"Science Stays True Here"
Advances in Ecological and Environmental Research (ISSN 2517-9454, USA) | Science Signpost Publishing

Forecasting EU Allowance Price Using Transactions of Central Entities

Sotirios Dimos[1], Eleni Evangelatou[1], Dimitris Fotakis[1], Andreas Mantis[1], Angeliki Mathioudaki[1]

1. School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou 9, 15780 Athens Greece.

Contact: Sotirios Dimos Email: sdimos@corelab.ntua.gr
Eleni Evangelatou Email: eevangel@corelab.ntua.gr
Dimitris Fotakis Email: fotakis@cs.ntua.gr
Andreas Mantis Email: amantis@corelab.ntua.gr
Angeliki Mathioudaki Email: tmathiou@corelab.ntua.gr

Received: March 11, 2020 / Accepted: April 12, 2020 / Published: Vol. 5, Issue 06, pp. 187-209, 2020

Abstract: European Union Emission Trading System (EU ETS) can be analyzed as a transaction network. Under the rules of ETS, firms must hold enough allowances to cover for their annual emissions. Excess allowances may be traded or banked for future use. The allowance price is considered as an indication of how well the policy meets its objectives. We aim for a better understanding of the structure of the network and of the way the market dynamics affect the EUA price. Our findings indicate that there exists a small subset of nodes which emerges as core transactors and accounts to roughly 1.8% of the entire network participants. We call it *All Time Almost Dominating Set*, *(ATADOM)*. Aiming to quantify its power, we provide a Vector AutoRegressive (VAR) model to forecast the EUA price. By using trading quantities restricted to ATADOM, along with significant price determinants from the literature, we find that it is possible to explain and to forecast EUA price by just tracking the behavior of the ATADOM set within EU ETS. Our work can be the basis of a compact evaluation tool to one of the most prominent environmental policies, the EU ETS.

Keywords: EU ETS; Trading Network; Core-Periphery; Dominating Set; EUA Price Forecasting; Vector AutoRegression

1. Introduction

The European Union Emission Trading System (EU ETS) [***1***], [***2***] was established in 2005 as a response to Kyoto Protocol [***3***]. The EU ETS is the largest cap-and-trade system in the world, regulating 31 European countries[1]. The system covers around 45% of the EU's Greenhouse gas (GhG) emissions by regulating approximately 11,000 stationary installations and 500 aviation

[1] 28 EU Member States plus Iceland, Norway and Liechtenstein

companies.

Under the cap-and-trade principal, the regulator (i.e. European Comission), sets a ceiling on the overall emissions and issues a corresponding volume of allowances (European Union Allowance or EUA). Each one of the allowances gives the holder the right to emit GhG equivalent to 1 tonne of carbon dioxide (CO_2). Any regulated emitter must possess and submit by the end of the regulating period an amount of allowances equivalent to its actual emissions[2]. Each year a decreasing proportion of EUAs is freely allocated. The rest is auctioned. Firms in excess of allowances can keep them for future use (EUAs are bankable) or sell them to another company that is short of allowances. Financial intermediaries (e.g. banks, brokers) can also participate in the EU ETS in order to ease the transactions. The most common type of trade is the future contracts [***2***], [***4***].

The price of the Allowance is an important aspect of any cap-and-trade system, as it shows how well the system performs. Since the first year of EU ETS operation, the system has been criticized mainly for the observed price instabilities [***4***], [***5***], [***6***] and its poor market design [***7***], [***8***]. Many studies on the drivers of the EUA price, show that energy prices, weather variations, offset usage, industrial activity and economic variations are significant EUA price determinants [***9***], [***10***], [***11***], [***12***], [***13***], [***14***]. Furthermore, many recent studies have used the above factors as an attempt to model the EUA price [***15***], [***16***], [***17***], [***18***] and to forecast its volatility [***19***], [***20***], [***21***].

A main characteristic of the EU ETS is the European Union Transaction Log (EUTL). EUTL is an electronic system that records every transaction made within the EU ETS. The data provided by the EUTL can be used to recreate the transaction network allowing for a better understanding of the market dynamics.

The work in this paper makes a significant contribution in analyzing the structure of a large transaction network as a way to capture its properties. We aim to gain a better understanding in the formation of the market as well as the EUA price, by reconstructing the EU ETS trading network. Technically, we simplify the network structure by considering a smaller set of participants that dominates the market, which we call *All Time Almost DOMinating set (ATADOM)*. We claim that such a subset, albeit tiny in size, incorporates all the necessary information about the trading volumes. To strengthen our claim, we quantify the allowance trading volume, by using the data of the actual transactions both for the total transaction network and for the smaller set of participants. Finally, we prove that this tiny set of participants is adequate, not only in understanding the EU ETS structure, but also in providing better forecasting results for the EUA price.

The remainder of this paper is organized as follows: Section 2 provides the reader with a summary of the characteristics of the European Union Transaction Log (EUTL) which is one of the main source of data. An analysis of allowance exchange network along with the definition and the

[2] Since 2013, failure to comply leads to 100€ fine for every missing EUA, along with the obligation to surrender the missing EUAs [***2***].

characteristics of ATADOM are provided in section 3. Section 4 presents the Vector AutoRegressive model used to forecast EUA's allowance price with its forecasting results. Finally, section 5 concludes our work.

2. The European Union Transaction Log

The main goal of the transaction log is to ensure the accurate accounting of the allowances within the ETS. The EUTL provides public access to all the transactions that have taken place up to three years before the current date. Every participant in the EU ETS, either with the obligation to surrender allowances for compliance or not, that wishes to trade allowances, may have many different (but at least one) *accounts* within the EUTL. In the dataset that we obtained from the EUTL in May 2019, we identified more than 40,000 accounts in total. We also identified about 16,500 different account holders (i.e. entities that open and manage an account). About 9,500 of them were associated with at least one open account in May 2019, about 13,000 of them had participated in at least one transaction in the past, and about 8,500 account holders both were associated with an open account and had participated in at least one transaction. The term *active* account (or account holder) refers to an account that had a certain activity (allowance exchange) at some time point. We highlight that an active account does not necessarily imply that this account found open on May of 2019. The term active may refer to the past.

Table 1: Number of EU ETS participants for each category

Category	active[a]		active & open[b]	
Governmental	54	0.40%	54	0.63%
Regulated	10,708	80.15%	7,793	91.46%
Financial	2,598	19.45%	674	7.91%

[a]Account Holders with at least one transaction
[b]Account Holders with at least one transaction in the EUTL on May 2019

The EUTL does not distinguish account holders according to their role in the system. In our data analysis, we consider an aggregating allowance activity per account holder and we classify account holders as regulated, governmental and purely financial, based on their main activity in the ETS. We consider (1) as *governmental*, the participants such as European Commission and the ministries and administrative bureaus of the countries members in the ETS, (2) as *regulated*, the participants that have the obligation to surrender allowances for their annual GhG emissions and (3) as *financial*, the entities which participate in the EU ETS to serve their own interests (e.g. brokers, investors) and trade allowances. The vast majority of the participants that have been active (80.15%) are regulated.

