State Space Model for the Gender Gap in the Labor Market Using GDP as a Covariate

Sahar Shirazi

2025-06-18

1. Introduction

This project employs advanced state space modeling techniques to analyze the gender gap in the Italian labor market. Specifically, it models the time series of the difference between male and female unemployment rates, denoted by y_t. The analysis is conducted using three models:

-Local Level Model
-Local Linear Trend Model
-Regression Model with Lagged GDP Growth as a Covariate

In the third model, we incorporate the lagged change in GDP (one-quarter lag) to examine the impact of economic growth on the gender gap.

Unemployment rate data is obtained from the Eurostat database: https://ec.europa.eu/eurostat/databrowser/view/lfsq_ergan/default/table?lang=en

The analysis is implemented in R using the dlm package for dynamic linear models and the eurostat package to access macroeconomic data directly from Eurostat.

Before proceeding to data preparation, we briefly review the mathematical framework of state-space models.

2. Mathematical Formula

General Linear Gaussian State-Space Model

The general form of a linear Gaussian state-space model consists of two main equations:

State Equation

$$X_t = G_t X_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}(0, W_t)$$

Observation Equation

$$Y_t = F_t X_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, V_t)$$

Initial state is also assumed Gaussian:

$$X_0 \sim \mathcal{N}(m_0, C_0)$$

where

-G_t is State Transition Matrix describing how the hidden state vector X_t evolves over time.

-F_t is an observation (Design) Matrix which maps the latent state vector X_t to the observed variable Y_t.

-V_t is an observation Noise Variance (or Covariance) which represents the variance of the observation error v_t.

Local Level Model (First Order Polynomial)

This is a special case of the state-space model where:

$$G_t = 1, \quad F_t = 1$$

Thus, the model becomes:

State Equation

$$X_t = X_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}(0, W)$$

Observation Equation

$$Y_t = X_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, V)$$

This model is also known as the **locally constant model** or **random walk plus noise**.

Local Trend Model (Second Order Polynomial)

The local trend model extends the local level model by including a slope term in the state:

State Equations

$$\begin{aligned} X_{t,1} &= X_{t-1,1} + X_{t-1,2} + \omega_{1,t} \\ X_{t,2} &= X_{t-1,2} + \omega_{2,t} \end{aligned}$$

Observation Equation

$$Y_t = X_{t,1} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, v)$$

Matrix Form

State vector:

$$X_t = \begin{bmatrix} X_{t,1} \\ X_{t,2} \end{bmatrix}$$

State transition:

$$X_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} X_{t-1} + \omega_t$$

Where

$$\omega_t \sim \mathcal{N}(0, W)$$

and W is a covariance matrix. Observation matrix:

$$F_t = [1 \quad 0]$$

Observation:

$$Y_t = F_t X_t + \nu_t$$

The component X_{t-2} (slope) provides a local increment for the level $X_{t,1}$.

Time-Varying State-Space Model: Simple Dynamic Regression

This is a regression model where the coefficients can vary over time:

Observation Equation

$$Y_t = X_{t,1} + X_{t,2} Z_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, v)$$

Written as:

$$Y_t = F_t X_t + \nu_t, \quad F_t = (1, Z_t)$$

State Equation

$$X_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} X_{t-1} + \omega_t$$

Where $\omega_t \sim \mathcal{N}(0, W)$

Time-Varying Regression with Lag (Autoregressive Coefficients)

This version allows each state to follow an AR(1) process.

Observation Equation

$$Y_t = X_{t,1} + X_{t,2} Z_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, v)$$

State Equation

$$X_t = \begin{bmatrix} \rho_{11} & 0 \\ 0 & \rho_{22} \end{bmatrix} X_{t-1} + \omega_t$$

Where $\omega_t \sim \mathcal{N}(0, W)$, and ρ_{11}, ρ_{22} are autoregressive parameters for each state variable.

