Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

Теперь мы рассмотрим модели с квадратичными объясняющими переменными указанного типа. Такая модель может быть установлена с использованием не модифицированного метода наименьших квадратов.

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

Однако привычная интерпретация параметра β_3 не применима, так как при изменении переменной X_2 на единицу переменная Y не изменяется на β_3. Это не может быть применено для X_2 без изменения X_2^2.

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

$$\frac{\mathrm{d}Y}{\mathrm{d}X_2} = \beta_2 + 2\beta_3 X_2$$

Дифференцируя уравнение по X_2, получаем скорость изменения Y по X_2. Таким образом, при изменении X_2 на единицу, Y изменится на величину ($\beta_2 + 2\beta_3 X_2$).

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

$$\frac{\mathrm{d}Y}{\mathrm{d}X_2} = \beta_2 + 2\beta_3 X_2$$

Это означает, что β_2 имеет интерпретацию отличную от интерпретации в обычной линейной модели , в которой при изменении X_2 , β_2 изменяется на Y.

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

$$\frac{\mathrm{d}Y}{\mathrm{d}X_2} = \beta_2 + 2\beta_3 X_2$$

Для частного случая , где X_2 = 0, β_2 интерпритируется как единичное изменения X_2 по *Y*. Для ненулевых значений X2 предельный эффект будет другим.

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

$$\frac{\mathrm{d}Y}{\mathrm{d}X_2} = \beta_2 + 2\beta_3 X_2$$

$$Y = \beta_1 + (\beta_2 + \beta_3 X_2) X_2 + u$$

β_3 имеет специальную интерпритацию. Если переписать модель,как показано на слайде, то β_3 можно интерпритировать как скорость изменения коэффициента перед X_2, на единицу изменения X_2.

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

$$\frac{dY}{dX_2} = \beta_2 + 2\beta_3 X_2$$

$$Y = \beta_1 + (\beta_2 + \beta_3 X_2) X_2 + u$$

Только β_1 имеет условно принятую интерпритацию. Обычно, когда X_2 = 0, β_1 =Y (кроме случайной составляющей) .

Квадратичные объясняющие переменные

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u$$

$$\frac{dY}{dX_2} = \beta_2 + 2\beta_3 X_2$$

$$Y = \beta_1 + (\beta_2 + \beta_3 X_2) X_2 + u$$

Возникает следующая трудность, связанная с тем, что , если $X_2 = 0$ находится вне диапозона данных, то возможное значение константы не имеет смысла. Если $X_2 = 0$ лежит вне диапозона данных, то такой же тип искажений может произойти и с оценкой, β_2.

Квадратичные объясняющие переменные

```
. gen SSQ = S*S

. reg EARNINGS S SSQ

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =   23.44
       Model |  6061.38243     2  3030.69122           Prob > F      =  0.0000
    Residual |  64267.5838   497  129.311034           R-squared     =  0.0862
-------------+------------------------------           Adj R-squared =  0.0825
       Total |  70328.9662   499  140.939812           Root MSE      =  11.372
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   .1910651   1.785822     0.11   0.915    -3.317626    3.699757
         SSQ |   .0366817   .0606266     0.61   0.545    -.0824344    .1557978
       _cons |   8.358401   12.86047     0.65   0.516    -16.90919    33.62599
----------------------------------------------------------------------------
```

Рассмотрим пример основанный на функции заработка. В таблице приведены результаты квадратичной регрессии заработка при обучении (Сумма квадратов (SSQ) определяется как квадрат обучения).

Квадратичные объясняющие переменные

```
. gen SSQ = S*S
```

```
. reg EARNINGS S SSQ

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =    500
-------------+------------------------------           F(  2,   497) =  23.44
       Model |  6061.38243     2  3030.69122           Prob > F      = 0.0000
    Residual |  64267.5838   497  129.311034           R-squared     = 0.0862
-------------+------------------------------           Adj R-squared = 0.0825
       Total |  70328.9662   499  140.939812           Root MSE      = 11.372
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   .1910651   1.785822     0.11   0.915    -3.317626    3.699757
         SSQ |   .0366817   .0606266     0.61   0.545    -.0824344    .1557978
       _cons |   8.358401   12.86047     0.65   0.516    -16.90919    33.62599
----------------------------------------------------------------------------
```

Коэффициент S означает, что для человека, не имеющего школьного образования, влияние каждого года обучения должно увеличивать почасовой заработок на 0,19 доллара США.