Table 1 shows the share of each category in the population of the participants.

In our analysis we exploit this rich source of data by calculating the allowance trading volumes of regulated and financial entities (see Table 6).

3. The EUA transaction network

The process of exchanging EUAs can be naturally modelled as a graph (or network) $G(V, E)$. In our analysis we consider a different network for every month. The set of nodes V are the firms or bureaus which conduct EUA transactions. We consider a single edge $\{v, w\} \in E$ between two nodes v and w, if they transact with each other at least once. We do not take into account the exact number of transactions or the overall amount of EUAs exchanged. The nodes are labeled as **regulated**, **governmental** or **financial**, according to their participating role in the EU ETS, as introduced in section 2. We note as $\mathcal{L}(v \in V)$ that label or category.

Given a node $v \in V$, we call the nodes adjacent or connected to v as the *neighbours* of v. We note as $\mathcal{N}(v) = \{w \in V: \{v, w\} \in E\}$ the set of v's neighbours. The number of edges adjacent to v is called degree of v and it is noted as $deg(v)$.

3.1 Almost Dominating Set

Previous studies [*7*], [*22*] have shown that within the EU ETS there is a small group of participants that they are notably active and transact with most of the rest. We are interested in finding such a group.

We define the *(node) coverage* of a subset D of V as the portion of nodes which are connected to or 'covered' by D. A *dominating set* for a graph $G(V, E)$ is a subset D of V such that every vertex not in D is adjacent to at least one member of D. Such a set exhibits 100% coverage. Of course, finding a dominating set is more meaningful when it contains the smallest number of vertices possible i.e. **a minimum dominating set**. Unfortunately finding the minimum dominating set is considered as a hard problem by the computer science community. This means that it is widely believed, that an algorithm which can perform such a computation efficiently for every possible graph does not exist. Technically, the problem: *"Given an arbitrary graph $G(V, E)$ and a positive integer d, is there a dominating set $D \subset V$ of cardinality d?"* is NP-complete (see problem [GT2] in [*23*]).

In our analysis, we are interested to locate quickly an adequate subset that makes a lot of transactions with the rest of the network. We do so heuristically via Algorithm 1 which computes an **Almost Dominating Set** $\mathcal{D}$. It initially excludes all nodes with at most one neighbour. Next, it orders the remaining nodes by their degree in decreasing order. Then Algorithm 1 repeatedly visits the ordered nodes one by one. A node is chosen if it has at least one neighbour that is not already visited (see lines 6, 7 of Algorithm 1).

Even though the outcome of Algorithm 1 is not a dominating set, it suffices for our purposes. We are just interested in identifying (a) a tiny subset (b) that "aggregates the information" embodied within the allowance exchange network. Concerning the former goal, the Almost Dominating sets $\mathcal{D}$ are indeed much smaller than the corresponding set of nodes V. As shown in Table 2 on average only 6.15% of the nodes are chosen by Algorithm 1. The cardinality of Almost Dominating sets, $|\mathcal{D}|$ varies from 19 to 160. In contrast the average cardinality of the number of nodes $|V|$ is 1,865.

Table 2 shows some statistics for the size of V and $\mathcal{D}$. The third line shows the size of the Almost Dominating set $|\mathcal{D}|$, relative to the size of nodes in the corresponding graph $G(V, E)$.

Algorithm 1: Greedy Algorithm for Calculation of Almost Dominating Set

Input: $G(V, E)$ a simple undirected unweighted graph
Output: $\mathcal{D}$ an Almost Dominating set for graph G
1: Function AlmostDomSet$G(V, E)$
2: $\quad \mathcal{D} \leftarrow \emptyset$
3: $\quad visited \leftarrow \{n \in V | deg(n) \leq 1\}$ *// all nodes with degree 1 are included in visited set*
4: $\quad$ Sort V in decreasing degree order
5: $\quad$ For $n \in V$ // algorithm visits nodes in decreasing degree order
6: $\quad\quad$ If $(\mathcal{N}(n) \backslash visited) \neq \emptyset$ *// n has at least 1 not visited neighbour*
7: $\quad\quad\quad \mathcal{D} \leftarrow \mathcal{D} \cup \{n\}$
8: $\quad\quad\quad visited \leftarrow visited \cup \mathcal{N}(n) \cup \{n\}$
9: $\quad\quad$ End if
10: $\quad$ End for
11: Return $\mathcal{D}$ the Almost Dominating Set
12: End function

The allowance exchange graphs differ significantly throughout the months. This can be seen in Table 2 where the entities exchanging EUAs vary from 242 to 6,993. Table 3 shows the average size of V and $\mathcal{D}$ for each different month. The largest graphs (more participants $|V|$) are on February and April. Most of the participants are regulated (more than 80% as shown in Table 1), and the fact that they acquire free allowances on February and they surrender allowances on April, results on larger graphs. However, many of the regulated entities only transact with just a governmental node (just one neighbor). Consequently these nodes do not transact with many other participants and Algorithm 1 does not choose them (see line 3). This fact is depicted in Table 2 (see $mean(\frac{|\mathcal{D}|}{|V|})$) where the size of Almost Dominating sets associated with February and April are even smaller relatively to the number of the corresponding graph's nodes.

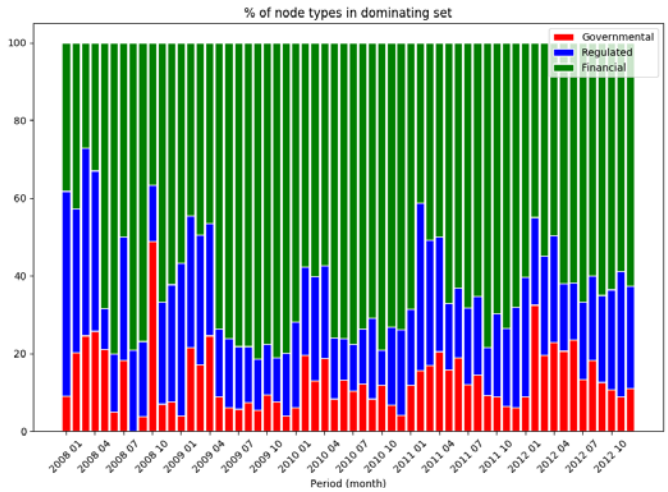
Table 2: Statistics for the number of nodes, |V|, the number of the members of the Almost Dominating sets, |D| and the for the monthly networks since January of 2006 until April 2016

	min	median	max	mean	std
$\|V\|$	242	966	6,993	1,865	1,810.46
$\|\mathcal{D}\|$	19	67	160	71	30.84
$\frac{\|\mathcal{D}\|}{\|V\|}$	0.32%	6.34%	12.57%	6.15%	2.99

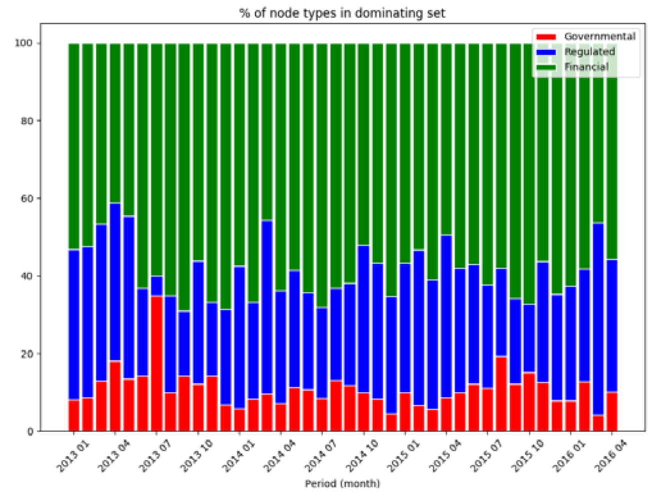
The Almost Dominating Sets are composed mainly of financial nodes as depicted in figure 1, where the green bars cover most of the plotting area. This is another indication of the fact that financial entities play an intermediate role as found by [*7*].

