

Чрезмерная реакция при меняющихся во времени параметрах

Пусть $F_{t,T}$ — прогноз (цена фьючерса), S_T — реализация. Coibion–Gorodnichenko связывают *ошибку прогноза* с его *пересмотром* и уровнем:

$$\underbrace{S_T - F_{t,T}}_{\text{forecast error}} = \alpha + \beta_1 \underbrace{(F_{t,T} - F_{t-k,T})}_{\text{forecast revision}} + \beta_2 F_{t-k,T} + \varepsilon_{t,h}.$$

- $\beta_1 > 0$ — *underreaction* (information rigidity: прогноз недостаточно двигается на новости);
- $\beta_1 < 0$ — *overreaction* (прогноз двигается слишком сильно и затем разворачивается).

Значимость β_1 означает, что ошибка предсказуема, то есть прогноз рынка $F_{t,T}$ можно улучшить.

Extended CG (2015) regression: WTI Crude Oil

Horizon in days	$h = 30$	$h = 60$	$h = 180$	$h = 360$
Constant	6.303*** (2.121)	15.978*** (4.229)	37.364*** (6.209)	48.855*** (4.797)
β_1 : forecast revision	0.166* (0.091)	0.130 (0.132)	-0.383* (0.219)	-0.719*** (0.186)
β_2 : lagged forecast	-0.092*** (0.031)	-0.229*** (0.063)	-0.524*** (0.090)	-0.672*** (0.067)
Observations	233	233	233	233

Note: Robust standard errors (HC1). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

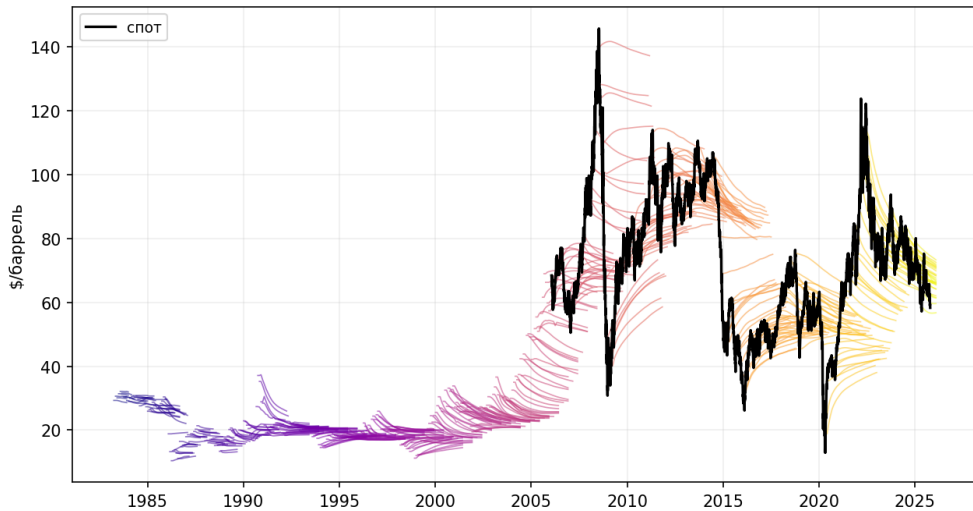
Sample period: January 2006 – September 2025. Each month we observe the spot price S_{t+h} on the contract's expiration day $t + h$.

Нашей целью является попробовать использовать модель для прогнозирования и построения торговых стратегий. Для начала попробуем оценить коэффициенты на скользящем окне и посмотрим на качество прогнозов модели

Данные. Месячные фьючерсные контракты на нефть WTI, дневная частота данных, период с 1980-х по настоящее время.

Оценка. Регрессию оцениваем на скользящем окне шириной ≈ 8 лет, отдельно для каждого горизонта h .