Квадратичные объясняющие переменные

```
. gen SSQ = S*S
```

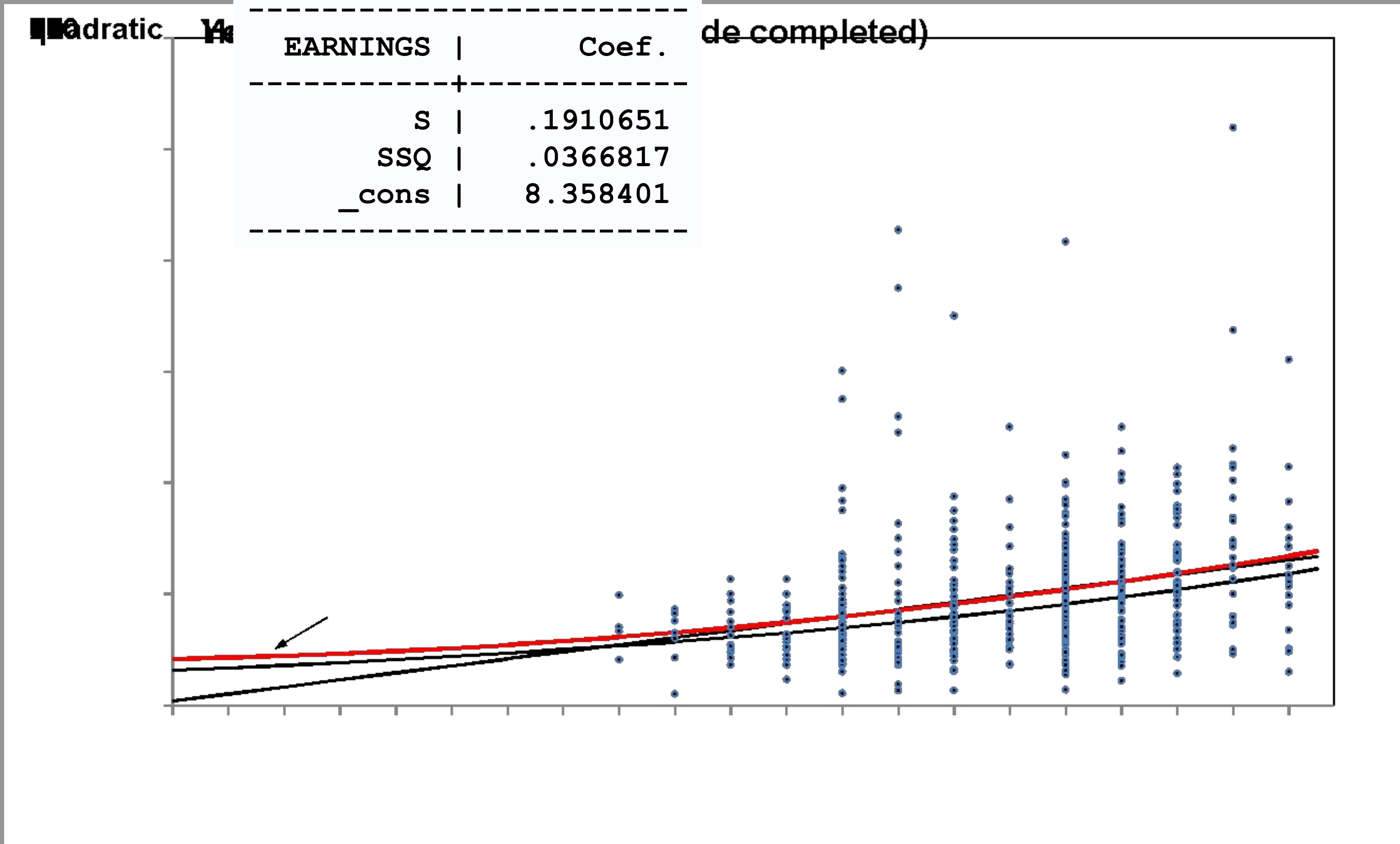
```
. reg EARNINGS S SSQ
```

```
----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,    497) =   23.44
       Model |  6061.38243     2  3030.69122           Prob > F      =  0.0000
    Residual |  64267.5838   497  129.311034           R-squared     =  0.0862
-------------+------------------------------           Adj R-squared =  0.0825
       Total |  70328.9662   499  140.939812           Root MSE      =  11.372
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   .1910651   1.785822     0.11   0.915    -3.317626    3.699757
         SSQ |   .0366817   .0606266     0.61   0.545    -.0824344    .1557978
       _cons |   8.358401   12.86047     0.65   0.516    -16.90919    33.62599
----------------------------------------------------------------------------
```

В буквальном смысле данное значение можно интерпретировать так, что человек без образования будет получать ежечасный заработок в размере 8,36 долл. США, что кажется невероятно высоким.

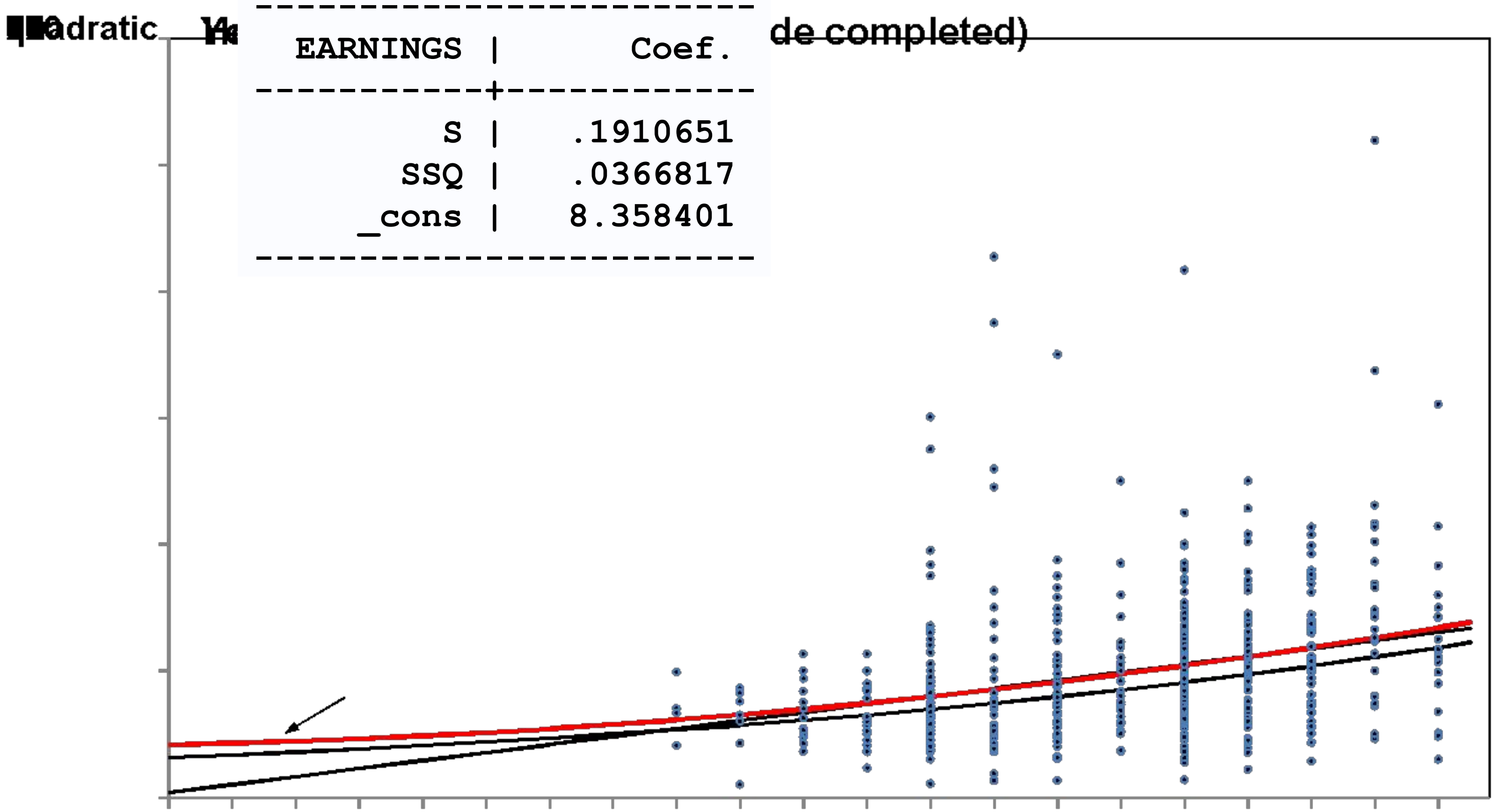
Квадратичные объясняющие переменные

EARNINGS	Coef.
S	.1910651
SSQ	.0366817
_cons	8.358401



На графике изображена квадратичная зависимость. По диапазону фактических данных график отлично подходит для наблюдений. График данной функции не сильно отличается от линейных и полулогарифмических моделей.

