

# Bayesian Time Series

Eric Lee, 11793249

2022-10-09

## INTRODUCTION

A time series is a stochastic process, whereby the future value of a series is unknowable with precision. Time series is defined as the study of a random variable throughout time. Unlike cross sectional data, time series data are not sampled from a population. Instead, we have a process that continues throughout time, and we sample from this process at different periods. Therefore, we cannot state time series data are independent, which means our ordinary least square estimates are not the best linear unbiased estimator. However, we can invoke the properties of stationarity and weak dependence in order use OLS.

## STATIONARITY AND WEAK DEPENDENCE

There are three conditions that must be met for a series to be considered stationary (Reale, 2022).

These being:

1.  $E(X_t) = \mu$ , the expected value of our series is constant and does not vary across time.
2.  $Var(X_t) = \sigma^2$ , the variance of  $X_t$  is constant and is not time dependent.
3.  $Cov(X_t, X_{t+h}) = g(h)$ , the covariance between two different points in time is a function of the gap between these two points,  $h$ , and not time  $t$ .

Weak dependence a substitute for the random sampling assumption in cross sectional data and is defined as:

$$Corr(X_t, X_{t+h}) = 0, \text{ as } h \rightarrow \infty$$

This states that the correlation between two data points should tend to zero fast enough as the distance between the the points increase. Placing these restrictions on our data makes it behave in a similar manner to cross sectional data. Therefore, stationarity and weak dependence plays a crucial role in time series analysis. If these conditions are not met, our OLS will not be BLUE. It is usually the case that we must manipulate the data e.g., taking the difference, so that these requirements are met.

## **FREQUENTIST VS BAYESIAN TIME SERIES**

Assuming we are using linear regression for the time series model, the main difference between the frequentist and Bayesian approach is that Bayesian allows us to incorporate other information using a prior. If we state our time series model as an auto regressive order one process, the model will be:

$$\begin{aligned} X_t &= \phi X_{t-1} + W_t \\ W_t &\sim N(0, \sigma_W^2) \\ |\phi| &\leq 1 \end{aligned}$$

In the classical frequentist approach, we will get the data and produce a likelihood function that is given by:

$$f(Y|\phi, \sigma^2) = (2\pi\sigma^2)^{\frac{-T}{2}} \exp\left(-\frac{(Y - \phi X)^T(Y - \phi X)}{2\sigma^2}\right)$$

We will try find the optimal parameter by taking the partial derivative *w. r. t*  $\phi$ , to derive an estimator. This estimator will be in the form:

