Essex Summer School in Social Science Data Analysis
Panel Data Analysis for Comparative Research

Modeling Nonstationary Time Series

Christopher Adolph

Department of Political Science
and
Center for Statistics and the Social Sciences
University of Washington, Seattle

What we're doing today

Next steps:

- Learn some (weak) techniques for identifying non-stationary time series (from previous lecture slides)
- Analyze non-stationary series using differences
- Analyze non-stationary series using cointegration

Differences & Integrated time series

Define $\Delta^d y_t$ as the dth difference of y_t

For the first difference $(d = 1)$, we write

$$\Delta y_t = y_t - y_{t-1}$$

For the second difference $(d = 2)$, we write

$$\Delta^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

or the difference of two first differences

or the difference in the difference

Differences & Integrated time series

For the third difference ($d = 3$), we write

$$\Delta^3 y_t = ((y_t - y_{t-1}) - (y_{t-1} - y_{t-2})) - (y_{t-1} - y_{t-2}) - (y_{t-2} - y_{t-3})$$

or the difference of two second differences

or the difference in the difference in the difference

This gets perplexing fast.

Fortunately, we will rarely need $d > 1$, and almost never $d > 2$.

Differences & Integrated time series

What happens if we difference a stationary AR(1) process ($|\phi_1| < 1$)?

$$y_t = y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference a stationary AR(1) process ($|\phi_1| < 1$)?

$$\begin{aligned} y_t &= y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= y_{t-1}\phi_1 - y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

Differences & Integrated time series

What happens if we difference a stationary AR(1) process ($|\phi_1| < 1$)?

$$
\begin{aligned}
y_t &= y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\
y_t - y_{t-1} &= y_{t-1}\phi_1 - y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\
\Delta y_t &= (1-\phi)y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t
\end{aligned}
$$

We still have an AR(1) process,
and we've thrown away some useful information – the levels in y_t – that our covariates x_t might have explained

Differences & Integrated time series

What happens if we difference a random walk?

$$y_t \quad = \quad y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference a random walk?

$$
\begin{aligned}
y_t &= y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\
y_t - y_{t-1} &= y_{t-1} - y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t
\end{aligned}
$$

Differences & Integrated time series

What happens if we difference a random walk?

$$\begin{aligned} y_t &= y_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= y_{t-1} - y_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \\ \Delta y_t &= \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

The result is AR(0), and stationary –
we could analyze it using ARMA(0,0), which is just LS regression!

When a single differencing removes non-stationarity from a time series y_t,
we say y_t is *integrated* of order 1, or I(1).

A time series that does not need to be differenced to be stationary is I(0).

This differencing trick comes at a price:
we can only explain changes in y_t, *not* levels,
and hence not the long-run relationship between y_t and $\mathbf{x}_t$.

Differences & Integrated time series

What happens if we difference an AR(2) unit root process?

$$y_t \quad = \quad 1.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference an AR(2) unit root process?

$$\begin{aligned} y_t &= 1.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= 1.5y_{t-1} - y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

Differences & Integrated time series

What happens if we difference an AR(2) unit root process?

$$\begin{aligned} y_t &= 1.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= 1.5y_{t-1} - y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ \Delta y_t &= 0.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

We get a stationary AR(2) process.
We could analyze this new process with ARMA(2,0).

We say that the original process is ARIMA(2,1,0),
or an integrated autoregressive process of order 2, integrated of order 1.

Differences & Integrated time series

Recall our GDP & Democracy example

$$\begin{aligned}\mathrm{GDP}_t &= \phi_1 \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t \\ \mathrm{GDP}_t &= 0.9 \times \mathrm{GDP}_{t-1} + 10 + 2 \times \mathrm{Democracy}_t + \varepsilon_t\end{aligned}$$

At year t, $\mathrm{GDP}_t = 100$ and the country is a non-democracy $\mathrm{Democracy}_t = 0$, and we were curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

But now suppose ϕ_1 is 1, β_0 is 0,
and we want to model the first difference, $\Delta\mathrm{GDP}_t$, instead of the level of GDP_t.

Differences & Integrated time series

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$\text{GDP}_t \quad = \quad \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t$$

Differences & Integrated time series

At year t, $\mathrm{GDP}_t = 100$ and the country is a non-democracy $\mathrm{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$\begin{aligned}
\mathrm{GDP}_t &= \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t \\
\mathrm{GDP}_t - \mathrm{GDP}_{t-1} &= \mathrm{GDP}_{t-1} - \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t
\end{aligned}$$

Differences & Integrated time series

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$
\begin{aligned}
\text{GDP}_t &= \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\text{GDP}_t - \text{GDP}_{t-1} &= \text{GDP}_{t-1} - \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\Delta\text{GDP}_t &= \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t
\end{aligned}
$$

Differences & Integrated time series

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$
\begin{aligned}
\text{GDP}_t &= \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\text{GDP}_t - \text{GDP}_{t-1} &= \text{GDP}_{t-1} - \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\Delta \text{GDP}_t &= \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\Delta \text{GDP}_t &= 0 + 2 \times \text{Democracy}_t + \varepsilon_t
\end{aligned}
$$

Differences & Integrated time series

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$\Delta\text{GDP}_t \quad = \quad 0 + 2 \times \text{Democracy}_t + \varepsilon_t$$

Works just as before – but gives only the one period change in GDP

Iterating, we get the cumulative change

We have to supply external information on the *levels* in order to get predictions of the level

The model doesn't know them

Moreover, the impact of lagged ε's here doesn't ever diminish over time

So long predictions are very unreliable

ARIMA(p,d,q) models

An ARIMA(p,d,q) regression model has the following form:

$$\begin{aligned}\Delta^d y_t &= \Delta^d y_{t-1}\phi_1 + \Delta^d y_{t-2}\phi_2 + \ldots + \Delta^d y_{t-p}\phi_p \\ &\quad +\varepsilon_{t-1}\rho_1 + \varepsilon_{t-2}\rho_2 + \ldots + \varepsilon_{t-q}\rho_q \\ &\quad +\mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t\end{aligned}$$

This just an ARMA(p,q) model applied to differenced y_t

The same MLE that gave us ARMA estimates still estimates $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\rho}}$, and $\hat{\boldsymbol{\beta}}$

We just need to choose d based on theory, ACFs and PACFs, and unit root tests (ugh)

ARIMA(p,d,q) models

Mechanically,
conditional forecasting and in-sample counterfactuals work just as before

Same code from last time will work; just change the d term of the ARIMA order to 1

But we need to be careful about forecasting too far into the future. . .

Example: Presidential Approval

We have data on the percent ($\times$ 100) of Americans supporting President Bush, averaged by month, over 2/2001–6/2006.

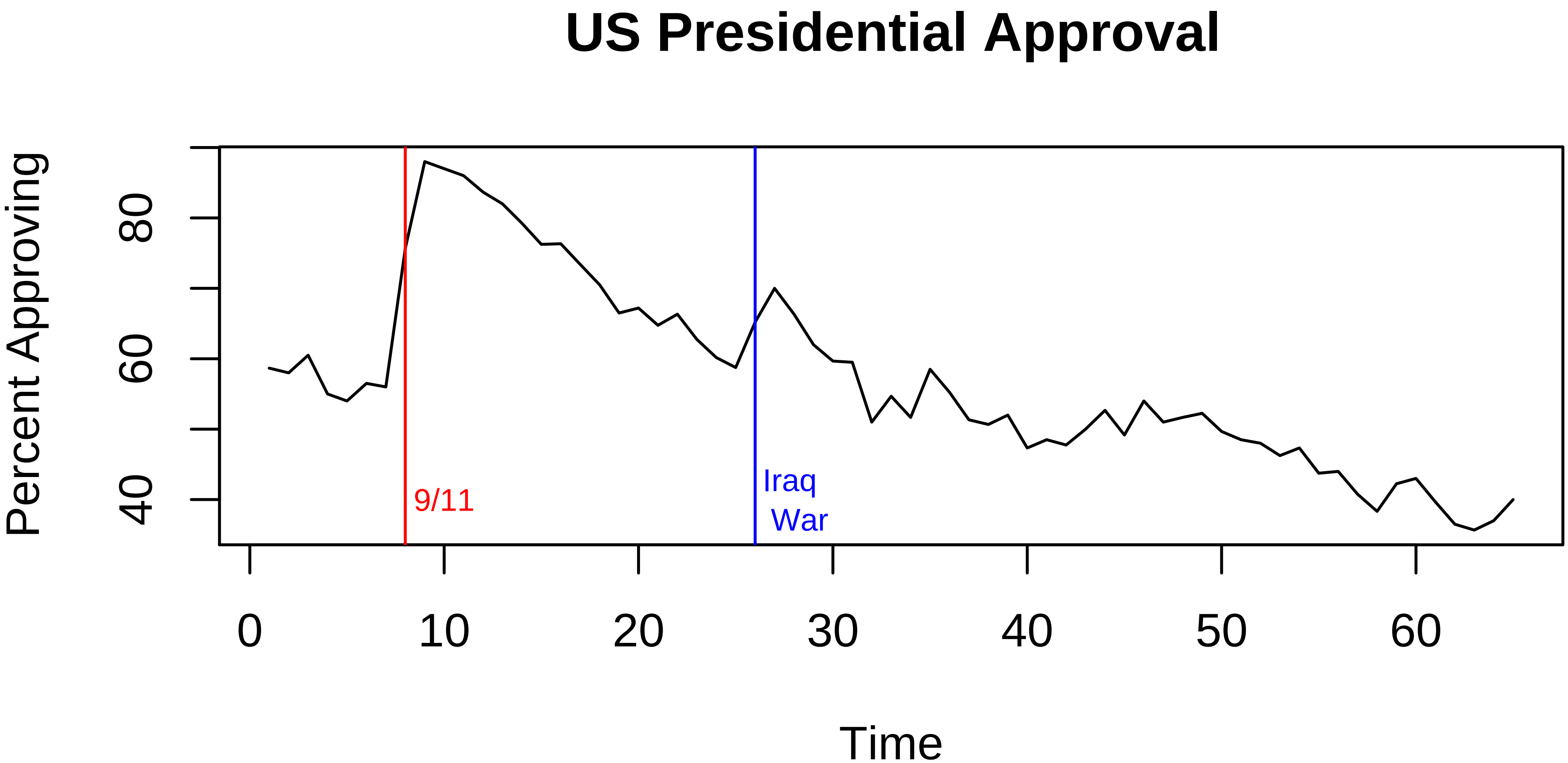
Our covariates include:

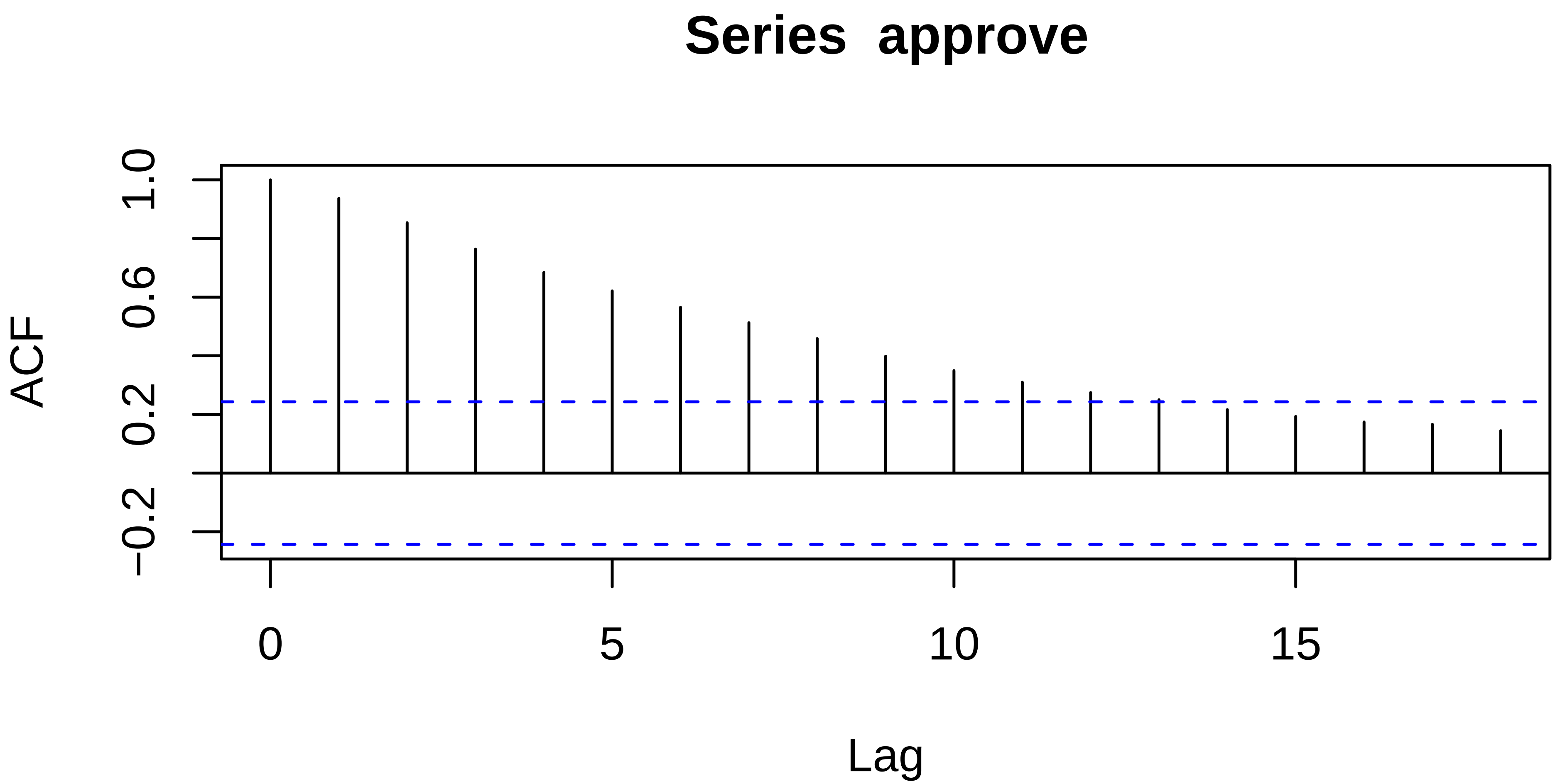
The average price of oil per month, in $/barrel