Квадратичные объясняющие переменные



EARNINGS	Coef.
S	.1910651
SSQ	.0366817
_cons	8.358401

Большинство исследований уравнения заработной платы принимают полулогарифмическую форму. Коэффициент наклона имеет простую интерпретацию, и модель не приводит к бессмысленным предсказаниям за пределами диапазона данных.

Квадратичные объясняющие переменные

Среднегодовые темпы роста (в процентах)

	Занятости	ВВП		Занятости	ВВП
Australia	2.57	3.52	Korea	1.11	4.48
Austria	1.64	2.66	Luxembourg	1.34	4.55
Belgium	1.06	2.27	Mexico	1.88	3.36
Canada	1.90	2.57	Netherlands	0.51	2.37
Czech Republic	0.79	5.62	New Zealand	2.67	3.41
Denmark	0.58	2.02	Norway	1.36	2.49
Estonia	2.28	8.10	Poland	2.05	5.16
Finland	0.98	3.75	Portugal	0.13	1.04
France	0.69	2.00	Slovak Republic	2.08	7.04
Germany	0.84	1.67	Slovenia	1.60	4.82
Greece	1.55	4.32	Sweden	0.83	3.47
Hungary	0.28	3.31	Switzerland	0.90	2.54
Iceland	2.49	5.62	Turkey	1.30	6.90
Israel	3.29	4.79	United Kingdom	0.92	3.31
Italy	0.89	1.29	United States	1.36	2.88
Japan	0.31	1.85			

Данные о темпах роста занятости и темпов роста ВВП , для 25 стран ОЭСР на данном слайде являются еще одним примером, на котором можно рассмотреть использование квадратичной функции.

Квадратичные объясняющие переменные

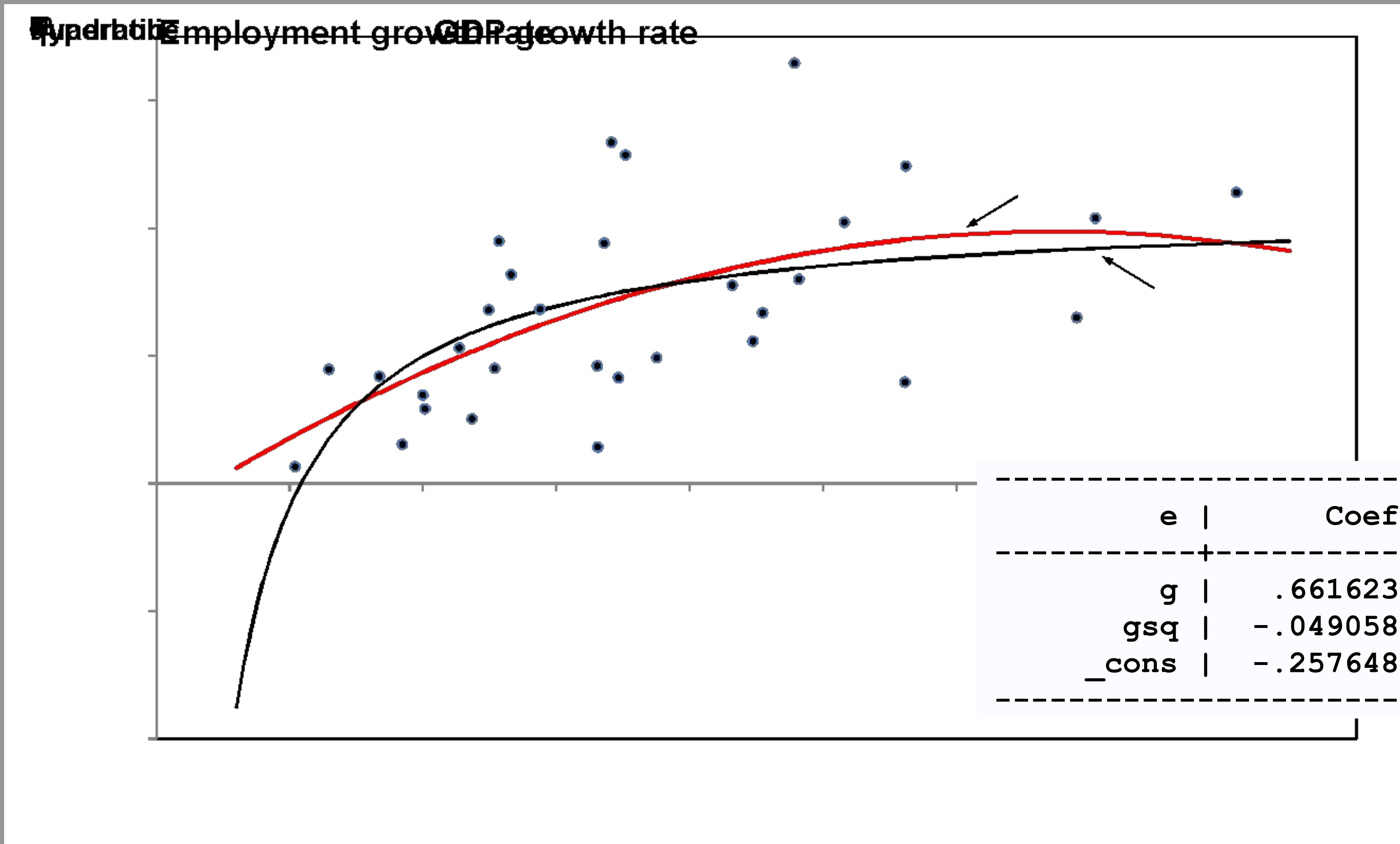
```
. gen gsq = g*g
```

```
. reg e g gsq

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =      31
-------------+------------------------------           F(  2,    28) =    7.03
       Model |  6.05131556     2  3.02565778           Prob > F      =  0.0034
    Residual |  12.0579495    28  .430641052           R-squared     =  0.3342
-------------+------------------------------           Adj R-squared =  0.2866
       Total |   18.109265    30  .603642167           Root MSE      =  .65623
------------------------------------------------------------------------------
           e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |   .6616232   .2988805     2.21   0.035     .0493942    1.273852
         gsq |  -.0490589   .0336736    -1.46   0.156    -.1180362    .0199185
       _cons |  -.2576489   .5845635    -0.44   0.663    -1.455073     .939775
------------------------------------------------------------------------------
```

G*sq* определяется как квадрат *g (темп роста ВВП*).

