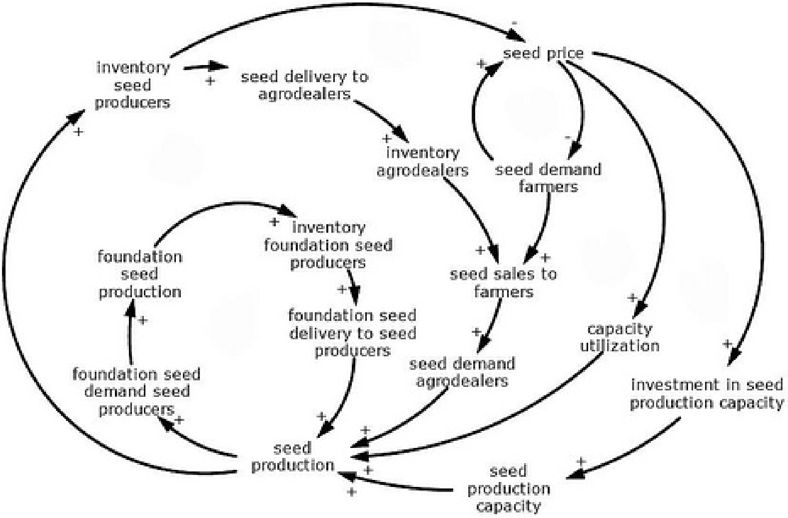
# A Fourier Approach to System Regressions

James Rising

April 8, 2011

# The Endogenous Universe




CAS-IP Case Study: The West African Seed Sector.

# Comparisons of Methods

## Econometric Methods for a Complicated World

| Method | Exogeneity | Restrictions | Validity |
|---|---|---|---|
| OLS | strict | linear, unidirectional | unbiased |
| GMM | $n$ vars | unidirectional | consistent |
| SEM | $n$ vars | 1 structural restriction | consistent |
| VAR | sampling | structural ordering | consistent |
| PSM | propensity choices ? | unidirectional | unbiased |
| CEM | none ? | coarsening | bounded |

# Comparisons of Methods

## Econometric Methods for a Complicated World

| Method | Exogeneity | Restrictions | Validity |
|---|---|---|---|
| OLS | strict | linear, unidirectional | unbiased |
| GMM | $n$ vars | unidirectional | consistent |
| SEM | $n$ vars | 1 structural restriction | consistent |
| VAR | sampling | structural ordering | consistent |
| PSM | propensity choices ? | unidirectional | unbiased |
| CEM | none ? | coarsening | bounded |
| FAR | none | relation tables | unbiased |

# Comparisons of Methods

## Econometric Methods for a Complicated World

| Method | Exogeneity | Restrictions | Validity |
|---|---|---|---|
| OLS | strict | linear, unidirectional | unbiased |
| GMM | $n$ vars | unidirectional | consistent |
| SEM | $n$ vars | 1 structural restriction | consistent |
| VAR | sampling | structural ordering | consistent |
| PSM | propensity choices ? | unidirectional | unbiased |
| CEM | none ? | coarsening | bounded |
| FAR | none | relation tables ? | unbiased ? |

# Cybernetic Relationships

The goal is to measure the simultaneous effects that signals have upon each other, assuming everything is endogenous.

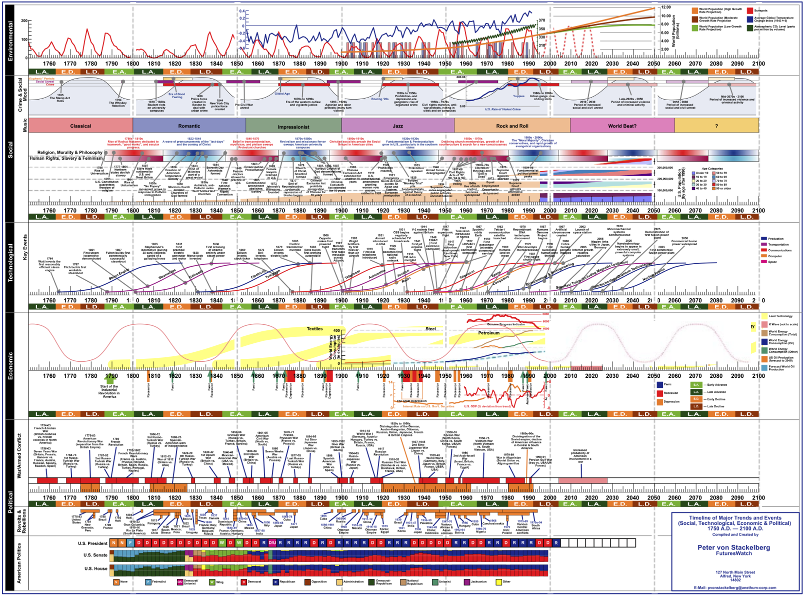
$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \begin{bmatrix} F_{\theta_{11}}\circ & F_{\theta_{12}}\circ & \cdots & F_{\theta_{1n}}\circ \\ F_{\theta_{21}}\circ & F_{\theta_{22}}\circ & \cdots & F_{\theta_{2n}}\circ \\ \vdots & \vdots & \ddots & \vdots \\ F_{\theta_{n1}}\circ & F_{\theta_{n2}}\circ & \cdots & F_{\theta_{nn}}\circ \end{bmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$

$F_{\theta_{ij}}$ is an operator (e.g., a scaling, delay, integral, or combination thereof), which describes how $y_j(t)$ contributes to $y_i(t)$.
The estimation problem is to determine $\hat{\theta}_{ij}$.