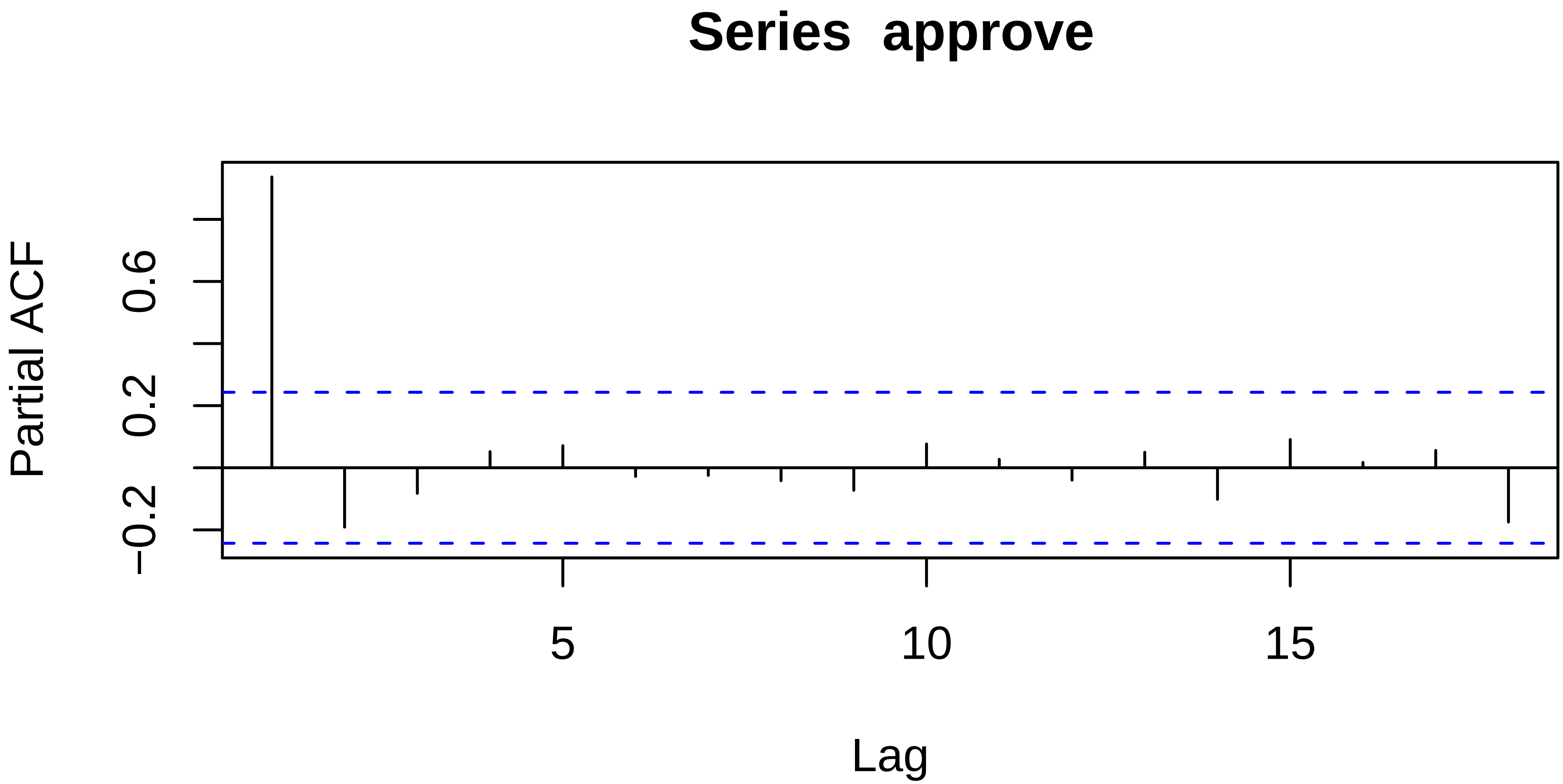
Dummies for September and October of 2001

Dummies for first three months of the Iraq War

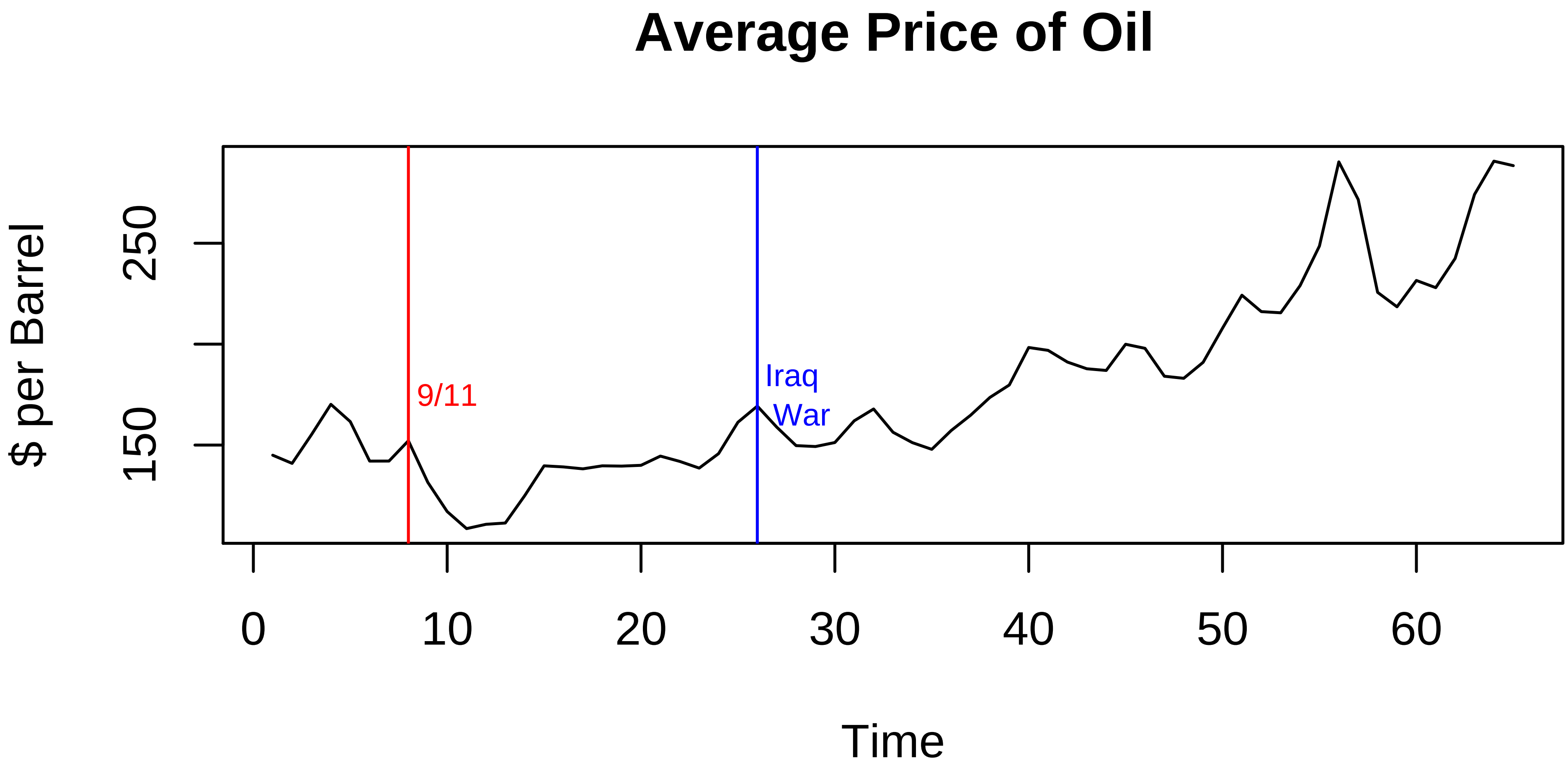
Let's look at our two continuous time series


US Presidential Approval
Percent Approving
40
60
80
9/11
Iraq
War
0
10
20
30
40
50
60
Time

