第 5 章 时间序列回归模型

# 第 5 章 时间序列回归模型

在本章节我们将会讨论时间序列的线性回归模型。线性回归模型的核心思路是：我们预测时间序列 y 时假设它与其它时间序列 x 之间存在线性关系。

例如，我们可以通过广告总花费 x 来预测月度销量 y；同样的，我们可以通过气温数据 x1 和星期数据 x2 来预测日耗电量 y。

**被预测变量** y 有时还称作回归变量、因变量或被解释变量。**预测变量** x 有时也叫作回归量、自变量或解释变量。在本书中我们称它们为“被预测变量”和“预测变量”。

library(fpp2)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

## ── Attaching packages ────────────────────────────────────────────── fpp2 2.4 ──

## ✔ ggplot2 3.3.6 ✔ fma 2.4   
## ✔ forecast 8.16 ✔ expsmooth 2.3

##

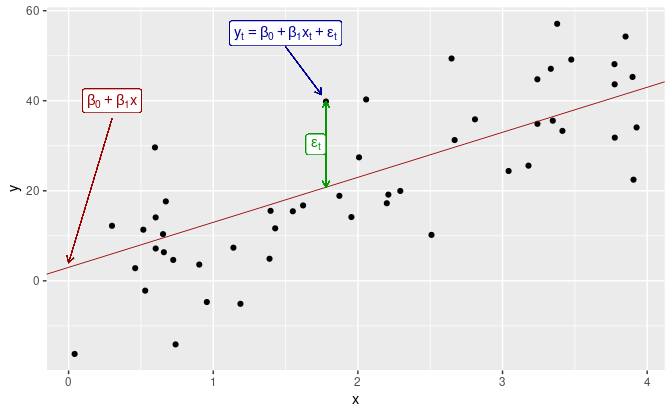
library(forecast)

5.1 线性模型

### 简单线性回归

最简单的线性回归模型假设被预测变量 y 和单个预测变量 x 之间存在如下线性关系：

图 [5.1](https://otexts.com/fppcn/regression-intro.html#fig:SLRpop1) 为示例数据在线性模型上的拟合结果。其中，系数 和 分别表示回归线的截距和斜率。截距项 β0 表示当 x=0 时 y 的预测值；斜率 β1 表示当 x 增加一个单位时，y 的平均变化。



示例数据在简单线性模型上的回归结果。

图 5.1: 示例数据在简单线性模型上的回归结果。

从上图可以看出，观测值并不全部落在回归线上，而是分布在回归线的周围。我们可以这样理解：每个观测值 yt 都包含可解释部分 和随机误差项 。随机误差项并不意味着错误，而是指观测值与线性模型的偏差。它捕捉到了除 外其他影响 的信息。

#### 示例：美国的消费支出

图[5.2](https://otexts.com/fppcn/regression-intro.html#fig:ConsInc)展示了1970年第一季度到2016年第三季度之间，美国实际个人消费支出 y 和实际个人可支配收入 x 的季度增长情况。

cbind('消费' = uschange[, "Consumption"],  
 '收入' = uschange[, "Income"]) %>%  
 autoplot(facets = TRUE, colour=TRUE) +  
 ylab("增长率 % ") + xlab("年份") +  
 theme(text = element\_text(family = "STHeiti"))

## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): Windows字体数据  
## 库里没有这样的字体系列  
  
## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): Windows字体数据  
## 库里没有这样的字体系列

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): Windows字体数据  
## 库里没有这样的字体系列

