УДК 004.94:005.7

1.4. Формонезависимое моделирование поверхности подразумеваемой волатильности на основе гауссовых смесей

Саданув Батырхан, Араратян Севак, Меркулов Вадим
Freedom Broker, г. Алматы, Казахстан

В работе представлена новая методика моделирования поверхности подразумеваемой волатильности, основанная на комбинации гауссовых смесей (Gaussian Mixture Models, GMM) и гауссовой регрессии смесей (Gaussian Mixture Regression). В отличие от традиционных методов, предполагающих фиксированную функциональную форму поверхности волатильности, предлагаемый фреймворк является формонезависимым (shape-agnostic) и способен гибко описывать сложную динамику волатильности для различных классов активов, включая широкий спектр форм улыбок — от классических U-образных до многомодальных W-образных и других, встречающихся в реальной торговой практике. Используя вероятностную выразительность GMM и прогностическую мощь регрессии, метод достигает точности, сопоставимой с моделями глубокого обучения, при этом превосходя их по вычислительной эффективности, не требуя обучения. Эмпирические эксперименты показывают, что предложенный подход превосходит существующие параметрические модели по гибкости и устойчивости, обеспечивая при этом точность, аналогичную нейросетевым решениям, но с заметно меньшими вычислительными затратами. Благодаря своей лёгкости, масштабируемости и интерпретируемости метод особенно подходит для высокочастотных финансовых приложений, предоставляя исследователям и практикам надежный инструмент для моделирования волатильности и управления рисками.

Введение

Поверхность подразумеваемой волатильности (IV-surface) является фундаментальным инструментом в количественных финансовых системах: она позволяет унифицировать котировки опционов с разными страйками и сроками, оценивать переоценки или недооценки, строить индексы IV и проводить управление рисками. Как подчеркивается в литературе, поверхность подразумеваемой волатильности представляет собой базовый элемент вычислительных финансов, лежащий в основе большинства моделей оценки и управления рисками [1]. Однако значение IV-поверхности выходит далеко за рамки лишь количественных систем. Она также играет существенную роль в фундаментальном анализе, поскольку отражает совокупные ожидания рынка относительно будущей неопределенности и риска. Например, форма и динамика IV-поверхности содержат информацию о сохраняющейся или нарастающей волатильности, что может быть связно с корпоративными событиями, изменениями дивидендной политики, технологическими шоками, макроэкономическими изменениями и другими событиями фундаментального или количественного характера [2]. В этом смысле IV-поверхность выступает не просто инструментом ценообразования, но и индикатором рыночных восприятий и ожиданий: она может быть использована при оценке фундаментальной стоимости компании, формировании премий за риск, а также отслеживании настроений инвесторов и потенциальной реакции рынка. Например, исследование показывает, что вытянутые кривые IV по срокам и страйкам отражают информацию о будущих доходностях акций и ожиданиях рынка [3]. Таким образом, интеграция IV-поверхности в анализ — как количественный, так и фундаментальный — позволяет получить более полное понимание рыночной динамики: не только с точки зрения распределения вероятностей будущей волатильности, но и с точки зрения сигналов, которые рынок посылает о рисках, неопределенности и изменении условий.

Для практических приложений важно не просто смоделировать поверхность подразумеваемой волатильности, но сделать это быстро, точно и без арбитражных нарушений, поскольку даже небольшие искажения в реальном времени могут вести к серьезным последствиям. В условиях краткосрочного/среднесрочного хеджирования, высокочастотной/внутридневной торговли задержки и шумы в модели могут привести к следующим нежелательным эффектам:

- ошибочной оценке риска — из-за неточной поверхности волатильности могут быть неправильно рассчитаны грек-параметры (дельта, вега, гамма) и, следовательно, хедж-позиции окажутся неадекватными;
- неэффективному распределению капитала — торговые или хеджевые решения могут базироваться на неверной оценке подразумеваемой волатильности, что ведет к избыточной или недостаточной экспозиции;
- арбитражным потерям или усилению систематического шума — модель с плохо сглаженной поверхностью может выдавать нелогичные точки (например, нарушение условий арбитража), что создаёт риск эксплуатации или потерь;

- искажению индексов волатильности и последующему ухудшению качества стратегий риск-менеджмента — если IV-показатели построены на шумной поверхности, то стресс-тесты, сценарное моделирование и расчёты хвостовых рисков становятся менее надежными.

Множество исследований подчеркивают важность скорости, точности и отсутствия арбитража при моделировании IV-поверхностей. Так, в работе «High-frequency dynamics of the implied volatility surface» автор указывает, что условия отсутствия арбитража (no-arbitrage) могут быть выражены через параметры модели, и уделяет внимание высокочастотной динамике IV-поверхности [4]. Другая работа «A Two-Step Framework for Arbitrage-Free Prediction of the Implied Volatility Surface» рассматривает задачу предсказания поверхности с учётом предотвращения арбитража и подчёркивает, что конструкция поверхности должна быть как гибкой, так и работоспособной для приложения в реальном времени [5]. В исследовании «The latency accuracy trade-off and optimization in implied volatility modelling» отмечает, что в таких системах существует явный компромисс между скоростью (латентностью) и точностью: чем выше скорость (меньше задержка), тем выше риск потери точности, что критично для количественного трейдинга [6]. Наконец, в статье «Imposing No-Arbitrage Conditions In Implied Volatility Surfaces Using Constrained Smoothing Splines» подробно анализируются техники обеспечения арбитраж-свободы при сглаживании поверхностей подразумеваемой волатильности, что важно для корректной работы риск-моделей [7].

