Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет» (СПбГУ)

№ договора ЕД-135482/Ф от 29.07.2024

СОГЛАСОВАНО
Председатель научной комиссии
в области экономики и бизнеса
_____Ж. В. Писаренко

УТВЕРЖДАЮ
И.о. Проректор по научной работе Е.В. Лебедева
_____ С.В. Микушев
«18» декабря 2024 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ, выполненной
в соответствии с техническим заданием по договору
№ ЕД-135482/Ф от 29.07.2024

по теме проекта
«УСОВЕРШЕНСТВОВАНИЕ МЕТОДОЛОГИИ РАННЕГО ПРЕДУПРЕЖДЕНИЯ ДЛЯ ФИНАНСОВЫХ РЫНКОВ С АПРОБАЦИЕЙ НА ФИНАНСОВЫХ РЫНКАХ СТРАН С ФОРМИРУЮЩЕЙСЯ РЫНОЧНОЙ ЭКОНОМИКОЙ»

Руководитель проекта Л.В. Гадасина

Исполнитель А.Б. Прохоров

Санкт-Петербург

2024

Реферат

Отчет 28 с., 7 ч., 5 рис., 5 табл., 46 источников.

СОСТАВНАЯ ЧАСТЬ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ ПРОЕКТА «УСОВЕРШЕНСТВОВАНИЕ МЕТОДОЛОГИИ РАННЕГО ПРЕДУПРЕЖДЕНИЯ ДЛЯ ФИНАНСОВЫХ РЫНКОВ С АПРОБАЦИЕЙ НА ФИНАНСОВЫХ РЫНКАХ СТРАН С ФОРМИРУЮЩЕЙСЯ РЫНОЧНОЙ ЭКОНОМИКОЙ»

Цель работы - разработка и применение новой онлайн-системы раннего предупреждения (EWS) о том, что в машинном обучении известно как дрейф концепций, в экономике - как смена режима, а в статистике - как точка изменения. Система выходит за рамки линейности, предполагаемой во многих традиционных методах, и устойчива к тяжелохвостности и зависимостям данных, что делает ее особенно подходящей для развивающихся рынков. Ключевым компонентом является эффективный механизм определения точки изменения для условной энтропии данных, а не для конкретного интересующего показателя. В сочетании с последними достижениями в области методов машинного обучения для многомерных случайных лесов этот механизм способен обнаруживать значительные сдвиги в передаче информации между взаимозависимыми временными рядами, когда традиционные методы терпят неудачу. В работе проводится исследование, когда это произойдет, используя моделирование, и приводится иллюстрация применения метода к товарному и фондовому рынкам Узбекистана.

Ключевые слова: раннее предупреждение, связка, локальный случайный лес, машинное обучение, статистика Ширяева-Робертса, побочные эффекты

Содержание

1 Введение

Финансовые рынки развивающихся экономик характеризуются необычайно высокой волатильностью, структурными разрывами, нелинейной зависимостью, асимметрией и сильной корреляцией в экстремумах, также известных как хвостовая зависимость. Есть свидетельства того, что эти стилизованные факты особенно ярко проявляются на развивающихся рынках (см., например, Chaudhuri and Wu, 2003; Chen and Ibragimov, 2019). Более того, современные финансовые рынки, как правило, имеют сложную структуру информационных потоков, часто представленных различными глобальными и локальными побочными эффектами и распространением, затрагивающими развивающиеся рынки и проявляющимися значительными задержками и затруднениями в передаче информации (см., например, Zhao et al., 2022; Khalfaoui et al., 2023; Ahmad et al., 2018). В этих условиях быстрое и надежное обнаружение изменений в информационных каналах, действующих на развивающихся рынках, является ключом к обеспечению стабильности рынков, эффективной экономической политике и предотвращению финансового заражения (см., например, Neaime, 2016; Smimou and Khallouli, 2015; Yang et al., 2023).

С точки зрения статистики область, разрабатывающая принципиальные способы построения систем раннего предупреждения (EWS), которые реагируют на изменения в случайном процессе с наименьшей задержкой, известна как обнаружение точки изменения. Эта литература восходит к Ширяеву (1961, 1963) и Roberts (1966), среди прочих, и к настоящему времени накоплен впечатляющий набор результатов оптимальности (см., например, Chen, 2019; Chu and Chen, 2019; Pergamenchtchikov and Tartakovsky, 2018, 2019). Важно отметить, что различные версии так называемой статистики Ширяева-Робертса (SR) оказались оптимальными с точки зрения минимизации задержки обнаружения, поскольку вероятность ложной тревоги приближается к нулю, для широкого спектра случайных процессов, включая автокоррелированные временные ряды финансовых доходов с неизвестным распределением до и после изменений. Это подстегнуло разработку онлайн-систем финансового надзора, основанных на статистике SR, где “онлайн” означает в режиме реального времени (см., например, Pepelyshev and Polunchenko, 2017).

Использование этих методологий в экономических приложениях и, в частности, для развивающихся рынков отстает. В эконометрике большая часть так называемой методологии структурных разрывов была “оффлайн”, то есть она использует исторические временные ряды для обнаружения изменений режима (см., например, Hamilton, 2016; Bai and Perron, 1998, 2003). Основное внимание было уделено получению последовательных тестов для определения количества и местоположения изменений режима в доступных данных, а не последовательному мониторингу точек изменений. Традиционная эконометрическая методология часто предполагает, что количество смен режимов должно быть известно или находиться в пределах фиксированного региона. Сравнение эффективности онлайн- и офлайн-методов требует пересмотра офлайн-методов, чтобы сделать их применимыми в онлайн-среде, для которой они не были разработаны. Поэтому в литературе мало рекомендаций по выбору между этими методами. Более того, традиционно экономика фокусировалась на линейных моделях передачи информации, представленных в таких концепциях и моделях, как Granger (1969) модели причинности, векторной авторегрессии (VAR) и динамической условной ковариации (DCC), которые

используют в основном параметрические спецификации и могут страдать от значительных ошибок в определении.

