Cosmological time delay estimation with Continuous Auto-Regressive Moving Average processes

Antoine D. Meyer (antoine.meyer@cfa.harvard.edu - adm18@ic.ac.uk)
Statistics Section, Department of Mathematics, Imperial College London
Predoc at CfA
In collaboration with David van Dyk, Hyungsuk Tak, Aneta Siemiginowska

CHASC
19 April 2022
The gravitational lensing phenomenon

- **Light rays are bent** by the strong gravitational field of the intervening object (i.e. the lens) on their way to Earth.
- Different rays travel along different paths of different distances, and therefore **arrive to the observer at different times** (separated by a delay $\Delta$).

![Diagram of gravitational lensing](image)

**Figure:** Source: Tak et al. [2017]

- Multiple copies of the original light curve, **brightness fluctuations are observed at different times** in the different copies.
- Refsdal [1964]: Estimates of time delays can be used to **constrain important cosmological parameters** such as $H_0$. 
A challenging statistical problem

Goal: estimate the time shift (x-axis) between multiple lensed light curves

Figure: Source: Tak et al. [2017]

Data is subject to:
- Irregular sampling (observational patterns)
- Seasonal gaps (celestial cycles)
- Brightness magnifications due to multiple effects (strong lensing, micro-lensing)
- Measurement errors (heteroskedastic)
Current methods:

**Two families of methods for time delay estimation:**

- Grid-based optimization methods (≈ non-parametric)
  - Minimize measure of distance between light curves, on a grid of $\Delta$ values
  - Produce uncertainty estimates using Monte Carlo simulation
  - Computationally expensive!

- Statistics modelling the stochastic variability of AGN light curves
  - More principled
  - "Direct" quantification of uncertainty in time delay parameter

**Time Delay Challenge (TDC, Liao et al. [2015]) surveyed and evaluated a variety of time delay estimation techniques**

- Winners: COSMOGRAIL collaboration (combine estimates from 4 different non-parametric techniques)
- Tak et al. [2017] did very well too! Bayesian method + DRW process to model AGN variability
Goals of this project

Goals of this project:
- Improve Bayesian time delay estimation method, based on Tak et al. [2017]
- Improve on the accuracy and applicability of time delay estimation method
- Address some of the computational limitations of current estimation strategies.

Current limitations:
- (Tak et al. [2017]) limited applicability of the DRW process to model AGN light curves.
- Difficult computation, in part due to multi-modality issues
- High sensitivity to initial guess for $\Delta$, most methods require method to compute plausible/starting value of $\Delta$. 
Our contributions

Limitation: use of the DRW to model AGN light curves restricts the range of observations on which time delay method can be applied

- Recent success in modeling AGN light curves with flexible CARMA\((p, q)\) in astrophysics literature [Kelly et al., 2014, Moreno et al., 2019].
- AGNs and particularly quasars are sources for which strong gravitational lensing is more likely to happen
- Generalization of DRW, which is a CARMA\((1,0)\).

**Main development:** update the intrinsic light curve model from DRW process to flexible **Continuous Auto-Regressive Moving Average (CARMA) process**

- Finer modeling tool can better fit a wider range of observations.
- Better fit to the lightcurve data → better accuracy

Our method: **TD-CARMA**

Tak et al. [2017]: **TD-DRW**
Contributions

**Additional development:** update the parameter inference algorithm from MCMC [Tak et al., 2017] to nested sampling (MultiNest)

- Deal with **multi-modality** of CARMA($p$, $q$) and time delay parameters
  - MultiNest identifies multiple modes in posterior distribution
  - MultiNest output can be used to quantify relative probability of the modes

- **Blind Search:** no initial value of $\Delta$ required
  - Most existing time delay methods are highly sensitive to initial values
  - Require an extra method to compute a plausible initial value
  - Tak et al. [2017] compute the expensive profile likelihood

- **Model Selection:** MultiNest estimates the Bayesian evidence (marginal distribution of the data) at no extra computational cost
Data

\[ D = \{ t_i, x_i, \delta^x_i, y_i, \delta^y_i \}_{i=1}^n \]

- Observation times \( t = \{ t_1, \ldots, t_n \} \)
- Observed magnitudes \( x = \{ x_1, \ldots, x_n \} \), and \( y \)
- Measurement errors \( \delta^x = \{ \delta^x_1, \ldots, \delta^x_n \} \) and \( \delta^y \) (standard deviation).

