PRINCETON UNIVERSITY

ORF 474: High Frequency Trading
Spring 2020
Robert Almgren

Lecture 7a

March 23, 2020

Grading change

- Old formula:
  - 60% HW
  - 20% midterm
  - 20% final HW
- New formula:
  - 70% HW
  - 10% midterm
  - 20% final HW
- Will compute better of old and new formulas
- P/D/F at student's individual option (no limits)

- Recent market events
- Review last week: futures, tick data, Kdb+
- Characterization of futures products

  diversification across maturities: entropy, etc
- Tick size

FRED ECONOMIC DATA | ST. LOUIS FED

ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

MY ACCOUNT

Search FRED

FRED® Economic Data | Information Services | Publications | Working Papers | Economists | About | St. Louis Fed Home

Categories > Money, Banking, & Finance > Financial Indicators > Stock Market Indexes

☆ S&P 500 (SP500)

DOWNLOAD

Observation:
2020-03-20: **2,304.92** (+ more)
Updated: Mar 20, 2020

Units:
Index,
Not Seasonally Adjusted

Frequency:
Daily,
Close

1Y | 5Y | 10Y | Max

2015-03 to 2020-03

EDIT GRAPH

FRED — S&P 500

Shaded areas indicate U.S. recessions

Source: S&P Dow Jones Indices LLC

fred.stlouisfed.org

https://fred.stlouisfed.org/series/SP500

VIX index is a good single source for market volatility

The Cboe Volatility Index - more commonly referred to as the "VIX Index" - is an up-to-the-minute market estimate of expected volatility that is calculated by using real-time prices of options on the S&P 500® Index listed on Cboe Exchange, Inc. ("Cboe Options") (Symbol: SPX). Only SPX options with Friday expirations are used to calculate the VIX Index. The VIX Index is calculated between 2:15 a.m. CT and 8:15 a.m. CT and between 8:30 a.m. CT and 3:15 p.m. CT. Only SPX options with more than 23 days and less than 37 days to the Friday SPX expiration are used to calculate the VIX Index. These SPX options are then weighted to yield a constant, 30-day measure of the expected volatility of the S&P 500 Index.

http://www.cboe.com/products/futures/vx-cboe-volatility-index-vix-futures/contract-specifications

Trading halts

Circuit breakers

The basic idea of stock-market circuit breakers is, like, some news happens, and the market reacts precipitously, and stocks fall 7%, and the market gets turned off for 15 minutes so that everyone can have some time to think and digest the news and see if they want to buy. On an average Tuesday afternoon, not everyone who might want to buy stocks is watching the market every minute. The computers are, sure, but some long-term investors are busy doing other things, reading 10-Ks or meeting with executives or whatever. Someone needs to call them up and say "hey not sure if you noticed but stocks are cheap now, you should buy some."

https://www.bloomberg.com/opinion/articles/2020-03-09/stuff-is-bad-so-stocks-are-down

Whole market

Market-Wide Circuit Breakers

Market-Wide Circuit Breakers FAQ[PDF]

The equities and options exchanges have procedures for coordinated cross-market trading halts if a severe market price decline reaches levels that may exhaust market liquidity. These procedures, known as market-wide circuit breakers ("MWCB"), may halt trading temporarily or, under extreme circumstances, close the markets before the normal close of the trading session. MWCBs provide for cross-market trading halts during a severe market decline as measured by a single-day decrease in the S&P 500 Index. A cross-market trading halt can be triggered at three circuit breaker thresholds that measure a decrease against the prior day's closing price of the S&P 500 Index -- 7% (Level 1), 13% (Level 2), and 20% (Level 3) (See NYSE, NYSE American and NYSE Arca Rule 7.12).

Individual stocks

Limit Up/Limit Down

On May 31, 2012, the SEC approved, on a pilot basis, a National Market System Plan, known as the Limit Up/Limit Down ("LULD") Plan, to address extraordinary market volatility.

The Plan is designed to prevent trades in NMS Stocks from occurring outside specified price bands, which are set at a percentage level above and below the average reference price of a security over the preceding five-minute period. The percentage level is determined by a security's designation as a Tier 1 or Tier 2 security. Tier 1 comprises all securities in the S&P 500, the Russell 1000 and select Exchange Traded Products (ETPs). Tier 2 comprises all other NMS securities, except for rights and warrants, which are specifically excluded from coverage. The Plan applies during regular trading hours of 9:30 am ET - 4:00 pm ET.

https://www.nyse.com/markets/nyse/trading-info

Floor closure

Updates on Trading at CME Group

CME Group closed its Chicago trading floor as of the close of business Friday, March 13, 2020, as a precaution to reduce large gatherings that can contribute to the spread of coronavirus in line with the advice of medical professionals. All products continue to trade on CME Globex. Read more below and check this page for any additional relevant trading updates.

On Monday, March 16, Cboe temporarily closed its Chicago trading floor and transitioned Cboe Options Exchange (C1) to an all-electronic mode. The transition took place as designed and without incident. This was a proactive measure and not in response to any confirmed cases of COVID-19 on the trading floor. Cboe will follow public health guidance when evaluating the reopening of the trading floor.

MarketWatch • Latest Watchlist Markets Investing Barron's Economy Personal Finance

BULLETIN Chevron and UTX are the sharpest decliners as the Dow drops by nearly 500 points →

Home > Markets > U.S. & Canada > **Market Extra**

Market Extra

As the stock market is in turmoil, here's what experts are watching for as the NYSE operates without humans for the first time

Published: March 23, 2020 at 9:15 a.m. ET

By Mark DeCambre

The New York Stock Exchange is going all-electronic on Monday, marking the first time that the centuries-old exchange will operate in regular hours without its legion of trusted flesh-and-blood floor traders.

"It's unprecedented, it's never happened before," said one NYSE official who declined to be identified because they weren't authorized to publicly talk about the exchange's operations. "Our markets have never been open while our trading floor's not open."

GET THE BASICS

The Global Derivatives Marketplace

Futures. Options. Forwards. Margins. Algorithms. Liquidity. If you think this is all above your head, think again.

Understanding derivatives starts with understanding one simple concept: risk. If you buy everyday products, own property, run a business or manage money for investors, risk is all around you every day. For some, risk stands between them and progress. For others, risk represents an opportunity to invest. If you've bought or sold a home, a car or shares of stock, you know that there is always a risk that the price of what you buy or sell today could shift depending on market movement.

In a derivatives marketplace, individuals and businesses everywhere are able to lock in a future price by putting it into a binding contract. These products are called futures and options – contractual agreements to buy or sell an amount of something at a fixed price at a future date.

This enables them to navigate business and financial risks or assume risk in an effort to make money when prices fluctuate.

A real world example

In the summer of 2012, the United States experienced its most severe drought since the Dust Bowl of the 1930s. The drought had perhaps the biggest impact in the Midwest Corn Belt. As the season became dryer and hotter, corn farmers and country elevators that store corn for later sale bought corn futures at a certain price, for a certain date of sale. This guaranteed a level of profit, which helped plan for a year in which production and supply of their crop would most certainly be lower than normal.

Futures trading helps businesses around the world navigate risk and uncertainty. By locking in future prices using futures products, they are better equipped to take on new opportunities, act more quickly and decisively, grow their businesses and ultimately help reduce costs for you, the consumer.

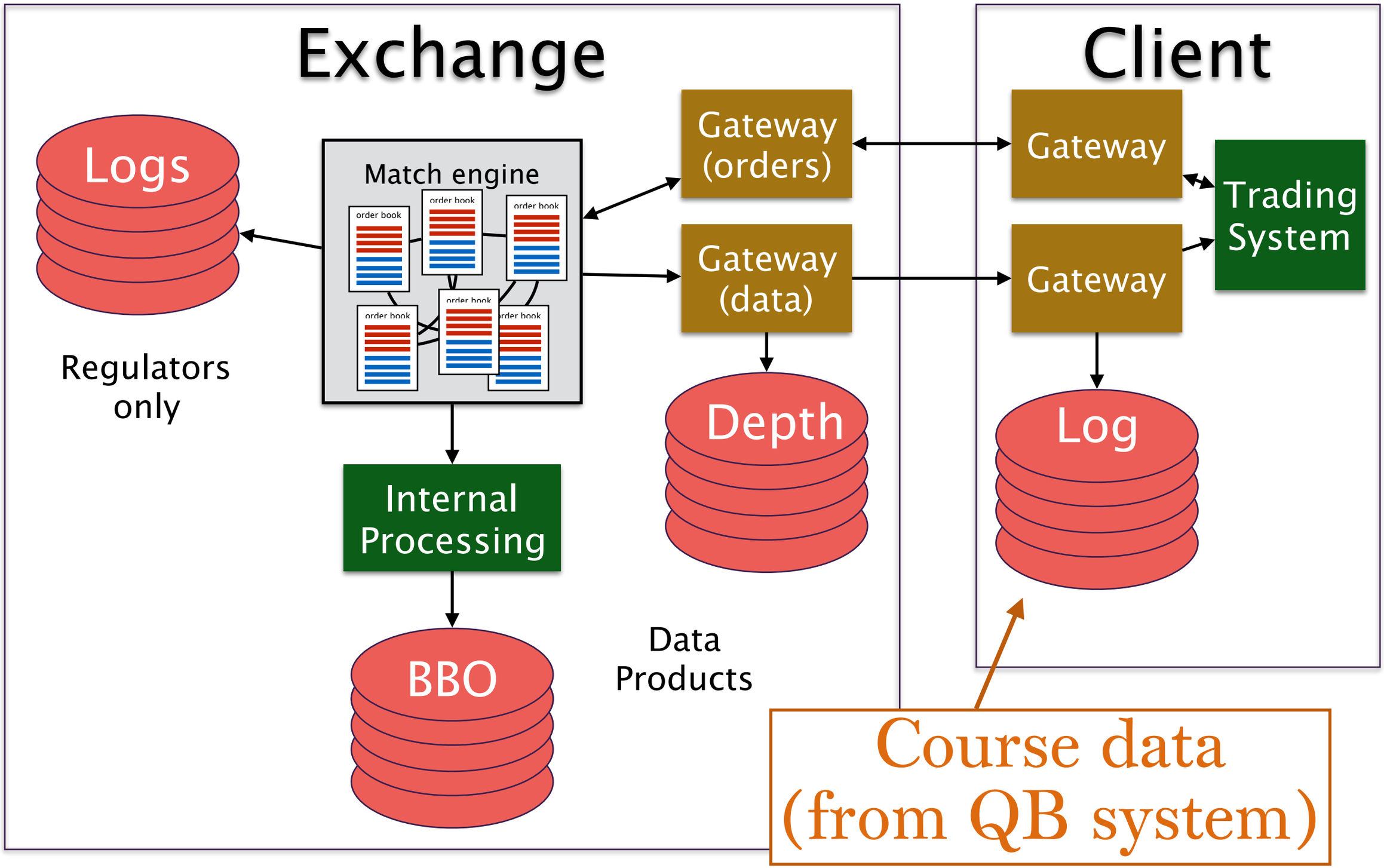
https://www.futuresfundamentals.org/get-the-basics/futures-and-options/ (CME group)