Между тем, в области машинного обучения были достигнуты значительные успехи в областях, связанных с задачей обнаружения точек изменения. Эта область машинного обучения, известная как "обнаружение дрейфа данных" (data drift detection), признает, что условное распределение целевой переменной с учетом объясняющих переменных может измениться, что кардинально повлияет на прогностическую способность механизма машинного обучения, обученного на данных из этого распределения (см., например, Hoens, 2012; Chen et al., 2022). Это выдвигает на первый план недавние успехи передовых методов машинного обучения, таких как случайные леса и искусственные нейронные сети, в оценке гибких непараметрических условных распределений с большими размерами набора условий (см., например, Breiman, 2001; Pospisil and Lee, 2018; Friedberg et al., 2021).

В настоящее время проводится большая актуальная работа на стыке машинного обучения и статистики. Например, Barros (2018) разработал статистическую оценку для обнаружения дрейфа концепции на основе критерия суммы рангов Вилкоксона для исследования стационарности концепции. Baidari и Nagaraj (2021) применяют тест Bhattacharyya (1943) для онлайн-обнаружения дрейфа. Wang (2023) использовал сдвиги в ковариационной матрице оценок для обнаружения точек изменения, что ограничивает внимание к линейным сдвигам; Ditzler и Polikar (2011) использовали сдвиги в распределении целевой переменной с течением времени в более обобщенно. Bifet и Gavalda (2007) фактически сделали то же самое, но сосредоточились на разработке метода адаптивного управления окнами (ADWIN) для мониторинга высокочастотных потоков данных. Raab et al. (2020) расширили ADWIN для использования статистики Колмогорова-Смирнова (KSWIN) и предоставили более стабильную и точную оценку. Tajeuna et al. (2023) разрабатывают метод обнаружения сдвигов режима в многомерных временных рядах посредством анализа совместно развивающегося представления временных рядов, называемого картографической сеткой, в то время как Guo et al. (2022) и Alanqary et al. (2021) разрабатывающие новые способы измерения изменений статистических свойств многомерных распределений. Эмпирической экономике еще предстоит извлечь выгоду из этих методологических достижений.

Цель данного исследования – продвинуть методологию раннего предупреждения, объединив эти достижения в области статистики и машинного обучения с применением в экономике и финансах, в частности, в условиях высокой волатильности и зависимостью от данных. В частности, мы разрабатываем новую версию методологии определения точки изменения, основанной на условной энтропии. Новизна заключается в том, что передача информации не ограничивается линейностью и включает изменения режима в любой характеристике распределения, а не только в среднем значении или дисперсии. На самом деле, среднее значение и дисперсия базового ряда могут быть даже недостаточно четко определены. Разрабатываемая нами EWS чувствительна к любому изменению условной энтропии на основе локальной оценки условной плотности случайного леса, которая связывает текущее состояние целевой переменной с информацией в объясняющих переменных, не учитываемой запаздываниями целевой переменной. В этой настройке мы напрямую измеряем величину неопределенности, которая не может быть объяснена ни запаздыванием целевого показателя, ни объясняющими переменными, не ограничивая взаимосвязи линейными и параметрическими или сосредотачиваясь на конкретной особенности целевого распределения.

В качестве отправной точки мы берем обычную линейную спецификацию, которая обычно используется для тестирования причинно-следственной связи по Грейнджеру, и преобразуем ее во временной ряд условных оценок энтропии, основанных на скользящем окне остатков от авторегрессии (AR) целевой переменной. Это приводит к появлению онлайн-последовательной статистики SR. Насколько нам известно, использование условной энтропии в статистике SR является новым и представляет интерес само по себе. Затем мы разрабатываем расширения этой базовой настройки, чтобы воспользоваться преимуществами, предоставляемыми оценкой случайного леса и тем фактом, что процедура основана на рангах данных, а не на самих данных. Это допускает нелинейность, конечную зависимость и бесконечные моменты времени, поскольку теперь мы имеем дело со статистической связкой, которая фиксирует нелинейную зависимость, присутствующую в данных, даже если данные имеют чрезвычайно большой объем. Представленные нами симуляции демонстрируют, что новый подход доминирует в применении статистики SR непосредственно к целевой переменной или остаткам процесса AR. Они также описывают случаи, когда использование рангов позволяет выявлять изменения режима, которые не могут быть обнаружены с помощью обычной статистики SR, и когда другие методы определения точки изменения терпят неудачу или работают неоптимально.

Работа организована следующим образом. В разделе 2 описывается этап, на котором определяются обозначения и такие понятия, как энтропия, оценка случайного леса и определение точки изменения. В разделе 3 описывается базовая версия предлагаемой процедуры. В разделе 4 описываются расширения базового метода для лучшего учета нелинейности и рангов. В разделе 5 представлены многочисленные модели. Раздел 6 содержит приложение. Раздел 7 содержит выводы.

2 Предварительные сведения и обзор литературы

2.1 Энтропия и передача информации

Предположим, что у нас есть целевой временной ряд Y_t и история объясняющих переменных, доступных на момент времени t, $\mathcal{X}_t=\{X_{t-1},X_{t-2},\ldots,X_{t-l}\}$ где каждая X_s представляет собой матрицу, содержащую d объясняющих переменных (или их лаги). Примером Y_t может быть составной фондовый индекс развивающегося рынка и $\mathcal{X}_t$ может содержать историю цен на сырьевые товары, которыми торгует страна. Наша цель – обнаружить сдвиги в передаче информации между целевой и объясняющей переменными, вызывая раннее предупреждение о состоянии рынка, описываемом целевой переменной. Обычно интенсивность передачи информации от X к Y *условной энтропии* Y *данного* $\mathcal{X}$ (см., например, Orlitsky, 2003), которую можно записать следующим образом:

$$H = -\int_{\mathcal{X},\mathcal{Y}} f(y,x)\cdot\log f(y\mid x)\,dy\,dx = -\int_{\mathcal{X}} f(x)\int_{\mathcal{Y}} f(y\mid x)\cdot\log f(y\mid x)\,dy\,dx = \mathbb{E}_{\mathcal{X}}\mathbb{E}\log f(y\mid x), \tag{1}$$

где $f(y|x)$ - условная плотность Y_t по $\mathcal{X}_t$, $f(y,x)$ – совместная плотность (Y_t,X_t) и $f(X_t)$ – предельная плотность X_t. Последнее равенство следует из определения ожиданий и закона взаимосвязанных ожиданий.