Table 3: Average number of nodes, |V| and average number of the members of Almost Dominating set, |D|, for each different month from January of 2006 until April 2016.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**
$mean(\|V\|)$	924	4081	2746	5921	1113	658
$mean(\|\mathcal{D}\|)$	69	67	105	103	66	55
$mean(\frac{\|\mathcal{D}\|}{\|V\|})$	7.74%	2.37%	4.05%	1.73%	6.66%	8.39%
	Jul	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
$mean(\|V\|)$	1238	582	638	964	1136	1761
$mean(\|\mathcal{D}\|)$	48	48	51	52	73	104
$mean(\frac{\|\mathcal{D}\|}{\|V\|})$	7.23%	8.23%	8.07%	7.13%	6.78%	6.24%



Phase II (2008 – 2012)



Phase III (up to April 2014)

Figure 1: Category composition of Almost Dominating Sets for each month

The Almost Dominating Sets $\mathcal{D}$ are indeed tiny (see Table 2 and Table 3). However we need a strong indication of its property of 'information aggregator'. A network of allowance exchange incorporates information about the transactions of allowances. So we need for the members of the Almost Dominating Set to be aware of most of the transactions that have taken place. To check if this is the case we introduce edge weights.

A weighted graph $G(V, E, w)$ is a graph $G(V, E)$ where each edge $\{a, b\} \in E$ is annotated with a numeric information, its weight $w(\{a, b\})$. The weight of that edge, is the total volume exchanged between entities a and b. We highlight that we do not designate the directions of the transactions. So if a transfers x allowances to b and then receives back y, then $w(\{a, b\}) = x + y$. The total weight of a set of edges is the summation of the weight of the edges belonging to that set $w(E_1 \subseteq E) = \sum_{e \in E_1} w(e)$.

Given a partition of a graph's nodes $\mathcal{D}$ and $V \backslash \mathcal{D}$ we partition the set of edges in the following three groups

$$E_{inside\,\mathcal{D}} = \{\{a, b\} \in E: a, b \in \mathcal{D}\}$$

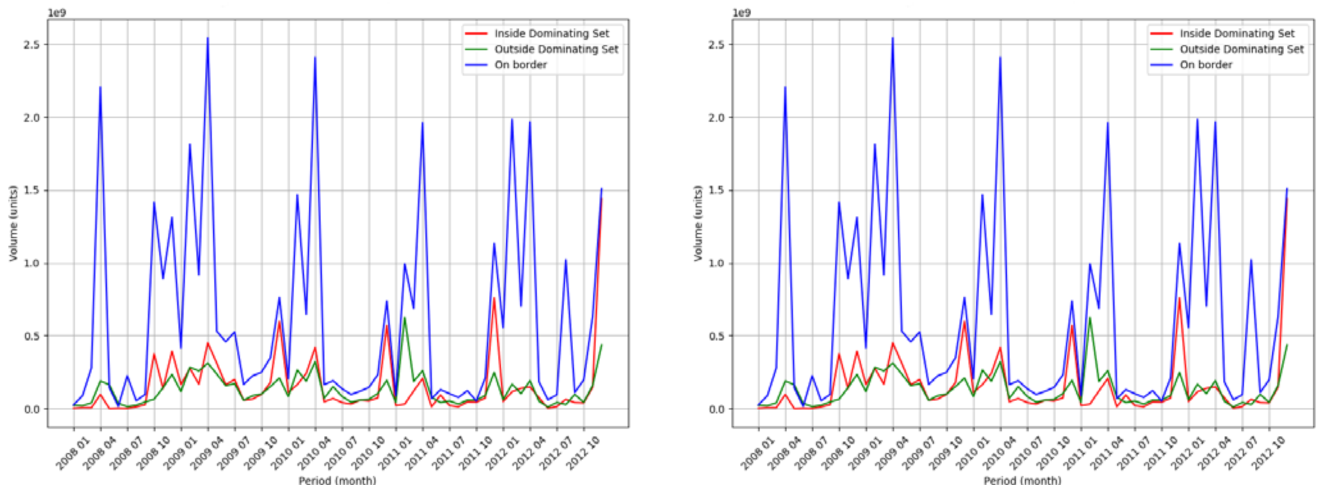
$$E_{border\,\mathcal{D}} = \{\{a, b\} \in E: a \in \mathcal{D} \quad \text{and} \quad b \in V \backslash \mathcal{D}\}$$

$$E_{outside\,\mathcal{D}} = \{\{a, b\} \in E: a, b \in V \backslash \mathcal{D}\}$$

We call the set $E_{border\,\mathcal{D}}$ the border of $\mathcal{D}$ and it consists of the edges with exactly one end point in $\mathcal{D}$. For the output of Algorithm 1 to encompass the information embodied in the original graph it must be the case where

$$w(E_{border\,\mathcal{D}} \cup E_{inside\,\mathcal{D}}) \approx w(E)$$

that it is most of the transacting volume falls either onto the border or inside the Almost Dominating set $\mathcal{D}$. Indeed this is true in our case as depicted in Figure 2 where the blue ($w(E_{border\ \mathcal{D}})$) and the red ($w(E_{inside\ \mathcal{D}})$) curves sum up to an adequate enough share of the total transacting volume. This corroborates that Almost Dominating Sets are "information aggregators" of the EU ETS transaction networks.



2nd phase 2008 – 2012 3rd phase up to April 2014

Figure 2: Total transacting volume inside, outside and on the border of the Almost Dominating sets.

3.2 ATADOM: All Time Almost DOMinating set

Section 3.1 provides strong indications that all allowance-exchange networks, considered on a monthly basis exhibit the core-periphery structure. In other words for each month there is a small set of entities whose behavior approximates the trading pattern of EU ETS. However instead of "tracking" a different set of entities every month, it would be of great interest if there exists a small but constant in time set of participants with the same properties as Almost Dominating sets introduced in section 3.1.