Во всех этих работах видно: модель поверхности подразумеваемой волатильности должна удовлетворять четырем ключевым требованиям:

1. высокая точность — чтобы адекватно отражать динамику рынка и позволять корректно выявлять переоценки/недооценки;
2. минимальная задержка — чтобы интегрироваться в системы реального времени или высокочастотную торговлю;
3. гибкость по форме — чтобы адаптироваться к различным рынкам, страйкам, срокам, ликвидности;
4. отсутствие арбитражных нарушений — чтобы поверхность была экономически корректной и не содержала нелогичных точек.

Существующие подходы и их ограничения

Классические параметрические подходы, прежде всего SVI и SABR, закрепились в практике из-за простоты, скорости и удобства калибровки. SVI задаёт улыбку для фиксированного срока экспирации в виде компактной параметризации с контролируемой асимптотикой по лог-страйку; в ряде работ показано, как накладывать условия отсутствия статического арбитража, чтобы получить экономически согласованные поверхности. Однако при всех достоинствах SVI остаётся жёстко заданной функциональной формой, что ограничивает ее способность описывать сложные и многомодальные профили рынка [8]. SABR, в свою очередь, дает стохастическую динамику волатильности и обеспечивает удобные асимптотические формулы для implied volatility, широко используемые на практике. Тем не менее точность асимптотик и арбитраж-свобода требуют осторожности (особенно в крыльях и при экстремальных страйках), а расширения, устраняющие арбитражные аномалии, усложняют численные процедуры и повышают вычислительную стоимость [9]. Непараметрические и полупараметрические методы сглаживания призваны повысить гибкость по форме. Важное направление — арбитраж-свободное сглаживание сплайнами: накладываются ограничения монотонности/выпуклости, совместимые с ценовыми неравенствами и условиями отсутствия арбитража. Эти методы сравнительно просты и быстры, но чувствительны к конфигурации данных и всё равно вводят структурные ограничения через класс сглаживающих функций [10]. Отдельный класс — многостадийные (multistage) или двухшаговые фреймворки. Типичный дизайн: на первом шаге извлекают компактное представление поверхности (факторы/признаки), на втором — восстанавливают полную поверхность с явным контролем арбитража. Такие подходы улучшают стабильность и совместимость с динамикой во времени, но опираются на относительно тяжёлые модели восстановления (включая глубокие сети) и сложные пайплайны [5]. С развитием машинного обучения получили распространение модели на нейросетях, в том числе с мягким или жёстким учётом no-arbitrage ограничений, а также генеративные модели (VAE/GAN/диффузионные) для имитации или до-заполнения поверхностей. Они способны подстраиваться под сложную геометрию IV-surface и часто дают высокую точность, но требуют существенных вычислительных ресурсов, длительного обучения, тонкой настройки и операционного сопровождения — что затрудняет применение в средах с ограничениями по латентности [11]. Ключевой практический вызов обостряется на коротких сроках до экспирации. Теория и эмпирика показывают, что близко к экспирации поведение улыбки резко зависит от структуры прыжков и свойств процесса цены; возникают характерные «взрывы»/крутые крылья, которые плохо поддаются жёстким параметризациям и требуют повышенной гибкости и аккуратной регуляризации [12]. Дополнительные сложности возникают вокруг корпоративных и макроэкономических событий. Перед отчетами по прибыли, заседаниями ФРС и другими релизами неопределённость увеличивается, а после — резко переоценивается; соответствующие деформации IV-поверхности документированы в академической литературе для акций и процентных ставок. Влияние проявляется как в уровнях IV, так и в форме улыбки/поверхности, усиливая требования к модели: она должна быть одновременно гибкой (по форме),

устойчивой (к шуму и пропускам данных) и операционно быстрой [13]. Таким образом, в текущем состоянии литературы наблюдается устойчивый компромисс: быстрые и простые схемы (SVI, SABR, сплайны) удобны и распространены, но накладывают сильные предпосылки на геометрию поверхности; более гибкие и точные ML/генеративные подходы уменьшают структурные ошибки, зато усложняют калибровку и повышают требования к вычислениям и поддержке, что особенно критично в эпоху высокочастотной торговли и алгоритмических систем с жесткими SLA по задержкам [8]. Наконец, даже обзорные работы подчеркивают, что выбор параметризации/функционала калибровки и численной процедуры напрямую влияет на устойчивость и арбитраж-свободу результирующей поверхности; а при скудных/шумных данных проблема усложняется еще сильнее. Это особенно заметно на коротких экспирациях, где «локальная» геометрия поверхности сильно меняется во времени, и вокруг событий, когда данные частичны и быстро устаревают [14].

Форма улыбки (или поверхности) подразумеваемой волатильности претерпела значительную эволюцию: если еще десятилетие-два назад доминировали сравнительно простые «U-образные» кривые (где implied volatility по мере удаления страйка от at-the-money возрастает), то сегодня на рынках можно наблюдать гораздо более сложные структуры — «W-образные», «S-образные», линейные участки, экспоненциальные хвосты, а при сильных событиях даже «многогорбые» формы или вовсе отсутствие улыбки. Так, исследование Алексcиу и др. фиксирует, что перед объявлениями прибыли у опционов с короткими сроками нередко возникают обратные (concave) кривые-формы, включая S- или W-виды, что резко отличается от привычной выпуклой улыбки [15]. Дополнительно, в теоретических и эмпирических исследованиях показано, что форма кривой подразумеваемой волатильности может принимать широкий спектр конфигураций — от классических U-образных до W-образных и осциллирующих структур с несколькими минимумами и максимумами. Такие многоэкстремальные профили возникают в зависимости от параметров распределения доходностей и структуры рыночных ожиданий, отражая многорежимность поведения волатильности и неоднородность потоков риска на рынке [16]. В частности, при очень коротком сроке до экспирации или перед корпоративным/макрособытием, когда риск наступления крупного прыжка или изменения режима существенно возрастает, распределение риска становится многомодальным, что провоцирует «двугорбые» или даже «трёхгорбые» улыбки [17]. Также важно отметить, что классические асимптотические исследования показывают, что при подходе к нулевому времени до экспирации или крайним страйкам распределение может демонстрировать гораздо более резкие изменения формы, чем простая квадратичная зависимость лог-страйка [18]. Таким образом, современная эмпирика и теория свидетельствуют о том, что модели, предусматривающие лишь U-образную или линейную форму улыбки, значительно уступают в адекватности реальной динамике рынка — требуется гораздо большая гибкость, чтобы корректно отражать, например, нелинейности, асимметрии, мультигорбость, эффекты ликвидности и режимов ожидания событий.