По сути, концепция энтропии измеряет неопределенность в отношении Y_t по отношению к информации $\mathcal{X}_t$. В контексте передачи сигналов и безусловного распределения, Hartley (1928) предназначена для наблюдения, какая информация может быть получена из n результатов, каждый из которых имеет разные вероятности p_i, и может быть определена как

$$-\sum_{i=1}^{n} p_i log_2 p_i$$

Она известна как энтропия Шеннона (Shannon, 1948). Чем выше это значение, тем выше неопределенность в отношении n реализаций. Наибольшая степень неопределенности соответствует равномерному распределению, p_i совпадают для всех i. Уравнение (1) можно рассматривать как обобщение этой концепции на непрерывные условные распределения, показывая средний объем информации, который необходимо предоставить, прежде чем можно будет передать информацию в Y, учитывая, что другая сторона знает информацию в $\mathcal{X}$.

2.2 Непараметрические и случайные леса

Важным компонентом вычисления H является оценка условной плотности. Гибкий способ оценки $f(y|x)$ без принятия спорных допущений о распределении заключается в использовании инструментов непараметрической эконометрики (см., например, Li и Racine, 2023). Однако большинство стандартных методов не работают, когда размерность $\mathcal{X}_t$ даже умеренно велика. Это известно как проклятие размерности. Кроме того, известно, что непараметрические методы плохо работают при экстремумах распределения, которые представляют особый интерес при работе с данными развивающихся рынков.

Альтернативный подход, который набирает популярность, но не был исследован в контексте EWS, основанных на энтропии, заключается в использовании методов машинного обучения. Например, Pospisil и Lee (2018) предлагают использовать случайный лес для оценки условной плотности (RFCDE). Оценка улучшает стандартный непараметрический метод определения плотности ядра (KDE) за счет использования адаптивных весов, полученных с использованием случайных лесов Breiman (2001). Основное преимущество здесь заключается в том, что случайные леса снимают проклятие размерности за счет использования случайных подвыборок всего набора обусловливающих переменных. Это позволяет RFCDE достигать согласованной оценки условной плотности с низкими вычислительными затратами и высокой точностью.

Затем интуитивно понятно, что скачок H, оцененный с помощью RFCDE в скользящем окне, должен указывать на увеличение неопределенности (уменьшение передачи информации от $\mathcal{X}$ к Y), и остается вопрос, как идентифицировать значительные скачки.

2.3 Определение момента изменения

В литературе по обнаружению точек изменения, насчитывающей не менее 60 лет (Shiryaev, 1961, 1963; Roberts, 1966), рассматривается агрегированная

последовательность коэффициентов правдоподобия, основанная на вероятности сбоя во времени t по сравнению с вероятностью отсутствия сбоя за историю наблюдений, когда плотности до и после изменения известны, но точка изменения - нет. Допущения об известных плотностях и операционных методах, которые рассматривают скользящее окно данных, а не всю историю, существенно смягчены. Однако в этой литературе не рассматривались параметры энтропийной или ранговой оценки передачи информации без знания истинных плотностей.

Статистика Ширяева-Робертса определяется следующим образом:

$$SR_t = \sum_{k=1}^{t} \prod_{i=k}^{t} \frac{f_0(\xi_i)}{f_\infty(\xi_i)}, \tag{2}$$

где ξ_i, i=1,…,t, - общие независимые случайные величины или векторы, f_0 - плотность ξ_i до изменения и f_∞ - плотность ξ_i после изменения. Легко получить рекурсивную версию

$$SR_t = (1 + SR_{t-1}) \frac{f_0(\xi_t)}{f_\infty(\xi_t)}. \tag{3}$$

Мы используем версию взвешенной статистики SR, предложенную Tartakovsky и Spivak (2021). Процедура допускает случайность в изменении H в том смысле, что при каждом t изменении может приниматься одно из m значений, а взвешенная статистика имеет вид

$$SR_t^w = \frac{1}{m} \sum_{i=1}^{m} SR_t(\Delta\mu_i), \tag{4}$$

где $SR_t(\Delta\mu_i)$ – значение SR_t, когда разница между средним значением $f_0(\xi_t)$ и средним значением $f\infty(\xi_t)$ составляет $\Delta\mu_i$. Это обеспечивает гораздо лучшую производительность в случаях, когда параметры распределения неизвестны. В остальном, построение статистики SR_t является стандартным (см., например, Pepelyshev и Polunchenko, 2017).

Сигнал прозвучит, когда SR_t превысит пороговое значение, контролирующее частоту ложных срабатываний, и, основываясь на ряде важных результатов, процедура SR является оптимальной в смысле минимизации средней задержки обнаружения (см., например, Polunchenko и Tartakovsky, 2010).

3 Новая процедура: базовая версия

Суть новой процедуры заключается в применении статистики Ширяева-Робертса к последовательности оценок энтропии H, оцениваемых на скользящем окне длины $\Delta_t=t_1-t_0$. Длина окна должна быть достаточно короткой, чтобы изменение режима было заметным, но достаточно длинной, чтобы не возникал шум (на практике мы используем окна длиной от 50 до 100 наблюдений). Сигнал EWS срабатывает, когда $SR^{w}{}_t$ оцениваемая с помощью Ht процесса превышает пороговое значение $A>0$, что указывает на значительный сдвиг в передаче информации от Y_t к X_t. В этом разделе обсуждается, как мы строим последовательность, статистику и пороговое значение.

Мы начнем с подгонки авторегрессионной модели для целевой переменной Y_t к скользящему окну длины Δ_t:

$$\hat{Y}_t = \hat{\alpha}_0 + \sum_{s=1}^{k} \hat{\alpha}_i Y_{t-s}, \tag{5}$$

где мы используем BIC для определения оптимального порядка AR порядка k. Это дает часть Y_t ортогональной истории целевой переменной, которая является непредсказуемой компонентой Y_t, которую мы рассматриваем в:

$$e_t = Y_t - \hat{Y}_t. \tag{6}$$

Остатки e_t процесса AR(k) отражают всю оставшуюся новую информацию целевой переменной Y_t, и мы хотим исследовать, как меняется распределение этой переменной в зависимости от объясняющих переменных $\mathcal{X}$. Мы хотим оставаться независимыми в отношении конкретного момента распространения, который изменился, или формы распространения.

