

# PREDICTION

True model

$$P_i = \beta_1 + \sum_{j=2}^k \beta_j X_{ji} + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \sum_{j=2}^k \hat{\beta}_j X_{ji}$$

В предыдущей последовательности мы увидели как прогнозировать цены на товары или активы, учитывая их структурные характеристики.(построение). В этом ряду мы обсудим свойства таких прогнозов.

## PREDICTION

True model

$$P_i = \beta_1 + \sum_{j=2}^k \beta_j X_{ji} + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \sum_{j=2}^k \hat{\beta}_j X_{ji}$$

Предположим, что даны примеры  $n$  наблюдений, как показано, мы установили модель ценообразования с  $k-1$  характеристиками.

# PREDICTION

True model

$$P_i = \beta_1 + \sum_{j=2}^k \beta_j X_{ji} + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \sum_{j=2}^k \hat{\beta}_j X_{ji}$$

Prediction conditional on  
 $\{X_2^*, X_3^*, \dots, X_k^*\}$

$$\hat{P}^* = \hat{\beta}_1 + \sum_{j=2}^k \hat{\beta}_j X_j^*$$

Предположим сейчас, что оно сталкивается с новым видом товара с характеристиками  $\{X_2^*, X_3^*, \dots, X_k^*\}$ . Учитывая (пример) результат выборки регрессии, естественно прогнозировать, что цена нового вида должна сводиться к третьему уравнению.

# PREDICTION

True model

$$P_i = \beta_1 + \sum_{j=2}^k \beta_j X_{ji} + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \sum_{j=2}^k \hat{\beta}_j X_{ji}$$

Prediction conditional on  
 $\{X_2^*, X_3^*, \dots, X_k^*\}$

$$\hat{P}^* = \hat{\beta}_1 + \sum_{j=2}^k \hat{\beta}_j X_j^*$$

Что же можно сказать о свойствах этого прогнозирования? Во-первых, естественно спросить справедливо ли оно в НЕсистематической переоценке или недооценке смысла фактической цены. Во-вторых, мы обеспокоены о вероятной точности прогноза

## PREDICTION

True model

$$P_i = \beta_1 + \beta_2 X_i + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Prediction conditional on  $X^*$

$$\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$$

Мы будем рассматривать пример, где товар имеет одну соответствующую характеристику и предположим, что мы установили простую регрессионную показанную модель. Следовательно, дадим новый вид товара с характеристиками  $X = X^*$ , эта модель дает нам прогнозируемую цену.

## PREDICTION

True model

$$P_i = \beta_1 + \beta_2 X_i + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Prediction conditional on  $X^*$

$$\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$$

Actual value of  $P^*$

$$P^* = \beta_1 + \beta_2 X^* + u^*$$

Допустим, что модель относится (обращается) к новому товару и, следовательно, к фактической цене, условно  $X = X^*$ ; оно образуется как показано, где  $u^*$  - это значение нарушения условия нового товара.

## PREDICTION

True model

$$P_i = \beta_1 + \beta_2 X_i + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Prediction conditional on  $X^*$

$$\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$$

Actual value of  $P^*$

$$P^* = \beta_1 + \beta_2 X^* + u^*$$

$$PE = P^* - \hat{P}^*$$

Мы объясним ошибку прогнозирования модели PE, как разницу между ценой фактической и прогнозируемой

## PREDICTION

True model

$$P_i = \beta_1 + \beta_2 X_i + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Prediction conditional on  $X^*$

$$\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$$

Actual value of  $P^*$

$$P^* = \beta_1 + \beta_2 X^* + u^*$$

$$PE = P^* - \hat{P}^* = (\beta_1 + \beta_2 X^* + u^*) - (\hat{\beta}_1 + \hat{\beta}_2 X^*)$$

Прогнозируемая ошибка получается путем подстановки фактической и прогнозируемой цены.



## PREDICTION

True model

$$P_i = \beta_1 + \beta_2 X_i + u_i$$

Fitted model

$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Prediction conditional on  $X^*$

$$\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$$

Actual value of  $P^*$

$$P^* = \beta_1 + \beta_2 X^* + u^*$$

$$PE = P^* - \hat{P}^* = (\beta_1 + \beta_2 X^* + u^*) - (\hat{\beta}_1 + \hat{\beta}_2 X^*)$$

$$E(PE) = E(\beta_1 + \beta_2 X^* + u^*) - E(\hat{\beta}_1 + \hat{\beta}_2 X^*)$$

Мы ожидаем.