Данные: hair graph



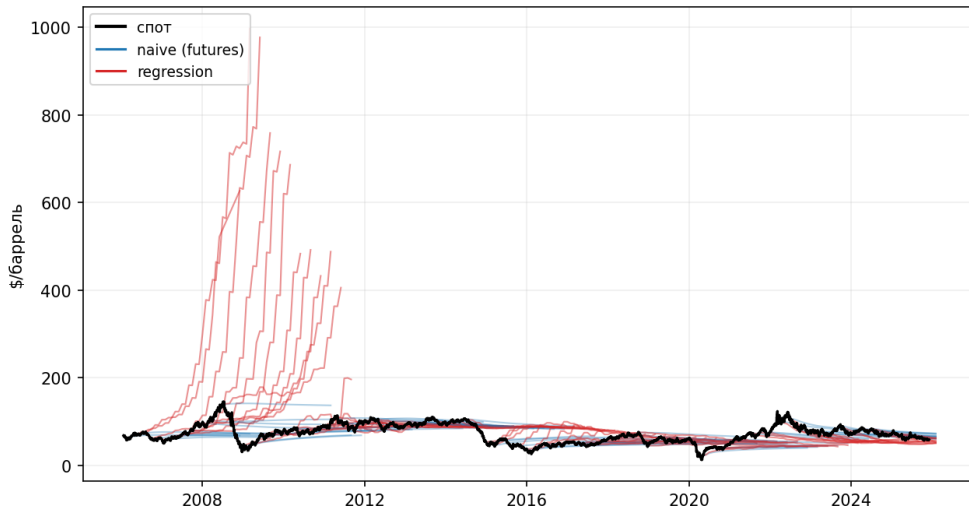
Обозначив $f_t = \log F_{t,T}$ и $S_T = F_{T,T}$:

$$f_T - f_t = \alpha + \beta_1(f_t - f_{t-k}) + \beta_2 f_{t-k} + \varepsilon,$$

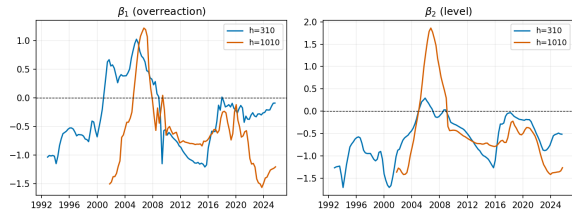
$$\Delta f_t = \alpha + \beta_1 \Delta f_{t-1} + \beta_2 f_{t-1} + \varepsilon,$$

$$f_T = \alpha + (1 + \beta_1)f_t + (\beta_2 - \beta_1)f_{t-1} + \varepsilon.$$

Прогнозы на скользящем окне



- Нас интересует β_1 (overreaction).
- Но на длинных горизонтах прогноз определяет параметр с уровнем $\beta_2 f_{t-k}$: $|\beta_2|$ растёт с h .
- К 2008 из-за продолжительного роста $\beta_2 > 0 \Rightarrow$ модель прогнозирует дальнейший огромный рост.

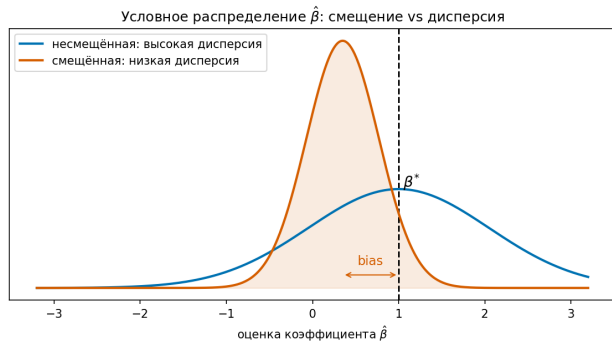


Variance-Bias tradeoff

$$\text{MSE}(\hat{\beta}) = \text{Bias}(\hat{\beta})^2 + \text{Var}(\hat{\beta}).$$

variance–bias tradeoff: допустив немного смещения (bias), можно снизить дисперсию и среднюю квадратичную ошибку оценки. Та же идея прослеживается во многих методах:

- ансамбли моделей;
- предпочтение простых моделей сложным;
- VAR vs Local Projections;
- Проклятие размерности.



При выполнении условия Гаусса-Маркова: OLS — **BLUE**: минимальная дисперсия *среди линейных несмещённых* оценок.

Но это не означает, что дисперсия будет минимальной среди всех estimators, включая смещенные (biased).

Пусть данных мало и оценка нестабильна. Добавим к выборке T *придуманных* наблюдений

$$Y_{\text{syn}} = 0, \quad X_{\text{syn}} = 1.$$

Что произойдёт с несмещённостью, эффективностью, состоятельностью?

