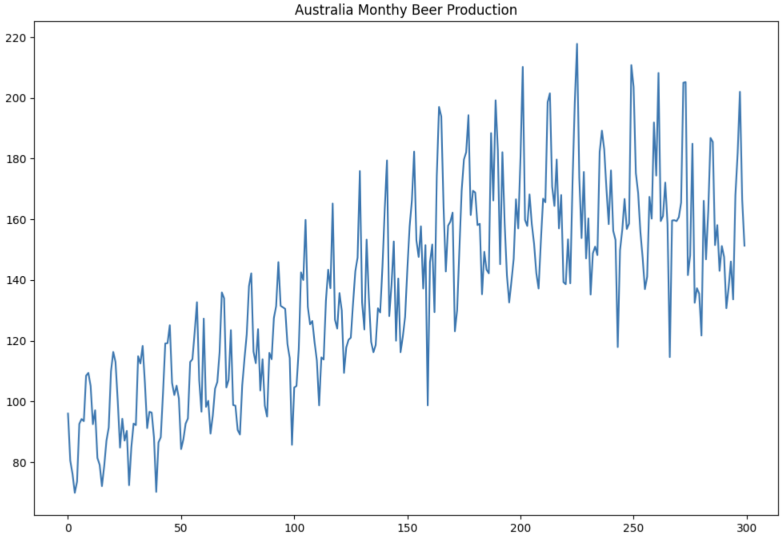
Improved Imputation of Missing Values in Time Series

University of Rochester StemForAll 2025

Alex Nappo, Spencer Lyudovyk, Showmee Zhou, Oscar Bernfield, Mingyu Zhang
Project Supervisors: Alex and Joshua Iosevich

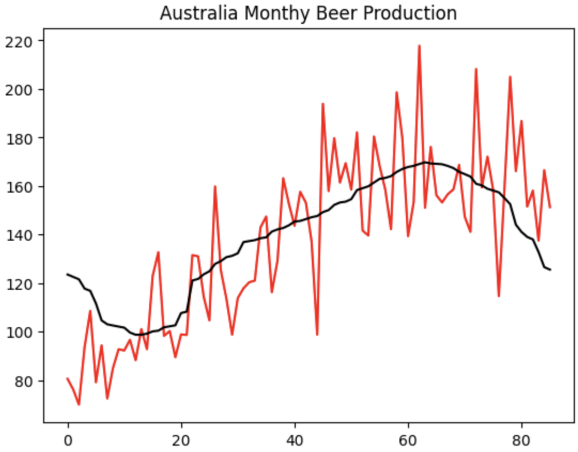
August 8, 2025

Australia Monthly Beer Production



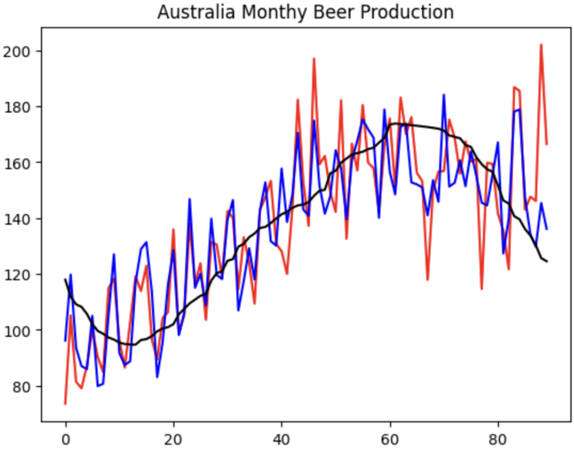
Classical imputation using trig polynomial regression

What if some values in our time series are lost? In the graph below, 100 of the original 300 values have been randomly removed. The original missing values are in red, and the imputed values are in black.



Imputation using signal recovery methods

This time, the original missing values are in red, the trig regression values in black, and the imputed values using signal recovery methods are in blue.



Imputation using signal recovery methods

- However, the question remains: how did we arrive at this graph and the imputation method shown, and can it be improved?

Signals and the Discrete Fourier Transform

- Let f be a signal of finite length, i.e. $f : \mathbb{Z}_N \to \mathbb{C}$

Signals and the Discrete Fourier Transform

- Let f be a signal of finite length, i.e. $f : \mathbb{Z}_N \to \mathbb{C}$
- Suppose that the Fourier transform of f is transmitted, where:

$$\hat{f}(m) = N^{-\frac{1}{2}} \sum_{x \in \mathbb{Z}_N} \chi(-x \cdot m) f(x); \ \chi(t) = e^{\frac{2\pi i t}{N}}$$

Signals and the Discrete Fourier Transform

- Let f be a signal of finite length, i.e. $f : \mathbb{Z}_N \to \mathbb{C}$
- Suppose that the Fourier transform of f is transmitted, where:

$$\hat{f}(m) = N^{-\frac{1}{2}} \sum_{x \in \mathbb{Z}_N} \chi(-x \cdot m) f(x); \; \chi(t) = e^{\frac{2\pi i t}{N}}$$

- Fourier Inversion states that f can be recovered by:

$$f(x) = N^{-\frac{1}{2}} \sum_{m \in \mathbb{Z}_N} \chi(x \cdot m) \hat{f}(m)$$

Exact Recovery Problem

- Now, suppose that the values $\{\hat{f}(m)\}_{m \in S}$ are not observed.
- Can f be recovered exactly from its discrete Fourier transform?
- The answer is yes (under some conditions)!