## PREDICTION

True model  $P_i = \beta_1 + \beta_2 X_i + u_i$

Fitted model  $\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Prediction conditional on  $X^*$   $\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$

Actual value of  $P^*$   $P^* = \beta_1 + \beta_2 X^* + u^*$

$$PE = P^* - \hat{P}^* = (\beta_1 + \beta_2 X^* + u^*) - (\hat{\beta}_1 + \hat{\beta}_2 X^*)$$

$$\begin{aligned} E(PE) &= E(\beta_1 + \beta_2 X^* + u^*) - E(\hat{\beta}_1 + \hat{\beta}_2 X^*) \\ &= \beta_1 + \beta_2 X^* + E(u^*) - E(\hat{\beta}_1) - X^* E(\hat{\beta}_2) \end{aligned}$$

$\beta_1$  и  $\beta_2$  приняты с фиксированными параметрами, поэтому ожидания на них не влияют. Также  $X^*$  принято считать фиксированной величиной и также ожидания не воздействуют. Тем не менее,  $u^*$ ,  $\hat{\beta}_1$  и  $\hat{\beta}_2$  случайные переменные.

## PREDICTION

True model  $P_i = \beta_1 + \beta_2 X_i + u_i$

Fitted model  $\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Prediction conditional on  $X^*$   $\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$

Actual value of  $P^*$   $P^* = \beta_1 + \beta_2 X^* + u^*$

$$PE = P^* - \hat{P}^* = (\beta_1 + \beta_2 X^* + u^*) - (\hat{\beta}_1 + \hat{\beta}_2 X^*)$$

$$\begin{aligned} E(PE) &= E(\beta_1 + \beta_2 X^* + u^*) - E(\hat{\beta}_1 + \hat{\beta}_2 X^*) \\ &= \beta_1 + \beta_2 X^* + E(u^*) - E(\hat{\beta}_1) - X^* E(\hat{\beta}_2) \\ &= \beta_1 + \beta_2 X^* - \beta_1 - X^* \beta_2 \end{aligned}$$

$E(u^*) = 0$  потому что  $u^*$  выявлено случайно из распределения для  $u$ , который мы приняли как нулевую совокупность. В соответствии с обычными предположениями МНК  $\hat{\beta}_1$  есть несмещенная оценка  $\beta_1$  и  $\hat{\beta}_2$  является несмещенной оценкой  $\beta_2$

## PREDICTION

True model  $P_i = \beta_1 + \beta_2 X_i + u_i$

Fitted model  $\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Prediction conditional on  $X^*$   $\hat{P}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$

Actual value of  $P^*$   $P^* = \beta_1 + \beta_2 X^* + u^*$

$$PE = P^* - \hat{P}^* = (\beta_1 + \beta_2 X^* + u^*) - (\hat{\beta}_1 + \hat{\beta}_2 X^*)$$

$$\begin{aligned} E(PE) &= E(\beta_1 + \beta_2 X^* + u^*) - E(\hat{\beta}_1 + \hat{\beta}_2 X^*) \\ &= \beta_1 + \beta_2 X^* + E(u^*) - E(\hat{\beta}_1) - X^* E(\hat{\beta}_2) \\ &= \beta_1 + \beta_2 X^* - \beta_1 - X^* \beta_2 = 0 \end{aligned}$$

Следовательно, ожидание ошибки прогнозирования равен нулю. Результат легко обобщается в случае, где множественные характеристики и новые товары включают в себя новую комбинацию

## Variance of prediction error

$$\sigma_{PE}^2 = \left\{ 1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\} \sigma_u^2$$

Показана дисперсия совокупности ошибки прогнозирования. Неудивительно, что это подразумевает, далее - значение  $X^*$  от среднего значения выборки, что дисперсия совокупности ошибки прогнозирования будет больше.

## Variance of prediction error

$$\sigma_{PE}^2 = \left\{ 1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\} \sigma_u^2$$

Также подразумевается, что снова неудивительно, что тем больше образец, тем меньше будет дисперсия совокупности ошибки прогнозирования с нижним лимитом  $\sigma_u^2$

## Variance of prediction error

$$\sigma_{PE}^2 = \left\{ 1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\} \sigma_u^2$$

При условии, что допущения модели регрессии действительны,  $\hat{\beta}_1$  и  $\hat{\beta}_2$  будут стремиться к их реальным значениям, так как образец становится больше, поэтому единственным источником ошибки в прогнозировании будет  $u^*$ , и по определению оно имеет дисперсию совокупности  $\sigma_u^2$ .

