Online Appendix for "Who gains from market fragmentation? Evidence from the early stage of the EU carbon market"

Estelle Cantillon and Aurélie Slechten

April 2023

This online Appendix provides details on:

- The procedure used to aggregate accounts at the level of ownership, and
- The procedure used to associate a price to the transactions recorded in the Community Independent Transaction Log (CITL).

1 Aggregation of accounts at the ownership level

1.1 Accounts dataset

Stage 1 – Data Import. The accounts dataset is constructed on the basis of the account data downloaded from the CITL, which records all accounts open in any national registry, at any point of time. We selected accounts open in the 24 registries that were active during phase I. The CITL account dataset classifies accounts in 3 categories: operator accounts for the firms subject to the EU ETS regulation (account code 120), country accounts (used for initial allocation, retrieval and other administrative operations), and third party accounts (account code 121), which are accounts open by any third party. These include exchanges, other financial intermediaries or any other individual interested in selling or buying allowances. We only download the information related to the operator or third party accounts. The original dataset contains 17,560 unique accounts, including 12,448 operator accounts and 5,112 third party accounts. Each line corresponds to one account and has the following fields:

- name: the account holder name (not unique),
- identifierInReg: this is a second identifier for the account (for example the name of the installation or the location),
- address fields,
- AccountTypeCode: 120 (for operators subject to the EU ETS) or 121 (for third parties),
- RegistryCode: 2 letter code for the country in which the account is open,

- AccountIdentifier: a number associated to each account and which is unique within a country,
- installationIdentifier: an installation ID for operator accounts.

We associate each operator account with an activity type, the name of a parent company (when available) and for each year of phase I, its initial allocation, its surrender, and its verified emissions. This is done using the compliance dataset downloaded from the CITL.

Stage 2 – Basic Cleaning. We create a unique account identifier (accountID) on the basis of the registry code and account identifier (two letters from registry code + account identifier).

The next step is to identify accounts that are effectively active during phase I. For this purpose, we use the raw transaction dataset downloaded from the CITL (described below) and search for the date of the first transaction recorded in the CITL for each account. 5,966 accounts are not active during the first phase of the EU ETS (first transaction date after 30 April 2008) and are dropped from our dataset. This leaves 10,808 operator accounts and 786 third party accounts (11,594 active accounts).

The original compliance data associate each operator account with one of 24 activity types. We aggregate these activity types into 8 sectors (Power & Heat, Refineries, Metals, Iron, Steel and Coke, glass, ceramics, pulp & paper, other), labeled 1 to 8 following Ellerman and Joskow (2008). We create an additional category (sector 9) for the third party accounts. Table 1 describes the sector distribution of active operator accounts in phase I along with the original activity types reported in the compliance dataset.

Stage 3 – Accounts Aggregation. Firms can hold multiple accounts. This is the case for operators that have several installations subject to the EU ETS. This is also the case for firms that want to hold accounts in different national registries to be able to trade on specific exchanges or to be able to trade when their national registry is still not operational. Some firms also have different accounts for different functions (trading, brokering, . . .). We create a unique firm ID for accounts corresponding to the same firms.

We first merge the accounts of exchanges and clearing members of ECX.[1]

- We merge the two accounts belonging to Powernext. These two accounts do not have the same name or the same address because the French exchange was initially Powernext (Caisse des depots et Consignation) and became Bluenext in December 2007.
- We merge the four accounts of Nord Pool (Nasdaq OMX).
- We merge the six accounts belonging to the Amsterdam Power Exchange (APX).
- We merge the three accounts of market maker SENDECO2.
- We merge the five accounts of POLPX.

[1] These are companies for which we can clearly identify the associated accounts. For the sake of simplicity, we want to give these companies a fixed Firm ID that will not vary across the algorithms used to merge the accounts.

Table 1: Number of active operator accounts in each industrial sector (and associated activity types)

Sector	Number of accounts	Corresponding activity types
Power & Heat (1)	7,150	Combustion installations with a rated thermal input exceeding 20 MW
		Combustion of fuels
Refineries (2)	151	Refining of mineral oil
		Mineral oil refineries
Metal, Iron, Steel & Cokes (3)	268	Metal ore (including sulphide ore) roasting or sintering installations
		Installations for the production of pig iron or steel
		Production of coke
		Production of pig iron or steel
		Production or processing of ferrous metals
		Production or processing of ferrous metals Production of primary aluminum
		Coke ovens
Cement & Lime (4)	541	Installations for the production of cement clinker in rotary kilns
		Production of cement clinker
		Production of lime, or calcination of dolomite/magnesite
Glass (5)	411	Manufacture of glass
		Installations for the manufacture of glass including glass fibre
Ceramics (6)	1,131	Manufacture of ceramics
		Installations for the manufacture of ceramic products by firing
Pulp & Paper (7)	807	Industrial plants for the production of pulp from timber or other
		Production of pulp
		Production of paper or cardboard
Other (8)	349	Production of ammonia
		Production of bulk chemicals
		Other activity opted-in pursuant to Article 24 of Directive 2003/87/EC

- We also merge the accounts of the clearing members of ECX (Firm IDs 12-34). ECX required traders to go through clearing members to trade on the exchange. We identify 25 accounts that trade with ECX. For these accounts we look at all their trading partner accounts and identify those belonging to the same company (similar name or address). We identify 114 accounts belonging to clearing members of ECX and merge them in 23 different Firm IDs.