Database content

- All CME futures tick data for 9 weeks
  Monday Sept 18 through Friday Nov 17, 2017
- All products (rates, energy, equity index, etc)
  Interest rates / non-rates on different services
- All outrights and spreads
- Very precise and reliable time stamps

Market data pipeline

why this data is very clean



Kdb+: Standard software on Wall Street

- Kdb+ or Q `https://kx.com`
- Academic program
  `https://kx.com/connect-with-us/kx-academic-program/`
- Descendant of APL
- Designed for large time series databases
- 32- and 64-bit client interface free to download

  we have 64-bit server academic license from Kx
- Programming language as well as database

WHY Kx?

Kx is a suite of enterprise-level products and solutions centered around kdb+, the world's fastest time-series database. Kdb+is optimized for ingesting, analyzing, and storing massive amounts of structured data. The combination of the columnar design of kdb+and its in-memory capabilities means it offers greater speed and efficiency than typical relational databases. Its native support for time-series operations vastly improves both the speed and performance of queries, aggregation and analysis of structured data.

"Kx helps us because it allows us to gather huge streams of data, in real-time, to analyze that data very quickly, and to make decisions when seconds count."

Matt Cadieux
Chief Information Officer, Aston Martin Red Bull Racing

TIME-SERIES DATABASE

Kx technology is designed to capture the essence of time in computing. Our native support for temporal and bi-temporal features is a rare capability. High precision nanosecond timestamps, time-ordered querying, uniquely performant aggregation across flexibly-defined time buckets, coupled with time-based table joins of unparalleled speed, allows kdb+to stand alone in time-series analytics.

FEATURED RESOURCES > LEARN MORE >

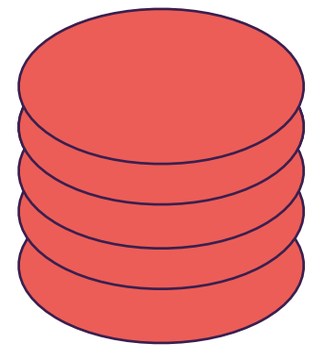
https://kx.com/discover

Q references

- Kx main code page
  `https://code.kx.com`
- "Q for Mortals"
  `http://code.kx.com/q4m3/`
- R connector
  `https://github.com/KxSystems/rkdb`
- Command line client
  (32-bit is OK, but does not work on Mac Catalina)
  `https://kx.com/connect-with-us/download/`

Command-line
client "q"

query (string)



data

R session with rkdb

Database server
orf474.princeton.edu

Kdb+, from Kx, is

- a high-performance historical time-series columnar database
- an in-memory compute engine
- a real-time streaming processor
- an expressive query and programming language called q

New

Simba ODBC 1.1

Search Box

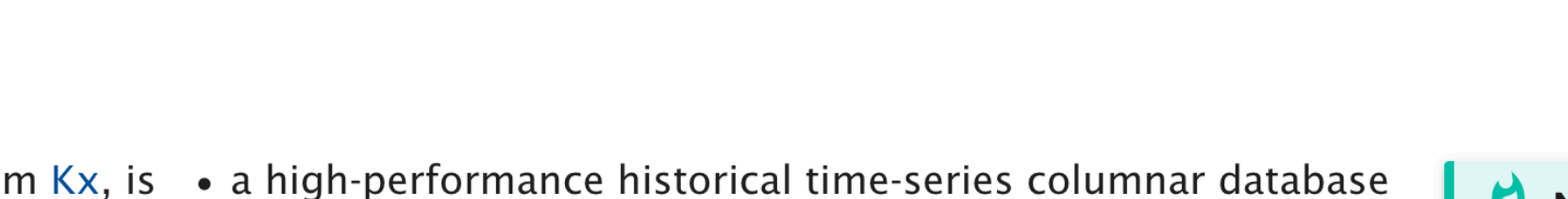
Surveillance latency

`man.q`

Q for Mortals 3.1

Kx Developer A visual environment used to manage, manipulate and explore massive datasets in real-time.

Free download



https://code.kx.com

Q basics

- variables and data types
  assignment
- arithmetic
  division, right-to-left precedence
- lists
- functions
- tables
- handles
- queries
- joins

query syntax (like SQL but not exactly)

Columns or functions of columns

`trade` or `quote`

Conditions on date, time, symbol, etc (comma for "and")

```
select ... from ... where ...
```

Grouping variable

```
select ... by ... from ... where ...
```

Q table syntax and joins

- Table creation (mostly for toy examples)
- Keyed tables

```
q)t:([] x: 1 2 2 3; y:`a`b`c`d )

q)t
x y
---
1 a
2 b
2 c
3 d

q)select last y by x from t
x| y
-| -
1| a
2| c
3| d
```

- lj left join
- aj as-of join

`lj` "left join"

- match on shared column
- right table must have key

```
q)t
x y
---
1 a
2 b
2 c
3 d

q)t1: ([] y:`a`c; z: 3.14 2.71 )

q)select last z by y from t1
y| z
-| ----
a| 3.14
c| 2.71

q)t lj select last z by y from t1
x y z
--------
1 a 3.14
2 b
2 c 2.71
3 d
```

lj example

```
q)h1:hopen `:orf474:6005

q)h1"([] inst:insts) lj select last name by inst from instinfo"
inst name
-------------------------------------
ZN   10 Yr Note
GE   Eurodollar
ES   E-mini S&P 500
ZF   5 Yr Note
CL   Crude Oil
GC   Gold
ZT   2 Yr Note
NQ   E-mini NASDAQ 100
ZB   U.S. Treasury Bond
6E   Euro FX
NG   Natural Gas (Henry Hub) Physical
6J   Japanese Yen
ZC   Corn
6B   British Pound
ZS   Soybean
ZQ   30 Day Federal Funds
6A   Australian Dollar
TN   Ultra 10 Yr Notes
YM   E-mini Dow ($5)
UB   Ultra U.S. Treasury Bond
HG   Copper
SI   Silver
RTY  R2000 Ind Mini
6C   Canadian Dollar
ZW   Wheat
ZL   Soybean Oil
6M   Mexican Peso
ZM   Soybean Meal
RB   RBOB Gasoline Physical
HO   NY Harbor ULSD
```

`aj` "as-of join"

typically

```
aj[ `seq; <trade table>; <quote table> ]
```

```
q)t:([] s: `float$1 5 6 7 9)
q)t
s
-
1
5
6
7
9
```

```
q)q:([] s:asc 5 ? 10.; b: 5 ? 20 )
q)q
s              b
---------------
2.37128796056 14
4.26917694043 17
5.67081043962 14
7.43328513578 17
9.99608161626 16
```

```
q)aj[ `s; t; q ]
s b
----
1
5 17
6 14
7 14
9 17
```

Q functions

```
q)f:{[x;y] x + 2*y }

q)f[1;2]
5
```

Function of 2 variables

```
q)g:f[1]

q)g 2
5

q)g 4
9
```

g[y] = f[1;y]

Futures spectrum across maturities

- Maturity = delivery date
- Some products trade only the "front month"
- Some trade across many maturities