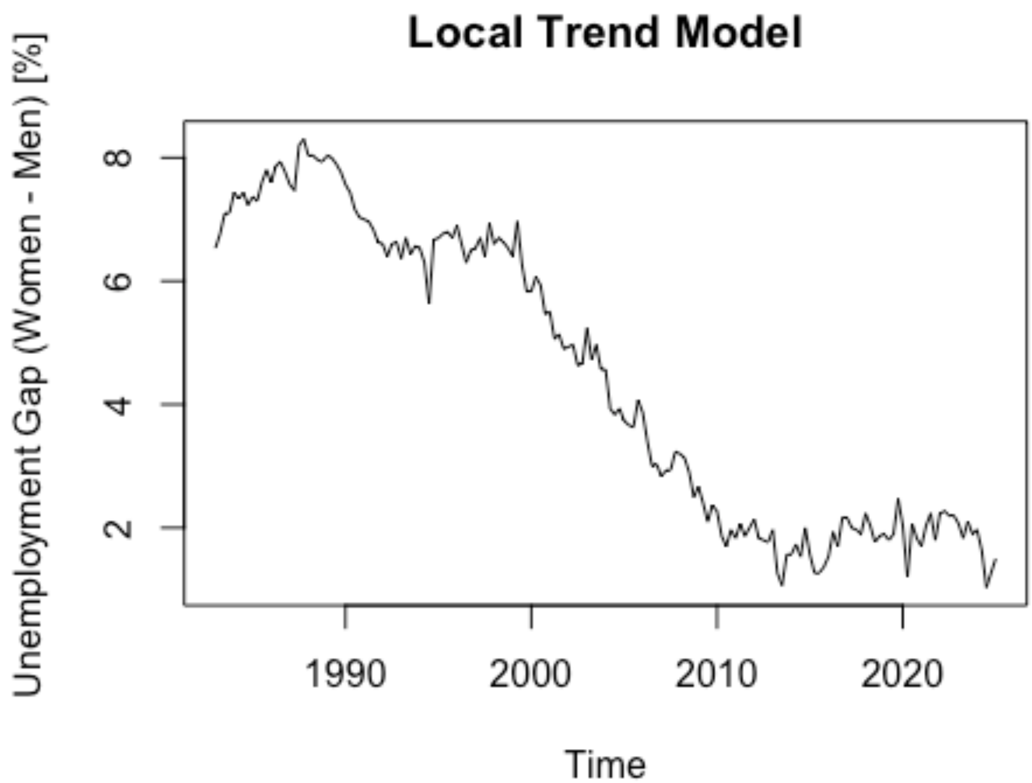
---

3. Data Preparation

In this part, we download monthly unemployment rates for women and men in Italy from Eurostat using the **eurostat** package. Then, we calculate the gender gap by subtracting male unemployment from female unemployment. Since our analysis works with quarterly data, we convert the monthly series to quarterly by averaging every three months. Next, we download quarterly GDP data for Italy and keep only the data starting from 1996. To make sure both the gender gap and GDP series match in time, we also trim the gender gap data to start from 1996. This way, both time series are aligned and ready for analysis.

```
library(eurostat)
DF=get_eurostat(id='une_rt_m',filters=list(
  geo='IT',s_adj='SA',age='TOTAL',unit='PC_ACT',sex='F') )

## Dataset query already saved in cache_list.json...

## Reading cache file
/var/folders/xg/n0p097vj2td1v939mpqjcpk40000gn/T//Rtmpy1z6cw/eurostat/b560d83
527f39bd40d0dcaebdb4e238a.rds

## Table  une_rt_m  read from cache file:
/var/folders/xg/n0p097vj2td1v939mpqjcpk40000gn/T//Rtmpy1z6cw/eurostat/b560d83
527f39bd40d0dcaebdb4e238a.rds

DM=get_eurostat(id='une_rt_m',filters=list(
  geo='IT',s_adj='SA',age='TOTAL',unit='PC_ACT',sex='M') )

## Dataset query already saved in cache_list.json...

## Reading cache file
/var/folders/xg/n0p097vj2td1v939mpqjcpk40000gn/T//Rtmpy1z6cw/eurostat/7a12bcd
dde08bf92722c0f57919c680b.rds

## Table  une_rt_m  read from cache file:
/var/folders/xg/n0p097vj2td1v939mpqjcpk40000gn/T//Rtmpy1z6cw/eurostat/7a12bcd
dde08bf92722c0f57919c680b.rds

#difference between unemployment rate for female and male in italy

y= ts(DF$values-DM$values, start = c(1983, 1), frequency = 12) #monthly data
y=aggregate(y, nfrequency = 4, FUN = mean,)   # quarterly data



plot(y, main = "Local Trend Model",
```

```r
     ylab = "Unemployment Gap (Women - Men) [%]",
     xlab = "Time")
```




```r
gdp=get_eurostat(id='namq_10_gdp',filters=list(geo='IT',na_item='B1GQ',s_adj=
'SCA',unit='PD20_EUR'))

## Dataset query already saved in cache_list.json...

## Reading cache file
/var/folders/xg/n0p097vj2td1v939mpqjcpk40000gn/T//Rtmpy1z6cw/eurostat/6e189fc
a71893132c200a2c0e145f9ce.rds

## Table  namq_10_gdp  read from cache file:
/var/folders/xg/n0p097vj2td1v939mpqjcpk40000gn/T//Rtmpy1z6cw/eurostat/6e189fc
a71893132c200a2c0e145f9ce.rds

gdp=ts(gdp$values,start=c(1975,1),frequency=4)
gdp=window(gdp,start = c(1996, 1))
y=window(y,start = c(1996, 1))
```