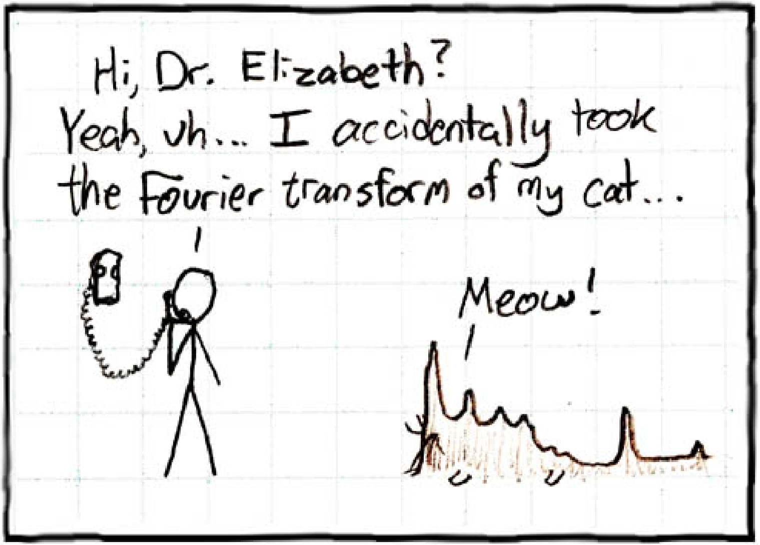
# Paradigms

- Endogeneity and variable errors are ubiquitous.
- Signal processing and system dynamics can help model economic processes.
- Underlying frequencies within different variables can represent interdependence.

# Underlying Frequencies

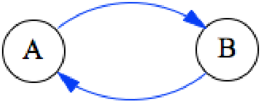


# Fourier Transforms



# Frequencies Show Feedback

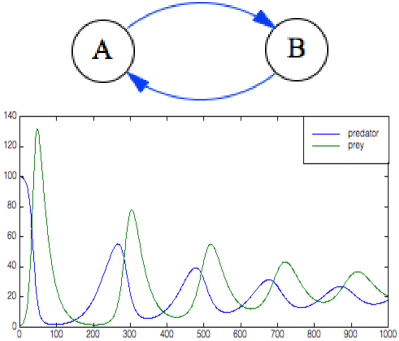
Feedback Relation



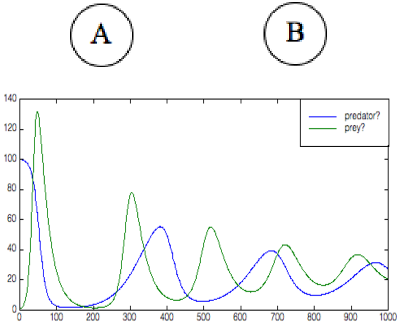

Unrelated Signals

# Frequencies Show Feedback

Feedback Relation

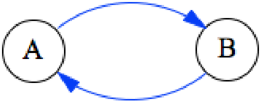


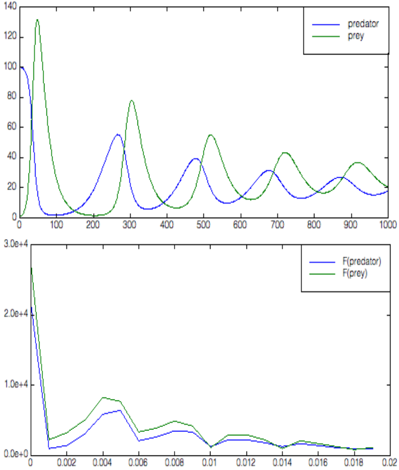

Unrelated Signals



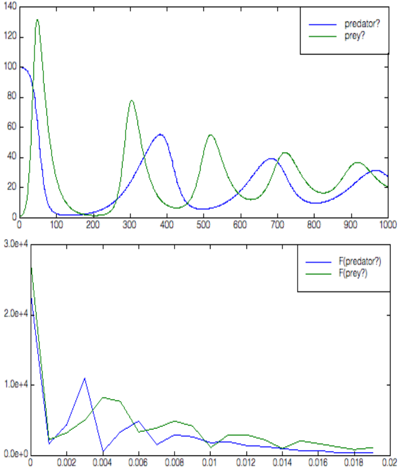
# Frequencies Show Feedback

## Feedback Relation



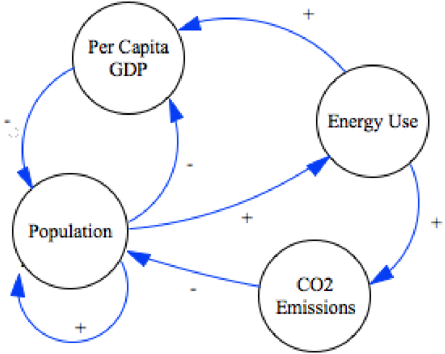




## Unrelated Signals

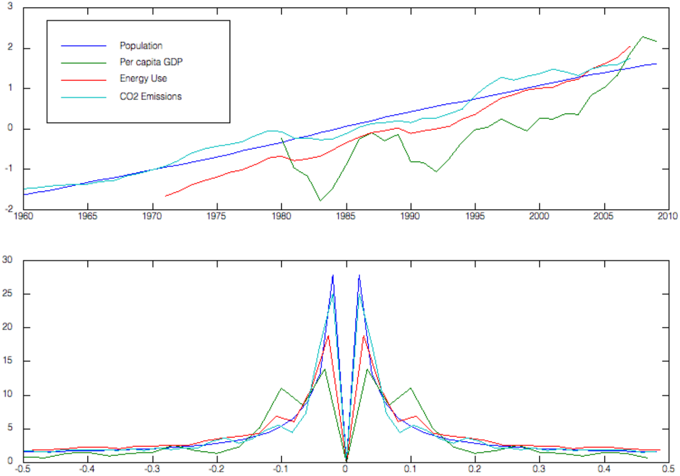



# Development and Energy in Brazil

How do population, income, and energy usage interact in the development of Brazil?



