

СИСТЕМА СТАТИСТИЧЕСКИХ ТЕСТОВ ДЛЯ БЫСТРОГО ПРОТОТИПИРОВАНИЯ МАКРОЭКОНОМИЧЕСКИХ РЕГРЕССИОННЫХ МОДЕЛЕЙ С ДЛИТЕЛЬНЫМ ГОРИЗОНТОМ ПРОГНОЗИРОВАНИЯ

А. С. Подгорный

Дальневосточный федеральный университет, г. Владивосток, Россия

E-mail: podgorny_as@dvfu.ru

В статье представлена программная система на языке Python для быстрого прототипирования и валидации макроэкономических регрессионных моделей, предназначенных для долгосрочного сценарного прогнозирования. При генерации прогнозов на 5–7 лет вперёд при ограниченной глубине обучающих данных (по январь 2025 г.) классические метрики качества (MAE, MSE, MAPE) становятся ненадёжными критериями оптимизации, сохраняя при этом диагностическую ценность для выявления грубых ошибок спецификации. В исследовании предлагается иерархическая методология валидации: экономическая интерпретируемость, статистическое тестирование свойств остатков, согласование с экспертными прогнозами. Разработанный модуль реализует автоматизированную генерацию признаков (лаги, разности, скользящие средние), интегрированную с комплексным набором статистических тестов. Применяются дифференцированные пороги значимости: строгие ($\alpha=0.05$) для тестов автокорреляции, критичных для накопления ошибок, и мягкие ($\alpha=0.10$) для тестов нормальности. Согласованность в VAR-системах верифицируется через анализ импульсных откликов. Результаты применения на российских макроэкономических данных демонстрируют практическую полезность фреймворка. Ретроспективная валидация показывает, что статистически отобранные модели демонстрируют меньшее накопление ошибок на длинных горизонтах по сравнению с MAPE-оптимизированными альтернативами.

SYSTEM OF STATISTICAL TESTS FOR FAST PROTOTYPING OF MACROECONOMIC REGRESSION MODELS WITH A LONG HORIZON OF FORECASTING

A. S. Podgorny

The article presents a Python-based software system for rapid prototyping and validation of macroeconomic regression models designed for long-term scenario forecasting. When generating forecasts for 5-7 years ahead with limited training data depth (up to January 2025), classical quality metrics (MAE, MSE, MAPE) become unreliable optimization criteria, while retaining diagnostic value for identifying severe specification errors. The study proposes a hierarchical validation methodology: economic interpretability, statistical testing of residual properties, and alignment with expert forecasts. The developed module implements automated feature generation (lags, differences, and moving averages) integrated with a comprehensive set of statistical tests. Differentiated significance thresholds are applied: strict ($\alpha=0.05$) for autocorrelation tests critical for error accumulation, and soft ($\alpha=0.10$) for tests

Введение

Макроэкономическое моделирование для долгосрочного прогнозирования представляет уникальные методологические вызовы, отличающие его от традиционной предиктивной аналитики. В задачах проектного финансирования, оценки рисков и стратегического планирования горизонты прогнозирования часто составляют 5–7 лет [1, 2]. Такие протяжённые горизонты фундаментально меняют парадигму валидации: традиционные метрики точности, вычисленные на исторических тестовых выборках, становятся ненадёжными критериями оптимизации, когда период прогноза существенно превышает доступную историю обучения.

Ключевая проблема возникает из фундаментальной асимметрии: имея 10–15 лет надёжных макроэкономических данных, мы должны генерировать сценарии на 5–7 лет вперёд. В этих условиях такие метрики, как MAE, MSE и MAPE, вычисленные на обучающей или ограниченной тестовой выборке, дают слабые гарантии относительно долгосрочного поведения модели. Модели, оптимизированные по этим метрикам, могут демонстрировать взрывную динамику, нереалистичные равновесные значения или несогласованные связи между переменными на длинных горизонтах [3].

Данное исследование отвечает на потребность в альтернативной системе валидации, основанной на статистических свойствах остатков регрессий и согласованности модели в рамках векторных авторегрессионных (VAR) систем. Предлагаемый подход ставит робастность спецификации выше точности точечного прогноза, позволяя выявлять модели, сохраняющие экономическую интерпретируемость и статистическую обоснованность на длительных горизонтах прогнозирования.