## Variance of prediction error

$$\sigma_{PE}^2 = \left\{ 1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\} \sigma_u^2$$

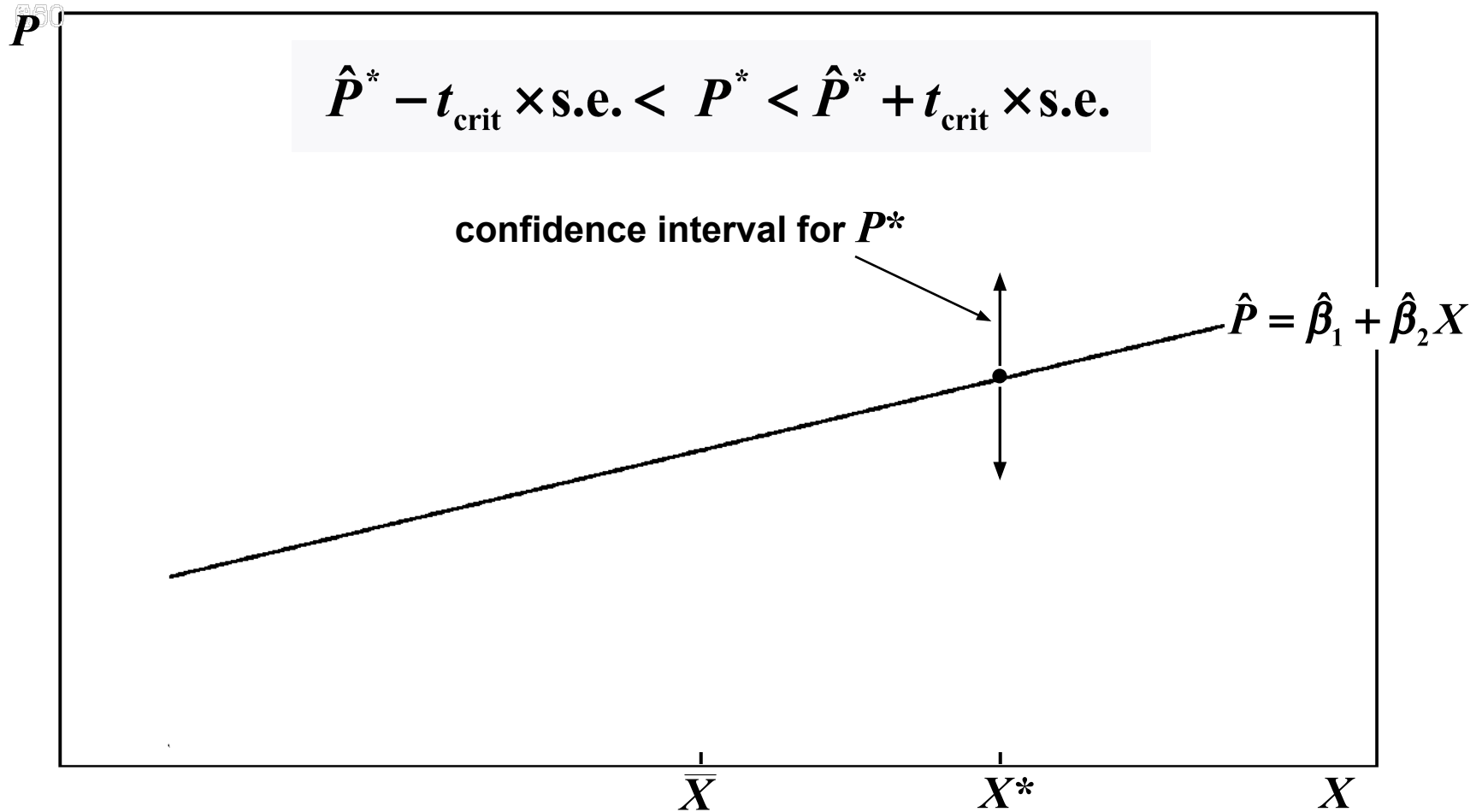
## Standard error

$$\text{s.e.}(PE) = \sqrt{\left\{ 1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\} \hat{\sigma}_u^2}$$

Стандартная ошибка прогнозирования вычисляется используя квадратный корень выражения для дисперсии совокупности, заменяя дисперсию  $u$  с оценкой, полученной при подгонке модели в период выборки.



