Detecting Edgeworth Cycles

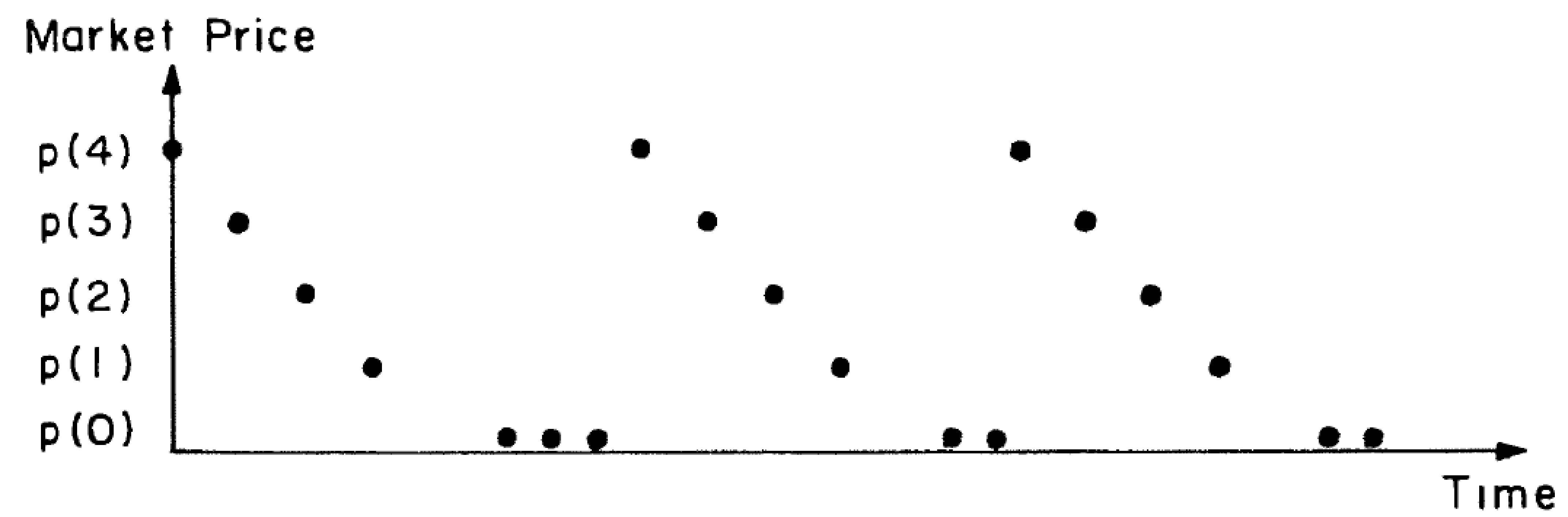
Timothy Holt (Università della Svizzera italiana)
Mitsuru Igami (Yale)
Simon Scheidegger (HEC Lausanne)

March 2024

Background

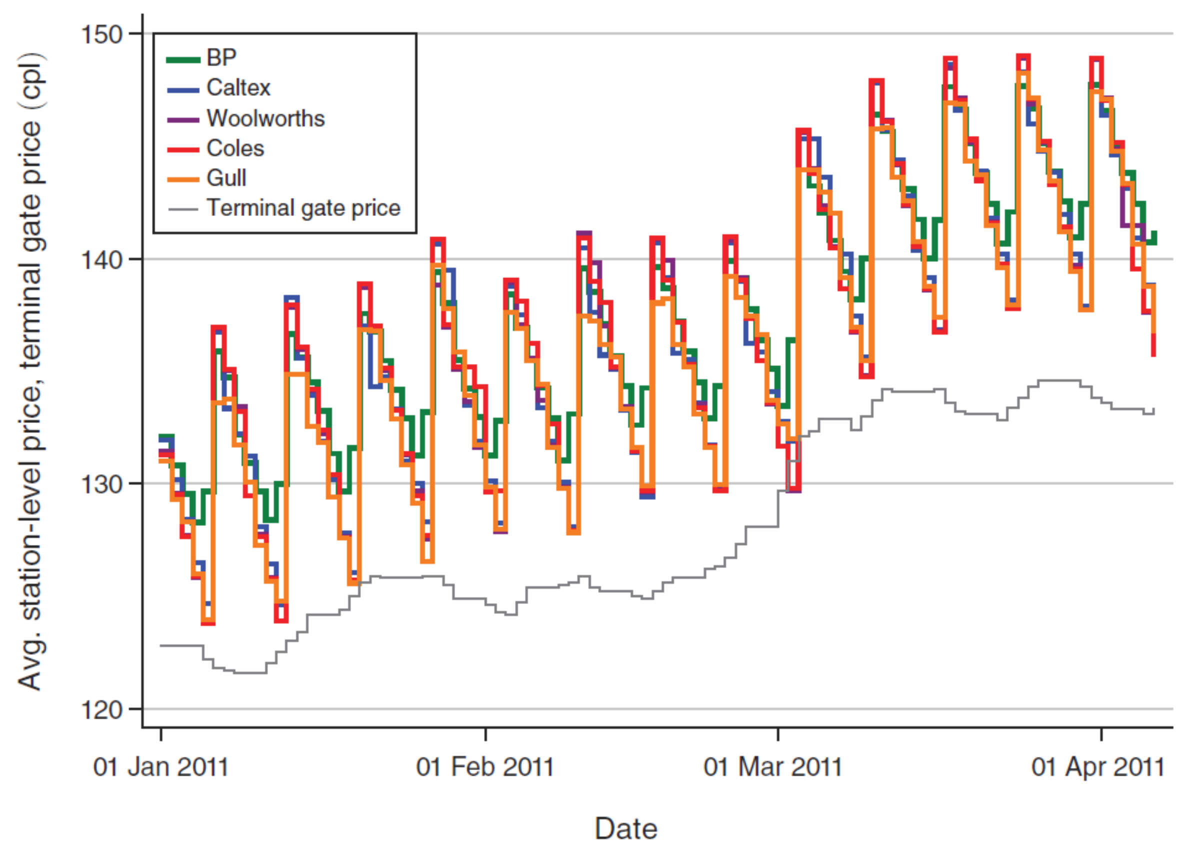
- My specialty: Empirical dynamic games
- Previous paper: Structural analysis of collusion ("Measuring the Incentive to Collude: The Vitamins Cartels, 1990–99,"2022 *REStud*, with Takuo Sugaya)
- Relatively easy to model explicit cartel agreements
- What about less explicit forms of cooperation, like price-leadership?

Edgeworth Cycles (1): Theory



Maskin & Tirole (1988 Econometrica)"A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles"

Edgeworth Cycles (2): Empirics



Byrne & de Roos (2019 AER)"Learning to Coordinate: A Study in Retail Gasoline"

Who Cares?

- Consumers, politicians, & government agencies: “Why are price increases bigger than decreases?”, “There must be some anti-competitive conspiracy”, “Let’s make real-time gas-price data publicly available”
- Price-fixing cases:

1. Clark & Houde (2014 JIE)“The Effect of Explicit Communication on Pricing: Evidence from the Collapse of a Gasoline Cartel”[Canada]
2. Foros & Steen (2013 Scandinavian J of Econ)“**Vertical Control and Price Cycles in Gasoline Retailing”**[Norway]
3. Wang (2008 RIO)“Collusive Communication and Pricing Coordination in a Retail Gasoline Market”[Australia]

Are Cycles Pro- or Anti-competitive?

- Mixed evidence on markup-cycle relationship:
- Positive correlation: Deltas (2008 JIE) [USA]; Clark & Houde (2014 JIE) [Canada]; Byrne (2019 RIO) [Australia]
- Negative correlation: Lewis (2009 J Law & Econ) [USA]; Zimmerman, Yun, & Taylor (2013 RIO) [USA]; Noel (2015 IJIO) [Canada]
- Potential reason 1: Intrinsic heterogeneity across countries & regions
- Potential reason 2: **Measurement/detection is non-trivial = THIS PAPER**

Why We Need *Good* Detection Methods

- Scalability: We just cannot eyeball all "big data."
- Reliability: Ad-hoc definitions could be inaccurate.
- Replicability: Systematic approach can be repeated/applied elsewhere.

More generally, this paper demonstrates…

- how economists can exploit "recent advances in machine learning"…
- …to solve pattern-recognition problems that actually matter.

Road Map

1. Theory
2. Four existing methods
3. Six new methods
4. Data & manual classification
5. Results (“horse race” + markups)

Co-authors in Switzerland

SCIENTIFIC ADVICE BY
SIMON SCHEIDEGGER (UNIL)

COMPUTATIONAL IMPLEMENTATION BY
TIMOTHY HOLT (USI)

Theory

Edgeworth (1925)

Maskin & Tirole (1988)

Theory & Measurement of Edgeworth Cycles

THEORETICALLY, EDGEWORTH CYCLES ARE CHARACTERIZED BY:

1. Cyclicality
2. Asymmetry
3. Stochasticity
4. Strategicness

EMPIRICALLY...

1. We propose methods to capture #1.
2. Existing papers focus on (& successfully capture) #2.
3. We do not use/require #3.
   - Both (stochastic & deterministic cycles) are documented.
   - Both raise antitrust concerns.
4. We do not use/require #4.
   - As long as price grid is fine, individual-station-level data should be sufficient to capture cycles.
   - Price grid must be fine in the Maskin-Tirole theory.
   - Price grid is fine in reality.

Four Existing Methods

1. Positive Runs vs. Negative Runs
2. Mean Increase vs. Mean Decrease
3. Negative Median Change
4. Many Big Price Increases

Method 1: Positive Runs vs. Negative Runs

Method 1: Positive Runs vs. Negative Runs ("PRNR"). Castanias and Johnson (1993) compare the lengths of positive and negative changes. We formalize this idea by classifying each station-quarter as cycling ($cycle_{i,t} = 1$) if and only if

$$mean\left(len\left(run^{+}\right)\right) < mean\left(len\left(run^{-}\right)\right) + \theta^{PRNR}, \qquad (1)$$

where $len\left(run^{+}\right)$ and $len\left(run^{-}\right)$ denote the lengths of consecutive (multi-day) price increases and decreases within quarter t, respectively. The means are taken over these "runs." $\theta^{PRNR} \approx 0$ is a scalar threshold, which we treat as a parameter.[14]

- ✓ Focus on asymmetry

Method 2: Mean Increase vs. Mean Decrease

Method 2: Mean Increase vs. Mean Decrease ("MIMD"). Eckert (2002) compares the magnitude of the mean increase and the mean decrease. Formally, station-quarter (i, t) is cycling if and only if

$$\left|mean_{d \in t}\left(\Delta p^{+}_{i,d}\right)\right| > \left|mean_{d \in t}\left(\Delta p^{-}_{i,d}\right)\right| + \theta^{MIMD}, \tag{2}$$

where $\Delta p^{+}_{i,d}$ and $\Delta p^{-}_{i,d}$ denote positive and negative daily price changes at station i (between days d and $d-1$), respectively, and $\theta^{MIMD} \approx 0$ is a scalar threshold. That is, a cycle is detected when the average price increase is greater than the average price decrease.

