Project HighLO and ROOT

May 16th

PPP Meeting

Philippe Debie

Project HighLO – Key team

Collaboration: WUR + UM + CERN + CORMEC

HighLO = **High** Energy Physics Tools in **L**imit **O**rder Book Analysis

WUR/UM/CORMEC		CERN
Joost Pennings (WUR/UM)	Marjolein Verhulst (WU/WEcR)	Axel Naumann
Koos Gardebroek (WUR)	Philippe Debie (WU/WEcR)	Lorenzo Moneta
Bedir Tekinerdogan (WUR)	Tarek AlSkaif (WUR)	Jonas Rembser
Cagatay Catal (WUR)		Danilo Piparo
Andres Trujillo-Barrera (CORMEC)		Han Dols

Project HighLO – Background info

Current finance research lacks the tools

- Huge datasets
- Everything is statistically significant

HighLO

- Adapt more capable software tools (ROOT)
- A new perspective on how to analyze data
- Search and detect market manipulation

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

HighLO

Project HighLO – Background info

Limit

Order

Book

Side	Price	Volume
Ask Side (Selling)	15	5 + 3 + 2 = 10
	14	4 + 2 = 6
	13	3 + 1 = 4
Bid Side (Buying)	12	2 = 2
	11	2 + 5 = 7
	10	5 = 5

Project HighLO – Background info

Financial market is governed by supply and demand

- Limit order book = Summary of all limit orders
- More sell limit orders → Price goes down
- More buy limit orders → Price goes up

Spoofing = Placing and cancelling limit orders to push the market (with no intention of execution)

Project HighLO – Research goal

Detect spoofing in the commodity futures market

1. Describe how spoofing works
2. Detect spoofing
3. Help regulators and lawmakers

Spoofing = Placing and cancelling limit orders to push the market
(with no intention of execution)

Int. Expert Group on Market Surveillance (IMS)

The Netherlands:

- Exchange: Euronext
- Regulator: Authority for Consumers and Markets
- Regulator: Authority Financial Markets

UK:

- Exchange: ICE Futures Europe

Germany:

- Exchange: Deutsche Börse (FSE, Eurex)
- Exchange: EEX

Switzerland:

- Regulator: FINMA
- Exchange: SIX Group

Italy:

- Regulator: CONSOB

EU:

- Regulator: ACER
- Regulator: ESMA

USA:

- Exchange: CME Group
- Regulator: CFTC

Into the Microseconds

Paper submitted to Management Science

WAGENINGEN
UNIVERSITY & RESEARCH

CORMEC
COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

HighLO

Research Questions

1. How to measure relationships on the microsecond level?

2. Can we measure which market is leading?
   And which market is following?

→ This paper is not about market manipulation

Data

	Corn Futures	Wheat Futures	Soybean Futures
CME Globex code	ZC	ZW	ZS
Data range	2019/07/01 – 2020/07/01		
Number of messages	758 Million	723 Million	1605 Million
Raw data	± 100 GB	± 100 GB	± 200 GB
In ROOT	5.3 GB	4.9 GB	9.9 GB

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

HighLO

Event-Based Impact Profile – The Concept

Trigger

Impact value

Idea:

Measure the changes in a time series
at fixed intervals before or after an event

Example:

What is the average price change in the Wheat market
50 ms after a price change in the Corn market

WAGENINGEN UNIVERSITY & RESEARCH   CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE   CERN   HighLO

Event-Based Impact Profile – The Construction

1. Collect a set of triggers

   → Price changes in the Corn market (800 k)

2. For each trigger, extract a time series

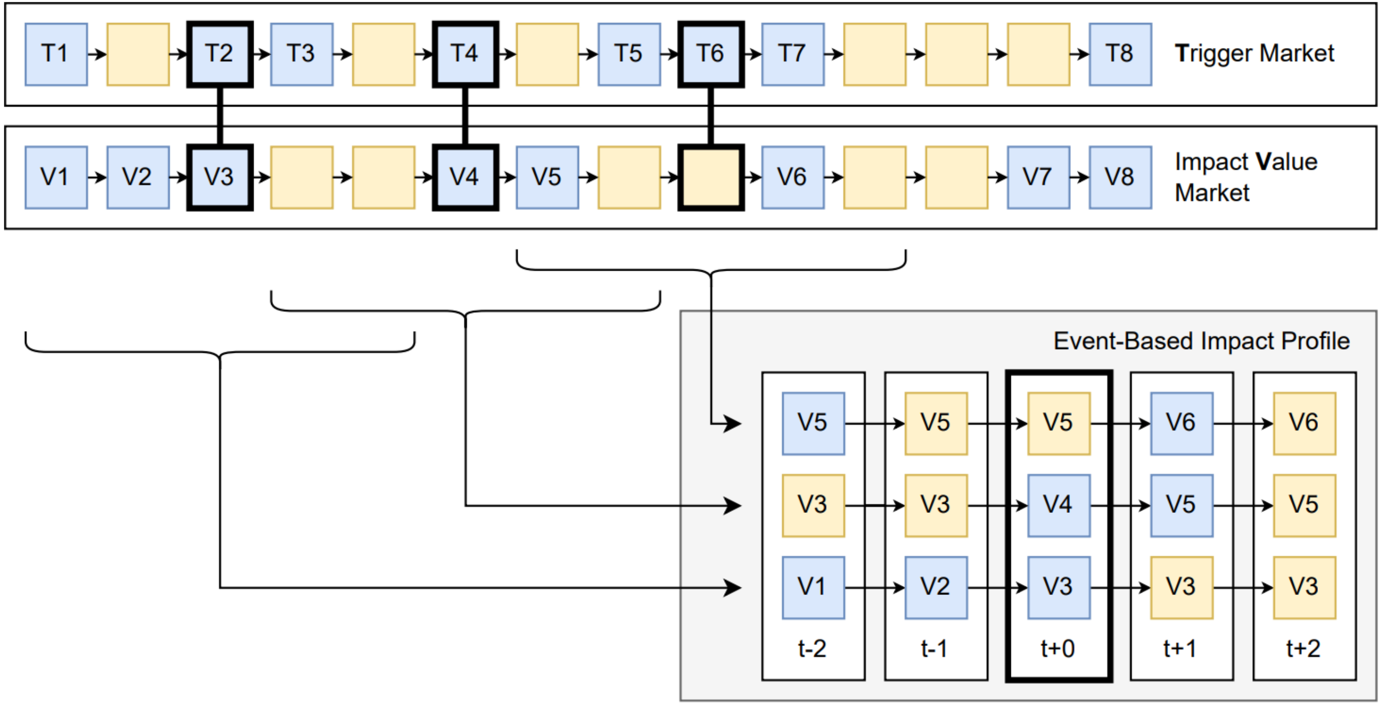
   → 200ms before to 200ms after the trigger