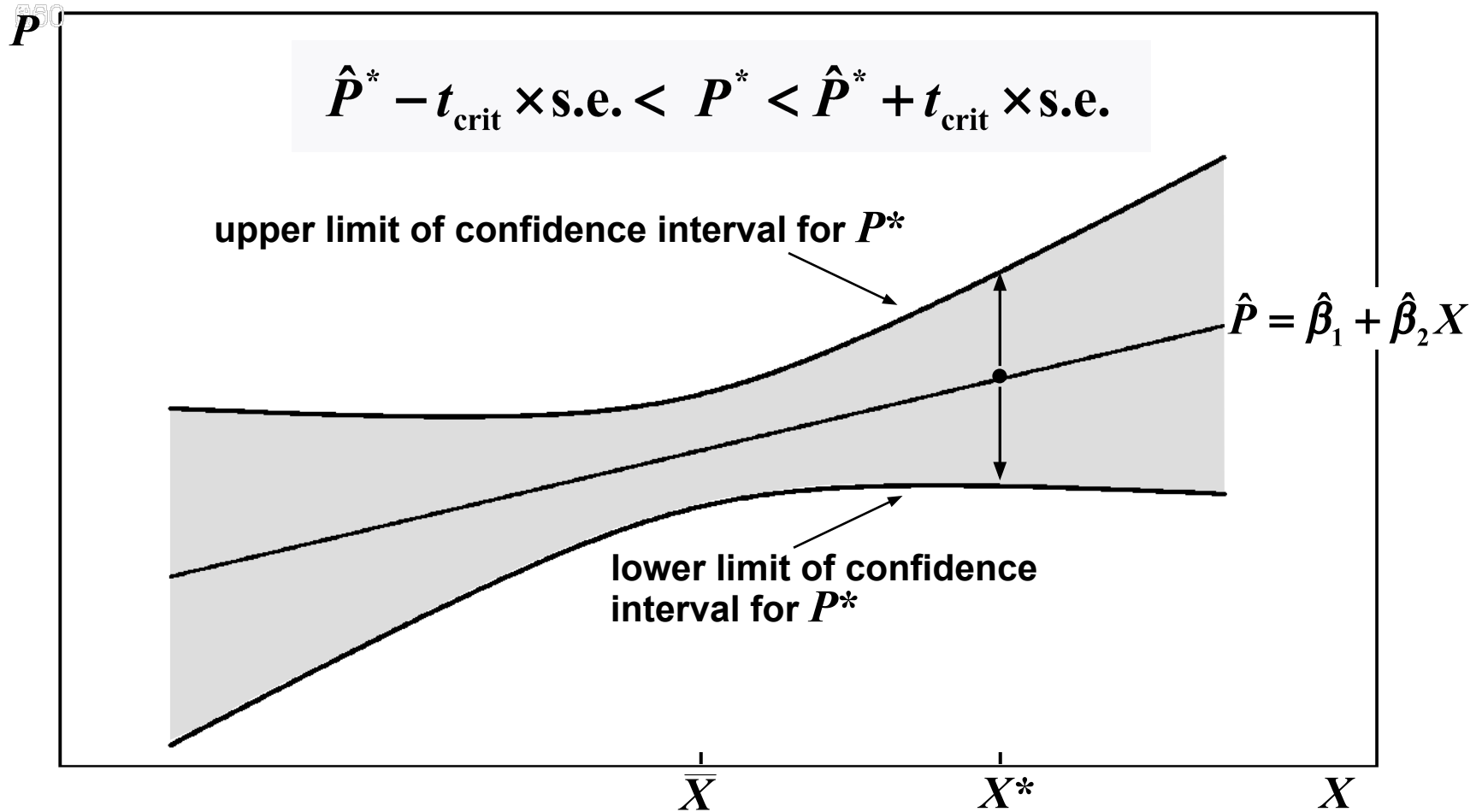
# PREDICTION



Следовательно, мы можем построить доверительный интервал для ожидания.

