# Intervention time-series model using transfer functions


**Xingwu Zhou & Nicola Orsini**

Biostatistics team, Department of Public Health Sciences


Sept 1, 2017

# Introduction

- I will present a Stata command **tstf** to estimate the intervention time series with transfer functions.
- The method has been described by Box and Tiao (1975, JASA).
- Estimation, inference, and graphs will be given for both the original data and the log-transformed data.
- The method will be illustrated using the Swedish National Tobacco Quitline (SRL)

# Why time series design?

# Why time series design?

- RCT: Randomized controlled trials are considered the ideal approach for assessing the effectiveness of interventions.

# Why time series design?

- RCT: Randomized controlled trials are considered the ideal approach for assessing the effectiveness of interventions.
- Time series analysis is a quasi-experimental design useful to evaluate the longitudinal effects of interventions on a population level.

# Why time series design?

- RCT: Randomized controlled trials are considered the ideal approach for assessing the effectiveness of interventions.
- Time series analysis is a quasi-experimental design useful to evaluate the longitudinal effects of interventions on a population level.
- Intervention time series analysis is widely used in areas like finance, economics, labor markets, transportation, public health and so on .

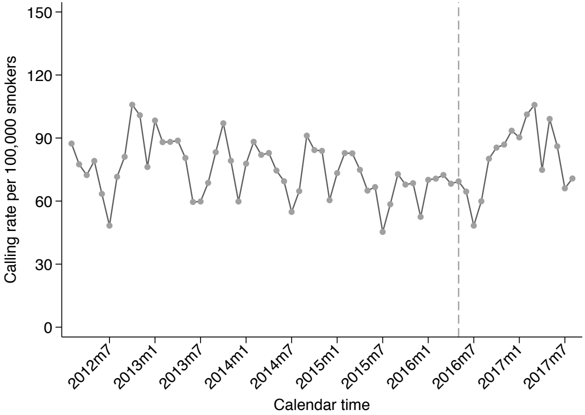
# SRL


The Swedish National Tobacco Quitline (SRL) established in 1998 is a nationwide, free service, providing telephone counseling for tobacco users who want to quit the habit.

# SRL continue


According to Swedish proposition 2015/16:82, started from May 2016, new cigarette packages sold in Sweden will have to display pictorial warnings together with text warnings and the SRL telephone number.

# SRL, Calling per 100,000 smokers



# SRL continue

- There is a need to understand how this new policy will affect the care-seeking behavior of Swedish tobacco users.

# SRL continue

- There is a need to understand how this new policy will affect the care-seeking behavior of Swedish tobacco users.
- To which extent this measure has been effective in inducing a behavioral change among Swedish tobacco users is not known.

# SRL continue

- There is a need to understand how this new policy will affect the care-seeking behavior of Swedish tobacco users.
- To which extent this measure has been effective in inducing a behavioral change among Swedish tobacco users is not known.
- A change in the inflow of calls received at the quitline may be used as an estimator of the population impact of policy measures.

# What already have in Stata?

Stata package **itsa** analyses interrupted time series using segmented regression.

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 XT_t \tag{1}$$

$\beta_0$ represents the baseline level at $T = 0$,
$\beta_1$ is interpreted as the change in outcome associated with a time unit increase
$\beta_2$ is the level change following the intervention
$\beta_3$ indicates the slope change following the intervention.

shortcomings: hard to reduce the auto-correlation among the residuals.

# Intervention time series

The intervention time series model (Box and Tiao, 1975, JASA) can be expressed as:

$$Y_t = M_t + X_t, \tag{2}$$

where $Y_t$ represents the monthly (log) calling rate per 100,000 smokers;

# Intervention time series

The intervention time series model (Box and Tiao, 1975, JASA) can be expressed as:

$$Y_t = M_t + X_t, \tag{2}$$

where $Y_t$ represents the monthly (log) calling rate per 100,000 smokers;

$X_t$ is a seasonal Box and Jenkin's ARIMA $(p, d, q)(P, D, Q)$ model which represents the baseline or background monthly (log) calling rate per 100,000 smokers throughout the selected interval of time;