Exact Recovery

- Let the support of a signal f be defined as

$$\text{supp}(f) = \{x \in \mathbb{Z}_N : f(x) \neq 0\}$$

Theorem (Matolcsi-Szucks/Donoho-Stark)

Let $f : \mathbb{Z}_N \to \mathbb{C}$ be supported in $E \subset \mathbb{Z}_N$. Suppose that $\hat{f}$ is transmitted but the frequencies $\{\hat{f}(m)\}_{m \in S}$ are unobserved, where $S \subset \mathbb{Z}_N$. Then f can be recovered exactly and uniquely if

$$|E| \cdot |S| < \frac{N}{2}$$

.

Logan's Phenomenon and L^1-Minimization Algorithm

Theorem

Let $f : \mathbb{Z}_N \to \mathbb{C}$ be supported in $E \subset \mathbb{Z}_N$. Suppose that $\hat{f}$ is transmitted but the frequencies $\{\hat{f}(m)\}_{m \in S}$ are unobserved, where $S \subset \mathbb{Z}_N$, with $|E| \cdot |S| < \frac{N}{2}$. Then f can be recovered exactly and uniquely. Moreover,

$$f = argmin_g ||g||_{L^1(\mathbb{Z}_N)} \text{with the constraint } \hat{f}(m) = \hat{g}(m), m \notin S$$

Logan's Phenomenon and L^1-Minimization Algorithm

Theorem

Let $f : \mathbb{Z}_N \to \mathbb{C}$ be supported in $E \subset \mathbb{Z}_N$. Suppose that $\hat{f}$ is transmitted but the frequencies $\{\hat{f}(m)\}_{m \in S}$ are unobserved, where $S \subset \mathbb{Z}_N$, with $|E| \cdot |S| < \frac{N}{2}$. Then f can be recovered exactly and uniquely. Moreover,

$$f = argmin_g ||g||_{L^1(\mathbb{Z}_N)} \text{with the constraint } \hat{f}(m) = \hat{g}(m), m \notin S$$

- Logan's celebrated result is the cornerstone of our further work

Shaping Logan to fit real-world data

- Datasets and signals are generally noisy, so it is useful to modify the constraint to allow small differences between g and f.

Shaping Logan to fit real-world data

- Datasets and signals are generally noisy, so it is useful to modify the constraint to allow small differences between g and f.
- Additionally, when dealing with time series data, referring to the support of a function is not practical. Rather, it is more realistic to define a relationship between values that are relatively small compared to the rest of the data set.

Shaping Logan to fit real-world data

- Datasets and signals are generally noisy, so it is useful to modify the constraint to allow small differences between g and f.
- Additionally, when dealing with time series data, referring to the support of a function is not practical. Rather, it is more realistic to define a relationship between values that are relatively small compared to the rest of the data set.

Definition

Let $u : \mathbb{Z}_N \to \mathbb{C}$. We say that u is L_p-concentrated on $A \subset \mathbb{Z}_N$ with the norm $\leq \epsilon$ if

$$\|u\|_{L^p(A_c)} \leq \frac{\epsilon}{N} \cdot \|u\|_{L^p(\mathbb{Z}_N)}$$

Previous Result

Theorem

Let $f : \mathbb{Z}_N \to \mathbb{C}$, and suppose that the values $\{f(x)\}_{x \in M}$ are unobserved, where M is a generic subset of $\mathbb{Z}_N$, of size $\leq \gamma_0 \frac{N}{\log(N)}$, where γ_0 is as in Talagrand's Theorem. Let

$$g = argmin_u ||\widehat{u}||_1 : ||u - f||_{L^1(M^c)} \leq \delta N^{-1} ||f||_{L^1(M^c)}. \tag{1}$$

Suppose that $\widehat{f}$ is ϵ-concentrated on $S \subset \mathbb{Z}_N$ such that

$$|S| < \frac{1}{16 C_T^2} \frac{N}{\log(N) \log\log(N)}. \tag{2}$$

Let $h = f - g$. Then if $h \neq 0$, then with probability $1 - o_N(1)$

$$\frac{1}{|M|} \sum_{x \in M} |h(x)| \leq (4\epsilon + 5\delta) \cdot \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|. \tag{3}$$