Методология

В данной работе мы предлагаем метод — на основе гауссовых смесей (GMM) в сочетании с регрессией — который удовлетворяет всем требованиям, упомянутым в предыдущих разделах: он является *shape-agnostic* (не накладывает жесткой формы), показывает точность, сравнимую с глубокими нейросетями, но выигрывает по скорости и ресурсным затратам. Эмпирические тесты показывают, что предложенный метод превосходит традиционные параметрические модели с точки зрения гибкости и аналогичен результатам сложных моделей в плане точности, при более легкой реализации и меньших вычислительных требованиях.

Bjerksund-Stensland Model: исходная подразумеваемая волатильность

На первом этапе собирается рыночная информация: премии опционов (call/put) для различных страйков K и сроков до экспирации τ. Для каждого опциона, особенно американского стиля (с учетом дивидендной доходности и возможности раннего исполнения), применяется аппроксимационная модель Bjerksund–Stensland, обеспечивающая аналитическую скорость при высокой точности. Этот метод даёт теоретическую цену опциона $C_{model}(\sigma)$ как функцию волатильности σ. Затем решается обратная задача: нахождение такой волатильности σ_{raw}, при которой модельная цена совпадает с рыночной,

$$C_{model}(\sigma_{raw}) \approx C_{mkt}.$$

В результате для каждой пары (K_i, τ_i) вычисляется значение $\sigma_{raw,i}$, формируя множество точек

$$\{(K_i, \tau_i) \to \sigma_{raw,i}\}.$$

Для американского опциона с дивидендами теоретическая цена по Bjerksund–Stensland (2002) выражается аналитически:

$$C_{BS}(S_0, K, \tau, \sigma) = \Phi(S_0, b(\tau), \beta_1, B_0) - \Phi(S_0, b(\tau), \beta_2, B_0) + A_1 S_0^{\beta_1} - A_2 K^{\beta_2},$$

где

$$\beta_{1,2} = \frac{1}{2} - \frac{r-q}{\sigma^2} \pm \sqrt{\left(\frac{q-r}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}},$$

а коэффициенты A_1, A_2 и граница раннего исполнения $b(t)$ задаются как

$$A_1 = \frac{b(T)^{1-\beta_1}}{\beta_1 - \beta_2}\left(\frac{\beta_2 - 1}{r - q + \frac{1}{2}\sigma^2\beta_2}\right), \quad A_2 = \frac{b(T)^{1-\beta_2}}{\beta_1 - \beta_2}\left(\frac{\beta_1 - 1}{r - q + \frac{1}{2}\sigma^2\beta_1}\right),$$

$$b(t) = B_0 + (B_\infty - B_0)\left(1 - e^{-\gamma(T-t)}\right), \quad B_\infty = \frac{\beta_2}{\beta_2 - 1}K, \quad B_0 = max\left(K, \frac{r}{q}K\right), \quad \gamma = -\frac{2r}{\sigma^2\beta_2(\beta_2 - 1)}.$$

Для извлечения σ_{raw} решается нелинейное уравнение

$$f(\sigma) = C_{BS}(S_0, K, \tau, \sigma) - C_{mkt}(K, \tau) = 0,$$

с помощью итерационного метода Ньютона–Рафсона,

$$\sigma_{n+1} = \sigma_n - \frac{C_{BS}(S_0, K, \tau, \sigma_n) - C_{mkt}(K, \tau)}{\frac{\partial C_{BS}}{\partial \sigma}(S_0, K, \tau, \sigma_n)}, \quad \frac{\partial C_{BS}}{\partial \sigma} \approx \frac{C_{BS}(\sigma + \Delta\sigma) - C_{BS}(\sigma - \Delta\sigma)}{2\Delta\sigma}.$$

Чтобы существенно ускорить процесс, нами предложен итеративный способ расчёта, который использует корреляцию соседних страйков и экспираций. Расчёт начинается с точки at-the-money (ATM), где волатильность наиболее устойчива и легко определяется. Далее итерации распространяются влево и вправо по страйкам:

$$\sigma_0(K_{i+1}, \tau_j) = \sigma_{raw}(K_i, \tau_j), \quad \sigma_0(K_{i-1}, \tau_j) = \sigma_{raw}(K_i, \tau_j),$$

а для соседних экспираций выполняется перекрёстная инициализация:

$$\sigma_{ATM}(\tau_{j+1}) \leftarrow \sigma_{ATM}(\tau_j).$$

Такой способ учитывает гладкость поверхности волатильности и уменьшает количество итераций до 2–3, даже для «плохих» стартовых значений, обеспечивая устойчивую сходимость без расхождений.

Эта схема демонстрирует особую эффективность на больших выборках и при потоковой обработке данных: все точки (K_i, τ_i) обрабатываются последовательно, но при этом каждая итерация использует локальную память предыдущих вычислений, что делает процесс линейным по времени и естественно параллелизуемым.

Благодаря такой оптимизации и использованию аналитической формы Bjerksund–Stensland вместо полного решения PDE, достигается высокая точность и минимальные временные затраты. Полученные значения σ_{raw} формируют плотную и устойчивую сетку исходной IV, служащую надёжной базой для дальнейшей калибровки поверхности в GMM/GMR-модели. Этот подход особенно эффективен для высокочастотных систем и больших наборов данных, где прямое численное решение для каждой точки по отдельности потребовало бы избыточных вычислений.