Fortunately enough there is such a set, we call it **A**ll **T**ime **A**lmost **DOM**inating set or *ATADOM* and it is computed by Algorithm 2 found in the Appendix. ATADOM basically consists of nodes that belong to many Almost Dominating sets (computed by Algorithm 1). Algorithm takes as input (a) a coverage threshold, $\vartheta \in (0,1)$, (b) a list of allowance exchange graphs $G[\ ]$ and (c) the corresponding to $G[\ ]$ list of Almost Dominating sets $D[\ ]$. It initially computes the union of all Almost Dominating sets $\mathcal{D}_{\text{union}}$. The algorithm then calculates the *appearance frequency* of entity

v, i.e. the number of Almost Dominating sets, v is a member. The elements of $\mathcal{D}_{\text{union}}$ are ordered decreasingly by their appearance frequency. Finally, for every input graph in $G[\ ]$, the algorithm iteratively includes in ATADOM the most frequent nodes of $\mathcal{D}_{\text{union}}$ until the former set covers at least a ϑ fraction of the nodes of the currently examined graph. For a more detailed representation of the aforementioned procedure, see Appendix.

Table 4: Composition of ATADOM by category

Category	ATADOM	
	entities	% share
governmental	25	10.64%
regulated	44	18.72%
financial	166	70.64%
sum	**235**	100%

Table 5: ATADOM's share in the allowance exchange from 2005 until May 2019

	transactions		allowances exchanged	
inside ATADOM	$163 * 10^3$	19.30%	$36.89 * 10^9$	31.97%
ATADOM's border	$555 * 10^3$	65.71%	$66.06 * 10^9$	57.26%
outside ATADOM	$127 * 10^3$	14.99%	$12.43 * 10^9$	10.77%
sum	$845 * 10^3$	100%	$115.38 * 10^9$	100%
involve ATADOM[a]	$718 * 10^3$	85.01%	$102.95 * 10^9$	89.23%

[a]At least one of the two traders is a member of ATADOM. In other words is the union of the transactions taking place inside or on the border of the ATADOM.

The outcome of the above algorithmic process with coverage threshold $\vartheta = 97\%$, is a compact aggregation of the EU ETS network. Table 4 reveals ATADOM's tiny size. Just 235 entities which make up only 1.76% of the registered participants (see Table 1 and Table 4) that have transacted at least once! Most of them (166 out of 235) are financial, which fact is another indication of their intermediate role. The information accumulation property of ATADOM is justified in Table 5. Amazingly enough, 85% of transactions are done through or within ATADOM! Its share in terms of transaction volume is even greater, reaching 89.23%.

4. EUA Price Forecast using EUTL information

Allowance prices are often viewed as an indication of how well the system is functioning [*4*]. Thus, the EUA price formation is a subject that interests many scientists throughout the years.

Although many factors about economic variations and energy prices have been studied extensively either as price determinants or for forecasting the EUA price, there is a gap concerning the volume of transactions [***4***], [***12***], [***16***], [***24***] as an EUA price driver.

Our purpose is to forecast the EUA future price, on a weekly basis, by exploiting as explanatory factors both known EUA price drivers (see [***9***], [***10***], [***11***], [***12***], [***13***], [***14***]) and quantities derived from the allowance trading network (see Table 6). Our results indicate that in terms of forecasting accuracy, by exploiting the EUTL transaction data we can achieve better results than exploiting only market fundamental factors. Additionaly, there is no actual gain by computing the EUTL variables from the whole network. In many cases forecasts are more accurate as we restrict these variables in ATADOM's behavior. This fact strengthens our claim that this small group of participants summarizes the EU ETS and provides qualitative information about the allowance trade.

More specifically, we went through the following steps: (i) we focused on the available data of Phase III of the EU ETS (2013 - 2020; available data until April 2016) (ii) the time series for each variable was assessed for stationarity; (iii) we considered many possible combinations of the variables; (iv) appropriate lag was determined using a lag-length selection criteria; (v) we used Vector Autoregressive (VAR) analysis to estimate various models from all possible combinations and lags; (vi) we performed a residuals analysis and VAR stability checking; (vii) we concluded to the best forecasting models for the EUA price. These steps were implemented in Matlab and sanity checks was performed using Gretl (Gnu Regression, Econometrics and Time-series Library)

4.1 Vector Auto-Regression Analysis

In our analysis, we consider multivariate time series $\mathbf{Y}_t = (y_{1,t}, \ldots, y_{k,t})^T$ with EUA price being a component of vector $\mathbf{Y}_t$. The other components are associated with possible interdependent factors. We adopt this approach to take advantage of possible cross-correlations with other variables. To that end, we use Vector Vector Auto-Regressive (VAR) models to represent the multivariate series in consideration.

A **stationary** multivariate (see e.g., [***25***], chap. 7] for an introduction to multivariate time series models and forecasting) time series $\mathbf{Y}_t$ can be represented by a VAR(p) model of p lags given by

$$\mathbf{Y}_t = \delta + \sum_{l=1}^{p} \mathbf{\Phi}_l \mathbf{Y}_{t-l} + \epsilon_t \tag{1}$$

If $\mathbf{Y}_t$ has k components, then $\mathbf{Y}_t, \delta, \epsilon_t$ are $k \times 1$ vectors and $\mathbf{\Phi}_l$ are $k \times k$ matrices. The residual vectors ϵ_t is a white noise of mutually independent random vectors following multivariate normal distribution with zero mean. If a multivariate time series Y_t can be approximated by a VAR(p) model, using an adequate number of observations, one can estimate the intercept vector δ

and the matrices $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_l$.

Once a VAR(p) model is estimated, it can be used to forecast future values. We let $\widehat{\mathbf{Y}}_t(\tau)$ denote the so-called τ-lead (or τ-ahead) forecast about time t, i.e., the forecast about time t made using information available up to time $t-\tau$. Given the realization of the actual observation $\mathbf{Y}_t$, the τ-ahead forecast error $e_t(\tau)$ is

$$e_t(\tau) = \mathbf{Y}_t - \widehat{\mathbf{Y}}_t(\tau) \tag{2}$$

4.2 Data Analysis and Forecast Assessment

In this paper, we considered known EUA price drivers as many studies suggest [***5***], [***9***], [***10***], [***11***], [***12***], [***13***], [***14***]. We collected a variety of data about known price determinants such as (i) data about the EUA price and energy futures prices aggregated from financial websites (ii) data about indicators for industrial production activity and economic variations; we also calculated transaction volumes from the data extracted from the EUTL. These quantities where calculated both for financial and regulated entities and represent the acquiring volume of allowances i.e. incoming volume, and the transferring volume of allowances i.e. outgoing volume. Finally, we also considered the total volume of allowances, i.e. the aggregate amount of allowance trading, for every group of entities. For our purpose, we considered two versions for all variables relevant to the EUTL: (a) calculated from all the EU ETS participants, which we call *whole EUTL data* for now on and (b) calculated solely from ATADOM's members, which we call *ATADOM data* for now on. All the data used in this work are on a weekly basis. For a detailed representation about the data and their sources see Table 6.

We tested many models consisted of various combinations of the variables in Table 6. Although the VAR model, equation (1), is a system of equations, in this paper we focused on forecasting the EUA price only.