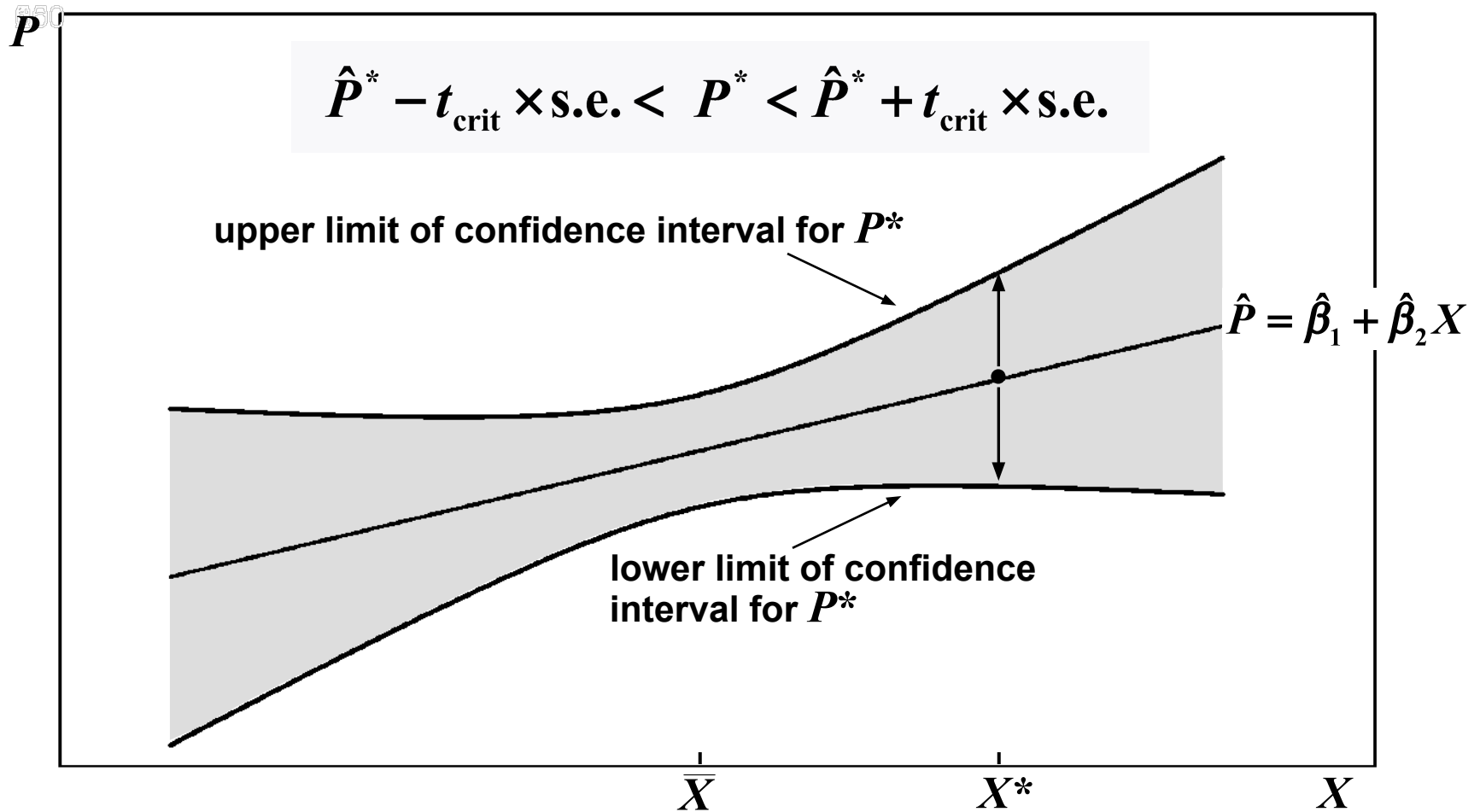
$t_{\text{crit}}$  критический уровень  $t$ , учитывая выбранный уровень значимости и количество степеней свободы, и  $\text{s.e.}$  является стандартной ошибкой прогнозирования.