With this first step, we create 34 unique Firm IDs corresponding to 140 accounts. For the rest of the accounts (11,454), we create a unique Firm ID using the information about parent companies and account names and addresses. We create a new variable *name_ address* that combines the name and main address of each account in lower case letters and without punctuation signs or accents that could reduce the effectiveness of the matching. We create the same type of variable for parent companies: *parentcompany_ temp*. As a matching procedure, we use the Stata command strgroup, which calculates the Levenshtein edit distance between all pairwise combinations of strings in *parentcompany_ temp* and *name_ address*. Strgroup then matches a string pair if their normalized edit distance is less than or equal to a certain threshold (we use 0.25). We then aggregate the accounts as follows:

- Accounts that have similar parent company are given the same firm ID: this generates 1,228 firm IDs.[2]
- Accounts that have similar account holder names and addresses are associated with the same firm ID: this brings our dataset to 6,651 different traders based on this criterion.

Combining both sources of information, we identify 6,619 unique firm IDs. We check manually the accounts that have similar names but are assigned different firm IDs (because they have different addresses), and based on similarities in addresses and web searches, we decide to give them the same firm ID (6,582 unique firm IDs). Finally, we identify some accounts that are assigned different Firm IDs (different name or parent company), but that have the same address and should have been merged (based on web searches and trading patterns). This process gives us 6,496 unique firm IDs.

Stage 4 – Trading Desks. The number of accounts obtained so far may still be overestimating the number of independent traders as fully-owned subsidiaries of the same firm, for example, will not be considered as belonging to the same trader ID under these criteria, unless they are located at the same address.

As a result, we complement this first approach by using transaction patterns to identify accounts belonging to the same firm. Specifically, we search for accounts that could serve as trading desks for groups of related firms, the idea being that the transactions of interest in this case are the transactions between the trading desk and third parties, whereas transactions between the trading desk and the accounts of the related firms are just internal transfers.

To do this, we consider that each of the 6,496 unique firm IDs in our account dataset is a node. Two nodes are connected if there was a transaction between them any time during phase I.

[2] The parent company field has a large number of missing variables.

Step 1: identification of potential trading desks (our "candidates")

We select those unique firm IDs that have a strictly positive betweenness meaning that they lie on at least one shortest path between two nodes, excluding exchanges and ECX clearing members (1,537 unique firm IDs). These are the candidates "trading desk".

Step 2: identification of potential partners for the trading desks

We identify 2,635 nodes j that have only one trading partner over phase I. These are the potential partners of our candidates.

Step 3: For each potential trading desk of step 1, identification of related accounts

Let i index the candidates "trading desk" identified in step 1. For each i, select those nodes j (from step 2) that are at distance 1 of i, meaning that there was at least one direct transaction between i and j.

Step 3 results in 602 candidates "trading desk". There are on average 19.66 related firms per candidate (min = 1, max = 119, median = 9). Each of these related firms are analyzed manually. Accounts are assigned the same Firm ID as their candidate "trading desk" if it is clear, on the basis on their names or other account or transaction characteristics, that these accounts are part of the same company as the trading desk account. This process further reduces the number of unique firm IDs to 5,983.

Note that the condition imposed in step 2 (only one trading partner) is conservative. If two subsidiaries of the same group trade directly with one another on top of using the trading desk, they will not be included in the list of potentially related firms for the trading desks. Likewise, if one of these subsidiaries also traded with a third party, it will not be included. As a result, it may lead us to continue to overestimate the number of independent traders and the number of transactions.