- ✓ Also focus on asymmetry

Method 3: Negative Median Change

Method 3: Negative Median Change ("NMC"). Lewis (2009) and many subsequent papers classify $cycle_{i,t} = 1$ if and only if

$$median_{d \in t} \left(\Delta p_{i,d} \right) < \theta^{NMC}, \tag{3}$$

where $\Delta p_{i,d}$ denotes price change between days d and $d-1$, and $\theta^{NMC} \approx 0$ is a scalar threshold. In other words, the significantly negative median change is taken as evidence of price cycles.

- ✓ Yet another way to measure asymmetry

Method 4: Many Big Price Increases

Method 4: Many Big Price Increases ("MBPI"). Byrne and de Roos (2019) identify price cycles with the condition

$$\sum_{d \in t} \mathbb{I}\left\{\Delta p_{i,d} > \theta_1^{MBPI}\right\} \geq \theta_2^{MBPI}, \tag{4}$$

where $\mathbb{I}\{\cdot\}$ is an indicator function that equals one if the condition inside the bracket is satisfied and zero otherwise. θ_1^{MBPI} and θ_2^{MBPI} are thresholds for "big" and "many" price increases, respectively. They set $\theta_1^{MBPI} = 6$ (Australian cents/liter) and $\theta_2^{MBPI} = 3.75$ (per quarter) in studying the WA data. Thus, many instances of big price increases are taken as evidence of price cycles.

✓ Captures not only asymmetry but both amplitude & frequency of cycles!

Four Existing Methods: Summary

- Essentially, simple "threshold models" with 1 or 2 parameter
- All focus on "asymmetry" but not "cyclicality."

Six New Methods

5. Fourier transform
6. Lomb-Scargle periodogram
7. Cubic splines
8. Long Short-Term Memory (LSTM)
9. Ensemble in random forests
10. Ensemble in LSTM

Method 5: Fourier Transform

Method 5: Fourier Transform ("FT"). Fourier analysis is a mathematical method for detecting and characterizing periodicity in time-series data. When a continuous function of time $g(x)$ is sampled at regular time intervals with spacing Δx, the sample analog of the Fourier power spectrum (or "periodogram") is

$$P(f) \equiv \frac{1}{N} \left| \sum_{n=1}^{N} g_n e^{-2\pi i f x_n} \right|^2, \tag{5}$$

where f is frequency, N is the sample size, $g_n \equiv g(n\Delta x)$, $i \equiv \sqrt{-1}$ is the imaginary unit (not to be confused with our gas-station index), and x_n is the time stamp of n-th observation. It is a positive, real-valued function that quantifies the contribution of each frequency f to the time-series data $(g_n)_{n=1}^{N}$.[15]

Method 5: Fourier Transform (cont.)

We focus on the highest peak of $P(f)$ and detect cycles if and only if

$$\max_f P_{i,t}(f) > \theta_{\max}^{FT}, \tag{6}$$

where $P_{i,t}(f)$ is the periodogram (5) of station-quarter (i,t), and $\theta_{\max}^{FT} > 0$ is a scalar threshold parameter.

- ✓ Suitable for regular cycles with deterministic frequency

Method 6: Lomb-Scargle Periodogram

Method 6: Lomb-Scargle Periodogram ("LS"). The Lomb-Scargle periodogram (Lomb 1976, Scargle 1982) characterizes periodicity in unevenly sampled time-series.[16] It has been extensively used in astrophysics because astronomical observations are subject to weather conditions and diurnal, lunar, or seasonal cycles. Formally, it is a generalized version of the classical periodogram (5):[17]

$$P^{LS}(f) = \frac{1}{2}\left\{ \frac{\left(\sum_n g_n \cos\left(2\pi f\left[x_n - \tau\right]\right)\right)^2}{\sum_n \cos^2\left(2\pi f\left[x_n - \tau\right]\right)} + \frac{\left(\sum_n g_n \sin\left(2\pi f\left[x_n - \tau\right]\right)\right)^2}{\sum_n \sin^2\left(2\pi f\left[x_n - \tau\right]\right)} \right\}, \tag{7}$$

where τ is specified for each frequency f as

$$\tau = \frac{1}{4\pi f}\tan^{-1}\left(\frac{\sum_n \sin\left(4\pi f x_n\right)}{\sum_n \cos\left(4\pi f x_n\right)}\right). \tag{8}$$

Method 6: Lomb-Scargle Periodogram (cont.)

We propose the following threshold condition to detect cycles:

$$\max_f P_{i,t}^{LS}(f) > \theta_{\mathbf{max}}^{LS}, \tag{9}$$

where $\theta_{\mathbf{max}}^{LS} > 0$ is a scalar threshold parameter.

- ✓ Like FT (Method 5), good for regular, deterministic cycles

Method 7: Cubic Splines

Method 7: Cubic Splines ("CS"). This method captures cycles' frequency in a less structured manner than FT and LS by using cubic splines (CS). A spline is a piecewise polynomial function. We smooth the discrete (daily) time-series by interpolating it with a cubic Hermite interpolater, which is a spline where each piece is a third-degree polynomial of Hermite form.[18] For each (i,t), we fit CS to its demeaned price series , $\overline{p}_{i,d} \equiv p_{i,d} - mean_{d \in t}\left(p_{i,d}\right)$, and count the number of times the fitted function $\overline{CS}_{i,t}(d)$ crosses the d-axis (i.e., equals zero). That is, we can count the number of real roots and detect cycles with the condition,

$$\#roots\left(\overline{CS}_{i,t}(d)\right) > \theta^{CS}_{root}, \tag{10}$$

where $\theta^{CS}_{root} > 0$ is a scalar parameter. Thus, we take any frequent oscillations (not limited to the sinusoidal ones as in FT or LS) as a sign of cycles.

✓ More flexible than spectral methods 5–6; good for irregular, stochastic cycles

Method 8: Long Short-Term Memory

- Recurrent neural networks (RNN) with LSTM
- De-facto "industry standard" for recognizing, modeling, & predicting speech, handwriting, language, polyphonic music, etc.
- Econometrically, a class of flexible, nonparametric models for time-series data

Method 8: LSTM is a recursive dynamic model whose behavior centers on (pairs of) two state variables:

$$\mathbf{s}_d^l = \underbrace{\tanh\left(\mathbf{c}_d^l\right)}_{\text{"output"}} \circ \underbrace{\Lambda\left(\boldsymbol{\omega}_1^l + \boldsymbol{\omega}_2^l \Delta p_d + \boldsymbol{\omega}_3^l \mathbf{s}_d^{l-1}\right)}_{\text{"output gate"}}, \text{ and} \tag{11}$$

$$\begin{aligned}\mathbf{c}_d^l = {} & \underbrace{\tanh\left(\boldsymbol{\omega}_4^l + \boldsymbol{\omega}_5^l \Delta p_d + \boldsymbol{\omega}_6^l \mathbf{s}_d^{l-1}\right)}_{\text{"input"}} \circ \underbrace{\Lambda\left(\boldsymbol{\omega}_7^l + \boldsymbol{\omega}_8^l \Delta p_d + \boldsymbol{\omega}_9^l \mathbf{s}_d^{l-1}\right)}_{\text{"input gate"}} \\ & + \mathbf{c}_d^{l-1} \circ \underbrace{\left[1 - \Lambda\left(\boldsymbol{\omega}_7^l + \boldsymbol{\omega}_8^l \Delta p_d + \boldsymbol{\omega}_9^l \mathbf{s}_d^{l-1}\right)\right]}_{\text{"forget gate"}},\end{aligned} \tag{12}$$

where $d = 1, 2, \cdots, D$ is our index of days, $\Delta p_d \equiv p_d - p_{d-1}$ (we set $\Delta p_1 = 0$), $\tanh(x) \equiv \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is the hyperbolic tangent function, $\circ$ denotes the Hadamard (element-wise) product, and $\Lambda(x) \equiv \frac{e^x}{1+e^x}$ is the cumulative density function (CDF) of the logistic distribution.[20] The $\boldsymbol{\omega}$s are weight parameters with the following dimensionality: (i) $\boldsymbol{\omega}_1^l$, $\boldsymbol{\omega}_2^l$, $\boldsymbol{\omega}_4^l$, $\boldsymbol{\omega}_5^l$, $\boldsymbol{\omega}_7^l$, and $\boldsymbol{\omega}_8^l$ are $B_l \times 1$ vectors; (ii) $\boldsymbol{\omega}_3^l$, $\boldsymbol{\omega}_6^l$, and $\boldsymbol{\omega}_9^l$ are $B_l \times B_l$ matrices. Thus, $\mathbf{B} \equiv (B_1, B_2, \cdots, B_L)$ determines the effective number of latent state variables and parameters, and hence the flexibility of the model.

Method 8: LSTM (cont.)

$$s^*\left(\mathbf{p}_{i,t};\boldsymbol{\theta}^{LSTM}\right)\equiv\omega_{10}+\omega_{11}s_D^L>0, \tag{13}$$

where $\boldsymbol{\theta}^{LSTM}\equiv(\boldsymbol{\omega},L,\mathbf{B})$ collectively denotes all parameters, including (i) the many weights in $\boldsymbol{\omega}\equiv\left(\left(\boldsymbol{\omega}_1^l,\boldsymbol{\omega}_2^l,\cdots,\boldsymbol{\omega}_9^l\right)_{l=1}^L,\omega_{10},\omega_{11}\right)$, (ii) the number of layers L, and (iii) the profile of the number of blocks in each layer, $\mathbf{B}$. We set $L=3$ and $\mathbf{B}=(16,8,4)$, and find $\boldsymbol{\omega}$ that approximately maximizes the accuracy of prediction (to be explained in section 3.3).[22] In summary, this method sequentially processes the daily price data in a nonparametric Markov model, and uses the terminal state s^* as a latent score to detect cycles.