3. Overlap these time series (align the triggers)

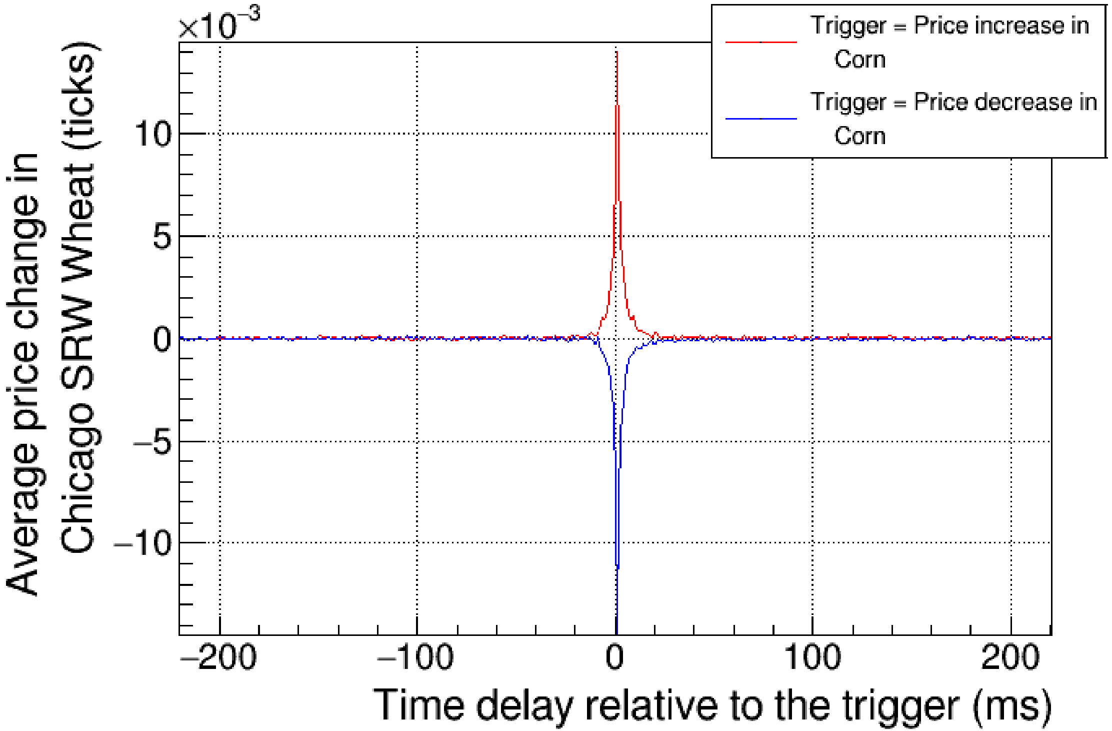
4. For each time delay, build a distribution