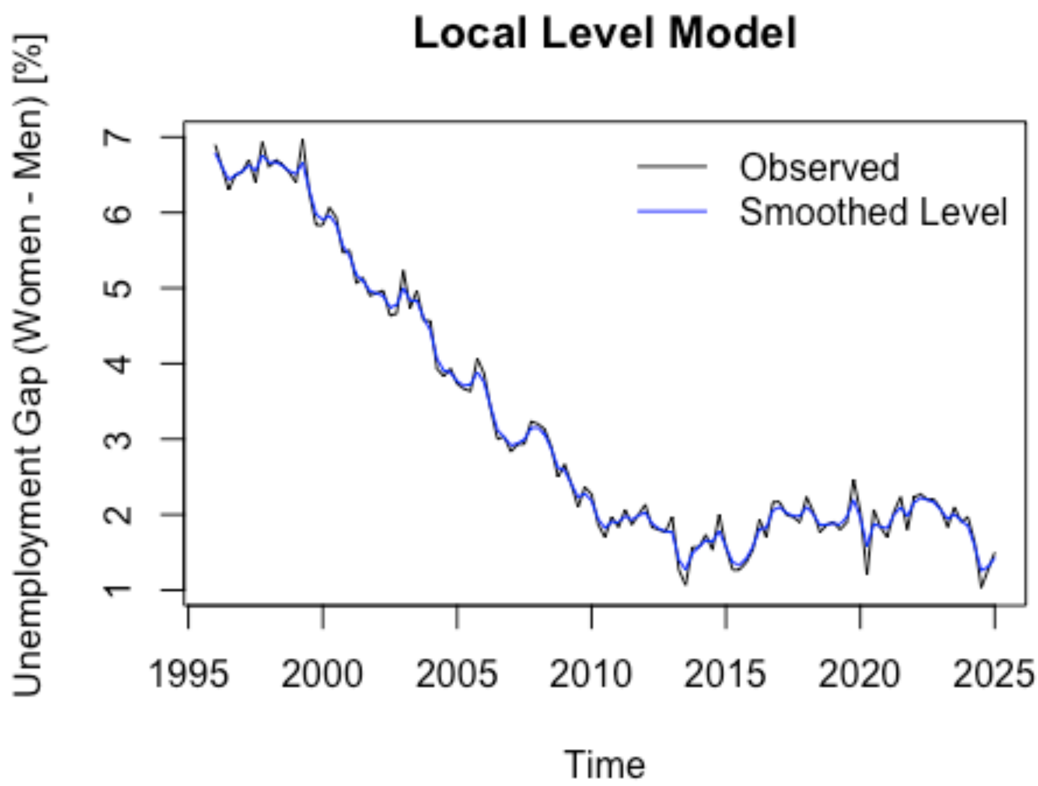
4. Models

4.1 Local Level Model

In the first section, a local level is modeled which assumes the gender gap follows a random walk with noise. In the following , $dlmModPoly(order = 1, \dots)$ and $dlmMLE(\dots)$creates and estimate this model. Moreover, $dlmSmooth(\dots)$ retrieves the smoothed state estimates (latent true gender gap over time). It is important to say that this model assumes no trend, only level changes.

In the local level model plot, the blue line is a denoised estimate of the true underlying gender gap and suggests short-term fluctuations around a drifting average.

```
fn1 <- function(par) {
  dlmModPoly(order = 1, dV = exp(par[1]), dW = exp(par[2]), C0 = exp(par[3]))
}

fit1 <- dlmMLE(y, rep(0, 3), build = fn1)
model1 <- fn1(fit1$par)
filtered1 <- dlmFilter(y, model1)
smoothed1 <- dlmSmooth(filtered1)
mu1 <- dropFirst(smoothed1$s)

plot(y, main="Local Level Model", ylab = "Unemployment Gap (Women - Men)
[%]",
     xlab = "Time")
lines(ts(mu1, start=start(y), frequency=4), col='blue')
legend("topright", legend = c("Observed", "Smoothed Level"),
       col = c("black", "blue"), lty = c(1, 1), bty = "n")
```