Useful tables on server

- `instinfo`: information about each instrument
- `sym2inst`: mapping from symbol to instrument
- `sym2type`: mapping from symbol to type
- `sym2expir`: mapping from symbol to expiration

  not guaranteed to always be correct
- `insts`: 30 most important symbols

Root of symbol "instrument"

inst	class	name	subclass	cur	notional	minpxincr	minpxincrval	dispfactor	matchalgo
KE	AG	KC Hard Red Winter (HRW) Wheat	Grain And Oilseed	USD	50	0.25	12.5	1	O
ZC	AG	Corn	Grain And Oilseed	USD	50	0.25	12.5	1	K
ZL	AG	Soybean Oil							
ZM	AG	Soybean Meal							
ZS	AG	Soybean							
ZW	AG	Wheat							
GF	AG	Feeder Cattle							
HE	AG	Lean Hog							
LE	AG	Live Cattle							
CL	EN	Crude Oil							
NG	EN	Natural Gas (Henry Hub) Physical	Natural Gas	USD	10	1	10	0.001	F
HO	EN	NY Harbor ULSD	Refined Products	USD	4.2	1	4.2	0.0001	F
RB	EN	RBOB Gasoline Physical	Refined Products	USD	4.2	1	4.2	0.0001	F
RTY	EQ	R2000 Ind Mini							
NIY	EQ	Nikkei/Yen							
NKD	EQ	Nikkei/USD							
EMD	EQ	E-mini S&P MidCap 400							
ES	EQ	E-mini S&P 500	US Index	USD	0.5	25	12.5	0.01	F
NQ	EQ	E-mini NASDAQ 100	US Index	USD	0.2	25	5	0.01	F
SMC	EQ	E-mini S&P 600 SmallCap	US Index	USD	1	10	10	0.01	F
YM	EQ	E-mini Dow ($5)	US Index						
6M	FX	Mexican Peso	Emer						
6A	FX	Australian Dollar	Majo						
6B	FX	British Pound	Majo						
6C	FX	Canadian Dollar	Majo						
6E	FX	Euro FX	Majo						
6J	FX	Japanese Yen	Majors	USD	12.5	0.5	6.25	1e-06	F
6N	FX	New Zealand Dollar	Majors	USD	10	1	10	0.0001	F
6S	FX	Swiss Franc	Majors	USD	12.5	1	12.5	0.0001	F
B1U	IR	30 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.03125	31.25	1	K
F1U	IR	5 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.015625	15.625	1	K
N1U	IR	10 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.015625	15.625	1	K
T1U	IR	2 Year USD Deliverable Interest Rate Swap	Deliverable Swaps	USD	1000	0.0078125	7.8125	1	K
GE	IR	Eurodollar	Stirs	USD	25	0.5	12.5	0.01	A
ZQ	IR	30 Day Federal Funds	Stirs	USD	41.67	0.5	20.835	0.01	K

Terminology for CME Globex trading system:

`inst` = generic name for product

`symbol` = `inst` + month code + last digit of year

F	G	**H**	J	K	**M**	N	Q	**U**	V	X	**Z**
Jan	Feb	**Mar**	Apr	May	**Jun**	Jul	Aug	**Sep**	Oct	Nov	**Dec**

`inst` = ES = SP500 futures

`symbol` = ESZ7 = Dec 2017 expiration

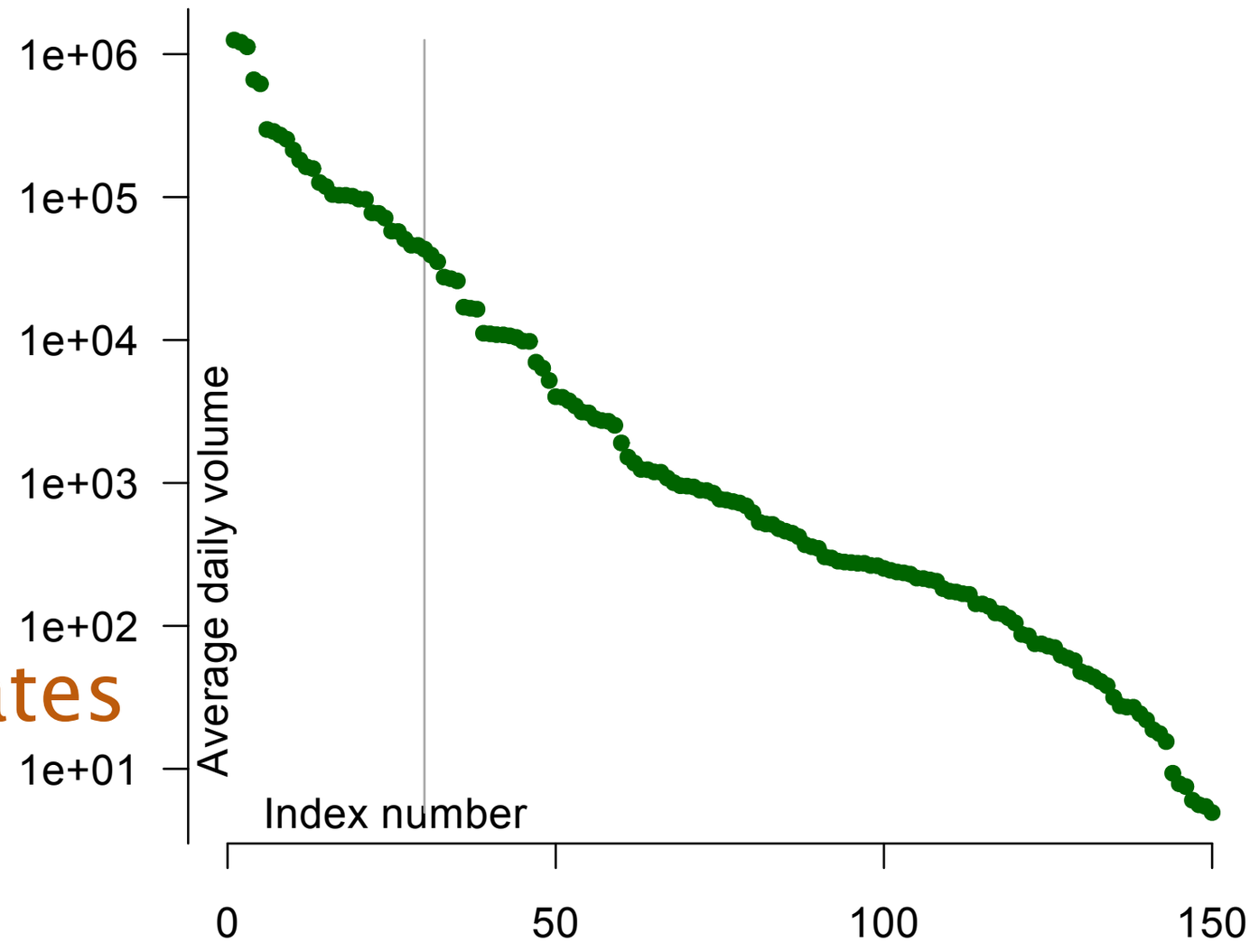
Most active instruments

```
q)h1"(`v xdesc select v:sum siz by inst:sym2inst[sym] from trade
        where date=2017.10.25,sym2type[sym]=`OT)
    lj select last name by inst from instinfo"
inst| v       name
----| ----------------------------------------------
ZN  | 1878740 10 Yr Note
GE  | 1480114 Eurodollar
ZF  | 962793  5 Yr Note
ZB  | 357519  U.S. Treasury Bond
ZT  | 312571  2 Yr Note
TN  | 147786  Ultra 10 Yr Notes
UB  | 125343  Ultra U.S. Treasury Bond
ZQ  | 81165   30 Day Federal Funds
...
```