# PREDICTION



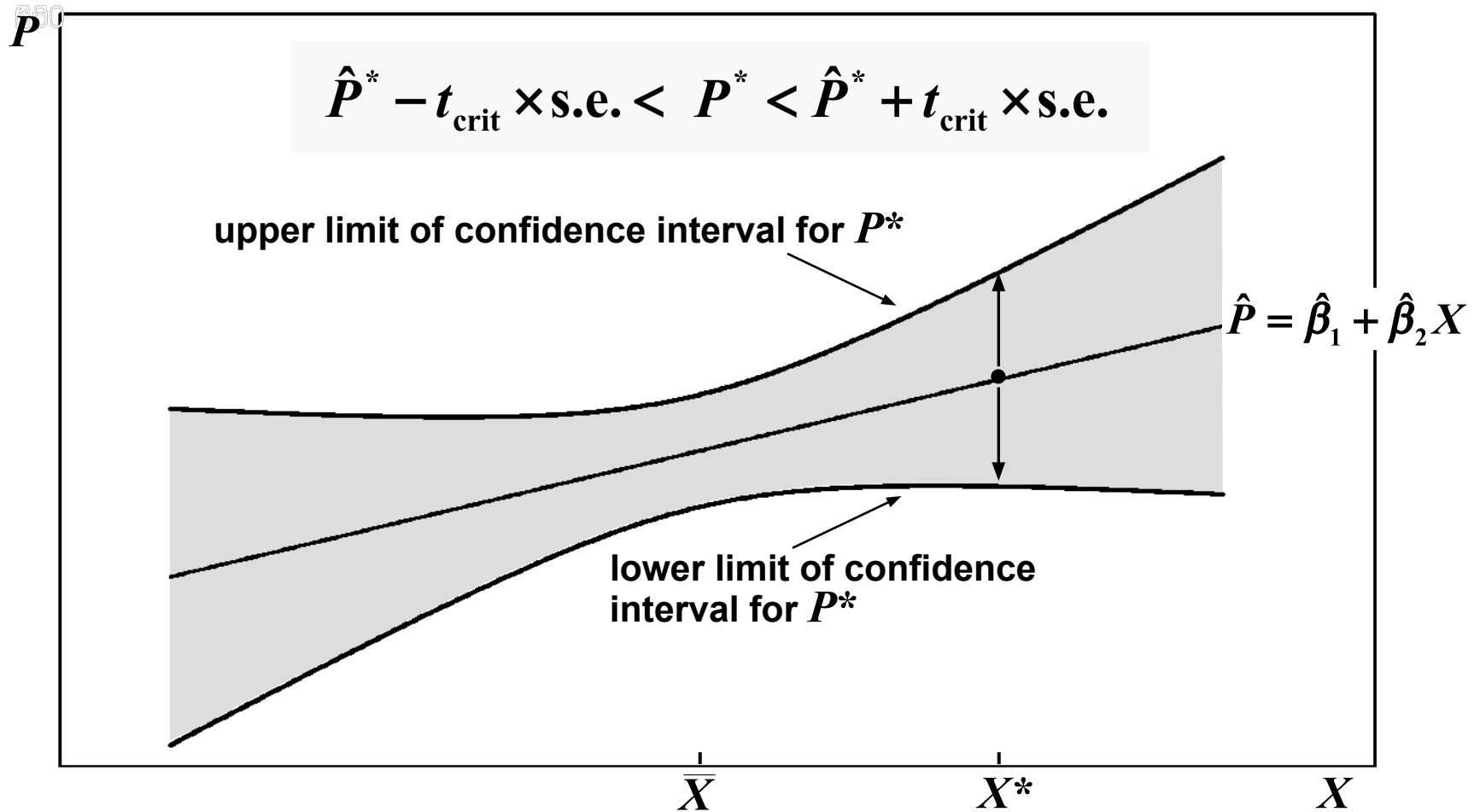
Доверительный интервал получен как функция. Как мы отметили из математического выражения, он становится шире, чем больше расстояние от  $X^*$  до среднего образца.

# PREDICTION



С несколькими объясняющими переменными выражение для ожидания дисперсии становится сложным

# PREDICTION



Следует отметить, что мультиколлинеарность не может оказывать отрицательного влияния на точность прогнозирования, даже если оценки коэффициентов имеют большие отклонения.

## PREDICTION

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Suppose  $X_2$  and  $X_3$  are positively correlated,  $\beta_2 > 0$ ,  $\beta_3 > 0$ .

Then  $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) < 0$ .

If  $\hat{\beta}_2$  is overestimated,  $\hat{\beta}_3$  is likely to be underestimated.

So  $(\hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^*)$  may be a good estimator of  $(\beta_2 X_2^* + \beta_3 X_3^*)$ .

Similarly, for other combinations.

Для простоты предположим, что есть две объясняющие переменные, что оба имеют положительные истинные коэффициенты и как показано, они положительно коррелированы; мы прогнозируем значение  $Y^*$ , учитывая значения  $X_2^*$  и  $X_3^*$ .

## PREDICTION

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Suppose  $X_2$  and  $X_3$  are positively correlated,  $\beta_2 > 0$ ,  $\beta_3 > 0$ .

Then  $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) < 0$ .

If  $\hat{\beta}_2$  is overestimated,  $\hat{\beta}_3$  is likely to be underestimated.

So  $(\hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^*)$  may be a good estimator of  $(\beta_2 X_2^* + \beta_3 X_3^*)$ .

Similarly, for other combinations.

Тогда, если эффект от  $X_2$  переоценен, поэтому  $\hat{\beta}_2 > \beta_2$ , эффект от  $X_3$  вероятно будет недооценен с  $\hat{\beta}_3 < \beta_3$ . Как следствие, эффекты ошибки могут в какой-то мере отмениться, в результате линейная комбинация может быть приближена к  $(\beta_2 X_2^* + \beta_3 X_3^*)$

## Simulation

$$Y = 10 + 2X_2 + 3X_3 + u$$

$$X_2 = \{1, 2, 3, 4, \dots, 17, 18, 19, 20\}$$

$$X_3 = \{2, 2, 4, 4, \dots, 18, 18, 20, 20\}$$

$$r_{X_2, X_3} = 0.9962$$

$$u \sim N(0,1)$$

$$\begin{aligned} Y^* &= \hat{\beta}_1 + \hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^* \\ &\cong \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3) X_2^* \end{aligned}$$

Это будет проиллюстрировано с моделированием, с моделью и показанными данными. Мы устанавливаем модель и делаем прогноз  $Y^* = \hat{\beta}_1 + \hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^*$

