie wurden der Seite blockexplorer.com entnommen. Die Schätzung wird aus der Difficulty und der Zeit, die zwischen der Veröffentlichung von zwei Blöcken liegt, erstellt. Im verwendeten Datensatz stammt wird Schätzung über den Bereich von je 5 Blöcken gebildet. Um eine reguläre Zeitreihe zu erhalten, wurde für jeden Tag der Durchschnitt ermittelt. Der Datenbereich liegt vom 09.02.2009-26.05.2013.

Die Variable ER_t entsprich dem Wechselkurs eines Bitcoins zum Zeitpunkt t in USD. Die Daten zugehörigen Daten stammen von der bitcoincharts.com [2]. Sie aggregieren die Daten von der Bitcoin Wechselstube Mt. Gox. Diese ist mit ca. 64% Marktanteil im Mai 2013 [1] die größte Wechselstube innerhalb des Bitcoinökosystems. Der dort ermittelte Wechselkurs kann folglich als eine Annäherung des tatsächlichen Wertes eines Bitcoin im Verhältnis zu einer gesetzlichen Währung angenommen werden. Als Bezugswährung wurde der USD gewählt, da der größte Teil der Bitcoin mit dieser Währung gehandelt werden[2]. Der Datensatz umfasst den gewichteten Wechselkurs für jeden Tag in dem Zeitraum vom 25.05.2011-23.05.2013. Im Zeitraum von 20.06.2011-25.06.2011 ist der Datensatz unvollständig. Die fehlenden Daten wurden, um eine Analyse zu ermöglichen, interpoliert.

Die Variable R_t ist eine kategorielle Variable mit einer entsprechenden Ausprägung für den zum Zeitpunkt t vorliegenden Reward. Ihr Wert lässt sich aus der Protokollspezifikation herleiten, welche in Kapitel 3.1 erläutert wurde.

Die Variable TC_t entspricht der durchschnittlichen Transaktionshöhe in BTC zum Zeitpunkt t. Die Daten stammen von blockchain.info, die diese aus der Blockchain entnommen und auf eine Tagesbasis verdichtet haben.

Für die Daten zur Effizienz E_t werden eigene Schätzungen verwendet. Als einfachste Schätzung wird angenommen, dass die Effizienz linear ansteigt. Eine weitere Annäherung soll auf Basis der in Tabelle 1 aufgeführten Werte mit den zusätzlichen Informationen zum erstmaligen Einsatz erfolgen. Mit der Annahme das die ersten drei Geräte Klassen schon verwendet werden, wird für den Bereich bis zum Auftreten der ersten ASICs eine lineare Steigerung der Effizienz angenommen. Für das erstmalige Auftreten wird der 01.02.2013 gewählt, ab diesem Zeitpunkt ist in dieser Schätzung die Zunahme der Effizienz doppelt so groß, wie im Bereich davor.

Im Zentrum der Betrachtung liegt der Zeitraum vom 25.05.2011-23.06.2013. In diesem Zeitraum ist Bitcoin schon populär genug, sodass die vorliegenden Daten genügend Aussagekraft besitzen und es genügend verwertbare Datenquellen gibt. So