   → 401 distributions, sequence of distributions

Event-Based Impact Profile – Diagram

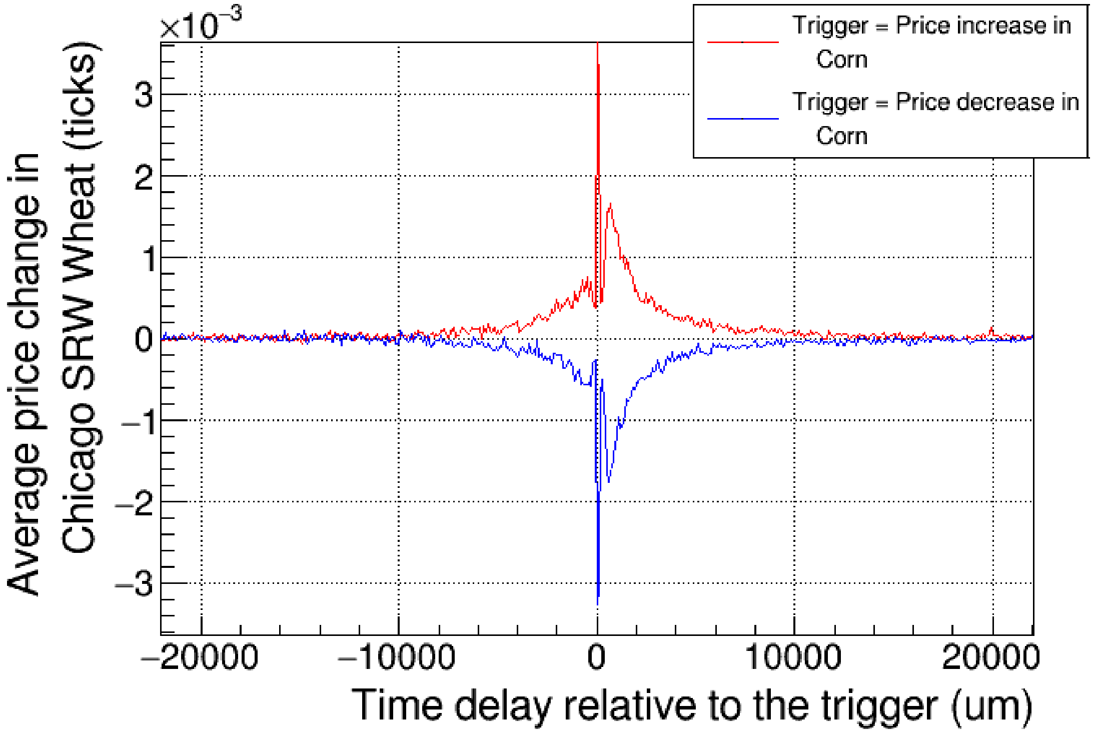


Trigger: Price change in Corn Futures
Impact: Price change in Wheat Futures
Resolution: 1 ms (x-axis spanning 440 ms)



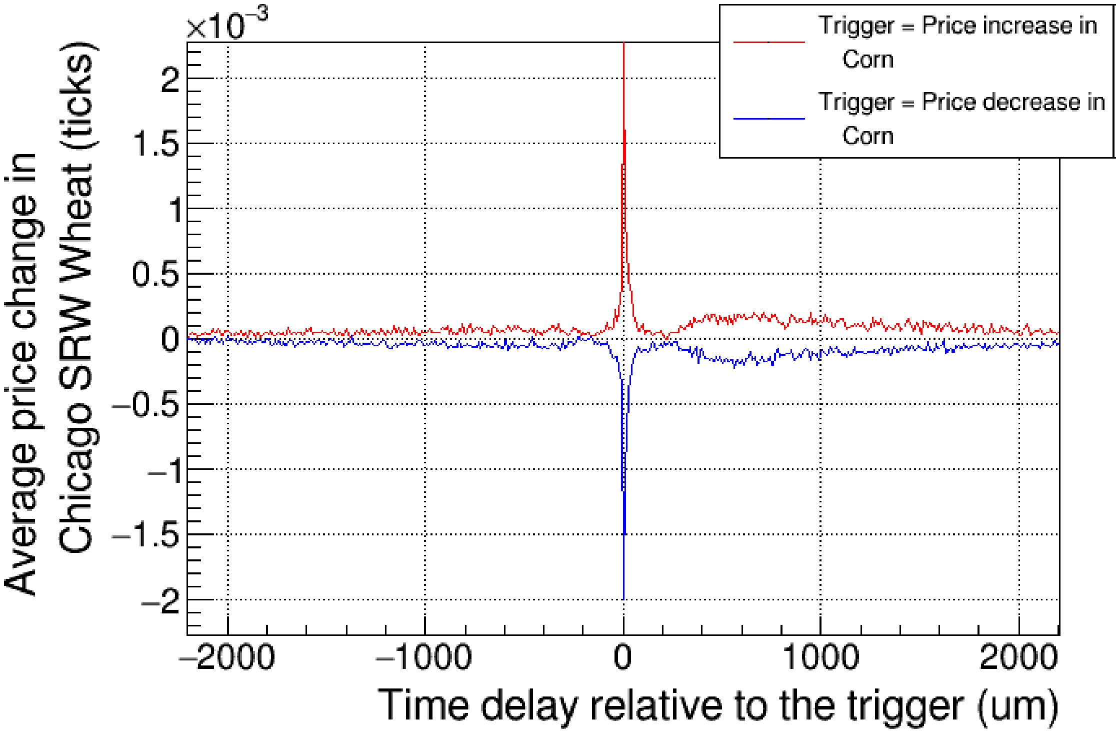
1 tick in Chicago SRW Wheat = 12.5 USD per contract (5000.0 BU)

Trigger: Price change in Corn Futures
Impact: Price change in Wheat Futures
Resolution: 100 µs (x-axis spanning 44 ms)