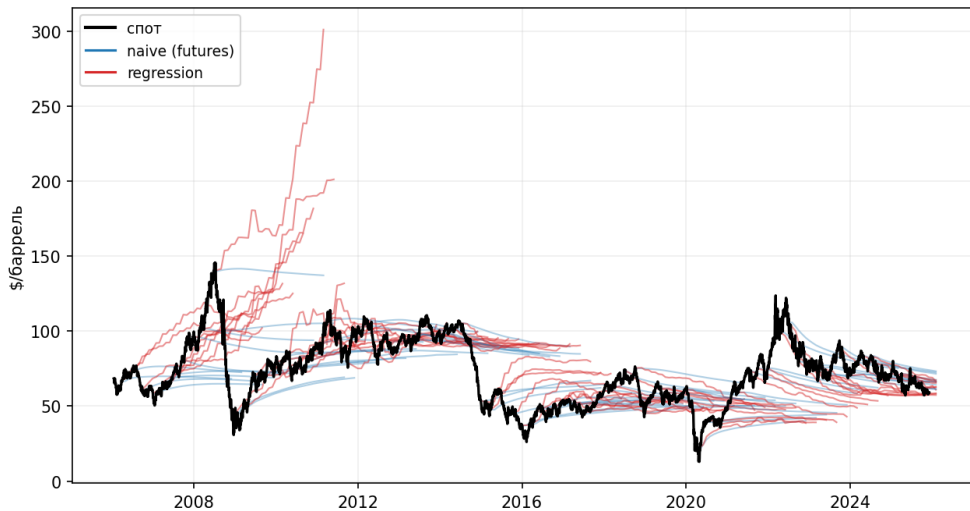
Оценка становится *смещённой* (сдвиг к 0) и *более эффективной* (дисперсия ниже); *состоятельность* сохраняется — с ростом реальной выборки вес синтетики стремится к нулю.

Общий вид: добавление $X_{\text{syn}} = \sqrt{\lambda} I$, $Y_{\text{syn}} = 0$ даёт

$$\hat{\beta} = (X'X + \lambda I)^{-1} X'Y = \arg \min_{\beta} \|Y - X\beta\|^2 + \lambda \|\beta\|^2 \quad (\text{Ridge}).$$

Добавление таких искусственных данных аналогично регуляризации с L_2 -штрафом — Ridge. λ — вес штрафа.

Прогнозы с регуляризацией



Представим, что есть прежняя выборка, на которой проводилось старое исследование. Добавляя к ней новые наблюдения, мы *обновляем* оценку.

Прежние данные имеют похожий эффект, что и синтетические: сдвигают результат к прежней оценке — только данные реальные. По сути коэффициенту задаётся *начальное распределение* до оценки.

y	x_1	x_2
0.03	0.10	4.2
−0.01	−0.05	4.3
0.02	0.08	4.1
−0.04	0.15	4.6
0.01	−0.02	4.5

серые — прежние данные,
синие — новые

Все вышеприведенные методы оценки — частные случаи **байесовского** подхода.

- **Frequentist (Частотный подход):**

- Истинный параметр β — фиксированная, но неизвестная константа.
- Оценка $\hat{\beta}$ — случайная величина, ее свойства зависят от выборки и самого метода оценки.

- **Bayes (Байесовский подход):**

- Параметр β — случайная величина (описывается распределением).
- *Prior* $p(\beta)$ отражает представления о параметре до оценки.
- *Posterior* $p(\beta | Y)$ обновляет эти знания реальными данными Y .

$$p(\beta | Y) = \frac{p(Y | \beta) p(\beta)}{p(Y)} = \frac{p(Y | \beta) p(\beta)}{\int p(Y | \beta) p(\beta) d\beta}$$

Пусть заданы нормальное правдоподобие и нормальное априорное распределение:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I) \quad \text{и} \quad \beta \sim \mathcal{N}(\beta_0, \Sigma_0)$$

Тогда апостериорное распределение также нормально: $\beta \mid Y \sim \mathcal{N}(\beta_1, \Sigma_1)$.

Оценка параметров:

$$\begin{aligned} \Sigma_1^{-1} &= \Sigma_0^{-1} + \sigma^{-2} X'X & \iff & \Sigma_{\text{OLS}}^{-1} = \sigma^{-2} X'X \\ \beta_1 &= \Sigma_1 (\Sigma_0^{-1} \beta_0 + \sigma^{-2} X'Y) & \iff & \beta_{\text{OLS}} = \Sigma_{\text{OLS}} (\sigma^{-2} X'Y) \end{aligned}$$

Связь с Ridge-регрессией

Если у нас нет начальных знаний ($\beta_0 = 0$) и ($\Sigma_0 = \frac{\sigma^2}{\lambda} I$), то апостериорное среднее в точности совпадает с Ridge-оценкой:

$$\beta_1 = (X'X + \lambda I)^{-1} X'Y = \hat{\beta}_{\text{Ridge}}$$

Последовательное обновление

Байесовский подход позволяет последовательно обновлять оценку β по мере поступления новых данных.