## Simulation

$$Y = 10 + 2X_2 + 3X_3 + u$$

$$X_2 = \{1, 2, 3, 4, \dots, 17, 18, 19, 20\}$$

$$X_3 = \{2, 2, 4, 4, \dots, 18, 18, 20, 20\}$$

$$r_{X_2, X_3} = 0.9962$$

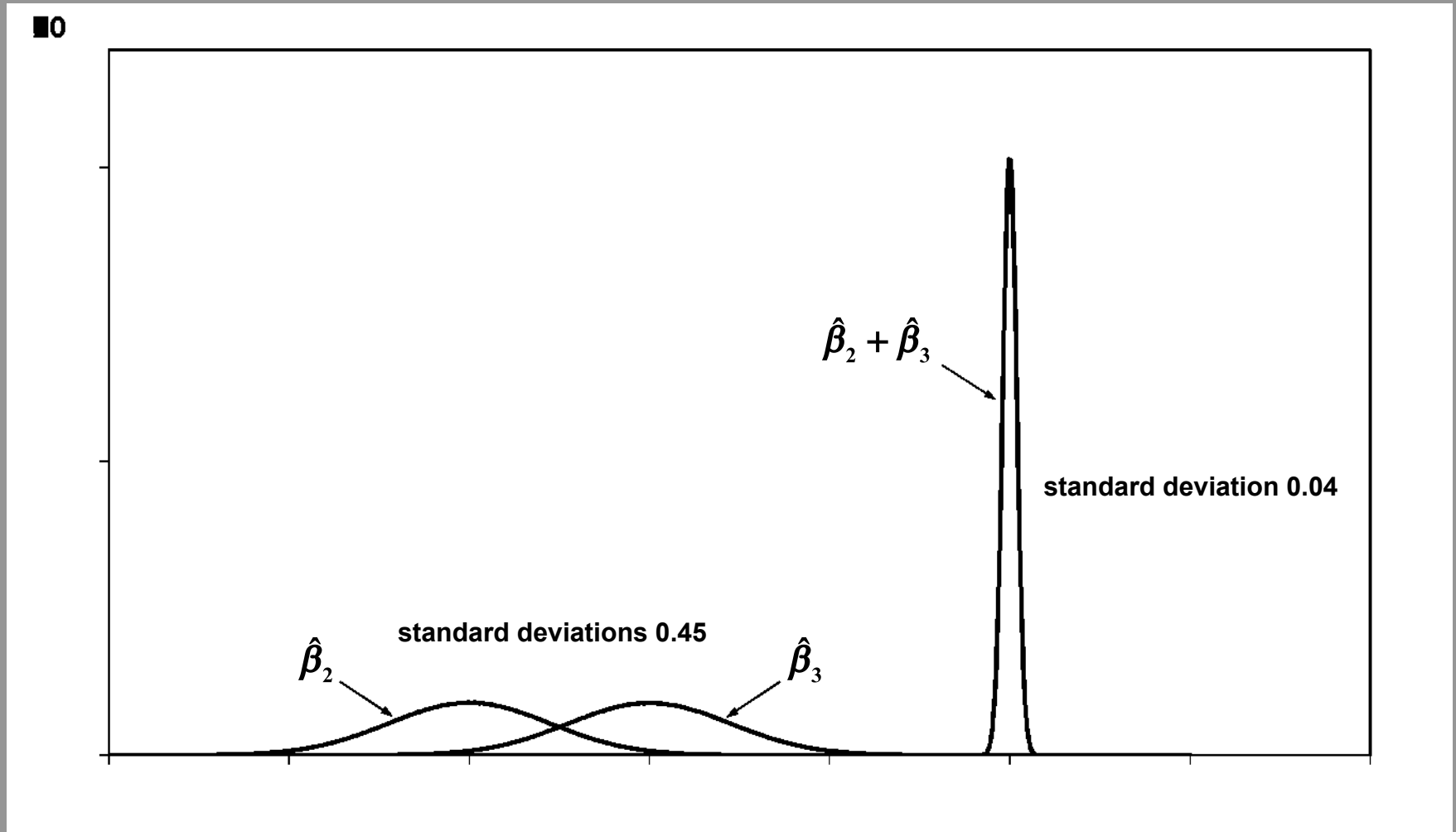
$$u \sim N(0,1)$$

$$\begin{aligned} Y^* &= \hat{\beta}_1 + \hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^* \\ &\cong \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3) X_2^* \end{aligned}$$

Поскольку  $X_2$  и  $X_3$  практически идентичны, они могут приблизиться к  $Y^* = \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3) X_2^*$ . Таким образом, точность прогноза зависит как близко  $(\hat{\beta}_2 + \hat{\beta}_3)$  к  $(\beta_2 + \beta_3)$ , то есть 5