Далее мы получаем оцененное значение регрессии для e_t с учетом задержек объясняющих переменных, содержащихся в $\mathcal{X}$ используя то же самое скользящее окно:

$$\hat{e}_t = \hat{\beta}_0 + \sum_{j=1}^{d} \sum_{s=1}^{l} \hat{\beta}_s^{(j)} X_{t-s}^{(j)}, \tag{7}$$

где $X_{t-s}(j)$ — одна из d доступных объясняющих переменных с лагом s, $s=1,\ldots,l, j=1,\ldots,d$. Оптимальный лаг l можно выбрать с помощью BIC. Очевидно, что эти два шага эквивалентны одноэтапной регрессии Y_t как на собственную историю, так и на историю ковариатов. Это соответствует знакомой нам концепции причинно-следственной связи по Грейнджеру, которую можно выявить с помощью совместного теста значимости $\hat{\beta}_s^{(j)}$.

Мы выходим за рамки причинно-следственной связи по Грейнджеру и вычисляем условную энтропию e_t при условии $\hat{e}_t$. Это позволяет учесть тот факт, что обнаружение точек изменения должно основываться на условном распределении $f(e_t|\hat{e}_t)$ вместо линейной

регрессии, что позволяет учитывать более сложные (чем линейные в среднем) модели зависимости.

Пусть $\hat{f}$ обозначает RFCDE условной плотности e при условии $\hat{e}$ с использованием скользящего окна размером Δt. Пусть $\hat{\mathcal{E}}_{[t_0, t_1]} = \{\hat{e}_{t_0}, \hat{e}_{t_0+1}, \ldots, \hat{e}_{t_1}\}$ обозначает множество $\hat{e}$ значений, попадающих в окно. Тогда оценку совокупной передачи информации в окне можно записать как сумму вкладов энтропии для значений $\hat{e}$ попадающих в окно:

$$\hat{H}_{[t_0, t_1]} = \sum_{\hat{e} \in \hat{\mathcal{E}}_{[t_0, t_1]}} \left[\int_{\mathbb{R}} \hat{f}(\varepsilon \mid \hat{e}) \cdot \log \hat{f}(\varepsilon \mid \hat{e}) \, d\varepsilon \right]. \tag{8}$$

Мы можем аппроксимировать интегралы слагаемых, используя стохастическое или детерминированное интегральное приближение, например, правило Симпсона, которое является детерминированным методом. Мы заинтересованы в раннем выявлении всплесков $\hat{H}_{[t_0, t_1]}$ по мере продвижения во времени. Обозначим последовательность оценок энтропии в скользящем окне размера Δt следующим образом

$$\hat{H}_{[t_0, t_0 + \Delta_t]}, \hat{H}_{[(t_0+1), (t_0+1) + \Delta_t]}, \cdots$$

Мы отмечаем, что шаг между значениями H в последовательности может превышать один период времени, но это задерживает обнаружение и ускоряет вычисления. Это временные ряды, к которым мы применяем статистику SR в базовой версии нашего подхода.

Алгоритм 1 описывает процедуру Ширяева-Робертса (Shiryaev-Roberts) для определения точки изменения энтропии, описанную выше. Алгоритм использует несколько входных данных помимо последовательности H, а именно набор значений сдвига $M_\Delta = \{\Delta\mu_1, \ldots, \Delta\mu_m\}$, порог A и параметры сглаживания α и β. Хотя нет строгого правила выбора значения порога A, принято рассматривать два критерия минимизации, а именно, среднюю задержку обнаружения (ADD), определяемую как $\mathbb{E}_\theta[\tau - \theta | \tau > \theta]$, и вероятность ложной тревоги (PFA), определяемую как $\mathbb{P}_\theta(\tau < \theta)$, где θ - фактический момент изменения, а $\tau = \mathrm{argmin}_t\{t | SR_t > A\}$ - момент времени его обнаружения.

Алгоритм 1. Процедура Ширяева-Робертса для определения точки изменения энтропии

Data: Time series $H_1, H_2, \ldots, H_n$; set of possible shifts in mean M_Δ; threshold A

$SR_0 = 0; \hat{\mu}_1 = H_1; \hat{\sigma}_1^2 = 1;$

while $2 \le t \le n$ **do**

$\xi_t = \dfrac{H_t - \hat{\mu}_{t-1}}{\hat{\sigma}_{t-1}};$

for $\Delta\mu_i \in M_\Delta$ **do**

$SR_t^{(i)} = (1 + SR_{t-1}^{(i)}) \cdot \exp\{\Delta\mu_i(\xi_t - \frac{\Delta\mu_i}{2})\};$

end for

$SR_t^w = \frac{1}{m} \sum_{i=1}^{m} SR_t^{(i)}$

if $SR_t^w > A$ **then**

$\tau = t;$

break

end if

$\hat{\mu}_t = \alpha\hat{\mu}_{t-1} + (1-\alpha)H_t;$

$\hat{\sigma}_t^2 = \beta\hat{\sigma}_{t-1}^2 + (1-\beta)(H_t - \hat{\mu}_t)^2;$

$t = t + 1;$

end while

Result: change-point τ

Pergamenchtchikov и Tartakovsky (2018, 2019) выявили, что процедура взвешенного SR приблизительно минимизирует ADD среди всех процедур обнаружения точек изменения с заданным PFA в окне фиксированной длины для данных, отличных от i.i.d., включая авторегрессии и модели типа GARCH с условной гетероскедастичностью. Однако эти величины, как правило, недоступны исследователю на практике, если он не имеет дело с синтетическими данными и не способен сгенерировать их для заданного набора этих параметров. Например, в разделе 5, где мы рассматриваем моделирование, мы можем установить, A выбирая значения ADD и PFA с помощью многочисленных соотношений, выведенных в литературе (см., например, Tartakovsky, 2016).

Параметры сглаживания α и β используются в конце каждой итерации алгоритма 1 для получения экспоненциального скользящего среднего, обновляющего условное ожидание и дисперсию H, балансируя прошлые значения и новую информацию. Естественное значение для α равно 0,5, что отражает равные веса старого значения среднего $\hat{\mu}_{t-1}$ и текущего H_t, в то время как для обновления дисперсии мы выбираем β=0.9, придавая больший вес оценке $\hat{\sigma}$ за предыдущий период и, таким образом, снижая обновление дисперсии. Большие значения приводят к более плавным траекториям моментов H.