Improved Result

Theorem (Improved)

Let $f : \mathbb{Z}_N \to \mathbb{C}$, and suppose that the values $\{f(x)\}_{x \in M}$ are unobserved, where M is a generic subset of $\mathbb{Z}_N$, of size $\leq \gamma_0 \frac{N}{\log(N)}$, where γ_0 is as in Talagrand's Theorem. Let

$$g = argmin_u ||\widehat{u}||_1 : ||u - f||_{L^1(M^c)} \leq \delta N^{-1} ||f||_{L^1(M^c)}. \tag{4}$$

Suppose that $\widehat{f}$ is ϵ-concentrated on $S \subset \mathbb{Z}_N$ such that

$$|S| < \frac{1}{16 C_T^2} \frac{N}{\log(N) \log\log(N)}. \tag{5}$$

Let $h = f - g$. Then if $h \neq 0$, then with probability $1 - o_N(1)$

$$\frac{1}{|M|} \sum_{x \in M} |h(x)| \leq \left(4\epsilon \frac{|S|}{N - \epsilon} + \left(4 \frac{|S|}{N} + 1 \right) \delta \right) \cdot \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|. \tag{6}$$

Improved Result

- The proof of our result uses the same framework as the previous theorem, but makes two key improvements.

Improved Result

- The proof of our result uses the same framework as the previous theorem, but makes two key improvements.
- First, we improved the bound on $\|\widehat{f}\|_1$.
  The previous bound was
  $$\|\widehat{f}\|_1 \leq N^{\frac{1}{2}}\|f\|_1.$$
  Our improved bound is
  $$\|\hat{f}\|_1 \leq \frac{|S|}{N-\epsilon} N^{\frac{1}{2}}\|f\|_1.$$

Improved Result

- Second, we improved the bound on $\|\widehat{h}\|_{L^1(S)}$.
  The previous bound was

$$\|\widehat{h}\|_{L^1(S)} \leq 5 \cdot \frac{N^{\frac{1}{2}} \cdot \delta \cdot \|f\|_{L^1(\mu)}}{4} + \frac{\|\widehat{h}\|_1}{4}.$$

  Our improved bound is

$$\|\widehat{h}\|_{L^1(S)} \leq \left(1 + 4\frac{|S|}{N}\right) \cdot \frac{N^{\frac{1}{2}} \cdot \delta \cdot ||f||_{L^1(\mu)}}{4} + \frac{\|\widehat{h}\|_1}{4}.$$

Improved Result

- Second, we improved the bound on $\|\widehat{h}\|_{L^1(S)}$.
  The previous bound was

$$\|\widehat{h}\|_{L^1(S)} \leq 5 \cdot \frac{N^{\frac{1}{2}} \cdot \delta \cdot \|f\|_{L^1(\mu)}}{4} + \frac{\|\widehat{h}\|_1}{4}.$$

  Our improved bound is

$$\|\widehat{h}\|_{L^1(S)} \leq \left(1 + 4\frac{|S|}{N}\right) \cdot \frac{N^{\frac{1}{2}} \cdot \delta \cdot ||f||_{L^1(\mu)}}{4} + \frac{\|\widehat{h}\|_1}{4}.$$

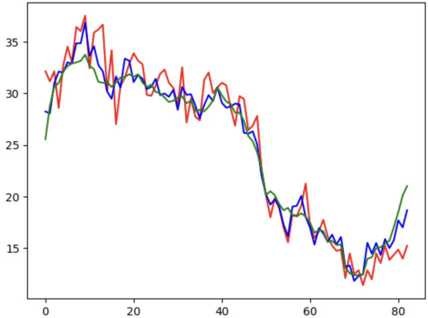
- Using these two improvements, we were able to create a tighter bound in the final inequality.

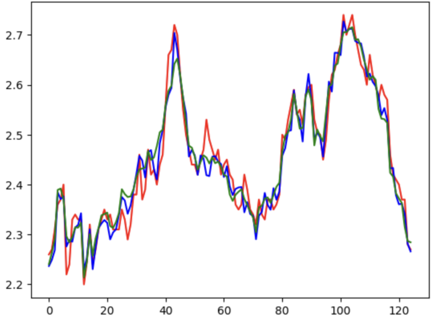
Numerical Experiments with L^1-minimization

- Throughout our numerical experiments, we found that applying the L^2 norm in the constraint was more effective for reducing error.

Numerical Experiments with L^1-minimization

- Throughout our numerical experiments, we found that applying the L^2 norm in the constraint was more effective for reducing error.
- The graphs below compare the L^1 optimizations where the line in red represents the original missing values, the line in blue the L^2 constraint, and line in green the L^1 constraint.