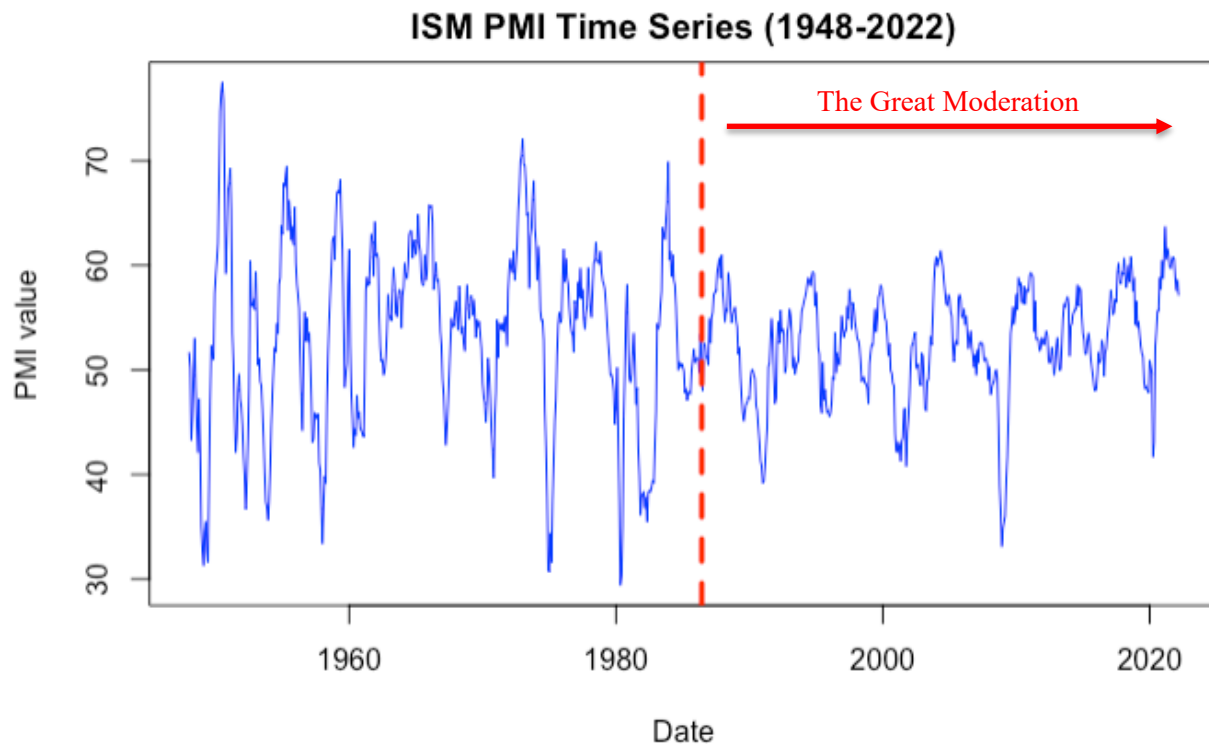
$$\hat{\phi} = (X^T X)^{-1} (X^T y)$$

In the Bayesian approach, we do not take the parameters of the model as a given. Instead, we treat it as a random variable that has its own distribution, which is what the prior is used for. Therefore, using Bayes Rule, we get:

$$\begin{aligned} f(\theta|x) &= \frac{f(x|\theta)f(\theta)}{f(x)} \\ f(\phi, \sigma^2|x) &= \frac{f(x|\phi, \sigma^2)f(\phi)f(\sigma^2)}{f(x)} \\ f(\phi, \sigma^2|x) &\propto f(x|\phi, \sigma^2)f(\phi)f(\sigma^2) \end{aligned}$$

## DATA

The data will be using to fit a Bayesian time series model is the Institute for Supply Management (ISM) Purchasing Manager Index (PMI). This is a diffusion index that is meant to be a barometer to the economy. Every month, the Institute for Supply Management surveys supply chain management professionals on ten different categories, which are, new orders, production, employment, supplier delivery, inventories, customers' inventories, prices, backlog orders, new export orders and imports (ISM Report on Business, 2022). The headline PMI score is made up of these ten categories. The PMI have a simple interpretation. Values over 50 means more than 50% of the survey participants saw general growth in the categories mentioned, which can be interpreted as a growth in economic activity. Conversely, values under 50% means less than 50% of the survey participants saw growth, which can be interpreted as a shrink in the economy. This is a very important indicator for financial market participants and is thought to be a leading indicator of the economy, i.e., the PMI score moves before the economy, although not perfectly. Below is a graph of the series.



From the graph of the series, it does appear to be stationary, as it oscillates between a mean value of 50. It is also evident that there are two distinct regimes present in the series. For example, the variance from 1948-1985 is different from the variance from 1985-2022, which voids the assumption of constant variance and makes the series non-stationary. The reason for the two distinct time periods can be explained by economic theory. The period from 1985-2022 is known as the great moderation. During this period, there was a decrease in volatility in a wide range of economic measures such as GDP, industrial production, and headline unemployment. The deduction in volatility can be attributed to policy makers, such as the United States Federal Reserve, taking a more active approach to managing the economy, through the manipulation of interest rate (Hakkio, n.d.). This in effect, dampened fluctuations in the business cycles. Although there were less recessions than before, it was more devastating when it occurred, which

is evident by the Asian Financial Crisis in 1997 and The Great Financial Crisis of 2008.

Therefore, due to the change in regimes, it will be inappropriate to use the whole series and our analysis will only be focused on the period pertaining to The Great Moderation.

## DERIVATION OF LIKELIHOOD FUNCTION

We will model our data through an autoregressive order one process which will be demeaned.

An AR (1) process with mean zero and Gaussian white noise can be written as:

$$\begin{aligned}X_t &= \phi X_{t-1} + W_t \\W_t &\sim N(0, \tau) \\X_t &\sim N(\phi X_{t-1}, \tau)\end{aligned}$$

For an AR (1) process, we can use recursive substitution to rewrite  $X_t$  as a sum of white noise processes. This is useful in deriving the mean, variance, and covariance of  $X_t$ .

$$\begin{aligned}X_t &= \phi X_{t-1} + W_t \\&= \phi(\phi X_{t-2} + W_{t-1}) + W_t \\&= \phi^2 X_{t-2} + \phi W_{t-1} + W_t \\&= \phi^k X_{t-k} + \sum_{i=0}^{k-1} \phi^i W_{t-i}\end{aligned}$$