4 Расширения нелинейной оценки на основе рангов

Базовая процедура, описанная в предыдущем разделе, использует оценку условной плотности e_t заданных $\hat{e}_t$ остатков процесса AR, где $\hat{e}_t$ представляет линейный диапазон запаздывающих объясняющих переменных $X_t^{(j)}$ в уравнении (7). Это согласуется с традиционным представлением о причинно-следственной связи Грейнджера, которая определяет передачу информации путем проверки совместной значимости запаздывающих регрессоров в линейной авторегрессионной модели. В предыдущем разделе мы расширили традиционный подход, построив гибкий, основанный на ядре оценщик $f(e_t|\hat{e}_t)$ и применив статистику SR к скользящему окну оценок энтропии, $\hat{H}_{[t_0,t_1]}$ полученных с использованием $\hat{f}(e_t|\hat{e}_t)$. Цель здесь – адаптировать подход к наиболее гибкой нелинейной спецификации, признавая хорошо документированное присутствие нелинейностей в данных финансового рынка (см., например, Caporale et al., 2020)

4.1 Локальный линейный лес

Одномерное обусловливание в $\hat{f}(e|\hat{e})$ существенно упрощает задачу непараметрической оценки плотности. Альтернативный вариант с учётом всего набора ковариаций $X_{t-s}^{(j)}$ по отдельности, без получения $\hat{e}_t$, приводит к непараметрической оценке $\hat{f}(e_t|\mathcal{X}_t)$, которая сильно страдает от «проклятия размерности» из-за высокой размерности $\mathcal{X}_t$, что делает оценку очень неточной. Проблема усугубляется компромиссом между быстрой задержкой обнаружения и точностью оценки плотности. В литературе по обнаружению точек изменения хорошо известно, что если мы увеличим скользящее окно, сделав оценку плотности более точной, то время до обнаружения точки изменения увеличится. Таким образом, многомерная непараметрическая оценка либо страдает от «проклятия размерности», либо обнаруживает точку изменения с неоптимальной задержкой.

Чтобы решить эту проблему, не увеличивая размерность оценки плотности, используем недавно разработанный метод локального линейного леса (LLF), предложенный Friedberg и др. (2021). LLF создаёт оценку условного среднего значения $\tilde{e}_t=\mathbb{E}[e_t|X_{t-1}=x_{t-1},\ldots,X_{t-l}=x_{t-l}]$, обеспечивая более высокую скорость сходимости, чем традиционные непараметрические оценки той же размерности, и предотвращая переобучение. Стандартная (нелокальная) оценка случайного леса (см., например, Breiman, 2001; Probst et al., 2019) чувствительна к выбору гиперпараметров и склонна к переобучению.

Таким образом, наше первое расширение базовой методологии заключается в использовании гибкой оценки LLF $\tilde{e}_t$ вместо подобранного значения $\hat{e}_t$ при построении $\hat{f}(e_t|\tilde{e}_t)$. Статистический показатель SR теперь основан на последовательности новых оценок H в скользящем окне $[t_0,t_1]$:

$$\hat{H}^{LLF}_{[t_0;t_1]} = \sum_{\tilde{e} \in \Psi_{[t_0;t_1]}} \left[\int_{\mathbb{R}} \hat{f}(\varepsilon \mid \tilde{e}) \cdot \log \hat{f}(\varepsilon \mid \tilde{e}) \, d\varepsilon \right], \tag{9}$$

где $\Psi_{[t_0;\, t_1]} = \{\tilde{e}_{t_0};\, \tilde{e}_{t_0+1};\, \ldots;\, \tilde{e}_{t_1}\}$. Остальная часть методологии остается неизменной.

4.2 Копулы

Наше второе расширение затрагивает проблему больших хвостов и зависимости от хвостов, в дополнение к проблеме нелинейности. Хорошо задокументировано, что финансовые рынки развивающихся экономик характеризуются асимметричным распределением с большим хвостом, что затрудняет применение стандартных эконометрических методов (см., например, Chen и Ibragimov, 2019). В качестве альтернативы мы следуем растущей литературе по оценке на основе копул в финансах (см., например, Ibragimov и Prokhorov, 2017) и переформулируем проблему оценки плотности с использованием рангов вместо фактических значений. Важно отметить, что для любого набора случайных величин их копулятивная функция содержит всю информацию о структуре зависимости между случайной величиной, а также между любыми преобразованиями случайных величин с сохранением ранга (см., например, Cherubini et al., 2011, обзор методов копулы в финансах). Следовательно, если мы используем копулу вместо исходного распределения данных в базовом методе, структура передачи информации останется неизменной, в то время как оценка теперь выполняется по рангам и H использует условную плотность копулы, которая устойчива к большим хвостам в маргинальных распределениях и четко определена, даже когда моменты исходных распределений могут быть бесконечными.

В предлагаемой модификации базовая оценка RFCDE $\hat{f}(e_t|\hat{e}_t)$ заменена оценкой RFCDE на основе рангов $\hat{f}(u_t|\hat{u}_t)$, где

$$u_t = \frac{1}{\Delta_t}\sum_{s=1}^{\Delta} \mathbb{1}\{e_{t-s} \le e_t\}, \qquad \hat{u}_t = \frac{1}{\Delta_t}\sum_{s=1}^{\Delta_t} \mathbb{1}\{\hat{e}_{t-s} \le \hat{e}_t\},$$

где, как и ранее, Δt – длина скользящего окна, по которому мы оцениваем плотности. Значения u_t и $\hat{u}_t$ известны как псевдонаблюдения; это значения между [0,1], соответствующие рангам наблюдений e_t и $\hat{e}_t$, соответственно, в скользящем окне, деленные на количество наблюдений в окне. Соответствующие оценки энтропии можно записать следующим образом

$$\hat{H}^{rank}_{[t_0;\, t_1]} = \sum_{\hat{u} \in \mathcal{U}_{[t_0;\, t_1]}} \left[\int_0^1 \hat{f}(w \mid \hat{u}) \cdot \log \hat{f}(w \mid \hat{u})\, dw\right], \tag{10}$$

Где $\mathcal{U}_{[t_0;\, t_1]} = \{\hat{u}_{t_0};\, \hat{u}_{t_0+1};\, \ldots;\, \hat{u}_{t_1}\}$ и $\hat{f}(u|\hat{u})$ могут быть проинтерпретированы как условная плотность копулы e_t по $\hat{e}_t$.