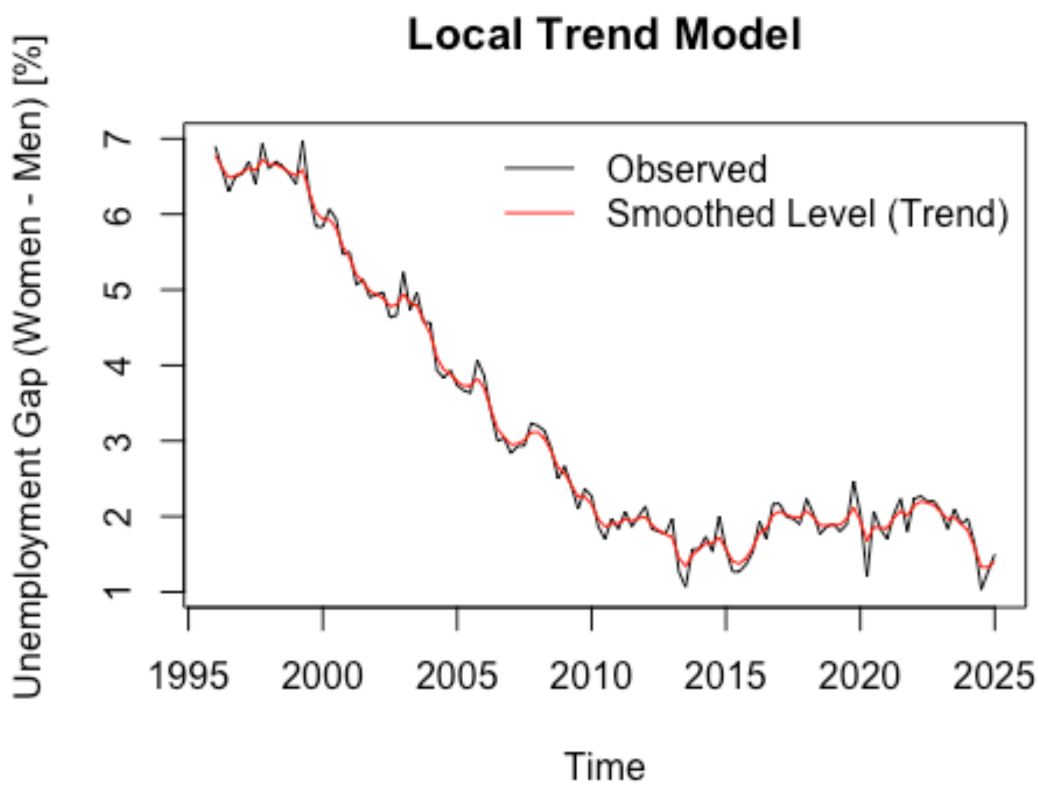
4.2 Local Linear Trend Model

In the second step, we can see the previous model but with an extension by allowing the gender gap to have a trend. $dlmModPoly(order = 2, \ldots)$allows both level and slope components. This model accounts for both persistent levels and gradual shifts in the gap, likely reflecting long-term structural changes.

The Local Linear Trend plot, Indicates a downward trend (red line)in the gap, suggesting structural improvements over time.

```
fn2 <- function(par) {
  dlmModPoly(order = 2, dV = exp(par[1]), dW = exp(par[2:3]), C0 =
diag(exp(par[4:5])))
}

fit2 <- dlmMLE(y, rep(0, 5), build = fn2)
model2 <- fn2(fit2$par)
smoothed2 <- dlmSmooth(dlmFilter(y, model2))
mu2 <- dropFirst(smoothed2$s[,1])

plot(y, main="Local Trend Model",
```

```r
     ylab = "Unemployment Gap (Women - Men) [%]",
     xlab = "Time")
lines(ts(mu2, start=start(y), frequency=4), col='red')
legend("topright", legend = c("Observed", "Smoothed Level (Trend)"),
       col = c("black", "red"), lty = c(1, 1), bty = "n")
```



4.3 Model with GDP Growth (1-lag)

This part of the project builds a state space model that explains the gender gap in unemployment in Italy using lagged GDP growth as a covariate. The aim is to see whether changes in the economy, measured by GDP growth, can help explain changes in the difference between male and female unemployment rates.

We start by computing the monthly gender gap (y_monthly) as the difference between female and male unemployment rates, and then convert it to quarterly data (y_quarterly) by averaging every three months. At the same time, we prepare the quarterly GDP time series (gdp_qtr) and calculate its log growth rate (log_gdp_growth), which represents economic growth from one quarter to the next.

Next, we align the gender gap and GDP growth series so they start at the same point in time. This ensures both series are compatible for modeling. The final response variable is the aligned gender gap (y), and the regressor matrix (X) includes both an intercept and the GDP growth.