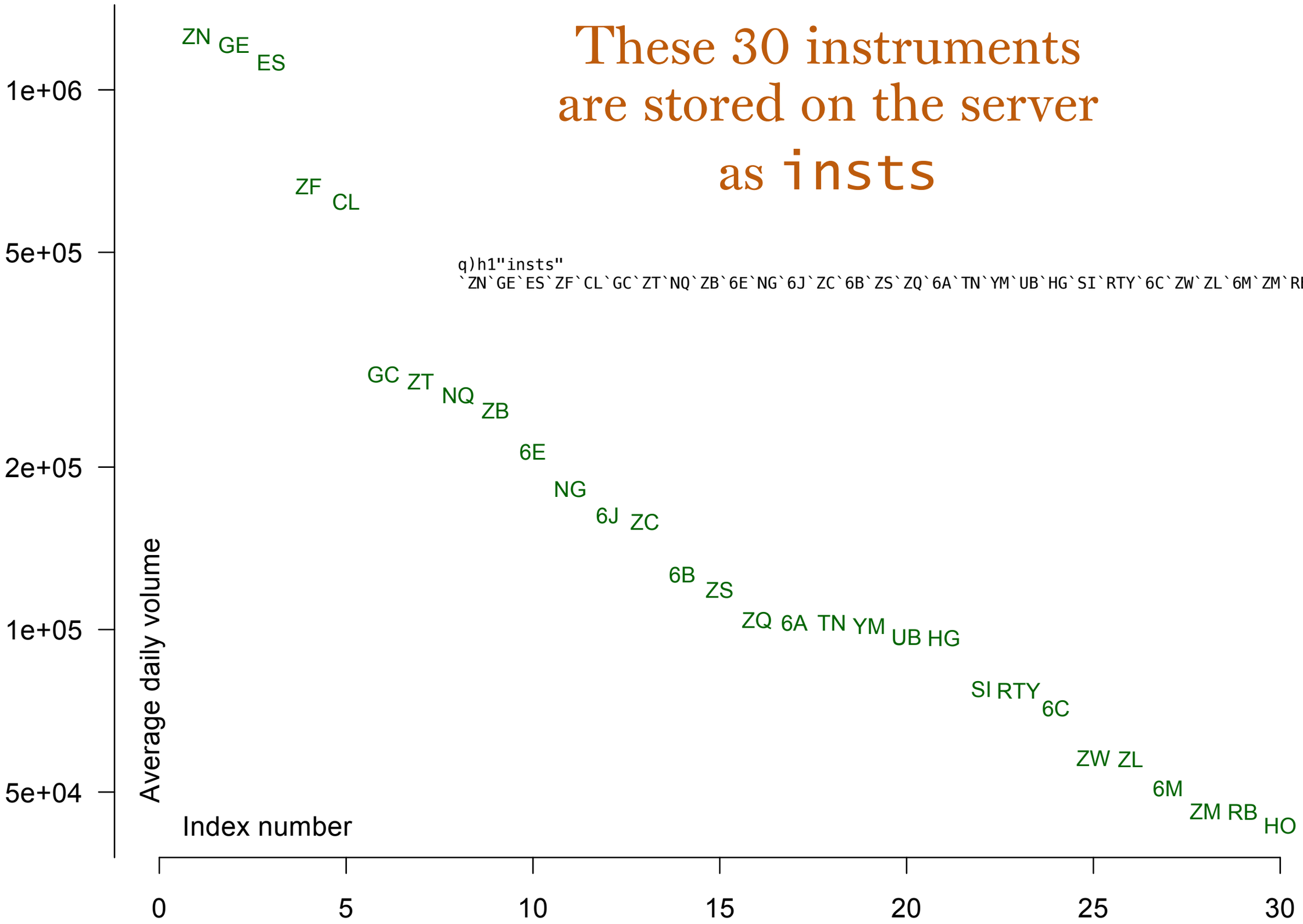
rates

```
q)h2"(`v xdesc select v:sum siz by inst:sym
        where date=2017.10.25,sym2type[sym]=
    lj select last name by inst from instin
inst| v       name
----| ----------------------------------
ES  | 2084658 E-mini S&P 500
CL  | 567716  Crude Oil
NQ  | 435947  E-mini NASDAQ 100
GC  | 365748  Gold
6J  | 205358  Japanese Yen
6E  | 204538  Euro FX
ZC  | 196873  Corn
YM  | 188698  E-mini Dow ($5)
...
```

non-rates



30 most active instruments



These 30 instruments are stored on the server as `insts`

```
q)h1"insts"
`ZN`GE`ES`ZF`CL`GC`ZT`NQ`ZB`6E`NG`6J`ZC`6B`ZS`ZQ`6A`TN`YM`UB`HG`SI`RTY`6C`ZW`ZL`6M`ZM`RB`HO
```

"Interesting" products

`Notional` = dollar value of unit change in contract price

`Minpxincr` = "tick"

`dispfactor` = price multiplier

```
q)h"insts"
`ZC`LE`ZS`ZW`CL`HO`NG`RB`ES`NQ`YM`6A`6B`6C`6E`6J`GE`ZF`ZT`ZN`UB`ZB`GC`HG`SI

q)h1"select from instinfo where inst in insts"
inst class name                                subclass          cur notional minpxincr minpxincrval dispfactor matchalgo
-------------------------------------------------------------------------------------------------------------------------
ZC   AG    Corn                                Grain And Oilseed USD 50       0.25      12.5         1          K
ZL   AG    Soybean Oil                         Grain And Oilseed USD 6        1         6            0.01       K
ZM   AG    Soybean Meal                        Grain And Oilseed USD 10       1         10           0.1        K
ZS   AG    Soybean                             Grain And Oilseed USD 50       0.25      12.5         1          K
ZW   AG    Wheat                               Grain And Oilseed USD 50       0.25      12.5         1          K
CL   EN    Crude Oil                           Crude Oil         USD 10       1         10           0.01       F
NG   EN    Natural Gas (Henry Hub) Physical    Natural Gas       USD 10       1         10           0.001      F
HO   EN    NY Harbor ULSD                      Refined Products  USD 4.2      1         4.2          0.0001     F
RB   EN    RBOB Gasoline Physical              Refined Products  USD 4.2      1         4.2          0.0001     F
RTY  EQ    R2000 Ind Mini                                        USD 0.5      10        5            0.01       F
ES   EQ    E-mini S&P 500                      US Index          USD 0.5      25        12.5         0.01       F
NQ   EQ    E-mini NASDAQ 100                   US Index          USD 0.2      25        5            0.01       F
YM   EQ    E-mini Dow ($5)                     US Index          USD 5        1         5            1          F
6M   FX    Mexican Peso                        Emerging Market   USD 0.5      10        5            1e-06      F
6A   FX    Australian Dollar                   Majors            USD 10       1         10           0.0001     F
6B   FX    British Pound                       Majors            USD 6.25     1         6.25         0.0001     F
6C   FX    Canadian Dollar                     Majors            USD 10       0.5       5            0.0001     F
6E   FX    Euro FX                             Majors            USD 12.5     0.5       6.25         0.0001     F
6J   FX    Japanese Yen                        Majors            USD 12.5     0.5       6.25         1e-06      F
GE   IR    Eurodollar                          Stirs             USD 25       0.5       12.5         0.01       A
ZQ   IR    30 Day Federal Funds                Stirs             USD 41.67    0.5       20.835       0.01       K
TN   IR    Ultra 10 Yr Notes                   US Treasury       USD 1000     0.015625  15.625       1          F
UB   IR    Ultra U.S. Treasury Bond            US Treasury       USD 1000     0.03125   31.25        1          F
ZB   IR    U.S. Treasury Bond                  US Treasury       USD 1000     0.03125   31.25        1          F
ZF   IR    5 Yr Note                           US Treasury       USD 1000     0.0078125 7.8125       1          F
ZN   IR    10 Yr Note                          US Treasury       USD 1000     0.015625  15.625       1          F
ZT   IR    2 Yr Note                           US Treasury       USD 2000     0.0078125 15.625       1          K
HG   MT    Copper                              Base              USD 2.5      5         12.5         0.0001     F
GC   MT    Gold                                Precious          USD 10       1         10           0.1        F
SI   MT    Silver                              Precious          USD 5        5         25           0.001      F
```

Scale factors: example

```
inst class name                                subclass          cur notional minpxincr minpxincrval dispfactor matchalgo
-----------------------------------------------------------------------------------------------------------------------
6C    FX    Canadian Dollar                    Majors            USD 10       0.5       5            0.0001     F
```

```
q)h"select v:sum siz by prc from trade where date=2017.10.26,sym=`6CZ7"
prc    | v
-------| ----
7780   | 8
7780.5| 66
7781   | 360
7781.5| 300
7782   | 485
7782.5| 665
7783   | 848
7783.5| 747
7784   | 903
7784.5| 1220
7785   | 1093
7785.5| 1293
..
```

minpxincr=0.5

Notional = USD 10:
Unit change in contract price
is worth $10

Dispfactor=0.0001: CAD/USD = 0.7782, etc

$$\Delta p = 1$$
$$= 1e\text{-}4 \text{ on } 1e5 \text{ CAD}$$
$$= \$10 \text{ USD}$$

Canadian Dollar Futures Contract Specs

View Another Product

Quotes | Settlements | Volume | Time & Sales | Contract Specs | Margins | Calendar

Futures | Options

Contract Unit	100,000 Canadian dollars
Trading Hours	Sunday - Friday 6:00 p.m. - 5:00 p.m. (5:00 p.m. - 4:00 p.m. Chicago Time/CT) with a 60-minute break each day beginning at 5:00 p.m. (4:00 p.m. CT)
Minimum Price Fluctuation	Outrights: .00005 USD per CAD ($5.00 USD). Consecutive Month Spreads (Globex only): 0.00001 USD per CAD (1.00 USD) All other Spread Combinations: 0.00005 USD per CAD (5.00 USD).
Product Code	CME Globex: 6C CME ClearPort: C1 Clearing: C1
Listed Contracts	Contracts listed for the first 3 consecutive months and 20 months in the March quarterly cycle (Mar, Jun, Sep, Dec)
Settlement Method	Deliverable
Termination Of Trading	9:16 a.m. Central Time (CT) on the business day immediately preceding the third Wednesday of the contract month (usually Tuesday).
Settlement Procedures	Physical Delivery Canadian Dollar Settlement Procedures