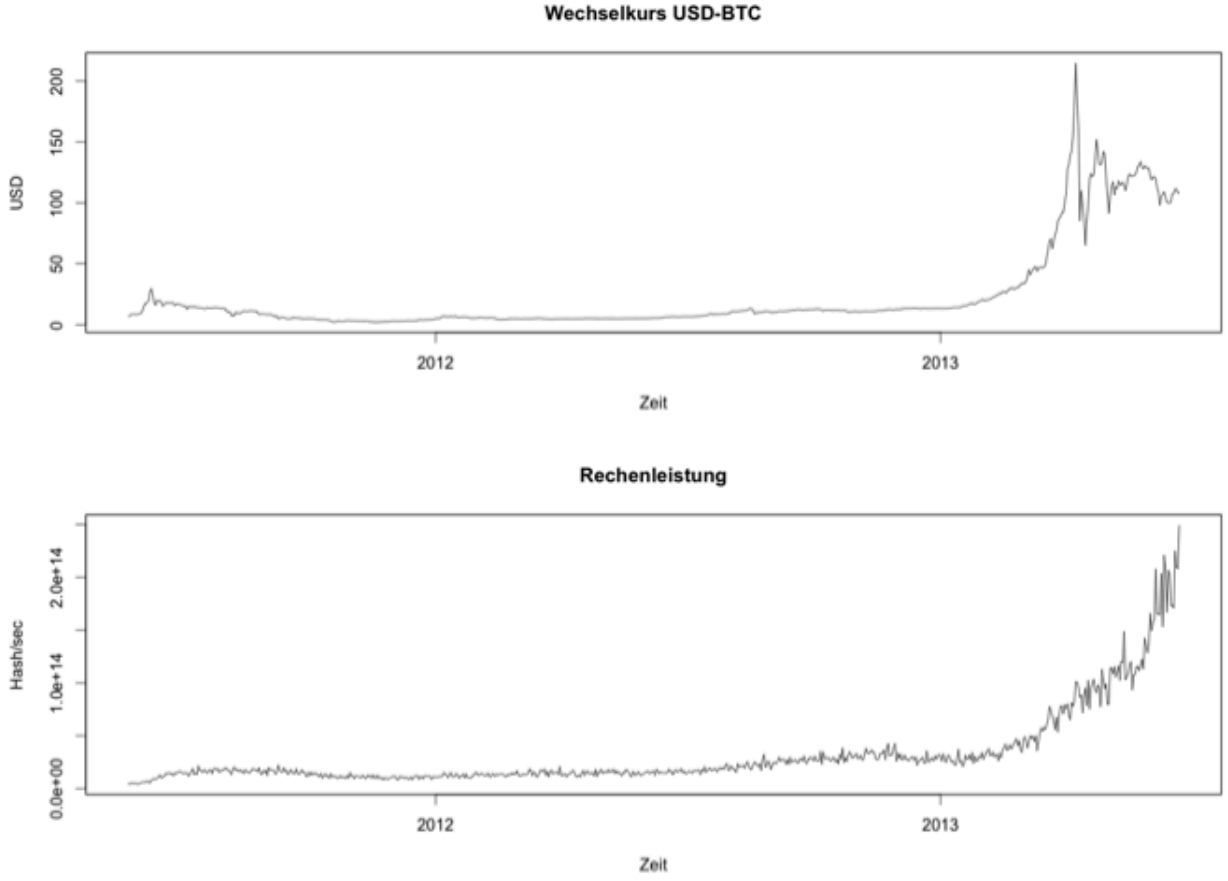

Abbildung 1: Wechselkurs und USD und Hashrate, basierend auf [5] und [2]

existieren Wechselkurse von MtGox erst seit Februar 2011 verfügbar. Bei allen Datensätzen handelt es sich um reguläre Zeitreihen, deren grundlegendes Zeitintervall einen Tag beträgt.

Für die Analysen werden zwei besondere Effekte identifiziert. Dies ist insbesondere die preisliche Entwicklung im April 2013, in Grafik 1 lässt sich hier eine starke Schwankung des Wechselkurses erkennen. Der zweite Effekte ist die zunehmende Verfügbarkeit von ASICs und die vermutlich stark ansteigende Effizienz der Mininggeräte, die potentiell einen starken Effekt auf die Analyse der Daten hat. Ein potentiell relevanter Punkt ist die Halbierung des Reward im November 2012.

Um differenzierte Analysen zu ermöglichen zwei verschiedene Zeiträume voneinander getrennt betrachtet. Hier beschreibt Z_1 den gesamten Zeitraum der vorliegenden Daten. Der Zeitraum Z_2, dessen Endpunkt am 01.02.2013 liegt, wird definiert um Analysen ohne die Effekte von ASICs zu ermöglichen.