- How flexible? Number of weight parameters (**ω**) = 2,165
- The most flexible of all stand-alone methods 1–8

Method 9: Ensemble in Random Forests

Method 9: Ensemble in Random Forests ("E-RF"). This method combines Methods 1–7 within random forests, a class of nonparametric regressions. Let

$$g_{i,t}^{m} \equiv LHS_{i,t}^{m} - RHS_{i,t}^{m} \tag{14}$$

denote a "gap," the scalar difference between the left-hand side (LHS) and the right-hand side (RHS) of the inequality that defines each method $m = 1, 2, \cdots, M$, excluding the threshold parameter, $\boldsymbol{\theta}^{m}$. For example, inequality (2) defines Method 2. Hence, $g_{i,t}^{2} = \left|mean_{d \in t}\left(\Delta p_{i,d}^{+}\right)\right| - \left|mean_{d \in t}\left(\Delta p_{i,d}^{-}\right)\right|$.[23] Let

$$\mathbf{g}_{i,t} \equiv \left(g_{i,t}^{m}\right)_{m=1}^{M} \tag{15}$$

denote their vector, where $M = 7$.[24] We construct a decision-trees classification algorithm

✓ Flexible aggregator that gets more information out of Methods 1–7

Method 9: Ensemble in Random Forests (cont.)

denote their vector, where $M = 7$.[24] We construct a decision-trees classification algorithm that takes $\mathbf{g}_{i,t}$ as inputs and predicts $cycle_{i,t} = 1$ if and only if

$$h\left(\mathbf{g}_{i,t}; \boldsymbol{\omega}^{RF}, \boldsymbol{\kappa}^{RF}\right) \equiv \sum_{k=1}^{K} \omega_k^{RF} \mathbb{I}\left\{\mathbf{g}_{i,t} \in R_k\right\} \equiv \sum_{k=1}^{K} \omega_k^{RF} \phi\left(\mathbf{g}_{i,t}; \boldsymbol{\kappa}_k^{RF}\right) > 0, \tag{16}$$

where K is the number of adaptive basis functions, ω_k^{RF} is the weight of the k-th basis function, R_k is the k-th region in the M-dimensional space of $\mathbf{g}_{i,t}$, and $\boldsymbol{\kappa}_k^{RF}$ encodes both the choice of variables (elements of $\mathbf{g}_{i,t}$) and their threshold values that determine region R_k.[25] Because finding the truly optimal partitioning is computationally infeasible, we use random forests algorithm to stochastically approximate it.[26] Thus, this method aggregates the seven threshold models in a flexible manner that permits interactions between $g_{i,t}^m$s. $\boldsymbol{\theta}^{RF} \equiv \left(\boldsymbol{\omega}^{RF}, \boldsymbol{\kappa}^{RF}\right) \equiv \left(\left(\omega_k^{RF}\right)_{k=1}^{K}, \left(\boldsymbol{\kappa}_k^{RF}\right)_{k=1}^{K}\right)$ is the full set of parameters.

Method 10: Ensemble in LSTM

Method 10: Ensemble in LSTM ("E-LSTM"). This method combines Methods 1–8 within an extended LSTM by incorporating $\mathbf{g}_{i,t}$ in (15) as additional variables in the laws of motion:

$$\mathbf{s}_d^l = \tanh\left(\mathbf{c}_d^l\right) \circ \Lambda\left(\boldsymbol{\omega}_1^l + \boldsymbol{\omega}_2^l \Delta p_d + \boldsymbol{\omega}_3^l \mathbf{s}_d^{l-1} + \boldsymbol{\omega}_{12}^l \mathbf{g}\right), \text{ and} \quad (17)$$

$$\begin{aligned} \mathbf{c}_d^l = & \tanh\left(\boldsymbol{\omega}_4^l + \boldsymbol{\omega}_5^l \Delta p_d + \boldsymbol{\omega}_6^l \mathbf{s}_d^{l-1} + \boldsymbol{\omega}_{13}^l \mathbf{g}\right) \circ \Lambda\left(\boldsymbol{\omega}_7^l + \boldsymbol{\omega}_8^l \Delta p_d + \boldsymbol{\omega}_9^l \mathbf{s}_d^{l-1} + \boldsymbol{\omega}_{14}^l \mathbf{g}\right) \\ & + \mathbf{c}_d^{l-1} \circ \left[1 - \Lambda\left(\boldsymbol{\omega}_7^l + \boldsymbol{\omega}_8^l \Delta p_d + \boldsymbol{\omega}_9^l \mathbf{s}_d^{l-1} + \boldsymbol{\omega}_{14}^l \mathbf{g}\right)\right], \end{aligned} \quad (18)$$

where $\left(\boldsymbol{\omega}_{12}^l, \boldsymbol{\omega}_{13}^l, \boldsymbol{\omega}_{14}^l\right)$ are the additional weight parameters for $\mathbf{g}_{i,t}$ (we suppress (i,t) subscript here). Unlike p_d, which varies across $D = 90$ days, $\mathbf{g}$ is constant for all d and l within (i,t). The other implementation details are the same as Method 8.

✓ Super-flexible aggregator that gets more out of Methods 1–8

Summary of 10 Methods

EXISTING METHODS

1. Positive Runs vs. Negative Runs (PRNR)
2. Mean Increase vs. Mean Decrease (MIMD)
3. Negative Median Change (NMC)
4. Many Big Price Increases (MBPI)

NEW METHODS

5. Fourier Transform (FT)
6. Lomb-Scargle Periodogram (LS)
7. Cubic Splines (CS)
8. Long Short-Term Memory (LSTM)
9. Ensemble in Random Forests (E-RF): Methods 1–7
10. Ensemble in LSTM (E-LSTM): Methods 1–8

Optimization of Parameter Values: Maximize Accuracy

$$\% \text{ correct}(\boldsymbol{\theta}) \equiv \frac{\sum_{(i,t)} \mathbb{I}\left\{\widehat{cycle}_{i,t}(\boldsymbol{\theta}) = cycle_{i,t}\right\}}{\# \textit{ all predictions}} \times 100, \tag{19}$$

where $\widehat{cycle}_{i,t}(\boldsymbol{\theta}) \in \{0,1\}$ is the algorithmic prediction for observation (i,t) at parameter value $\boldsymbol{\theta}$, and $cycle_{i,t} \in \{0,1\}$ is the manual classification label (data). We analogously define two types of prediction errors, "false negative" and "false positive."[27] Thus,

$$\boldsymbol{\theta}^* \equiv \arg\max_{\boldsymbol{\theta}} \; \% \text{ correct}(\boldsymbol{\theta}). \tag{20}$$

characterizes the optimized (or "trained") model for each method.[28]

Data & Manual Classification

Training humans before training machines

Summary Statistics

Dataset	(1) Western Australia	(2) New South Wales	(3) Germany
Sample period (yyyy/mm/dd)	2001/1/3 – 2020/6/30	2016/8/1 – 2020/7/31	2014/6/8 – 2020/1/7
Number of gasoline stations	821	1,226	14,780
Number of calendar quarters	77	15	26
Number of station-quarters	25,463	9,693	353,086
Of which:			
Labeled as "cycling" by 3 RAs	0 (0.0%)	6,878 (71.0%)	14,116 (39.6%)
Labeled as "cycling" by 2 RAs	0 (0.0%)	906 (9.4%)	7,173 (20.1%)
Labeled as "cycling" by 1 RA	15,007 (61.1%)	759 (7.8%)	6,280 (17.6%)
Not labeled as "cycling" by any RA	9,562 (38.9%)	1,150 (11.9%)	8,116 (22.7%)
Total manually labeled	24,569 (100.0%)	9,693 (100.0%)	35,685 (100.0%)
Not manually labeled	894	0	317,401

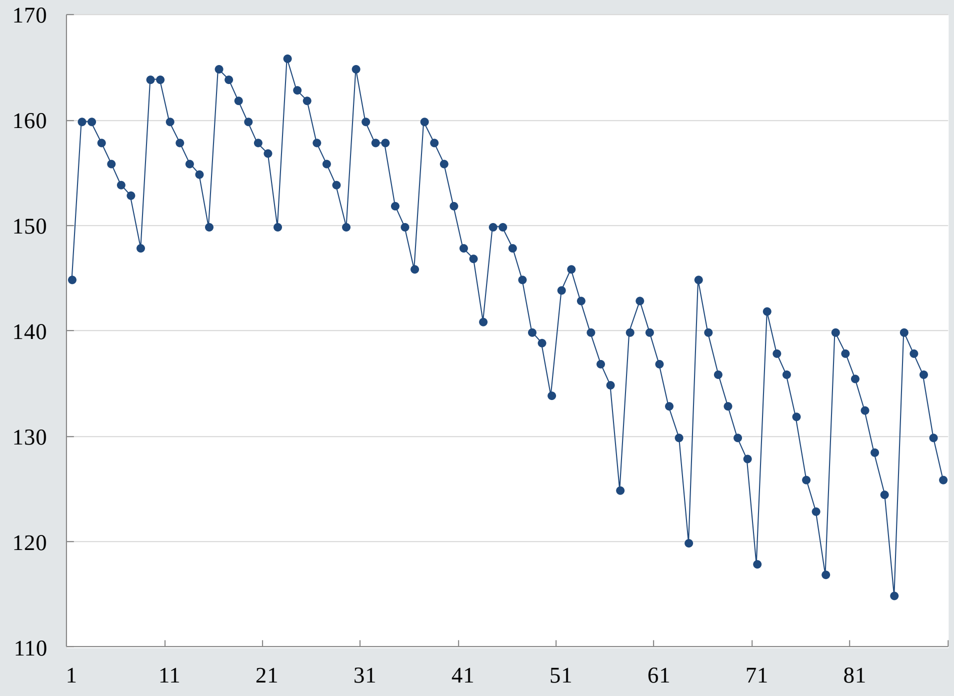
Note: Each "manually labeled" station-quarter observation in the WA data is single-labeled as either "cycling," "maybe cycling," or "not cycling," whereas the NSW and German data are triple-labeled. See Appendix A for details.