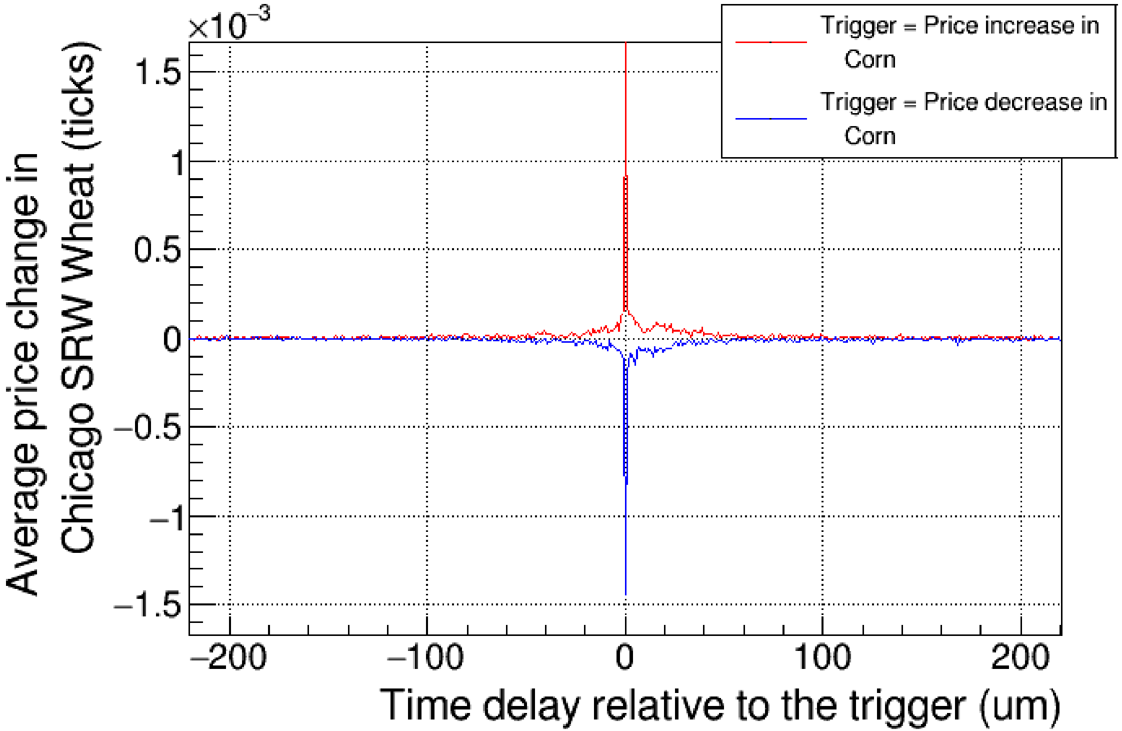
1 tick in Chicago SRW Wheat = 12.5 USD per contract (5000.0 BU)

Trigger: Price change in Corn Futures
Impact: Price change in Wheat Futures
Resolution: 10 µs (x-axis spanning 4.4 ms)



1 tick in Chicago SRW Wheat = 12.5 USD per contract (5000.0 BU)

Trigger: Price change in Corn Futures
Impact: Price change in Wheat Futures
Resolution: 1 µs (x-axis spanning 0.44 ms)



1 tick in Chicago SRW Wheat = 12.5 USD per contract (5000.0 BU)

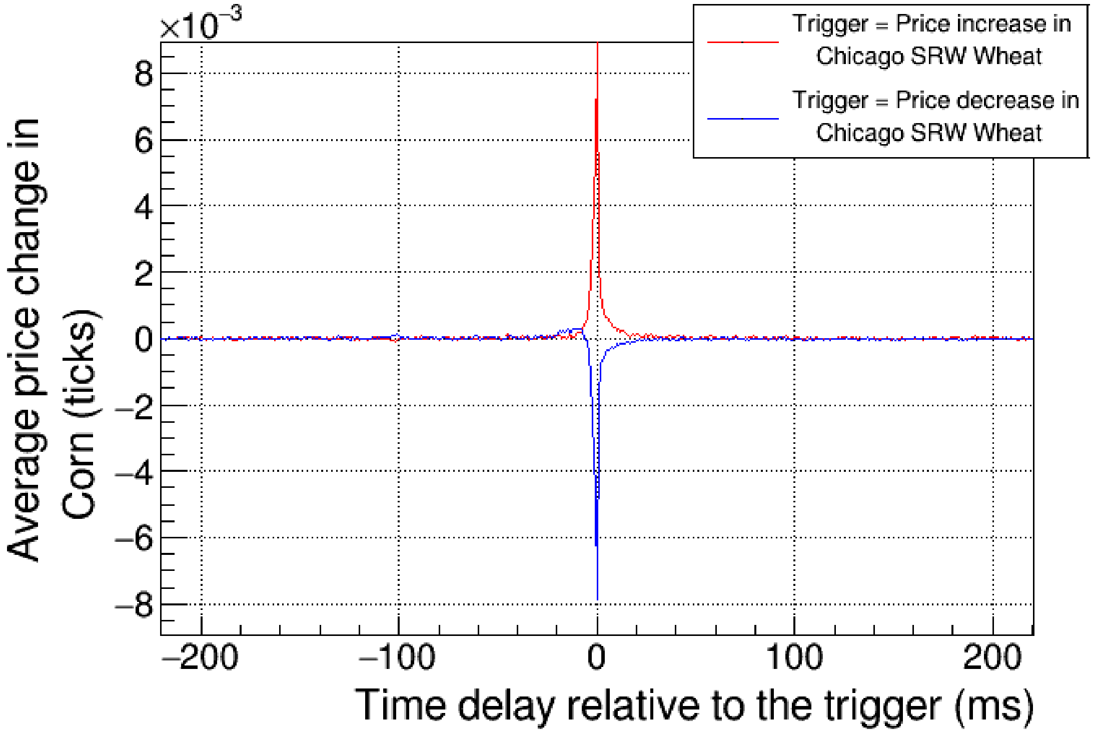
Summary 1: Impact of Corn on Wheat

Summary

1. Price increase occurs in 2 parts
    a. Instantaneously
    b. Starting 0.5 milliseconds after
2. Asymmetric, price increase in Wheat follows price increase in Corn