5.2 Regressionsanalyse

Für die Regressionsanalysen werden die zuvor vorgestellten Datensätze verwendet. Hierbei wird ausgehend von einem einfachen Modell schrittweise ein komplexeres Modell entwickelt. Da Zeitreihen analysiert werden, sind mit einer hohen Wahrscheinlichkeit die Residuen nicht unabhängig voneinander.

Wechselkurs

Es wird ein einfaches Modell $M1$ konstruiert, welches den Zusammenhang zwischen Hashrate und Wechselkurs beschreibt.

$$\text{Hahsrate}_t = \beta_0 + \beta_{ER}\text{ER}_t + u_t \tag{5}$$

Das Ergebnis der Analyse ist in Tabelle 3 als Modell M_1 aufgeführt. Der hohe Wert des Bestimmtheitsmaßes lässt zunächst vermuten, dass das Modell passend gewählt wurde. Bei genauerer Betrachtung des gesamten Beobachtungszeitraum fällt jedoch auf, dass der Wechselkurs größeren Schwankungen unterworfen ist und auch einige Ausreißer aufweist. Diese stammen aus starken Schwankungen des Wechselkurses innerhalb von wenigen Tagen, im Juni 2011 und Anfang April 2013 aufgetreten sind. Im Zeitraum Z_2 ist das Bestimmtheitsmaß jedoch sehr gering, hier scheint der Wechselkurs kaum einen Erklärungswert zu haben. Da das Bestimmtheitsmaß für Trends anfällig ist, welche im Zeitraum Z_1 ab April zu beobachten sind, kann dem Modell zunächst kein hoher Erklärungswert zugeordnet werden.

Bei Betrachtung der Regression fällt zudem auf, dass die Residuen zunehmen. Es legt die Vermutung nahe, dass Heteroskedastizität vorliegt, und somit eine Verletzung der Annahmen des Regressionsmodells fehlerhaft sein könnten. Um