Manual Classification

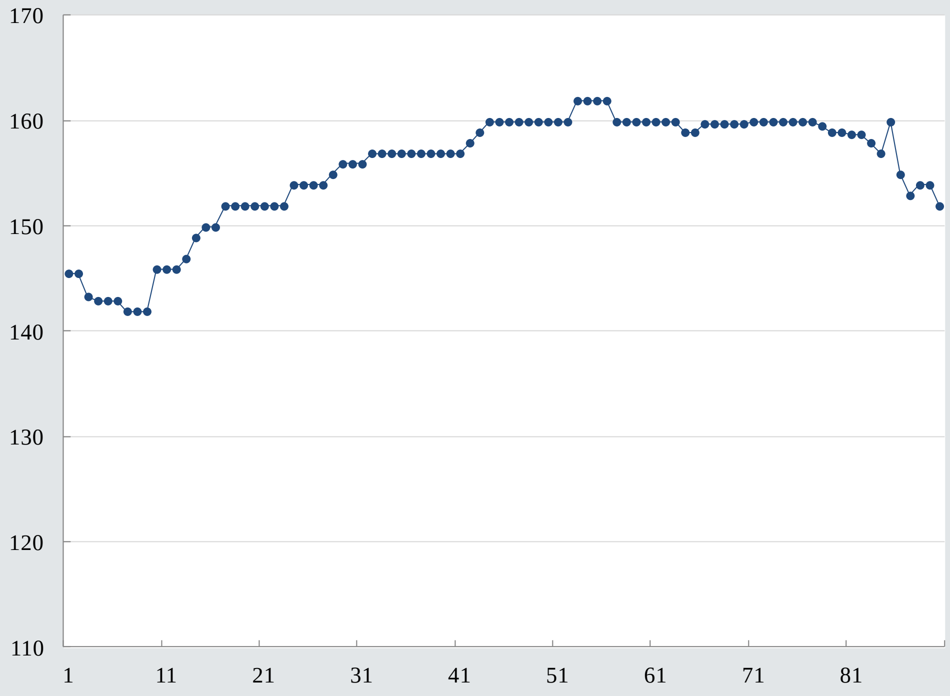
- **Western Australia (WA)**: Team 1 (two RAs) single-labeled 24,569 station-quarters in 260 hours.
- **New South Wales (NSW)**: Team 2 (three RAs) triple-labeled 9,693 station-quarters in 210 hours.
- **Germany**: Teams 2 & 3 (three + three = six RAs) triple-labeled 35,685 station-quarters in 480 hours.
- All RAs are either graduate or undergraduate students at Yale University, majoring in economics, mathematics, or statistics. Wage = US$13.50/hour
- Total labor cost = US$12,825