1. Оптимальное объединение двух сигналов

Если есть два сигнала z_1, z_2 с ошибками σ_1^2, σ_2^2 , то оптимально:

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2 \quad \Longleftrightarrow \quad \frac{1}{\sigma_{\hat{x}}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

2. Последовательно обновление

Задача обновления сводится к нахождению k_t :

$$\hat{\beta}_t = (1 - k_t) \underbrace{\hat{\beta}_{t-1}}_{\text{история}} + k_t \underbrace{z_t}_{\text{сигнал}}, \quad \text{где } k_t = \frac{\sigma_{\beta, t-1}^2}{\sigma_{\beta, t-1}^2 + \sigma_{\varepsilon}^2}$$

Фильтр Калмана позволяет находить скрытые состояния (например β) из сигналов.

- 1 **Prediction (Прогноз):** Пересчет Prior. Параметр β_t и наша неопределенность относительно параметра $P = \text{Var}(\beta_t \mid Y_{1:t-1})$ смещаются во времени согласно заданному моделью процессу.
- 2 **Update (Обновление):** Учет новых данных происходит с учетом шума сигнала $R = \text{Var}(\varepsilon_t)$.

Связь с теоремой Байеса

Шаг Update — это перемножение Prior и Likelihood:

$$\underbrace{p(\beta_t \mid Y_{1:t})}_{\text{Posterior}} \propto \underbrace{\mathcal{N}(y_t \mid \beta_t, R)}_{\text{Likelihood (Новый сигнал)}} \times \underbrace{\mathcal{N}(\beta_t \mid \hat{\beta}_{t|t-1}, P)}_{\text{Prior (Прогноз модели)}}$$

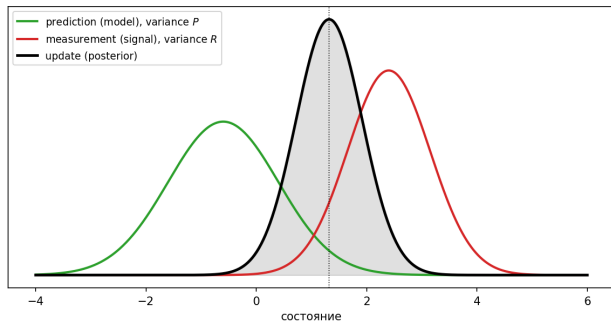
Максимум этого Posterior (его среднее) дает точно ту же рекурсию с весом $k_t = \frac{P}{P+R}$.

Обновлённая оценка:

$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + K_t(y_t - X_t\hat{\beta}_{t|t-1})$$

$$K_t = P_{t|t-1}X_t^T(X_tP_{t|t-1}X_t^T + R)^{-1}$$

Вес K_t (Kalman gain) больше, если априорная неуверенность P велика или сигнал точен (малый R).



Модель пространства состояний (State-Space)

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, H) \quad (\text{уравнение наблюдения})$$

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q) \quad (\text{уравнение состояния})$$

1. Шаг прогноза (Prediction)

$$a_{t|t-1} = T_t a_{t-1|t-1}$$

$$P_{t|t-1} = T_t P_{t-1|t-1} T_t^T + R_t Q R_t^T$$

2. Шаг обновления (Update)

$$K_t = P_{t|t-1} Z_t^T (Z_t P_{t|t-1} Z_t^T + H)^{-1}$$

$$a_{t|t} = a_{t|t-1} + K_t (y_t - Z_t a_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - K_t Z_t P_{t|t-1}$$

Специализация: $\alpha_t = \beta_t$, $Z_t = X_t$, $T_t = I$, $R_t = I$, $H = R$.

$$y_t = X_t \beta_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, R) \quad (\text{уравнение наблюдения})$$

$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q) \quad (\text{уравнение состояния})$$

1. Шаг прогноза (Prediction)

$$\hat{\beta}_{t|t-1} = \hat{\beta}_{t-1|t-1}$$

$$P_{t|t-1} = P_{t-1|t-1} + Q$$

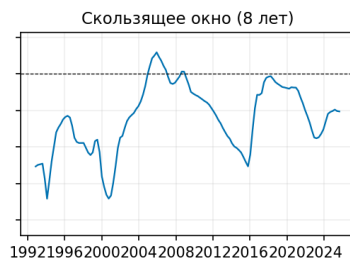
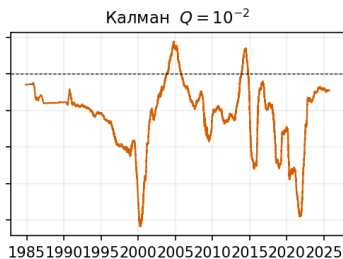
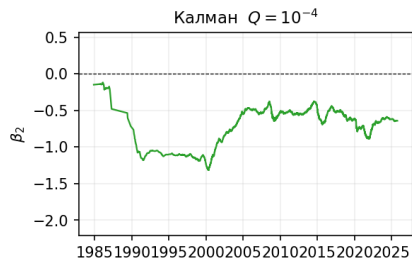
2. Шаг обновления (Update)