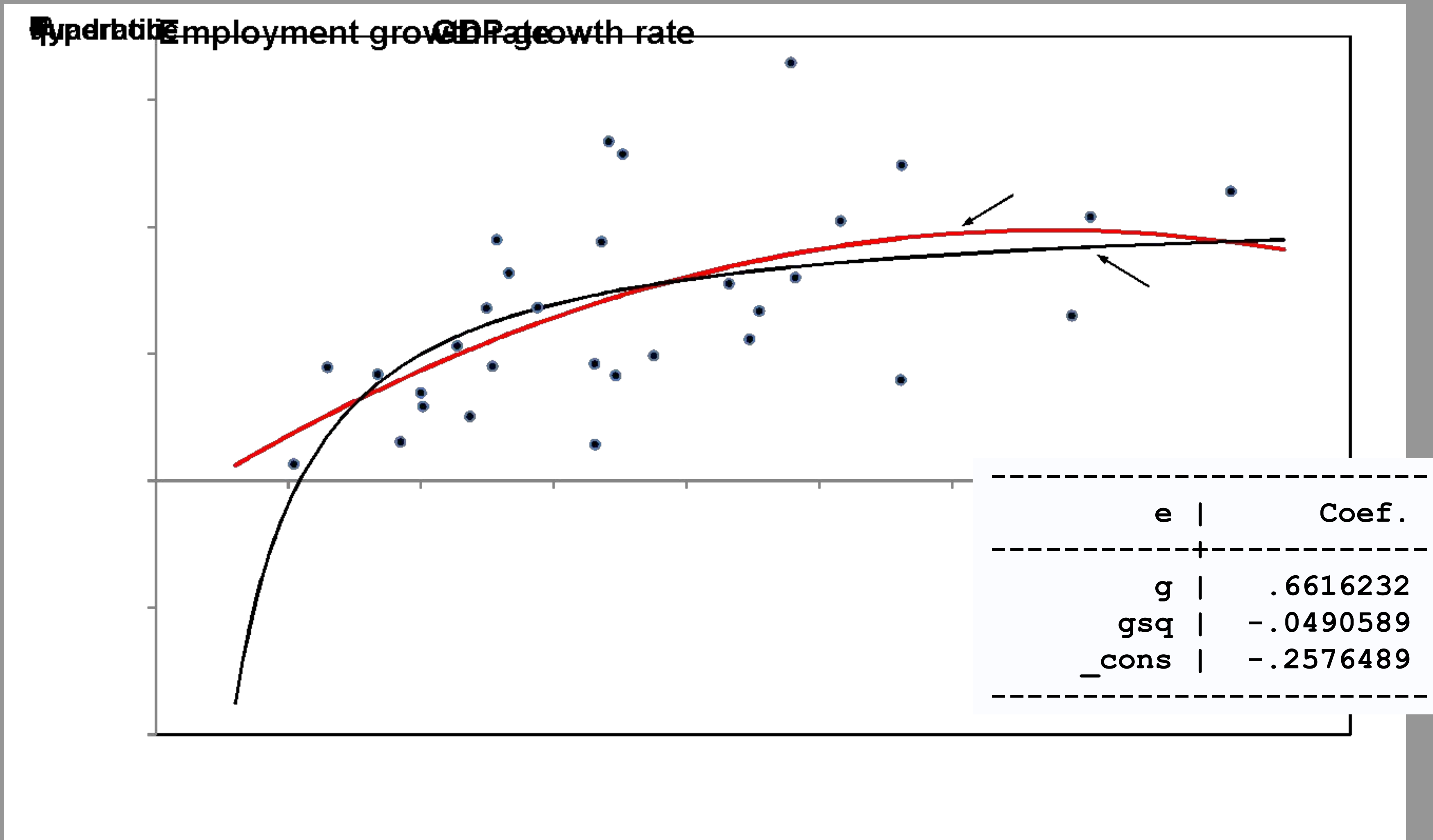
Квадратичные объясняющие переменные



e	Coef.
g	.6616232
gsq	-.0490589
_cons	-.2576489

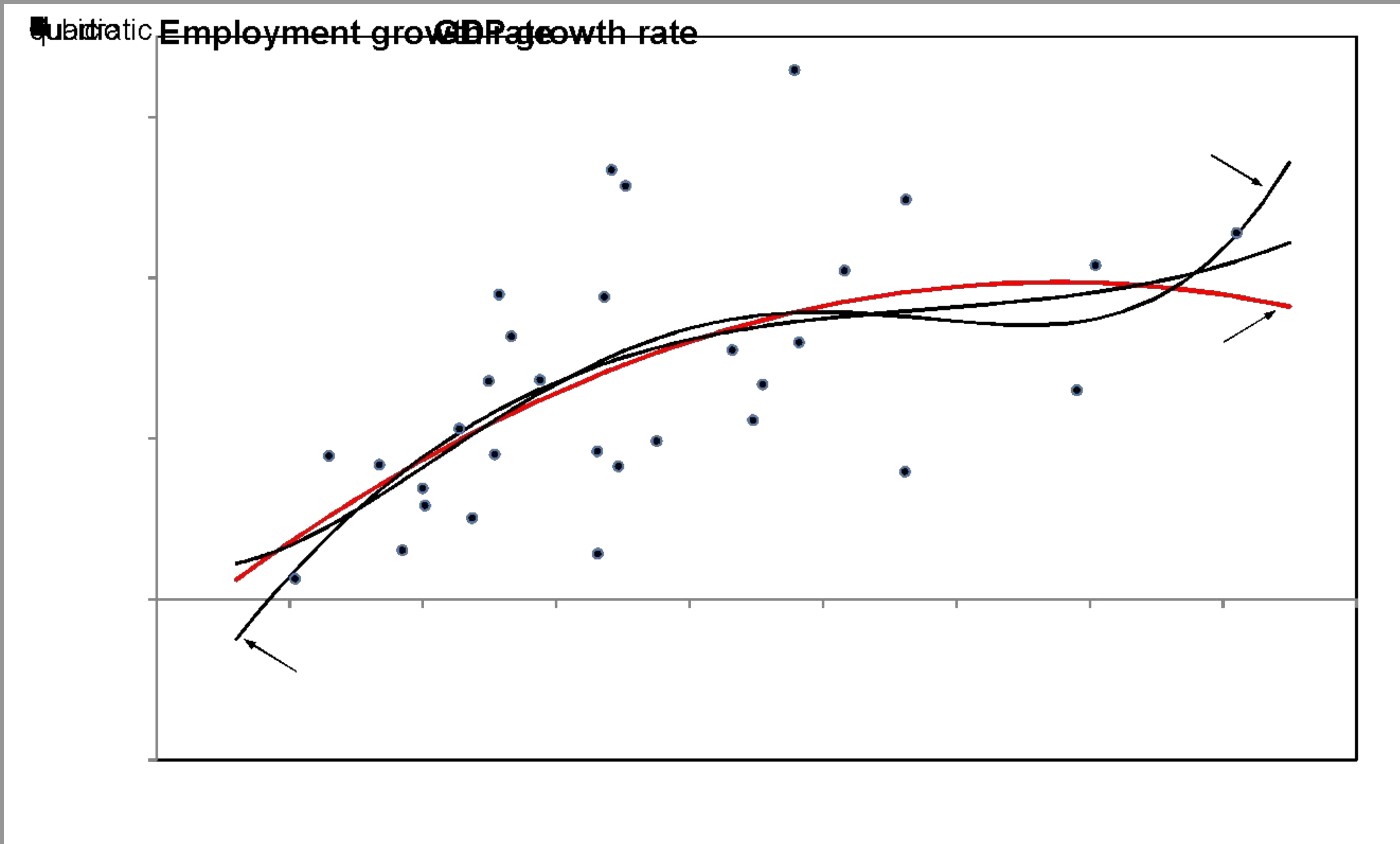
Квадратичная модель является улучшением гиперболической модели. Квадратичная модель является наиболее подходящей для низких значений темпа роста ВВП, поскольку по сравнению с другими моделями квадратичная не дает черезвычайно больших отрицательных прогнозных значений темпа роста занятости.