**Figure:** Source: Tak et al. [2017]
Time Delay Estimation Framework

Assumptions of the time delay model:

- Assumption I: \( x \) and \( y = \) discrete realizations of unobserved continuous lightcurves \( x(t) \) and \( y(t) \) (true source magnitudes), \( t \in \mathbb{R} \).
- Assumption II: \( y(t) \) is time and magnitude shifted version of \( x(t) \).

Let’s translate them into our model:

1. Strong lensing effect 1: Time shift

\[
y(t) = x(t - \Delta)
\]  

2. Strong lensing effect 2: different average magnitudes

\[
y(t) = x(t - \Delta) + \theta_0
\]  

3. Micro lensing effect: extrinsic long-term variability

\[
y(t) = x(t - \Delta) + w_m(t - \Delta)\theta
\]

- \( w_m(t - \Delta) := \{1, t - \Delta, \ldots, (t - \Delta)^m\} \) vector of polynomial time variables
- \( \theta = \{\theta_0, \ldots, \theta_m\} \) micro-lensing coefficients
**Time Delay Estimation Framework**

**Plan:**

1. **Reconstruct intrinsic light curve** from its lensed counterparts
2. Model stochastic variability in intrinsic light curve with CARMA process

**Reconstructing the intrinsic light curve:**

- Given $\Delta$ and $\theta$, construct measurements for the intrinsic light curve, denoted $z = \{z_j\}_{j=1}^{2n}$.
- Measured at times $t^\Delta = \{t_i\}_{i=1}^n \cup \{t_i - \Delta\}_{i=1}^n$.

![Graph showing light curve measurements](image-url)
Time Delay Estimation Framework

\[
z_j = \begin{cases} 
  x_i & \text{for some } i \text{ if } t_j^\Delta \text{ is in } t, \\
  y_i - \mathbf{w}_m(t_j - \Delta)\theta & \text{for some } i \text{ if } t_j^\Delta \text{ is in } t - \Delta,
\end{cases}
\]  

(4)

Similarly, for the vector measurement error standard deviations \( \{\delta^z_i\}_{i=1}^{2n} \):

\[
\delta^z_j = \begin{cases} 
  \delta^x_i & \text{for some } i \text{ if } t_j^\Delta \text{ is in } t, \\
  \delta^y_i & \text{for some } i \text{ if } t_j^\Delta \text{ is in } t - \Delta.
\end{cases}
\]  

(5)
Time Delay Measurement Framework

We constructed our intrinsic light curve:

- Given $\Delta$ and $\theta$, we have $\{t^\Delta, z, \delta z\}$ for the intrinsic light curve.
- Assumption: $z$ = discrete realization of unobserved continuous process $z(t)$.

Let’s model the stochastic variability in the intrinsic lightcurve:

- Parametric model, with parameter vector $\Omega$, i.e. define $p(z|\Delta, \theta, \Omega)$
- Tak et al. [2017]: CARMA(1,0) (DRW) process.
- We generalize this to CARMA($p, q$) processes.