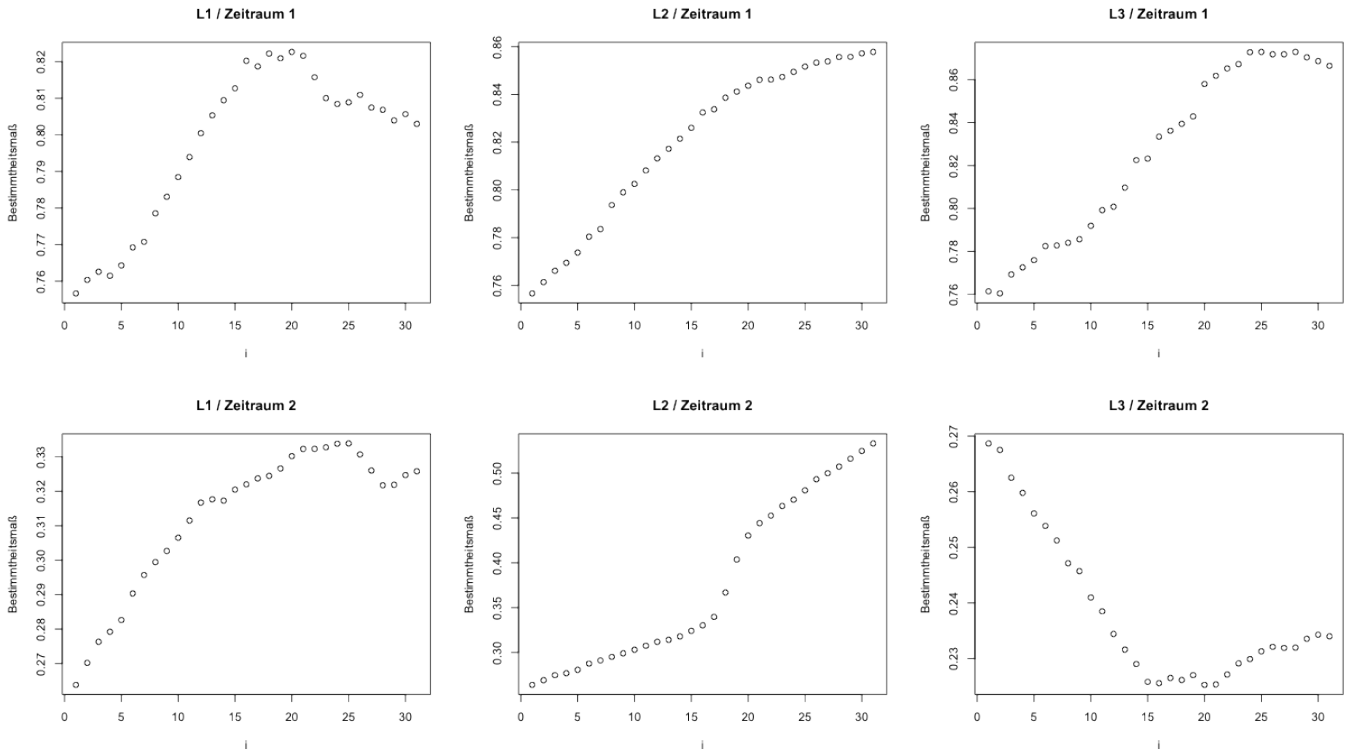

Abbildung 2: Bestimmtheitsmaß der einzelnen Lagstrukturen

dies zu bestätigen kann ein White oder ein Breusch-Pagan-Test durchgeführt werden. Der White-Test hat die Nullhypothese der Homoskedastizität. Diese kann hier nicht beibehalten werden, sodass davon ausgegangen werden muss, dass Heteroskedasitizität vorliegt. Diese kann verschiedene Ursachsen haben. Eine wahrscheinliche Ursache ist, dass das Modell noch nicht ausreichend spezifiziert ist. Zudem wäre die Verwendung eines robusten Schätzers sinnvoll.

Reaktionszeit

In dem einfachen Modell wird nicht berücksichtigt, dass eine Steigerung der Rechenkapazität nicht sofort bereitgestellt werden kann, wenn sich der Wechselkurs erhöht. Genauso lässt sich vermuten, dass ein Abfallen des Wechselkurses nicht sofort zum Abschalten der verfügbaren Rechenkapazität führt.

Dies führt zu einem Modell, in dem die Hashrate nicht abhängig von einem aktuellen Wechselkurs ist, sondern von einem in der Vergangenheit liegenden Wechselkurs. Der Effekt tritt verzögert ein. Dies wird auch als Lag bezeichnet. In der formalen Darstellung wird dazu der Lagoperator L verwendet. Er verschiebt die Beobachtung um eine bestimmtes Zeitintervall. So gilt dann:

$$L^i X = X_{t-i} \tag{6}$$

Nun lässt sich das zuvor entwickelte Modell um die Berücksichtigung des Lag zum Modell $L1$ erweitern:

$$\text{Hashrate}_t = \beta_0 + \beta_{ER} L^i \text{ER}_t + u_t \tag{7}$$

Um zu untersuchen, bei welchem Lag das Modell optimal ist, wird die Regression für jeden Wert von 0 bis 30 angewandt. Die Ergebnisse sind in Abbildung 2 dargestellt und verkürzt in Tabelle 2. Der maximale Wert des Bestimmheitsmaßes tritt bei einem Lag von 19 auf. Die Verzögerung der Reaktion der Hashrate auf den Wechselkurs würde damit 19 Tage betragen. Bei einer Betrachtung des Zeitraumes Z_2 kann eine Verzögerung von 25 Tagen festgestellt werden. Jedoch lässt sich in Z_2 die Hashrate weiterhin nur schlecht durch den Wechselkurs beschreiben.