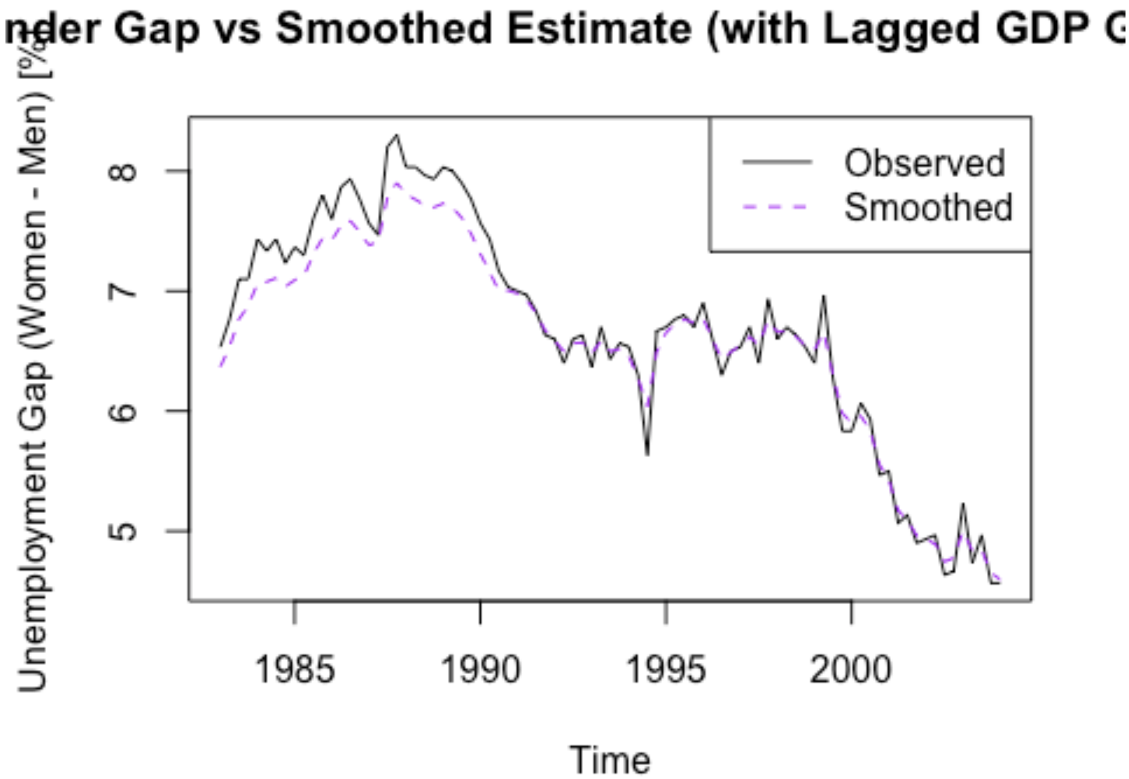
We define a state space regression model using the dlm function, where the observation depends on a time-varying intercept and the lagged GDP growth. The model parameters (variances of the noise terms) are estimated using Maximum Likelihood Estimation (dlmMLE). After estimation, we build the model with the best parameters and apply filtering and smoothing to recover the hidden trend (the level) behind the gender gap, now influenced by economic growth.

```
library(eurostat)
library(dlm)
library(ggplot2)

# Step 1: Monthly gender gap
y_monthly <- ts(DF$values - DM$values, start = c(1983, 1), frequency = 12)

# Step 2: Convert to quarterly
y_quarterly <- aggregate(y_monthly, nfrequency = 4, FUN = mean)

# Step 3: GDP time series
gdp_qtr <- ts(gdp, start = c(1975, 1), frequency = 4)

# Step 4: Compute log GDP growth
log_gdp_growth <- diff(log(gdp_qtr))  # starts from 1975 Q2

# Step 5: Shift y to match starting from GDP growth (which starts 1 period
later)
y_quarterly <- window(y_quarterly, start = start(log_gdp_growth))

## Warning in window.default(x, ...): 'start' value not changed

# Step 6: Intersect and align both series automatically
aligned <- ts.intersect(y = y_quarterly, x = log_gdp_growth)

# Step 7: Extract aligned components
y <- aligned[, "y"]
X <- cbind(1, aligned[, "x"])  # add intercept

# check length
stopifnot(length(y) == nrow(X))

# Step 8: Regressor matrix
X <- cbind(1, log_gdp_growth)


# Model function with observation and state variances
fn3 <- function(par) {
  dlm(
```

```
    FF = matrix(c(1, 0), nrow = 1),                    # Observation matrix