Stage 5 – Traders connected to an exchange. We first manually match the accounts associated with the partners of Climex Alliance (i.e. Vertis, Wallich & Matthes and STX), based on similarities in name, parent company and identifierInReg.[3] We then focus on the other accounts that had at least one transaction with an exchange during phase I (as exchange-based transactions are the focus of our empirical analysis). Let i index a trader connected to an exchange (i.e. who traded at least once on an exchange during phase I). For each i, we identify all its trading partners, and look for similarities between the partners' names and trader i's identifierInReg or parent company: This stage helps us identify accounts that are opened by firms to trade exclusively on an exchange. Powernext in particular required traders to hold an account in the French registry and several firms opened an account in the French registry for the exclusive purpose of trading on Powernext. Very often, the name of the account is the name of the person in charge of trading on the exchange, but the identifierInReg may contain information regarding the associated company (e.g. Account named "SPRENGERS", with identifierInReg "STATKRAFT"; or account named "VISSER-DE JONG", with

[3]The Climex Alliance is a pan-European spot carbon emissions exchange. It includes the Dutch exchange New Values (trades cleared through APX), the Spanish exchange SENDECO2, the UK power exchange UKPX (a subsidiary of APX), the carbon emissions broker STX Services (Dutch and Belgian markets), and the Hungarian investment group Vertis Environmental Finance.

identifierInReg "Vattenfall Energy Trading Netherlands N.V."). Unlike stage 4, stage 5 does not require that the trading partner of i only trades with i.

We identify 539 firm IDs that trade on exchanges. For each of them, we identify the potential related accounts among their trading partners. On average, we find 87.83 related accounts per trader (min = 0, max = 297 and median = 68). Each of these related firms are analyzed manually and accounts are merged if it is clear, on the basis on their names or other account or transaction characteristics, that these accounts do indeed correspond to a group. We find 49 candidates that meet those criteria. This process further reduces the number of unique firm IDs to 5,902.

Stage 6 – Similarities in names among trading partners. We identify the trading partners of all the accounts who traded at least once with a different company during phase I (3,942 unique Firm IDs). For each trader, we identify the trading partners who share similarities in names, identifierInReg or parent company. We use the same command as in stage 3 (strgroup), but we impose a lower threshold (0.5 instead of 0.25). The objective is to make sure that our criteria in stage 3 were not too restrictive, and ensure that we did not miss intra-firm transactions. 326 traders share similarities with at least one of their trading partners (min=1, max=60, median=10). We find 69 traders that should be merged with some of their trading partners.[4] Finally, we consider Firm IDs containing only one third party account (121). To ensure that these accounts are not related to another Firm ID, we identify their trading partners and check that this third party account is not a trading desk that we missed in stage 4. We identify 34 candidates that should be merged with their trading partners. At the end of stage 6, the total number of unique firm IDs is 5,729.

We redefine the sector of exchanges to 10 instead of 9. There are 26 accounts associated with an exchange (and 11 firm IDs). We also import NACE codes for all operator accounts (when available). The NACE codes come from the data downloaded from the EUETS.INFO database, which rebuilds the CITL data and associates a NACE code to the operator accounts based on the activity recorded in the CITL.

Final stage. In our final accounts dataset, each unique accountID is associated with the name of the account holder (Name), an identifier (identifierInReg, which is a second name), a firm ID (a number), a registry (two letter code), an address, a parent company (equal to the name if missing), an account type (120 or 121), an opening date (first appearance in the CITL data), and, for operator accounts an industrial sector (number code + sector name + NACE + NACE description), initial allocations of allowances, surrender and verified emissions levels.

1.2 Firms dataset

As our main analysis is at the firm level, we build a second dataset that contains only firm-level information. There are 5,499 unique firm IDs with an account opening date before 31 May 2007.

[4] During this process, we also identify two accounts that were merged in stage 3 (same parent company DELIPAPIER). However, DELIPAPIER and LPC PRODUITS PAPIERS merged in 2012. We therefore create a separate Firm ID for LPC PRODUITS PAPIERS (note that these accounts did not trade with each other in phase I, so the number of transactions in the transaction dataset is not affected).

Accounts characteristics (from the accounts dataset) are aggregated at the firm ID level as follows:

- Number of accounts associated with the firm ID,
- Opening date is taken to be the opening date of the first account opened in the CITL,
- The date of the first transaction is taken to be the date of the firm's first transaction observed in our final transaction dataset (see below),
- Yearly initial allocations, surrender and certified emissions are aggregated across all accounts belong to the firm,
- The number of transactions they have had since their opening date, normalized by the number of days between the opening date and 31 May 2007 (expressed in number of trades per year during which active),
- Average transaction size,
- 10 sector dummies (sectori dummy = 1 if the firm has an operator account associated with sectori) + a main sector dummy which is equal to the sector which surrendered the largest number of allowances over the whole phase. If the total number of allowances for a firm ID is zero, we use the number of installations per sector to determine the main sector,
- A dummy, *Electricity Producer,* to identify the firms related to electricity and gas supply (NACE 35). A firm is classified as *Electricity Producer* if the emissions from installations with NACE 35 represent more than 80% of this firm's total emissions,
- A firm type, with three categories: "operator", "exchange" and "other". The *operator* category represents the compliance traders (i.e. if the firm has at least one operator account). The *exchange* category corresponds to sector 10 (defined above), while the category *other* includes all the other non-compliance traders (sector 9), i.e. financial intermediaries - clearing members, market makers or brokers - or other participants such as speculators. There are 5,254 firm IDs with firmtype "operator", 234 with firmtype "other", and 11 exchanges.[5]