Lag	$\overline{R}^2$		
	Z_1		Z_2
i	L1	L3	L1
19	0,8209328	0,8428792	0,3266273
20	0,8226692	0,8580144	0,3302058
21	0,8216148	0,8618345	0,3322976
...	...	...	
24	0,8084419	0,8728072	0,3338191
25	0,8088961	0,8729230	0,3338939
26	0,8109418	0,8718293	0,3307162
27	0,8074654	0,8718409	0,3260285
28	0,8068609	0,8729342	0,3217474
29	0,8039479	0,8704309	0,3218941

Tabelle 2: Lagstrukturen

Die Abhängigkeit von einem diskreten Wert in der Vergangenheit zu formulieren ist nicht realistisch, daher wird das Modell so modifiziert, dass der Einfluss verschiedener Wechselkurse an unterschiedlichen Zeitpunkte berücksichtigt werden kann. Hierzu bietet sich ein Distributed-Lag-Modell an,

welches im allgemeinen Fall folgendermaßen formuliert werden kann:

$$Y_t = \alpha + B(L)X_t + u_t \qquad \text{mit} \qquad B(L) = \sum_{i=0}^{s} \beta_i L^i \quad (8)$$

Der Term $B(L)$ beschreibt eine bestimmte Lagstruktur. Neben der angegebenen existieren noch weitere Varianten. Es kann zwischen endlichem und unendlichem Lagstrukturen unterscheiden werden. Für die Darstellung des zu untersuchenden Sachverhaltes ist es sinnvoll eine endliche zu wählen. Bei einer Entscheidung über Anschaffung bzw. Abschaltung, sind nur Zeitpunkte relevant, die nicht allzu weit vom aktuellen Zeitpunkt entfernt liegen. Auch wenn die Verzögerung durch eine Lieferzeit ist zumeist endlich. Mit der oben formulierten Lag-Struktur $B(L)$ kann das Modell $L2$ formuliert werden.

$$\text{Hashrate}_t = \beta_0 + \sum_{i=0}^{s} \beta_{ER_i} L^i \text{ER}_t + u_t \quad (9)$$

In diesem Modell muss noch bestimmt werden, wie weit in die Vergangenheit die Regressoren einen Einfluss ausüben, indem ein passendes s gewählt wird. Der optimale Wert wird auf die gleiche Weise bestimmt wie zuvor. Es ist ein kontinuierlicher Anstieg des des Bestimmtheitsmaßes zu beobachten; jedoch sind die durch die Lagstruktur ergänzten Parameter nicht signifikant. Durch die hohe Anzahl der Parameter, ist davon auszugehen dass eine Überanpassung vorliegt.

Eine Lagstruktur, wie sie zuvor untersucht wurde, bietet viel Flexibilität. Damit die Gefahr der Überanpassung reduziert werden kann, muss die Anzahl der Parameter gesenkt werden. Um dennoch eine Abhängigkeit von mehreren Variablen der Vergangenheit zu erreichen, kann der Durchschnitt eines Ausschnitts der Vergangenheitswerte verwendet werden. Für diesen muss dann nur noch ein Parameter geschätzt werden. Das daraus resultierende Modell $L3$ wird dann in folgender Form formuliert:

$$\text{Hashrate}_t = \beta_0 + \beta_{ER_a} \frac{1}{s} \sum_{i=1}^{s} L^i \text{ER}_t + \beta_{ER} \text{ER}_t + u_t \quad (10)$$

Wird der gesamte Beobachtungszeitraum beachtet, so kann ein Lag von 28 festgestellt werden. Wird jedoch nur der Zeitraum Z_2 betrachtet, so ist kein Lag mehr zu erkennen. Durch die Form der Lagstruktur kommt es zu einer Abhängigkeit der einzelnen Werte für den Wechselkurs untereinander. Da das Ergebnis nicht eindeutig ist, kann ein verzögerter Einfluss des Preises auch hier nicht bestätigt werden.