# Data Series


Population
Per capita GDP
Energy Use
CO2 Emissions
3
2
1
0
-1
-2
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
30
25
20
15
10
5
0
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5

# Agriculture and Development Results

- Multiple solutions exist (most of them are of the form $P = P$).
- No solution gives four unique equations (not sure why).

| | Population ($P$) | P.C. GDP ($Y$) | Energy Use ($E$) | $CO_2$ Emis. ($C$) |
|---|---|---|---|---|
| $P =$ | $2.858\underline{/-0.30}\ P$ | | $0.197\underline{/3.08}\ E$ | $1.455\underline{/3.00}\ C$ |
| $Y =$ | $1205\underline{/-2.37}\ P$ | $3\underline{/-0.18}\ Y$ | $80\underline{/-0.45}\ E$ | $1156\underline{/0.81}\ C$ |
| $E =$ | $11.162\underline{/-0.90}\ P$ | $.003\underline{/-1.98}\ Y$ | $1.310\underline{/0.81}\ E$ | $9.862\underline{/2.32}\ C$ |
| $C =$ | $1.69\underline{/-0.04}\ P$ | | $0.188\underline{/-2.70}\ E$ | |

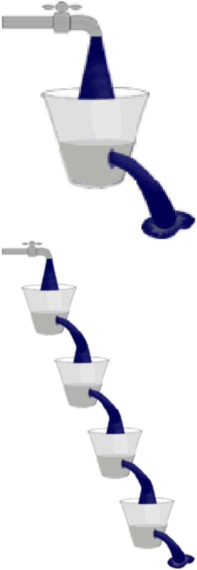
# Global Weather

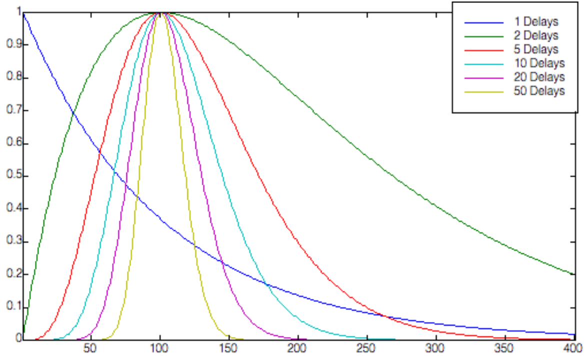
Does weather in one area predict weather in another?

Looking at 126 weather stations, with between 10-20 years of daily temperature data.

Model each region as related to 1-3 other stations, by first-order delays.

# First-Order Delays





# Weather Interconnections Map

# Example: La Guardia Airport Station

Causal Weather:

| Station | Delay | Fraction |
|---|---|---|
| Halifax, NS | 26.0476 | 57.1% |
| Bangalore, India | 0.1244 | 10.8% |

Caused Weather:

| Station | Delay | Percent | Other Causals |
|---|---|---|---|
| Boston, MA | 0.0014 | 79.9% | Beijing (16.1%) |
| Toronto, ON | 1.5549 | 66.4% | Riyadh (7.5%) |
| Tallinn, Estonia | 12.4923 | 74.1% | |
| Karachi, Pakistan | 2.3987 | 21.1% | Brussels (48.9%) |

# Transfer Functions

Any **Linear, Time-Invariant** (LTI) system can be described as a collection of **filters**.

Filters are characterized by their **impulse response** ($h(t)$), the output of the filter when an impulse ($\delta(t)$) is input.

The output of the filter to a general input, $x(t)$, is $y(t) = x(t) * h(t)$, defined as

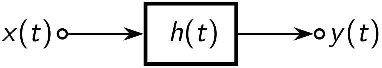
$$\int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$$

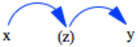
Alternatively, the filter can be characterized by its **frequency response** ($H(\omega) = \mathcal{F}h(t)$). Then the response of the filter to an arbitrary input is

$$Y(\omega) = X(\omega)H(\omega)$$

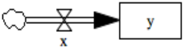
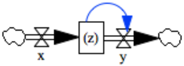
where $Y(\omega) = \mathcal{F}y(t)$, and $X(\omega) = \mathcal{F}x(t)$.

# Fourier Transform Properties

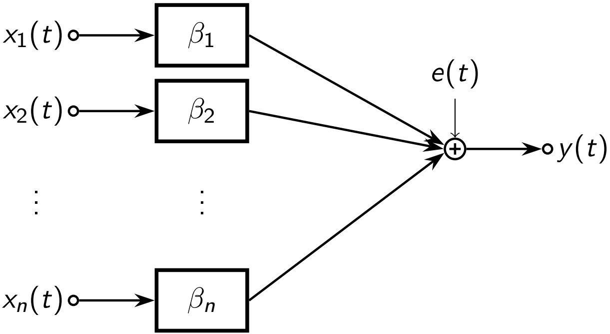



| property | time $(h(t) * x(t))$ | frequency $(H(\omega)X(\omega))$ |
|---|---|---|
| linearity  | $\beta_1 x(t) + \beta_2 z(t)$ | $\beta_1 X(\omega) + \beta_2 Z(\omega)$ |
| composition  | $h(t) * g(t) * x(t)$ | $H(\omega)G(\omega)X(\omega)$ |
| feedback  | $\alpha(L)y(t) = \theta(L)x(t)$ | $\frac{H(\omega)X(\omega)}{1+G(\omega)H(\omega)}$ |