In our regressions, we use three dummy variables that are constructed using the information from the firms dataset:

1. *Small Compliance Trader:* Dummy variable equal to 1 if the variable firmtype is "operator" and the average annual allocation of this firm is lower than 1 million tCO2.
2. *Energy Sector*: Dummy variable equal to 1 if the variable firmtype is "operator" and if the firm's NACE is 35 (dummy variable *Electricity Producer* = 1) or if the firm belongs to sector 2 (oil refineries). This is similar to the approach used by Balietti (2016).
3. *Non-compliance trader*: Dummy variable equal to 1 if the variable firmtype is "other".

[5] Barclays Bank plc and Fortis have an operator account for their heat generation installation (no allowance received, very small amount surrender). We do not apply the rule that a firm type is by default operator if it has at least one operator account for Barclays and Fortis.

2 Procedure to associate a price to the transactions in the CITL

2.1 Initial cleaning of the transactions dataset

The CITL records all transactions that take place between traders. These data are public with a 3-year lag. We downloaded all phase I transactions that took place from 1 January 2005 (the first transaction took place on Jan 11, 2005) until 30 April 2008. There are 116,576 such transactions. The original dataset contains, for each transaction, a unique transaction ID, the name, account type code, registry code and account identifier of the buyer and the seller, a time stamp for the transaction, the size of the transaction, and a transaction type code (initial allocation, surrender, transaction). As our focus is on trades between third parties and operators, we drop initial allocations and surrenders (58% of the total number of transactions). We also drop intra-firm transactions (14,660 transactions, 12.57% of all transactions), using the unique firm IDs defined in our accounts dataset.

This leaves us with 34,742 transactions.

The dataset contains a unique transactionID, the firm and account IDs of the buyer and the seller, a time stamp (+ year, month and day) and the number of allowances transferred. It also includes the name of the exchange if the buyer or the seller is one of the exchanges identified in stage 3 (Firm IDs 1 to 11).

We provisionally associate each transaction with a trading venue variable using the account dataset and our knowledge of the accounts associated with an exchange. The association is provisional for two reasons. First, for exchanges for which transactions go through a clearing member (e.g. ECX), nothing guarantees that this clearing member exclusively transacts with the exchange. Second, several exchanges (Powernext, ECX and EEX) also offered clearing services for OTC spot transactions. 18,907 transactions are provisionally associated with an exchange. We also create a dummy variable *RealTransaction* that is provisionally set to 1 and is meant to identify those transactions in the CITL that correspond to real transactions (and not deposits and late withdrawals, or transactions that could not be reconciled with tick data of the corresponding exchange).

At this stage, our transactions dataset does not have any information on transaction prices. In this section, we detail the procedure to identify real transactions on exchanges and associate a price to every real transaction that takes place on an exchange and OTC. This process is specific to each trading venue.

2.2 Transactions on Powernext

10,322 transactions are provisionally associated with Powernext in the CITL. We also have tick data from Powernext containing 5,073 transactions. One buyer – seller transaction in the tick data corresponds to two transactions in the CITL: one from the buyer to Powernext account and one from Powernext to the seller. Deliveries on Powernext usually appear within 15' in the CITL. For each side of transactions and each day, we perform the following tests to identify potentially problematic cases.

- **Test 1**: For each day, we check that the number of buying transactions is equal to the number of selling transactions in the CITL

- **Test 2**: For each day, we check that the number of transactions in the CITL is equal to (twice) the number of transactions in the tick data.

These two tests identify 87 problematic days. We checked the transactions for these days manually, both in the CITL and in the tick data and identified 5 possible reasons for the mismatch:

1. Clearing delays: Powernext received allowances from the seller on one day but only delivered allowances the day after. We change the day of those delayed transactions in the CITL transactions dataset. 149 lines are affected.

2. Transfer mistakes: We also identified potential transfer mistakes whereby a trader wrongly transfers allowances to Powernext.[6] For these transactions, the *RealTransaction* is turned to zero. 88 lines are affected.

3. Split transfers: For a given transaction in the tick data, the seller transfers the sold allowances in two waves (close to one another in time).[7] This concerns 4 transactions in the CITL. For these cases, we replace the number of allowances for the first of the two split transactions with the total number and we drop the second transaction. We thus drop two transactions.

4. Transactions in the tick data that do not appear in the CITL: 5 transactions are in this case. They are dropped because our primary source of information is the CITL (these transactions take place after May 2007 and are not part of our final dataset any ways).