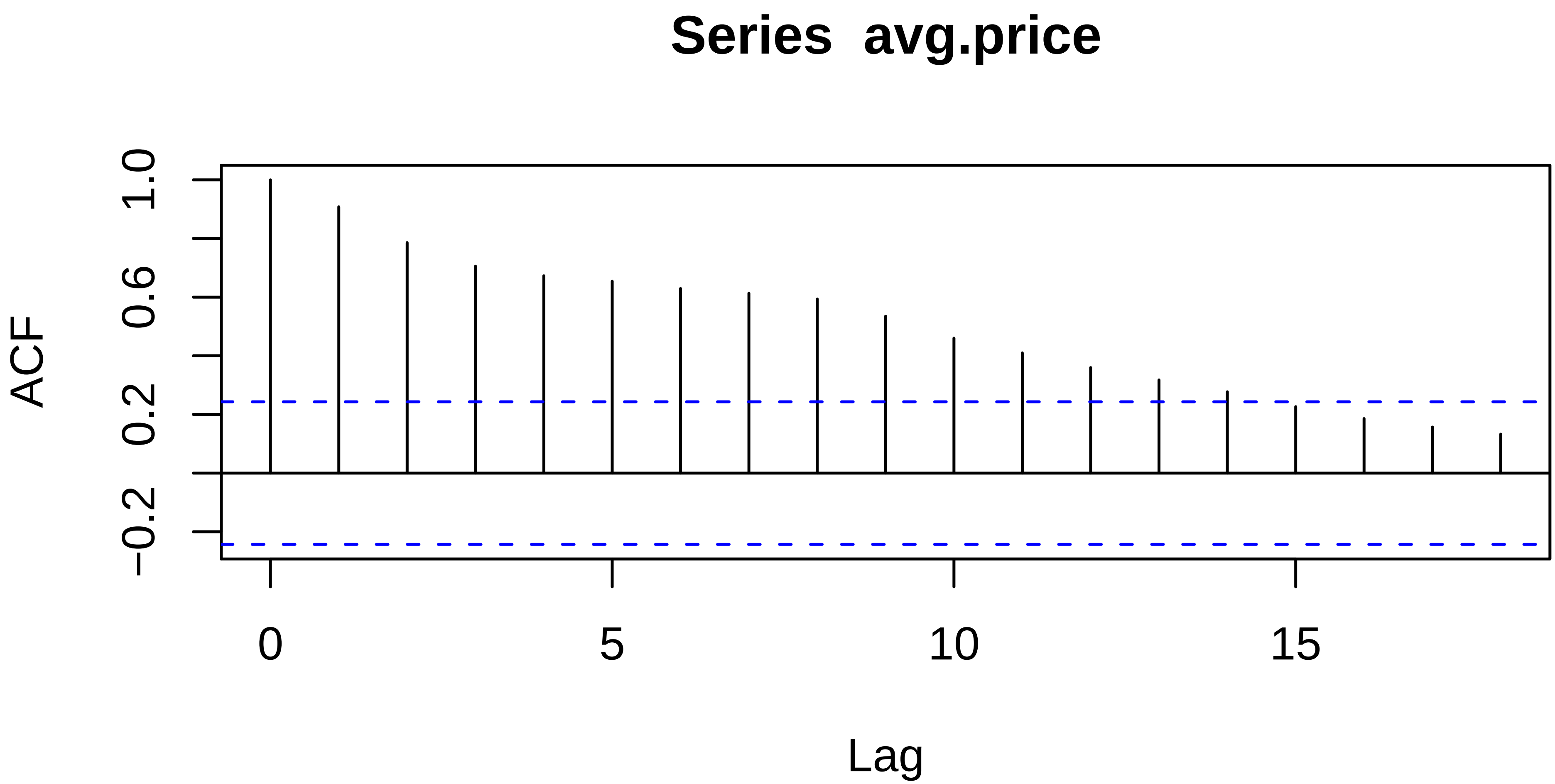

Series approve
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag



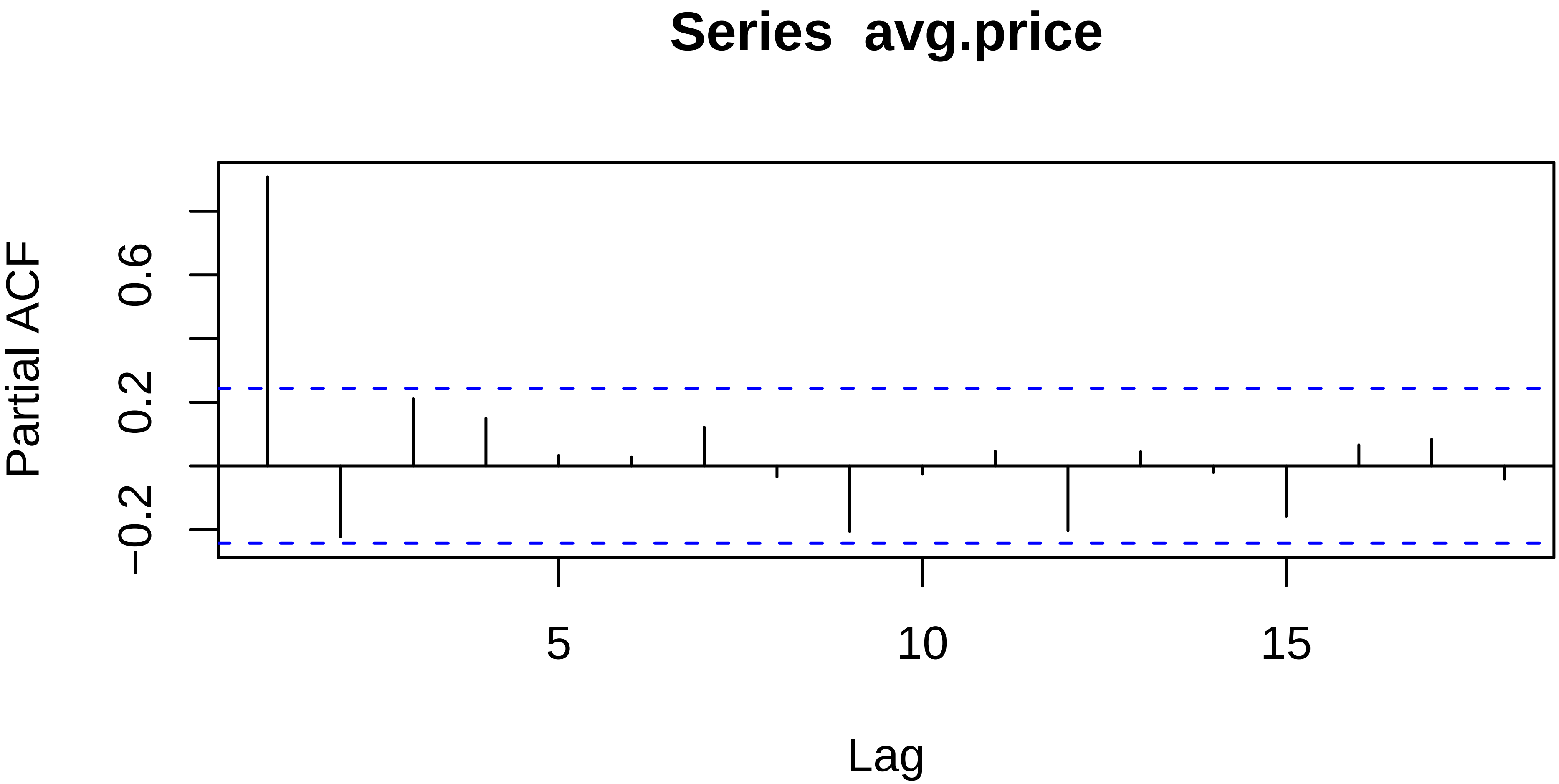
Series approve
Partial ACF
-0.2
0.2
0.6
5
10
15
Lag