Examples (1): WA

CYCLING

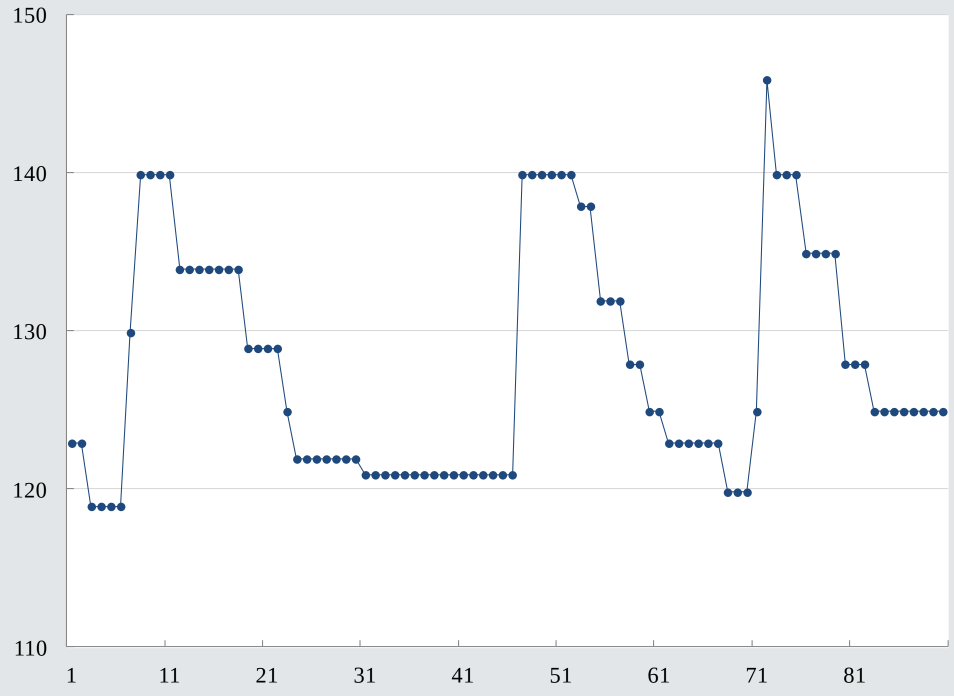


NOT CYCLING

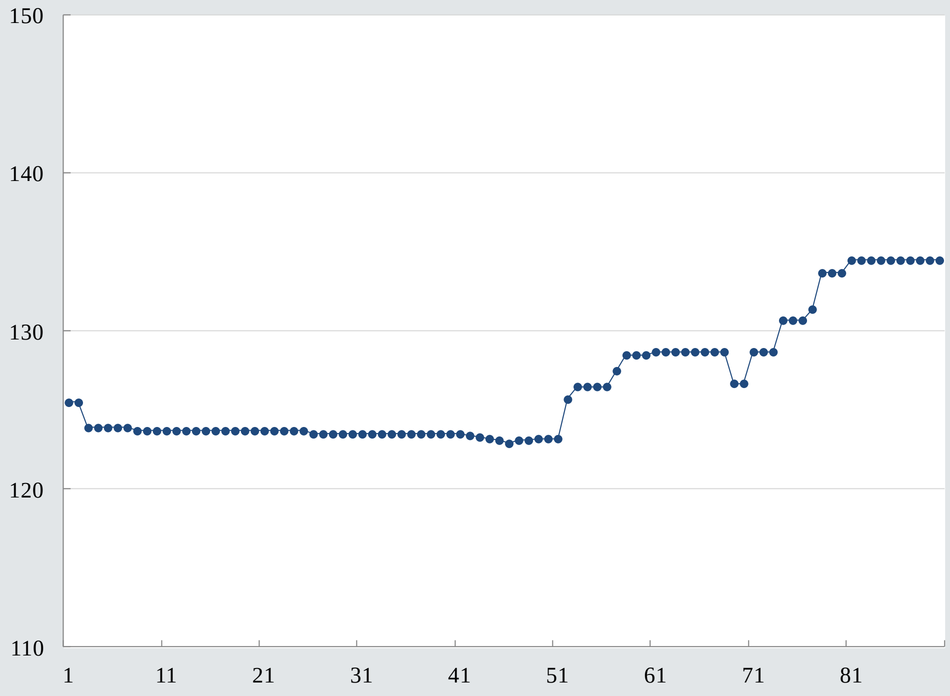


Examples (2): NSW

CYCLING

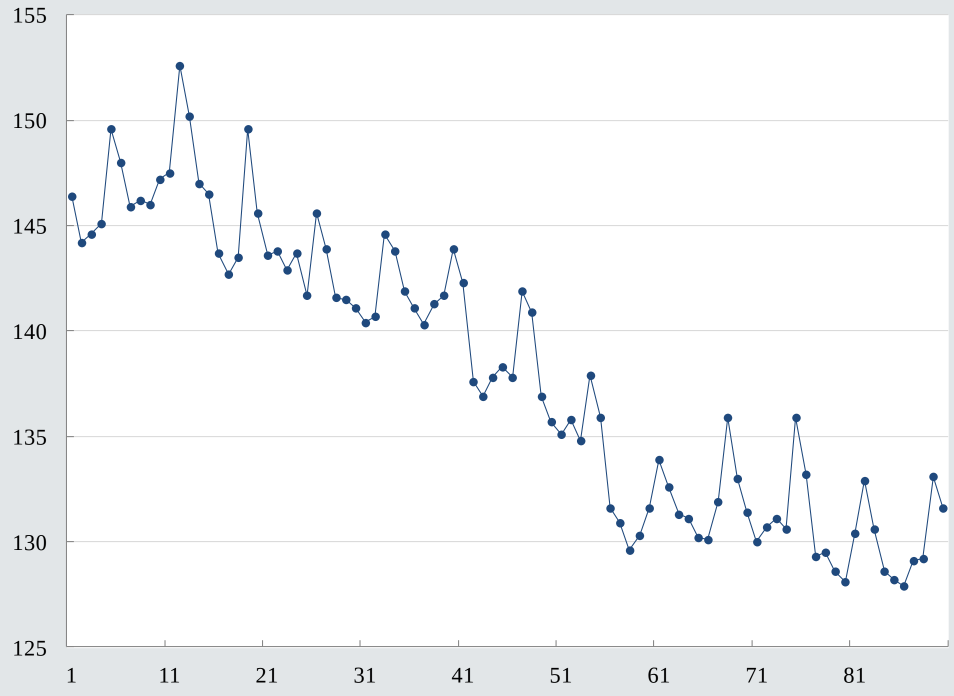


NOT CYCLING

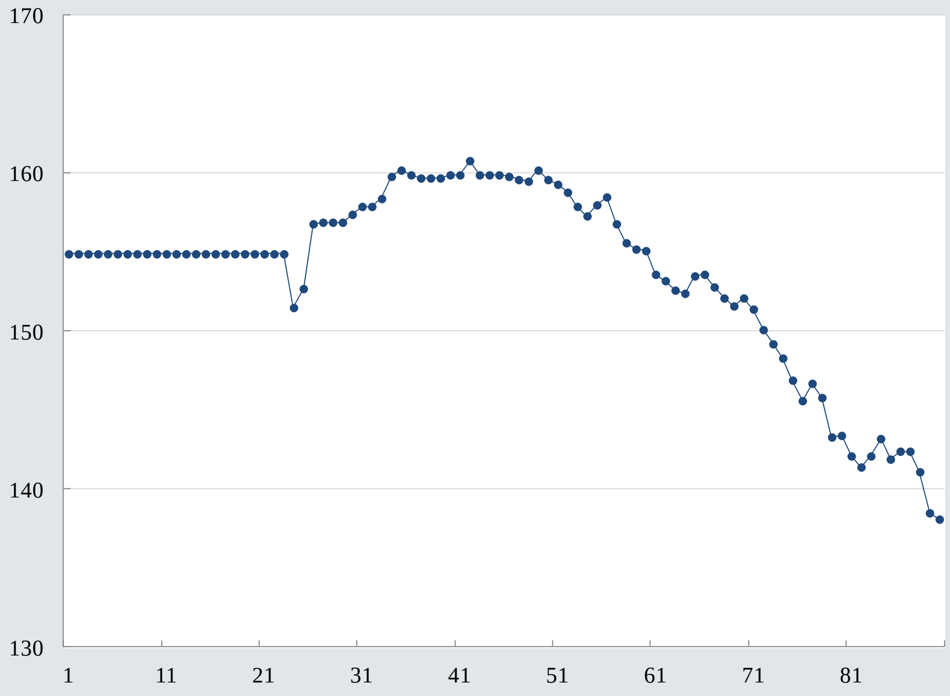


Examples (3): Germany

CYCLING



NOT CYCLING



Results

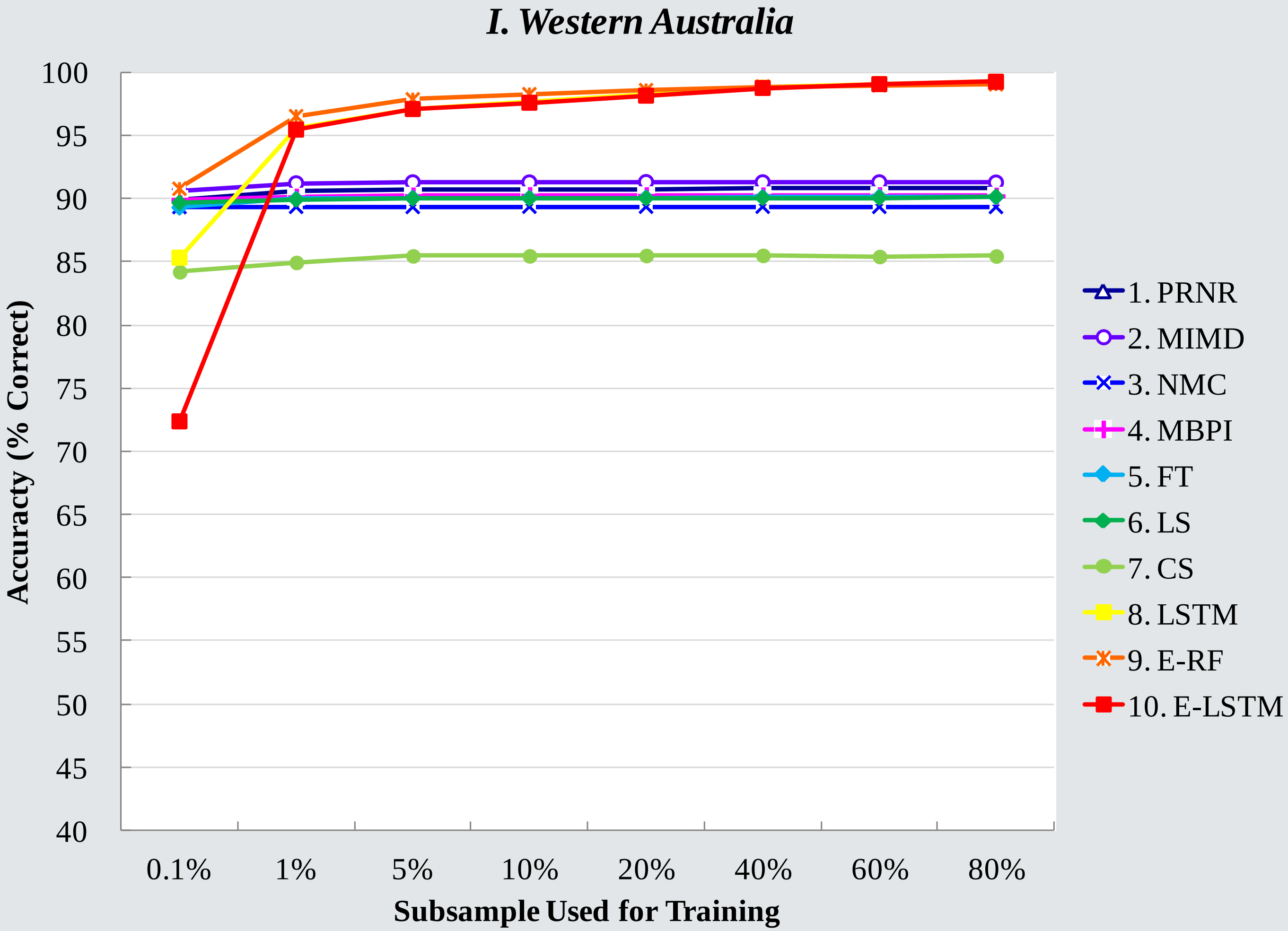
1. Accuracy “horse race”
2. How much data do we need?
3. Markups & cycles

Accuracy Comparison in WA (= easy)

Most methods perform near/above 90%.

Methods 8–10 perform near/above 99%.

Method 7 lags behind.

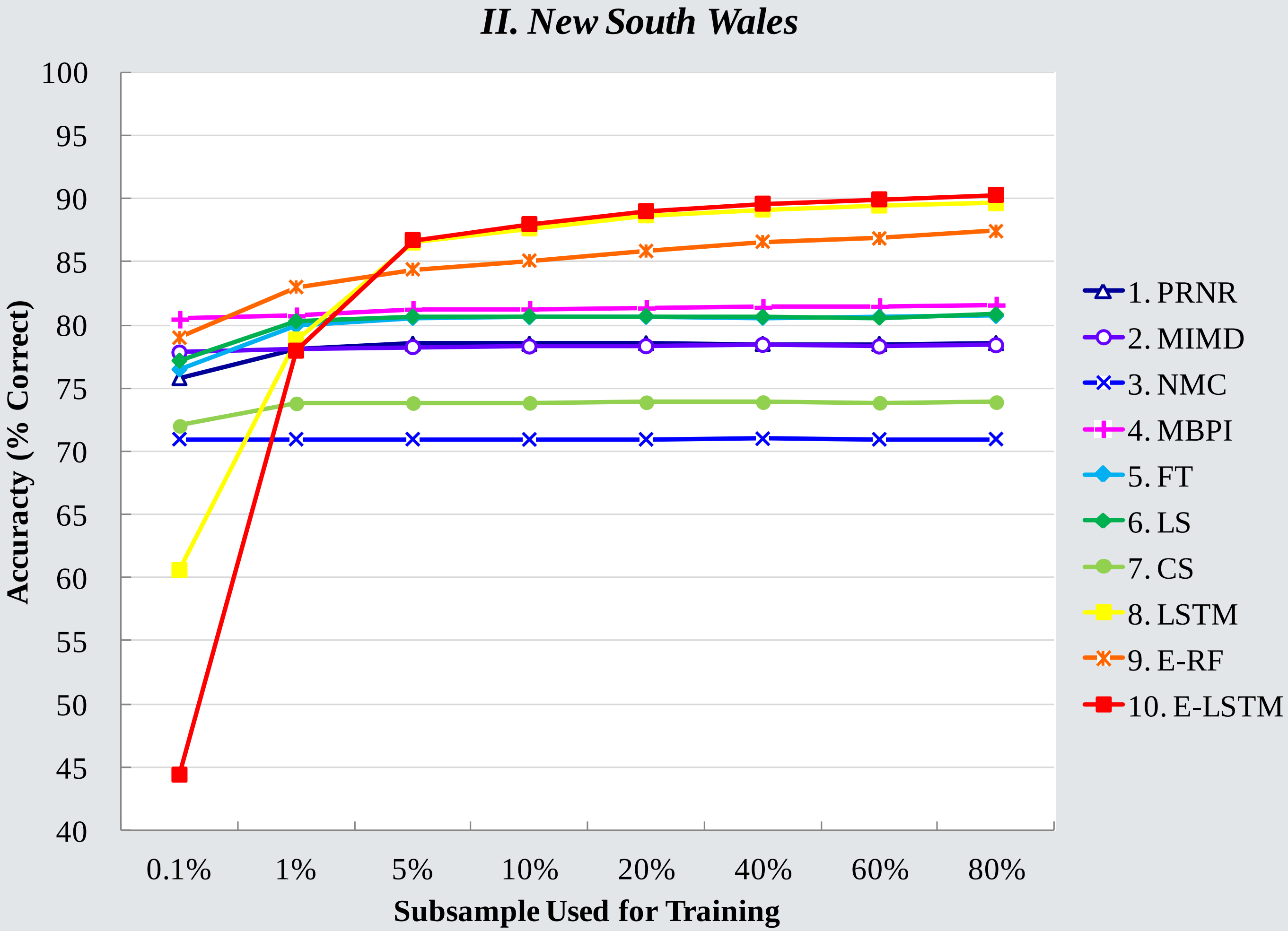


Accuracy Comparison in NSW (= medium)

Most methods perform near/above 80%.

Methods 8–10 perform near/above 85%–90%.

Methods 3 & 7 give degenerate predictions



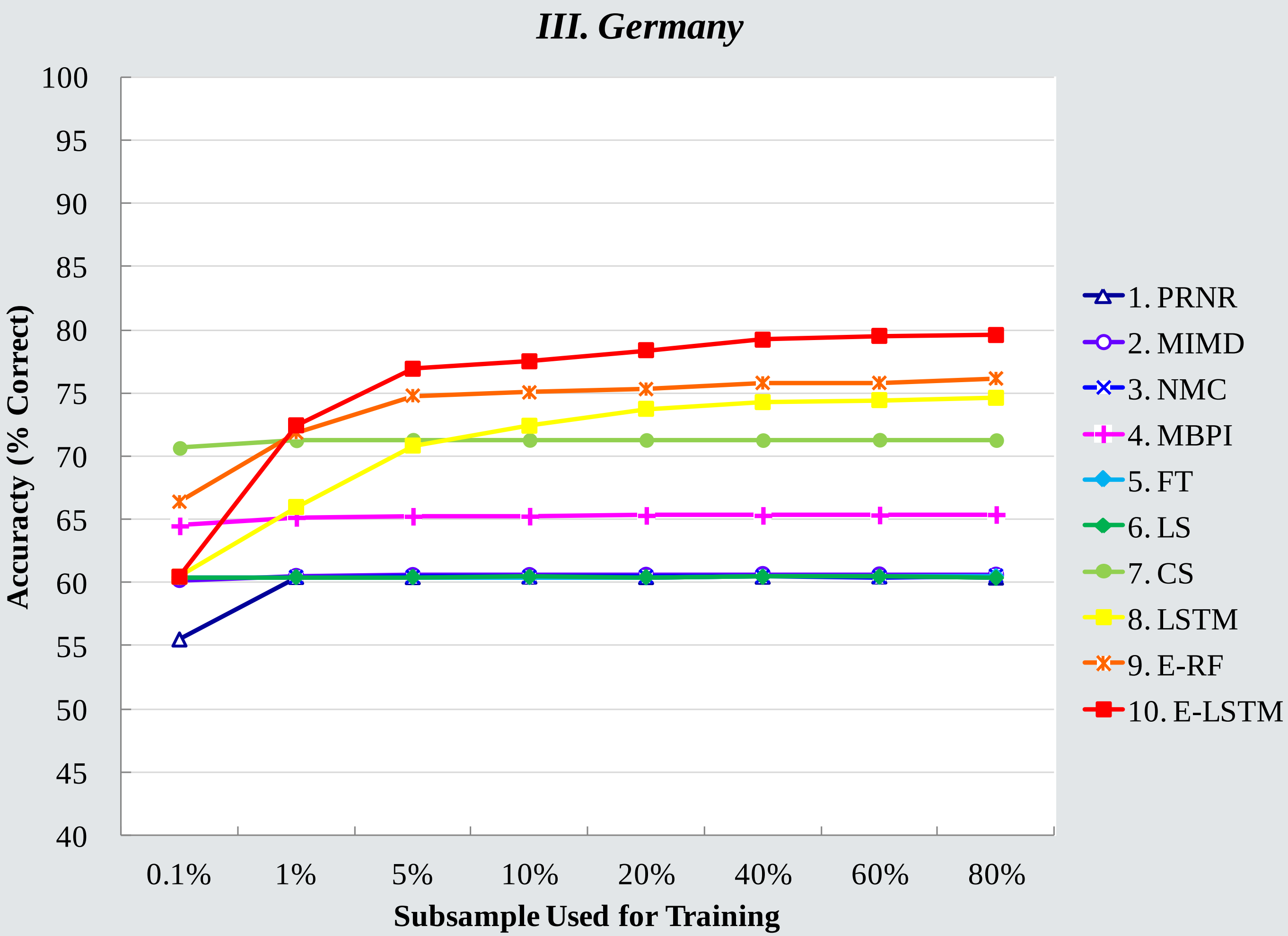
Accuracy Comparison in Germany (= hard)

Most methods ***fail***.