# Relational Filters

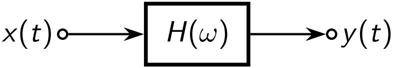
| relation | time $(h(t))$ | frequency $(H(\omega))$ |
|---|---|---|
| complex scaling<br> | $\beta$ | $2\pi\beta\delta(\omega)$ |
| integration<br> | $u(t)$ | $\pi\left(\frac{1}{i\pi\omega} + \delta(\omega)\right)$ |
| time delay<br> | $\delta(t - t_0)$ | $e^{-it_0\omega}$ |
| first order delay<br> | $e^{-t/\tau}$ | $\frac{\tau}{1+i\tau\omega}$ |
| moving average | $\mathbf{1}\{\|x\| < T_1\}$ | $\frac{2sin\omega T_1}{\omega}$ |

# Normal Regressions As Signal Processing

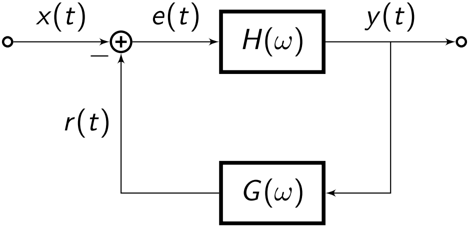


# Filters and Feedback Loops

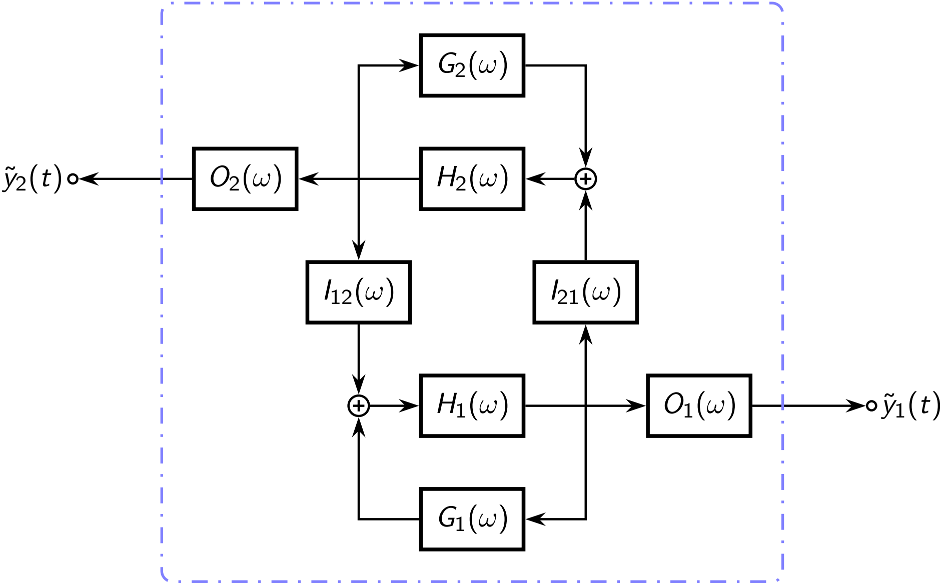
Simple (MA) Filter:



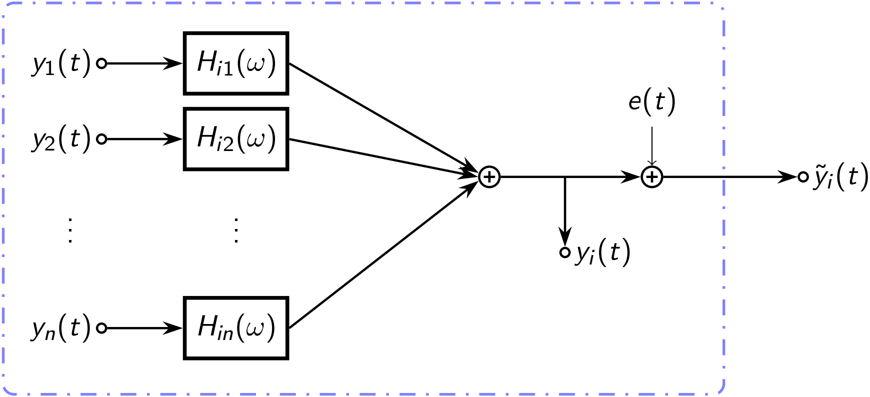

Feedback (ARMA) System:



# Feedback-Feedback Loops

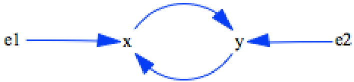


# System Kernel



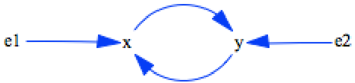
# Errors and System Dynamics

Autoregressive (Error-Dominated) System

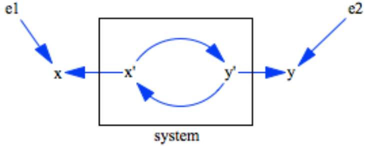


# Errors and System Dynamics

Autoregressive (Error-Dominated) System




System Dynamical (Self-Contained) System

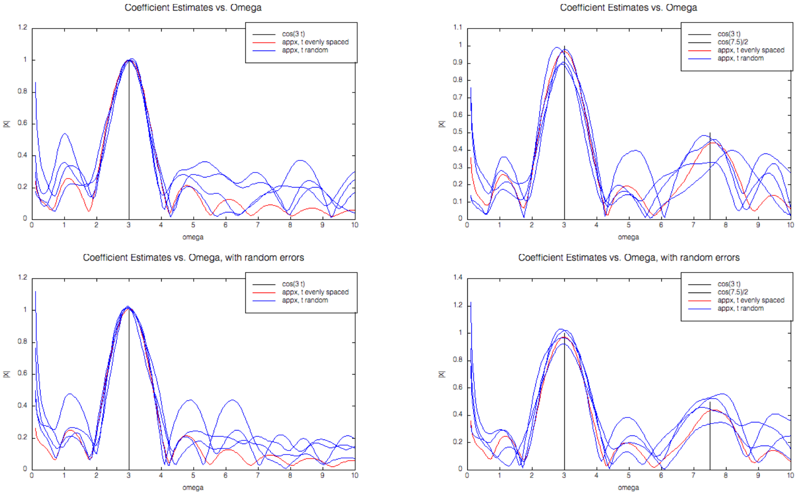


# Scattered Fourier Analysis

Let $y\{t\}$ be a continuous functions sampled at scattered points. What $Y(\omega)$ best approximates it?