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

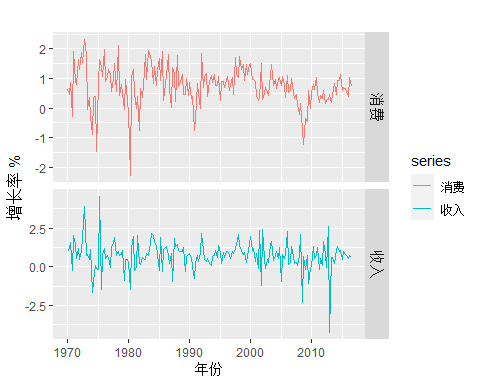


图 5.2: 美国实际个人消费支出和实际个人可支配收入的增长情况。

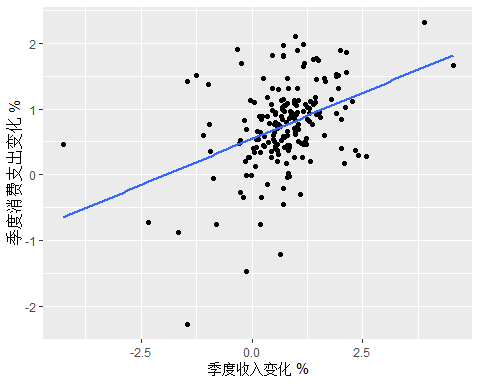
图[5.3](https://otexts.com/fppcn/regression-intro.html#fig:ConsInc2)表明了消费变化与收入变化的散点图，估计的回归线为：（我们在 y 头上加一个‘帽’，表示为 ,这表示模型对 y 的预测值。）

uschange %>%  
 as.data.frame() %>%  
 ggplot(aes(x=Income, y=Consumption)) +  
 ylab("季度消费支出变化 %") +  
 xlab("季度收入变化 %") +  
 geom\_point() +  
 geom\_smooth(method="lm", se=FALSE)+  
 theme(text = element\_text(family = "STHeiti"))+  
 theme(plot.title = element\_text(hjust = 0.5))

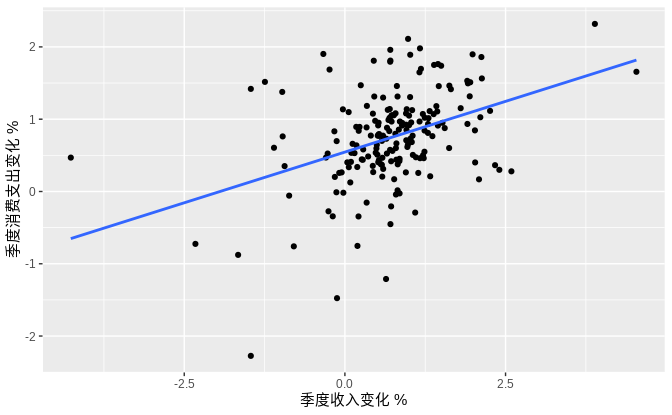
## `geom\_smooth()` using formula 'y ~ x'

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列



#> `geom\_smooth()` using formula 'y ~ x'



季度消费支出变化与个人收入变化的拟合回归线和散点图。

图 5.3: 季度消费支出变化与个人收入变化的拟合回归线和散点图。

在R中，可以使用 tslm() 函数对模型进行估计：

tslm(Consumption ~ Income, data=uschange)

##   
## Call:  
## tslm(formula = Consumption ~ Income, data = uschange)  
##   
## Coefficients:  
## (Intercept) Income   
## 0.5451 0.2806

#>  
#> Call:  
#> tslm(formula = Consumption ~ Income, data = uschange)  
#>  
#> Coefficients:  
#> (Intercept) Income  
#> 0.545 0.281

在[5.2](https://otexts.com/fppcn/least-squares.html#least-squares)中，我们会详细阐述 tslm() 计算参数的原理。

拟合出的回归线斜率为正，反映了收入与消费之间的正相关关系。斜率系数表明，x 每增加一个单位（个人可支配收入增加1个百分点），会导致 y 平均增加0.28 个单位（个人消费支出平均增加0.28 个百分点）。或者说，当 x （个人可支配收入增长百分比）为1时，y（个人消费支出增长百分比）为 0.55+0.28×1=0.83。

截距项即为 x=0 时 y 的值。在本例中，当 x=0（即个人可支配收入没有变化）时，y 的预测值为0.55（个人消费支出平均增加0.55%）。虽然 x=0 时没有意义，但截距项仍然是模型中的非常重要的一部分。假如模型中没有截距项，斜率系数很可能是一个错误的估计值。除非强制要求回归线通过原点，否则模型中应始终包含截距项。下面我们假设模型中总是包含截距项。