Gaussian Mixture Models and Regression: калибровка поверхности подразумеваемой волатильности

После получения исходных точек подразумеваемой волатильности формируется обучающая выборка

$$D = \{(K_i, \tau_i, \sigma_i^{raw})\}_{i=1}^N$$

где K — страйк, τ — время до экспирации, σ^{raw} — исходная волатильность.

На этом наборе обучается Gaussian Mixture Model (GMM) с числом компонент $M \approx 15$. Каждая компонента описывает локальную область поверхности, а вся смесь аппроксимирует вероятностную структуру распределения точек (K, τ, σ^{raw}). Исследования показывают, что гауссовые смеси высокой размерности способны гибко описывать сложные формы улыбок и поверхностей, включая асимметричные, многомодальные и осциллирующие зависимости [16].

Совместная плотность смеси записывается как

$p(K, \tau, \sigma) = \sum_{m=1}^M \pi_m N([K, \tau, \sigma]^\top | \mu_m, \Sigma_m)$,

где π_m — вес компоненты под индексом $m \Rightarrow (\Sigma_m \pi_m = 1)$, μ_m — вектор средних, Σ_m — ковариационная матрица.

После обучения GMM применяется Gaussian Mixture Regression (GMR), позволяющая получить условное распределение волатильности при фиксированных (K, τ). Для каждой компоненты m ковариационная структура разбивается на блоки:

$$\Sigma_m = \begin{bmatrix} \Sigma_m^{uu} & \Sigma_m^{u\sigma} \\ \Sigma_m^{\sigma u} & \Sigma_m^{\sigma\sigma} \end{bmatrix}, \qquad \mu_m = \begin{bmatrix} \mu_m^u \\ \mu_m^\sigma \end{bmatrix}, \quad u = [K, \tau]^\top$$

Тогда условное распределение $\sigma \mid u$ выражается как смесь нормальных распределений

$$p(\sigma \mid u) = \sum_{m=1}^M h_m(u)\, N(\sigma \mid \mu_{m|u}(u), \Sigma_{m|u}),$$

где апостериорные веса

$$h_m(u) = \frac{\pi_m\, N(u \mid \mu_m^u, \Sigma_m^{uu})}{\sum_{\ell=1}^M \pi_\ell\, N\left(u \mid \mu_\ell^u, \Sigma_\ell^{uu}\right)},$$

Условные параметры каждой компоненты определяются как

$$\mu_{m|u}(u) = \mu_m^\sigma + \Sigma_m^{\sigma u}(\Sigma_m^{uu})^{-1}(u - \mu_m^u),$$
$$\Sigma_{m|u} = \Sigma_m^{\sigma\sigma} - \Sigma_m^{\sigma u}(\Sigma_m^{uu})^{-1}\Sigma_m^{u\sigma}$$

Из этого распределения вычисляется условное математическое ожидание и дисперсия подразумеваемой волатильности:

$$\sigma_{calib}(K,\tau) = E[\sigma \mid K,\tau] = \sum_{m=1}^{M} h_m([K,\tau]^\top)\, \mu_{m|u}([K,\tau]^\top),$$

$$Var_{calib}(K,\tau) = \sum_{m=1}^{M} h_m([K,\tau]^\top)\big(\Sigma_{m|u} + \mu_{m|u}([K,\tau]^\top)^2\big) - (\sigma_{calib}(K,\tau))^2$$

Экспериментальные результаты

Для эмпирической оценки эффективности предложенного метода проведен ряд экспериментов на исторических данных опционов пяти инструментов из разных секторов экономики — AAPL (технологический сектор), KO (потребительский сектор), JPM (финансовый сектор), XLB (промышленный/сырьевой сектор) и BKNG (транспортно-сервисный сектор). Выбор этих активов обусловлен их разной ликвидностью, уровнем волатильности и характером движения базового актива, что позволяет проверить устойчивость и универсальность предложенного подхода на гетерогенных рыночных данных.

Для каждого актива использовались рыночные котировки опционов со всеми доступными страйками и сроками экспирации от 1 дня до 2-х лет, что охватывает как краткосрочные, так и среднесрочные горизонты. На основе этих данных рассчитывались исходные значения подразумеваемой волатильности σ_{raw} с использованием модели Bjerksund–Stensland, после чего выполнялась кластеризация GMM.

Методика эксперимента

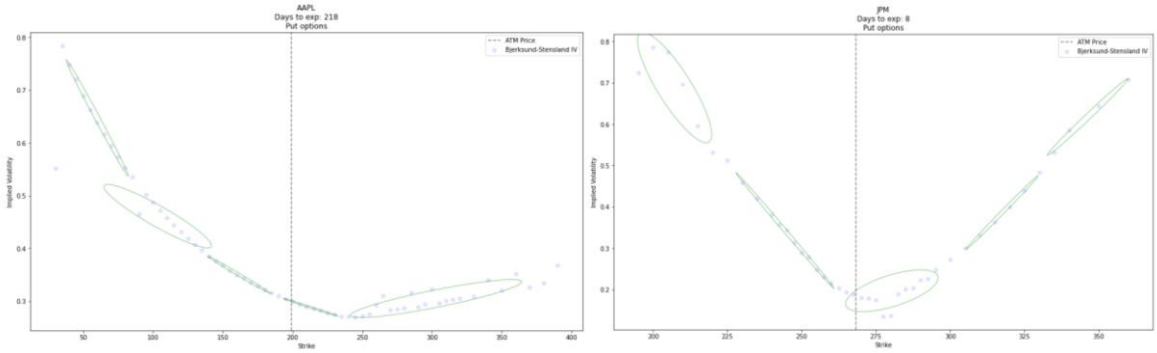

Рисунок 1: Кластеризация наблюдаемых значении подразумеваемой волатильности одной экспирации алгоритмом Gaussian Mixture Models (GMM) для AAPL (218 дней до экспирации) и JPM (8 дней до экспирации)

Далее выполняется калибровка поверхности методом Gaussian Mixture Regression (GMR) с вычислением дисперсии регрессии. Анализ распределения этой дисперсии позволяет выявлять области повышенной неопределенности по страйку и сроку до экспирации — именно там, где наблюдаются наиболее активные рыночные изменения. В перспективе эти зоны можно дополнительно сопоставлять с распределением открытых интересов (open interest) для уточнения локальных центров рыночной активности и подтверждения направлений движения.