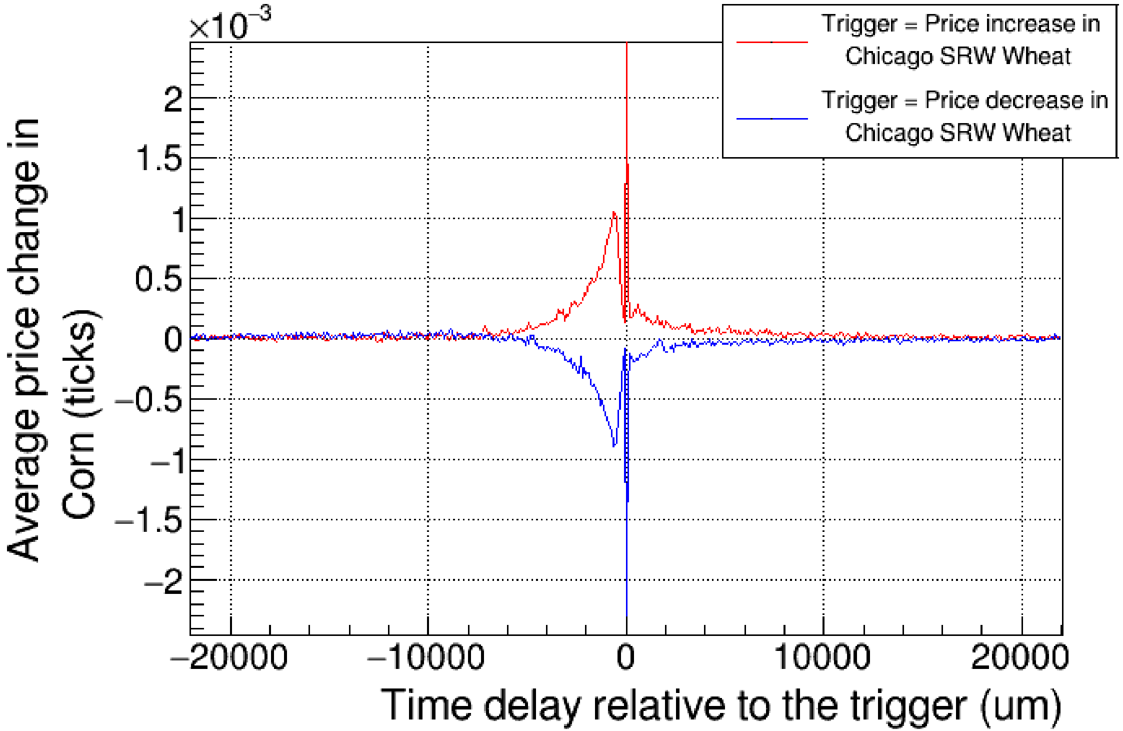
Next question: What if we inverse the trigger and impact contract?

Trigger: Price change in Wheat Futures
Impact: Price change in Corn Futures
Resolution: 1 ms (x-axis spanning 440 ms)



1 tick in Corn = 12.5 USD per contract (5000.0 BU)

Trigger: Price change in Wheat Futures
Impact: Price change in Corn Futures
Resolution: 100 µs (x-axis spanning 44 ms)



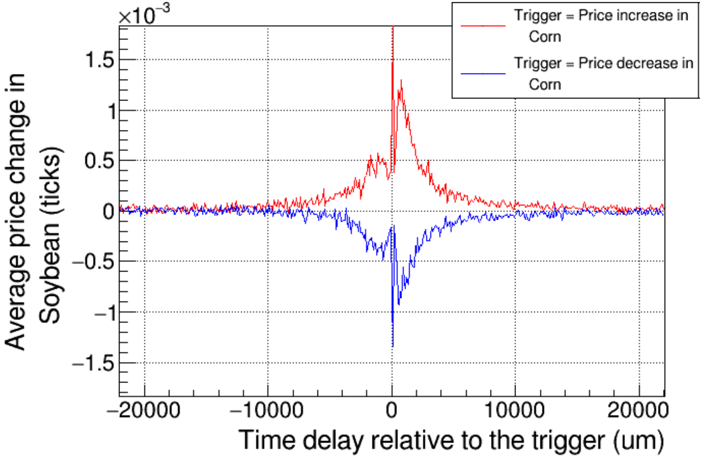
1 tick in Corn = 12.5 USD per contract (5000.0 BU)

Summary 2: Impact of Wheat on Corn

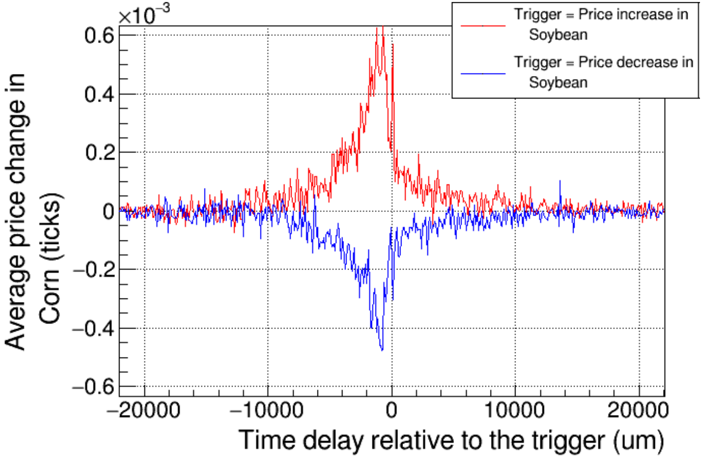
Summary

1. Price increase occurs in 2 parts
2. Asymmetric, price increase in Wheat follows price increase in Corn

Next question: How are relations between other markets?