Method 10 achieves 80%, followed by Methods 8–9.

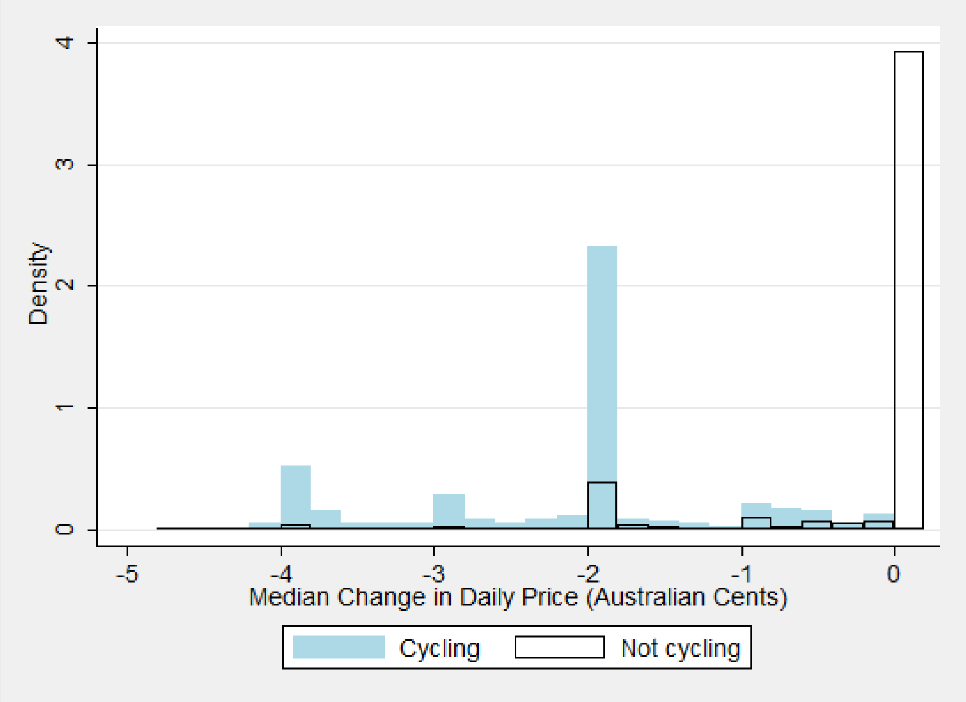
Method 7 does O.K.

Among existing methods, only Method 4 gives non-degenerate predictions.

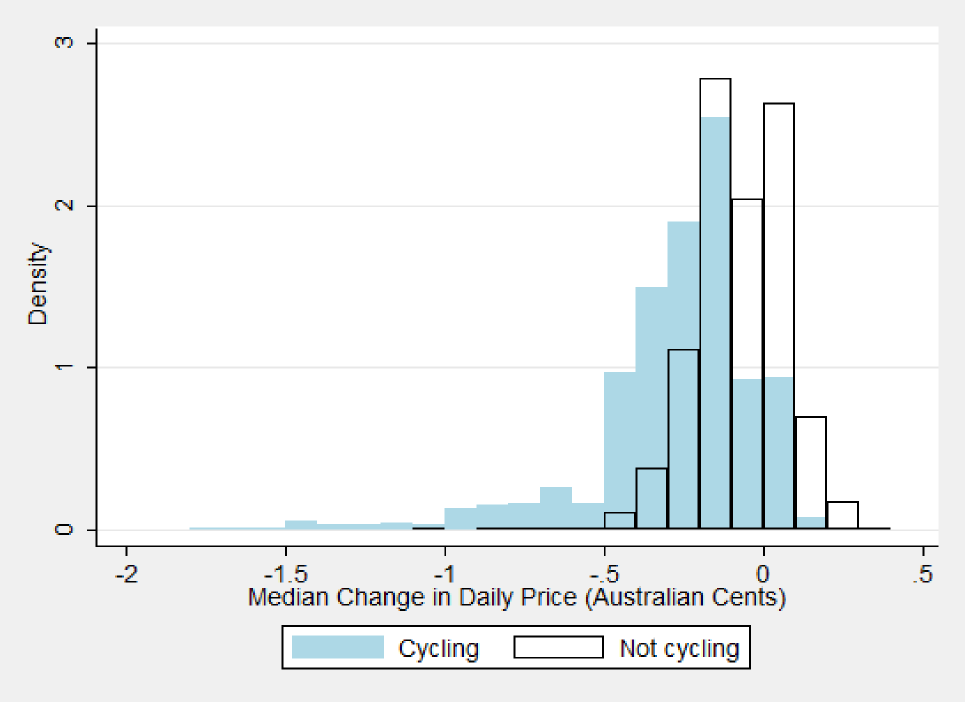


Obvious Question 1:
Why Do Existing Methods (1–4) Work So Well in Australia…?

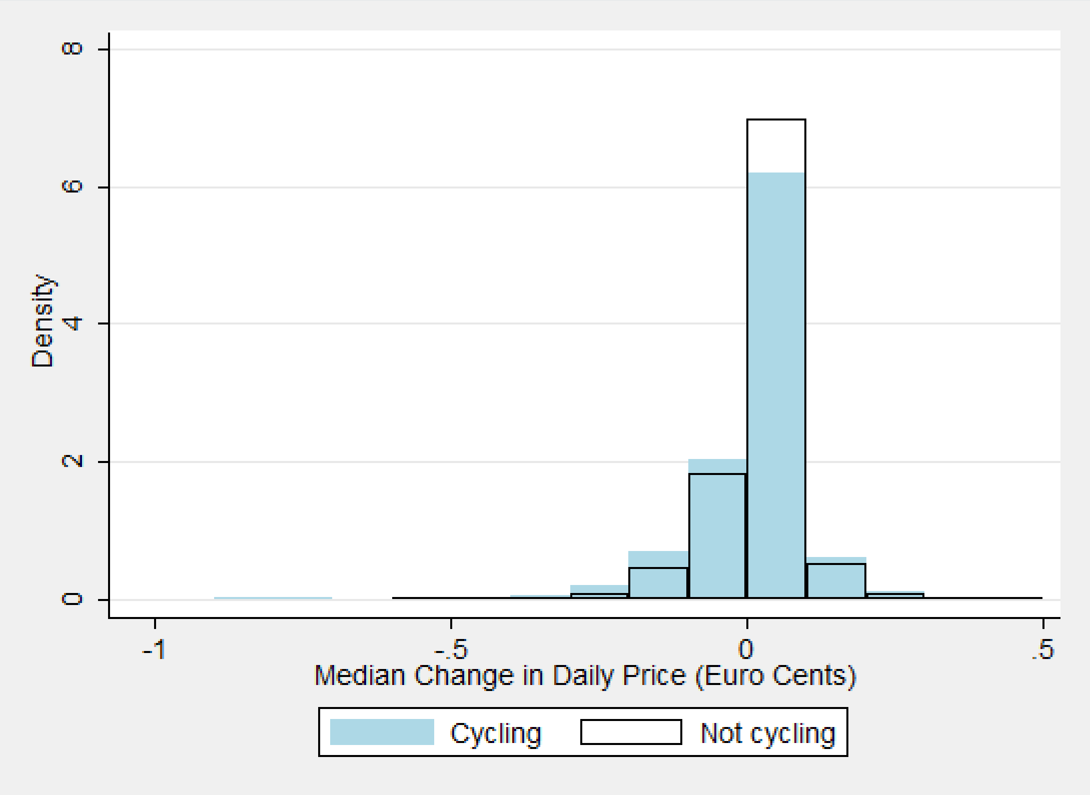
WA



NSW



Obvious Question 1:
...But Totally Fail in Germany?



Obvious Question 2:
How Much (Manually Labeled) Data Do We Need?

- Methods 1–7 & 9 perform surprisingly well with only 0.1% of the data (= 25, 10, & 36 observations in WA, NSW, & Germany, respectively).
- Methods 8 & 10 need more data but eventually outperform the others.
- The "critical" data size is about 1%–5% of the samples (= several hundred observations = tens of RA hours = a few hundred US dollars): Economical!

Markups & Cycles (1): WA

Method	(0) Manual	(1) PRNR	(2) MIMD	(3) NMC	(4) MBPI	(5) FT	(6) LS	(7) CS	(8) LSTM	(9) E-RF	(10) E-LSTM
I. Western Australia (# manually labeled observations: 24,569)											
Cycling											
# obs.	15,007	14,462	14,620	16,147	16,941	16,223	15,774	15,953	15,011	14,994	14,999
Mean	11.86	12.07	12.21	11.66	11.46	11.88	12.03	11.78	11.86	11.86	11.86
Std. dev.	4.01	3.80	3.74	3.98	4.13	3.87	3.85	4.04	4.01	4.01	4.01
Not cycling											
# obs.	9,562	10,107	9,949	8,422	7,628	8,346	8,795	8,616	9,558	9,575	9,570
Mean	9.47	9.30	9.05	9.52	9.73	9.08	8.94	9.35	9.47	9.47	9.47
Std. dev.	4.97	5.04	4.98	5.22	5.20	5.18	5.03	5.02	4.97	4.97	4.96
Difference											
Mean diff.	2.39	2.77	3.16	2.14	1.73	2.80	3.09	2.43	2.39	2.39	2.39
Welch's t	39.53	46.74	53.80	32.96	25.64	43.53	50.02	38.67	39.53	39.55	39.60
D. F.	17,247	17,771	17,314	13,648	12,134	13,263	14,608	14,723	17,236	17,282	17,295
p value	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$