Цель исследования и иерархия валидации

Основная цель – разработка программного фреймворка для статистической валидации макроэкономических регрессионных моделей, предназначенных для генерации долгосрочных сценариев. Методология использует иерархический подход к валидации с тремя уровнями приоритета.

Уровень 1 – Экономическая интерпретируемость: знаки коэффициентов и направления влияния должны соответствовать экономической теории. Спецификация, в которой рост ключевой ставки ведёт к росту цен на жильё, отвергается независимо от статистических свойств.

Уровень 2 – Статистические тесты: диагностика остатков должна подтверждать, что предпосылки модели выполняются в достаточной степени для надёжной экстраполяции. Спецификации, не прошедшие критические тесты, подлежат модификации.

Уровень 3 – Согласование с экспертными прогнозами: выходы модели должны быть согласованы с прогнозами Центрального банка и консенсус-прогнозами, выступая как проверка здравого смысла, а не как цель оптимизации.

Данная иерархия разрешает потенциальные конфликты: если статистические тесты указывают на нестационарность, но модификация спецификации нарушит экономическую интерпретируемость, мы фиксируем ограничение и

применяем модель с явными границами неопределённости, а не принимаем экономически неправдоподобную альтернативу.

Диагностическая роль классических метрик

Хотя классические метрики непригодны как критерии оптимизации для долгосрочного прогнозирования, они сохраняют важную диагностическую ценность. Фреймворк вычисляет MAE, MSE и MAPE не для выбора модели, а как «санитарный фильтр» для выявления грубых ошибок спецификации. Модель с приемлемыми результатами статистических тестов, но катастрофически высоким MAE на обучении, вероятно, содержит ошибки реализации или обработки данных.

Ключевое различие — между оптимизацией и диагностикой: мы не выбираем модели, минимизирующие MAPE, но отвергаем модели, где MAPE превышает правдоподобные границы. Такое асимметричное использование метрик отражает фундаментальное понимание: низкая ошибка на исторических данных не является ни необходимым, ни достаточным условием хорошего долгосрочного прогноза, тогда как экстремально высокая ошибка достаточна для отклонения.

Ретроспективный анализ подтверждает этот подход: модели, отобранные по статистическим критериям в 2015 году, сравнивались с MAPE-оптимизированными альтернативами за период 2015–2024. Статистически отобранные модели показали меньшее накопление ошибок, особенно в периоды структурных сдвигов (2020, 2022), когда переобученные модели генерировали экстремальные прогнозы.

Батарея статистических тестов

1. Тесты нормальности. Тесты Шапиро–Уилка и Харке–Бера выявляют отклонения распределения остатков от гауссовского. Хотя ненормальность влияет на доверительные интервалы, а не на точечные прогнозы, она может указывать на пропущенные нелинейности. Применяется мягкий порог $\alpha = 0.10$, поскольку влияние на долгосрочное прогнозирование опосредовано.

2. Тесты гетероскедастичности. Тесты Бройша–Пагана, Голдфелда–Квандта, Уайта и Глейзера проверяют постоянство дисперсии ошибок. Связь с долгосрочным прогнозом здесь прямая: если дисперсия зависит от уровня переменных, при выходе прогноза за исторические диапазоны доверительные интервалы становятся либо опасно узкими, либо взрывными. Применяется порог $\alpha = 0.05$.

3. Тесты автокорреляции. Тесты Дарбина–Уотсона, Бройша–Годфри и Льюнга–Бокса наиболее критичны для сценарного моделирования. При итеративном прогнозировании каждая ошибка становится входом для следующего периода. Если ошибки автокоррелированы, они накапливаются систематически, а не компенсируют друг друга. Применяется строгий порог $\alpha = 0.05$; статистика Дарбина–Уотсона вне диапазона 1.5–2.5 требует пересмотра спецификации.

4. Тест стационарности. Расширенный тест Дики–Фуллера (ADF) применяется к остаткам для проверки стационарности. Нестационарные остатки означают, что ошибки могут трендить или случайно блуждать, делая многолетние прогнозы бессмысленными. Если ADF не отвергает единичный корень, сначала

пробуем модификацию спецификации (взятие разностей, альтернативная структура лагов).