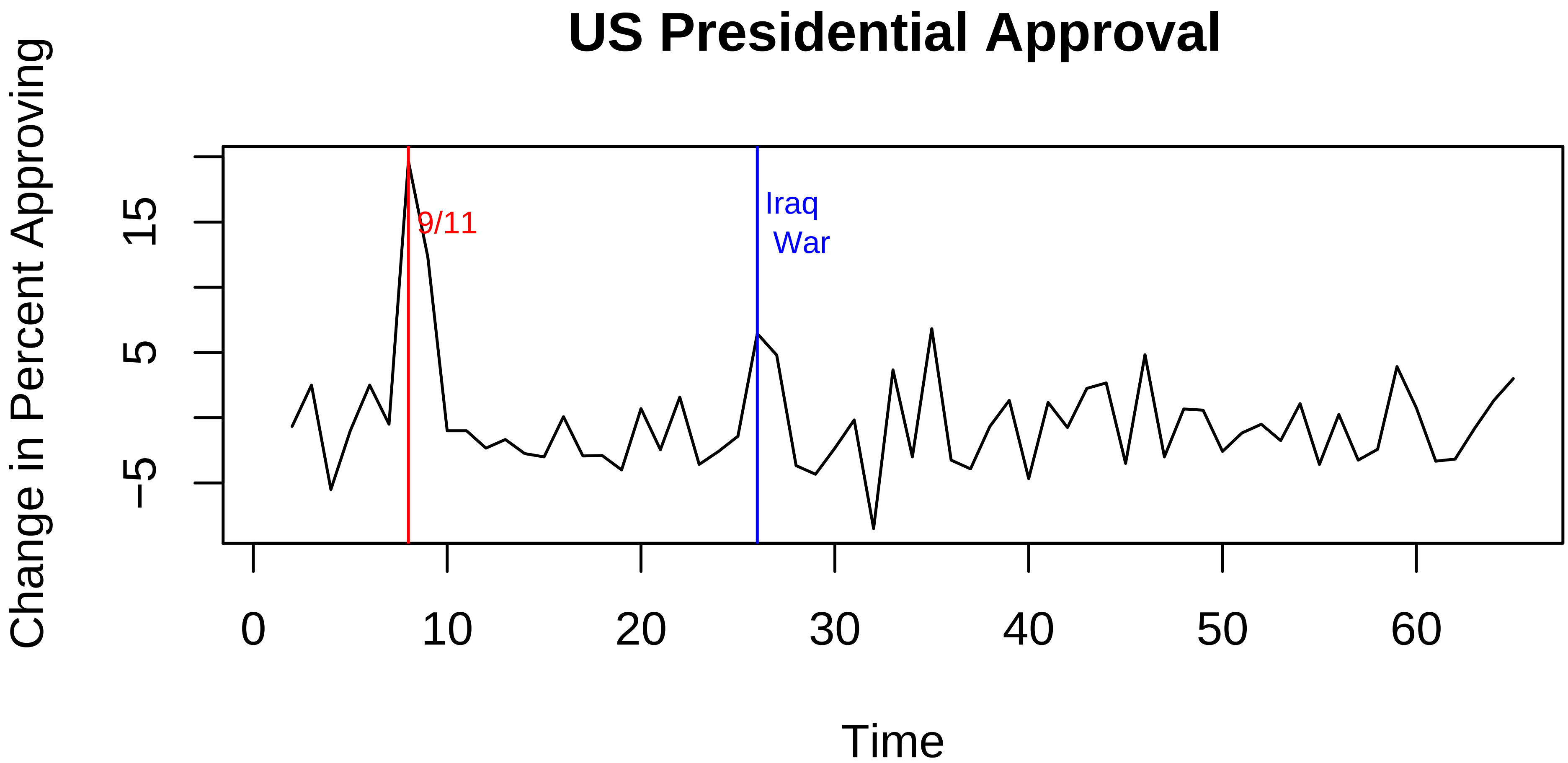


Average Price of Oil
$ per Barrel
150
250
9/11
Iraq
War
0
10
20
30
40
50
60
Time

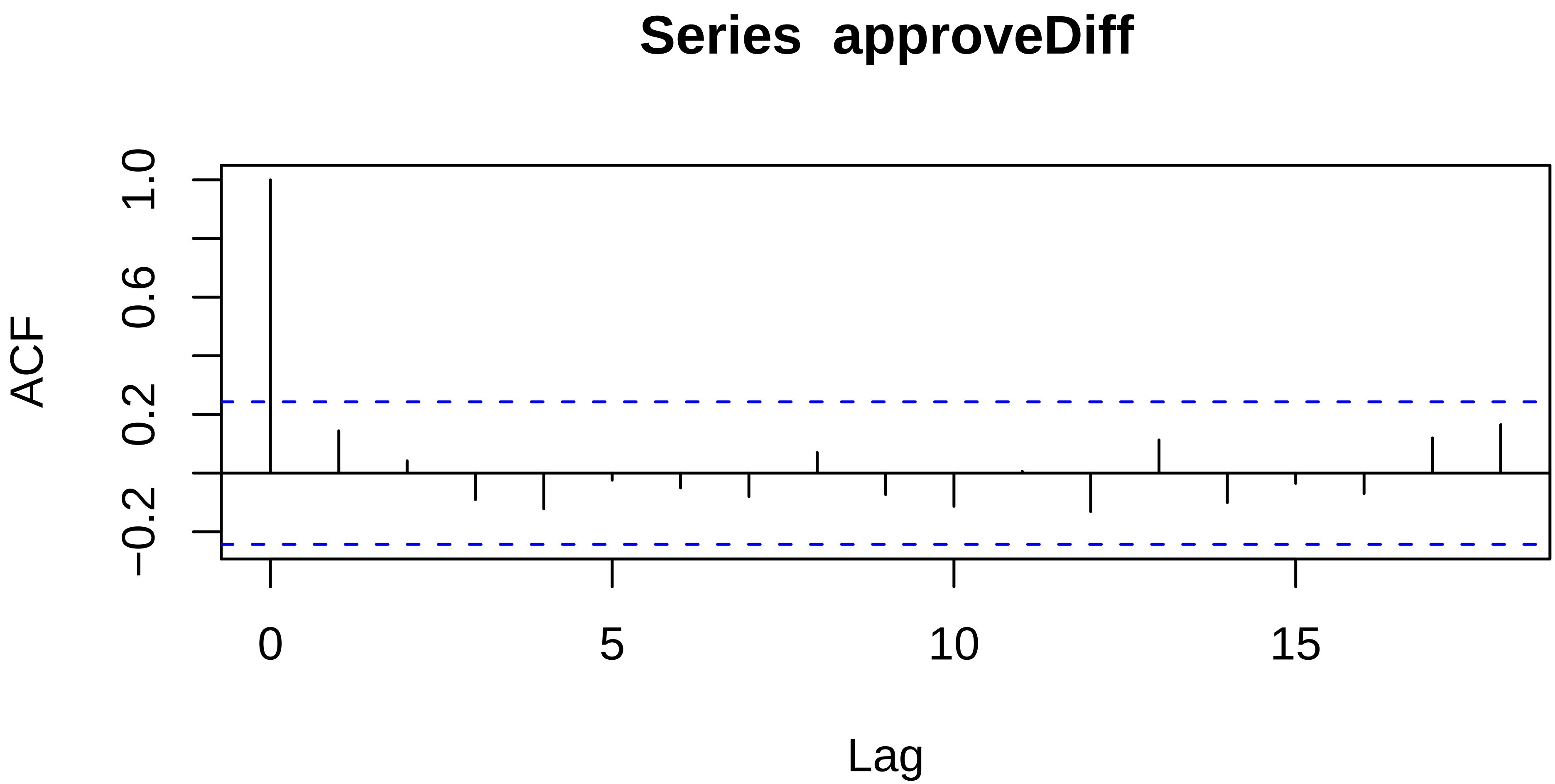


Series avg.price
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

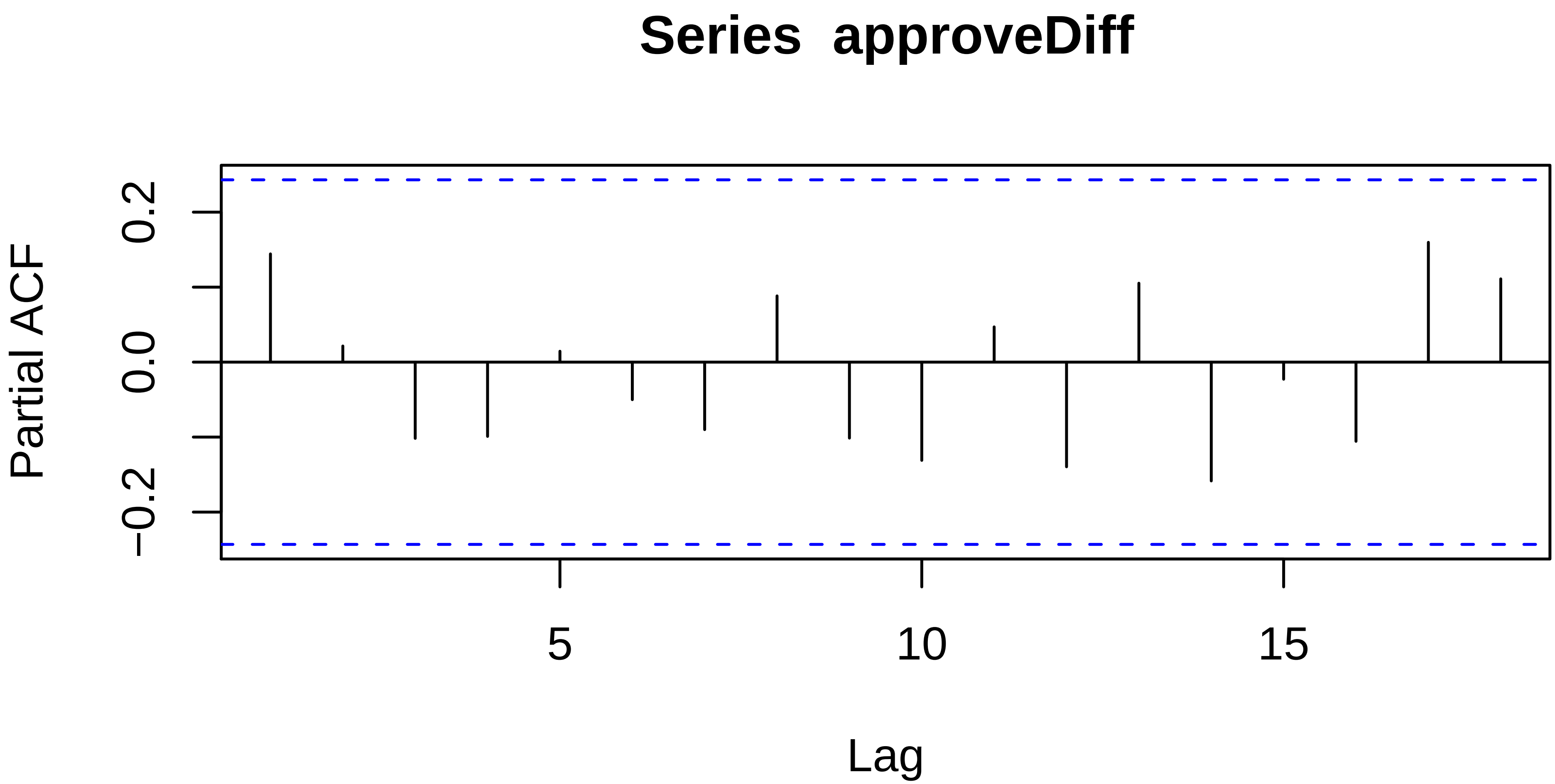

Series avg.price
Partial ACF
-0.2
0.2
0.6
5
10
15
Lag



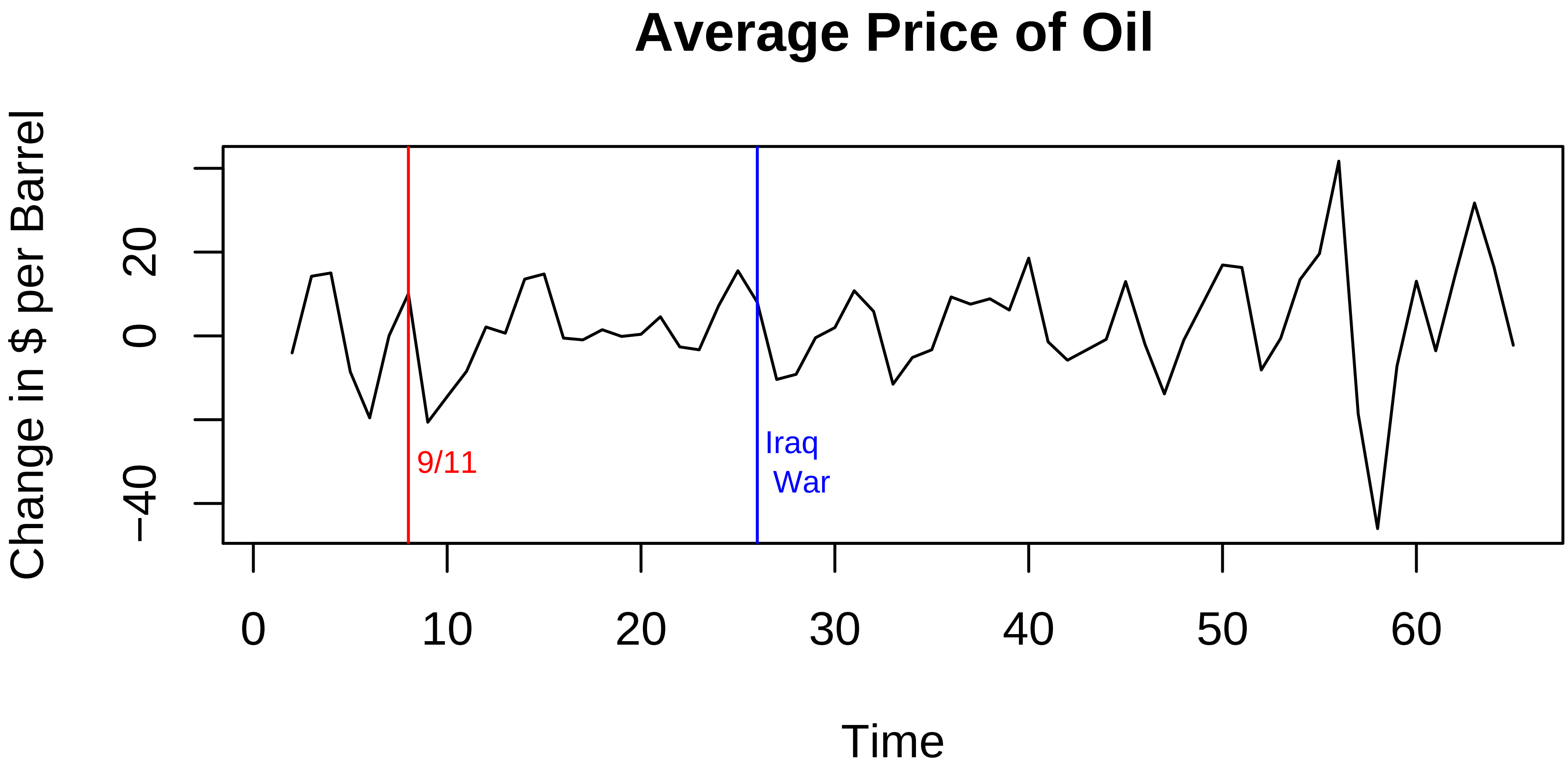
US Presidential Approval
Change in Percent Approving
Time
9/11
Iraq
War
−5
5
15
0
10
20
30
40
50
60

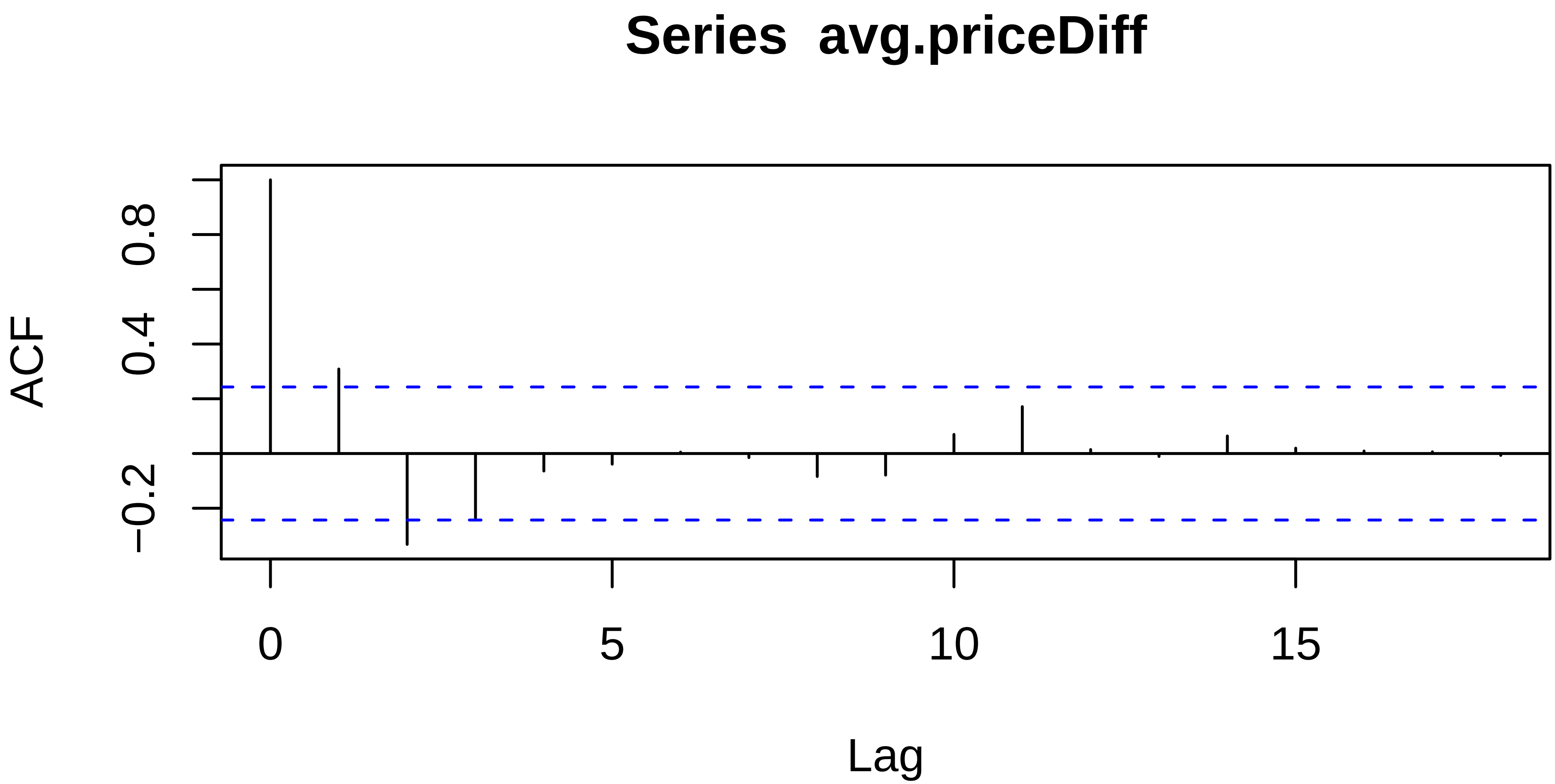


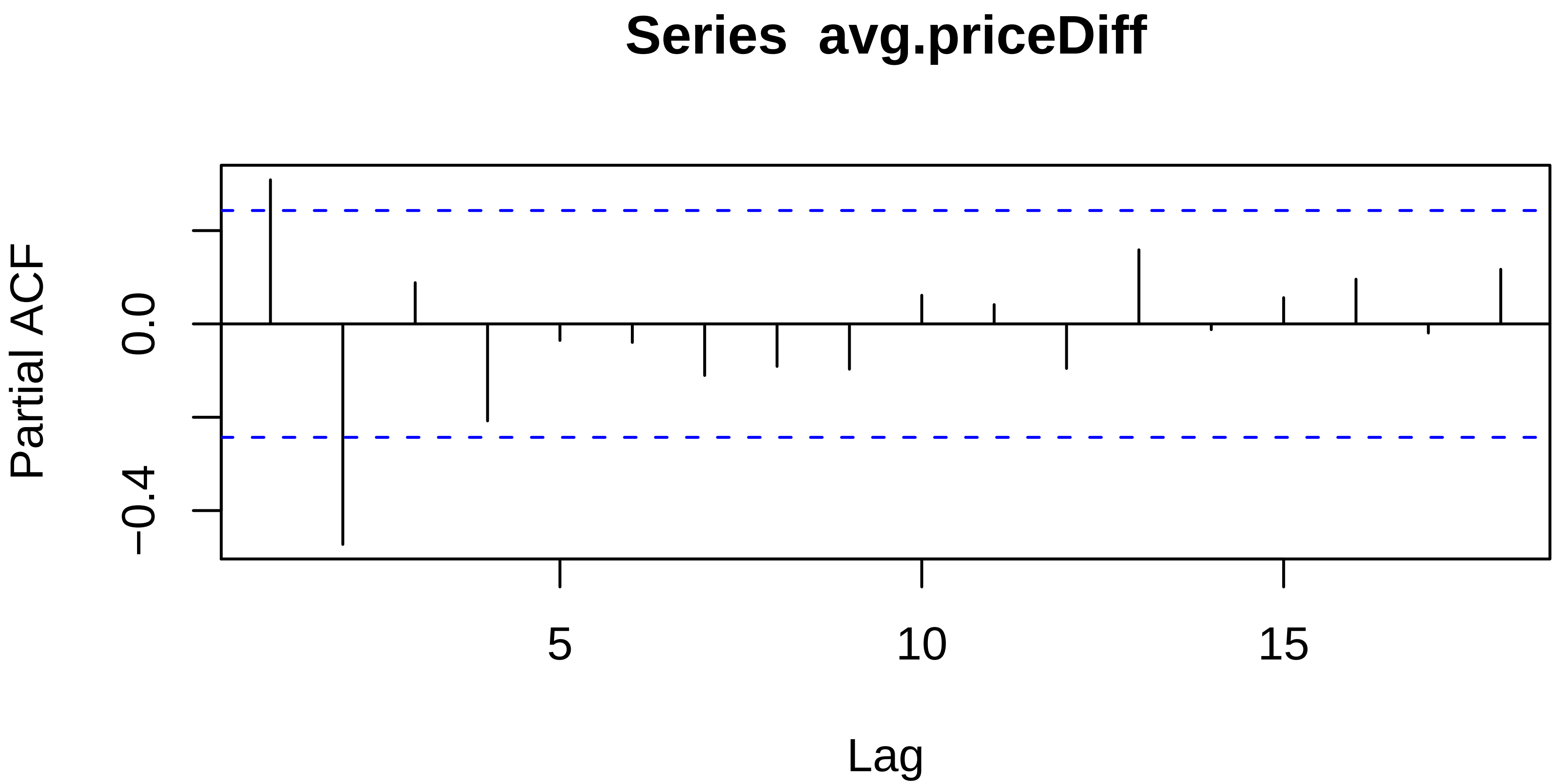
Series approveDiff
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag



Series approveDiff
Partial ACF
0.2
0.0
−0.2
5
10
15
Lag


Average Price of Oil
Change in $ per Barrel
−40
0
20
9/11
Iraq
War
0
10
20
30
40
50
60
Time


Series avg.priceDiff
ACF
0.8
0.4
−0.2
0
5
10
15
Lag



Series avg.priceDiff
Partial ACF
0.0
-0.4
5
10
15
Lag

Example: Presidential Approval

Many suspect `approve` and `avg.price` are non-stationary processes

Theoretically, what does this mean? Could an approval rate drift anywhere?

Note a better dependent variable would be the logit transformation of approve,
$\log(\texttt{approve}/(1 - \texttt{approve}))$,
which is unbounded and probably closer to the latent concept of support

And extending `approve` out to $T = \infty$ would likely stretch the concept too far for a democracy with regular, anticipated elections

We'll ignore this to focus on the time series issues

Example: Presidential Approval

To a first approximation, we suspect `approve` and `avg.price` may be non-stationary processes

We know that regressing one I(1) process on another risks spurrious correlation

How can we investigate the relationship between these variables?