Квадратичные объясняющие переменные



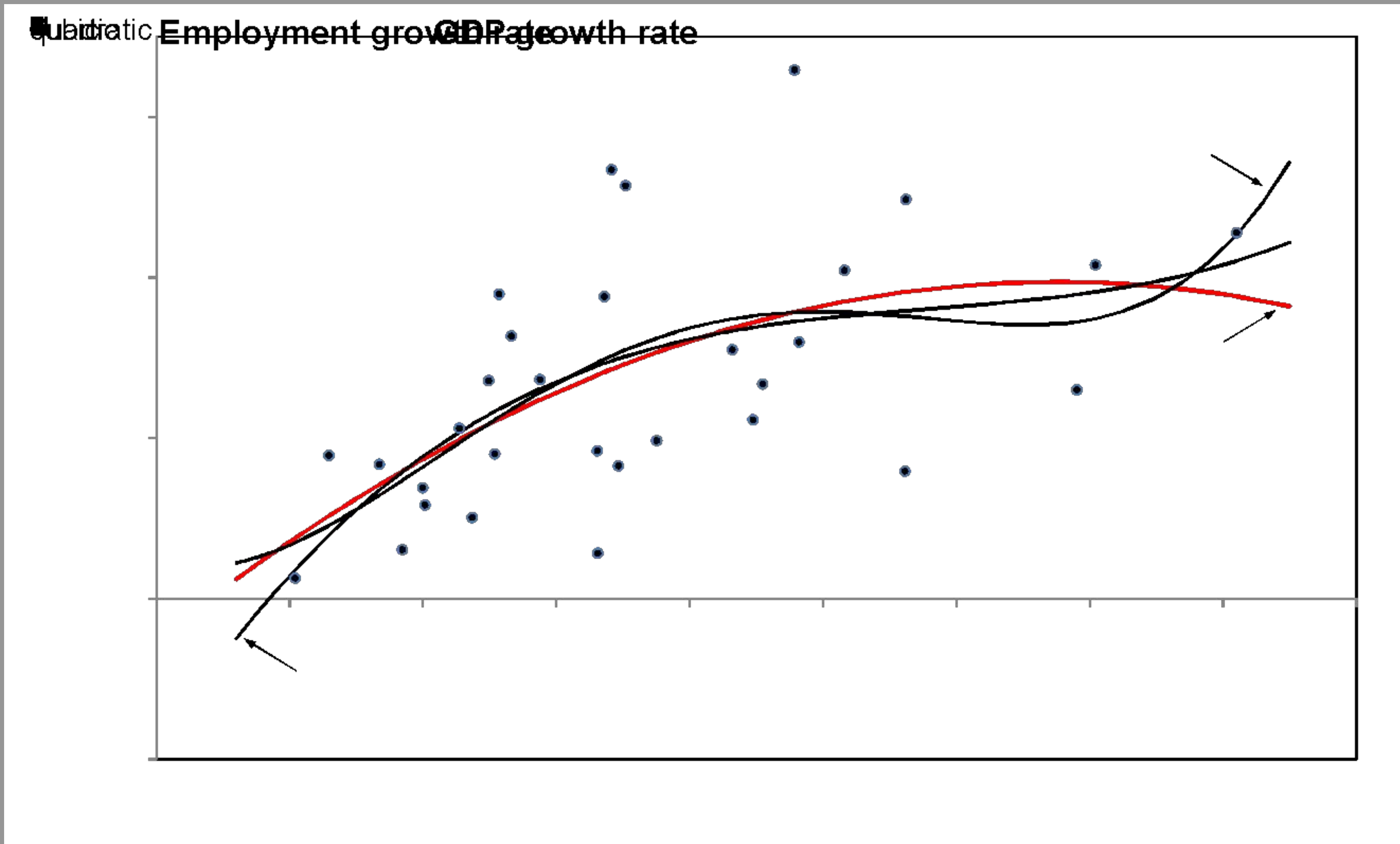
Единственным недостатком является то, что на графике рассчитанное значение e начинает падать, когда g превышает 7.