$$K_t = P_{t|t-1} X_t^T (X_t P_{t|t-1} X_t^T + R)^{-1}$$

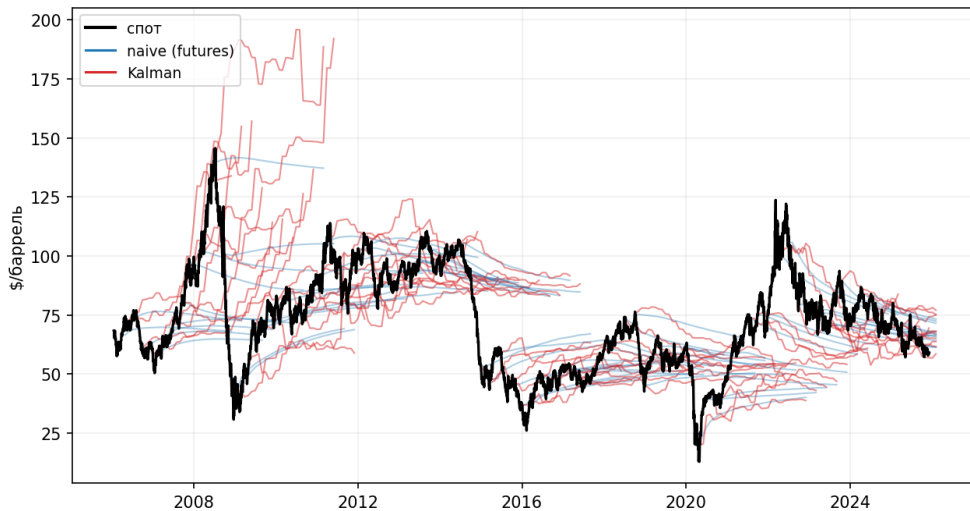
$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + K_t (y_t - X_t \hat{\beta}_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - K_t X_t P_{t|t-1}$$

Оценка β_2 : фильтр Калмана и скользящее окно



Прогнозы фильтра Калмана ($Q = 10^{-4}$)



Прогноз $\hat{y}_{t,h}$ сам по себе не позиция. Его превращают в сигнал и позицию правилом $q_t = g(\hat{y}_{t,h})$; для торговли важно *направление*, а не точное значение, поэтому правило может быть проще модели. Например, пороговое правило:

$$q_t = \begin{cases} +1, & \hat{y}_{t,h} > \tau, \\ 0, & |\hat{y}_{t,h}| \leq \tau, \\ -1, & \hat{y}_{t,h} < -\tau. \end{cases}$$

Критерии оценки стратегий: эффективность

Торговую стратегию оценивают комплексно по метрикам доходности, риска и стабильности прогноза, а не только по накопленной прибыли:

Средняя доходность:
$$\bar{R} = \frac{1}{n} \sum_{t=1}^n R_t$$

Коэффициент Шарпа:
$$SR = \sqrt{m} \frac{\bar{R} - r_f}{\hat{\sigma}_R} \quad (m — \text{фактор годовой частоты})$$

Доля попадания:
$$\text{Hit-Rate} = \frac{1}{n} \sum_{t=1}^n 1\{\text{sign}(\hat{y}_t) = \text{sign}(r_{t+1})\}$$

- **Sharpe Ratio** нормирует избыточную доходность на волатильность портфеля.
- **Hit-Rate** показывает качество бинарного прогноза направления знака (up/down).

Критерии оценки стратегий: риск и издержки

Для практического применения стратегии критически важно оценить устойчивость к просадкам, транзакционные издержки и рыночный риск:

Макс. просадка (MDD):
$$\text{MDD} = \max_t \frac{\max_{s \leq t} V_s - V_t}{\max_{s \leq t} V_s} \quad (\text{где } V_t \text{ — капитал})$$

Связь с рынком (например, S&P 500) отделяют регрессией CAPM:

$$R_t - r_f = \alpha + \beta (R_t^m - r_f) + \varepsilon_t$$

Цель стратегии: обычно выступать элементом диверсифицированного портфеля. Для этого ищут значимую α (доходность сверх рынка) при минимальной рыночной экспозиции β .