- In general, $\frac{n}{2}$ unique frequencies can exactly reproduce $y\{t\}$, so how do we choose them?
- We may want to choose frequencies strategically to find interactions.

# Example of Identification Looseness



# Fundamental Equations

Let $\alpha_{ik}$ be fourier coefficient $\tilde{y}_i\{t\}$, measured at $\omega_k$. Define $\gamma_{ik}$ as the corresponding internal variable coefficient, so

$$\alpha_{ik} = \gamma_{ik} + \epsilon_{ik}$$

According to the diagram above, in the frequency domain,

$$\gamma_{ik} = \sum_j H_{ij}(\omega, \theta)\gamma_{jk}$$

where $H(\omega, \theta)$ has a supposed functional form, and $\theta$ is the parameter we want to estimate.

# An Eigenapproach

This can be rewritten

$$\lambda_k \vec{\gamma_k} = H(\omega, \theta)\vec{\gamma_k} \ \forall k$$

where $\lambda$ is added to allows different coefficients to be determined by the system up to a multiplicative constant; here $H(\omega, \theta)$ is $N \times K$, and $\vec{\gamma_k}$ is $N \times 1$.
This is an eigenequation, which has $N$ solutions. That is,

$$\gamma_k = c_k(\omega_k) v_{I(k)}(\theta)$$

where $I(k)$ maps every element $k$ to a value from 1 to $N$.
In other words, the $K$ vectors $\vec{\alpha_k}$ need to be partition into $N$ collections. The estimation can proceed within each partition independently.

# The Partitioning Problem

The number of ways to partitions of these values can get very large (called the Stirling numbers of the second kind and denoted $\left\{ n \atop k \right\}$).

| n\k | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| **2** | 1 | | | | | | | |
| **3** | 3 | 1 | | | | | | |
| **4** | 7 | 6 | 1 | | | | | |
| **5** | 15 | 25 | 10 | 1 | | | | |
| **6** | 31 | 90 | 65 | 15 | 1 | | | |
| **7** | 63 | 301 | 350 | 140 | 21 | 1 | | |
| **8** | 127 | 966 | 1701 | 1050 | 266 | 28 | 1 | |
| **9** | 255 | 3025 | 7770 | 6951 | 2646 | 462 | 36 | 1 |
| **10** | 511 | 9330 | 34105 | 42525 | 22827 | 5880 | 750 | 45 |
| **11** | 1023 | 28501 | 145750 | 246730 | 179487 | 63987 | 11880 | 1155 |
| **12** | 2047 | 86526 | 611501 | 1379400 | 1323652 | 627396 | 159027 | 22275 |
| **13** | 4095 | 261625 | 2532530 | 7508501 | 9321312 | 5715424 | 1899612 | 359502 |
| **14** | 8191 | 788970 | 10391745 | 40075035 | 63436373 | 49329280 | 20912320 | 5135130 |
| **15** | 16383 | 2375101 | 42355950 | 210766920 | 420693273 | 408741333 | 216627840 | 67128490 |

# Bias and Consistency

After finding the optimal partition, it is easy to get the estimated $\hat{H}$ (by choosing $\lambda$ so $\hat{H} \neq I_n$) and $\hat{\gamma}$, so

$$\hat{\gamma} = \hat{H}\hat{\gamma}$$

$$\alpha = \hat{\gamma} + \hat{\epsilon}$$

These can be rearranged to give

$$\hat{H}\alpha = H\alpha + (\hat{H} - I_n)\hat{\epsilon} - (H - I_n)\epsilon \implies$$

$$\hat{H} = H + \left((\hat{H} - I_n)\hat{\epsilon} - (H - I_n)\epsilon\right)\alpha'(\alpha\alpha')^{-1}$$

So, $\hat{H}$ is unbiased and consistent if $E(\hat{H} - I_n)\hat{\epsilon} = (H - I_n)\epsilon$.

# Bias and Consistency

Or, equivalently, if

$$E\hat{\epsilon}(\hat{\gamma}+\hat{\epsilon})' = E\epsilon(\gamma+\epsilon)'$$

By the model, $E\epsilon\gamma' = 0$, and it seems plausible that

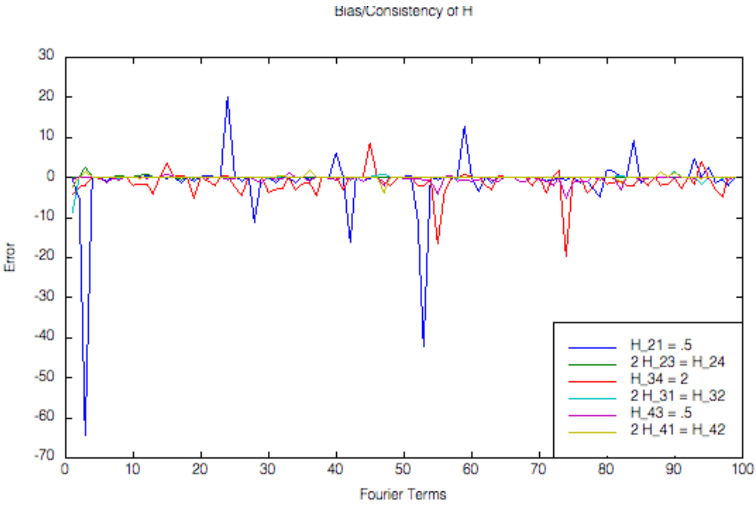
$$E\hat{\epsilon}\hat{\epsilon}' = E\epsilon\epsilon'$$

(that is, that $\sigma_\epsilon^2$ can be consistently estimated).
So, the remaining criteria is that

$$E(\hat{H} - I_n)\hat{\epsilon}\hat{\gamma} = 0$$

Rather than selecting $\lambda$ arbitrarily as above, we can select $\lambda$ to minimize this value, thereby minimizing the bias.