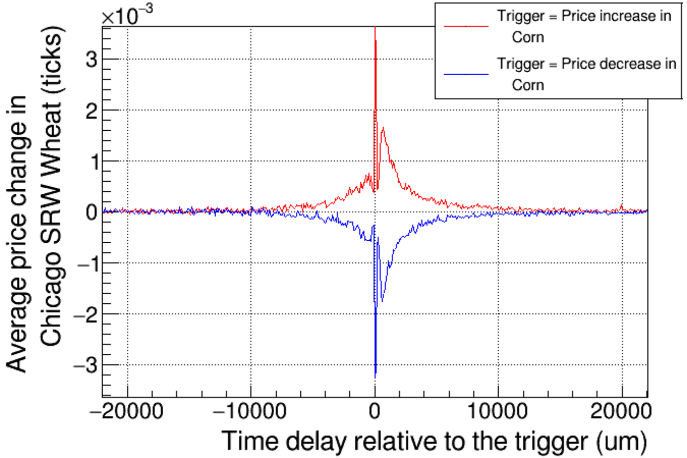


1 tick in Soybean = 12.5 USD per contract (5000.0 BU)

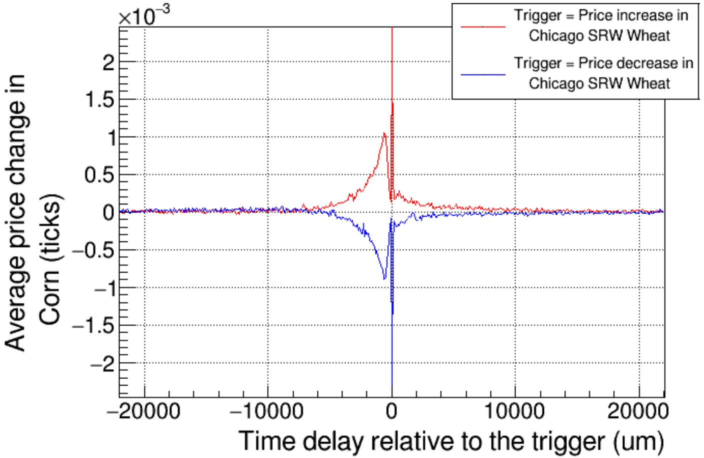


1 tick in Corn = 12.5 USD per contract (5000.0 BU)

Soybean follows Corn

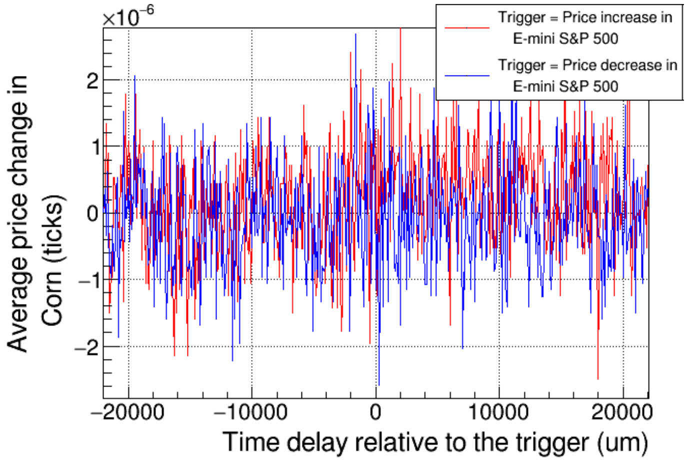


1 tick in Chicago SRW Wheat = 12.5 USD per contract (5000.0 BU)

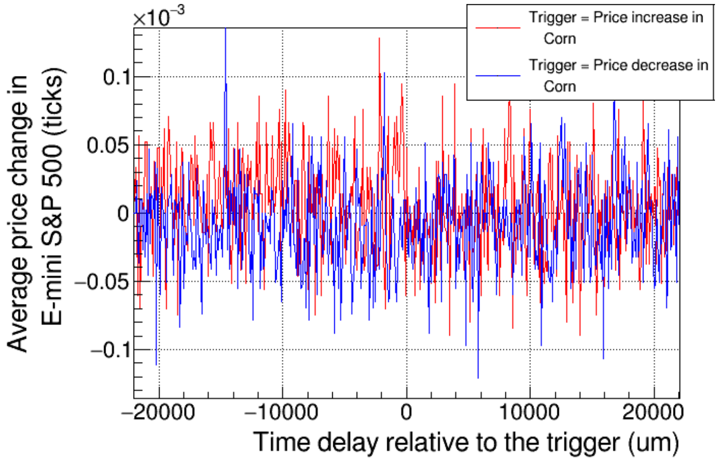


1 tick in Corn = 12.5 USD per contract (5000.0 BU)

Wheat follows Corn



1 tick in Corn = 12.5 USD per contract (5000.0 BU)