Reward

In der vorangegangenen Untersuchung wurde angenommen, dass die Belohnung für das Finden eines Blockes, bei konstant einer Höhe liegt. Ausgehend von der theoretischen Herleitung wird jedoch durch die Halbierung des Rewards die Einahmen signifikant verringert und sollte somit Auswirkungen auf die Hashrate haben. Daher wird das Modell M_1 um eine nominelle Variable zu Modell M_2 erweitert. Es wird für jeden Wert den der Reward annehmen kann, eine Dummy-Variable eingeführt. In Zeiträumen Z_1 und Z_2 ist es ausreichend, die Variable D50 zu verwenden, die den Wert eins annimmt, wenn die Belohnung 50 BTC beträgt.

$$\text{Hashrate}_t = \beta_0 + \beta_{ER} \text{ER}_t + \gamma 50 \text{D50}_t + u_t \quad (11)$$

Sowohl im Zeitraum Z_1, als auch im Zeitraum Z_2 hat die Halbierung der Belohnung einen negativen Effekt auf die Hashrate. Die Attraktivität Rechenleistung aufzuwenden steigt mit der Halbierung der Belohnung. Die ist ein Widerspruch zu dem theoretischen Überlegungen. Ein nicht berücksichtigter Trend könnte dieses Ergebnis verursachen. Dies bedeutet, dass wiederum das Modell noch nicht richtig spezifiziert ist und ein Einfluss der Belohnung nicht gezeigt werden kann. Aus der Herleitung ist erkennbar, dass der Reward multiplikativ in den Umsatz eingeht. Aus diesem Grund wird folgendes Modell M_3 formuliert

$$\text{Hashrate}_t = \beta_0 + \beta_{ER} \text{ER}_t + \gamma_{50} \text{D50}_t + \delta \text{D50}_t \cdot \text{ER}_t + u_t \quad (12)$$

Der t-Test zeigt, dass die Interaktion zwischen Reward und Preis nicht signifikant ist. Im Zeitraum Z_2 geht die Reward, wie im Modell zuvor negativ ein. Somit kann auch der multiplikative Einfluss des Rewards nicht gezeigt werden.

Transaktionsgebühren

Ein zusätzlicher Anreiz, ergänzend zu den Rewards für jeden Block, sind die Transaktionsgebühren. Bei der Analyse des Datensatzes fällt auf, dass Sie im Vergleich zu dem Betrag, der beim Finden eines Blockes ausgeschüttet, wird nur sehr klein sind. Dies wird durch eine Hinzunahme des Parameters zum Modell M_5 überprüft.

$$\text{Hashrate}_t = \beta_0 + \beta_{ER} \text{ER}_t + \beta_{TC} \text{TC}_t + u_t \quad (13)$$

Der t-Test ergibt, dass im Zeitraum Z_1 diese Variable insignifikant ist. Sie scheint somit keinen Einfluss auf die Rechenleistung des Netzwerkes zu haben. Jedoch geht sie in Z_2 zu einem kleinen Teil in die Hashrate signifikant mit ein. Es ist weiter zu untersuchen, von welchen weiteren Variablen, die Transaktionskosten abhängig sind, sodass dieser Einfluss erklärbar wird.

Effizienz

In diesem Modell wurde der im Zeitverlauf zu beobachtende Effizienzgewinn von Miningequipment nicht berücksichtigt, welcher in Kapitel 3.3 entdeckt wurde. Um diesen zu berücksichtigen kann der Zuwachs in einem ersten Schritt strikt linear geschätzt ($E1$). Dies ist insbesondere im Zeitraum Z_2, auf Grund der Vernachlässigung von ASICs, eine relativ gute Approximation.

Mit dem Aufkommen von ASICs ist diese Annahme nicht haltbar. Um dies zu berücksichtigen wird die in Kapitel 5.1 Schätzung verwendet. In der Analyse wird diese mit $E2$ bezeichnet. Um diesen Sachverhalt zu untersuchen wird das Modell $M4$ wie folgt formuliert.

$$\text{Hashrate}_t = \beta_0 + \beta_E \tilde{\text{E}}_t + \beta_{ER} \text{ER}_t + u_t \quad (14)$$

Die Regression ergibt, dass im Zeitraum Z_2, die Modellgüte sich durch den linearen Trend signifikant verbessert. Unter der Vorraussetzung, dass die Schätzung der Effizienz angemessen ist, besteht ein starker Hinweis darauf, dass dieser Zusammenhang besteht. In Z_1 wird β_0 insignifikant, sodass das Modell entsprechend angepasst wird.