Квадратичные объясняющие переменные



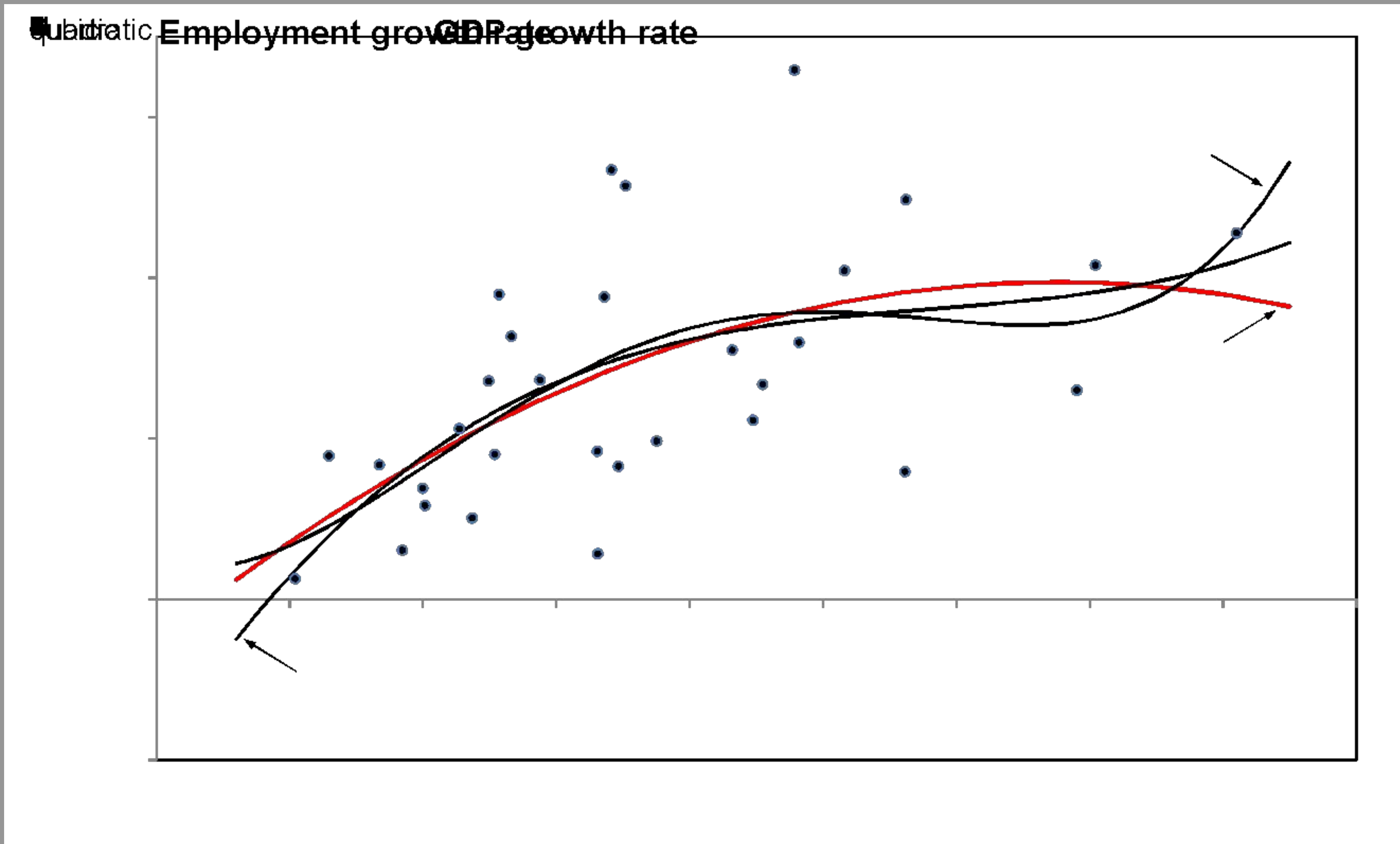
Зачем останавливаться на квадратичной модели? Почему бы не рассмотреть кубический, или квартичный, или многочлен еще более высокого порядка? Как правило, есть несколько веских причин не делать этого.

Квадратичные объясняющие переменные



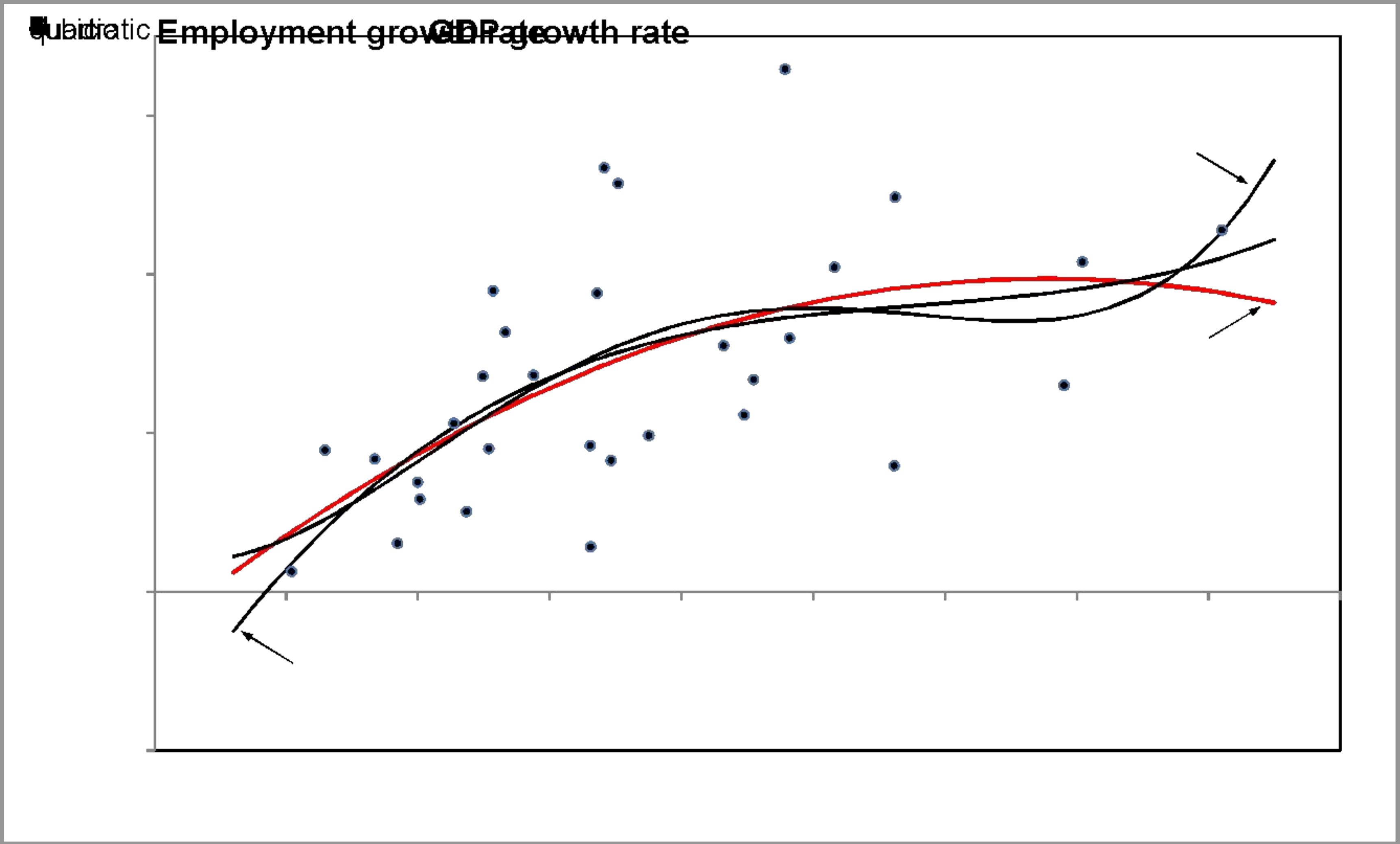
Использование квадратичных моделей, связано с уменьшением крайних эффектов, исключение которых является задачей экономической теории. Экономическая теория нечасто сталкивается с зависимостями, которые могут быть качественно объяснены кубическими или более высокими полиномами.