Match algorithms

```
q)h2"(`matchalgo xasc select inst,name,matchalgo from instinfo where inst in insts) lj matchalgos"
inst name                                matchalgo algoname
-----------------------------------------------------------------------
GE   Eurodollar                          A         Allocation
CL   Crude Oil                           F         FIFO
NG   Natural Gas (Henry Hub) Physical    F         FIFO
HO   NY Harbor ULSD                      F         FIFO
RB   RBOB Gasoline Physical              F         FIFO
RTY  R2000 Ind Mini                      F         FIFO
ES   E-mini S&P 500                      F         FIFO
NQ   E-mini NASDAQ 100                   F         FIFO
YM   E-mini Dow ($5)                     F         FIFO
6M   Mexican Peso                        F         FIFO
6A   Australian Dollar                   F         FIFO
6B   British Pound                       F         FIFO
6C   Canadian Dollar                     F         FIFO
6E   Euro FX                             F         FIFO
6J   Japanese Yen                        F         FIFO
TN   Ultra 10 Yr Notes                   F         FIFO
UB   Ultra U.S. Treasury Bond            F         FIFO
ZB   U.S. Treasury Bond                  F         FIFO
ZF   5 Yr Note                           F         FIFO
ZN   10 Yr Note                          F         FIFO
HG   Copper                              F         FIFO
GC   Gold                                F         FIFO
SI   Silver                              F         FIFO
ZC   Corn                                K         Split FIFO and Pro-Rata
ZL   Soybean Oil                         K         Split FIFO and Pro-Rata
ZM   Soybean Meal                       K         Split FIFO and Pro-Rata
ZS   Soybean                            K         Split FIFO and Pro-Rata
ZW   Wheat                              K         Split FIFO and Pro-Rata
ZQ   30 Day Federal Funds                K         Split FIFO and Pro-Rata
ZT   2 Yr Note                          K         Split FIFO and Pro-Rata
```

Pro rata

Time priority

Mixed

Most active sym for each "interesting" contract

```
q)h1:hopen `:orf474:6003
q)h2:hopen `:orf474:7003

q)h1"select first sym,first v by inst from
      update inst:sym2inst[sym] from
       `v xdesc
         select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym] in insts,sym2type[sym]=`OT"
inst| sym  v
----| ------------
GE  | GEZ8 206841
TN  | TNZ7 147786
UB  | UBZ7 125342
ZB  | ZBZ7 357455
ZF  | ZFZ7 962785
ZN  | ZNZ7 1878572
ZQ  | ZQF8 18482
ZT  | ZTZ7 311933

q)h2"select first sym,first v by inst from
      update inst:sym2inst[sym] from
       `v xdesc
        select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym] in insts,sym2type[sym]=`OT"
inst| sym   v
----| -------------
6A  | 6AZ7  145711
6B  | 6BZ7  151376
6C  | 6CZ7  136228
6E  | 6EZ7  202545
6J  | 6JZ7  204642
6M  | 6MZ7  80084
CL  | CLZ7  506586
ES  | ESZ7  2082705
GC  | GCZ7  363223
HG  | HGZ7  100796
HO  | HOZ7  35471
NG  | NGZ7  82811
NQ  | NQZ7  435242
RB  | RBZ7  37096
RTY | RTYZ7 95737
SI  | SIZ7  88525
YM  | YMZ7  188541
ZC  | ZCZ7  141346
ZL  | ZLZ7  44246
ZM  | ZMZ7  35141
ZS  | ZSX7  54143
ZW  | ZWZ7  42794
```

For almost all products,
the first quarterly expiration
is the most active
(Dec 2017 = Z7, in Oct 2017)

When we talk about "the" symbol
for a particular instrument,
we mean the symbol that is
most actively traded

What Crude Oil products trade?

Port 7000-7009 for non-interest rate

Always put date condition first

Select products in the "CL" family using `sym2inst`

```
q)h:hopen `:orf474:7004

q)h"select v:sum siz by sym from trade where date=2017.10.24,sym2inst[sym]=`CL"
sym           | v
--------------| ----
CL:BF F8-G8-H8| 174
CL:BF F8-J8-N8| 1
CL:BF G8-H8-J8| 111
CL:BF H8-J8-K8| 85
CL:BF H8-M8-U8| 4
CL:BF J8-K8-M8| 15
CL:BF J9-K9-M9| 6
CL:BF K8-M8-N8| 739
CL:BF K9-M9-N9| 320
CL:BF M8-U8-Z8| 5
CL:BF M8-Z8-M9| 53
CL:BF M9-Z9-M0| 76
CL:BF N8-Q8-U8| 20
..
```

These are strange symbols! "Butterflies"

Classify by contract type using `sym2type`

```
q)h"select n:count distinct sym by typ:sym2type[sym] from trade where date=2017.10.24,sym2inst[sym]=`CL"
typ| n
---| ---
BF | 22
IC | 11
OT | 20
SP | 121
```

Butterflies (spreads of spreads)

Outrights ("normal" futures)

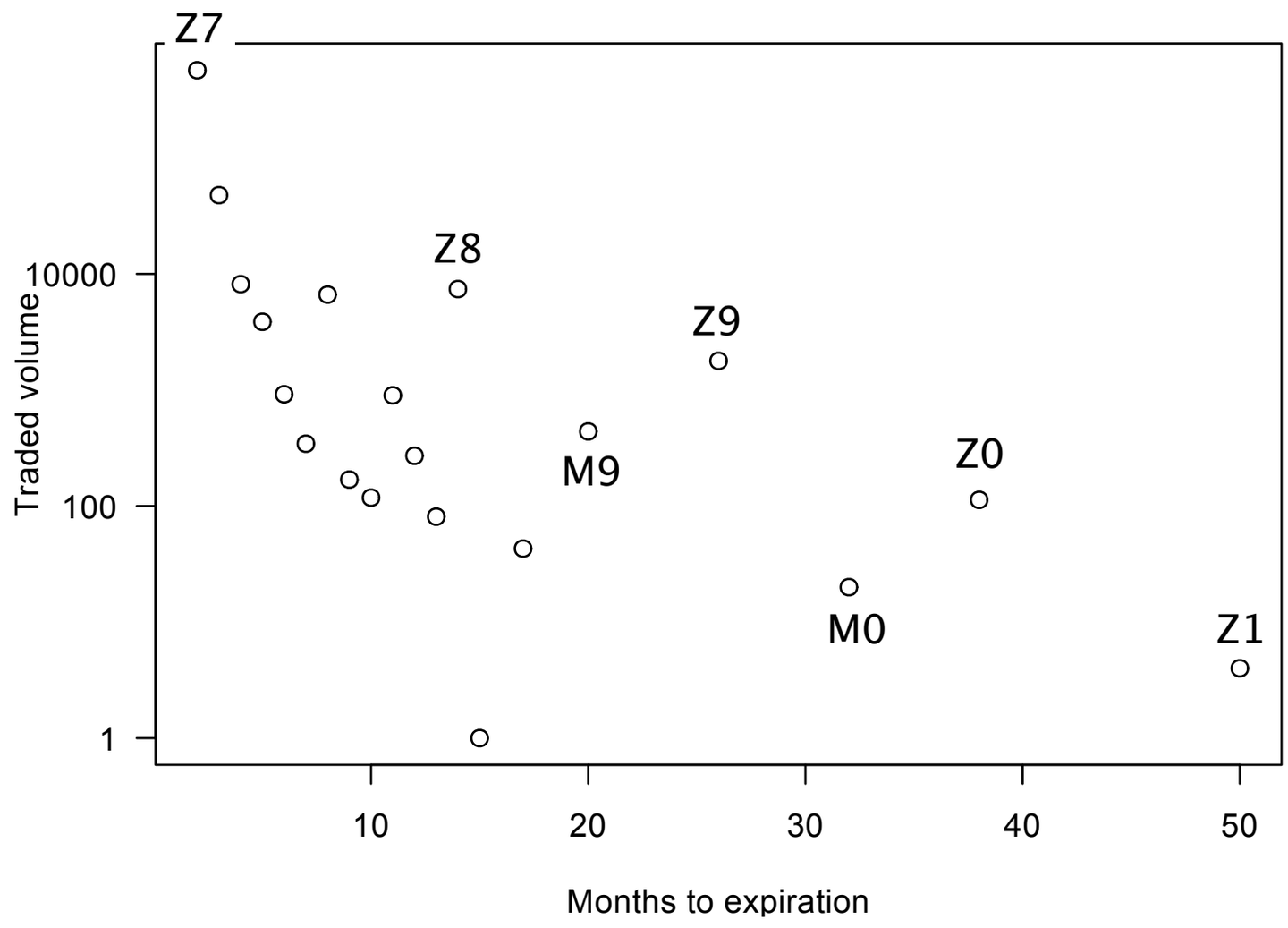
Calendar spreads

Distribution of traded volume for CL outrights

Add the expiration using `sym2expir` and sort by that

Select only outright symbols (you almost always want to do this)

```
q)h"`expir xasc update expir:sym2expir[sym] from
        select v:sum siz by sym from trade where date=2017.10.24,sym2inst[sym]=`CL,sym2type[sym]=`OT"
sym | v      expir
----| --------------
CLZ7| 568398 2017.12
CLF8| 47706  2018.01
CLG8| 8156   2018.02
CLH8| 3864   2018.03
CLJ8| 917    2018.04
CLK8| 344    2018.05
CLM8| 6632   2018.06
CLN8| 169    2018.07
CLQ8| 118    2018.08
CLU8| 898    2018.09
CLV8| 271    2018.10
CLX8| 81     2018.11
CLZ8| 7399   2018.12
CLF9| 1      2019.01
CLH9| 43     2019.03
CLM9| 439    2019.06
CLZ9| 1780   2019.12
CLM0| 20     2020.06
CLZ0| 113    2020.12
CLZ1| 4      2021.12
```



