Модель множественной регрессии

теперь t – порядковый номер наблюдения.

стохастич.

$$X_t = \beta_0 + \beta_1 Z_{t_1} + \beta_2 Z_{t_2} + \ldots + \mathcal{E}_t$$

переменные регрессоры (объясняющие)

это в лучшем случае сл. величина (надеемся, что они независимы)

или же это сл. пр. / сл. поле

$\mathcal{E}_t : M[\mathcal{E}_t] = \varnothing$

$D[\mathcal{E}_t] = \sigma^2$ – const

гомоскедастичность

$cov(\mathcal{E}_t, \mathcal{E}_s) = \varnothing$, $t \neq s$

$\forall$ $X_t = \beta_0 + \beta_1 Z_{t_1} + \mathcal{E}_t$ — Z_{t_1} — сл. в.

Если рассмотрим МНК-оценку $\hat{\beta}_1 = \dfrac{cov(X_1, Z_1)}{\sigma^2_{Z_1}}$

коэф. корреляции Пирсона

Следовательно $\hat{\beta}_1 \dfrac{\sigma_{Z_1}}{\sigma_x} = \dfrac{cov(X_1, Z_1)}{\sigma_{Z_1} \cdot \sigma_x} = r_{xz_1}$, если X и Z_1 из нормальных распределений.

$\approx 0{,}8 \approx 0{,}9$

В простейшей подстановке задачи мн-й регрессии можно получить МНК-оценки $\hat{\beta}$ и $\hat{\sigma}$, которые обладают теми же св-ми, что и в предыдущих выводах.

$\exists$ получим $\hat{\beta}$ для этой модели. $\exists$ $\hat{\beta}_1$ достаточно большая, ее t-статистика подсказывает, что $\hat{\beta}_1$ ст. значима.

$\exists$ $M[\Delta X_t] = const \neq 0$

$\exists$ $M[\Delta Z_1] = const = \varnothing$

Следовательно можем положить, что есть зависимость от времени.

Строим регрессию

$$\left. \begin{aligned} X_t &= \gamma_0 + \gamma_1 t + \mathcal{E}_t' \\ Z_{t_1} &= \eta_0 + \eta_1 t + \mathcal{E}_t'' \end{aligned} \right\} \Rightarrow$$ переход к „очищенным" переменным

Получаем очищенные аналоги

$$\text{DS процесс} \left(\begin{aligned} X_t^{(\text{r})} &= X_t - (\hat{\gamma}_0 - \hat{\gamma}_1 t) \\ Z_{t_1}^{(\text{r})} &= Z_{t_1} - (\hat{\eta}_0 - \hat{\eta}_1 t) \end{aligned} \right\}$$ снова регрессия $X_t^{(r)}$ на $Z_{t_1}^{(r)}$

r — redefind

$$X_t^{(r)} = \beta_0 + \beta_1 Z_{t_1}^{(r)} + \mathcal{E}_t''' \Rightarrow \hat{\beta}_1' \text{ – ст. незначима}$$

$\approx 0{,}0\ldots$ (t-статист. очень малая)

На „очищенных" переменных мы обнаружили, что регрессия X_t на Z_{t_1} – ложная.

$\hat{\beta}_1 \dfrac{\sigma_{Z_1^{(r)}}}{\sigma_{X^{(r)}}} = r'_{xz_1}$ — частный коэф. коррел.

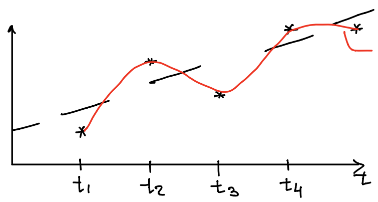
(исключены влияния других перем. регресс)

если $X, Z \sim N(0; \sigma)$, то

$$r'_{xz_1}(X, Z_1 | Z_2) = \frac{r(X, Z_1) - r(X, Z_2) \cdot r(Z_1, Z_2)}{\sqrt{1 - r^2(X, Z_2)} \cdot \sqrt{1 - r^2(Z_1, Z_2)}}$$

Проблема идентификации стр-ры модели лин. регрессии

Прибавлять перем. регрессии и смотреть на улучшение приближ. к объясн. переменной.



В будущем не сможем предсказать.

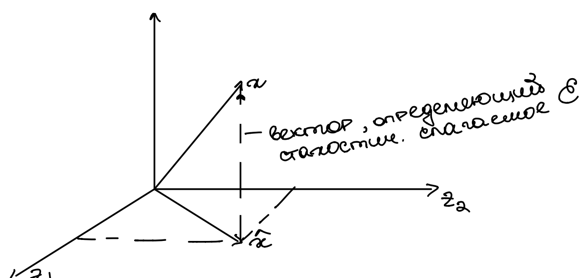
Фиктивные переменные (dummy variables) — вводимые переменные для баланса рассматриваемой системы.

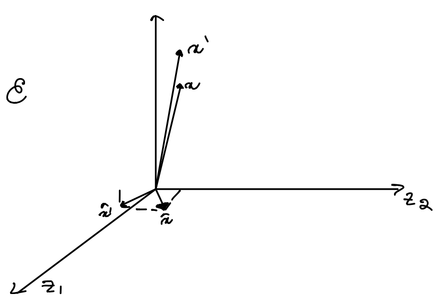
Информационные критерии: (включать или нет) (только для Гауссовских процессов)

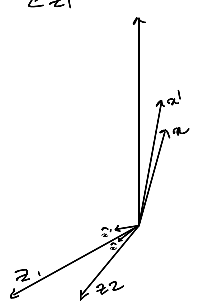
1) AIC
2) BIC (байесовский)

Основные проблемы с лин. регрессией:

1) спецификация модели
2) „ложная" регрессия
3) мультиколлинеарность регрессоров







После поворота z_2 теперь проецирование на гипотенузу

При появлении связи оценки разваливаются.
Как оценивать эту связь?
Copula — инструмент для апроксимации распределений.
=> можно вытащить оценку связи.

Следующие классы моделей ARCH гетероскедастичная модель $D[\varepsilon_t] = \sigma_t^2$

generelaized — GARCH

CH - conditional heteroscedastic

$$X_t = AR + u_t$$
$$u_t = \varepsilon\sqrt{\alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_p u_{t-p}^2}$$
$$\sigma_t^2 = \alpha_0 + \dots$$

20 апреля - определение 14:00