# Empirical Test



# Preview of Method

How It Works:

1. Scattered Fourier Analysis
2. Frequencies Partitioning
3. Eigenvalue Determination

# Preview of Method

How It Works:

1. Scattered Fourier Analysis
2. Frequencies Partitioning
3. Eigenvalue Determination

How You Work It:

```
data = load_data(17, 29);

[omegas, coeffs, four_errors] = sysfour(data, 1000, 10);
[hhs, scale_errors] = sysrel_scale(coeffs);
```

# Fourier Transforms

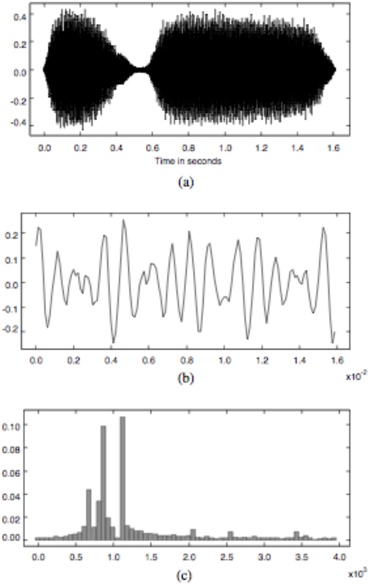



Figure 7.9: (a) A 1.6 second train whistle. (b) A 16 msec segment of the train whistle. (c) The Fourier series coefficients for the 16 msec segment.

# Fourier Transforms

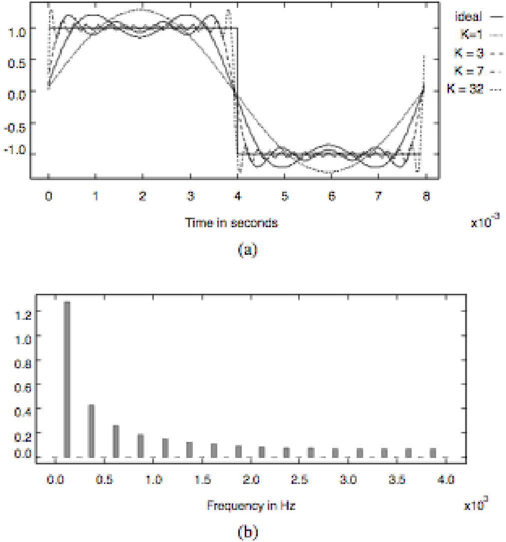



Figure 7.7: (a) One cycle of a square wave and some finite Fourier series approximations. (b) The amplitudes of the Fourier series terms for the square wave.

# Efficient Recomposition

A function represented by Fourier coefficients $\vec{Y}$ at scattered frequencies $\vec{\omega}$ can be calculated efficiently at scattered points $\vec{t}$ using

$$\vec{y} = \Re\{\vec{Y} e^{\vec{\omega}' \vec{t} i}\}$$

Similarly, coefficients $\vec{Y}$ can be determined independently (that is, not a full optimal combination) using

$$\vec{Y} = \vec{y} e^{-i \vec{t}' \vec{\omega}}$$

(all are row vectors).

# Least-Squares Solution

Consider a scattered function $y\{t\}$, and let $\vec{\omega}$ be a set of frequencies that will compose the Fourier approximation. The problem of Fourier approximation is to minimize the errors:

$$\min_{\vec{Y}} \sum_{i=1}^{n} (y_i - \Re\{\vec{Y} e^{\vec{\omega}' t_i i}\})^2 =$$

$$\min_{\vec{Y}, \vec{Y^\star}} \left( \vec{y} - \frac{1}{2} \left( \vec{Y} e^{\vec{\omega}' \vec{t} i} + \vec{Y^\star} e^{\vec{-\omega}' \vec{t} i} \right) \right) \left( \vec{y} - \frac{1}{2} \left( \vec{Y} e^{\vec{\omega}' \vec{t} i} + \vec{Y^\star} e^{\vec{-\omega}' \vec{t} i} \right) \right)'$$

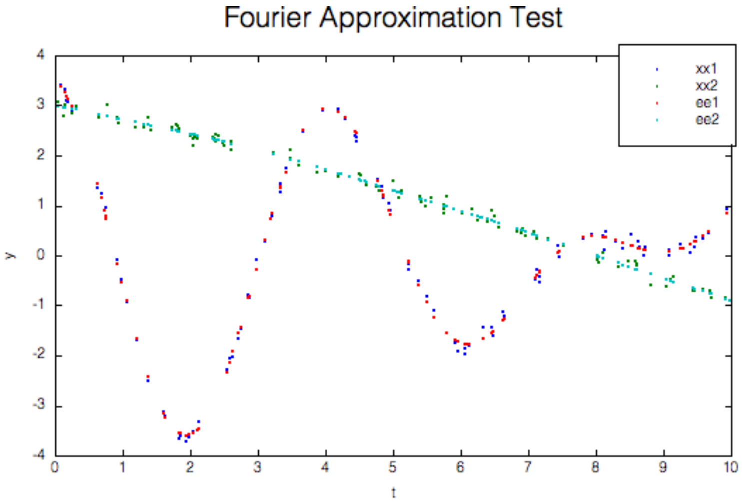
$$\Longrightarrow 2\vec{y} e^{\vec{t}' \vec{\omega} i} = \vec{Y} e^{\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i} + \vec{Y^\star} e^{-\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i}$$

$$\Longrightarrow \vec{Y^\star} = \left( 2\vec{y} e^{\vec{t}' \vec{\omega} i} - \vec{Y} e^{\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i} \right) (e^{-\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i})^{-1}$$

