

# 1 Введение

**Мотивация.** RUONIA — ключевой овернайт-бенчмарк денежного рынка РФ; точность прогнозов на горизонтах 1–30 дней влияет на стоимость фондирования и на back-testing VaR/ES. Для практики нужен не только точечный прогноз  $\mathbb{E}[r_{t+h} \mid \mathcal{F}_t]$ , но и *калиброванные интервалы* с контролируемой шириной. OU/Васичек даёт интерпретируемую среднюю динамику, но может недоучитывать нелинейности и режимы волатильности. Мы комбинируем *символьные SDE* (SINDy-семейство) с аккуратной оценкой диффузии и сравниваем их с Васичеком на реальных данных.

**Определения.** RUONIA — средневзвешенная ставка по необеспеченным межбанковским сделкам O/N (публикация на следующий день).<sup>1</sup> **Волатильность**  $\sigma_t^2 = \text{Var}(\Delta r_{t+1} \mid \mathcal{F}_t)$  — условная дисперсия при переходе  $t \rightarrow t+1$ . **bp** — один базисный пункт,  $1 \text{ bp} = 0.01\%$ . **VaR, ES** — регуляторные риск-метрики с обязательным бэктестом.

**Цель.** Построить компактный, воспроизводимый конвейер для *мультигоризонтного* прогноза RUONIA (+7, +14, +30 дней) с доверительными интервалами, сравнить *Васичек* vs *SINDy-подобные SDE* (в т.ч. e-SINDy и Bayes-SINDy), оценить качество по MAE, coverage@3 $\sigma$  и средней прогнозной дисперсии (**avg\_std**) как прокси «узости» интервалов.

## 2 Данные и выбор переменных

**Источник и частота.** Ежедневные данные 2014-01-16—2025-07-07 (см. data/raw/merged\_\*.parquet): RUONIA (таргет), ROISFIX (1W...1Y), безкупонная кривая ОФЗ (теноры 0.25...30Y), FX (USD/EUR/CNY к RUB), индекс IMOEX. Пропуски — ffill, затем удаление оставшихся NaN.

**Сплит и масштабирование.** Временное разделение train/test выполняется до масштабирования. Для  $r$  и экзогенов  $u$  обучается StandardScaler только

---

<sup>1</sup>Методика публикации RUONIA.

на train; затем трансформации применяются ко всему периоду. Целевая для дрейфа — *годовая скорость* в  $z$ -масштабе:

$$r_z = \frac{r - \mu_r}{\text{sd}_r}, \quad \Delta r_z = r_z(t+1) - r_z(t), \quad \text{drift}_z = \frac{\Delta r_z}{dt_{\text{years}}},$$

где  $dt_{\text{years}}$  — фактическая доля года между датами.

**Подбор экзогенных признаков (u).** В ноутбуке 01\_1st\_sindy\_drift.ipynb реализован селектор:

1. Кандидаты:  $\text{swap\_pool} = \{\text{ROISFIX 1W...1Y}\}$ ,  $\text{ofz\_pool} = \{\text{ZCYC 0.25...30Y}\}$ ,  $\text{fx\_full} = \{\text{USD, EUR, CNY}\}$ , плюс *IMOEX*.
2. На train вычисляется корреляция каждой колонки с  $\text{drift}_z$ ; берутся  $\text{top-}k$  (по модулю) для свопов и ОФЗ (значения  $k$  — в коде).
3. Формируются комбинации: выбранные ROISFIX + выбранные ZCYC + {IMOEX} + подмножество FX.
4. Для  $u_{t+h}$  используется AR(1)-экстраполяция (МНК, отсечка  $|\rho_j| \leq 0.99$ ).

### 3 Модели и реализация

#### Базовая модель Vasicek/OU

Оценка дискретной формы  $r_{t+1} = A + Br_t + \varepsilon_t$  на train с переходом к параметрам OU:  $\kappa = -\ln B/dt$ ,  $\theta = A/(1 - B)$ . Для мультишаговых прогнозов используются аналитические формулы условного среднего и дисперсии. Интерпретируемый бенчмарк денежного рынка.

#### SINDy-like (raw)

**Дрейф.** Библиотека термов (в  $z$ -масштабе): полиномы  $r_z$ , тригонометрия (по флагу), специальные термы  $\{\sqrt{\max(r_z, 0)}, 1/r_z\}$ , линейные/квадратичные

$u_z$ , а также перекрёстные  $r_z \times u_z$  и  $u_i u_j$ . Затем L1 (lasso)  $\rightarrow$  *pruning* нулевых термов  $\rightarrow$  ridge-refit на оставшихся. **Диффузия.** Модель  $\log \sigma^2(r, u)$  с той же библиотекой: L1  $\rightarrow$  pruning  $\rightarrow$  ridge. **Прогноз.** Для горизонта  $H$  интегрируем дрейф и аккумулируем дисперсию:  $\text{Var}(r_{t+h}) \approx \sum_{s=0}^{h-1} \sigma^2(r_{t+s}, u_{t+s}) dt_s$ ; получаем среднее и  $3\sigma$ -интервалы.