### 多元线性回归

当预测变量有两个甚至更多时，模型被称为**多元线性回归模型**。多元线性回归模型的一般形式如下：

其中，y 是被预测变量， 是 k 个预测变量，每个预测变量都必须为数值型变量。系数 分别衡量了在保持其他所有预测变量不变的情况下，该预测变量对被预测变量的影响程度。因此，系数衡量了对应预测变量对被预测变量的*边际影响*。

#### 示例：美国的消费支出

图[5.4](https://otexts.com/fppcn/regression-intro.html#fig:MultiPredictors)显示了可用于预测美国消费支出的其他预测变量。这些预测变量包括工业产出和个人储蓄的季度变化百分比，以及失业率的季度变化。由于我们预测消费支出时不仅考虑了个人收入，还考虑了其他预测变量，因此建立多元线性回归模型可能会产生更准确的预测结果。



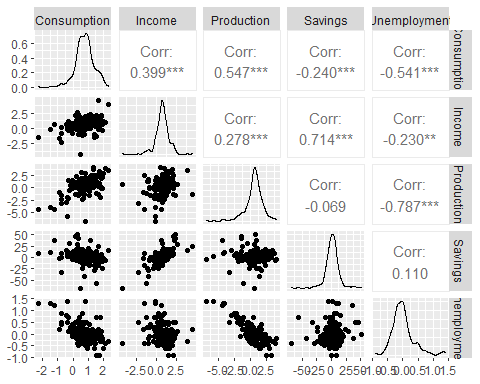
1970年第一季度到2016年第三季度之间，工业生产和个人储蓄的季度变化百分比以及失业率的季度变化情况。

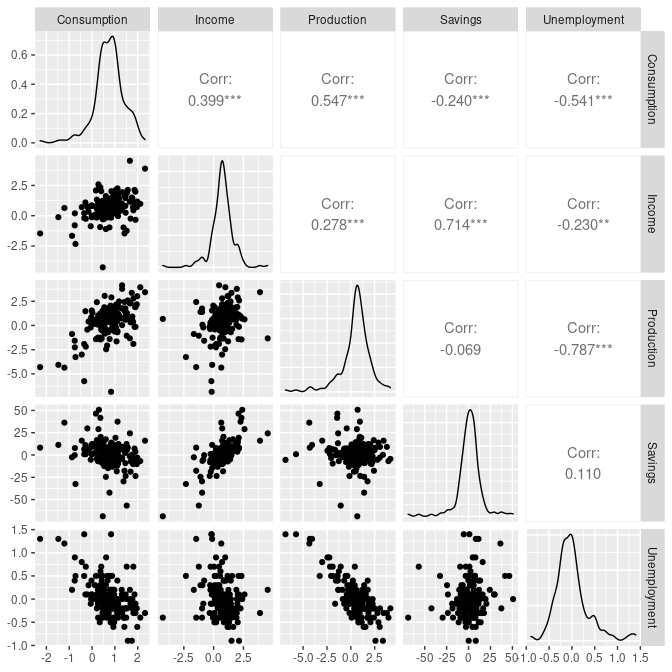
图 5.4: 1970年第一季度到2016年第三季度之间，工业生产和个人储蓄的季度变化百分比以及失业率的季度变化情况。

图[5.5](https://otexts.com/fppcn/regression-intro.html#fig:ScatterMatrix)是五个变量的散点图矩阵。其中，第一列显示了被预测变量（消费）与其他预测变量的关系。该图表明，居民收入与工业生产产值存在正相关关系，与储蓄和失业率存在负相关关系。相关关系的强度由相关系数来表示。其余的散点图和相关系数表明各个预测变量之间的关系。

uschange %>%  
 as.data.frame() %>%  
 GGally::ggpairs()

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2





美国消费支出和预测变量的散点图矩阵。

图 5.5: 美国消费支出和预测变量的散点图矩阵。

### 假设条件

当我们想要使用线性回归模型（式[(5.1)](https://otexts.com/fppcn/regression-intro.html#eq:lm)）时，需要对变量做出一些基本假设。