    V = exp(par[1]),                                    # Observation noise
    GG = diag(2),                                       # State transition
matrix
    W = diag(exp(par[2:3])),                            # State noise
    m0 = rep(0, 2),                                     # Initial state
    C0 = diag(1e7, 2),                                  # Initial state
covariance
    JFF = matrix(c(1, 2), nrow = 1),                    # Time-varying
regression: intercept + covariate
    X = X                                               # Covariate matrix (with
intercept + GDP growth)
  )
}
##step 9:estimate parameter MLE
# Initial guess for parameters: log variances
init_par <- rep(0, 3)

# Fit the model
fit3 <- dlmMLE(y, parm = init_par, build = fn3, hessian = TRUE)

# Checking convergence(which should be 0)
fit3$convergence
```

```
## [1] 0
```

```
# Step 10: Build Final Model with Estimated Parameters
# Build model using estimated parameters
model3 <- fn3(fit3$par)

#Step 11: Filtering and Smoothing
# Filter and smooth the series
filtered3 <- dlmFilter(y, model3)
smoothed3 <- dlmSmooth(filtered3)

# Extract the smoothed level component (state 1)
mu3 <- dropFirst(smoothed3$s[, 1])

#Step 12: Plot the Results

# Plot observed and smoothed estimates
ts.plot(y, mu3, col = c("black", "purple"),
        lty = 1:2, main = "Gender Gap vs Smoothed Estimate (with Lagged GDP
Growth)",
        ylab = "Unemployment Gap (Women - Men) [%]",
        xlab = "Time")
legend("topright", legend = c("Observed", "Smoothed"), col = c("black",
"purple"), lty = 1:2)
```