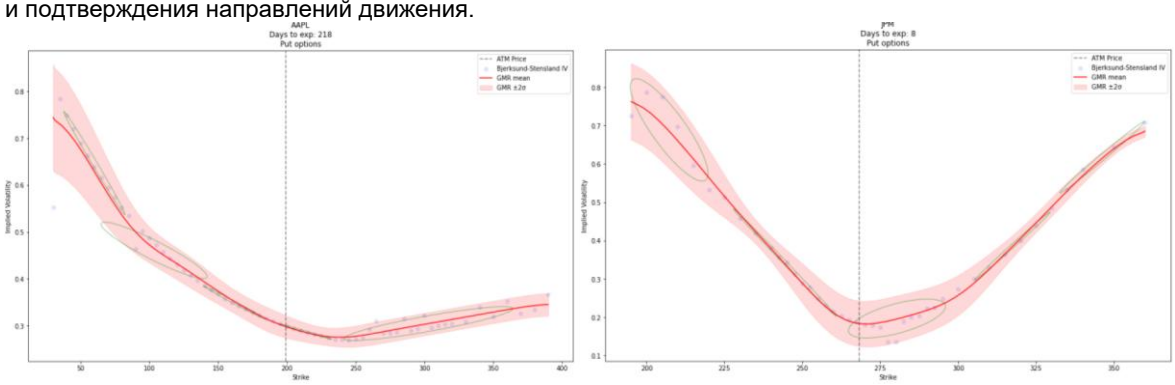

Рисунок 2: Гауссовая регрессия наблюдаемой подразумеваемой волатильности на обученной GMM модели для AAPL (218 дней до экспирации) и JPM (8 дней до экспирации). Показаны средняя линия и зона 2-х стандартных отклонении регрессии.

В трёхмерной постановке моделирования подразумеваемая волатильность σ_{calib} рассматривается как функция двух независимых переменных — страйка K и времени до экспирации τ:

$$\sigma_{calib} = f(K,\tau).$$

После получения исходных точек $\big(K_i, \tau_i, \sigma_{raw,i}\big)$, обученная смесь Гауссов (GMM) описывает их вероятностное распределение в трёхмерном пространстве (K,τ,σ).

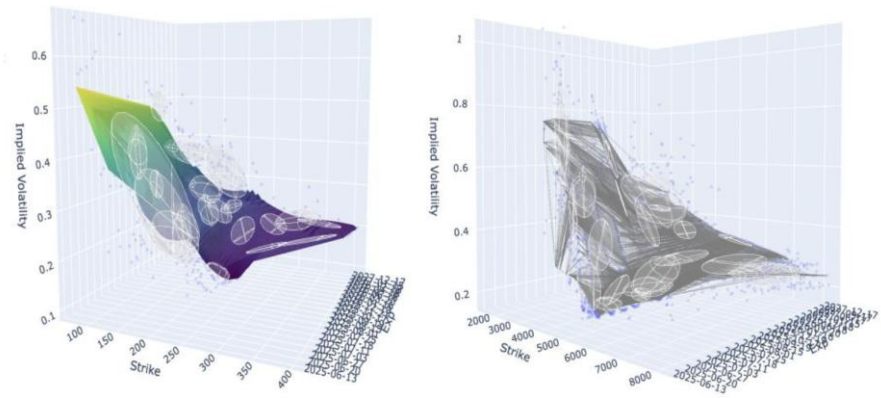

Рисунок 3: Гаусовая поверхность подразумеваемой волатильности и гауссовы смеси для JPM Call опционов и BKNG Put опционов на 12 июня 2025.

Далее с помощью Gaussian Mixture Regression (GMR) вычисляется условное распределение $\sigma|(K,\tau)$. Для каждой пары (K,τ) берётся математическое ожидание этого распределения:

$$\sigma_{calib}(K,\tau) = E[\sigma \mid K,\tau].$$

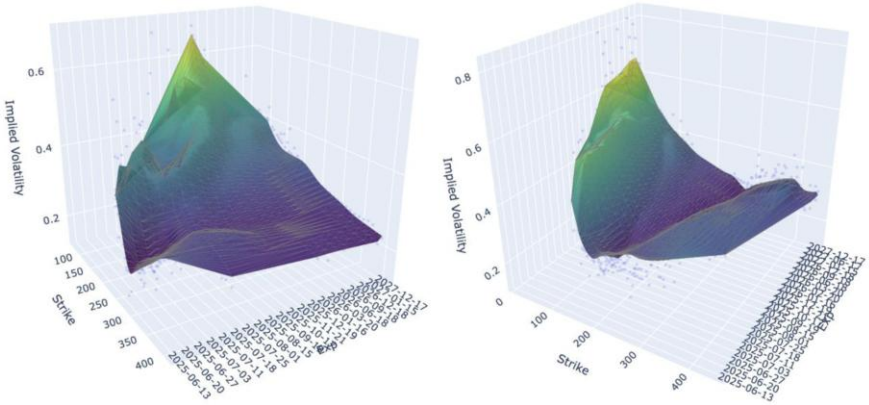

Рисунок 4: Поверхность подразумеваемой волатильности (Implied Volatility Surface) JPM и AAPL Put опционов на 12 июня 2025

Это создаёт непрерывную трёхмерную поверхность implied volatility, где оси K и τ независимы, а по вертикали отображается прогнозная волатильность. Визуально она выглядит как плавная поверхность без разрывов и «ступеней» между сроками — не набор отдельных улыбок, а единая функция, обеспечивающая согласованность во времени и по страйку.