Table 6: List of the variables used in our analysis along with the corresponding data sources

Variable Name	Description	Data Source
EUA	Futures price of EUA	EEX
Coal	Futures price of coal	investing.com
Ngas	Futures price of natural gas	
Oil	Futures prices of oil	
IPI	Industrial Production Index	Eurostat
ESI	Economic Sentiment Index	
TVolReg	Total allowance volume exchanged by regulated entities	EUTL
TVolFin	Total allowance volume exchanged by financial entities	
InVolReg	Incoming allowance volume of regulated entities	
OutVolReg	Outgoing allowance volume of regulated entities	
InVolFin	Incoming allowance volume of financial entities	
OutVolFin	Outgoing allowance volume of financial entities	

Testing for Stationarity: Stationarity is critical to develop a VAR model. Hence, as a first step, we had to assure that our data is stationary by removing trends and seasonality from every variable. We used the Augmented Dickey-Fuller (ADF) test [***26***], to test if our time series are stationary. The null hypothesis is that the time series is nonstationary, thus rejecting the null hypothesis indicates that the series does not need transformation to achieve stationarity. We log-transformed and differentiated the data as needed and we re-applied the ADF test to the transformed time series. We should highlight that some of the explanatory variables (see EUTL data in Table 6) are in a much larger scale than the EUA price. Hence, log-transformations are useful in our analysis as they reveal the relative (or percentage) changes on the original scale.

Lag-Length Selection: Lags refer to the number of previous observations in a time series to be included in the VAR model. Selection of the appropriate lag-length is an important step of the procedure, as selecting too few may lead to autocorrelated residuals while selecting too many may cause overfitting. Lag length selection for VAR models can be based either on ad-hoc choice or in a variety of statistical criteria [***27***]. We chose the appropriate number of lags by the Akaike

Information Criterion (AIC), Schwarz Bayesian Information Criterion (BIC), and the Hannan-Quinn criterion (HQ). For every considered set of variables we estimated a VAR model for one up to six lags.

Performing Residuals diagnostics: Once a VAR model has been constructed, the next step is to determine if the model provides an adequate description of the data. This is performed by testing the residuals of the model, i.e. the differences between the actual and the fitted values. Here, we had to examine residuals for autocorrelation, heteroscedasticity and normality. Namely, we used (a) the Ljung-Box Q test (null hypothesis: no autocorrelation) to check for residual autocorrelation, (b) the Engle's ARCH test (null hypothesis: no heteroscedasticity) to check for residual heteroscedasticity and (c) the Kolmogorov-Smirnov test (null hypothesis: residuals are normally distributed) to test for normally distributed residuals. If a model satisfies all the above conditions, it can proceed to the stability checking test.

Assessing Stability of selected models: Following the residual diagnostics, we had to check the stability of the selected models, i.e. whether a model is a good representation of the evolution of the time series over the sampling period. The stability of the VAR model is visually inspected by Figure 3. The dots represent the roots of the characteristic polynomial associated to $\boldsymbol{\Phi}$ of equation 1. If the roots lie inside the unit circle, then the VAR system satisfies stability condition.

Assessing Forecast Accuracy: Goodness of fit based on the residuals analysis is not a reliable indication for forecast performance. The forecast accuracy can only be evaluated by how well a model performs on data not used when fitting the model. Thus, it is common practice to separate the available data sample in training set, which is used to estimate the parameters of a model, and in testing set, which is used to evaluate the forecast performance.

In our work we used two versions of training-testing sets: (i) used the 80% of the observations to estimate the model (training set) and the rest 20% for forecasting (testing set) and (ii) time series cross-validation, where starting with the 80% of the observations we expanded the window of training set with one observation at a time and we re-estimate the model. The forecast accuracy is computed by the average forecast errors over the test sets. A forecast error is the difference between the observed value and its forecast.

In order to measure the forecast accuracy of our models, we will abuse a little the notation in (2) for the τ-ahead forecast error to note only the forecast error for the EUA price instead of the whole error vector. We used the Mean Absolute Error (MAE) of the one-step ahead forecast errors

$$\mathrm{MAE}(n) = \sum_{t=1}^{n} |e_t(1)| \tag{3}$$

and the Mean Squared Error (MSE) of the one-step ahead forecast errors

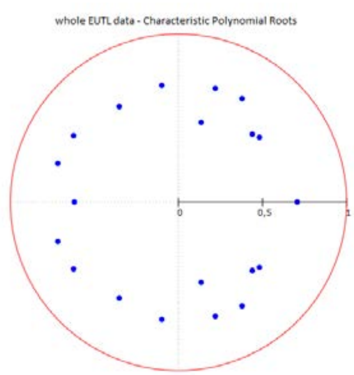
$$\mathrm{MSE}(n) = \sum_{t=1}^{n} e_t(1)^2 \tag{4}$$

The number n in equations (3), (4) is how many forecasts are made and we call it the *forecasting horizon*.

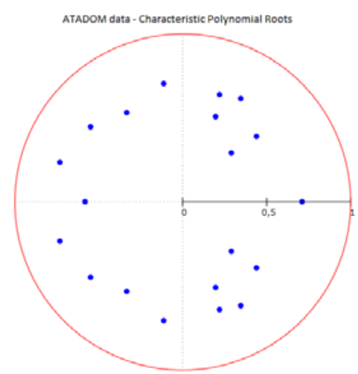
4.3 Discussion of the Forecast Results

The model that satisfies the residuals diagnostics based on the tests mentioned in the previous section, is a 5-dimentional VAR with 4 lags and the variables are shown in Table 8. Namely, *EUA* stands for allowance price, *IPI* for Industrial Production Index, *ESI* for Economic Sentiment Index, *TVolFin* for total EUA exchange where at least one transactor is financial entity, *TVolReg* for total EUA exchange where at least one transactor is regulated entity.

The PValues of the Ljung-Box Q test, the Eangle's ARCH test and the Kolmogorov-Smirnoff test, are shown in Table 7. In Figure 6 and in Figure 5 one can also see the normal plots of the residuals for the whole EUTL dataset and the ATADOM set respectively. The unit roots of the VAR models both for the whole EUTL dataset and the ATADOM set lie inside the unit circle, as one can see in Figure 3.



Unit Roots for VAR whole EUTS



Unit Roots for VAR ATADOM

Figure 3: VAR Characteristic polynomial roots

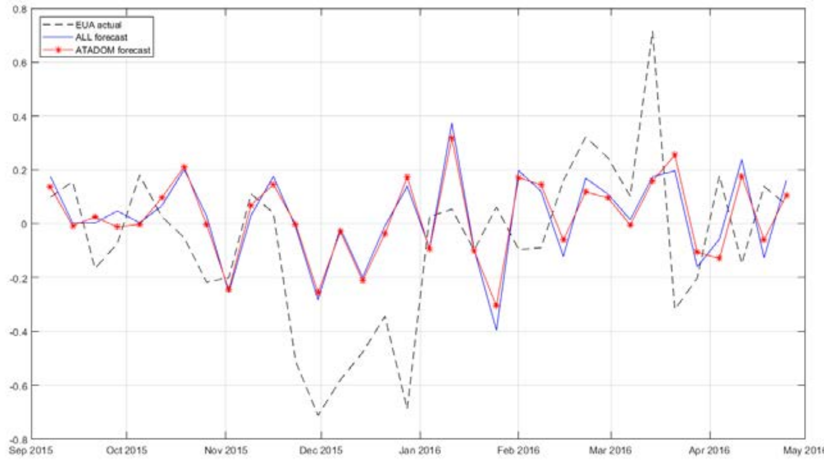


Figure 4: Forecast comparison of the two chosen models

Overall, we justify the claim that the forecast can improve by tracking the allowance trading volumes of the ATADOM. In Table 8, we summarize the one-step ahead forecast errors for the selected model. Mean errors of the model estimated with the ATADOM set are smaller comparatively to the model estimated with information from the whole EUTL data. In Figure 4 one can visually observe the forecast comparison between the models with the whole EUTL dataset (blue line) and the ATADOM set (red line) along with the actual observations for the forecast horizon.