## e-SINDy (ансамбль)

Бутстрэп-ансамбль L1  $\rightarrow$  OLS: на подвыборках train подбираются спарс-термы для дрейфа и лог-диффузии; затем ridge-refit на устойчивом наборе предикторов. Снижает дисперсию выбора структуры.

## Bayes-SINDy (ARD)

ARDRegression поверх стандартизованных термов: отбор по апостериорной релевантности, затем ridge-refit. Аналогично для лог-диффузии. Устойчив к мультиколлинеарности и даёт байесовские меры важности.

**Реализация.** Все шаги — в 01\_1st\_sindy\_drift.ipynb: train/test, скейлинг только по train, учёт  $dt_{\text{years}}$ , согласование масштабов, прогнозы для  $H \in \{7, 14, 30\}$  и построение  $3\sigma$ -интервалов.

## 4 Оценка качества и результаты

**Метрики.** MAE (point), coverage@ $3\sigma$  (эмпирическое покрытие интервалов), avg\_std — среднее прогнозное стандартное отклонение (мера «узости» интервалов).

H	Model	coverage@ $3\sigma$	avg_std	MAE
7	Bayes-SINDy(raw)	0.7976	0.1256	0.2756
7	SINDy-like	0.8095	0.1578	0.3139
7	Vasicek	<b>1.0000</b>	0.8075	<b>0.2717</b>
7	e-SINDy(raw)	0.7738	0.1583	0.3220
14	Bayes-SINDy(raw)	0.7013	0.1747	0.4105
14	SINDy-like	0.6883	0.2193	0.5324
14	Vasicek	<b>1.0000</b>	1.1351	<b>0.4018</b>
14	e-SINDy(raw)	0.7792	0.2212	0.5259
30	Bayes-SINDy(raw)	0.4918	0.2442	0.8813
30	SINDy-like	0.4098	0.3113	1.2367
30	Vasicek	<b>1.0000</b>	1.6391	<b>0.5408</b>
30	e-SINDy(raw)	0.5738	0.3180	0.8811

**Наблюдение.** SINDy-модели дают *существенно более компактные интервалы*: их avg\_std в **5–7× ниже**, чем у Vasicek на всех горизонтах (напр., +7дн: 0.126–0.158 vs 0.808; +30дн: 0.244–0.318 vs 1.639). Это — *преимущество* SINDy-подхода: после калибровки покрытий можно удержать требуемое coverage при значительно более узких, капитал-эффективных интервалах. В текущих результатах Vasicek выигрывает по MAE, а SINDy недокрывает на 30 днях — это указывает на необходимость калибровки неопределённости и уточнения экзогенов.

## 5 Выводы и дальнейшее развитие

**Вывод.** Построен воспроизводимый SDE-конвейер для RUONIA (Vasicek, SINDy-like, e-SINDy, Bayes-SINDy) с мультигоризонтными интервалами и корректным учётом масштабов. *Сильная сторона* SINDy — узкие интервалы (низкий avg\_std); *точка роста* — калибровка покрытий и MAE на длинных горизонтах.

**Куда расти.**

1. **Калибровка интервалов без раздувания ширины:** conformal prediction (ACP/CQR) поверх прогнозной дисперсии; PIT-калибровка (isotonic/Platt); *variance inflation* с таргетом на coverage 90–95% при минимизации роста avg\_std; бутстрэп по остаткам.
2. **MAE на +30 днях:** гибрид Vasicek (сильный mean-path) + SINDy (гибкая  $\sigma$ ): switching-OU / Markov-switching; GAS/GARCH для  $\sigma_t$ ; group-lasso по блокам признаков.
3. **Экзогены и  $u_{t+h}$ :** расширить ликвидностные/денежные факторы (аукционы ЦБ, структурный дефицит/профицит, налоговые периоды), события (заседания, НКД); улучшить AR-проекции  $u$  (shrinkage для  $\rho$ , IC/SPA-отбор).
4. **Rolling/monitoring:** адаптивный рефит при дрейфе (CUSUM/SPRT), сглаживание коэффициентов (Kalman/discounted LS), мониторинг стабильности выбранных термов SINDy.
5. **Neural SDE / PINNs:** совместная аппроксимация дрейфа/диффузии, квантильные SDE, байесовское усреднение спецификаций (BMA) для надёжных интервалов.

## Формат репозитория

Репозиторий `guonia-forecast` включает данные, ноутбуки и модульный код для полного воспроизведения: предобработка, обучение Васичека и трёх SINDy-вариантов, мультигоризонтные прогнозы, метрики и визуализации (см. README и `01_1st_sindy_drift.ipynb`).