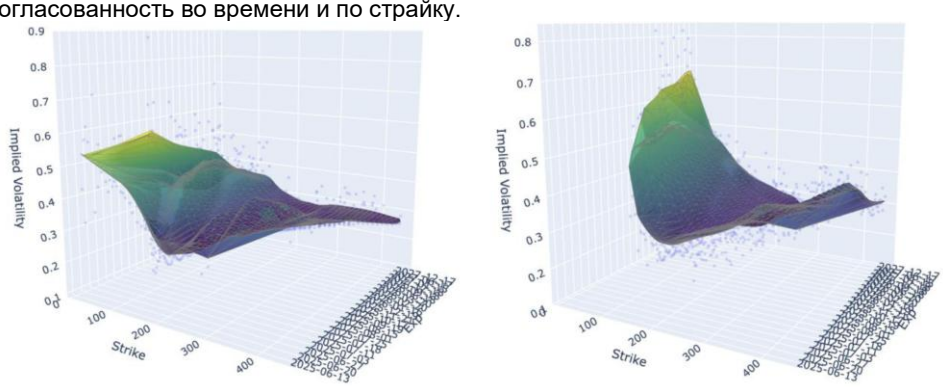

Рисунок 5: Поверхность подразумеваемой волатильности (Implied Volatility Surface) опционов AAPL Call и Put на 12 июня 2025

Такой подход особенно удобен для анализа динамики крыльев и временной структуры волатильности: на одной 3D-поверхности можно сразу видеть, где рынок переоценивает дальние опционы, как кривизна меняется с приближением к экспирации, и как распределяется риск вдоль всей сетки (K,τ). В отличие от параметрических моделей, где каждая экспирация моделируется отдельно, здесь поверхность восстанавливается целиком — с учётом межвременных зависимостей и плавности между сроками.

Погрешности и дисперсионный анализ

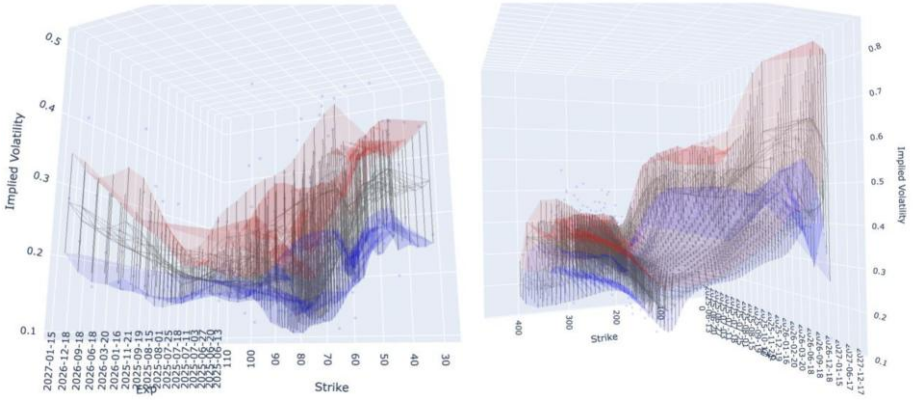

Рисунок 6: Распределение отклонения регрессии для KO Call и AAPL Put опционов

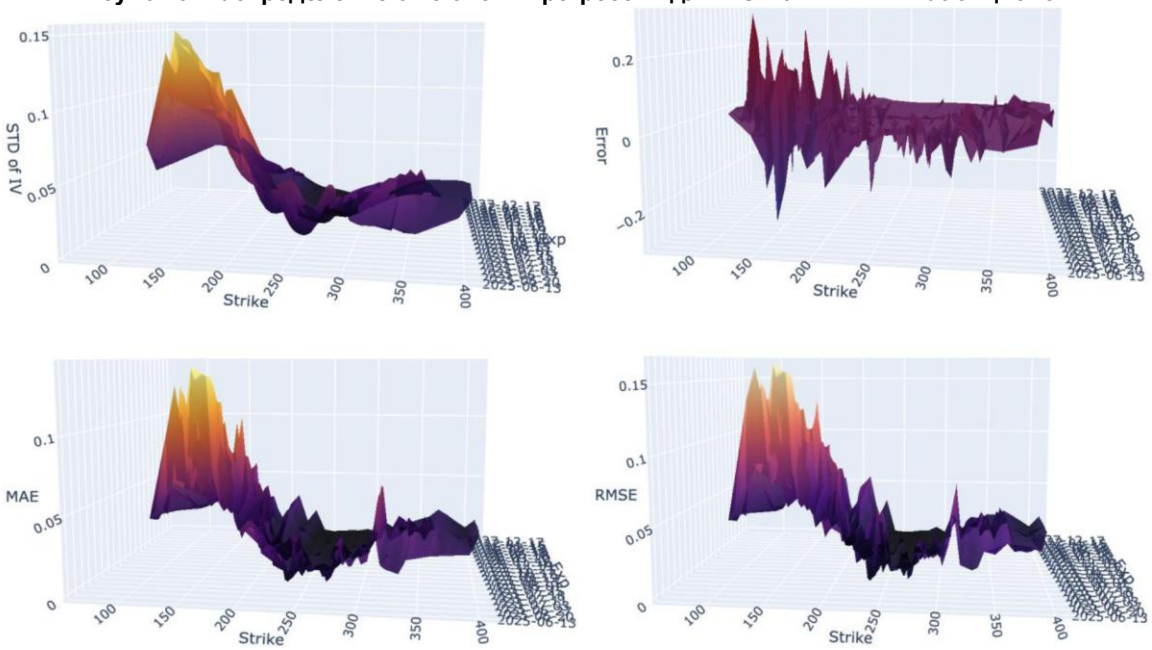

Рисунок 7: Стандартное отклонение регрессии, ошибка (наблюдение – регрессия), Mean Absolute Error (MAE) и Root Mean Square Error (RMSE) для JPM и KO Put опционов на 12 июня 2025