Likelihood function of model parameters $(\Delta, \theta, \Omega)$ given observed data

$$L(\Delta, \theta, \Omega) = p(x, y|\Delta, \theta, \Omega)$$
$$= p(z|\Delta, \theta, \Omega)$$ (6)
DRW and CARMA\((p, q)\) processes

- A **Damped Random Walk (DRW)** process with mean \(\mu\) is the solution to the following stochastic differential equation:

\[
dX(t) = -\frac{1}{\omega}(X(t) - \mu)dt + \epsilon(t)
\]  

- \(\epsilon(t) \sim N(0, \sigma^2)\) is a white noise process.
- \(\omega = \text{timescale for the process to revert to its long-term mean}\)
- \(\text{DRW} = \text{CARMA}(1,0)\)

- A **CARMA\((p, q)\)** process is the solution to the following stochastic differential equation:

\[
d^p y(t) + \alpha_{p-1} d^{p-1} y(t) + \cdots + \alpha_0 y(t) = \beta_q d^q \epsilon(t) + \cdots + \epsilon(t)
\]  

- \(\alpha = \{\alpha_0, \ldots, \alpha_{p-1}\} = \text{auto-regressive parameters}\)
- \(\beta = \{\beta_0, \ldots, \beta_{q-1}\} = \text{moving-average coefficients}\)
DRW and CARMA($p, q$) processes

In time-space (Auto-Correlation Function):

- **DRW auto-correlations**: single exponentially decaying auto-correlation function

- **CARMA auto-correlations**: weighted sum of exponentially decaying auto-correlation functions and exponentially damped sinusoidal functions

In frequency-space (Power Spectral Density)

- **CARMA**: multiple breaks and frequencies (QPOs) in the PSD
DRW and CARMA: Auto-Covariance Functions

ACF of a DRW process: \((t_i > t_j)\)

\[
R(t_i - t_j) = \frac{\sigma^2 \omega}{2} e^{-\frac{(t_i-t_j)}{\omega}}
\]  

(9)

- exponentially decaying with e-folding timescale \(\omega\)

ACF of a CARMA\((p, q)\) process: \((t_i > t_j)\)

\[
R(t_i - t_j) = \sigma^2 \sum_{k=1}^{p} \left[ \sum_{l=0}^{q} \beta_l r_k^l \right] \left[ \sum_{l=0}^{q} \beta_l (-r_k)^l \right] \exp \left( r_k \tau \right) - 2\Re(r_k) \prod_{l=1, l \neq k} \left( r_l - r_k \right) \left( r_l^* + r_k \right). 
\]  

(10)

- \(r_k\) = roots of auto-regressive polynomial \(A(z) = \sum_{k=0}^{p} \alpha_k z^k\)
- ACF of CARMA is a weighted sum of:
  - exponentially decaying components (when \(r_k\) is real)
  - exponentially damped sinusoids (when \(r_k\) is complex)
- Enforce \(\Re(r_k) < 0\) for stationarity
DRW and CARMA: Power Spectrum Density

PSD of a DRW process:

\[
P(f) = \sigma^2 \frac{1}{(\frac{1}{\omega})^2 + (2\pi f)^2}
\]  

- Lorentzian centered at 0, with a break frequency at \(1/2\pi \omega^2\)

PSD of a CARMA process:

\[
P(f) = \sigma^2 \left| \sum_{j=0}^{q} \beta_j (2\pi if)^j \right|^2 \frac{\left| \sum_{k=0}^{p} \alpha_k (2\pi if)^k \right|^2}{\left| \sum_{k=0}^{p} \alpha_k (2\pi if)^k \right|^2}
\]  

- Weighted sum of Lorentzian functions
- Lorentzian centered at 0 → break frequency (when \(r_k\) is real)
- Lorentzian centered away from 0 → Quasi-Periodic Oscillation (QPO, when \(r_k\) is complex).