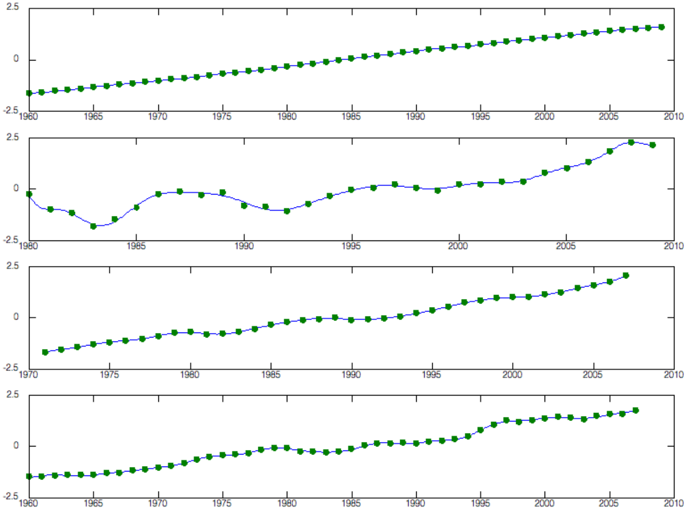
$$\Longrightarrow 2\vec{y} e^{-\vec{t}' \vec{\omega} i} = \left( 2\vec{y} e^{\vec{t}' \vec{\omega} i} - \vec{Y} e^{\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i} \right) (e^{\vec{-\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i})^{-1} (e^{-\vec{\omega}' \vec{t} i} e^{-\vec{t}' \vec{\omega} i})$$
$$+ \vec{Y} (e^{\vec{\omega}' \vec{t} i} e^{-\vec{t}' \vec{\omega} i})$$

$$\Longrightarrow 2\vec{y} \left( e^{-\vec{t}' \vec{\omega} i} - e^{\vec{t}' \vec{\omega} i} (e^{\vec{-\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i})^{-1} (e^{-\vec{\omega}' \vec{t} i} e^{-\vec{t}' \vec{\omega} i}) \right) =$$
$$\vec{Y} \left( e^{\vec{\omega}' \vec{t} i} e^{-\vec{t}' \vec{\omega} i} - e^{\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i} (e^{-\vec{\omega}' \vec{t} i} e^{\vec{t}' \vec{\omega} i})^{-1} (e^{-\vec{\omega}' \vec{t} i} e^{-\vec{t}' \vec{\omega} i}) \right)$$

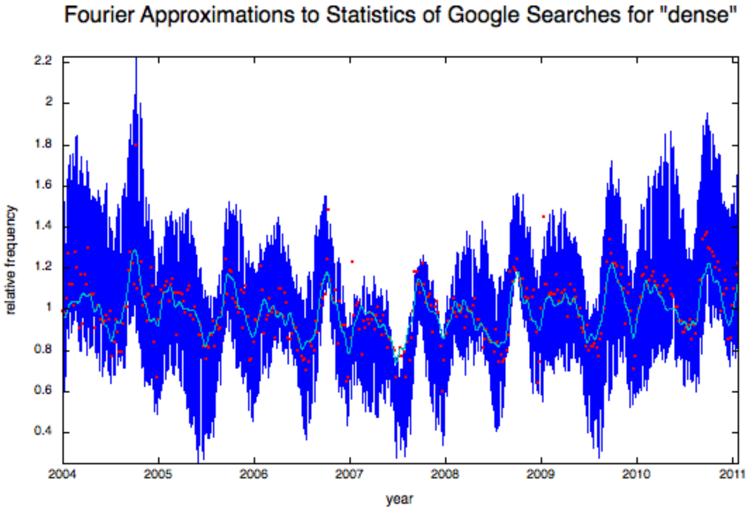
# Results: Identification



# Results: Brazilian Signals



# Results: Choosing Frequencies



# Solving the Partition's Equation

We minimize the least squares error within each partition to find $c_k$ and $v_i$ (assume these are constants, i.e. $H_{ij}(\omega) = H_{ij}$):

$$\epsilon_{ik} = \alpha_{ik} - c_k v_i \implies \mathcal{L} = \sum_i \sum_k (\alpha_{ik} - c_k v_i)(\alpha^\star_{ik} - c^\star_k v_i)$$

We define $c_k = a_k + ib_k$ and $v_i = x_i + iy_i$, and solve. The result is,

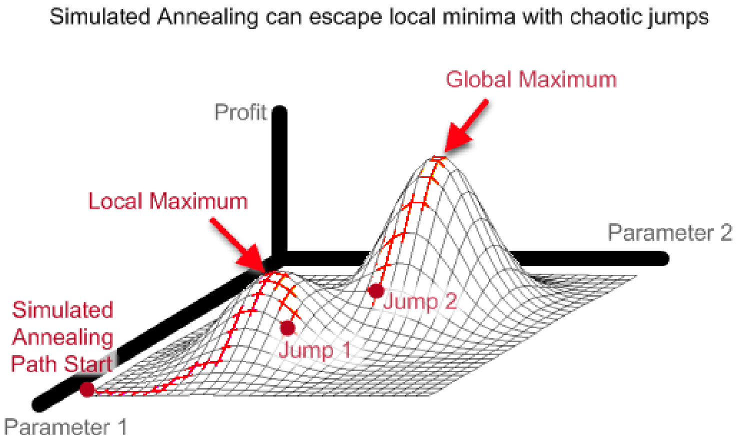
$$\frac{\vec{v}^{\star\prime}\left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)\vec{v}}{\vec{v}'\vec{v}^\star}\vec{v} = \left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)\vec{v}$$

This is another eigenequation. The left coefficient of $\vec{v}$ is the Rayleigh quotient, equal to the eigenvalue of $\left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)$.