首先，我们假设线性模型是对现实情况的合理近似；也就是说，预测变量和被预测变量之间的关系基本满足这个线性方程。

其次，我们对误差项 (ε1,…,εT) 做出如下假设：

* 期望为零；否则预测结果会产生系统性偏差。
* 随机误差项彼此不相关；否则预测效果会很差，因为这表明数据中尚有很多可用信息没有包含在模型中。
* 与预测变量不相关；若误差项与预测变量相关，则表明模型的系统部分中应该包含更多信息。

为了方便得到预测区间，我们还需要假设随机误差项服从方差为 σ2 的正态分布。

线性回归模型还有一个重要的假设是预测变量 x 不是随机变量。在进行模拟实验时，我们可以控制每个 x 的值（所以 x 不会是随机的）并观察 y 的结果值。但在实际生活中，我们只能得到观察数据（包括商业和经济学中的大多数数据），而不能控制 x 的值。因此，我们需要做出如上假设。

## 5.2 最小二乘估计

在实际问题中，我们有一系列的观察值，但是我们不知道模型系数 β0,β1,…,βk 的具体值。因此，我们需要利用模型对这些参数进行估计。

最小二乘估计方法通过最小化残差平方和来确定模型的各个参数。也就是说，我们通过最小化下式来确定 β0,β1,…,βk 的估计值：T∑t=1ε2t=T∑t=1(yt−β0−β1x1,t−β2x2,t−⋯−βkxk,t)2.

由于它的目标是最小化残差平方和，因此被称为**最小二乘估计**。寻找最优参数的过程，一般被称为“拟合”模型，或者被称为模型的“学习”或者“训练”。图[5.3](https://otexts.com/fppcn/regression-intro.html#fig:ConsInc2)中的拟合线就是通过该方法得到的。

参数的估计值，一般用 β0,…,βk 来表示。相关方程在[5.7](https://otexts.com/fppcn/regression-matrices.html#regression-matrices)节中会详细阐述。

tslm()函数可以将时间序列数据拟合到线性回归模型中。它和广泛用于线性模型的lm()函数非常相似，但不同的是tslm()函数可用于时间序列数据。

#### 示例：美国消费支出

美国消费的多元线性回归模型为：yt=β0+β1x1,t+β2x2,t+β3x3,t+β4x4,t+εt,其中， y 是实际个人消费支出的百分比变化， x1 是实际个人可支配收入的百分比变化， x2 是工业产值的百分比变化， x3 是个人储蓄的百分比变化， x4 是失业率的变化。

以下是关于拟合模型的一些信息。Coefficients的第一列是每个 β 的估计值，第二列是估计值的标准误差（通过重复估计类似数据集上的 β 系数可以获得标准差）。标准误差是 β 的不确定性度量。

fit.consMR <- tslm(Consumption ~ Income + Production + Unemployment + Savings,  
 data=uschange)  
summary(fit.consMR)

##   
## Call:  
## tslm(formula = Consumption ~ Income + Production + Unemployment +   
## Savings, data = uschange)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.88296 -0.17638 -0.03679 0.15251 1.20553   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.26729 0.03721 7.184 1.68e-11 \*\*\*  
## Income 0.71449 0.04219 16.934 < 2e-16 \*\*\*  
## Production 0.04589 0.02588 1.773 0.0778 .   
## Unemployment -0.20477 0.10550 -1.941 0.0538 .   
## Savings -0.04527 0.00278 -16.287 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3286 on 182 degrees of freedom  
## Multiple R-squared: 0.754, Adjusted R-squared: 0.7486   
## F-statistic: 139.5 on 4 and 182 DF, p-value: < 2.2e-16

#>   
#> Call:  
#> tslm(formula = Consumption ~ Income + Production + Unemployment +   
#> Savings, data = uschange)  
#>   
#> Residuals:  
#> Min 1Q Median 3Q Max   
#> -0.8830 -0.1764 -0.0368 0.1525 1.2055   
#>   
#> Coefficients:  
#> Estimate Std. Error t value Pr(>|t|)   
#> (Intercept) 0.26729 0.03721 7.18 1.7e-11 \*\*\*  
#> Income 0.71448 0.04219 16.93 < 2e-16 \*\*\*  
#> Production 0.04589 0.02588 1.77 0.078 .   
#> Unemployment -0.20477 0.10550 -1.94 0.054 .   
#> Savings -0.04527 0.00278 -16.29 < 2e-16 \*\*\*  
#> ---  
#> Signif. codes:   
#> 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
#>   
#> Residual standard error: 0.329 on 182 degrees of freedom  
#> Multiple R-squared: 0.754, Adjusted R-squared: 0.749   
#> F-statistic: 139 on 4 and 182 DF, p-value: <2e-16