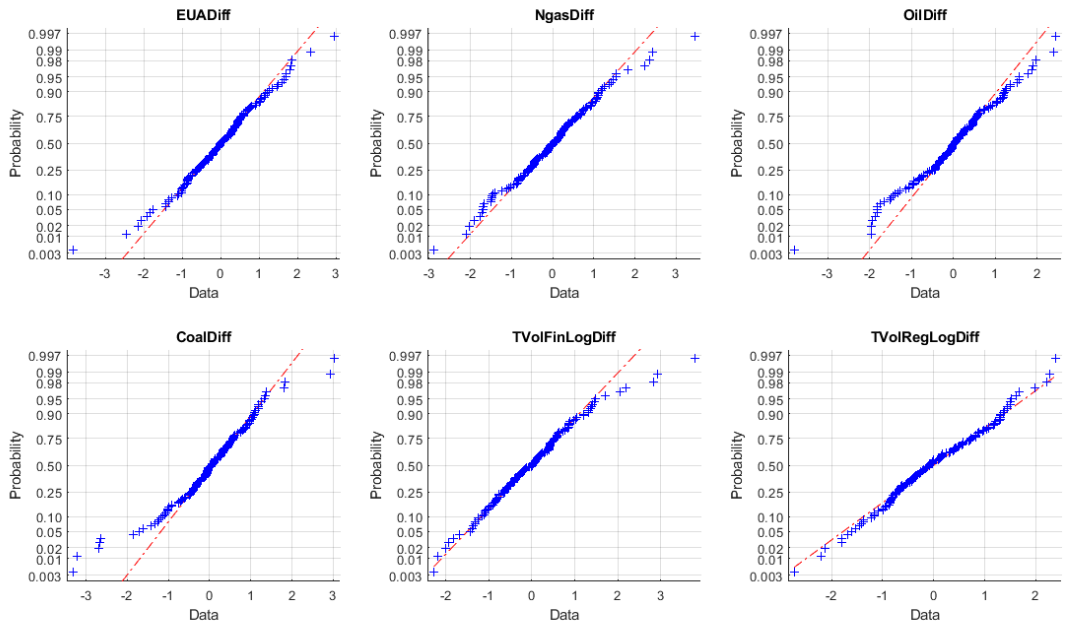


Figure 5: Normal plots of the residuals of the VAR model ATADOM

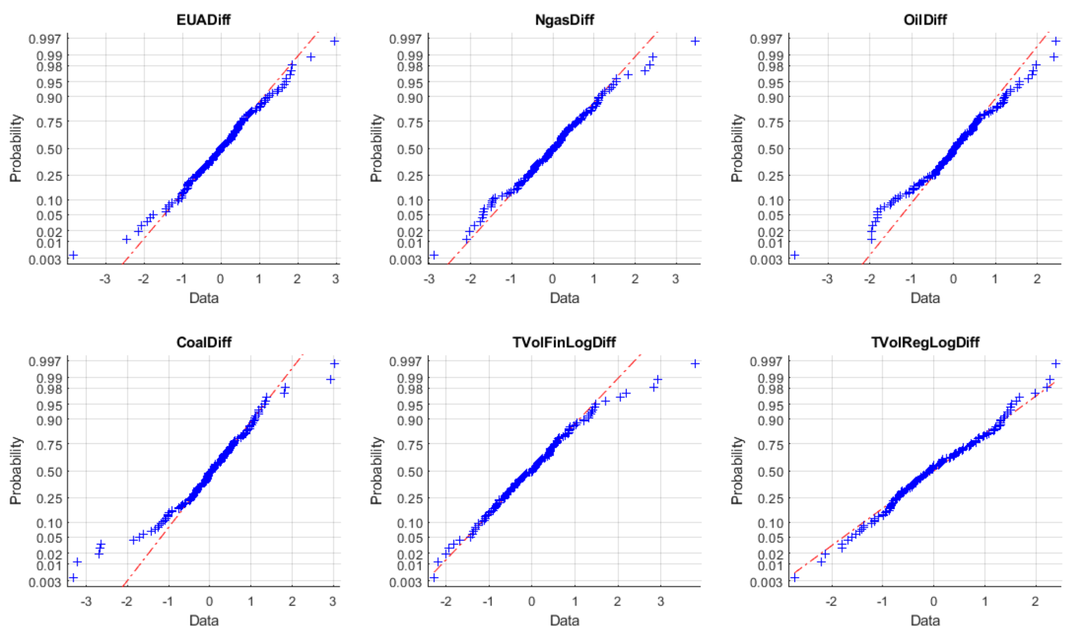


Figure 6: Normal plots of the residuals of the VAR model whole EUTL data

Table 7: Residual diagnosis p-values for the statistical tests

	whole EUTL			ATADOM		
	Kolmogorov-Smirnov test	**Ljung-Box test**	**ARCH test**	**Kolmogorov-Smirnov test**	**Ljung-Box test**	**ARCH test**
EUA	0.6891	0.2137	0.3584	0.0243	0.3286	0.2206
Ngas	0.945	0.4765	0.4691	0.8593	0.263	0.4457
Oil	0.6093	0.9589	0.9884	0.1358	0.25	0.8883
Coal	0.2909	0.9888	0.1244	0.3438	0.2531	0.0528
TVolFin	0.7726	0.0649	0.0585	0.6776	0.1223	0.9041
TVolReg	0.6446	0.5848	0.2221	0.8546	0.2838	0.4438

Although factors about energy futures prices, economic variations and production activity have been studied extensively either as price determinants or for forecasting the EUA price, the actual total volume of transactions, as many scientists suggest [***4***], [***5***], [***12***], [***16***], [***24***] was not tested as a factor for explaining or forecasting the EUA price. The analysis of the EU ETS market structure through the actual level of trading, is an interesting topic for further investigation, that may shed some light to its complex nature. To the best of our knowledge, our work is the first to compute trading volumes from the EUTL and use their amount as an explanatory factor for modeling the EUA price. Moreover, our work also indicates that it is more helpful to track transactions of ATADOM than the whole EUTL network and that not only suffices to forecast the EUA price, but in reality improves the forecast accuracy.

Table 8: Phase 3 Forecast errors for the best VAR models of those considered. If we restrict our data collection within ATADOM the forecast is improved.