5. Transactions in the CITL that do not appear in the tick data: Some trades do not appear in the thick data of Powernext. The venue variable for these transactions is replaced by OTCPowernext (96 lines).

Once this is done, we pursue to match Powernext transactions with a price. To do this, we first drop momentarily the 96 transactions that do not appear in the tick data and the 88 potential transfer mistakes. We then match each transaction with a tick price using the sequence of the transaction within the day and volume. Specifically, we first consider buyers and for each day order the buy transactions on Powernext according to the CITL timestamp. We do the same for the tick data. Then we match them based on their sequence number and transaction volume.[8] No other mismatch is identified in the process. The final number of transactions on Powernext (with a price) is equal

[6] For example on June 28, 2005, a seller transfers 5,000 allowances to Powernext at 11:03 am and Powernext returns those allowances at 11:17 am, without any corresponding transactions taking place in the tick data.

[7] For example on Jan 4, 2006, for a transaction of 10,000 allowances that took place in the tick data, Powernext first transferred 1,000 allowances at 11:35 am to a trader and then 9,000 allowances to the same trader at 11:36 am.

[8] We match on sequence within the day rather than on the timestamp because there is 15-20' time lag between the time stamp in Powernext tick data and the CITL data (the clearing lag is not exactly the same all the time). The match is done both on sequence and volume to identify potential volume discrepancy. A single volume mismatch is identified in the process on Feb 6, 2007 where the CITL volume is 15,000 whereas the tick data volume is 10,000 (single transaction that day). We kept the CITL volume since the CITL is our source data.

to 10,136. There are 96 OTC transactions cleared by Powernext that are not associated with a transaction-specific price. We also have 88 non-real transactions.

2.3 Transactions on Nord Pool

There are 919 transactions that go through Nord Pool in the CITL. We also have tick data for both spot and future transactions on Nord Pool.[9]

For each trading day (interpreted within the clearing window, from t-2 up to t+3), we manually match transactions on the CITL with spot transactions in the tick data, based on volume and sequence within the day (time stamp), making sure that a purchase appears in the CITL after it appears in the tick data (usually at t+2 or t+3) and that a sale appears in the CITL before it appears in the tick data (usually at t-1 or t).[10] Transactions in the CITL often correspond to multiple transactions in the tick data because a buyer (or seller) transacted with multiple sellers (buyers) on the exchange, often implying different prices for these transactions. These CITL transactions are split for each corresponding tick data transaction so as to associate the right price for each allowance.

723 out of the 838 spot transactions in the tick data are matched in this way and a price is associated with these transactions.[11] We also use the tick data to confirm the market venue or switch the market venue variable to OTCNordPool if needed (note that for Nord Pool even OTC transactions are associated with a price).

Settled futures transactions also appear in the CITL transaction dataset at the time of settlement. When the trader is short, allowances for settled transactions appear up to three weeks before settlement. When the trader is long in the futures market, allowances appear on his account within a couple of days of settlement. We identified 110 CITL transactions for which we are sure that they correspond to settled futures. The price we associate with these matched transactions is the closing price of the future on the last day of trading but it is not used in the analysis.

The final transaction dataset for NordPool-related transactions (NordPool or OTCNordPool) contains a total of 1,086 transactions (instead of the original 919) including 723 spot transactions associated with a price (335 sales and 388 purchases), 110 futures settlements associated with a price and 253 unmatched transactions (for which we keep the market venue as "NordPool" but set the *RealTransaction* dummy to 0).

Note that we had to drop two transactions from the CITL dataset (out of the original 919):

- TransactionID NL4157: this transaction associated with the transaction NL4156 corresponds

[9] The tick data file (for the first phase of the EU ETS) contains both spot and futures (15,120 lines). The variable "ContractTicker" indicates if it is a spot (e.g. EUAD0102-06) or a future (EUADEC-05, EUADEC-06, EUADEC-07, EUAMAR-06, EUAMAR-07). This file contains 838 lines for spot transactions (2 lines correspond to one transaction). These include both exchange-mediated and OTC / exchange-cleared transactions. The rest (14,282 lines) concerns futures.

[10] There are a number of exceptions to this rule where we accepted sale transactions in the CITL that were within an hour of a few hours of the sale transaction in the tick data when it was clear that these were matching transactions.

[11] Matching is not always possible because the net position of Nord Pool does not return to zero at the end of each day or around settlement dates there are too many transactions of the same size.

to a unique purchase on Nord Pool (transfer in two legs within the same day: 206,134 + 243,866 = 450,000 units). We set the amount of the first leg to 450,000 (NL4156) and dropped the second leg (NL4157).

- TransactionID DK1210: Within the same afternoon, the trader transferred 62,000 units (SE637) to Nord Pool for futures settlement and then got back 12,000 units from Nord Pool (DK1210). We assume that this trader changed their mind or made a mistake in the transfer. We decided to set the amount of transaction SE637 to 40,000 and to drop transaction DK1210.