Result with L_2 Norm

Theorem (Improved)

Let $f : \mathbb{Z}_N \to \mathbb{C}$, and suppose that the values $\{f(x)\}_{x \in M}$ are unobserved, where M is a generic subset of $\mathbb{Z}_N$, of size $\leq \gamma_0 \frac{N}{\log(N)}$, where γ_0 is as in Talagrand's Theorem. Let

$$g = argmin_u ||\widehat{u}||_1 : ||u - f||_{L^2(M^c)} \leq \delta N^{-1} ||f||_{L^2(M^c)}. \tag{7}$$

Suppose that $\widehat{f}$ is ϵ-concentrated on $S \subset \mathbb{Z}_N$ such that

$$|S| < \frac{1}{16 C_T^2} \frac{N}{\log(N) \log\log(N)}. \tag{8}$$

Let $h = f - g$. Then if $h \neq 0$, then with probability $1 - o_N(1)$

$$\frac{1}{|M|} \sum_{x \in M} |h(x)| \leq \left(4\epsilon \frac{|S|}{N - \epsilon} + \left(4 \left(\frac{|S|}{N} \right)^{\frac{1}{2}} + 1 \right) \delta \right) \cdot \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |f(x)|. \tag{9}$$

Experiments with Proportions

- We also found that the "type" of data that is missing plays a large role in how accurately it can be recovered.

Experiments with Proportions

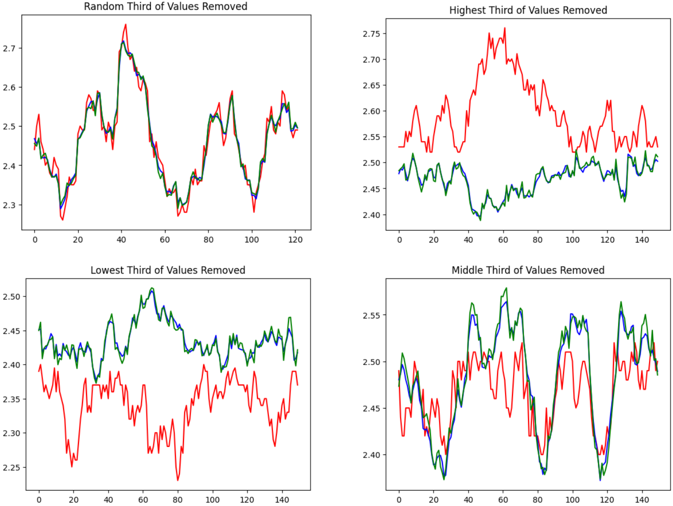
- We also found that the "type" of data that is missing plays a large role in how accurately it can be recovered.
- In particular, removing the highest third of values resulted in significantly more error than removing a random third of values, suggesting this data has structure which is important for accurate recovery.

Experiments with Proportions

- We also found that the “type” of data that is missing plays a large role in how accurately it can be recovered.
- In particular, removing the highest third of values resulted in significantly more error than removing a random third of values, suggesting this data has structure which is important for accurate recovery.
- The result is similar when the lowest values are removed (a lot more error than random removal). However, there is less error when the middle values are removed.

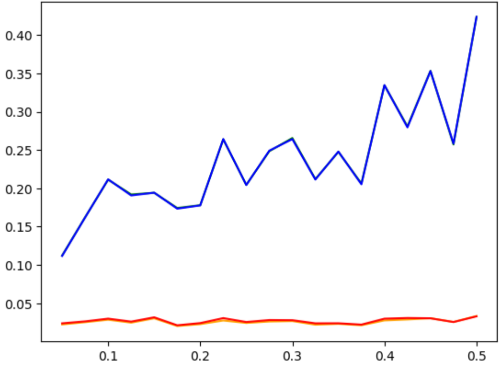
Experiments with Proportions

- The actual data is in red, and the imputed data is in green (L^1 optimizer) and blue ($L^1\epsilon$ optimizer).



Experiments with Proportions

- The graph below shows how error in recovery increases when the proportion p of the data which is removed increases. The proportion of values removed is on the x axis. The line in blue represents error when the highest values are removed, and the line in red represents errors when values are randomly chosen for removal.



References

W. Burstein, A. Iosevich, A. Mayeli, and H. Nathan, *Fourier minimization and time series imputation*, (arXiv:2506.19226), (2025).

D. Donoho and P. Stark, *Uncertainty principle and signal processing*, SIAM Journal of Applied Math., (1989), Society for Industrial and Applied Mathematics, volume 49, No. 3, pp. 906-931.

W. Hagerstrom, *A number of perspectives on signal recovery*, University of Rochester Honors Thesis (2025).

A. Iosevich and A. Mayeli, *Uncertainty Principles, Restriction, Bourgain's Λ_q theorem, and Signal Recovery*, Applied and Computational Harmonic Analysis 76 (2025): 101734.