Strategy 1: ARIMA(0,1,0), first differencing

Example: Presidential Approval

We load the data, plot it, with ACFs and PACFs

Then perform unit root tests

```
> PP.test(approve)

        Phillips-Perron Unit Root Test

data:  approve
Dickey-Fuller = -2.839, Truncation lag parameter = 3, p-value = 0.2350

> adf.test(approve)

        Augmented Dickey-Fuller Test

data:  approve
Dickey-Fuller = -3.957, Lag order = 3, p-value = 0.01721
alternative hypothesis: stationary
```

Example: Presidential Approval

```
> PP.test(avg.price)

        Phillips-Perron Unit Root Test

data:  avg.price
Dickey-Fuller = -2.332, Truncation lag parameter = 3, p-value = 0.4405

> adf.test(avg.price)

        Augmented Dickey-Fuller Test

data:  avg.price
Dickey-Fuller = -3.011, Lag order = 3, p-value = 0.1649
alternative hypothesis: stationary
```

Example: Presidential Approval

We create differenced versions of the time series, and repeat

```
> adf.test(na.omit(approveDiff))

        Augmented Dickey-Fuller Test

data:  na.omit(approveDiff)
Dickey-Fuller = -4.346, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

> adf.test(na.omit(avg.priceDiff))

        Augmented Dickey-Fuller Test

data:  na.omit(avg.priceDiff)
Dickey-Fuller = -5.336, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Example: Presidential Approval

We estimate an ARIMA(0,1,0), which fit a little better than ARIMA(2,1,2) on the AIC criterion

```
Call:
arima(x = approve, order = c(0, 1, 0),
      xreg = xcovariates, include.mean = TRUE)

Coefficients:
      sept.oct.2001  iraq.war  avg.price
             11.207     5.690     -0.071
s.e.          2.519     2.489      0.034