The above equation can be approximated by  $\sum_{i=0}^{k-1} \phi^i W_{t-i}$ , as  $\phi^k X_{t-k}$  tends to zero as we take the limit of  $k$  to infinity. This is provided that  $|\phi| < 1$ , and that  $W_t$  has a finite variance and zero mean as the geometric series will not converge otherwise. Therefore, the variance and expected value of  $X_t$  can be given by:

$$\begin{aligned}
X_t &= \sum_{i=0}^{k-1} \phi^i W_{t-i} \\
E(X_t) &= 0 \\
Var(X_t) &= E[(X_t - E(X_t))^2] \\
&= E[X_t^2] \\
&= E[(W_t + \phi W_{t-1} + \phi^2 W_{t-2} + \phi^3 W_{t-3} \dots)^2] \\
&\propto E[(W_t^2 + \phi^2 W_{t-1}^2 + \phi^4 W_{t-2}^2 + \phi^6 W_{t-3}^2 \dots)] \\
&= \sigma^2 [1 + \phi^2 + \phi^4 + \phi^6 \dots] \\
&= \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} \\
&= \frac{\sigma^2}{1 - \phi^2} \\
&= \frac{1}{\tau(1 - \phi^2)}
\end{aligned}$$

Now we know the variance of the AR (1) model, we can try ascertaining the likelihood function in the numerator of the Bayes formula. The joint likelihood of the data  $f(x_t, x_{t-1}, \dots, x_1 | \phi, \sigma^2)$  can be simplified as our AR (1) model assumes that  $x_t$  is only dependent on the previous value  $x_{t-1}$  and a parameter  $\phi$ . Essentially, we can think of the joint likelihood as being chained by the previous value at each period. However, if our chain is defined by

$f(x_t | x_{t-1}), f(x_{t-1} | x_{t-2}), f(x_{t-2} | x_{t-3}) \dots f(x_2 | x_1), f(x_1 | ???)$ , the last value  $x_1$  is not

dependent on any previous value (MLE of a Gaussian AR(1) Process, n.d.). Therefore, our

likelihood function is dependent on knowing the value of  $x_1$  which is distributed  $x_1 \sim$

$N\left(0, \frac{1}{\tau(1-\phi^2)}\right)$ , with the variance being derived above. This means our likelihood function has

two parts. The first part is a product of all  $x_{t-n+1}$ , and the second is  $x_t$ . To find  $f(x_1)$ , we can

input the above variance into the PDF of the normal to get:

$$\begin{aligned}
f(x_t, x_{t-1}, \dots, x_1 | \phi, \tau) &= f(x_t | x_{t-1}, \phi, \tau) f(x_{t-1} | x_{t-2}, \phi, \tau) \dots f(x_2 | x_1, \phi, \tau) f(x_1) \\
x_1 &\sim N(0, \frac{1}{\tau(1-\phi^2)}) \\
f(x_1) &= \sqrt{\frac{\tau(1-\phi^2)}{2\pi}} \exp\left(-\frac{\tau(1-\phi^2)}{2} x_1^2\right)
\end{aligned}$$

Assuming our  $X_t$  is Gaussian, we can formally write the joint likelihood of the AR (1) process

as:

$$\begin{aligned}
f(x_t, x_{t-1}, \dots, x_2 | x_1, \phi, \tau) &= f(x_1 | \phi, \tau) \prod_{t=2}^n f(x_t | x_{t-1}, \tau, \phi) \\
&= \sqrt{\frac{\tau(1-\phi^2)}{2\pi}} \exp\left(-\frac{\tau(1-\phi^2)}{2} x_1^2\right) \cdot \prod_{t=2}^n \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2} (x_t - \phi x_{t-1})^2\right) \\
&= \sqrt{1-\phi^2} \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left(-\frac{\tau}{2} \sum_{t=2}^n (x_t - \phi x_{t-1})^2 + (1-\phi^2) x_1^2\right) \\
l(x_t, x_{t-1}, \dots, x_2 | x_1, \phi, \tau) &= \frac{1}{2} \log(1-\phi^2) + \frac{n}{2} \log\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2} \left(\sum_{t=2}^n (x_t - \phi x_{t-1})^2 + (1-\phi^2) x_1^2\right)
\end{aligned}$$