Таблица 1. Ошибки аппроксимации модели для разных активов и типов опционов

Актив	Тип опциона	MAE	RMSE	rror Mean
AAPL	Call	0.02609	0.03310	0.00047
	Put	0.02635	0.03336	-0.00011
KO	Call	0.03914	0.04906	0.00054
	Put	0.00427	0.05422	0.00140
JPM	Call	0.03416	0.04338	0.00289
	Put	0.02810	0.03692	-0.00008
XLB	Call	0.02667	0.03311	0.00111
	Put	0.02248	0.02816	-0.00213
BKNG	Call	0.02345	0.03248	-0.00099
	Put	0.02377	0.03302	-0.00096

$$MAE_{avg} = 0.02545,\ \ RMSE_{avg} = 0.0376, Error\ Average = 0.00021.$$

MAE_{avg} = 0.02545 и $RMSE_{avg}$ = 0.0376 означают, что средняя ошибка аппроксимации составляет 2.5–3.7 % от масштаба волатильности. Это — очень высокая точность для моделей IV-поверхностей, особенно без обучения нейросети. Такой уровень ошибки показывает, что модель корректно

восстанавливает реальную геометрию улыбки и временной структуры, при этом оставаясь стабильной на всех страйках и сроках.

Средний Error Average = 0.00021 (≈ 0.02 %) дополнительно подтверждает отсутствие систематического смещения: положительные и отрицательные ошибки взаимно компенсируются, что говорит о нейтральной генерализации — модель не переоценивает и не недооценивает волатильность в среднем.

Сравнительный анализ

В большинстве опубликованных исследований для оценки моделей поверхности подразумеваемой волатильности применяются индексные данные — например, SPX, STOXX 50, NIKKEI 225 и другие высоколиквидные индексы. Эти данные отличаются стабильностью и плотностью страйков, что делает задачу сглаживания и аппроксимации существенно проще.

В нашем случае подход был протестирован в более сложных условиях: мы провели анализ на гетерогенном наборе активов — от технологических и потребительских акций (AAPL, KO) до финансовых (JPM), сырьевых (XLB) и сервисных компаний (BKNG), представляющих сектор онлайн-туризма и бронирования. Это позволило проверить устойчивость модели при разных профилях ликвидности, дисперсии волатильности и неровной структуре страйков.

Таблица 2. Сравнение средних точностей нейронных моделей и нашего метода.

Модель	Год	Тип подхода	MAE	RMSE
GMM + GMR (предлагаемая)	2025	Непараметрическая, без обучения, no-arbitrage	0.025	0.038
Neural Correction Model (Grith) [19]	2023	Нейросетевая коррекция параметрических моделей	0.030	0.055
Deep Volatility Network (Stacy) [20]	2023	LSTM / GRU / CNN	0.056 – 0.114	0.060 – 0.120
GAN-Enhanced IV Reconstruction [21]	2025	GAN (генеративная модель)	0.039	0.049
Two-Step Arbitrage-Free DNN (Zhang et al.) [22]	2021	DNN + PCA / VAE (двухшаговый фреймворк)	≈ 0.040	≈ 0.050

Таким образом, хотя сопоставление результатов с другими работами носит ориентировочный характер, по порядку величин ошибок наша модель показывает себя на уровне лучших опубликованных нейросетевых решений. Более того, тот факт, что она сохраняет стабильную точность даже на сложных активах, демонстрирует её робастность и реальную практическую применимость.

Выводы

Предложенный подход, основанный на комбинации Gaussian Mixture Models (GMM) и Gaussian Mixture Regression (GMR), продемонстрировал, что формонезависимое, универсальное моделирование поверхности подразумеваемой волатильности может быть реализовано без компромисса между точностью и вычислительной эффективностью. Модель достигает средней ошибки менее 4 % при практически нулевом смещении, что соответствует уровню state-of-the-art методов, при этом полностью исключая необходимость в глубоком обучении и сложной оптимизации. Такой результат особенно значим, учитывая, что большинство существующих исследований опирается на относительно «лёгкие» индексные данные (SPX, STOXX 50 и т. д.), тогда как наш метод был протестирован на разнообразном наборе активов с разной структурой страйков, ликвидностью и сроками до экспирации — от технологических акций до финансовых и сырьевых инструментов.

Эта универсальность подтверждает высокую робастность модели, её способность сохранять гладкость и точность даже при значительных изменениях рыночных условий, а также отсутствие переобучения на локальных областях поверхности. Кроме того, GMM/GMR-подход обладает естественной интерпретируемостью: каждая компонентная плотность имеет статистическое значение, а регрессионная структура позволяет получать не только центральную оценку, но и условную дисперсию, отражающую уровень неопределенности по страйкам и срокам. Это превращает модель в надёжный инструмент количественного анализа, способный не только восстанавливать структуру implied volatility, но и выявлять зоны рыночной напряжённости, где сосредоточена ключевая динамика. В совокупности эти свойства делают предложенный метод эффективным, воспроизводимым и практически применимым решением для задач ценообразования, хеджирования и риск-менеджмента в современных финансовых системах.

Литература

1. Homescu C. Implied volatility surface: construction methodologies and characteristics. arXiv preprint arXiv:1107.1834. 2011.