Проверка согласованности VAR-системы

Тестирование отдельных уравнений, будучи необходимым, недостаточно для обеспечения системной согласованности. Фреймворк реализует анализ импульсных откликов для проверки межуравненной динамики: шок одной переменной (например, +100 б.п. к ключевой ставке) распространяется через систему, и траектории откликов проверяются на экономическую правдоподобность [4].

Критерии проверки включают: (1) правильные знаки откликов — рост ставки должен снижать спрос на жильё и цены; (2) правдоподобный тайминг — корректировка ипотечной ставки должна предшествовать реакции цен на жильё; (3) возврат к равновесию — шоки не должны порождать перманентные сдвиги уровней в стационарных переменных; (4) разумная амплитуда — отклики должны укладываться в исторически наблюдавшиеся диапазоны.

Спецификации, прошедшие индивидуальные статистические тесты, но порождающие неправдоподобные импульсные отклики, отвергаются. Этот критерий операционализирует требование экономической интерпретируемости на системном уровне, обеспечивая генерацию внутренне согласованных сценариев.

Программная реализация

Фреймворк статистического тестирования реализован на Python (исходный код: https://github.com/andreypodgorny10/macro_sims). Модуль принимает спецификации моделей в виде словарей, отображающих зависимые переменные на пары «признак–коэффициент», поддерживая декларативное определение моделей и быстрое переключение между спецификациями.

1. Автоматизированная генерация признаков. Имена признаков кодируют метаданные преобразований через суффиксы: $_Lk$ для лага k , $_Dk$ для разности порядка k , $_Mak$ для скользящего среднего с окном k . Спредовые признаки (например, $MR_spread_key_rate$ для спреда ипотечной ставки над ключевой) вычисляются динамически.

2. Оценка и управление коэффициентами. Спецификации содержат преоценённые OLS-коэффициенты, обеспечивая быстрое прототипирование без повторной оценки. Метод `calculate_stats` переоценивает модели на текущих данных, сохраняя обновлённые коэффициенты. Такой двойственный подход поддерживает как воспроизводимую генерацию сценариев (фиксированные коэффициенты), так и мониторинг моделей (выявление дрейфа коэффициентов при поступлении новых данных).

3. Обработка структурных разрывов через DUMMY-переменные. DUMMY-переменные (например, $DUMMY_{20+}$ для периода COVID, $DUMMY_{22+}$ для санкционного шока 2022) изолируют аномальные периоды при оценке коэффициентов. Критически важно: эти переменные обнуляются при генерации прогнозов — мы проецируем динамику «нормального режима». Будущие кризисные сценарии моделируются через стохастические шоки в симуляционном слое CIR++ [5, 6], а не через детерминированную активацию DUMMY, сохраняя вероятностную природу долгосрочной неопределённости.

Применение к российским макроэкономическим данным

Фреймворк применялся к российским макроэкономическим показателям: инфляция ИПЦ, рост реального ВВП, ключевая ставка ЦБ, ипотечные ставки, индексы цен на жильё (первичный и вторичный рынки), строительные затраты, ставки RUONIA и показатели NPL для розничного, корпоративного и финансового секторов. Данные охватывают период 2007 — январь 2025 с месячной периодичностью.

Итеративное уточнение на основе статистических тестов позволило получить набор спецификаций со следующими характеристиками: статистика Дарбина–Уотсона в диапазоне 1.5–2.5 для всех уравнений; p-value теста Льюнга–Бокса превышает 0.10 на лаге 10; ADF отвергает единичный корень ($p < 0.05$) в остатках для 12 из 14 уравнений. Два уравнения (NPL_corp и cost_m_yoy) демонстрируют пограничную стационарность и отмечены для усиленного мониторинга.

Проверка импульсных откликов подтвердила правдоподобную динамику: шок ключевой ставки +200 б.п. порождает рост ипотечной ставки в течение 1–2 месяцев, замедление роста цен на жильё в течение 3–6 месяцев и рост NPL по розничным кредитам в течение 6–12 месяцев — что согласуется с таймингом трансмиссионного механизма, наблюдавшимся в исторических эпизодах.