## PRIORS

For our priors we decided to choose uninformative ones. Since  $\phi$  is bounded by -1 to 1, we set the prior for it as zero, as it seems like a good middle ground. The prior for  $\tau$  was more difficult to set, thus we went with the standard 0.01 values.

$$\begin{aligned}
\phi &\sim N(\mu_0, \tau_0) \\
\mu_0 &= 0 \\
\tau_0 &= 0.01
\end{aligned}$$

$$\begin{aligned}
\tau &\sim \text{gamma}(a, b) \\
a &= 0.01 \\
b &= 0.01
\end{aligned}$$

## METROPOLIS HASTINGS ALGORITHM

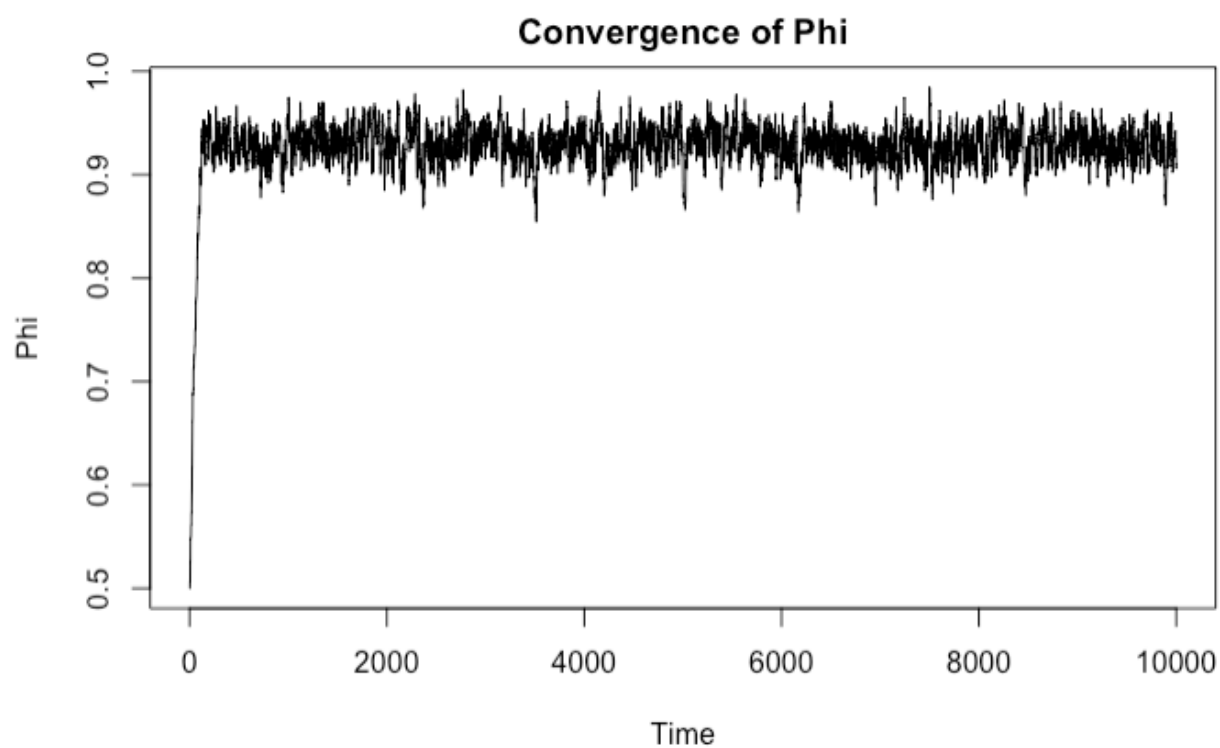
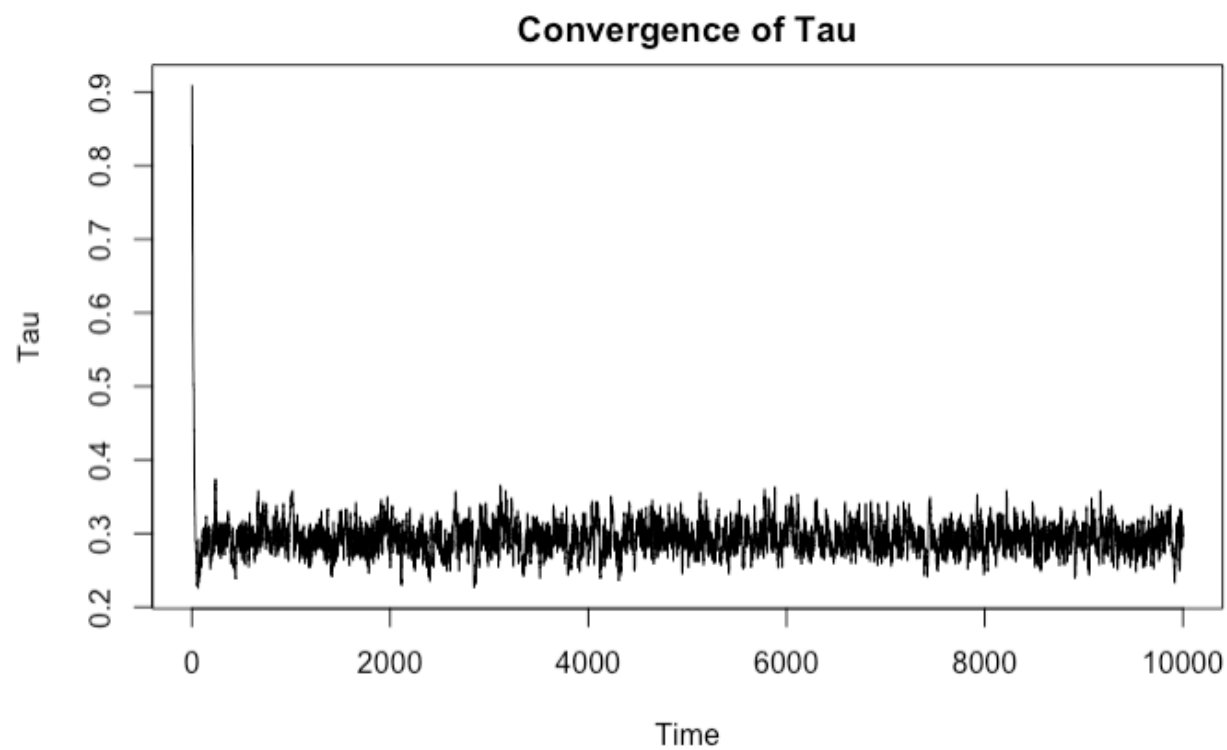
Now we have our posterior distribution to a proportionality constant we can use the Metropolis Hastings Algorithm. Our Metropolis Hastings Algorithm will function in a similarly manner to a Gibbs Sampler. Meaning, we will have initial values of both  $\phi$  and  $\tau$ . For example, in the first iteration of the log likelihood function for  $\phi$ , we will assume  $\tau$  is known and is set to  $\tau_{initial}$ . If the proposed  $\phi$  is accepted, then we will input this to the log likelihood function of  $\tau$ . We will repeat this for  $n$  number of iterations. Essentially, the algorithm is like the Gibbs Sampler except not every new value gets accepted.