Characteristics such as multiple break-like features and QPOs have been observed in AGN optical data Kelly et al. [2014], Ryan et al. [2019]
Likelihood

**CARMA and DRW process are Gaussian → $p(z|\Delta, \theta, \Omega)$ is Gaussian**

- Typically, the computation of the likelihood of an $n$-point realization of a Gaussian process requires the inversion of an $n \times n$ covariance matrix → scales $O(n^3)$.

\[
L(\Delta, \theta, \Omega) = p(z|\Delta, \theta, \Omega) \\
= \prod_{i=1}^{2n} p(z_i|z_{<i}, \Delta, \theta, \Omega) \\
\propto \prod_{i=1}^{2n} \frac{1}{\text{Var}(z_i|z_{<i}, \Delta, \theta, \Omega)} \times \exp\left(-\frac{1}{2} \left(\frac{z_i - \mathbb{E}(z_i|z_{<i}, \Delta, \theta, \Omega)}{\text{Var}(z_i|z_{<i}, \Delta, \theta, \Omega)}\right)^2\right)
\]

- But CARMA processes are special!
- They admit a linear state-space representation that allows to compute the likelihood in linear time $O(n)$!
Computing the likelihood: state-space representation

Linear state-space representation of a CARMA\((p,q)\) process denoted by \(z(t)\) is:

\[
\begin{align*}
    z(t) &= bx(t) + \delta(t), \\
    dx(t) &= Ax(t)dt + edW(t)
\end{align*}
\]

- Latent state process \(x(t)\) governs the underlying dynamics of the system
- Observation equation gives relationship between latent process \(x\) and observed process \(z(t)\).
- \(\delta(t)\) measurement error process.

**Kalman Filter algorithm** efficiently computes \(\{E(z_i|z_{<i}, \Delta, \theta, \Omega)\}_{i=1}^n\) and \(\{\text{Var}(z_i|z_{<i}, \Delta, \theta, \Omega)\}_{i=1}^n\) with linear complexity \(O(n)\).

- CARMA processes modelling is scalable to large datasets
Bayesian Inference

We operate under the Bayesian paradigm:

- Quantify the uncertainty in model parameters via their (joint) posterior distribution
- Bayes’ Theorem:

\[
p(\Delta, \beta, \Omega | D) = \frac{L(\Delta, \beta, \Omega)p(\Delta, \beta, \Omega)}{p(D)}
\]  
(14)

- \(p(\Delta, \beta, \Omega | D)\) \rightarrow posterior distribution
- \(p(D) \equiv Z\rightarrow marginal\) distribution of the data (can be used for Bayesian model comparison).
- \(p(\Delta, \beta, \Omega)\) \rightarrow prior distribution (we choose uniform priors)

Prior distributions:

- \(\Delta \sim [t_1 - t_n, t_n - t_1], \mu \sim [-30, 30], \theta \sim [-M, M]\) with \(M\) large
- \(\alpha, \beta, \sigma\) sampled on log-scale, \([-15, 15]\).
Posterior sampling

We produce a sample of the posterior distribution using the MultiNest implementation of Nested Sampling (NS).

- Standard MCMC have trouble sampling from multi-modal posteriors
- MultiNest is designed to sample from multi-modal posterior posteriors

Most time delay estimation techniques are highly sensitive and require the input of an initial guess for $\Delta$

- Need auxiliary method or prior knowledge to find plausible initial value for $\Delta$
- Tak et al. [2017] compute expensive profile likelihood to find good initial guess.

MultiNest does not require the input of an initial value

- You just need to specify the boundaries of the parameter space!
- Method is blind search and standalone
Bayesian Model Selection

Model selection problem:
- what is the \((p, q, m)\) triplet that best fits the data?
- CARMA\((p, q)\) models are non-nested.

MultiNest evaluates the Bayesian evidence \(Z\):

\[
Z = \int_{\Theta} p(D|\theta)p(\theta)d\theta
\]  

(15)

- \(Z\) is a measure of the "goodness of fit" of a model to the data \(D\).
- **Incorporates Occam’s razor:** more complex models, i.e. models defined on higher-dimensional parameter spaces, are penalized if they do not sufficiently improve the fit to the data.
- No extra computational cost: \(Z\) is computed jointly to the posterior sampling
Tackling multi-modality with MultiNest

Two multi-modality issues:

- CARMA($p, q$) parameters $\alpha, \beta, \sigma$
- Time Delay parameter $\Delta$

Multi-modal posterior distributions are difficult to sample from. But MultiNest can help!