2.4 Transactions on EXAA

Auctions take place weekly and transfers of allowances appear on the same day or at t-1, t-2, ... t-5 days in the CITL. EXAA requests that traders deposit the allowances that they plan to sell ahead of the auction to be able to place a sell bid. As a result, t minus something transactions correspond deposits of allowances from a trader to the EXAA account (checked systematically). The equilibrium price on EXAA is determined as the equilibrium price of a double auction. At the end of the auction day t (or at t+1), EXAA transfer the allowances to the buyers or back to the sellers who did not manage to sell their allowances on that day.

There are 789,931 allowances traded on the EXAA platform during the relevant time period and 3,114,670 (sum of units in the CITL divided by 2) allowances traded in the CITL dataset (with EXAA as the counterpart). To reconcile transactions on EXAA with the transactions on the CITL, we generate two new variables: a week variable (from 1 to 52 that records the week of the transaction within the year) and a variable "auction" that tells us to which "Tuesday auction" the transaction belongs (based on weeks during which there have been transactions in the CITL involving EXAA). There are 122 different auctions. We compute the net position (deposit − withdrawal) for the EXAA account for each auction. For the vast majority of auctions, this position is zero, i.e. after each auction, participants withdraw their remaining allowances.There are few exceptions at the end of the first phase (15 days − 27 lines in the transactions dataset, from 27 December 2007 to 29 April 2008). This means that we can treat CITL transactions during a week as referring to the particular auction of that week.

Handling of transactions for auctions for which the net position of EXAA = 0. We create a temporary dataset for those 107 auctions for which EXAA has a net zero position. In this dataset, for each EXAA trader, we create a new variable that records their net position (deposits − withdrawals) for each auction. If this variable is positive (resp. negative), this indicates that the participant is a seller (resp. a buyer) during this auction. The volume of each trade is given by the absolute value of the position of each participant. For all these trades, the central counterparty is EXAA. If the net position is zero, we consider that the participant does not trade on that day (we thus drop 41 auctions for which the total deposit = total withdrawal for each participant). At the end of this process, our temporary dataset contains data (buyerID, sellerID, auction, net position) for 66 auctions and 170 lines that correspond to real transactions.

We then merge our temporary dataset with our transactions dataset to replace the transactions that appear in the CITL with the actual transactions to which these flows in the CITL correspond. In particular this means that, for each auction, there is only one transaction per trader (multiple deposits / withdrawals are eliminated and the transaction volume is replaced by the net position).

Handling of transactions for auctions for which the net position of EXAA is different from 0. For each trader, we compute their net position for those 15 auctions. A net positive position means that that trader bought or that he took back some of the allowances he had left with EXAA in one of the past auctions, and mutatis mutandis for the case of a net negative position. We cannot therefore guarantee that these transactions correspond to real transactions. We simply aggregate, for each auction and each trader, their CITL transactions into a net transaction. This replaces the original 27 lines in the CITL transactions by 17 for which the *RealTransaction* dummy is set to 0.

Finally, we associate a price to each real transaction using the auction equilibrium price from EXAA. The day of the transaction is replaced by the day of the auction (the time stamp is left unchanged). In our final transaction dataset, we have 187 lines for EXAA instead of the original 591: 170 real transactions with a price and 17 non-real transactions without price.

2.5 Transactions on SENDECO2

There are 605 transactions that go through SENDECO2 in the CITL. SENDECO2 requires traders to deposit allowances on the exchange account prior to submitting a sale order. Most traders on SENDECO2 were small traders who would deposit allowances shortly before their trade or request their purchased allowances shortly after their trade. We analyze manually the transactions on SENDECO2 (605 lines) and identify different issues:

1. Transfer mistakes, or no buyer found, or only part of the deposit has been sold (transfers between SENDECO2 and the same trader within a short time interval). We turn the variable *RealTransaction* to 0 (64 lines).

2. "Delayed sale": it might be the case that a seller deposit allowances but it takes a few days to find a buyer. As a convention, we use the date of the purchase (transfer from SENDECO2 to the buyer) as the transaction date. This concerns 82 lines and we change 40 dates.

3. Split transfers: One deposit by a seller is associated with two or more purchase transactions. As before, we use the date of the purchase (transfer from SENDECO2 to the buyer) as the transaction date. When the purchase dates vary, we split the original deposit and associate the purchase dates with each part of the deposit. This concerns 9 original deposits and we create 10 additional transactions.

4. 37 transactions on SENDECO2 are transactions between SENDECO2 and APX (another exchange). Both APX and SENDECO2 were part of Climex Alliance, which implies that the two exchanges were probably sharing their order books. Transfers between these two

exchanges are therefore considered as internal transfers and the dummy *RealTransaction* is turned to 0.