sigma^2 estimated as 12.4:  log likelihood = -171.2,  aic = 350.5
```

Example: Presidential Approval

To interpret the model, we focus on historical counterfactuals

What would Bush's approval have looked like if 9/11 hadn't happened?

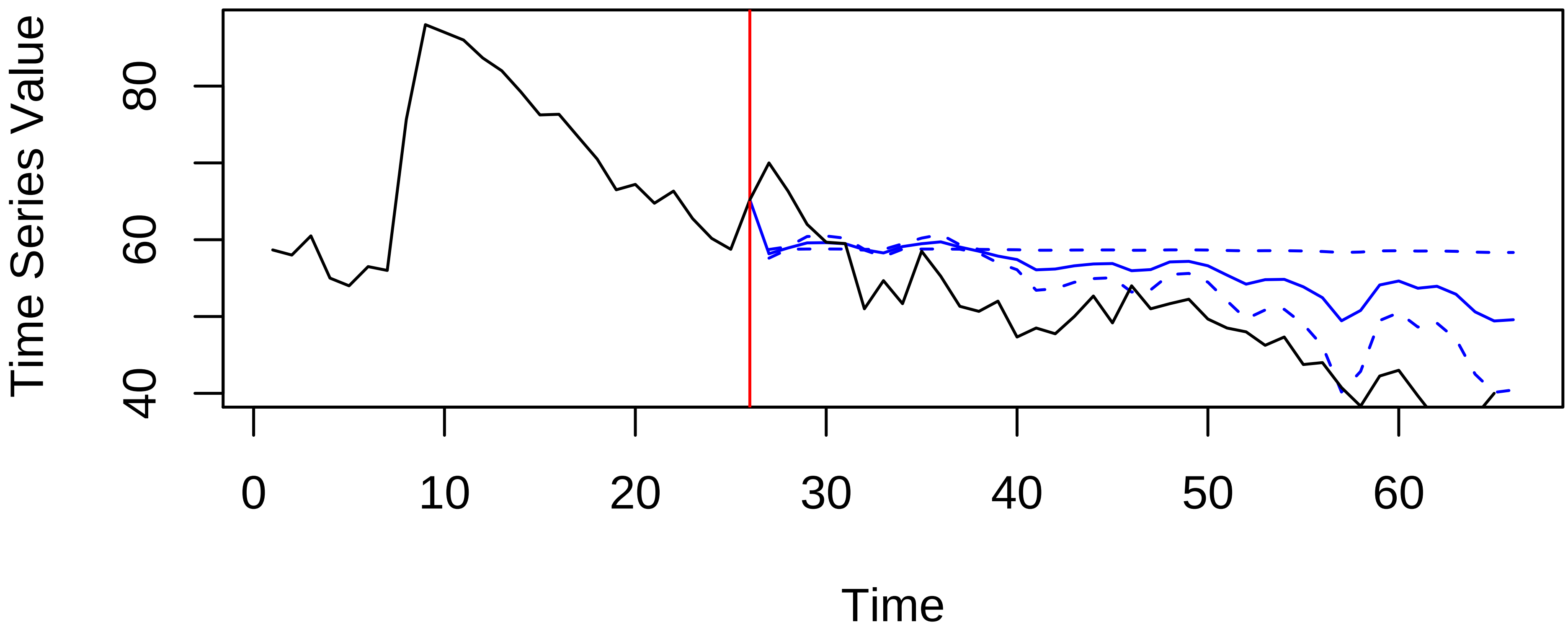
What if Bush had not invaded Iraq?

What if the price of oil had remained at pre-war levels?

Naturally, we only trust our results so far as we trust the model

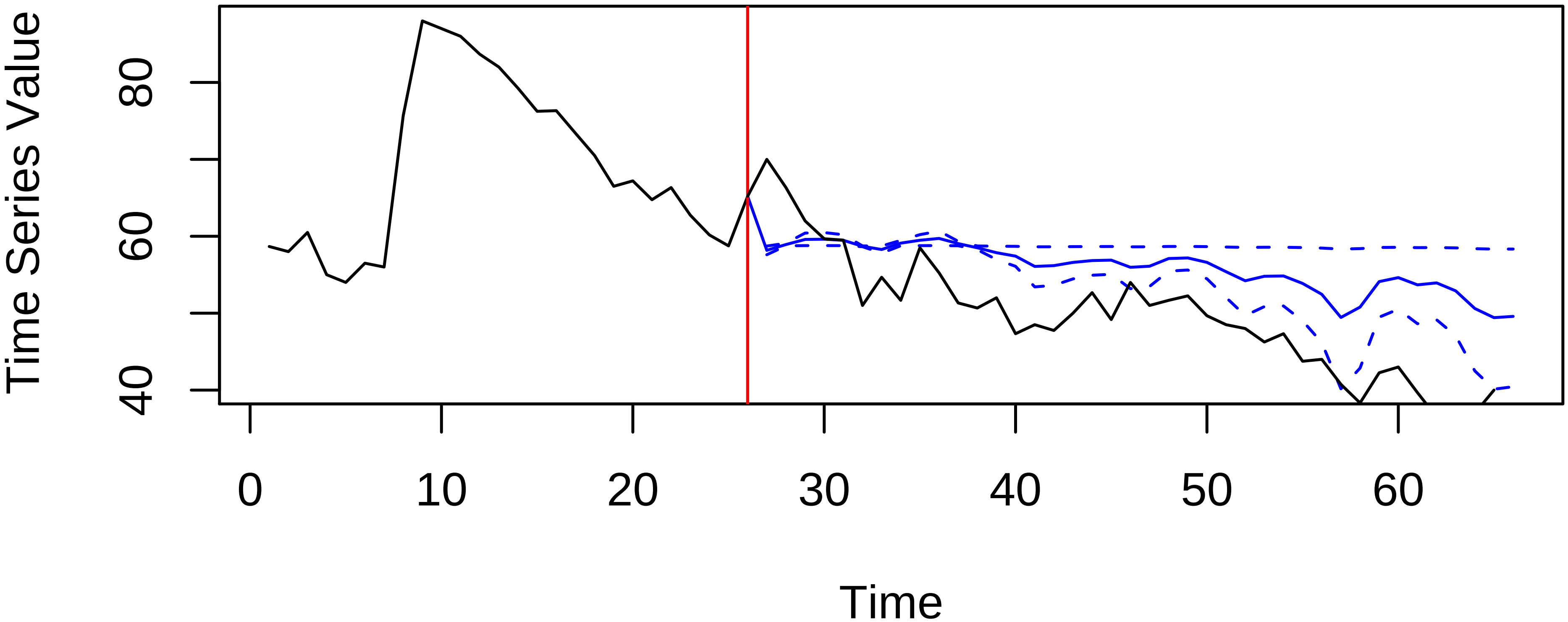
(which is not very much – we've left out a lot, like unemployment, inflation, boundedness of approve, . . .)

We simulate counterfactual approval using Zelig's implementation of ARIMA



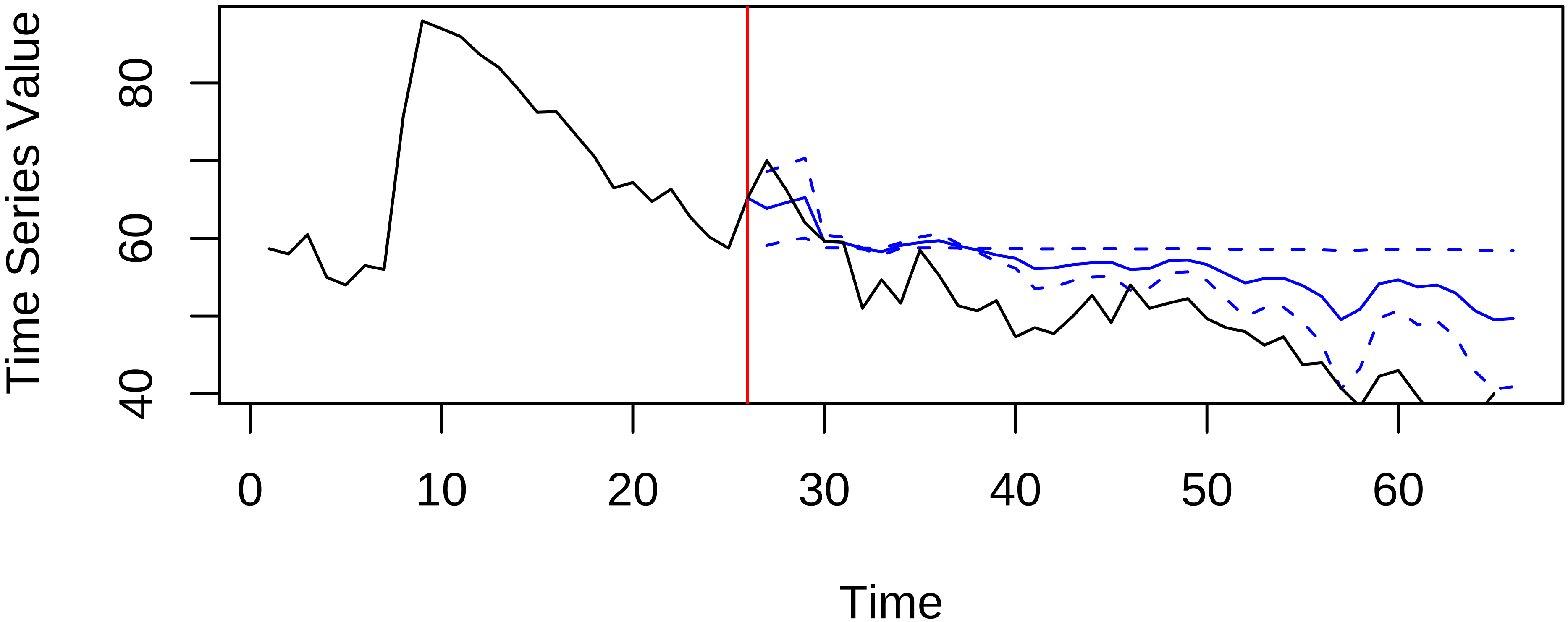
In blue: Predicted Bush approval without Iraq

In black: Actual approval



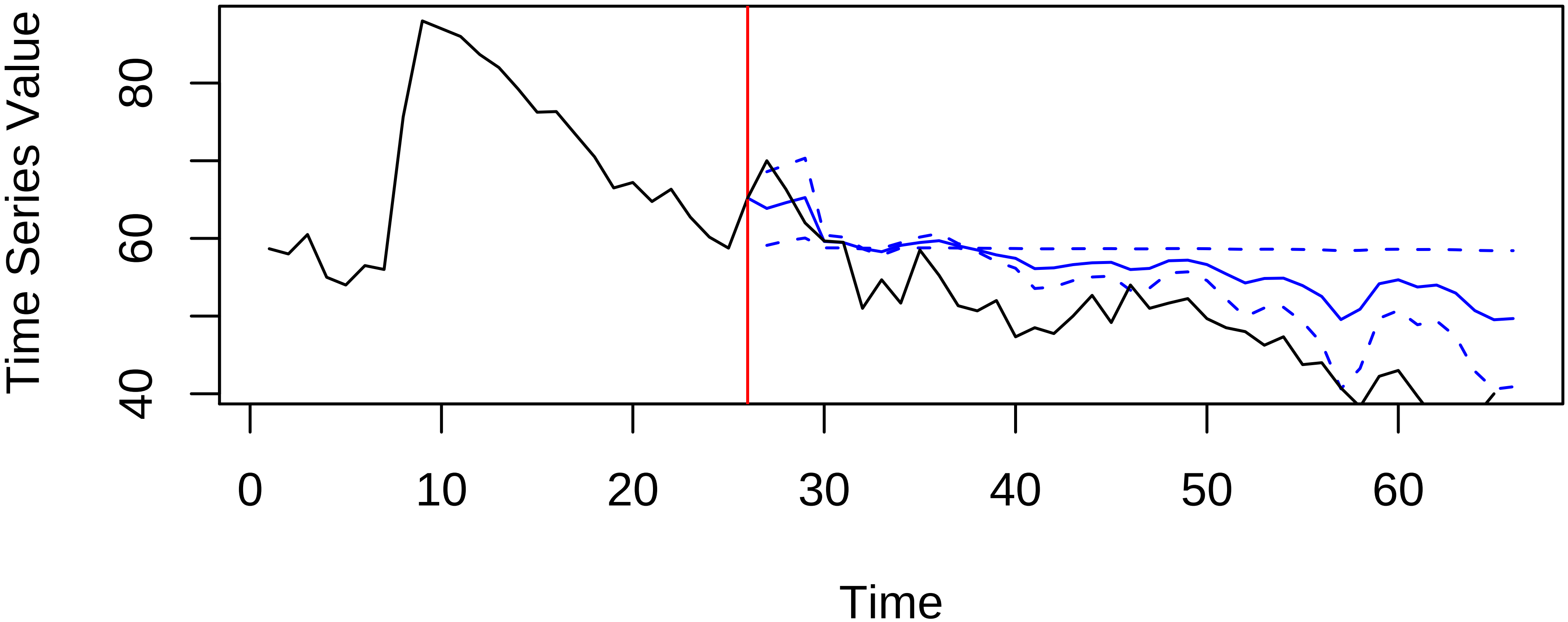
At first, starting the war in Iraq appears to help Bush's popularity

Then, it hurts – a lot. Sensible result. So are we done?

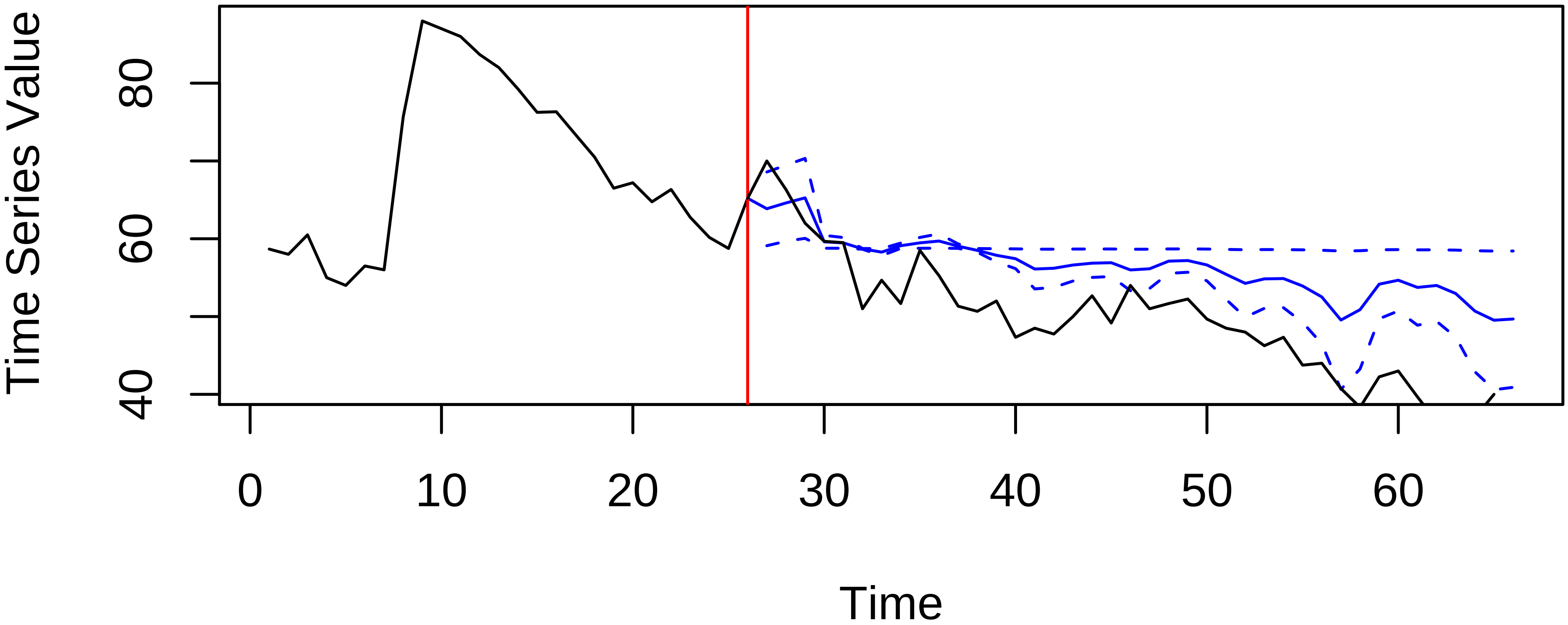


In blue: Predicted Bush approval with Iraq war

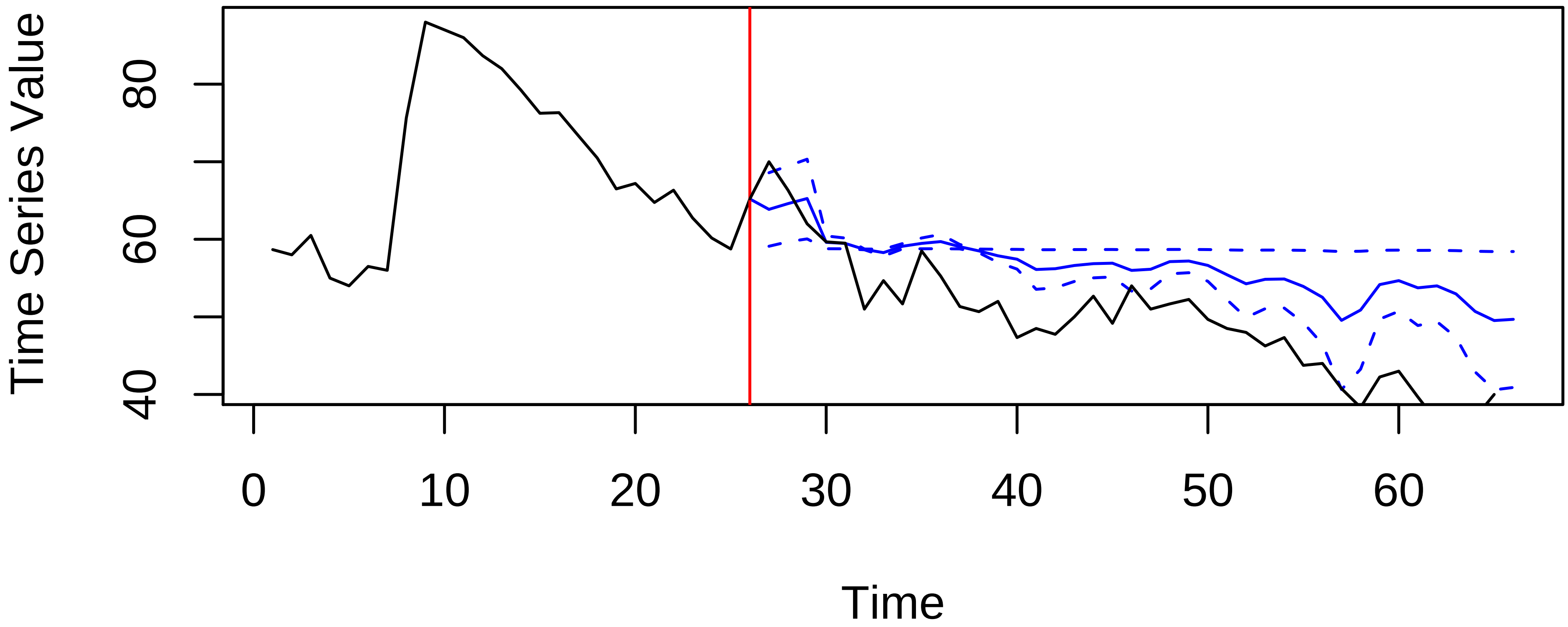
In black: Actual approval



Wait – can the model predict the long run approval rate?

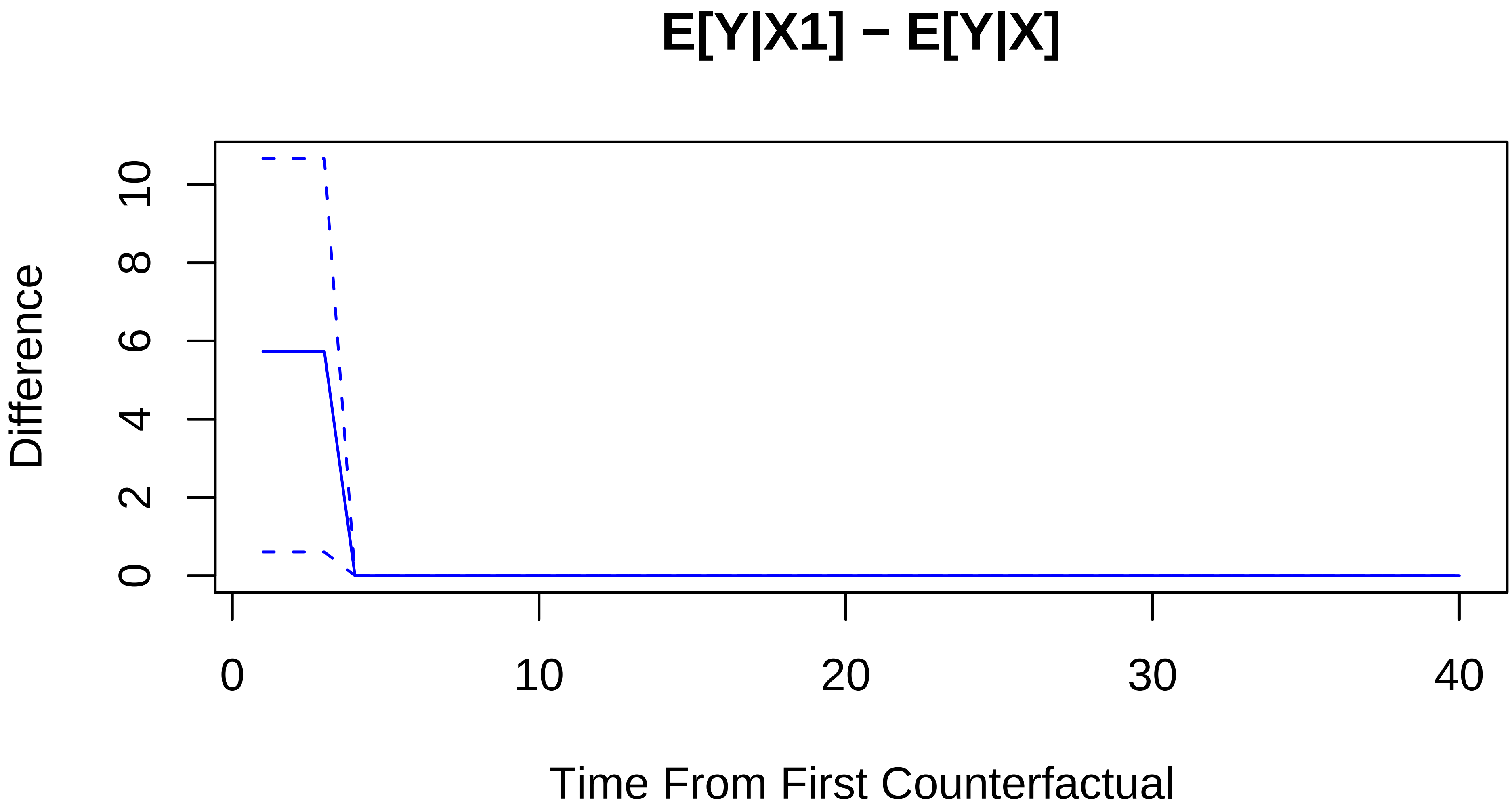


Wait – can the model predict the long run approval rate? Not even close



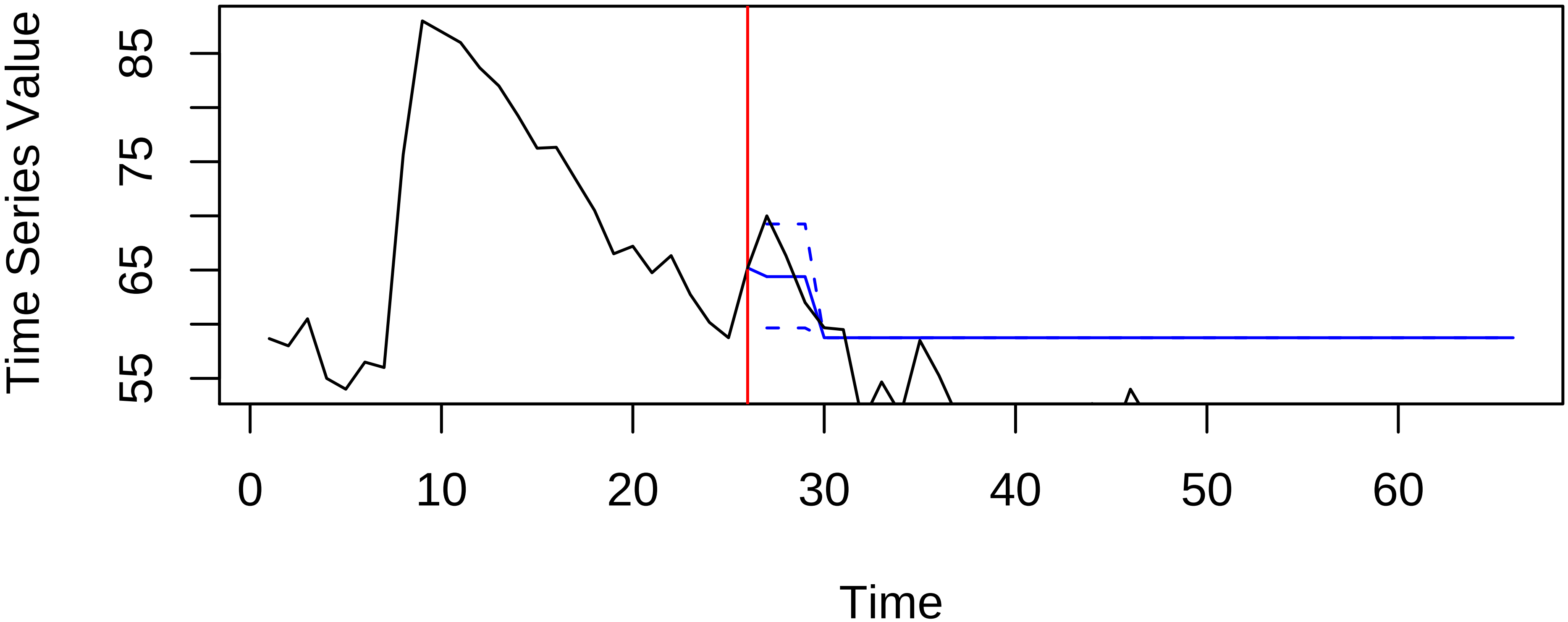
The model fit well for the first few months, then stays close to the ex ante "mean" approval

But reality (which is I(1)) drifts off into the cellar



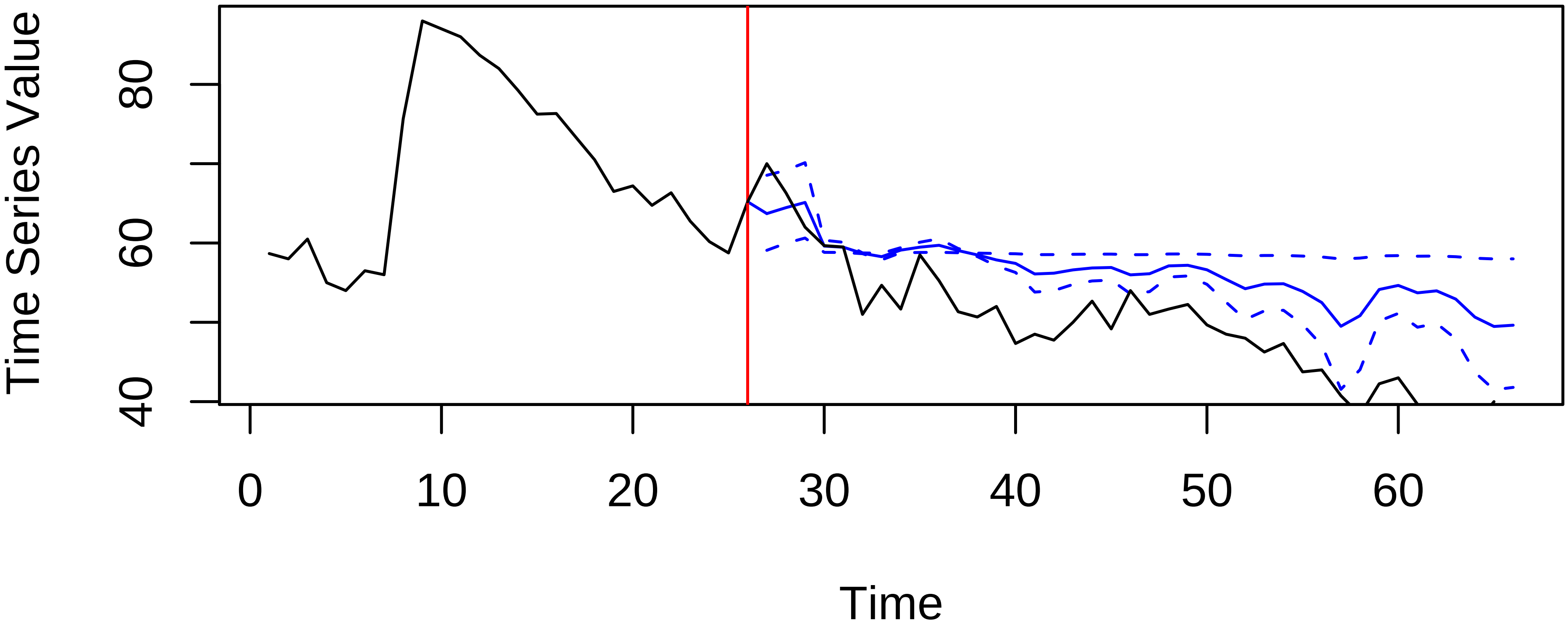
First differences show that all the action is in the short-run

Long-run predictions are not feasible with unit root processes



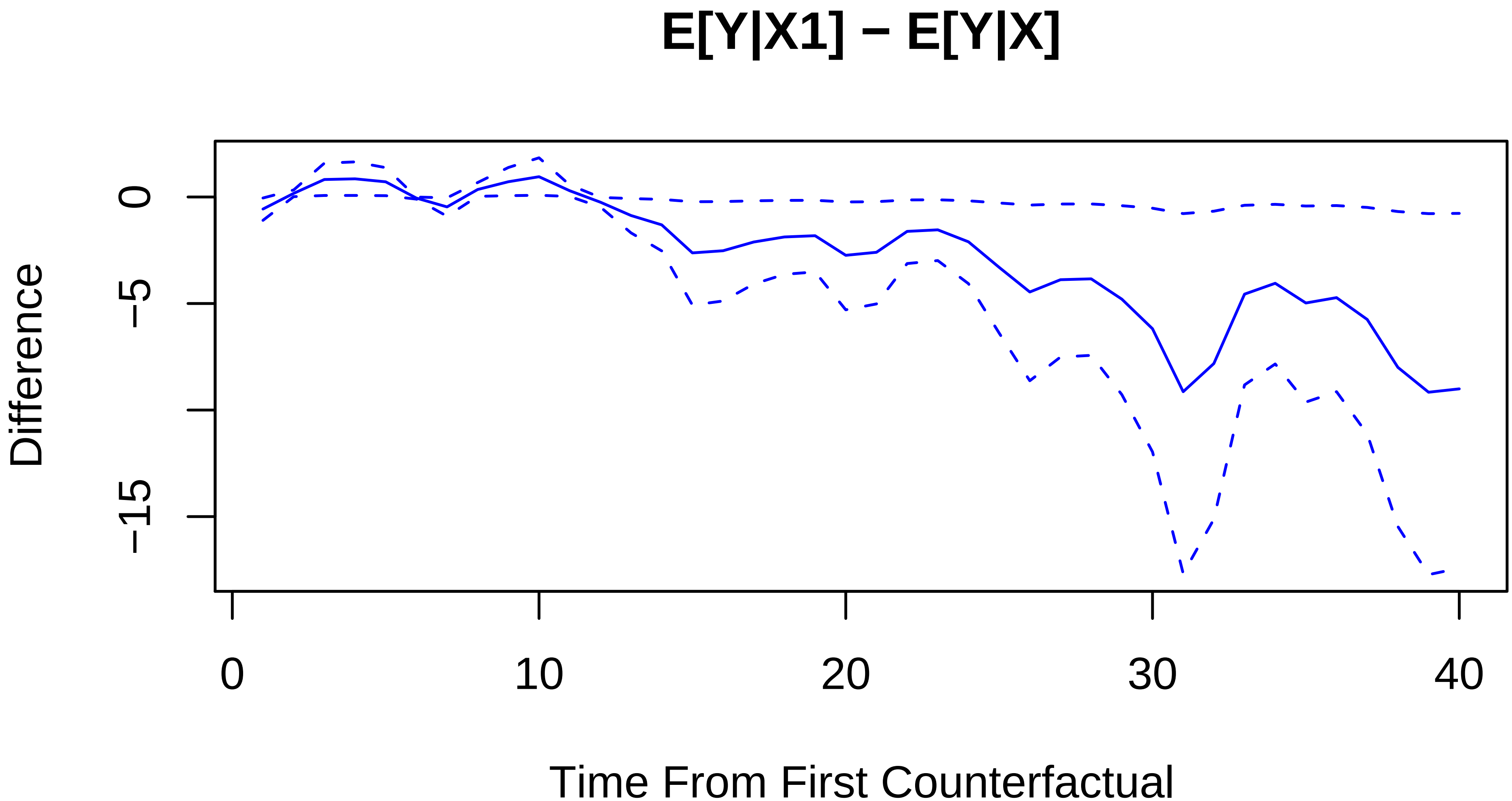
Suppose Oil had stayed at its pre-war price of $161/barrel

Then Bush's predicted popularity looks higher than the data



But wait – here are the factual “predictions” under the actual oil price

Miss the data by a mile


E[Y|X1] − E[Y|X]
Difference
0
−5
−15
0
10
20
30
40
Time From First Counterfactual

Limits of ARIMA

ARIMA(p,1,q) does a good job of estimating the short run movement of stationary variables

But does a terrible job with long-run levels

No surprise: The model includes no level information

While the observed level could drift anywhere

Limits of ARIMA

Using Δy_t as our response has a big cost

Purging all long-run equilibrium relationships from our time series

These empirical long-run relationships may be spurious (why we're removing them)

But what if they are not? What if y_t and x_t really move together over time?

Then removing that long-run relationship removes theoretically interesting information from our data

Since most of our theories are about long-run levels of our variables, we have usually just removed the *most* interesting part of our dataset!

Aside: Multiple Time Series

In our stationary time series example (road accidents),
we had a single continuous time series and a binary covariate

In our nonstationary example (approval),
we have two continuous time series

We chose to model approval as a function of oil prices,
but we could have reversed this,
and modeled oil prices as a function of approval

Why didn't we?

Aside: Multiple Time Series

In our stationary time series example (road accidents),
we had a single continuous time series and a binary covariate

In our nonstationary example (approval),
we have two continuous time series

We chose to model approval as a function of oil prices,
but we could have reversed this,
and modeled oil prices as a function of approval

Why didn't we? We had a theory-based model:
we don't think oil markets are driven by American opinions about their president

Theory informs the specification of "structual models"

Structural models could, of course, encompass multiple equations, as in SEM

But what if we don't have (or trust) a single theory about temporal relationships
among multiple continuous variable?

Aside: Multiple Time Series

There are atheoretical approaches to multiple continuous time series

Consider the following Vector Autoregression (VAR):

$$\text{approval}_t \quad = \quad \beta_0 + \sum_{j=0}^{J} \beta_{1j} \text{oil}_{t-j} + \sum_{k=1}^{K} \beta_{2k} \text{approval}_{t-k} + \varepsilon_t$$

Aside: Multiple Time Series

There are atheoretical approaches to multiple continuous time series

Consider the following Vector Autoregression (VAR):

$$
\begin{aligned}
\text{approval}_t &= \beta_0 + \sum_{j=0}^{J} \beta_{1j}\text{oil}_{t-j} + \sum_{k=1}^{K} \beta_{2k}\text{approval}_{t-k} + \varepsilon_t \\
\text{oil}_t &= \delta_0 + \sum_{j=0}^{J} \delta_{1j}\text{approval}_{t-j} + \sum_{k=1}^{K} \delta_{2k}\text{oil}_{t-k} + \eta_t
\end{aligned}
$$

Aside: Multiple Time Series

There are atheoretical approaches to multiple continuous time series

Consider the following Vector Autoregression (VAR):

$$\begin{aligned} \text{approval}_t &= \beta_0 + \sum_{j=0}^{J} \beta_{1j} \text{oil}_{t-j} + \sum_{k=1}^{K} \beta_{2k} \text{approval}_{t-k} + \varepsilon_t \\ \text{oil}_t &= \delta_0 + \sum_{j=0}^{J} \delta_{1j} \text{approval}_{t-j} + \sum_{k=1}^{K} \delta_{2k} \text{oil}_{t-k} + \eta_t \end{aligned}$$

This setup allows the price of oil to affect many things:
the future price of oil, through $\delta_{2k}, \quad k = 1, \ldots, K$;
the current approval rate, through β_{10}; and
the future approval rating, through $\beta 1 j, \quad j = 1, \ldots, J$

A parallel set of effects is possible for approval ratings

Aside: Multiple Time Series

$$\begin{aligned} \text{approval}_t &= \beta_0 + \sum_{j=0}^{J} \beta_{1j}\text{oil}_{t-j} + \sum_{k=1}^{K} \beta_{2k}\text{approval}_{t-k} + \varepsilon_t \\ \text{oil}_t &= \delta_0 + \sum_{j=0}^{J} \delta_{1j}\text{approval}_{t-j} + \sum_{k=1}^{K} \delta_{2k}\text{oil}_{t-k} + \eta_t \end{aligned}$$

We can use this system of equations to model
the short run effects of shocks in any variable on all other variables

Those shocks should gradually die out in stationary series

Note the absence of MA terms. We could add them, making a VARMA model

Note the absence of binary covariates. We could add them, too.

Of course, this VAR assumes stationarity of both oil and approval

That's a problem – so we need, as before, to difference these variables first

Aside: Multiple Time Series

$$\Delta\text{approval}_t = \psi_0 + \sum_{j=1}^{J} \psi_{1j} \Delta\text{oil}_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta\text{approval}_{t-k} + u_t$$