Markups & Cycles (2): NSW

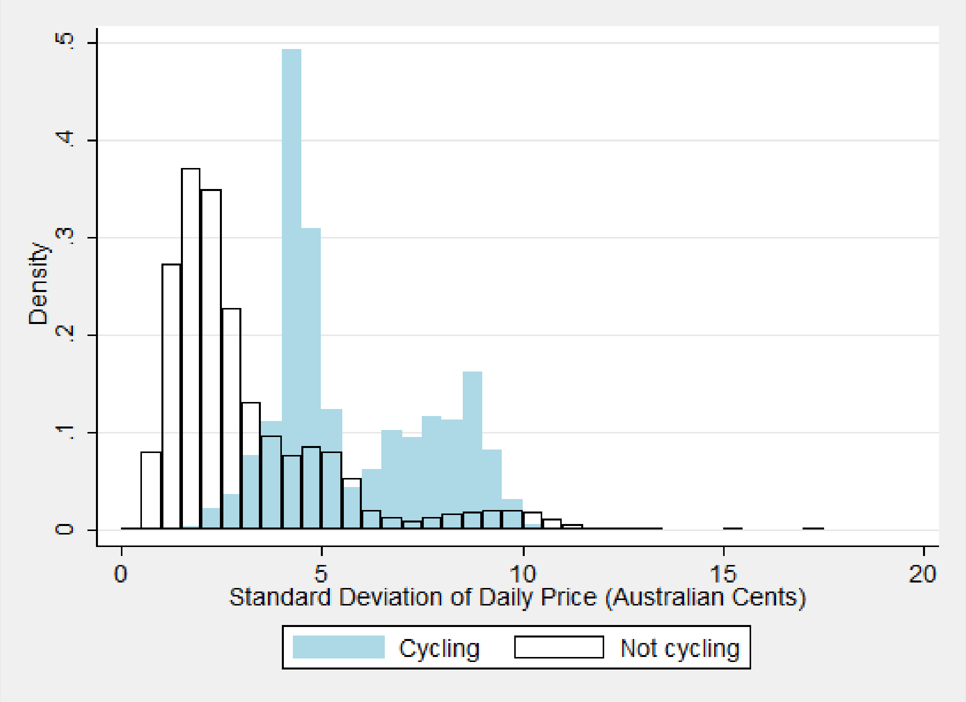
Method	(0) Manual	(1) PRNR	(2) MIMD	(3) NMC	(4) MBPI	(5) FT	(6) LS	(7) CS	(8) LSTM	(9) E-RF	(10) E-LSTM
II. New South Wales (# manually labeled observations: 9,693)											
Cycling											
# obs.	6,878	8,324	8,038	9,693	7,303	7,704	7,994	9,253	7,052	6,961	7,183
Mean	12.03	11.73	12.35	11.66	12.48	11.76	11.81	11.58	12.19	12.07	12.13
Std. dev.	5.51	5.80	5.58	6.04	5.48	5.89	5.84	5.99	5.54	5.53	5.56
Not cycling											
# obs.	2,815	1,369	1,655	0	2,390	1,989	1,699	440	2,641	2,732	2,510
Mean	10.76	11.25	8.33	–	9.18	11.28	10.97	13.48	10.25	10.64	10.33
Std. dev.	7.10	7.31	7.01	–	6.92	6.56	6.85	6.79	7.01	7.08	7.08
Difference											
Mean diff.	1.27	0.48	4.02	–	3.30	0.48	0.84	−1.90	1.94	1.43	1.80
Welch's t	8.50	2.31	21.94	–	21.24	2.97	4.70	−5.76	12.80	9.48	11.55
D. F.	4,266	1,663	2,106	–	3,423	2,870	2,252	472	3,939	4,103	3,648
p value	$< .001$	.021	$< .001$	–	$< .001$	.003	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$

Markups & Cycles (3): Germany

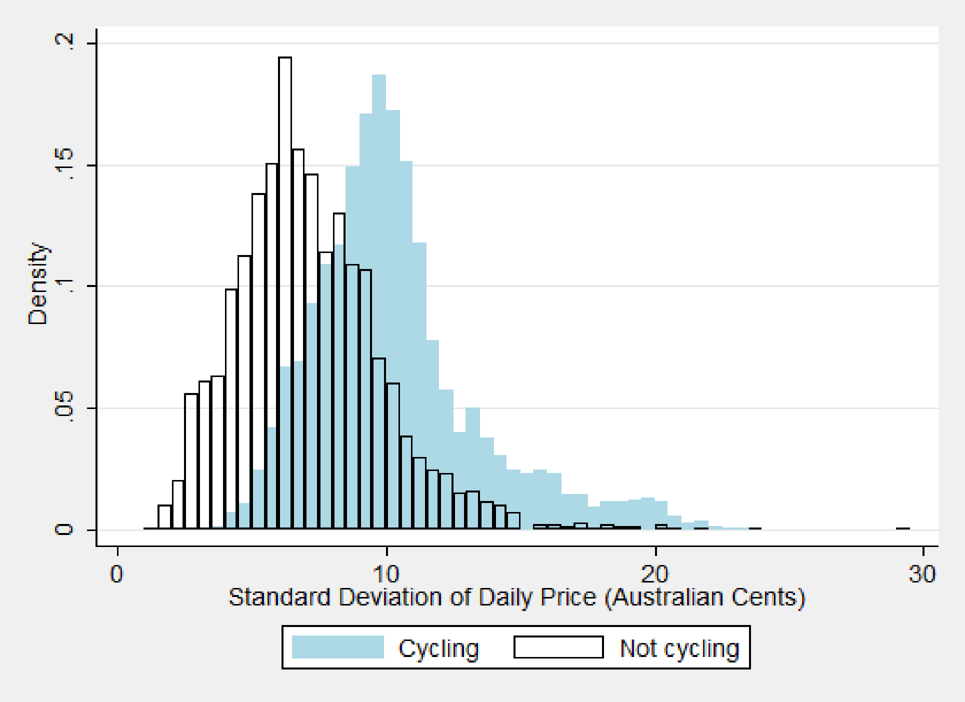
Method	(0) Manual	(1) PRNR	(2) MIMD	(3) NMC	(4) MBPI	(5) FT	(6) LS	(7) CS	(8) LSTM	(9) E-RF	(10) E-LSTM
III. Germany (# manually labeled observations: 35,685)											
Cycling											
# obs.	14,116	0	1,013	72	8,763	7	7	14,281	11,762	13,574	15,299
Mean	98.18	–	99.57	99.67	98.73	114.11	115.64	98.19	98.38	98.16	98.18
Std. dev.	3.57	–	6.96	3.26	3.84	32.10	31.40	3.60	3.60	3.59	3.51
Not cycling											
# obs.	21,569	35,685	34,672	35,613	26,922	35,678	35,678	21,404	23,923	22,111	20,386
Mean	98.65	98.46	98.43	98.46	98.38	98.46	98.46	98.65	98.50	98.65	98.68
Std. dev.	4.37	4.08	3.96	4.08	4.15	4.05	4.05	4.36	4.30	4.34	4.45
Difference											
Mean diff.	−0.47	–	1.14	1.21	0.35	15.65	17.18	−0.46	−0.12	−0.49	−0.50
Welch's t	−11.11	–	5.19	3.14	7.26	1.29	1.45	−10.86	−2.77	−11.55	−11.86
D. F.	33,984	–	1,031	71	15,941	6	6	34,110	27,415	32,697	35,595
p value	< .001	–	< .001	.002	< .001	.245	.197	< .001	.006	< .001	< .001

Obvious Question 3: Why Margins at *Cycling* (i, t) > Margins at *Non-Cycling* (i, t) in Australia…?

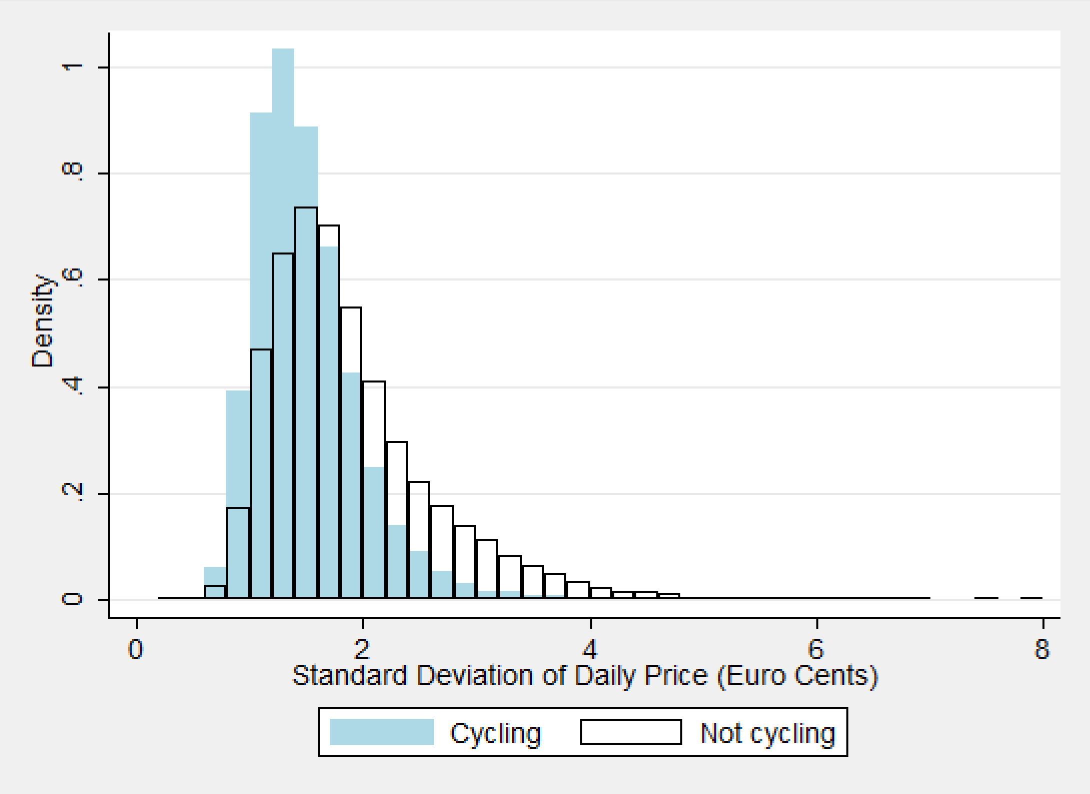
WA



NSW

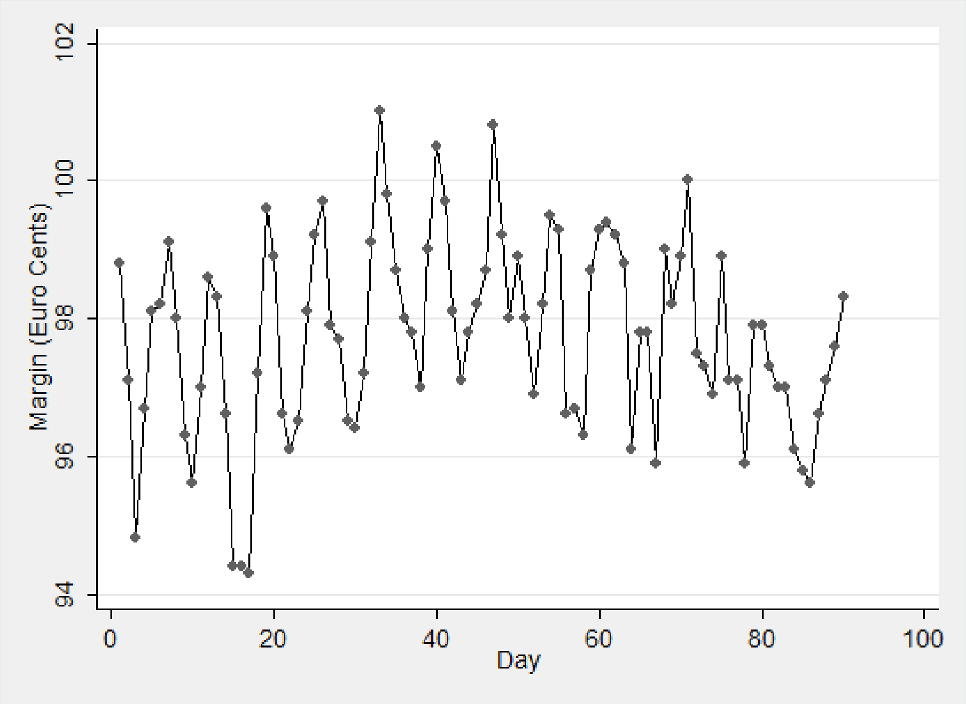


Obvious Question 3: ...And Why Margins at *Cycling* (*i*, *t*) < Margins at *Non-Cycling* (*i*, *t*) in Germany?