Ранговые оценки, приведенные в уравнении (10), инвариантны к преобразованиям переменных с сохранением ранга и, таким образом, делают большие хвосты в предельных

распределениях неуместными. Это позволяет дополнительно сосредоточиться на изменениях в характеристиках распределения, которые выходят за рамки первых двух моментов величин. Например, мы можем рассматривать ситуации, когда первые два момента могут отсутствовать из-за чрезвычайной волатильности на рынках и когда линейные корреляции слишком упрощены в качестве меры зависимости, например, когда данные зависят от хвоста.

5 Моделирование методом Монте-Карло

5.1 Проектирование моделирования

В этом разделе мы сравниваем эффективность базовых методов и методов расширения с альтернативными процедурами и между собой, используя синтетические данные. Мы используем два распространённых в литературе сценария, а именно прекращение передачи и инверсию передачи. Примеры двух используемых нами схем:

(a) Прекращение передачи информации: в точке изменения передача информации от X и Z к Y прекращается. Мы рассматриваем следующий процесс генерации данных (DGP):

$$Y_t = \begin{cases} 0.5 \ln X_{t-1} + 0.2 Z_{t-1}^2 + 0.3 v_t, & t \leq 500 \\ \xi_t, & t > 500 \end{cases}$$

где

$$v_1, \ldots, v_{500} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \xi_5 01, \ldots, \xi_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(5,36)\ X_1, \ldots, X_T \overset{\text{i.i.d.}}{\sim} Exp(3)$$

и

$$Z_1, \ldots, Z_T \overset{\text{i.i.d.}}{\sim} \Gamma(3,1).$$

(b) Инверсия передачи информации: в точке изменения поток информации от X и Z к Y меняется на противоположный. DGP, который мы используем, выглядит следующим образом

$$Y_t = \begin{cases} 0.5 \ln X_{t-1} + 0.2 Z_{t-1}^2 + 0.3 v_t, & t \leq 500 \\ \xi_t, & t > 500 \end{cases}$$

где

$$X_t = \begin{cases} \upsilon_t, & t \leq 500 \\ 0.3 Y_{t-1} + 0.1 \psi_t, & t > 500 \end{cases}$$

$$Z_t = \begin{cases} \delta_t, & t \leq 500 \\ 0.6 Y_{t-1}^2 + 0.1 \zeta_t, & t > 500 \end{cases}$$

и где

$$v_1, \ldots, v_{500}, \psi_{501}, \ldots, \psi_T, \zeta_{501}, \ldots, \zeta_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \qquad \xi_{501}, \ldots, \xi_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(5,9)$$

и

$$\upsilon_1, \ldots, \upsilon_{500} \overset{\text{i.i.d.}}{\sim} Exp(3), \delta_1, \ldots, \delta_{500} \overset{\text{i.i.d.}}{\sim} \Gamma(3,1)$$

Сводная статистика смоделированных данных с использованием двух DGP'ов представлена в таблице 1.

Таблица 1: Сводная статистика на синтетических данных

режим	**среднее**	**std**	**асимметрия**	**эксцесс**	$corr(Y_t;Y_{t-1})$	**rank** $corr(Y_t;Y_{t-1})$
Прекращение передачи информации						
перед изменением	2.541	2.768	-2.106	5.812	-0.062	-0.040
после изменения	5.126	5.968	0.110	0.241	-0.076	-0.080
Инверсия передачи информации						
перед изменением	2.541	2.768	-1.527	2.973	0.040	-0.040
после изменения	5.126	2.99	-0.192	0.137	0.007	-0.064

5.2 Сравнение с традиционным подходом

Мы начнём с сравнения нового подхода с традиционными методами SR, то есть без использования энтропии. Традиционные альтернативы, которые мы рассматриваем, следующие: (а) применение статистики SR для обнаружения изменения среднего значения целевой переменной Y_t напрямую; (б) применение статистики SR для обнаружения изменения среднего значения остатков e_t от модели AR в уравнении (6); (в) применение статистики SR для обнаружения изменения среднего значения подобранных значений $\hat{e}_t$ из уравнения (7). Мы сосредоточимся на прекращении передачи информации.

Спецификация DGP, рассматриваемая в этом разделе, следующая:

$$Y_t = \begin{cases} 0.5 \ln X_{t-1} + 0.2 Z_{t-1}^2 + 0.1 v_t, & t \leq 500 \\ \xi_t, & t > 500 \end{cases}$$