Weitere Faktoren

Der Einfluss weiterer zuvor motivierten Faktoren der Risikowahrnehmung und Energiekosten werden hier nicht weiter untersucht. Für die zukünftige Forschung bietet sich an eine

Zeitraum		Z_1			Z_2	
Variablen	M_1	M_2	M_3	M_1	M_2	M_3
β_0	1,098e+13 ***	2,064e+13 ***	2,025e+13 ***	1,049e+13 ***	1,976e+13 ***	2,879e + 13 ***
β_{ER}	8,822e+11 ***	7,955e+11 ***	7,988e+11 ***	8,486e+11 ***	6,091e+11 ***	−2,165e+10
γ_{25}						
γ_{50}		−1,024 + e+13 ***	−8,651e+13 **		−8,047e+11 ***	−1,719e+13 **
δ			−1,757e+11			6,448e+11
$\overline{R}^2$	0,7567	0,7642	0,7642	0,2638	0,3649	0,3549
p-value F-Test	< 2,2e−16	< 2,2e−16	< 2,2e−16	< 2,2e−16	< 2,2e < −16	< 2,2e−16

Tabelle 3: Ergebnisse der Regressionsanalysen, Modelle M1-M3

Zeitraum		Z_1		Z_2	
Variablen	M_5	M_5	M_4	M_5	M_4
β_0			1,042e+12 ***	3,819e+12 ***	1,039e+13 ***
β_{ER}	3,706e+10 ***	6,461e+11 ***	8,706e+11 ***	5,975e+11 ***	7,628e+11 ***
β_{E1}	4,441e+11 ***			2,845e+10 ***	
β_{E2}		4,166e+10 ***			
β_{TC}			3,553e+10		4,871e+10
$\overline{R}^2$	0,8795	0,8872	0,7567	0,7168	0,2778
p-value F-Test	< 2,2e−16	< 2,2e−16	< 2,2e−16	< 2,2e−16	< 2,2e−16

Tabelle 4: Ergebnisse der Regressionsanalysen, Modelle M4 und M5

Schätzung dieser vorzunehmen. Auf Grundlage von diesen kann dann deren Einfluss genauer untersucht werden.

5.3 Prognosemodell

Die vorgestellten Modelle versuchen einen Zusammenhang zwischen abhängiger und unabhängigen Variablen zu zeigen. Da Lageffekte nicht eindeutig feststellbar eignen sich diese Modelle nur schlecht für Prognosen. In dem Blog [7] entwickelt der Autor Modelle, um die zukünftige Entwicklung der Rechenleistung im Netzwerk zu prognostizieren. Hierzu verwendet der Autor auch die Zeitreihenanalyse. Es werden 5 Modelle vorgestellt. Vier für die Vorhersage von einen Zeitraum von einer bis vier Wochen und ein Modell, welches spezielle Abweichungen im voraus erkennen soll. Bei den vier Prognosemodellen verwendet der Autor autoregressive Modelle. In denen wird die Zielvariable in Abhängigkeit zu seinen Vergangenheitswerten modelliert. In diesem Fall ist die Rechenleistung abhängig von der Rechenleistung in der Vergangenheit und der des vergangenen Wechselkurses. Das Modell formuliert er folgendermaßen:

$$\begin{aligned}\text{Hashrate}_t = \beta_{H1} L^1 \text{Hashrate}_t + \beta_{H11} L^{11} \text{Hashrate}_t + \\ \beta_{ER1} L^1 \text{ER}_t + \beta ER11 L^{11} \text{ER}_t + u_t\end{aligned} \quad (15)$$

Nach Analysen des Autors liegen die tatsächlichen Werte zum großen Teil innerhalb der des 0,95% Konfidenzintervalls seiner Prognose. Fehler treten vor allen dann auf, wenn spezielle Ereignisse eintreten, wie die Halbierung des Rewards oder die Auslieferung großer Mengen von ASICs zum einem diskreten Zeitpunkt (Batch). Dieses wird versucht durch ein Canarymodell zu entdecken, welches ausschließlich eine Abhängigkeit zwischen Wechselkurs und Rechenleistung abbildet. Autoregressive Modelle haben nur einen geringen bis gar keinen Erklärungswert auf die tatsächlichen Einflussfaktoren auf die Zielvariable. [16] Sie sind schwierig zu interpretieren und daher mit Vorsicht zu verwenden. Vorteilhaft jedoch ist ihre einfache Berechnung.