2. Chan DK-C, Fong W-M, Lam KF, Yung K. Informed Option Trading and Term Structure. Auckland University of Technology — Auckland Centre for Financial Research. 2019.
3. Park H, Kim B, Kim D. Informed options trading on the implied volatility surface: a cross-sectional approach. Auckland (NZ): Auckland University of Technology, Auckland Centre for Financial Research; 2019.
4. Baldacci B. High-frequency dynamics of the implied volatility surface. arXiv preprint arXiv:2012.10875, 2020.
5. Zhang Y, Li J, Zhang W. A Two-Step Framework for Arbitrage-Free Prediction of the Implied Volatility Surface. arXiv preprint arXiv:2106.07177; 2021.
6. Ulrich M., Zimmer L., Merbecks C. *Implied volatility surfaces: a comprehensive analysis using half a billion option prices.* Review of Derivatives Research. 2023;26(2-3):135–169. DOI: 10.1007/s11147-023-09195-5.
7. Laurini MP. Imposing no-arbitrage conditions in implied volatility surfaces using constrained smoothing splines. *Applied Stochastic Models in Business and Industry*. 2011;27(6):649-659. DOI: 10.1002/asmb.877.
8. Gatheral J, Jacquier A. Arbitrage-free SVI volatility surfaces. Quantitative Finance. 2014;14(1):59-71. doi:10.1080/14697688.2013.819986.
9. Hagan P., Kumar D., Lesniewski A., Woodward D. Managing Smile Risk. Wilmott Magazine. 2002;(84):84-108. DOI:10.1007/s10203-007-0091-4.
10. Borak S, Fengler M, Härdle W. DSFM fitting of implied volatility surfaces. Discussion Paper SFB 649, Humboldt-Universität zu Berlin; 2005.
11. Ackerer D., Tagasovska N., Vatter T. Deep Smoothing of the Implied Volatility Surface. NeurIPS 2020 Conference Paper; 2020.
12. Mijatović A, Tankov P. A new look at short-term implied volatility in asset price models with jumps. arXiv preprint arXiv:1207.0843; 2012.
13. Driessen JP, Jorgensen BJ. The effects of uncertainty and liquidity on the cross-section of stock returns. Columbia Business School — Research Paper; 2006.
14. Fengler M. Arbitrage-free smoothing of the implied volatility surface. SFB 649 Discussion Paper No. 2005-019, Humboldt-Universität zu Berlin; 2005.
15. Alexiou L., Goyal A., Kostakis A., Rompolis L. Pricing Event Risk: Evidence from Concave Implied Volatility Curves. *Derivatives & Risk Management Meeting 2022*; 2022.
16. Glasserman P, Pirjol D. W-Shaped Implied Volatility Curves and the Gaussian Mixture Model. SSRN Electron J. 2021;(Paper 3951426).
17. Keller-Ressel M. W-shaped implied volatility curves in a variance-gamma mixture model. arXiv preprint arXiv:2209.14726; 2022.
18. De Marco S., Hillairet C., Jacquier A. Shapes of implied volatility with positive mass at zero. arXiv preprint arXiv:1310.1020; 2013.
19. van der Meer FK. Modeling equity option implied volatility surfaces using non-parametrically corrected parametric models. Rotterdam: Erasmus University Rotterdam, Erasmus School of Economics; 2024.
20. Stacy P. Machine-Learning Based Modeling of the Implied Volatility Surface. Rotterdam: Erasmus University Rotterdam, Erasmus School of Economics; 2023.
21. Ge Y, Wang Y, Liu J, Wang J. GAN-Enhanced Implied Volatility Surface Reconstruction for Option Pricing Error Mitigation. *IEEE Access*. 2025; eISSN 2169-3536. doi:10.1109/ACCESS.2025.3619553.
22. Zhang W, Li L, Zhang G. A Two-Step Framework for Arbitrage-Free Prediction of the Implied Volatility Surface. *Quantitative Finance*. 2023;23(1):21-34. doi:10.1080/14697688.2022.2135454.

Ключевые слова

Поверхность подразумеваемой волатильности, формонезависимое моделирование, модель гауссовых смесей, гауссовая регрессия смесей, модель Бьерксунда–Стенсланда, непараметрическая калибровка, арбитражно-свободная волатильность, ценообразование опционов, количественные финансы, динамика улыбки волатильности, высокочастотное моделирование, управление рисками, эффективная аппроксимация поверхности

Садуанов Батырхан, Араратян Севак, Меркулов Вадим,
Freedom Broker, г. Алматы, Казахстан

Key words

Implied volatility surface, shape-agnostic modeling, Gaussian Mixture Model, Gaussian Mixture Regression, Bjerksund–Stensland model, non-parametric calibration, arbitrage-free volatility, option pricing, quantitative finance, volatility smile dynamics, high-frequency modeling, risk management, efficient surface fitting

Batyrkhan Saduanov, Sevak Araratyan, Vadim Merkulov, Form-independent modeling of the implied volatility surface based on Gaussian mixtures

DOI: 10.34706/DE-2025-05-04
JEL classification: F36 Финансовые аспекты экономической интеграции.

Abstract

This paper introduces a new method for modeling the implied volatility surface based on the combination of Gaussian Mixture Models (GMM) and Gaussian Mixture Regression (GMR). Unlike traditional approaches that assume a fixed functional form of the volatility surface, the proposed framework is shape-agnostic and can flexibly capture the complex volatility dynamics observed across different asset classes, including a wide range of smile shapes — from classical U-shaped to multimodal W-shaped and other non-standard patterns observed in real trading environments. By leveraging the probabilistic expressiveness of GMMs and the predictive power of regression, the method achieves accuracy comparable to deep learning models while outperforming them in computational efficiency, requiring no training phase. Empirical results demonstrate that the proposed approach surpasses existing parametric models in flexibility and robustness, providing accuracy on par with neural-network-based solutions but with significantly lower computational cost. Due to its lightness, scalability, and interpretability, the method is particularly well-suited for high-frequency financial applications, offering practitioners and researchers a reliable and efficient tool for volatility modeling and risk management.