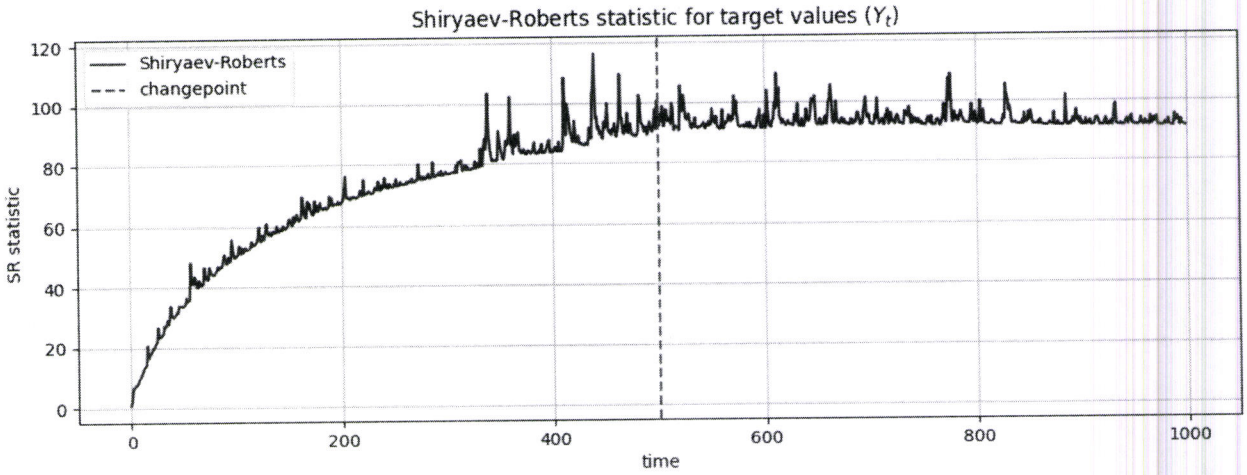
где

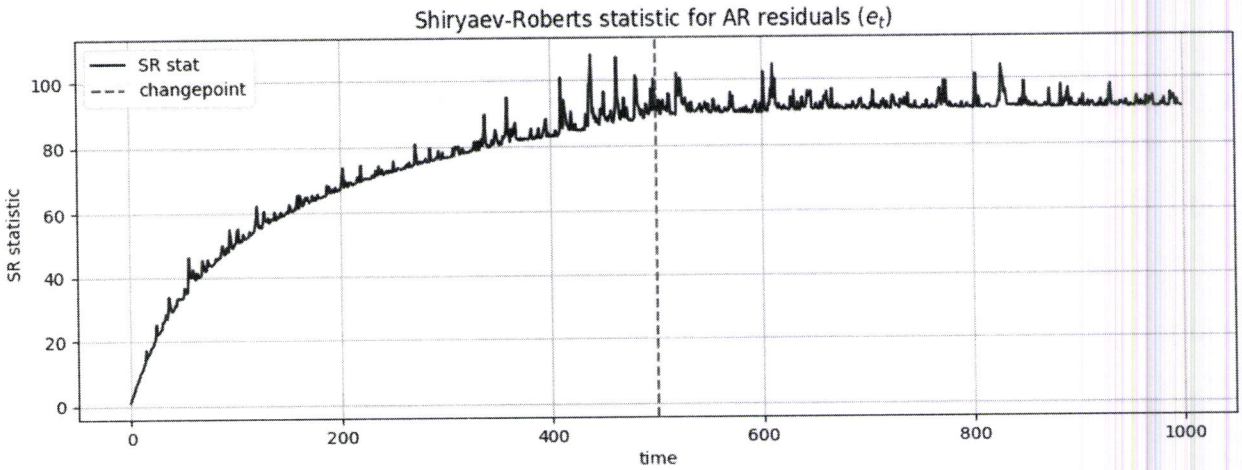
$$\nu_1, \ldots, \nu_{500} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1),\ \xi_{501}, \ldots \xi_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(2.35, 1.5),\ X_1, \ldots, X_T \overset{\text{i.i.d.}}{\sim} Exp(3),$$

и

$$Z_1, \ldots, Z_T \overset{\text{i.i.d.}}{\sim} \Gamma(3,1)$$

Процесс аналогичен прекращению информационного потока в разделе 5.1, за исключением того, что теперь среднее значение сдвигается на меньшую величину, слишком малую для учета традиционной процедурой. Все параметры для вычисления статистики SR одинаковы для всех реализаций традиционной статистики SR. Длина окна составляет Δt=50.