# Time series background

A Box and Jenkin's ARIMA $(p, d, q)$ model can be written as

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\epsilon_t, \tag{3}$$

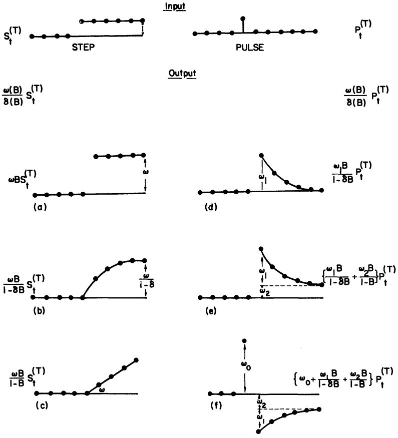
where $B$ is the back-shift operator such that $BX_t = X_{t-1}$, $d$ is the number of trend differences, $\phi_p(B)$ and $\theta_q(B)$ are the polynomials in $B$ of order $p$ and $q$ separately, that is $\phi_p(B) = 1 + \sum_{i=1}^{p} \phi_i B^i$, $\theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$. If we consider seasonality, Model (3) can be modified as

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t, \tag{4}$$

where $D$ is the number of seasonal differences, $s$ is the seasonal period, and $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are polynomials in $B^s$ of order $P$ and $Q$, respectively.

# Transfer function

Two types of interventions: step and pulse interventions (Box and Tiao, 1975)




[a] (a), (b), (c) show the response to a step input for various simple transfer function models; (d), (e), (f) show the response to a pulse for some models of interest.

# Transfer function: continue

$M_t$ represents the additive change in the log calling rate due to the intervention. In other words $M_t$ is the log rate ratio at certain time $t$.

$M_t$ is a transfer function of intervention (or dummy) variable $I_t$,

$$I_t = \begin{cases} 1 & \text{if } t >= T_0, \\ 0 & \text{otherwise .} \end{cases}$$

# Transfer function: continue

$M_t$ represents the additive change in the log calling rate due to the intervention. In other words $M_t$ is the log rate ratio at certain time $t$.

$M_t$ is a transfer function of intervention (or dummy) variable $I_t$,

$$I_t = \begin{cases} 1 & \text{if } t >= T_0, \\ 0 & \text{otherwise }. \end{cases}$$

A flexible yet parsimonious form of $M_t$ could be a "first order" dynamic process of $I_t$, e.g., $M_t = \delta M_{t-1} + \omega I_t$. The value of the transfer function $M_t$ is as following:

$$M_t = \begin{cases} \frac{\omega(1-\delta^{k+1})}{1-\delta} & \text{if } k >= 0, \\ 0 & \text{otherwise }, \end{cases}$$

# Transfer function: continue

Compared to the calling rates that would have been observed in the absence of intervention, the relative change in the calling rate $k$ months after intervention is a non-linear function of two parameters $(\omega, \delta)$

$$\mathrm{RR}_k = \begin{cases} \exp(\omega(\frac{1-\delta^{k+1}}{1-\delta})) & \text{if } k \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

# Transfer function: continue

The limits of the function are
$RR(\omega, \delta, k \to 0) = exp(\omega)$ "immediate" intervention effect occurring exactly the intervention month;

# Transfer function: continue

The limits of the function are
$RR(\omega, \delta, k \to 0) = exp(\omega)$ "immediate" intervention effect occurring exactly the intervention month;

$RR(\omega, \delta, k \to large) = exp(\frac{\omega}{1-\delta})$ "permanent or long term" effect after k months;

# Transfer function: continue

The limits of the function are
$RR(\omega, \delta, k \to 0) = exp(\omega)$ "immediate" intervention effect occurring exactly the intervention month;

$RR(\omega, \delta, k \to large) = exp(\frac{\omega}{1-\delta})$ "permanent or long term" effect after k months;

The parameter $\delta$ provides information about how quickly the rate ratio converges toward its long-term value. The cloest $\delta$ to 0, the quickest is the convergence.

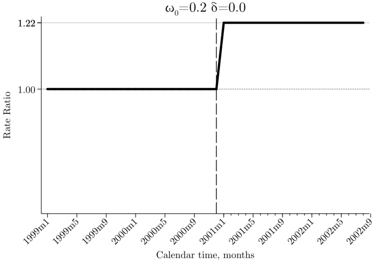
# Transfer function: continue

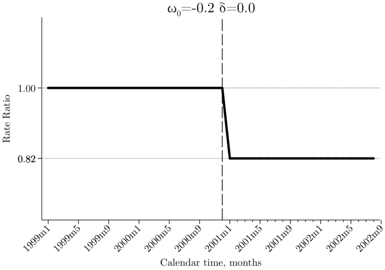
(1) No intervention effect:

If $\omega = 0$ the $RR_k = 1$ regardless of the magnitude and sign of $\delta$

(2) Immediate (either positive or negative) and no further effect over time:

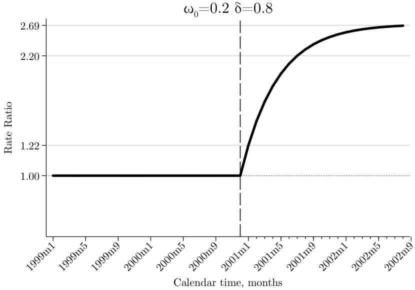
If $\delta = 0$ the $RR_k = exp(\omega)$

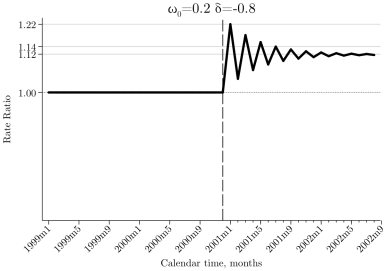


# Transfer function: continue

(3) Immediate positive effect plus a smooth and gradual increasing effect over time or an oscillating effect over time



# Algorithm

Let $\theta = (\theta_1', \theta_2')'$, $\theta_1$ includes the parameters from the transfer function such that $\theta_1 = (\delta, \omega)'$; and $\theta_2$ includes the parameters coming from the time series models.

(1) Create the likelihood function in Mata

(a) Use $\theta_1$ to calculate $m_t$, where $m_t = \delta * m_{t-1} + \omega * I_t$;

(b) Update $y_t^* = y_t - m_t$;

(c) Use $\theta_2$ and $y_t^*$ to calculate the likelihood, by calling Stata **arima** command

(d) Return the likelihood

(2) Maximize the likelihood function using Mata **optimize()**

(i) Use the BFGS algorithm

(3) Inference on the intervention effects

(i) Delta method using **nlcom**

(4) Tabulate the estimated results, plot the graphs and so on

# Syntax

```
// Simple step function
tstf lograte after
```

# Syntax

```
// Simple step function
tstf lograte after

// Controlling the ARIMA and transfer parameters
tstf lograte after, arima(1,0,1) sarima(0,1,0,12) t(1,0)
```

# Syntax

```
// Simple step function
tstf lograte after

// Controlling the ARIMA and transfer parameters
tstf lograte after, arima(1,0,1) sarima(0,1,0,12) t(1,0)

// Graphs and tabulated
tstf lograte after, arima(1,0,1) sarima(0,1,0,12)  t(1,0) method(ML)  ///
greffect gredate tabulate
```

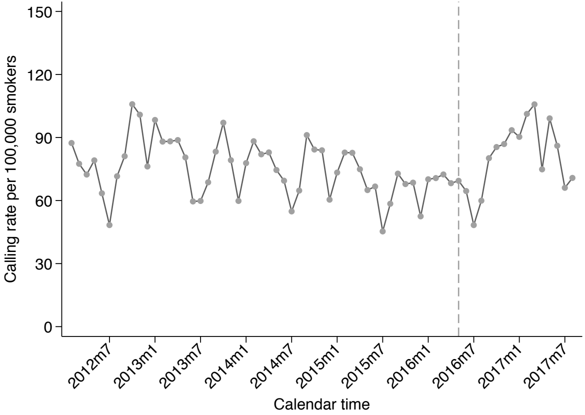
# Syntax

```
// Simple step function
tstf lograte after

// Controlling the ARIMA and transfer parameters
tstf lograte after, arima(1,0,1) sarima(0,1,0,12) t(1,0)

// Graphs and tabulated
tstf lograte after, arima(1,0,1) sarima(0,1,0,12)  t(1,0) method(ML)  ///
greffect gredate tabulate

// Graphs and tabulated, exponentitation
tstf lograte after, arima(1,0,1)  sarima(0,1,0, 12)   t(1,0) method(ML)  ///
gre grd eform tabulate
```

# Steps when using tstf command

- Choose the orders of the (seasonal) ARIMA background through the pre-intervention time points;
- Choose an intervention model;
- Estimate the model through ML and get the common dynamic effect of the intervention;
- Inference on the effect $k$-months after each intervention.

# Data: Calling rate per 100,000 smokers




Larger pictorial warnings on cigarette packs (May 2016)

# Output

```
// Simple step function
   tstf lograte after

ARIMA regression with a transfer function            No. of obs     =          67
Optimization = CSS-ML
Log likelihood = 26.707402
Sample = 2012m2 - 2017m8   Intervention starts = 2016m5
------------------------------------------------------------------------------
     lograte |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARIMA        |
         ar1 |    .499028   .1063989     4.69   0.000     .2904899    .7075661
       _cons |   4.307029   .0438631    98.19   0.000     4.221059    4.392999
-------------+----------------------------------------------------------------
TRANSFER     |
       omega |   .0429717   .0866812     0.50   0.620    -.1269204    .2128638
------------------------------------------------------------------------------
```