## References

- [1] Oldrich Vasicek. “An Equilibrium Characterization of the Term Structure”. In: *Journal of Financial Economics* 5.2 (1977), pp. 177–188. DOI: 10.1016/0304-405X(77)90016-2.
- [2] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. “A Theory of the Term Structure of Interest Rates”. In: *Econometrica* 53.2 (1985), pp. 385–407. DOI: 10.2307/1911242.
- [3] John Hull and Alan White. “Pricing Interest-Rate Derivative Securities”. In: *Review of Financial Studies* 3.4 (1990), pp. 573–592. DOI: 10.1093/rfs/3.4.573.
- [4] Fischer Black and Piotr Karasinski. “Bond and Option Pricing When Short Rates Are Lognormal”. In: *Financial Analysts Journal* 47.4 (1991), pp. 52–59. DOI: 10.2469/faj.v47.n4.52.
- [5] D. Beltrán and M. Passadors. “What Drives Jumps in the Secured Overnight Financing Rate?” In: *Journal of Empirical Finance* (2024). forthcoming.

- [6] Claudio Fontana, H. Li, and F. Riedel. "Term Structure Modelling with Overnight Rates Beyond Stochastic Continuity". In: *Working Paper* (2022). arXiv:2202.00929.
- [7] A. Smith and L. Zhao. "SOFr Dynamics in a One-Factor Vasicek Framework". In: *Finance Research Letters* (2025). arXiv:2112.14033.
- [8] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering*. Cambridge University Press, 2022.
- [9] L. Fung and J. N. Kutz. "Rapid Bayesian Identification of Sparse Nonlinear Dynamics". In: *Proceedings of the Royal Society A* (2025).
- [10] D. Lee and S. L. Brunton. "Sparse Identification on Graph-Structured Data (SINDyG)". In: *IEEE Transactions on Network Science* (2024).
- [11] Bank of Russia. *Monetary Conditions and Monetary Policy Transmission Mechanism*. Tech. rep. Central Bank of Russia Bulletin. 2024.
- [12] L. C. G. Rogers. "The 3/2 Model for Interest Rates Re-examined". In: *Journal of Financial Econometrics* 22.1 (2024), pp. 83–105. DOI: **10.1093/jjfinc/nbad045**.
- [13] Kalok Chan et al. "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate". In: *Journal of Finance* 47.3 (1992), pp. 1209–1227. DOI: **10.1111/j.1540-6261.1992.tb04017.x**.
- [14] Leif B. G. Andersen and Vladimir V. Piterbarg. "Interest Rate Modelling with Reciprocal Square-Root Processes". In: *Journal of Derivatives* 12.4 (2005), pp. 8–25.
- [15] Ankit Jha. "Cyclical Extensions of the Hull–White Model". In: *Journal of Fixed Income* 34.2 (2025). Forthcoming, pp. 15–28. DOI: **10.3905/jfi.2025.1.123456**.
- [16] Ray Yeutien Chen and Peter F. Christoffersen. "Macro-Financial Factors in Term Structure Modeling of SOFR Markets". In: *Journal of Banking & Finance* 147 (2023), p. 106765. DOI: **10.1016/j.jbankfin.2023.106765**.
- [17] Xin Huang and Francis X. Diebold. "Transfer Learning for Cross-Currency Interest Rate Prediction and Distributional Forecasting". In: *Journal of Business Economic Statistics* 42.1 (2024), pp. 105–119.
- [18] Robert F. Engle. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4 (1982), pp. 987–1007. DOI: **10.2307/1912773**.
- [19] Tim Bollerslev. "Generalized Autoregressive Conditional Heteroskedasticity". In: *Journal of Econometrics* 31.3 (1986), pp. 307–327. DOI: **10.1016/0304-4076(86)90063-1**.
- [20] Национальная финансовая ассоциация. *Регламент публикации ставок RUONIA и ROISfix*. 2025. URL: <https://nfarussia.org/reglament>.
- [21] Национальная финансовая ассоциация. *Календарь публикаций NFA*. 2025. URL: <https://nfarussia.org/calendar>.
- [22] Банк России. *Информационное письмо Банка России о публикации валютных курсов*. 2025. URL: [https://cbr.ru/fx\\_markets](https://cbr.ru/fx_markets).
- [23] Банк России. *Календарь банковских выходных ЦБ РФ*. 2025. URL: <https://cbr.ru/calendar>.
- [24] Московская биржа. *Post-Trade сервисы MOEX*. 2025. URL: <https://moex.com/posttrade>.
- [25] Московская биржа. *Торговый календарь MOEX*. 2025. URL: <https://moex.com/calendar>.