Обсуждение

Переход от метрик точности к статистическому тестированию отражает фундаментальную переориентацию, адекватную задачам долгосрочного прогнозирования. Вместо вопроса «Насколько точно модель предсказывает следующее наблюдение?» методология спрашивает: «Порождает ли спецификация хорошо себя ведущие остатки, поддерживающие надёжную экстраполяцию?» Это согласуется с требованиями генерации сценариев, где правдоподобные, внутренне согласованные траектории важнее точности точечного прогноза [7].

Дифференцированные пороги значимости признают, что не все нарушения предпосылок одинаково разрушительны. Автокорреляция напрямую накапливается при итеративном прогнозировании, требуя строгого контроля. Умеренная ненормальность влияет на доверительные интервалы, но не на точечные прогнозы, допуская мягкие пороги. Такой нюансированный подход избегает как избыточного отклонения полезных моделей, так и принятия фундаментально порочных спецификаций [8].

Иерархическая структура валидации даёт чёткие ориентиры при конфликте критериев. Экономическая интерпретируемость как первичный фильтр предотвращает статистически элегантные, но экономически бессмысленные спецификации. Статистическое тестирование как вторичный фильтр обеспечивает надёжность экстраполяции. Согласование с экспертными прогнозами как третичная проверка защищает от систематического смещения, не ограничивая гибкость модели.

Заключение

В данном исследовании представлен Python-фреймворк для статистической валидации макроэкономических регрессионных моделей, предназначенных

для долгосрочного сценарного прогнозирования. Ключевые вклады включают: (1) иерархическую методологию валидации, приоритизирующую экономическую интерпретируемость, статистические свойства и согласование с экспертами; (2) дифференцированные пороги значимости, отражающие относительную важность нарушений предпосылок; (3) интеграцию анализа импульсных откликов для проверки согласованности VAR-системы; (4) ретроспективное свидетельство в пользу статистического отбора по сравнению с оптимизацией по метрикам.

Модульная архитектура поддерживает итеративное уточнение и будущую разработку библиотеки. Планируемые расширения включают автоматизированный поиск спецификаций под управлением статистических критериев, расширенное тестирование коинтеграции для нестационарных систем и интеграцию с фреймворками стохастических симуляций для генерации вероятностных сценариев.

Результаты демонстрируют, что строгое статистическое тестирование в рамках принципиальной иерархической структуры обеспечивает надёжную основу для выбора моделей в контекстах, где традиционные метрики точности неадекватны. Приоритизируя диагностику остатков и системную согласованность над минимизацией ошибки прогноза, практики могут выявлять спецификации, робастные к требованиям экстраполяции долгосрочного макроэкономического прогнозирования.

СПИСОК ЛИТЕРАТУРЫ

1. *Bekaert G., Engstrom E., Xu N. R.* The time variation in risk appetite and uncertainty // *Management Science*. 2022. Vol. 68. Iss. 6. P. 3975-4004.
2. *Fernandez-Villaverde J., Guerron-Quintana P. A.* Uncertainty shocks and business cycle research // *Review of Economic Dynamics*. 2020. Vol. 37. P. 118-146.
3. *Antolin-Díaz J., Petrella I., Rubio-Ramírez J. F.* Structural scenario analysis with SVARs // *Journal of Monetary Economics*. 2021. Vol. 117. P. 798-815.
4. *Lütkepohl H.* *New Introduction to Multiple Time Series Analysis*. Berlin: Springer, 2005. 764 p.
5. *Akhtari B., Li H.* The Cox-Ingersoll-Ross process under volatility uncertainty // *Journal of Mathematical Analysis and Applications*. 2024. Vol. 531. Iss. 1. Part 1. P. 2-31.
6. *Chen R.-R., Scott L.* Multi-Factor Cox-Ingersoll-Ross models of the term structure: estimates and tests from a Kalman filter model // *The Journal of Real Estate Finance and Economics*. 2003. Vol. 27. P. 143-172.
7. *Stock J. H., Watson M. W.* Vector Autoregressions // *Journal of Economic Perspectives*. 2001. Vol. 15. No. 4. P. 101-115.
8. *Wooldridge J. M.* *Introductory Econometrics: A Modern Approach*. 7th ed. Boston: Cengage Learning. 2020. 848 p.