1 tick in E-mini S&P 500 = 12.5 USD per contract (50.0 IPNT)

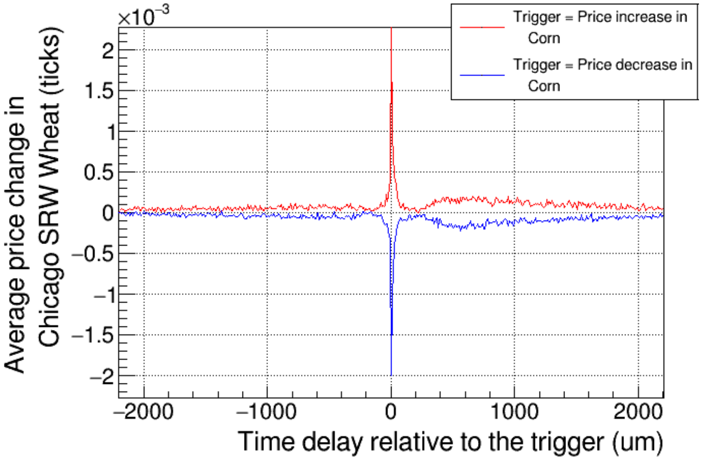
Conclusion: Research question

Observations:

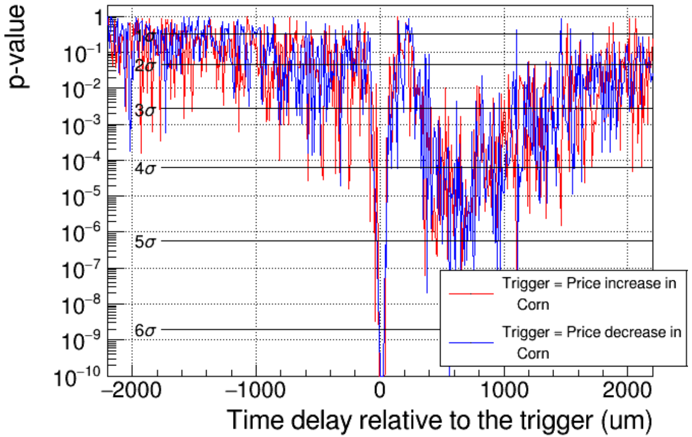
1. The market responds in 2 steps

2. **Price** changes in Wheat and Soybean follow
   Price changes in Corn

Trigger: Price change in Corn Futures
Impact: Price change in Wheat Futures
Resolution: 10 μs (x-axis spanning 4.4 ms)



1 tick in Chicago SRW Wheat = 12.5 USD per contract (5000.0 BU)



Conclusion: The methodology

Advantages compared to an impact analysis using VAR models

1. No data fitting or user-chosen modelling parameters
2. Measure past and future time correlations
3. Linear computational cost and trivial multi-threading
4. Not limited by the complexity of the model

→ Interdisciplinary collaboration led to new techniques

How does HighLO contribute to ROOT?

Project 1: Interactive Dashboard to Explore High-Dimensional Histograms

How does HighLO Contribute to ROOT?

Project 2: RDataframe with Time Series Support

New Operations for RDataframe

Slides from PPP 118

RDataframe operations

- Define using lead and lag (differentiation)
- Persistent data objects (integration and more)
- Resample a time series

Proof of concept: https://github.com/philippe554/root

Lead and Lag

```
ROOT::RDataFrame rdf(50);

auto r = rdf
   .DefineSlotEntry("foo", [](unsigned int slot, ULong64_t entry){return static_cast<int>(entry);})
   .Define("bar", [](int foo){return foo * foo;}, {"foo"})
   .MovingCache<int, int>({"foo", "bar"})
   .Define("D", [](int bar1, int bar2){return bar2 - bar1;}, {"bar", "bar"}, {-1, 0})
   .Display({"foo", "bar", "D"});

r->Print();
```

```
+-----+-----+-----+---+
| Row | foo | bar | D |
+-----+-----+-----+---+
| 1   | 1   | 1   | 1 |               → Note that it skipped the first entry
+-----+-----+-----+---+
| 2   | 2   | 4   | 3 |
+-----+-----+-----+---+
| 3   | 3   | 9   | 5 |
+-----+-----+-----+---+
| 4   | 4   | 16  | 7 |
+-----+-----+-----+---+
```

Persistent Define

```
ROOT::RDataFrame rdf(10);

auto r = rdf
   .DefineSlotEntry("foo", [](unsigned int slot, ULong64_t entry){return static_cast<int>(entry);})
   .Define("D", [](){return gRandom->Exp(1);})
   .DefinePersistent("time", [](double& time, double D){time += D;}, {"D"})
   .DefinePersistent("state", [](std::string& state, int foo){state = state + std::to_string(foo);}, {"foo"})
   .Display({"foo", "D", "time", "state"});

r->Print();
```

```
+-----+-----+----------------+----------------+--------------+
| Row | foo | D              | time           | state        |
+-----+-----+----------------+----------------+--------------+
| 0   | 0   | 0.00025828445 | 0.00025828445 | "0"          |
+-----+-----+----------------+----------------+--------------+
| 1   | 1   | 1.8145581      | 1.8148164      | "01"         |
+-----+-----+----------------+----------------+--------------+
| 2   | 2   | 1.2636598      | 3.0784762      | "012"        |
+-----+-----+----------------+----------------+--------------+
| 3   | 3   | 0.054243873    | 3.1327201      | "0123"       |
+-----+-----+----------------+----------------+--------------+
| 4   | 4   | 1.4624994      | 4.5952195      | "01234"      |
+-----+-----+----------------+----------------+--------------+
| 5   | 5   | 0.72366079     | 5.3188803      | "012345"     |
+-----+-----+----------------+----------------+--------------+
```