In this plot, we compare the actual gender gap data (black line) with the smoothed estimates (green line) from the model. The plot shows that economic conditions (measured by GDP growth) play a role in shaping the unemployment gender gap. Periods of economic decline or slow growth often see a divergence between male and female unemployment, and the model captures that interaction well. The green line's ability to mimic the actual gender gap suggests that incorporating GDP into the state space framework improves predictive and explanatory power.

To better understand how GDP growth affects the gender gap, we look at the second component of the state vector, which represents the coefficient β for lagged GDP growth. This coefficient changes over time, so we use the smoothed values from the model to see how it behaves. Plotting this coefficient helps us see whether GDP growth tends to increase or reduce the gender gap in unemployment.

```
beta_gdp <- dropFirst(smoothed3$s[, 2])
plot(beta_gdp, type = "l", col = "blue",
     main = "Time-Varying Coefficient of GDP Growth on Gender Gap",
     ylab = "β (GDP Coefficient)", xlab = "Time")
abline(h = 0, col = "red", lty = 2)
```

The final plot *the time-varying coefficient for lagged GDP growth* shows that the values are consistently positive, especially from 1991 onwards, where the coefficient becomes stable and remains above zero. This suggests that economic growth (as measured by GDP increase) has a consistent and positive impact on reducing the gender gap in unemployment. In other words, during periods of higher GDP growth, the difference between male and female unemployment rates tends to decrease — likely because women benefit more from economic improvements in the labor market. The stabilization of the coefficient after 1991 may reflect structural changes in the economy or policy interventions that made the relationship between GDP and the gender gap more consistent over time.

---

5. Conclusion

Over the past two decades, the gender gap in unemployment in Italy has gradually declined and stabilized, reflecting deep structural shifts in the labor market, evolving gender equality policies, and broader societal changes. The use of state space models in this project provided a powerful analytical framework to uncover these trends, effectively filtering out short-term noise and revealing underlying dynamics. Among the models applied, the local linear trend model successfully captured long-term improvements, while the regression model incorporating lagged GDP growth offered the most insight into short-term fluctuations.

The time-varying coefficient for lagged GDP growth was consistently positive and stabilized around 1991, suggesting that economic growth plays a steady and beneficial role in narrowing the gender gap in unemployment. This finding implies that periods of GDP expansion tend to support more equal employment outcomes, potentially because women gain increased access to job opportunities during times of economic growth. To sustain this positive trend, policymakers should support inclusive job creation, invest in sectors with high female participation, and ensure women benefit equally from economic growth. Measures like vocational training, flexible work policies, and reducing hiring bias can further strengthen gender equality in employment.