对于预测来说，我们并不是特别关心估计系数的 t 值和 p 值。 t 值是估计系数 β 与其标准误差的比值；估计结果的最后一列是 p 值，若消费支出和相关的预测变量没有显著的关系时， p 值将会很大。这在检验及研究各预测变量对被预测变量是否有显著影响时很有用，但对于预测本身并不特别有用。

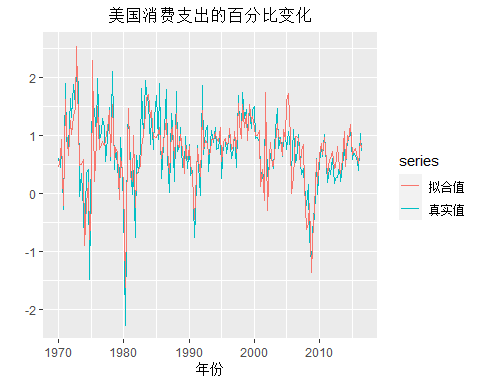
### 拟合值

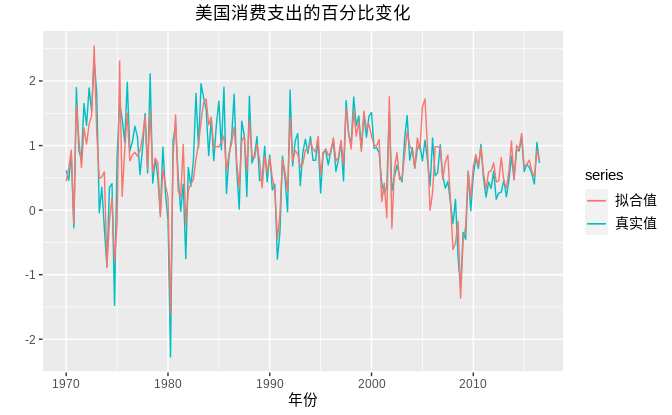
我们可以利用回归方程中的估计系数并将误差项设置为零来预测 y 。我们通常将模型写成如下形式：yt=β0+β1x1,t+β2x2,t+⋯+^βkxk,t.(5.2)将训练样本中 x1,t,…,xk,t （其中 t=1,…,T ）的值代入模型中，我们将会得到 yt 的预测值，即为模型的*拟合值*。需要注意的是，这是模型估计得到的训练样本的预测值，而不是 y 未来真实值的预测值。

下图是美国消费支出百分比变化的拟合值与真实值的比较。图[5.6](https://otexts.com/fppcn/least-squares.html#fig:usfitted1)中的时间序列图表明拟合值非常接近实际数据。这一点可以通过散点图[5.7](https://otexts.com/fppcn/least-squares.html#fig:usfitted2)中显示的明显的正相关关系来验证。

autoplot(uschange[,'Consumption'], series="真实值") +  
 autolayer(fitted(fit.consMR), series="拟合值") +  
 xlab("年份") + ylab("") +  
 ggtitle("美国消费支出的百分比变化") +  
 theme(text = element\_text(family = "STHeiti"))+  
 theme(plot.title = element\_text(hjust = 0.5))