Квадратичные объясняющие переменные



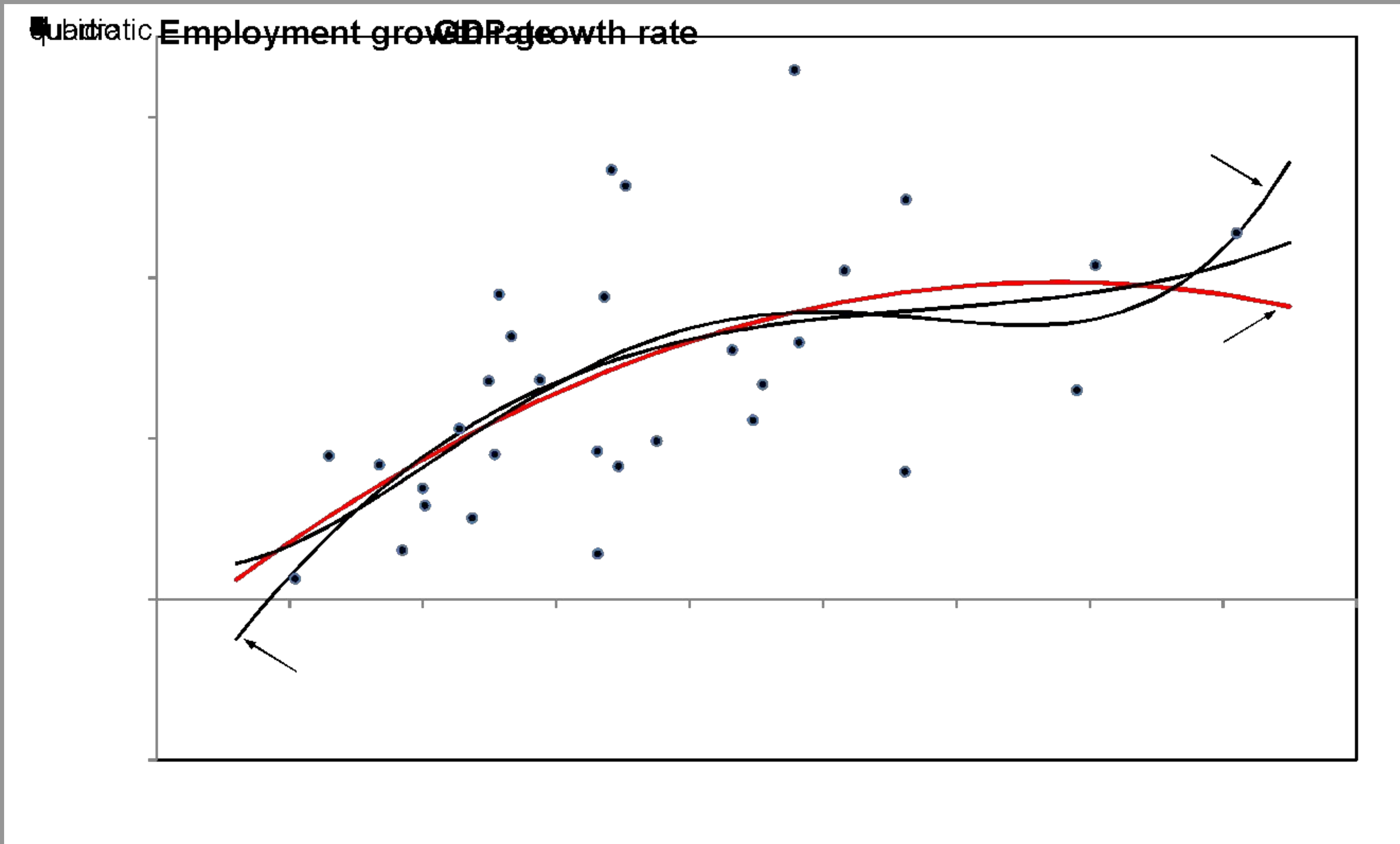
Во-вторых, не имеет смысла рассматривать полиномы более высоких порядков, поскольку это не дает дополнительной информации.

Квадратичные объясняющие переменные



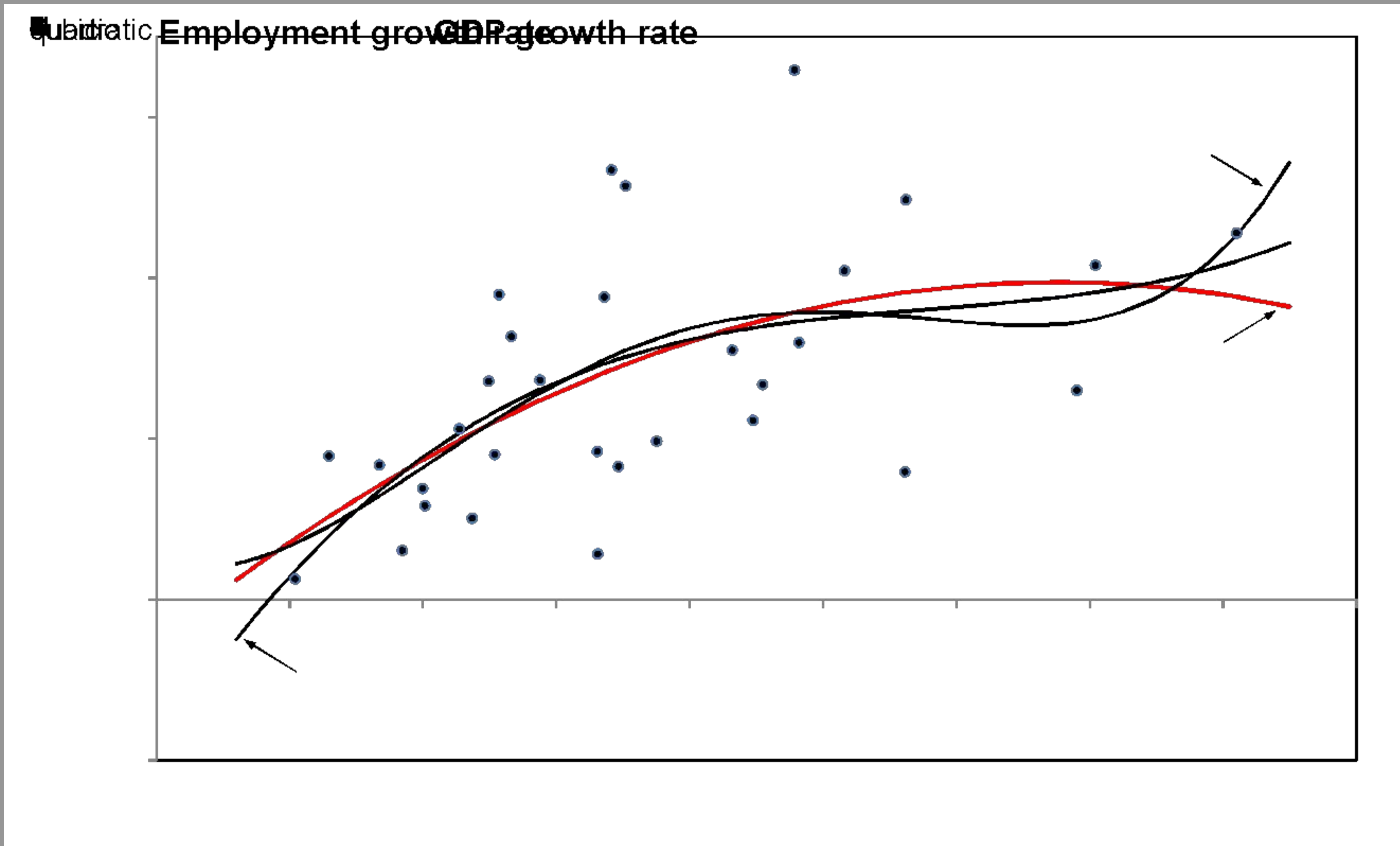
В-третьих, если выборка не очень мала, то графики полиномов более высокого порядка вряд ли будут сильно отличаться от графика квадратичной модели.