$$\begin{aligned} f(\tau, \phi|x) &\propto f(x_t, x_{t-1}, \dots, x_2|x_1, \phi, \tau) f(\tau|a, b) f(\phi|\mu_0, \tau_0) \\ &= \sqrt{1 - \phi^2} \left( \frac{\tau}{2\pi} \right)^{n/2} \exp \left( -\frac{\tau}{2} \sum_{t=2}^n (x_t - \phi x_{t-1})^2 + (1 - \phi^2) x_1^2 \right) \\ &\quad \cdot \frac{b^a}{\Gamma(a)} \tau^{a-1} e^{-b\tau} \cdot \left( \frac{\tau_0}{2\pi} \right)^{1/2} \exp \left( -\frac{\tau_0}{2} (\phi - \mu_0)^2 \right) \end{aligned}$$

Below, we can see the convergence of  $\phi$  and  $\tau$ .  $\phi$  has converged to a value of 0.93 and  $\tau$  0.29, which translates to a variance of 3.41. When we compare our parameters obtained using a Bayesian method to the traditional method, they are similar, as indicated by the table. Note, the model output which these figures are based on are in the appendix.

	Bayesian Method	Traditional Method using arima (1, 0, 0) function
Variance	3.4	3.4
Phi	0.93	0.93
Standard Error	0.02	0.017





Therefore, the final model for the AR (1) process of our ISM PMI time series data is:

$$X_t = 0.93X_{t-1} + W_t$$

This is assuming the series is an AR (1) process, which it may not be. As you can see  $\phi$  is less than one and indicates a stationary process, and because it is stationary the variance is finite.