$$\Delta\text{oil}_t = \zeta_0 + \sum_{j=1}^{J} \zeta_{1j} \Delta\text{approval}_{t-j} + \sum_{k=1}^{K} \zeta_{2k} \Delta\text{oil}_{t-k} + v_t$$

Now this is a VAR on differenced time series
(I've changed parameters to emphasize this)

Now the recent changes in each variable can influence
subsequent changes in all variables

This model still does not have anything to say about the long run

If you want to know more about VAR models, start with Box-Steffensmeier Ch. 4;
there's a huge literature in econometrics

Cointegration

Consider two time series y_t and x_t:

$$
\begin{aligned}
x_t &= x_{t-1} + \varepsilon_t \\
y_t &= y_{t-1} + 0.6x_t + \nu_t
\end{aligned}
$$

where ε_t and ν_t are (uncorrelated) white noise

x_t and y_t are both: AR(1) processes, random walks, non-stationary, and I(1).

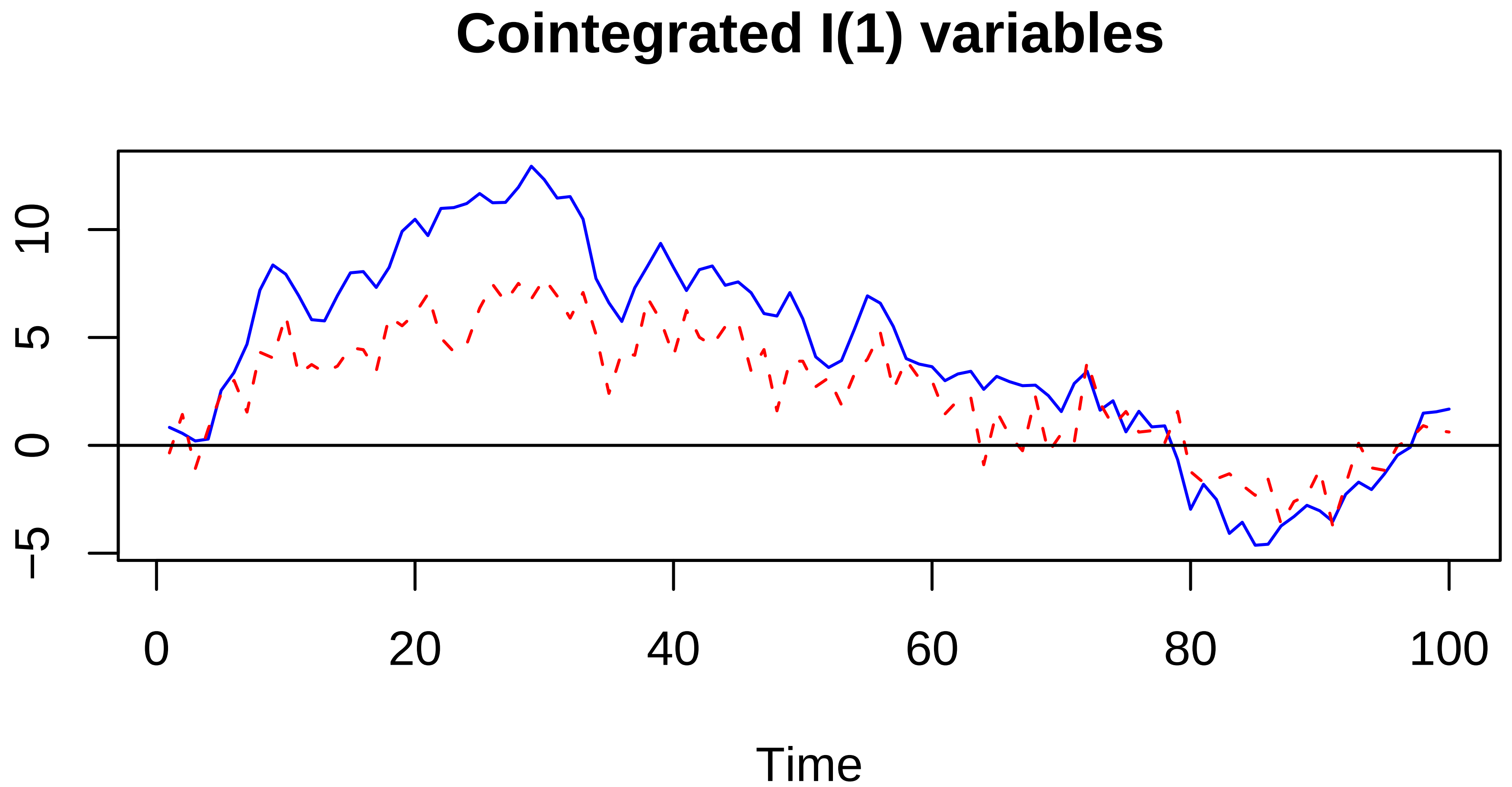
They are not spuriously correlated, but genuinely causally connected

Neither tends towards any particular level, but each tends towards the other

A particularly large ν_t may move y_t away from x_t briefly,
but eventually, y_t will move back to x_t's level

As a result, they will move together through t indefinitely

x_t and y_t are said to be *cointegrated*


Cointegrated I(1) variables
10
5
0
−5
0
20
40
60
80
100
Time

Cointegration

Any two (or more) variables y_t, x_t are said to be cointegrated if

1. each of the variables is I(d), $d \geq 1$

Cointegration

Any two (or more) variables y_t, x_t are said to be cointegrated if

1. each of the variables is I(d), $d \geq 1$ . . . usually, both are assumed to be I(1)

Cointegration

Any two (or more) variables y_t, x_t are said to be cointegrated if

1. each of the variables is I(d), $d \geq 1$ . . . usually, both are assumed to be I(1)

2. there is some *cointegrating vector* $\boldsymbol{\alpha}$ such that

$$\begin{aligned} z_t &= [y_t, x_t]'\boldsymbol{\alpha} \\ z_t &\sim I(0) \end{aligned}$$

   or in words, there is some linear combination of the non-stationary variables which is stationary

There may be many cointegrating vectors;
the cointegration rank r gives their number

Cointegration: Engle-Granger Two Step

Several ways to find the cointegration vector(s) and use it to anaylze the system

Simplest is Engle-Granger Two Step Method

Just repeated application of linear regression!

Works best if cointegration rank is $r = 1$ *and*

serial correlation is ARIMA(p,d,0) with clearly established p

that is, more complex or uncertain serial correlation can cause bias,
as with any other least square time series model

Fancier estimation techniques will address this limitation

Cointegration: Engle-Granger Two Step

Step 1: Estimate the cointegration vector by least squares with no constant:

$$y_t = \alpha_1^* x_{t-1} + \alpha_2^* x_{t-2} + \ldots + \alpha_K^* x_{t-K} + z_t$$

This gives us the cointegration vector $\alpha = (1, -\alpha_1^*, -\alpha_2^*, \ldots - \alpha_K^*)$

and the long-run equilibrium path of the cointegrated variables, $\hat{z}_t$

We can test for cointegration by checking that $\hat{z}_t$ is stationary

Note that the usual unit root tests work, but with different critical values

This is because the $\hat{\alpha}$'s are very well estimated: "super-consistent"
(converge to their true values very fast as T increases)

Cointegration: Engle-Granger Two Step

Step 2: Estimate an Error Correction Model

After obtaining the equilibrium $\hat{z}_t$'s and confirming they are I(0), we can estimate a particularly useful specification known as an *error correction model*, or ECM

ECMs simultaneously estimate long- and short-run effects for a system of cointegrated variables

Better than ARIMA(p,d,0) because we don't throw away level information

ECMs are simple generalizations of VARs in the differences of our time series

Interestingly, ECMs can also be estimated with least squares

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$\Delta y_t \quad = \quad \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$
\begin{aligned}
\Delta y_t &= \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t \\
\Delta x_t &= \zeta_0 + \gamma_2 \hat{z}_{t-1} + \sum_{j=1}^{J} \zeta_{1j} \Delta y_{t-j} + \sum_{k=1}^{K} \zeta_{2k} \Delta x_{t-k} + v_t
\end{aligned}
$$

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$\begin{aligned}
\Delta y_t &= \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t \\
\Delta x_t &= \zeta_0 + \gamma_2 \hat{z}_{t-1} + \sum_{j=1}^{J} \zeta_{1j} \Delta y_{t-j} + \sum_{k=1}^{K} \zeta_{2k} \Delta x_{t-k} + v_t
\end{aligned}$$

These equations are the "error correction" form of the model

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$\begin{aligned}\Delta y_t &= \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t \\ \Delta x_t &= \zeta_0 + \gamma_2 \hat{z}_{t-1} + \sum_{j=1}^{J} \zeta_{1j} \Delta y_{t-j} + \sum_{k=1}^{K} \zeta_{2k} \Delta x_{t-k} + v_t\end{aligned}$$

These equations are the "error correction" form of the model

Like the VAR on which it is based (the one we saw earlier!),
it shows the short-run relationships across all our time series

Unlike a VAR, an error correction model (ECM) also captures
how y_t and x_t respond to deviations from their long run relationship

(Technically, the above is a Vector ECM or VECM model,
which is the cointegrated generalization of VAR)

Cointegration: Engle-Granger Two Step

Let's focus on the evolution of Δy_t as a function of its lags, lags of Δx_t, and the "error" in the long-run equilibrium, $\hat{z}_{t-1}$:

$$\Delta y_t = \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

Cointegration: Engle-Granger Two Step

Let's focus on the evolution of Δy_t as a function of its lags, lags of Δx_t, and the "error" in the long-run equilibrium, $\hat{z}_{t-1}$:

$$\Delta y_t = \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

$\gamma < 0$ must hold for at least one γ:
This is the speed of adjustment back to equilibrium;
larger negative values imply faster adjustment

This is the central assumption of cointegration:
In the long run, y_t and x_t cannot diverge too much

So short-run differences must be made up later by convergence

For example, y_t (or x_t) must *eventually reverse course* after a big shift away from x_t

A negative γ_1 shows how quickly y_t reverses back to x_t

Cointegration: Engle-Granger Two Step

Recall our cointegrated time series, y_t and x_t:

$$
\begin{aligned}
x_t &= x_{t-1} + \varepsilon_t \\
y_t &= y_{t-1} + 0.6 x_t + \nu_t
\end{aligned}
$$

To estimate the Engle-Granger Two Step for these data, we do the following in R:

```
set.seed(123456)

# Generate cointegrated data
e1 <- rnorm(100)
e2 <- rnorm(100)
x <- cumsum(e1)
y <- 0.6*x + e2

# Run step 1 of the E-G two step
coint.reg <- lm(y ~ x -1)
coint.err <- residuals(coint.reg)
```

```r
# Check for stationarity of the cointegration vector
punitroot(adf.test(coint.err)$statistic, trend="nc")

# Make the lag of the cointegration error term
coint.err.lag <- coint.err[1:(length(coint.err)-2)]

# Make the difference of y and x
dy <- diff(y)
dx <- diff(x)

# And their lags
dy.lag <- dy[1:(length(dy)-1)]
dx.lag <- dx[1:(length(dx)-1)]

# Delete the first dy, because we are missing lags for this obs
dy <- dy[2:length(dy)]

# Estimate an Error Correction Model with LS
ecm1 <- lm(dy ~ coint.err.lag + dy.lag + dx.lag)
summary(ecm1)
```

```
Call:
lm(formula = x ~ y - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4565 -0.7754  0.3567  1.7542  5.7091

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
y  1.43472    0.05568   25.77   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.737 on 99 degrees of freedom
Multiple R-squared:  0.8702,Adjusted R-squared:  0.8689
F-statistic: 663.9 on 1 and 99 DF,  p-value: < 2.2e-16

> punitroot(adf.test(coint.err)$statistic, trend="nc")
Dickey-Fuller
 6.551997e-05
```

```
Call:
lm(formula = dy ~ coint.err.lag + dy.lag + dx.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9553 -0.5375  0.1538  0.7042  2.3240

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.02267    0.10381   0.218    0.828
coint.err.lag -0.96617    0.15864  -6.090 2.45e-08 ***
dy.lag        -1.05776    0.10848  -9.751 6.21e-16 ***
dx.lag         0.81035    0.11223   7.221 1.33e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.026 on 94 degrees of freedom
Multiple R-squared:  0.5456,Adjusted R-squared:  0.5311
F-statistic: 37.62 on 3 and 94 DF,  p-value: 4.624e-16
```

Cointegration: Johansen estimator

Alternatively, we can use the urca package, which handles unit roots and cointegration analysis:

```
# Create a matrix of the cointegrated variables
cointvars <- cbind(y,x)

# Perform cointegration tests
coint.test1 <- ca.jo(cointvars,
                   ecdet = "const",
                   type="eigen",
                   K=2,
                   spec="longrun"
                   )
summary(coint.test1)            # Check the cointegration rank here

# Using the output of the test, estimate an ECM
ecm.res1 <- cajorls(coint.test1,
                    r = 1,          # Cointegration rank
                    reg.number = 1) # which variable(s) to put on LHS
                                    # (column indexes of cointvars)
summary(ecm.res1$rlm)
```

Cointegration: Johansen estimator

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , without linear t

Eigenvalues (lambda):
[1]  3.105e-01  2.077e-02 -1.400e-18

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  2.06  7.52  9.24 12.97
r = 0  | 36.44 13.75 15.67 20.20

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             y.l2   x.l2 constant
y.l2      1.00000   1.00    1.000
x.l2     -0.58297  10.13   -1.215
```

```
constant -0.02961 -50.24  -38.501

Weights W:
(This is the loading matrix)

           y.l2      x.l2   constant
y.d -0.967715 -0.001015 -1.004e-18
x.d  0.002461 -0.002817 -2.899e-19
```

Cointegration: Johansen estimator

```
Call:
lm(formula = substitute(form1), data = data.mat)

Residuals:
   Min     1Q Median     3Q    Max 
-2.954 -0.536  0.150  0.712  2.318 

Coefficients:
      Estimate Std. Error t value Pr(>|t|)    
ect1    -0.968      0.158   -6.13  2.0e-08 ***
y.dl1   -1.058      0.108   -9.82  4.1e-16 ***
x.dl1    0.809      0.112    7.26  1.1e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.02 on 95 degrees of freedom
Multiple R-squared: 0.546,      Adjusted R-squared: 0.532 
F-statistic: 38.1 on 3 and 95 DF,  p-value: 2.97e-16
```

Example: Approval

Return to our Bush approval example, and estimate an ECM equivalent to the ARIMA(0,1,0) model we chose:

```
Residuals:
   Min     1Q Median     3Q    Max 
-7.140 -1.675 -0.226  1.643  5.954 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
ect1           -0.1262     0.0301   -4.20  9.4e-05 ***
sept.oct.2001  19.5585     2.1174    9.24  5.4e-13 ***
iraq.war        5.0187     1.6243    3.09   0.0031 ** 
approve.dl1    -0.3176     0.0945   -3.36   0.0014 ** 
avg.price.dl1  -0.0505     0.0259   -1.95   0.0561 .  
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1 

Residual standard error: 2.67 on 58 degrees of freedom
Multiple R-squared: 0.63,       Adjusted R-squared: 0.598 
F-statistic: 19.8 on 5 and 58 DF,  p-value: 1.91e-11
```

Cointegration: Final Thoughts

Cointegration and ECMs give us a way to cope with nonstationary time series without throwing away levels information

They provide information on short-run effects *and* long-run tendencies towards equilibrium

They *do not* tell us exact long-run destinations,
because for nonstationary series there isn't one

Could you use ECM to talk about long-run equilibria in stationary time series?

Many methodologists think this is possible and useful