# Output

```
//Controlling the ARIMA and transfer parameters
tstf lograte after, arima(1,0,1)  sarima(0,1,0,12)   t(1,0) ///
                    method(ML)  gre grd eform tabulate

ARIMA regression with a transfer function           No. of obs     =          67
Optimization = ML
Log likelihood = 42.170879
Sample = 2012m2 - 2017m8   Intervention starts = 2016m5
------------------------------------------------------------------------------
     lograte |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARIMA        |
         ar1 |   .8558569   .1183034     7.23   0.000     .6239865    1.087727
         ma1 |  -.5901846   .1807572    -3.27   0.001    -.9444622   -.2359069
-------------+----------------------------------------------------------------
TRANSFER     |
       delta |   .9054777   .0727619    12.44   0.000      .762867    1.048088
       omega |   .0472376   .0212599     2.22   0.026      .005569    .0889062
------------------------------------------------------------------------------
```

# Output, continue

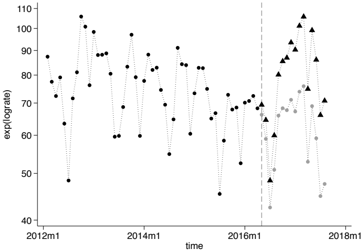
```
// tabulate the confidence intervals
tstf lograte after, arima(3,1,2)  sarima(1,0,0,12)   t(1,0) ///
                    method(ML)    eform tabulate
```

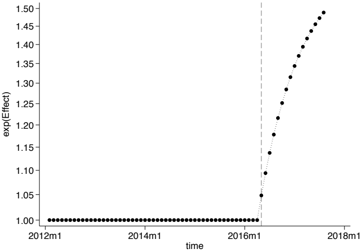
Table of effects k units of time after intervention

| time | k | exp(Eff) | LB | UB | P-value |
|---|---|---|---|---|---|
| 2016m5 | 0 | 1.05 | 1.01 | 1.09 | 0.026 |
| 2016m6 | 1 | 1.09 | 1.02 | 1.18 | 0.017 |
| 2016m7 | 2 | 1.14 | 1.03 | 1.26 | 0.011 |
| 2016m8 | 3 | 1.18 | 1.05 | 1.32 | 0.006 |
| 2016m9 | 4 | 1.22 | 1.07 | 1.39 | 0.003 |
| 2016m10 | 5 | 1.25 | 1.09 | 1.44 | 0.002 |
| 2016m11 | 6 | 1.28 | 1.11 | 1.49 | 0.001 |
| 2016m12 | 7 | 1.32 | 1.13 | 1.54 | 0.001 |
| 2017m1 | 8 | 1.34 | 1.14 | 1.58 | 0.000 |
| 2017m2 | 9 | 1.37 | 1.16 | 1.62 | 0.000 |
| 2017m3 | 10 | 1.39 | 1.17 | 1.65 | 0.000 |
| 2017m4 | 11 | 1.42 | 1.19 | 1.69 | 0.000 |
| 2017m5 | 12 | 1.44 | 1.19 | 1.73 | 0.000 |
| 2017m6 | 13 | 1.46 | 1.20 | 1.76 | 0.000 |
| 2017m7 | 14 | 1.47 | 1.20 | 1.80 | 0.000 |

# Output, continue

```
// graphs
tstf lograte after, arima(3,1,2)  sarima(1,0,0,12)   t(1,0) ///
                        method(ML)  gre grd eform tabulate
```




The calling rate gradually and significantly ($p$-value $< 0.001$) increased after the introduction of the larger pictorial warnings on the cigarette packs.

# Summary

- We are working on a Stata package **tstf** to estimate intervention time series model with transfer functions
- We focus on the transfer functions with two parameters (shape and scale parameter), the background is a seasonal time series model
- Estimation, inference, and graphs are given for both the original data and the log-transformed data

# Future work

- Keep on working the **tstf** package for multiple interventions
- Power calculations
- Time-vary confounders

# Team members

XingWu Zhou, Postdoc Biostatistics, Karolinska Institutet

Alessio Crippa, PhD student Biostatistics, Karolinska Institutet

Rosaria Galanti, Professor Epidemiology, Centre of Epidemiology and Community Medicine

Nicola Orsini, Associate Professor Medical Statistics, Karolinska Institutet

# References

(1) Box, G.E. P., Tiao G. C. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association. 1975. 70(349), 70-79.

(2) Shadish, W. R., Cook, T. D., & Campbell, D. T. Experimental and quasi-experimental designs for generalized causal inference. 2002. Belmont, CA: Wadsworth Cengage Learning.