"Entropy" and "Effective number"

```
q)h2"`v xdesc select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym]=`ES,sym2type[sym]=`OT"
sym | v
----| -------
ESZ7| 2082705
ESH8| 1903
ESM8| 37
ESU8| 13

q)h2"`v xdesc select v:sum siz by sym from trade where date=2017.10.25,sym2inst[sym]=`CL,sym2type[sym]=`OT"
sym | v
----| ------
CLZ7| 506586
CLF8| 42942
CLG8| 4990
CLZ8| 4021
CLM8| 3065
CLH8| 2737
CLJ8| 1000
CLZ9| 956
CLK8| 400
CLN8| 308
CLM9| 277
CLU8| 199
CLQ8| 101
CLZ0| 91
CLV8| 23
CLX8| 13
CLZ1| 5
CLM0| 1
CLZ2| 1
```

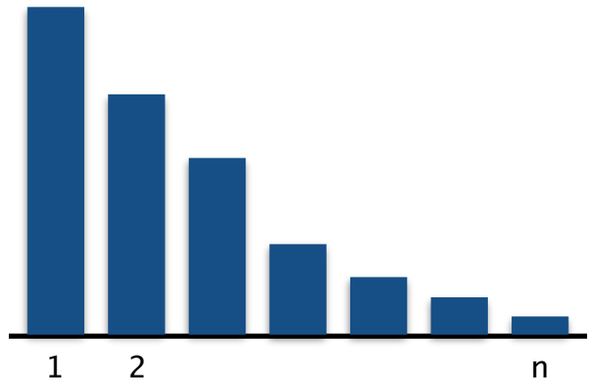
ES has "few" active symbols

CL has "many" active symbols
but how many?

Should we count these tiny ones?

Diversity measures

Given $p_1, \ldots, p_n$ with $p_1 + \cdots + p_n = 1$



What is the "effective number" of values?

Ecology -- number of species in an environment

- Simpson index

Antitrust -- number of companies in an industry

- Herfindahl-Hirschman index

Signal processing -- information content

- Shannon entropy

Futures markets -- characterization of products

Antitrust

THE UNITED STATES DEPARTMENT of JUSTICE

HERFINDAHL-HIRSCHMAN INDEX

The term “HHI” means the Herfindahl–Hirschman Index, a commonly accepted measure of market concentration. The HHI is calculated by squaring the market share of each firm competing in the market and then summing the resulting numbers. For example, for a market consisting of four firms with shares of 30, 30, 20, and 20 percent, the HHI is 2,600 ($30^2 + 30^2 + 20^2 + 20^2 = 2{,}600$).

The HHI takes into account the relative size distribution of the firms in a market. It approaches zero when a market is occupied by a large number of firms of relatively equal size and reaches its maximum of 10,000 points when a market is controlled by a single firm. The HHI increases both as the number of firms in the market decreases and as the disparity in size between those firms increases.

Diverging Trends in National and Local Concentration*

Esteban Rossi-Hansberg
Princeton University

Pierre-Daniel Sarte
Federal Reserve Bank of Richmond

Nicholas Trachter
Federal Reserve Bank of Richmond

This version: January 30, 2020
First version: September 12, 2018

Using U.S. NETS data, we present evidence that the *positive* trend observed in *national* product-market concentration between 1990 and 2014 becomes a *negative* trend when we focus on measures of *local* concentration. We document diverging trends for several geographic definitions of local markets. SIC 8 industries with diverging trends are pervasive across sectors. In these industries, top firms have contributed to the amplification of both trends. When a top firm opens a plant, local concentration declines and remains lower for at least 7 years. Our findings, therefore, reconcile the increasing national role of large firms with falling local concentration, and a likely more competitive local environment.

JEL codes: D22, D43, L16, L22, R12.

Let $S^{I,G}_{e,i,\ell,t}$ denote the nominal sales of enterprise e in industry i at location ℓ in year t, and $S^{I,G}_{e,i,g,t} = \sum_{\ell \in g} S^{I,G}_{e,i,\ell,t}$ its sales in the broader geography g (i.e. the sum of all its establishments' sales across all latitude-longitude pairs ℓ in geography g). The index I refers to the industrial level of aggregation (i.e. by SIC 2, SIC 4, SIC 6, or SIC 8). The index G indicates the geographic level of aggregation (i.e. by ZIP code level, CBSA level, County level, or the whole U.S.) that we use to define a location ℓ. We then denote by $s^{I,G}_{e,i,g,t}$ this enterprise's share of all sales in industry i located in geography g at date t for the levels of aggregation I and G. We adopt as our benchmark measure of market concentration the Herfindahl-Hirschman Index (HHI),

$$C^{I,G}_{i,g,t} = \sum_{e} \left(s^{I,G}_{e,i,g,t} \right)^2,$$

where $C^{I,G}_{i,g,t} \in [1/N^{I,G}_{i,g,t}, 1]$ is the sales concentration, and $N^{I,G}_{i,g,t}$ the number of enterprises in industry i and geography g at time t. In the online-only supplement to this paper, we also consider alternative measures of

https://www.princeton.edu/~erossi/DTNLC.pdf

DIVERSITY AND EVENNESS: A UNIFYING NOTATION AND ITS CONSEQUENCES[1]

M. O. HILL[2]

School of Plant Biology, University College of North Wales, Bangor, Caernarvonshire, Wales

Abstract. Three commonly used measures of diversity, Simpson's index, Shannon's entropy, and the total number of species, are related to Rényi's definition of a generalized entropy. A unified concept of diversity is presented, according to which there is a continuum of possible diversity measures. In a sense which becomes apparent, these measures provide estimates of the effective number of species present, and differ only in their tendency to include or to ignore the relatively rarer species. The notion of the diversity of a community as opposed to that of a sample is examined, and is related to the asymptotic form of the species-abundance curve. A new and plausible definition of evenness is derived.

When we say that the humid tropics are more diverse than the tundra, we mean that there are more species there. More precisely, we mean that the species in the humid tropics have on average lower proportional abundances than those in the tundra—a fact which is amply visible to the naked eye and which can be demonstrated by the use of any measure of diversity we care to devise. But there is little point in merely confirming the obvious; the purpose of determining diversity by a numerical index is rather to provide a means of comparison between less clear-cut cases. Unfortunately, when we look for a suitable numerical definition, we find that no particular formula has a pre-eminent advantage, and that different authors have plausibly proposed different indices. In the ensuing confusion, Hurlbert (1971) has despaired, declaring diversity to be a nonconcept. Fortunately his despair is premature, and when carefully defined according to an appropriate notation, diversity can be as unequivocal as any other ecological parameter.

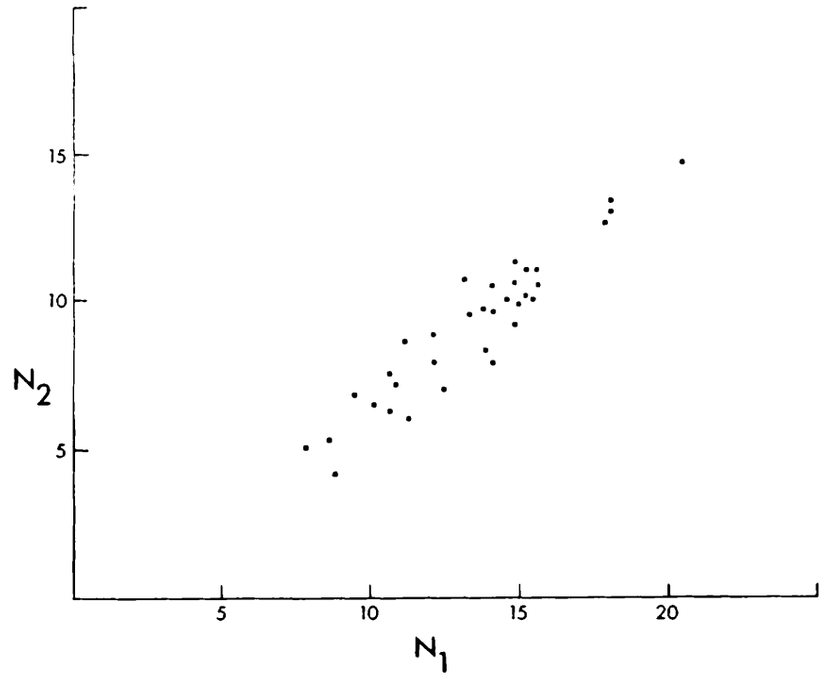


FIG. 3. Scatter diagram of N_1 versus N_2, computed for 30 cm × 30 cm dry-weight samples in a pasture in North Wales.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

6. CHOICE, UNCERTAINTY AND ENTROPY

We have represented a discrete information source as a Markoff process. Can we define a quantity which will measure, in some sense, how much information is "produced" by such a process, or better, at what rate information is produced?

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant.

Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

Mean of $(x_1, \ldots, x_n)$ under $(p_1, \ldots, p_n)$

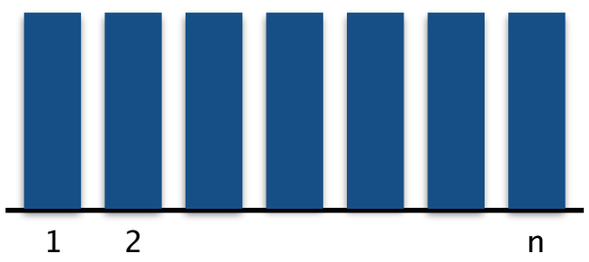
$$\langle x \rangle \;=\; \sum p_j \, x_j$$

Generalized mean of $(x_1, \ldots, x_n)$ under $(p_1, \ldots, p_n)$

$$M_q(x) \;=\; \langle x^q \rangle^{1/q} \;=\; \left(\sum p_j \, x_j^q \right)^{1/q}$$

Effective number of order q

$$N_q(p) \;=\; \frac{1}{M_q(p)} \;=\; \frac{1}{\left(\sum p_j^{q+1} \right)^{1/q}}$$



$$p_j = \frac{1}{n} \quad \Longrightarrow \quad N_q(p) = n$$

Shannon entropy (q=0)

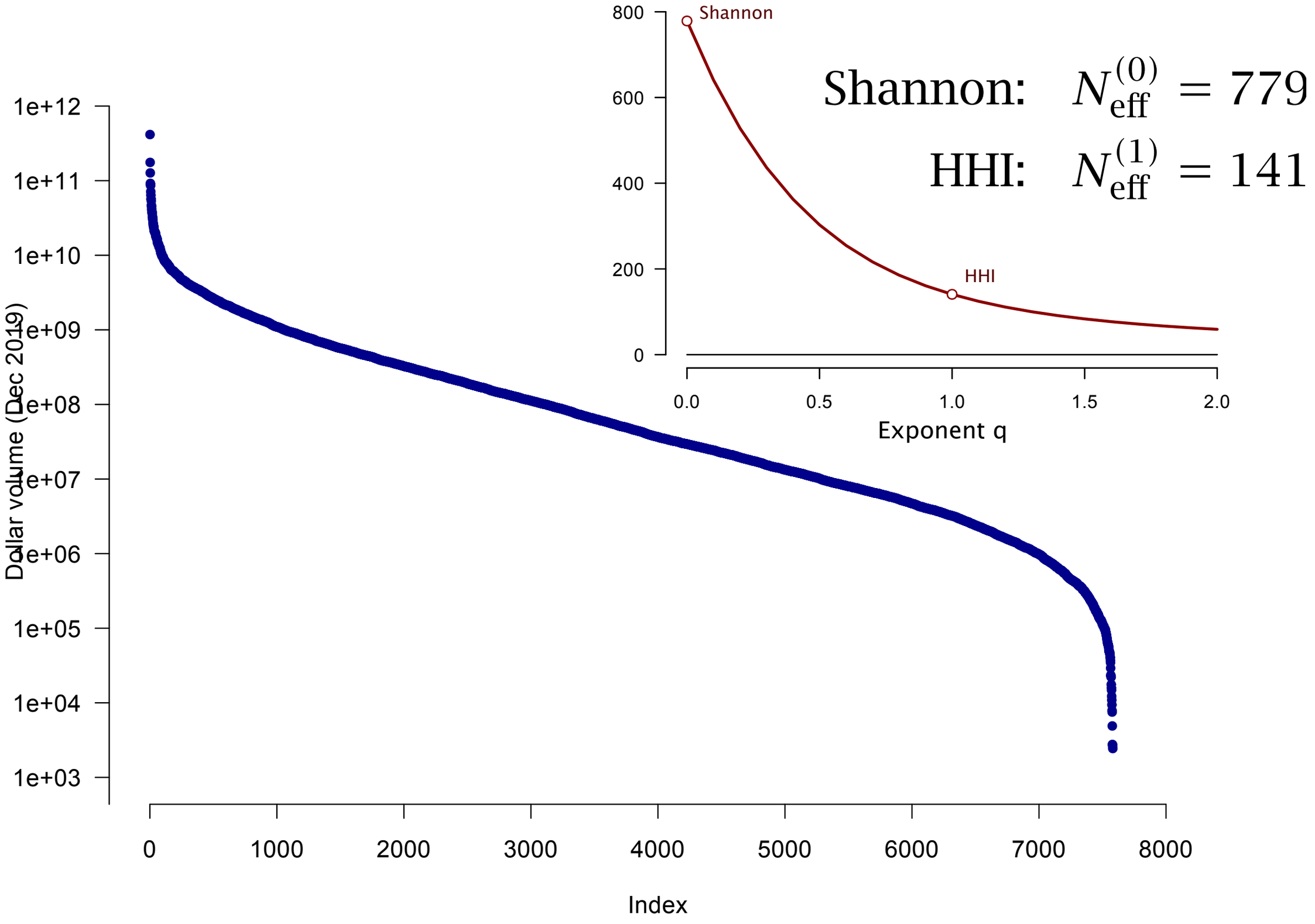
$$N_0(p) \;=\; \exp \sum -p_j \log p_j$$

Simpson index (ecology)
Herfindahl-Hirschman index (economics) (q=1)

$$N_1(p) \;=\; \frac{1}{\sum p_j^2}$$

$$N_\infty(p) \;=\; \frac{1}{\max p_j}$$

US equity volume, Dec 2019



$$\text{Shannon:} \quad N_{\text{eff}}^{(0)} = 779$$

$$\text{HHI:} \quad N_{\text{eff}}^{(1)} = 141$$

Apply to a single instrument

```
q)h2"{[dmin;dmax;inst]
    exec exp sum neg p*log p from
    update p:v % sum v from
    select v:sum siz by sym from trade where date within (dmin;dmax),
             sym2inst[sym]=inst,sym2type[sym]=`OT
       }[2017.10.01;2017.10.31;`CL]"

2.8131619113
```

Apply to multiple instruments

```
q)h2"{[dmin;dmax;inst]
     exec exp sum neg p*log p from
     update p:v % sum v from
     select v:sum siz by sym from trade where date within (dmin;dmax),
              sym2inst[sym]=inst,sym2type[sym]=`OT
          }[2017.10.01;2017.10.31] each `CL`ES"

2.8131619113 1.00890548575
```

Apply to all instruments

```
q)h2"`Neff xasc ([] inst:insts;
    Neff:{[dmin;dmax;inst]
       exec exp sum neg p*log p from update p:v % sum v from
        select v:sum siz by sym from trade where date within (dmin;dmax),
           sym2inst[sym]=inst,sym2type[sym]=`OT}[2017.10.01;2017.10.31]
    each insts)"

inst Neff
------------------
RTY  1.00034196781
6M   1.00452172163
ES   1.00890548575
YM   1.0097401893
NQ   1.01294970827
6J   1.02330012816
6B   1.02431960184
6A   1.03225770414
6C   1.03362941301
GC   1.05753871273
SI   1.06632280704
6E   1.07518727869
HG   1.2257078794
ZW   1.98039363339
ZC   2.31684118338
ZM   2.46968492734
ZL   2.47885402264
CL   2.8131619113
ZS   2.875605368
RB   3.51759025319
NG   3.6786605864
HO   3.68320053807
```

Most products have one active maturity

```
q)h1"`Neff xasc ([] inst:insts;
    Neff:{[dmin;dmax;inst]
       exec exp sum neg p*log p from update p:v % sum v from
        select v:sum siz by sym from trade where date within (dmin;dmax),
           sym2inst[sym]=inst,sym2type[sym]=`OT}[2017.10.01;2017.10.31]
    each insts)"
inst Neff
------------------
UB   1.00009914399
ZF   1.00019282326
ZN   1.00136113178
ZB   1.00166131796
ZT   1.0019205291
ZQ   11.422332455
GE   15.6192927728
```