Model Variables: EUA, Ngas, Oil, Coal, TvolFin, TvolReg				
	MAE	**reMAE**	**MSE**	**reMSE**
whole network	0.25185	0.25766	0.09713	0.10265
ATADOM restricted	0.24555	0.24924	0.09607	0.10063

MAE/MSE: Mean Absolute/Square error of 1-step-ahead forecasts

reMAE/reMSE: MAE/MSE when VAR model is repeatedly estimated

5. Conclusions

In this work, we analyzed the EU ETS from a network-based approach. The process of exchanging EUAs can be naturally modelled as a network. Thus, by using the EU Transaction Log (EUTL) database we reconstructed the entire transaction network in a monthly basis. We considered a classification of the entities/nodes, based on their role in the market, as (a) governmental, (b) regulated and (c) financial.

According to [***7***] and [***19***], the EU ETS network exhibits a core-periphery structure, i.e., that almost all the nodes are connected primarily with a subset of highly connected nodes (the core). Following this observation we searched for a subset of nodes that could summarize the important aspects of the entire network. To that end, we wanted to test the hypothesis that there exists a subset of nodes that could summarize the entire behaviour of the network. We show that such a subset exists for each month and therefrom we tested out if we could find just one, small and constant in time set having the same properties. Fortunately enough we prove that such a set also exists and we call it **ATADOM**. Basically, ATADOM consists of nodes that belong to many alomst dominating sets. Surprisingly, ATADOM is consists of just 235 entities, which stands up for the 1.76% of the participants. The majority of the entities within the ATADOM are financial, which indicates their intermediary role. We also show that the 85% of the transactions are done through or within the ATADOM.

In order to strengthen the claim that the ATADOM embodies the structure and the characteristics of the entire transaction network, we perform a Vector AutoRegression analysis (VAR), aiming to forecast the EUA price. We first quantified the allowance trading volume, both for the subnetwork restricted to ATADOM and for the entire network. We then used the above calculated variables, along with other EUA price determinants (such as energy prices, Industrial and Economic Indicators) from the literature and we estimated VAR models for Phase III of the EU ETS. Focusing solely on the the EUA price, we provided one-step-ahead forecast errors of the selected model. Our conclusions indicate that the forecast models that include variables based on the ATADOM set, performs better than the corresponding based on the entire network.

Overall, we show that the ATADOM, though tiny in size, is sufficient enough, not only in understanding the EU ETS structure, but also in providing better forecast results for the EUA price.

Our main methodological contribution is the calculation of the ATADOM. This small subset of participants may be used as a tool either to locate future drawbacks and fix instabilities, or to forecast the EUA price. The unique feature of the ATADOM is its simplified nature that may provide reliable information about the entire network quickly. Furthermore, using ATADOM in analyzing the large and complex network of the EU ETS, may be the basis for investigating other similar market networks and educing useful remarks and conclusions for their structure and operation. An interesting direction for further research is to test if the ATADOM remains a good aggregator in the future, by exploiting new data from the EUTL registry. Additionaly, another major research direction may be seen in testing other forecasting techniques, or non-parametric methods such as Artificial Neural Networks, and observe if ATADOM set continues to perform well in predicting the EUA price.

Funding

This research is carried out/funded in the context of the project “Applications of Reverse Greedy Mechanisms to Social Choice problems” (MIS 5004766) under the call for proposals “Supporting researchers with emphasis on new researchers” (EDULLL 34). The project is co-financed by Greece and the European Union (European Social Fund- ESF) by the Operational Programme Human Resources Development, Education and Lifelong Learning 2014-2020.

References

[1] European Parliament, “Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC.”, 2003, https://eur-lex.europa.eu/eli/dir/2003/87/oj.

[2] EU ETS Handbook, "European Union Emission Trading System Handbook.," 2015.

[3] United Nations, "Kyoto Protocol to the United Nations Framework Convention on Climate Change.," 1998.

[4] Denny Ellerman, Claudio Marcantonini, and Aleksandar Zaklan, "The European Union Emissions Trading System: Ten Years and Counting.," *Review of Environmental Economics and Policy*, vol. 10, no. 1, pp. 89–107, 2015.

[5] Beat Hintermann, Sonja Peterson, and Wilfried Rickels, "Price and Market Behavior in Phase II of the EU ETS: A Review of the Literature.," *Review of Environmental Economics and Policy*, vol. 10, no. 1, pp. 108–128, 2015.

[6] Ralf Martin, Mirabelle Muuls, and Ulrich Wagner, "The Impact of the European Union Emissions Trading Scheme on Regulated Firms: What Is the Evidence after Ten Years?," *Review of Environmental Economics and Policy*, vol. 10, no. 1, pp. 129–148, 2015.

[7] Andreas Karpf, Antoine Mandel , and Stefano Battiston, "Price and network dynamics in the European carbon market.," *Journal of Economic Behavior & Organization*, vol. 153, pp. 103 – 122, 2018.

[8] Simone Borghesi and Andrea Flori, EU ETS Facets in the Net: How Account Types Influence the Structure of the System, 2016, FEEM Working Paper No. 008.

[9] A. Christiansen, A. Arvanitakis, K. Tangen, and H. Hasselknippe, "Price determinants in the EU emissions trading scheme.," *Climate Policy*, vol. 5, no. 1, pp. 15–30, 2005.

[10] Frank Convery and Luke Redmond, "Market and Price Developments in the European Union Emissions Trading Scheme.," *Review of Environmental Economics and Policy*, vol. 1, no. 1, pp. 88–111, 2007.

[11] Joost Kanen, *Carbon Trading & Pricing.*, 1st ed.: Environmental Finance Publications, 2006.

[12] Emilie Alberola, Julien Chevallier, and Benoit Cheze, "Price drivers and structural breaks in European

carbon prices 2005-2007," *Energy Policy*, vol. 36, no. 2, pp. 787 - 797, 2008.

[13] Maria Mansanet-Bataller, Angel Pardo, and Enric Valor, "CO2 Prices, Energy and Weather.," *The Energy Journal*, vol. 28, no. 3, pp. 73–92, 2007.

[14] Gernot Klepper and Sonja Peterson, "Emissions Trading, CDM, JI, and More: The Climate Strategy of the EU.," *The Energy Journal*, vol. 27, no. 2, pp. 1–26, 2006.

[15] Anna Creti, Pierre-André Jouvet , and Valérie Mignon, "Carbon price drivers: Phase I versus Phase II equilibrium?," *Energy Economics*, vol. 34, no. 1, pp. 327 – 334, 2012.

[16] Beat Hintermann, "Allowance price drivers in the first phase of the EU ETS.," *Journal of Environmental Economics and Management*, vol. 59, no. 1, pp. 43 – 56, 2010.

[17] Wilfried Rickels, Dennis Görlich, and Sonja Peterson, "Explaining European emission allowance price dynamics: evidence from phase II.," *German Economic Review*, vol. 16, no. 2, pp. 181-202, 2015.

[18] Chune Young Chung, Minkyu Jeong, and Jason Young, "The Price Determinants of the EU Allowance in the EU Emissions Trading Scheme.," *Sustainability*, vol. 10, no. 11, 2018.