- MultiNest identifies modes in the posterior distribution
- Partitions the parameter space into regions $\Theta_i$ on which the separated modes are supported
- Evaluates the local-evidence $Z_i$ of each mode $i$, defined as:

$$Z_i = \int_{\Theta_i} p(D|\theta)p(\theta)d\theta.$$ (16)

- Compute the relative probability $p_i$ of mode $i$:

$$p_i = \int_{\Theta_i} p(\theta|D)d\theta = \frac{1}{Z} \int_{\Theta_i} p(D|\theta)p(\theta)d\theta = \frac{Z_i}{Z}.$$
Time Delay Challenge dataset
Generated under the DRW model

Figure: Doubly-lensed simulated quasar dataset from Time Delay Challenge (TDC)
Time Delay Challenge (TDC) results

- We fit TD-CARMA\((p, q, m)\) with \(p = \{2, 3, 4\}, q = \{0, 1, 2, 3\}\) and \(m = \{1, 2, 3\}\).

<table>
<thead>
<tr>
<th>Model</th>
<th>(\hat{\Delta})</th>
<th>SD((\hat{\Delta}))</th>
<th>ln((Z))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Truth</strong></td>
<td>5.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-DRW(3) [Tak et al., 2017]</td>
<td>6.33</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>TD-DRW(3) (This work)</td>
<td>6.343</td>
<td>0.267</td>
<td>692.32</td>
</tr>
<tr>
<td>TD-CARMA((4, 3, 2))</td>
<td>6.296</td>
<td>0.255</td>
<td>699.11</td>
</tr>
<tr>
<td>TD-CARMA((4, 2, 2))</td>
<td>6.282</td>
<td>0.252</td>
<td>698.98</td>
</tr>
<tr>
<td>TD-CARMA((4, 1, 2))</td>
<td>6.287</td>
<td>0.259</td>
<td>698.49</td>
</tr>
<tr>
<td>TD-CARMA((2, 1, 2))</td>
<td>6.254</td>
<td>0.230</td>
<td>698.44</td>
</tr>
<tr>
<td>TD-CARMA((2, 0, 2))</td>
<td>6.269</td>
<td>0.242</td>
<td>698.39</td>
</tr>
</tbody>
</table>

**Table**: Posterior mean and standard deviation for \(\Delta\) under the five models with highest Bayesian log-evidence, comparing to true value and estimates from DRW methods, as reported by Tak et al. [2017] and computed by our own code.
Dataset Application - HS2209

HS 2209 +1914

Magnitude (relative)

HJD - 2400000.5 [day]


COSMOGRAIL.org
HS2209: Results

<table>
<thead>
<tr>
<th>Technique</th>
<th>Reference</th>
<th>$\hat{\Delta}$</th>
<th>SD($\hat{\Delta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined estimate (COSMOGRAIL)</td>
<td>Eulaers et al. [2013]</td>
<td>-20.0</td>
<td>5</td>
</tr>
<tr>
<td>Difference-smoothing (modified)</td>
<td>Kumar et al. [2015]</td>
<td>-22.9</td>
<td>5.3</td>
</tr>
<tr>
<td>$\Delta$CARMA(3, 2, 3)</td>
<td>This work</td>
<td>-21.96</td>
<td>1.448</td>
</tr>
<tr>
<td>$\Delta$CARMA(2, 1, 3)</td>
<td>This work</td>
<td>-21.74</td>
<td>1.423</td>
</tr>
</tbody>
</table>
DRW process on HS2209 data

- Inconclusive results using *timedelay* package from Tak et al. [2017].
- With our code, MultiNest finds 10 modes in the posterior distribution of $\Delta$. Modes for $\Delta$ include $[-14.30, -11.72, 16.44, 17.50, 20.38, 43, 64]$ days.