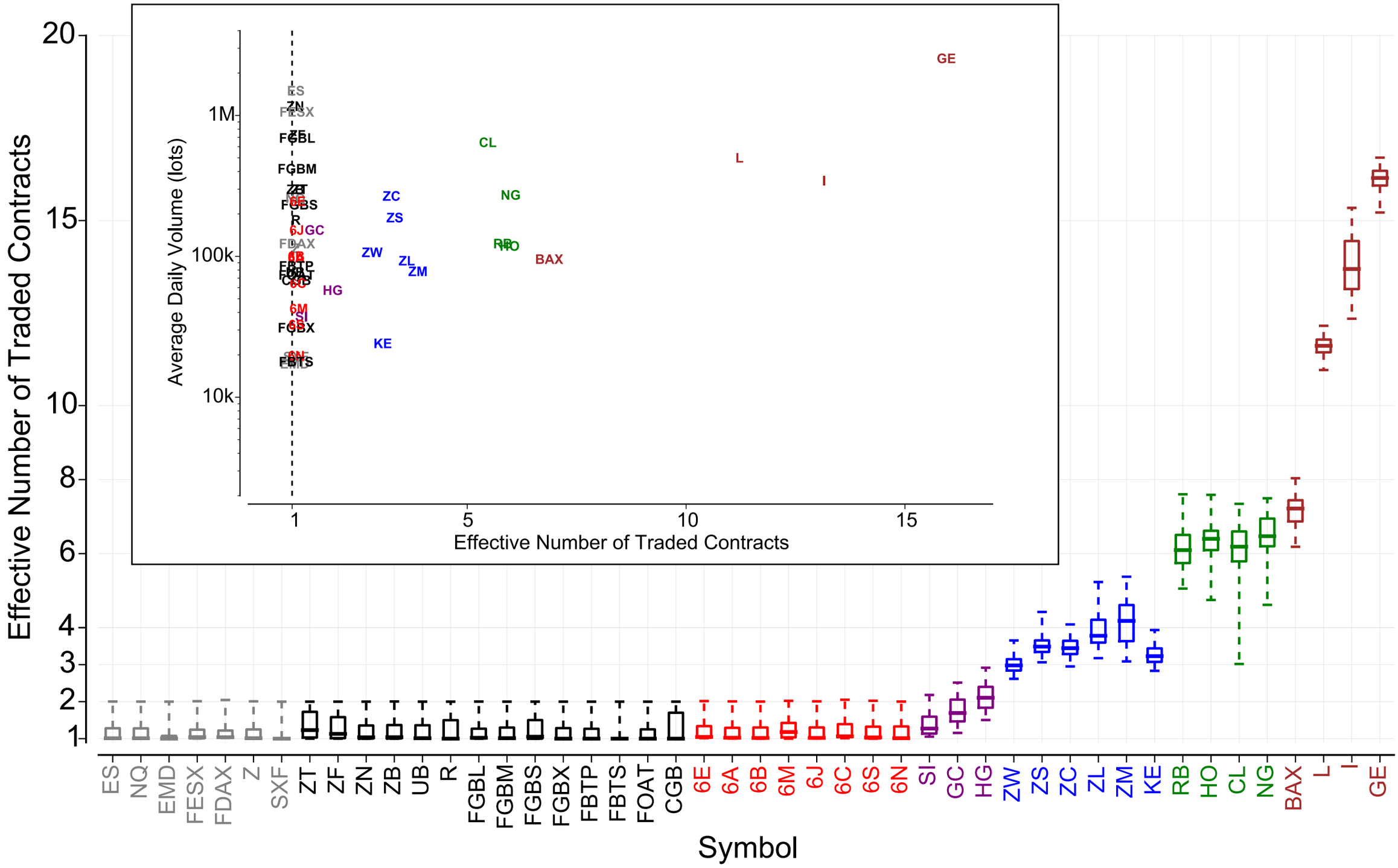
Most products have one active maturity

A few products (energy, agriculturals, Eurodollars) have multiple active maturities

Entropy and Futures Contracts

Isaac Carruthers

June 29, 2015



Tick size

- "Natural" bid-ask spread

  Δ_{natural}

  Bid-ask spread if prices were continuous

  Nonzero because of

    trading costs

    risk of informed traders (adverse selection)

    inventory risk

  Difficult to estimate directly

- Exchange-specified discrete price grid

  Δ_{exchange}

  Market-dependent

    US equities: 0.01 by Reg NMS

    futures: set by exchange for arbitrary reasons

  `minpxincr` in instinfo table

  Nonzero to aggregate liquidity at price points and for back office operations

Large-tick vs small-tick

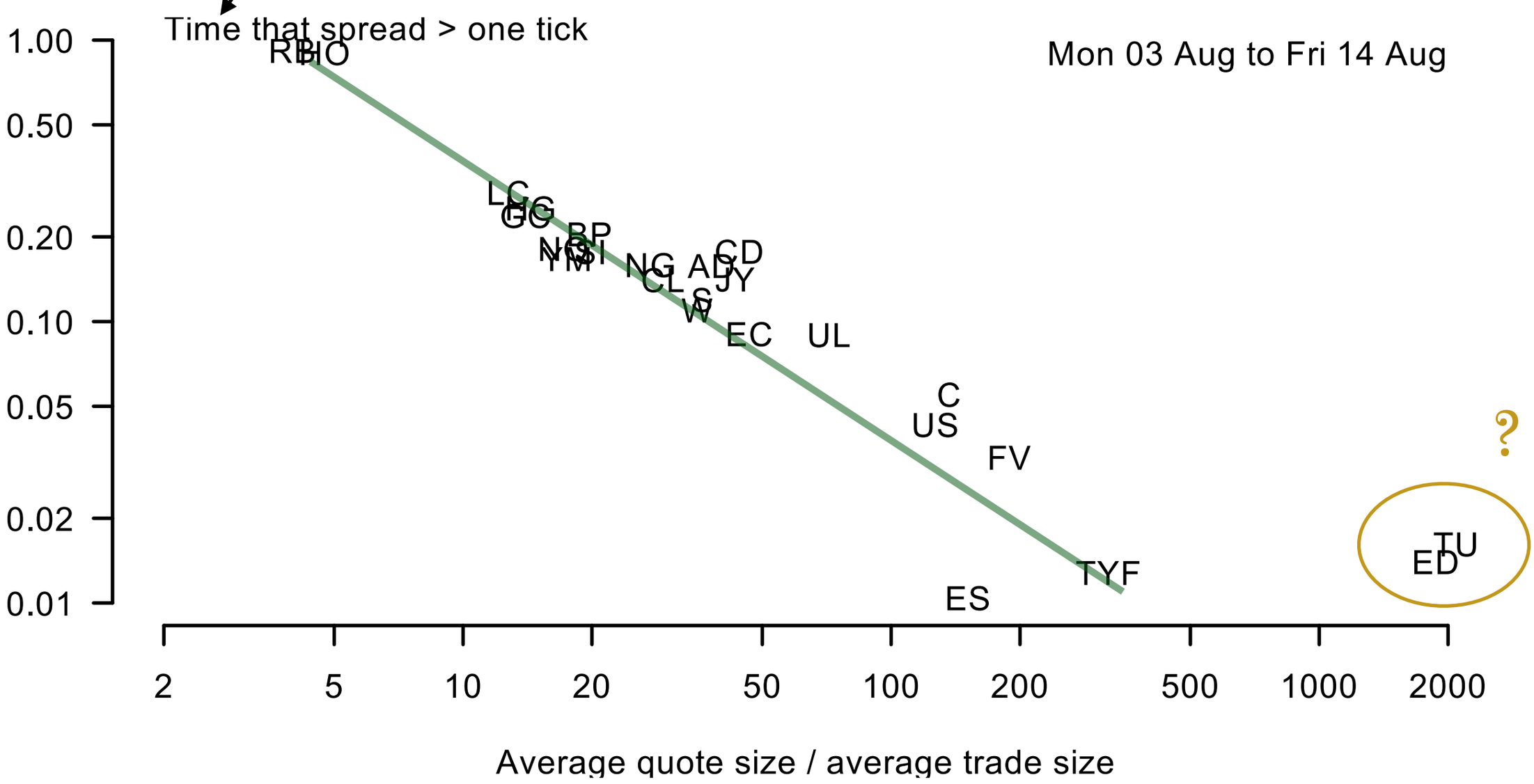
- Large-tick: $\Delta_{\text{exchange}} \gg \Delta_{\text{natural}}$
  large quote sizes
  lots of reversion
- Small-tick: $\Delta_{\text{exchange}} \ll \Delta_{\text{natural}}$
  small quote sizes
  little reversion
- "Ideal" $\Delta_{\text{exchange}} \sim \Delta_{\text{natural}}$
  prices move naturally

Tick size correlates with many other properties

Properties that correlate with tick size

- Average bid-ask spread
  (or fraction of "time" spread is larger than 1 tick)
  - > 1 tick: not large tick
  - = 1 tick: degrees of "large-tickness"
- Average quote size / average trade size
  - large when liquidity aggregates on price levels
  - also depends on match algorithm
- Reversion
  - trades: measure of trading costs (effective spread)
  - quote midpoint: different measure

"Time" = fraction of trades



Average spread, sampled by trade times

```
q)h2"select n:count i by sprd:ask-bid from
    {[d;s] aj[ `seq;
      select seq from trade where date=d,sym=s;
      select seq,bid,ask from quote where date=d,sym=s ] }
        [2017.10.25;`ESZ7]"
sprd| n
----| ------
0   | 2
25  | 152605
50  | 55

q)h2"exec avg ask-bid from
    {[d;s] aj[ `seq;
      select seq from trade where date=d,sym=s;
      select seq,bid,ask from quote where date=d,sym=s ] }
        [2017.10.25;`ESZ7]"

25.0086793046
```

Spread is almost always its min

"select" returns table
"exec" returns number

Use function on server, to avoid typing same arguments twice

ES has spread always minimum

```
q)h2"select n:count i by sprd:ask-bid from
    {[d;s] aj[ `seq;
      select seq from trade where date=d,sym=s;
      select seq,bid,ask from quote where date=d,sym=s ] }
        [2017.10.25;`RBZ7]"
sprd| n
----| ----
1   | 7453
2   | 8140
3   | 6668
4   | 3365
5   | 1291
6   | 544
7   | 178
8   | 74
9   | 34
10  | 18
11  | 16
12  | 4
13  | 2
20  | 3
23  | 20
```

```
q)h2"exec avg ask-bid from
    {[d;s] aj[ `seq;
      select seq from trade where date=d,sym=s;
      select seq,bid,ask from quote where date=d,sym=s ] }
        [2017.10.25;`RBZ7]"
2.51743976987
```

RB has spread generally larger than minimum