[19] Eva Benz and Stefan Trück, "Modeling the price dynamics of CO2 emission allowances.," *Energy Economics*, vol. 31, no. 1, pp. 4–15.

[20] Carolina García-Martos, Julio Rodríguez, and María Jesús Sánchez, "Modelling and forecasting fossil fuels, CO2 and electricity prices and their volatilities.," *Applied Energy*, vol. 101, pp. 363-375, 2013.

[21] Suk Joon Byun and Hangjun Cho, "Forecasting carbon futures volatility using GARCH models with energy volatilities.," *Energy Economics*, vol. 40, pp. 207–221, 2013.

[22] Regina A. Betz and Tobias S. Schmidt, "Transfer patterns in Phase I of the EU Emissions Trading System: a first reality check based on cluster analysis.," *Climate Policy*, vol. 16, no. 4, pp. 474–495, 2016.

[23] Michael R. Garey and David S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.* New York, NY, USA: W. H. Freeman & Co, 1979.

[24] Akira Maeda, "Impact of banking and forward contracts on tradable permit markets.," *Environmental Economics and Policy Studies*, vol. 6, no. 2, pp. 81–102, 2004.

[25] Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci, *Introduction to the Time Series Analysis and Forecasting.*: John Wiley & Sons, Inc, 2008.

[26] David A. Dickey and Wayne A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root.," *Journal of the American statistical association*, vol. 74, no. 366a, pp. 427–431, 1979.

[27] R.W. Hafer and Richard G. Sheehan, "The sensitivity of VAR forecasts to alternative lag structures.," *International Journal of Forecasting*, vol. 5, no. 3, pp. 399 – 408, 1989.

Appendix

A. Computation of ATADOM

Algorithm 2: Computing the All Time Almost DOMinating set

Input: $G[\ ]$ is the list of the monthly graphs, G.length is the size of the list equals to the number of graphs, $G[t].V$ and $G[t].E$ are the nodes and the edges of the t-th monthly graph
Input: $\mathcal{D}[\ ]$ is the list of the almost dominating sets where $\mathcal{D}[t]$ is associated with graph $G[t]$ and computed by Algorithm 1
Input: ϑ is the coverage threshold of the returned ATADOM
Output: ATADOM, the All Time Almost Dominating Set
1: **function**: computeATADOM($G[\ ]$, $\mathcal{D}[\ ]$, ϑ)
2: $\mathcal{D}_{\text{union}} \leftarrow \bigcup_{t=1}^{G.\text{length}} \mathcal{D}[t]$; // *Union of all almost dominating sets*
3: $\forall v \in \mathcal{D}_{\text{union}}$ compute its *appearance frequency*, i.e. count the sets of list $\mathcal{D}[\ ]$ in which v is a member.
4: *sort* all nodes of $\mathcal{D}_{\text{union}}$ in decreasing order by their appearance frequency
5: ATADOM $\leftarrow \emptyset$;
6: **for** t from 1 to G.length **do**
7: $v \leftarrow$ the most frequent element of $\mathcal{D}_{\text{union}} \setminus$ ATADOM
8: **while** ATADOM connected with less than ϑ fraction of $G[t].V$ **do**
9: ATADOM $\leftarrow$ ATADOM $\cup$ $\{v\}$;
10: $v \leftarrow$ the *next* most frequent element of $\mathcal{D}_{\text{union}} \setminus$ ATADOM
11: **end while**
12: **end for**
13: **return** ATADOM;
14: **end function**

B. Descriptive Statistics

The volume and wallet variables are presented in millions. The *Dom* in front of a variable indicates that it refers to an *ATADOM* variable. Δ indicates transformation using first differences. "log" indicates transformation using the natural logarithm.

Table 9: Raw variables descriptive statistics for Phase 3

Variable	n	Min	Mean	Median	Max	InterQuartile Range
EUA	173	3	6	6	8.6	2.2
Oil	173	28.9	82.2	99.7	118.9	55.6
Coal	173	43.5	68.6	73	90	19.5
IPI	173	95	98.3	97.6	102.7	3.3
ESI	173	87.4	101	103	107.7	4.7
FinWallet	173	534.8	935.6	838.8	1,788.10	162.6
RegWallet	173	384.1	1,199.70	1,136.20	2,846.80	873.1
InVolFin	173	10.9	101	61.5	1,812.10	54
OutVolFin	173	8.1	109.4	57.8	1,905.30	64
InVolReg	173	2.9	65.6	21.3	582.5	58.8
OutVolReg	173	2.1	79.5	15.5	1,001.80	34.2
TVolFin	173	19.3	210.4	121	3,717.30	114.4
TVolReg	173	6.5	145.1	42.5	1,005.80	129
DomFinWallet	173	69.9	336.4	270.8	819.4	127.8
DomRegWallet	173	54.9	263.6	254.2	624.9	156
DomInVolFin	173	5.9	66.3	38.5	1,239.30	38.3
DomOutVolFin	173	5.9	70.5	34.5	1,355.70	36.4
DomInVolReg	173	0.7	17.3	8.4	228.7	12.4
DomOutVolReg	173	0.3	20.5	7.4	261.7	13.1
DomTVolFin	173	12.4	113.4	60.1	2,086.10	56.9
DomTVolReg	173	1.8	37.5	15.9	478.2	26.2

Table 10: Transformed variables descriptive statistics for Phase 3

Variable	N	Min	Mean	Median	Max	InterQuartile Range
Δ(Coal)	173	-5.6	-0.2	-0.1	4.6	1
Δlog(DomFinWallet)	173	-1.1	0	0	0.3	0.1
Δlog(DomInVolFin)	173	-3.5	0	0	2.7	1
Δlog(DomInVolReg)	173	-3.9	0	0	2.9	1.2
Δlog(DomOutVolFin)	173	-2.3	0	0	4	1.1
Δlog(DomOutVolReg)	173	-3.6	0	0	4	1.5
Δlog(DomRegWallet)	173	-0.7	0	0	0.6	0
Δlog(DomTVolFin)	173	-2.5	0	0	2.6	0.9
Δlog(DomTVolReg)	173	-2.5	0	0	2.3	1.3
Δ(ESI)	173	-0.5	0.1	0.1	0.8	0.2
Δ(EUA)	173	-1.3	0	0.1	0.8	0.3
Δlog(FinWallet)	173	-0.6	0	0	0.1	0
Δlog(InVolFin)	173	-3.3	0	0.1	2.5	0.9
Δlog(InVolReg)	173	-3.2	0	0	2.6	1.2
Δ(IPI)	173	-0.4	0	0	0.7	0.2
Δ(NGas)	173	-6.1	-0.2	-0.2	9.1	1.9
Δ(Oil)	173	-10.2	-0.4	-0.5	5.6	3.1
Δlog(OutVolFin)	173	-2.1	0	0	2.7	0.9
Δlog(OutVolReg)	173	-3.8	0	0.1	3.2	1
Δlog(RegWallet)	173	-1.2	0	0	0.4	0
Δlog(TVolFin)	173	-2.4	0	0	2.5	0.8
Δlog(TVolReg)	173	-3.1	0	0	2	1