Furthermore, for small errors, the matrix $\left(\sum_k \vec{\alpha_k}\vec{\alpha_k}^{\star\prime}\right)$ will be nearly rank 1, and have only one eigensolution that is not near 0. Let this $\vec{v}$ be called the **characteristic vector**. Then $c_k = \frac{\vec{\alpha_k}'\vec{v}}{\vec{v}'\vec{v}^\star}$.

# Finding Optimal Partitions

Problem: There are many local optima in the space of possible partitions, but we want the global optimum (or something near it).
Solution: Simulated Annealing (a tunneling algorithm)



# Partitioning Algorithm

1. Propose an initial partition, $V^\star$.
2. Calculate the characteristic vector, and sum the errors. Call the total error $e^\star$.
3. Set the temperature to an initial value, $T$. Begin the refining loop:
4. Propose an exchange of vectors between two randomly chosen partitions, generating a new $V'$.
5. Calculate the new characteristic vectors and total error, $e'$
6. If $e' < e^\star$, or a random number $r \in [0, 1]$ satisfies $r < e^{-\frac{e^\star - e'}{T}}$, then set $e^\star \leftarrow e'$, $V^\star \leftarrow V'$.
7. Decrease temperature slightly and repeat from step 4 until satisfied.

# Recovering $\lambda$

To recover $H$, note that for any matrix with eigen vectors $(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) = V$,

$$H = V\Lambda V^{-1}$$

We can use $\lambda$ to set up to $n-1$ arbitrary elements of the $H$ matrix to 0 (thereby getting $n$ unique relationships):

$$H_{ij} = \sum_k \left( \sum_l V_{il}\Lambda_{lk} \right) (V^{-1})_{kj} = \sum_k (V_{ik}\lambda_k)(V^{-1})_{kj}$$

$$= \sum_k (V^{-1'})_{jk} V_{ik}\lambda_k$$

At least one value in $H_{ij}$ must be set to 1, and more values are required if some elements have no interdependence on other elements.

# Generalizing to $H(\omega)$

We assumed $H(\omega) = H$, a constant, to determine the partition's characteristic vectors. To apply this method to a different filter, we require that $H(\omega)$ can be decomposed into two parts:

$$H(\omega, \theta) = C(\omega)V(\theta)$$

For example, for the first-order delay filter,

$$H(\omega) = \frac{\alpha\tau}{\omega\tau i - 1} = \left(\frac{1}{\omega\bar{\tau} i - 1}\right)(\alpha\bar{\tau})$$

If $\tau = \bar{\tau}$, a known parameter.

# Next Steps

- Determine how to interpret results.
- Determine variance on error and bias.
- Determine efficient method of selecting $\lambda$.

# Multiple Takes

Suppose that there is a single underlying system which determines the behavior of variables in many regions– but perhaps with some different parameters. How do I deal with that?
For example, economic development might be assumed to have the same underlying system in different countries, but with strengths specific to each. This might also allow the method to construct all-endogenous estimates for returns to schooling.

# The Problem

An infinite number of Fourier representations $Y\{\omega\}$ can reproduce the values of $y\{t\}$.

Methods of Choosing $\{\omega\}$:

- Evenly spaced $\{\omega\}$, simultaneous solution $\longleftarrow$ imprecise
- Randomly select $\{\omega\}$, simultaneous solution, repeat indefinitely $\longleftarrow$ unreliable

# The Problem

An infinite number of Fourier representations $Y\{\omega\}$ can reproduce the values of $y\{t\}$.
Methods of Choosing $\{\omega\}$:

- Evenly spaced $\{\omega\}$, simultaneous solution $\longleftarrow$ imprecise
- Randomly select $\{\omega\}$, simultaneous solution, repeat indefinitely $\longleftarrow$ unreliable
- Identify peak frequency, subtract off, and repeat $\longleftarrow$ may mis-identify peaks near each other
- Demean data, identify next peak on left and subtract off, and repeat $\longleftarrow$ may mis-identify ridges leading to peaks

# The Advantage

By most means, different signals will suggest different frequencies. But when different signals exhibit nearly identical frequencies, we want to try to use the same frequency for both.
Methods of Reconciling between Signals:

- Combine all frequencies, select a random new, determine its optimal errors, and compare to other selections.
- For each signal, switch out single frequencies for those identified in other signals, and see if these frequencies are selected.

# Questions?

To Athena, with an incense of aromatic herbs:
You who put a dance in the heart and glory in embattlements,
You can put the sting of mania into a mortal soul!
Athletic Maiden with a heart sublime,
Slayer of the Gorgon, fugitive of the bridal bed,
Mother of Art in all your abundance, catalyst of progress!
You bring folly to the corrupt and a sense of purpose to the pure!

Exit the edict of exogeneity;
Enter the era of inward tranquility.

Eliminate your erroneous instruments, and find a new frequency,
Before perspective and prejudice can take on any parity.