5.4 Beurteilung der Modelle

Die Modelle M_1 - M_4 weisen an vielen einigen Schwachpunkte auf und haben daher nur wenig Aussagekraft. Auch Lag-Strukturen sind nicht eindeutig festzustellen. Das Modell M_5 kann unter Vorbehalt positiv eingeschätzt werden. Zentral für die Güte des Modells sind die Schätzungen für die aggregierten Parameter. Eine genauere Analyse kann dazu führen, dass diese deutlich verbessert werden. Ansätze dafür findet sich zu einen in die Verkaufzahlen bestimmter Geräte, wie zum Beispiel durch Meiners [18] untersucht wurden. Zum anderen durch eine Analyse verschiedener Miningpools, wie sie in [8] durchgeführt werden. Um die Qualität des autoregressiven Modells genauer bestimmen zu können, müssen zunächst weitere Daten gesammelt und diese dann analysiert werden.

6. FAZIT

Durch eine Analyse des Bitcoinprotokolls und des Bitcoinökosystems konnten verschiedene Einflussfaktoren auf die im Netzwerk aufgewandte Rechenleistung identifiziert werden. Mit den vorliegenden Daten lässt sich ein Modell entwickeln, welches unter Vorrausetzung einer guten Schätzung, die Rechenleistung im Bitcoinnetzwerk durch den Wechselkurs und die Effizienz der Mininggeräte beschreibt. Für eine zukünftige Untersuchung bietet sich eine genauere Schätzung der Effizienz, sowie eine ausführliche Untersuchung des Einflusses der Risikoaffinität der Miner an.

7. REFERENCES

[1] http://bitcoincharts.com/charts/volumepie/.
[2] http://bitcoincharts.com/markets/.
[3] http://bitcoinmagazine.com/asicminer-starts-hashing/.
[4] http://blockchain.info/de/charts/market-cap.
[5] http://blockexplorer.com/q/getdifficulty.
[6] http://launch.avalon-asics.com/.
[7] http://organofcorti.blogspot.de/.

[8] http://organofcorti.blogspot.de/2013/06/pool-and-network-electricity-use$_2$5.*html*.
[9] https://en.bitcoin.it/wiki/blocks.
[10] https://en.bitcoin.it/wiki/mining.
[11] https://en.bitcoin.it/wiki/mining$_h$*ardware*$_c$*omparison*.
[12] https://en.bitcoin.it/wiki/transaction.
[13] https://en.bitcoin.it/wiki/transaction$_f$*ee*.
[14] http://www.bitkom.org/de/presse/8477$_7$6044.*aspx*.
[15] http://www.energy.eu/.
[16] P. Hackl. *Einführung in die Ökonometrie*. Pearson Studium, 2005.
[17] J. Kroll, I. Davey, and E. Felten. The economics of bitcoin mining or, bitcoin in the presence of adversaries. 2013.
[18] J. Meiners. Bitcoin hardware. 2013.
[19] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2009.
[20] M. Rosenfeld. Analysis of hashrate-based double-spending. 2012.
[21] C. Sorge and A. Krohn-Grimberghe. Bitcoin: Eine erste einordnung. *Datenschutz und Datensicherheit*, 2012.

Alle Webseiten wurden zuletzt am 28.06.2013 abgerufen.

8. ACKNOLOWLEDGEMENTS

Ich danke insbesondere meinen unbekannten Gutachtern für ihr sehr ausführliches und qualitativ hochwertiges Feedback. Auch möchte ich Jens Meiners danken, der mir nützliche Hinweise zur Recherche von Daten zu aktuellen Bitcoin Mining-Hardware geben konnte.