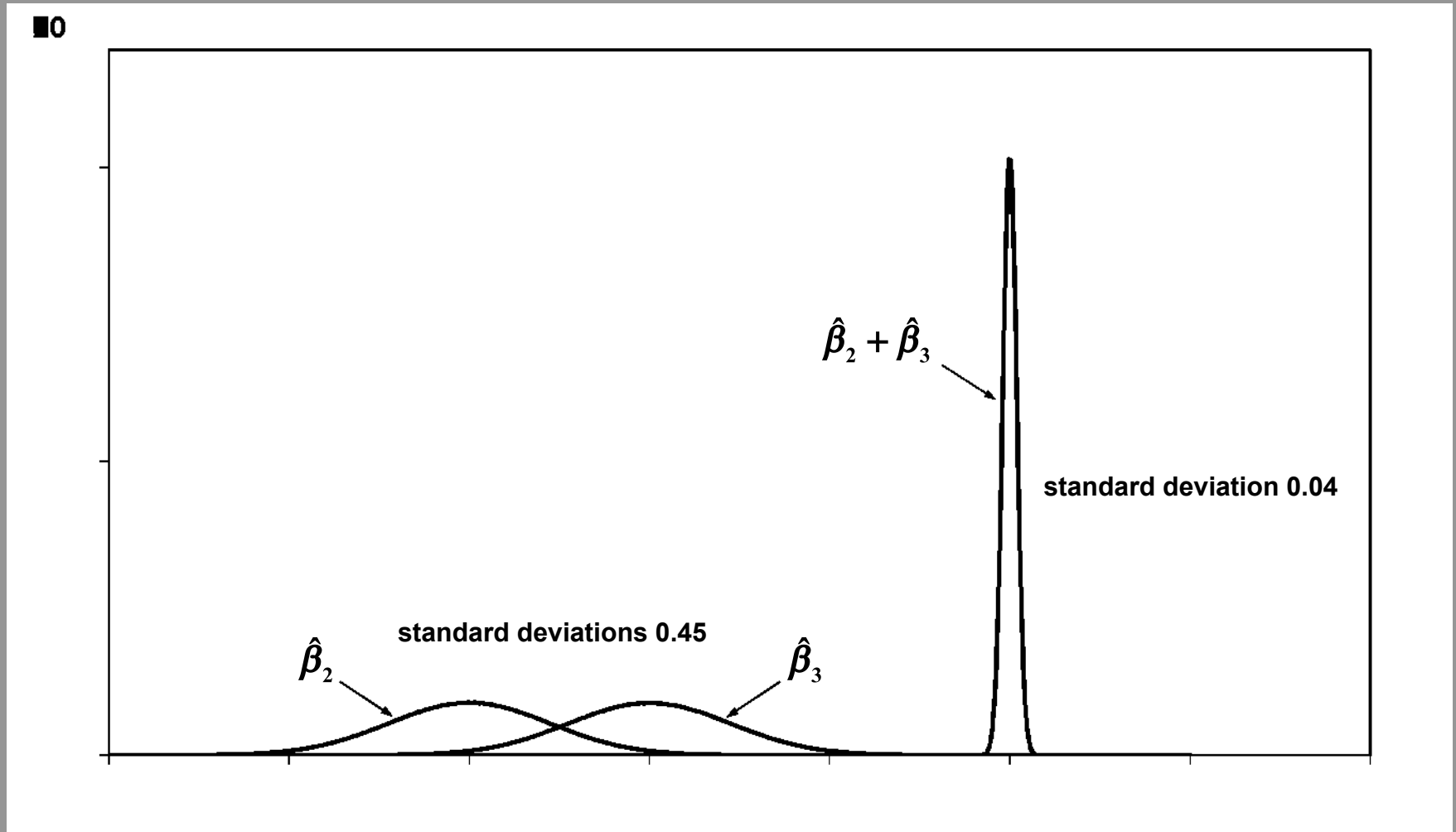


# PREDICTION



Фигура показывает распределение  $\hat{\beta}_2$  и  $\hat{\beta}_3$  для 10 миллионов образцов. Их распределения имеют относительно широкие отклонения вокруг их истинных моделей, как и следовало ожидать, учитывая мультиколлинеарность. Фактические стандартные отклонения их распределений составляют 0,45.

# PREDICTION



Фигура также показывает сумму их распределения. Как и ожидалось, он распределяется примерно на 5, но с гораздо более низким стандартным отклонением 0,04, несмотря на мультиколлинеарность, влияющую на точечные оценки отдельных коэффициентов