Квадратичные объясняющие переменные



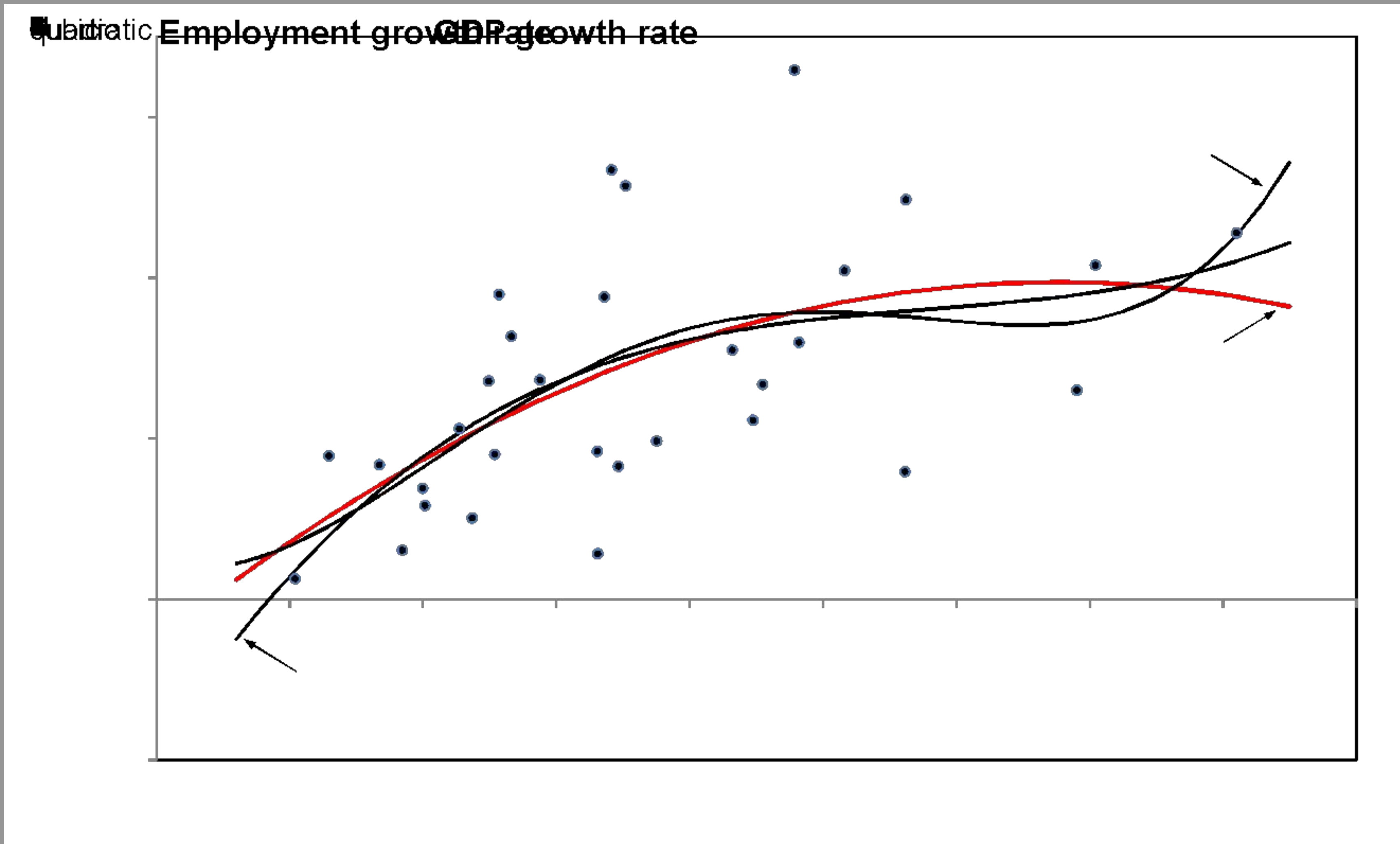
Для сравнения с квадратичной регрессией на графике показаны кубическая и регрессия четвертого порядка. В основном диапазоне данных, от g = 1,5 до g = 5, построения кубического графика и графика четвертого порядка очень похожи на квадратичный график, что подтверждает положения предыдущих слайдов.

Квадратичные объясняющие переменные



Для квадратичной формы R^2 = 0.334. Для кубической формы R^2 = 0.345, а для полинома четвертого порядка R^2 = 0.355, что демонстрирует лишь незначительные изменения.

Квадратичные объясняющие переменные



Если увеличиваются значения темпов роста ВВП, то наклон кубической кривой сначала уменьшается, а затем увеличивается. Кривая четвертого порядка фактически уменьшает значения темпов роста ВВП в диапозоне от 5 до 7, а затем наблюдается необоснованный прирост.