Resample a Time Series

Slides from PPP 118

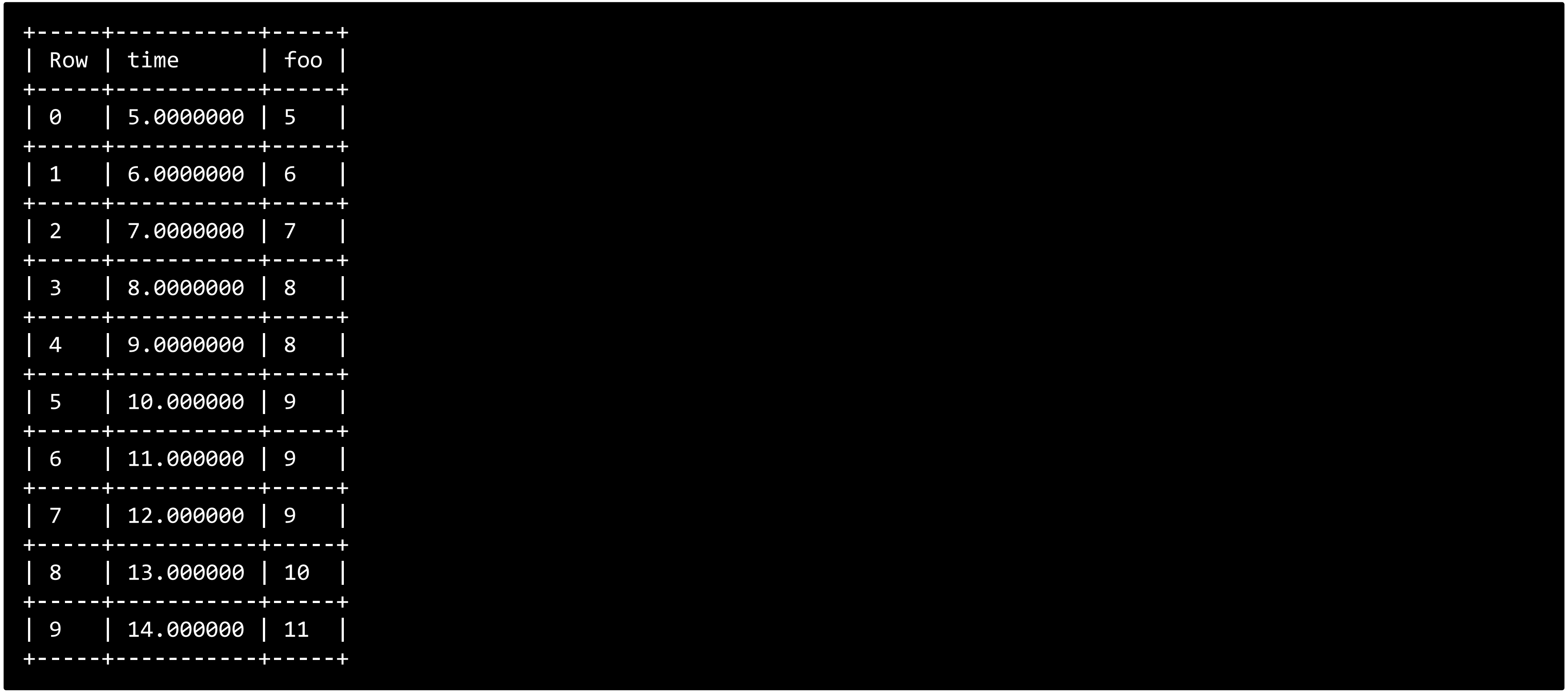
```cpp
ROOT::RDataFrame rdf(50);

auto r = rdf
   .DefineSlotEntry("foo", [](unsigned int slot, ULong64_t entry){return static_cast<int>(entry);})
   .Define("D", [](){return gRandom->Exp(1);})
   .DefinePersistent("time", [](double& time, double D){time += D;}, {"D"})
   .Resample<double, double, int>("time", 1, 5, 15, {"time", "foo"})
   .Display({"time", "foo"}, 10);

r->Print();
```

Resample a Time Series

Slides from PPP 118



```
+-----+-----------+-----+
| Row | time      | foo |
+-----+-----------+-----+
| 0   | 5.0000000 | 5   |
+-----+-----------+-----+
| 1   | 6.0000000 | 6   |
+-----+-----------+-----+
| 2   | 7.0000000 | 7   |
+-----+-----------+-----+
| 3   | 8.0000000 | 8   |
+-----+-----------+-----+
| 4   | 9.0000000 | 8   |
+-----+-----------+-----+
| 5   | 10.000000 | 9   |
+-----+-----------+-----+
| 6   | 11.000000 | 9   |
+-----+-----------+-----+
| 7   | 12.000000 | 9   |
+-----+-----------+-----+
| 8   | 13.000000 | 10  |
+-----+-----------+-----+
| 9   | 14.000000 | 11  |
+-----+-----------+-----+
```

Discussion