## APPENDIX

Hakkio, C. S. (n.d.). *The Great Moderation*. Federal Reserve History.

<https://www.federalreservehistory.org/essays/great-moderation>

*ISM Report On Business*. (2022). ISM. <https://www.ismworld.org/supply-management-news-and-reports/reports/ism-report-on-business/>

*MLE of a Gaussian AR(1) Process*. (n.d.). SDU Education.

<http://course.sdu.edu.cn/G2S/eWebEditor/uploadfile/20140110134920017.pdf>

Reale, M. (2022, September). *Introduction to AR, MA and ARMA Models*. Marco Reale's Lecture Slides.

<https://learn.canterbury.ac.nz/mod/resource/view.php?id=2364749&redirect=1>

## CODE

```
# Loading PMI data
data <- read.csv("ISM-MAN.csv")
data$Date <- as.Date(data$Date, format = "%Y-%m-%d")

# plotting series
plot(data$Date, data$PMI,
      xlab = "Date",
      ylab = "PMI value",
      main = "ISM PMI Time Series (1948-2022)",
      type = "l",
      col = "blue")
abline(v = 6000, col="red", lwd=3, lty=2)

# selecting subset of the data
library(forecast)

data <- data[data$Date > "1985-01-01", ]

### Metropolis Hastings Algorithm

# de meaning the series
data$PMI <- data$PMI - mean(data$PMI)

# my likelihood
log.likelihood <- function(tau, phi, data){
```

```

n <- length(data)
xt <- data[2:n] #  $x(T)$ 
xt1 <- data[1:n-1] #  $x(T-1)$ 
joint.likelihood = 0.5*log(1-phi^2) + (n/2)*log(tau/(2*pi)) -
  (tau/2)*(sum((xt - phi*xt1)^2)+(1-phi^2)*data[1]^2)
return(joint.likelihood)
}

# phi prior (normal)
phi.prior <- function(phi,mu0,tau0){
  return(dnorm(phi,mu0,sd=1/sqrt(tau0),log=T))
}

# tau prior (gamma)
tau.prior <- function(tau,a,b){
  return(dgamma(tau,a,b,log=T))
}

# arima(data$PMI,order=c(1,0,0))

# initial values
N <- 10000
tau <- 1
phi <- 0.5
tau.vec <- phi.vec <- numeric(length=N)

for (i in 1:N){
  # sample mu, proposal = normal
  new.phi <- rnorm(1,phi,0.01)
  logR <- (log.likelihood(tau,new.phi,data$PMI) - log.likelihood(tau,phi,data$PMI)) +
    (phi.prior(new.phi,0,0.01) - phi.prior(phi,0,0.01))
  if (log(runif(1,0,1)) < logR){phi <- new.phi}
  phi.vec[i] <- phi

  new.tau <- exp(rnorm(1,log(tau),0.05))
  logR <- (log.likelihood(new.tau, phi, data$PMI) - log.likelihood(tau, phi, data$PMI)) +
    (tau.prior(new.tau, 0.01, 0.01) - tau.prior(tau,0.01,0.01)) +
    (log(new.tau) - log(tau))
  if (log(runif(1,0,1)) < logR){tau <- new.tau}
  tau.vec[i] <- tau
}

ts.plot(tau.vec,
  ylab = "Tau",
  main = "Convergence of Tau")

```

```

ts.plot(phi.vec,
        ylab = "Phi",
        main = "Convergence of Phi")

variance <- 1/mean(tau.vec[1000:N])
phi <- mean(phi.vec[1000:N])
sprintf("Bayesian variance: %s", round(variance, 2))

sprintf("Bayesian phi: %s", round(phi, 2))

sprintf("Bayesian SE: %s", round(sqrt(variance/9000), 2))

arma(data$PMI, order=c(1,0,0))

```

## COMPARING BAYESIAN METHOD WITH TRADITIONAL METHOD

### Parameter estimate using Bayesian method

```
## [1] "Bayesian variance: 3.4"
```

```
## [1] "Bayesian phi: 0.93"
```

```
## [1] "Bayesian SE: 0.02"
```

### Parameter estimate using traditional method

```
## Call:
```

```
## arima(x = data$PMI, order = c(1, 0, 0))
```

```
##
```

```
## Coefficients:
```

```
##          ar1  intercept
```

```
##          0.9282    0.0497
```

```
## s.e.      0.0173    1.1818
```

```
##
```

```
## sigma^2 estimated as 3.396: log likelihood = -906.45, aic = 1818.91
```