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列





美国消费支出真实值与拟合值的时间序列图。

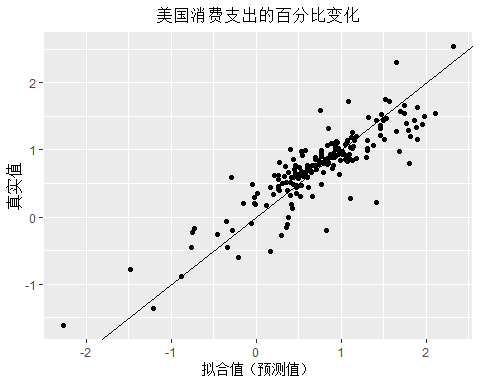
图 5.6: 美国消费支出真实值与拟合值的时间序列图。

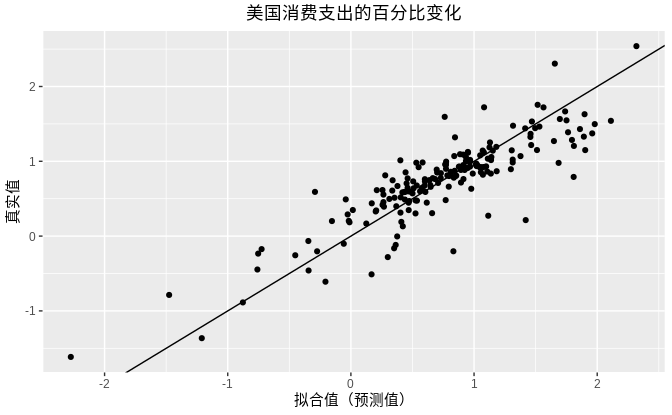
cbind(Data=uschange[,"Consumption"], Fitted=fitted(fit.consMR)) %>%  
 as.data.frame() %>%  
 ggplot(aes(x=Data, y=Fitted)) +  
 geom\_point() +  
 xlab("拟合值（预测值）") +  
 ylab("真实值") +  
 ggtitle("美国消费支出的百分比变化") +  
 geom\_abline(intercept=0, slope=1)+  
 theme(text = element\_text(family = "STHeiti"))+  
 theme(plot.title = element\_text(hjust = 0.5))

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## Windows字体数据库里没有这样的字体系列





美国消费支出真实值与拟合值的关系图。

图 5.7: 美国消费支出真实值与拟合值的关系图。

### 拟合优度

一般用可决系数（R2）评价线性回归模型对数据的拟合程度。它可以通过计算观测值 y 和预测值 ^y 之间的相关性来得出。或者，通过下式计算：R2=∑(^yt−¯y)2∑(yt−¯y)2,可决系数反映了回归模型所能解释的被预测变量的变异占被预测变量总变异的比例。

在简单线性回归模型中， R2 也等于 y 和 x 的相关系数的平方（假设存在截距项）。

预测值越接近于真实值， R2 则会越接近于1。相反，若预测值和真实值不相关，则 R2=0 （假设存在截距项）。在其它情况下， R2 的值则会处在0和1之间。

但是仅仅利用 R2 来衡量模型是远远不够的。因为当增加解释变量的个数时， R2 值将会不断增加，但这并不意味着更好的模型效果。目前并不存在衡量 R2 值好坏的规则， R2 值的有效性需要视具体情况而定。因此，利用模型在测试集上的预测结果来衡量模型好坏比直接根据 R2 大小来衡量模型更加有效。

#### 示例：美国消费支出

图[5.7](https://otexts.com/fppcn/least-squares.html#fig:usfitted2)绘制了实际消费支出与拟合值的关系曲线。变量之间的相关系数为 r=0.868 ， R2=0.754 。从 R2 值来看，该模型的拟合效果较好，模型可以解释消费支出75.4%的变异。而[5.1](https://otexts.com/fppcn/regression-intro.html#regression-intro)节中，利用简单回归模型对数据进行拟合，结果显示其 R2 值为0.16。因此，添加三个额外的预测变量可以使模型解释消费数据中更多的变化。

### 回归的标准误差

另外一个衡量模型拟合效果的指标是残差的标准偏差，通常称之为“残差标准误差”。在上例中，模型的残差标准误差为0.329。它可以通过下式来计算：

^σe= ⎷1T−k−1T∑t=1e2t,(5.3)其中， k 是模型中预测变量的个数。需要注意的是，由于需要估计的参数个数为 k+1 （截距项和 k 个解释变量），因此上式中分母为 T−k−1 。

模型的标准误差和平均误差有一定联系。我们可以将标准误与 y 的均值或标准差做对比，得到一些关于模型精度的结论。

在生成被预测变量的预测区间时，标准误差将十分有用。我们会在[5.6](https://otexts.com/fppcn/forecasting-regression.html#forecasting-regression)节中详细讨论该内容。