5. There are 134 transfers from a trader to SENDECO2 (sell transaction) that we could not match to a purchase transaction. The dummy *RealTransaction* is turned to 0 for those transactions.

6. For the purchase transactions that could not be matched with a sell transaction:

   (a) Most of these purchase transactions are from one-time buyers, or traders who are buyers only. We decide to keep them as real transactions.

   (b) A few transfers SENDECO2-traders occurring at the end of April 2008 corresponds to traders emptying their account on SENDECO2. We identify those by looking for traders who made an initial deposit in Phase I and for which the only "purchase transaction" was in April 2008. For those transactions (8 lines), we turned the dummy *RealTransaction* to 0.

This leaves us with 621 transactions on SENDECO2, including 408 real transactions. We have daily SENDECO2 price data between 1 February 2006 and 30 April 2008.

2.6 Transactions on ECX

ECX only offered trading in futures. This means that ECX transactions that appear in our dataset only concern settlements for these contracts, not spot transactions. We use them exclusively to build the network of trading relations and compute our centrality measures. There are 4,966 transactions related to ECX in the original CITL transaction dataset. 210 are transactions between ECX and a clearing member and 4,756 are transactions between clearing members and other traders. Not all transactions between ECX clearing members and other traders correspond to a transaction on ECX (ECX clearing members also provide clearing and brokering for other market participants). The main cleaning task therefore is to reconstruct from the transactions in which clearing members are involved, those that went through ECX.

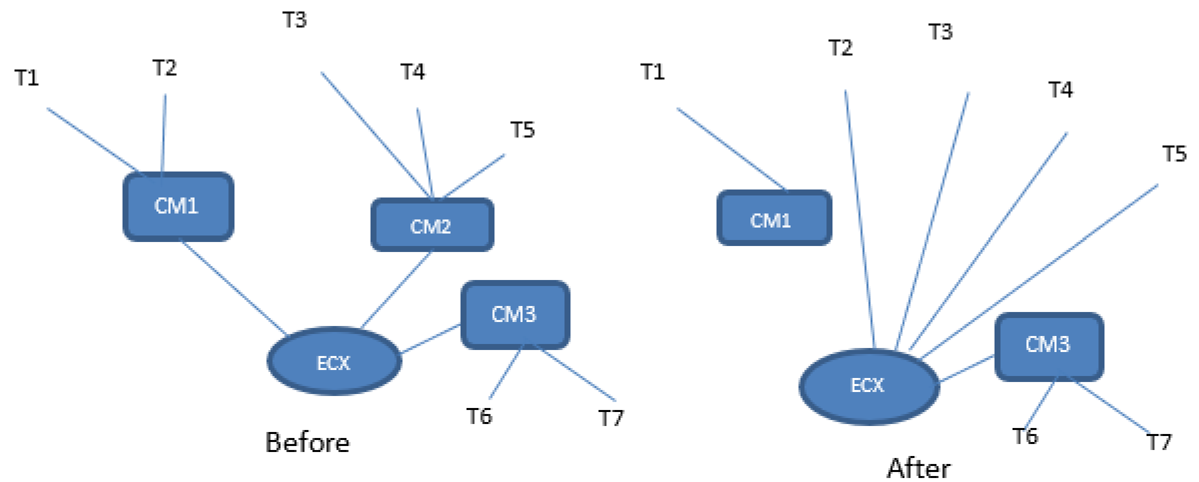
<u>Transactions between ECX and clearing members:</u>

We do a number of sanity checks on the data. First, we check that, at each maturity date, the acquired volume is equal to the sold volume for ECX. Second, we check, for each maturity, that allowances sales are indeed recorded at times t-1 or t and purchases are recorded t+1 and later. There are 11 transactions for which a sale and purchase between a clearing member and ECX happen within a short time interval (these transactions happen on Mar 26, 2007, Dec 14, 2007 and Dec 19, 2006). We interpret these transactions as mistakes or changes of mind and classify them as non-real transactions.

<u>Transactions between clearing members and other traders:</u>

We next try to identify which of the 4,756 transactions between clearing members and other traders can be associated with transactions made on ECX (there are originally 210 transactions

Figure 1: Illustration of the changes in the structure of transactions on ECX



between clearing members and ECX). To do this, we work clearing member by clearing member. We replace the venue variable by OTC for those transactions that take place outside of a maturity month since we can be sure that they do not correspond to ECX settlement transactions. For the other transactions and for each maturity, we manually match the trader-clearing member transactions with the clearing member – ECX transactions, based on transaction size and time (making sure a transfer from a clearing member to ECX is preceded by a transfer from the trader to the clearing). For those matched transactions, we replace the two transactions (the one between the trader and the clearing member and the one between the clearing member and ECX) by a single transaction between the trader and ECX.