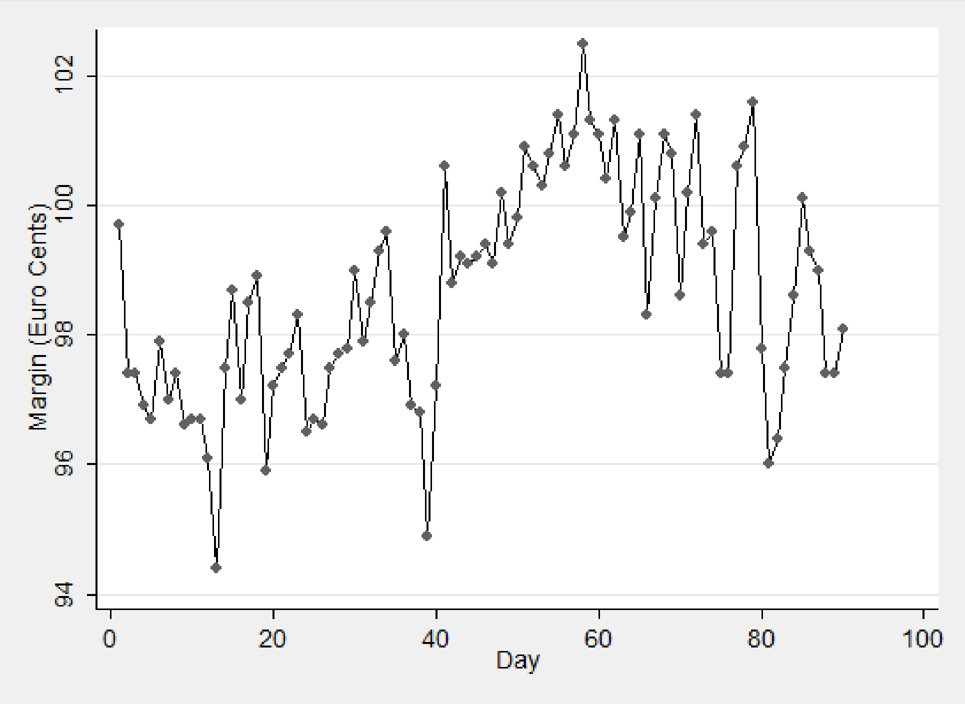


Obvious Question 4:
But, How Can "Cycles" Be Less Volatile Than "Non-Cycles"?

"CYCLING" EXAMPLE
(MEAN=98.13 & STD.DEV.=1.59)



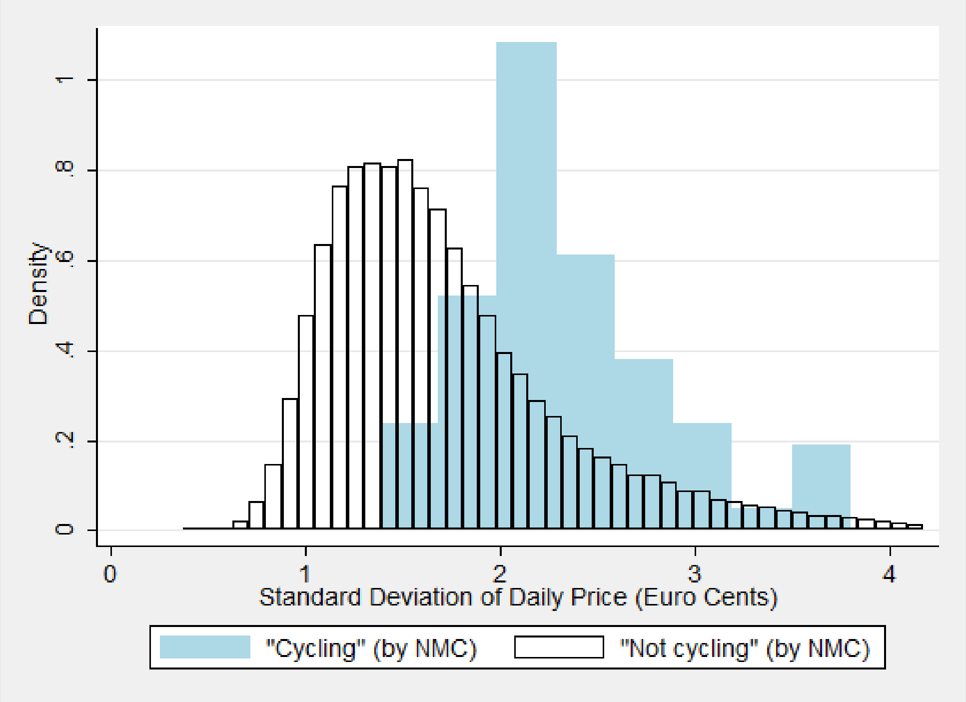
"NON-CYCLING" EXAMPLE
(MEAN=98.13 & STD.DEV.=1.59)



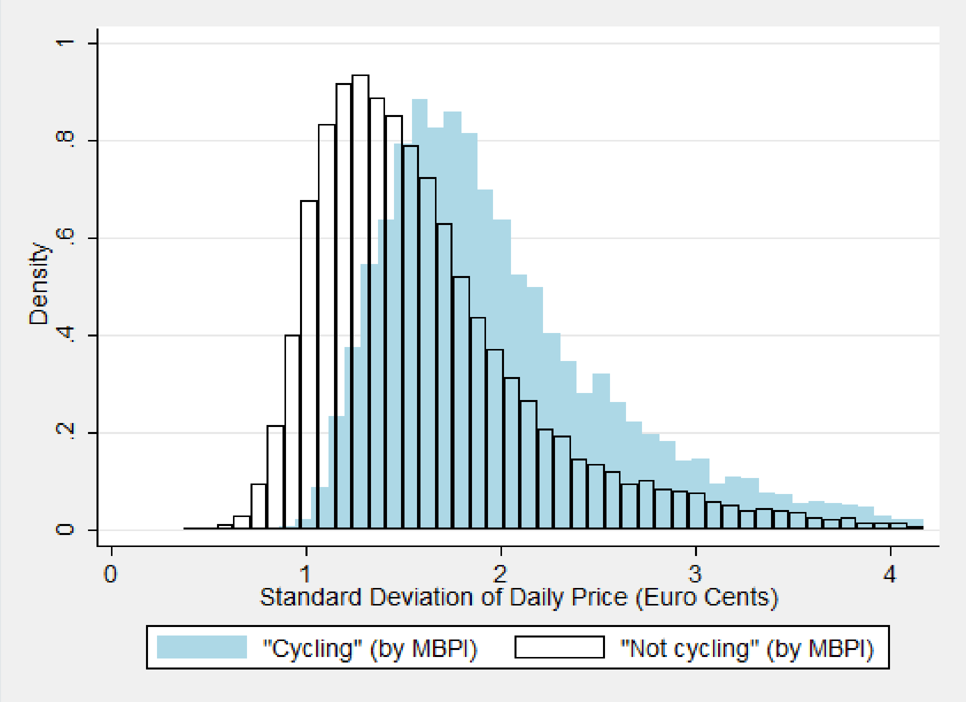
Hint: Human RAs recognize multi-day up-downs as "cycles" & daily zig-zags as noise.

Obvious Question 5:
Why Did Existing Methods Find Corr(margin,cycle) > 0?

METHOD 3: NEGATIVE MEDIAN CHANGE



METHOD 4: MANY BIG PRICE INCREASES



Hint: Their threshold conditions tend to pick up high-volatility (≈ high-mean) cases.

Obvious Question 6:
Human RAs Focus on "Cyclicality" But Not "Asymmetry."
Maybe "Asymmetric Cycles" Do Feature Higher Margins?

Method	(0) Manual	(00) Manual + asymmetry
	III. Germany	
Cycling		
# obs.	14,116	4,265
Mean	98.18	98.01
Std. dev.	3.57	3.39
Not cycling		
# obs.	21,569	31,420
Mean	98.65	98.53
Std. dev.	4.37	4.16
Difference		
Mean diff.	−0.47	−0.52
Welch's t	−11.11	−9.05
D. F.	33,984	6,153
p value	$< .001$	$< .001$

Answer: No.

Conclusion

- We formalize 4 existing methods, propose 6 new methods, & empirically assess their performances in WA, NSW, & Germany.
- **Methodologically**: (1) difficulty of cycle detection varies across countries/regions; (2) Existing methods work well in WA & NSW but mostly fail in Germany, because not all German cycles fit Edgeworth-style asymmetry → Distinguish between “asymmetry” & “cyclicality”
- (3) Nonparametric/machine-learning methods (esp. LSTM & E-LSTM) achieve highest accuracy (99%, 90%, & 80%, respectively) at reasonable labor cost.

Conclusion (cont.)

- **Substantively**: Whether researchers find a positive or negative statistical relationship between gas stations' profit margins & the existence of cycles could critically depend on their choice of "operational definitions" & detection methods.
- Because the discovery of "facts" inform subsequent policy interventions, these (seemingly innocuous) methodological considerations are consequential & directly policy-relevant.

Recommendation for Researchers/Practitioners

1. Manually label 100 observations for cyclicality.
2. Calibrate/optimize Method 4 (MBPI) for detecting cycles.
3. If needed, use Methods 5 (FT) or 6 (LS) for clearly defining cycles.
4. If these methods do not work, additionally label 200–400 observations and try Methods 7 (CS), 9 (E-RF), 8 (LSTM), & 10 (E-LSTM) in the increasing order of complexity/accuracy.
5. After automating cycle-detection, classify cycling observations by asymmetry: (a) Edgeworth, (b) inverse-Edgeworth, & (c) symmetry.
6. Compare prices & markups between subsamples (defined in the above).