![Graph showing the posterior distribution of $\Delta$](image)

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\Delta}$</th>
<th>SD($\hat{\Delta}$)</th>
<th>ln($Z$)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD-CARMA(3, 2, 3)</td>
<td>-21.96</td>
<td>1.448</td>
<td>2760.24</td>
<td>0.601</td>
</tr>
<tr>
<td>TD-CARMA(4, 2, 3)</td>
<td>-21.95</td>
<td>1.403</td>
<td>2759.83</td>
<td>0.399</td>
</tr>
<tr>
<td>TD-CARMA(2, 1, 3)</td>
<td>-21.74</td>
<td>1.423</td>
<td>2752.52</td>
<td>$4.4 \times 10^{-4}$</td>
</tr>
<tr>
<td>TD-DRW(3)</td>
<td>20.23</td>
<td>0.934</td>
<td>2536.03</td>
<td>$4.23 \times 10^{-98}$</td>
</tr>
</tbody>
</table>
SDSS J1001+5027 doubly lensed quasar

SDSS J1001 +5027

A shifted by -119.3 days and -0.25 mag

B

Mag A

2005 2006 2007 2008 2009 2010 2011

53500 54000 54500 55000 55500

HJD - 2400000.5 [day]

Mercator : 239 epochs
HCT : 143 epochs
Maidanak : 61 epochs
J1001: Multi-modality of CARMA parameters

- Two modes in the posterior distribution of CARMA parameters are identified by MultiNest.
- One mode corresponds to a frequency in the PSD ($f = 2$), but this frequency falls below the measurement noise level (so we are discarding modes/models that feature the frequency).

Detection of frequency can dramatically reduce uncertainty in time delay $\Delta$ ($SD(\hat{\Delta}) = 0.686$ without freq, 0.224 with) $\rightarrow$ but only if we believe the frequency exists!
J1001: Results

<table>
<thead>
<tr>
<th>Technique</th>
<th>Reference</th>
<th>$\hat{\Delta}$</th>
<th>SD($\hat{\Delta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined estimate (COSMOGRAIL)</td>
<td>Kumar et al. [2013]</td>
<td>119.1</td>
<td>3.3</td>
</tr>
<tr>
<td>Gaussian Processes</td>
<td>Hojjati et al. [2013]</td>
<td>117.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Difference-smoothing (modified)</td>
<td>Kumar et al. [2015]</td>
<td>119.7</td>
<td>1.8</td>
</tr>
<tr>
<td>$\Delta$CARMA(2,1,3)</td>
<td>This work</td>
<td>120.18</td>
<td>0.749</td>
</tr>
<tr>
<td>$\Delta$CARMA(4,3,2)</td>
<td>This work</td>
<td>120.93</td>
<td>1.015</td>
</tr>
</tbody>
</table>

**Technique**

- Combined estimate (COSMOGRAIL)
- Gaussian Processes
- Difference-smoothing (modified)
- $\Delta$CARMA(2,1,3)
- $\Delta$CARMA(4,3,2)
DRW process on J1001 data

- Inconclusive results using *timedelay* package from Tak et al. [2017].
- With our code, MultiNest finds 20 modes in the posterior distribution of $\Delta$. Modes for $\Delta$ include $[122.8, 127.6, 130.5, 132.8]$ days.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\Delta}$</th>
<th>SD($\hat{\Delta}$)</th>
<th>$\ln(Z)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD-CARMA(4, 3, 2)</td>
<td>120.93</td>
<td>1.015</td>
<td>2761.25</td>
<td>0.416</td>
</tr>
<tr>
<td>TD-CARMA(2, 1, 3)</td>
<td>120.18</td>
<td>0.749</td>
<td>2744.05</td>
<td>$4.1 \times 10^{-8}$</td>
</tr>
<tr>
<td>TD-OU(3)</td>
<td>132.71</td>
<td>0.750</td>
<td>1803.24</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Future Research

Improvements of the method:
- Speed up likelihood computation using *celerite* model [Foreman-Mackey et al., 2017] (same complexity).
- Multi-band light curves?

Applications of the method:
- $H_0$ estimation
- Time delays arising in reverberation mapping

THANK YOU!