Figure 1 shows the impact of this operation on the structure of transactions in our dataset. For matched transactions, the clearing member is removed from the transaction graphs. OTC transactions cleared or brokered by clearing members remain (in the figure this corresponds to the transaction between T1 and the clearing member 1). Some transactions between clearing members and ECX can remain (these correspond to transactions that could not be matched and is illustrated by the transactions between ECX, CM3 and traders T6 and T7).

At the end of this procedure, the trading venue of 4,501 transactions between clearing members and traders is turned to OTC.

This leaves us with 344 transactions (333 real transactions and 11 non-real transactions) where ECX is the trading venue.

2.7 Transactions on EEX, APX, CMCEK, GME, CEB and POLPX

All these exchanges require traders to deposit allowances on the exchange account prior to submitting a sale order and transfers of unsold or purchased allowances were only made at the request of traders. We do not have tick data for these exchanges. As a result, transactions in the CITL cannot

Table 2: Real Transactions by trading venue

Venue	Real	Non-Real	Real (% total)
Powernext	10,136	88	99
NordPool	738	253	74
EXAA	170	17	91
SENDECO2	408	213	66
ECX	333	11	97
EEX	0	550	0
APX	0	696	0
CEB	188	5	97
GME	0	59	0
POLPX	7	8	46
CMCEK	0	9	0
OTC	20,321	15	100
OTCPowernext	96	0	100
OTCNordPool	95	0	100
Total	***32,492***	***1,924***	***94***

Table 3: Real transactions without price

Venue	Real transactions without a price	%
SENDECO2	8	0.0
CEB	188	0.09
POLPX	7	0.0
OTC	20,321	98.6
OTCPowernext	96	0.5
Total	***20,620***	***100***

be reconciled with transactions on these exchanges. There are 550 transactions associated with EEX, 59 transactions associated with GME, 696 transactions associated with APX, and 9 transactions associated with CMCEK in the CITL transaction dataset. We turn the *RealTransaction* dummy to 0 for these exchanges.

Most traders on CEB and POLPX were small traders who would deposit allowances shortly before their trade or request their purchased allowances shortly after their trade. It is therefore possible to use the deposits and withdrawals from CEB and POLPX accounts to reconstruct the transactions on those exchanges.

At the end of this process, we have 34,416 transactions, including 32,492 that are classified as real transactions and 1,924 that are classified as non-real. The distribution of transactions across trading venues is shown in Table 2.

2.8 Point Carbon Index

Some real transactions are not associated with a price, as shown in Table 3. For all these transactions, we create a new variable *proxyprice* and set the proxy price equal to the daily closing Point

Table 4: :Number of real transactions associated with a price or proxy price by trading venue

	Number of real transactions			
Venue	**With price**	**With proxy price**	**Without price**	**Total**
Powernext	9,452	0	0	9,452
NordPool	719	0	0	719
EXAA	168	0	0	168
SENDECO2	0	274	0	274
ECX	221	0	0	221
CEB	0	173	0	173
POLPX	0	3	0	3
OTC	0	15,208	751	15,960
OTCPowernext	0	90	0	90
OTCNordPool	83	0	0	83
Total	***10,643***	***15,578***	***752***	***27,143***

Carbon Index (Datastream).[12]

2.9 Final transactions dataset and market-wide transaction price

Finally, we exclude from our dataset all the transactions after 31st May 2007. The final number of transactions is 28,548, including 27,143 real transactions (95.08%) and 1,405 non-real transactions. The number of transactions associated with a price or a proxy price is detailed in Table 4.

To compute the price advantage defined in the paper, we need a measure of the **hypothetical market-wide transaction price** of day t. As a proxy, we use the volume-weighted median transaction price of the day based on the exchange spot transactions with a transaction-specific price (transactions on Powernext, Nord Pool, EXAA, and SENDECO2 - 10,503 transactions)[13] and OTC transactions (including OTC transactions cleared by Powernext and Nord Pool - 16,133 transactions).

References

[1] Balietti, A.C. (2016). Trader types and volatility of emission allowance prices. Evidence from EU ETS Phase I. *Energy Policy*, 98, pp. 607-620.

[2] Ellerman, A.D., and Joskow, P. (2008). The European Union's CO2 Cap-and-Trade System in Perspective. *Pew Center on Global Climate Change Report.* (http://www.pewclimate.org/eu-ets)

[12]771 real OTC transactions could not be associated with a proxy price because for those days the Point Carbon Index is missing.

[13]Note that for Nord Pool, the number of transactions shown in Table 4 includes 609 spot transactions and 110 futures transactions. The 110 futures transactions are excluded when computing the volume-weighted median transaction price