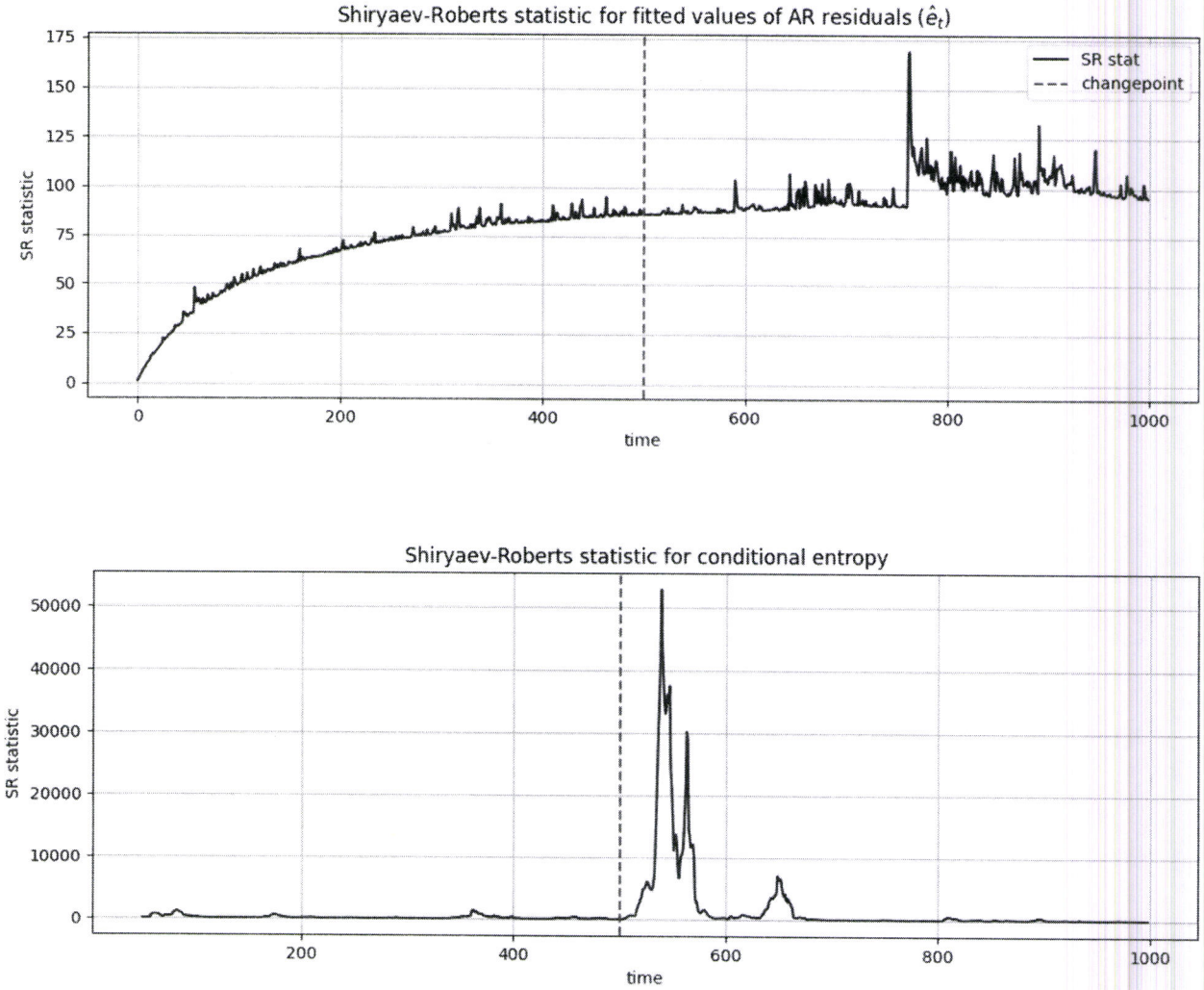


Рисунок 1: Статистика Ширяева-Робертса для прекращения передачи информации

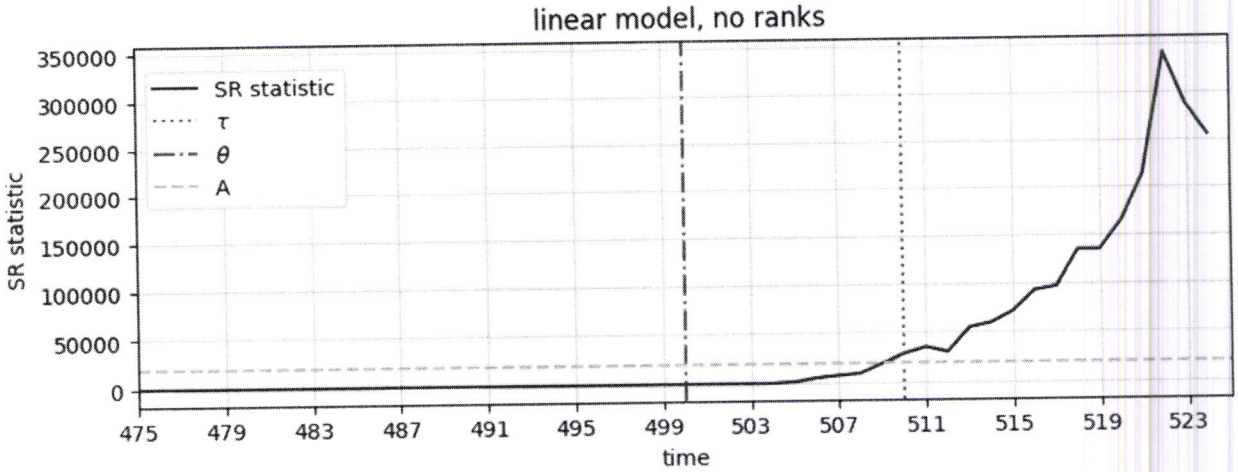
На рисунке 1 показаны значения соответствующей статистики SR и местоположение точки изменения (вертикальная пунктирная линия). Из рисунка видно, что традиционные способы применения процедуры SR не позволяют обнаружить точку изменения. Статистика SR для трех из четырех цифр растет со временем и выравнивается вблизи точки изменения, но никакие всплески не являются достаточно большими, чтобы вызвать тревогу. С другой стороны, процедура, основанная на энтропии, показанная на нижней правой панели, вызвала немедленную и сильную тревогу.

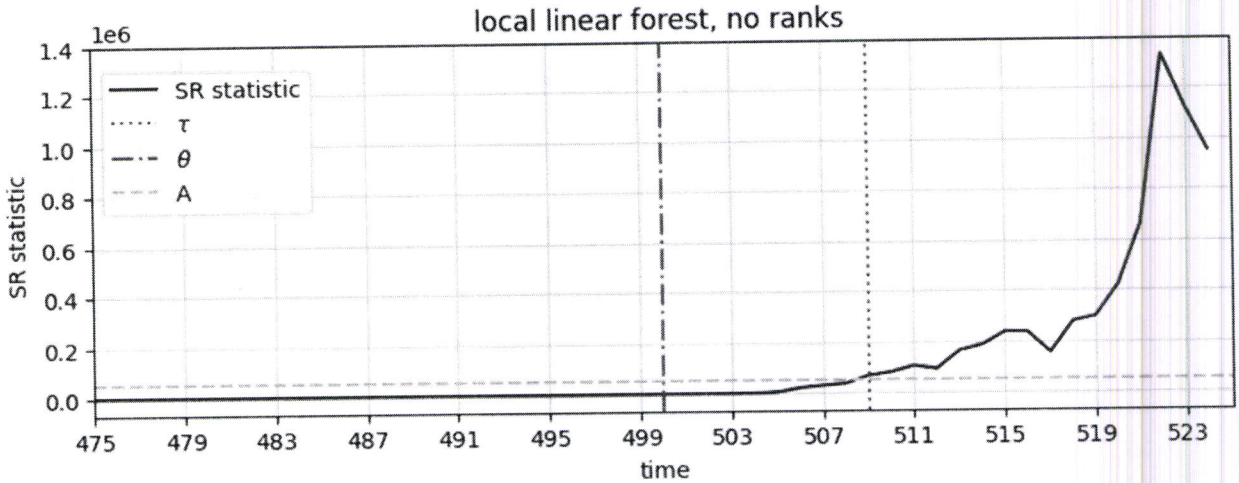
Эта разница в производительности обусловлена ограниченностью традиционных подходов. При применении к конкретным переменным, в отличие от энтропии, они, по-видимому, нацелены на сдвиги специфического характера, например, сдвиги в среднем значении, дисперсии, автокорреляции. В данном случае сдвиг в среднем слишком мал. Ключевым преимуществом нашего подхода является то, что он учитывает скачки в *передаче информации*. В этом смысле вместо анализа различных сдвигов в статистике SR, каждый из которых проверяет наличие определенного смещения, мера, основанная на энтропии, напрямую учитывает стабильность структуры зависимости.

5.3 Нелинейность и ранги

Теперь мы переходим к обнаружению прекращения и инверсии передачи информации с помощью расширений подхода, основанного на энтропии, то есть с помощью локальных линейных лесов и рангов. Моделирование проводится так, как описано в разделе 5.1. Результаты моделирования представлены на рисунках (2)-(3), где θ обозначает момент изменения, τ обозначает момент обнаружения изменения, а A — пороговое значение, используемое для статистики Ширяева-Робертса. Длина окна составляет Δt=50.

Из рисунков (2)-(3) следует, что все расширения базовых методов, основанных на энтропии, очень хорошо работают во всех четырех настройках. Нет четкого порядка в производительности между четырьмя настройками: (a) базовая версия (линейная модель, без рангов); (b) расширение с помощью LLF (локальный линейный лес, без рангов); (c) расширение с рангами $\hat{e}_t$ e^t (линейная модель, ранги); (d) расширение с рангами $\tilde{e}_t$ (локальный линейный лес, ранги). Как для завершения, так и для инверсии все четыре способны подавать сигнал тревоги в течение 7-15 периодов времени, используя соответствующие пороговые значения. Таблица 2 суммирует этот вывод.





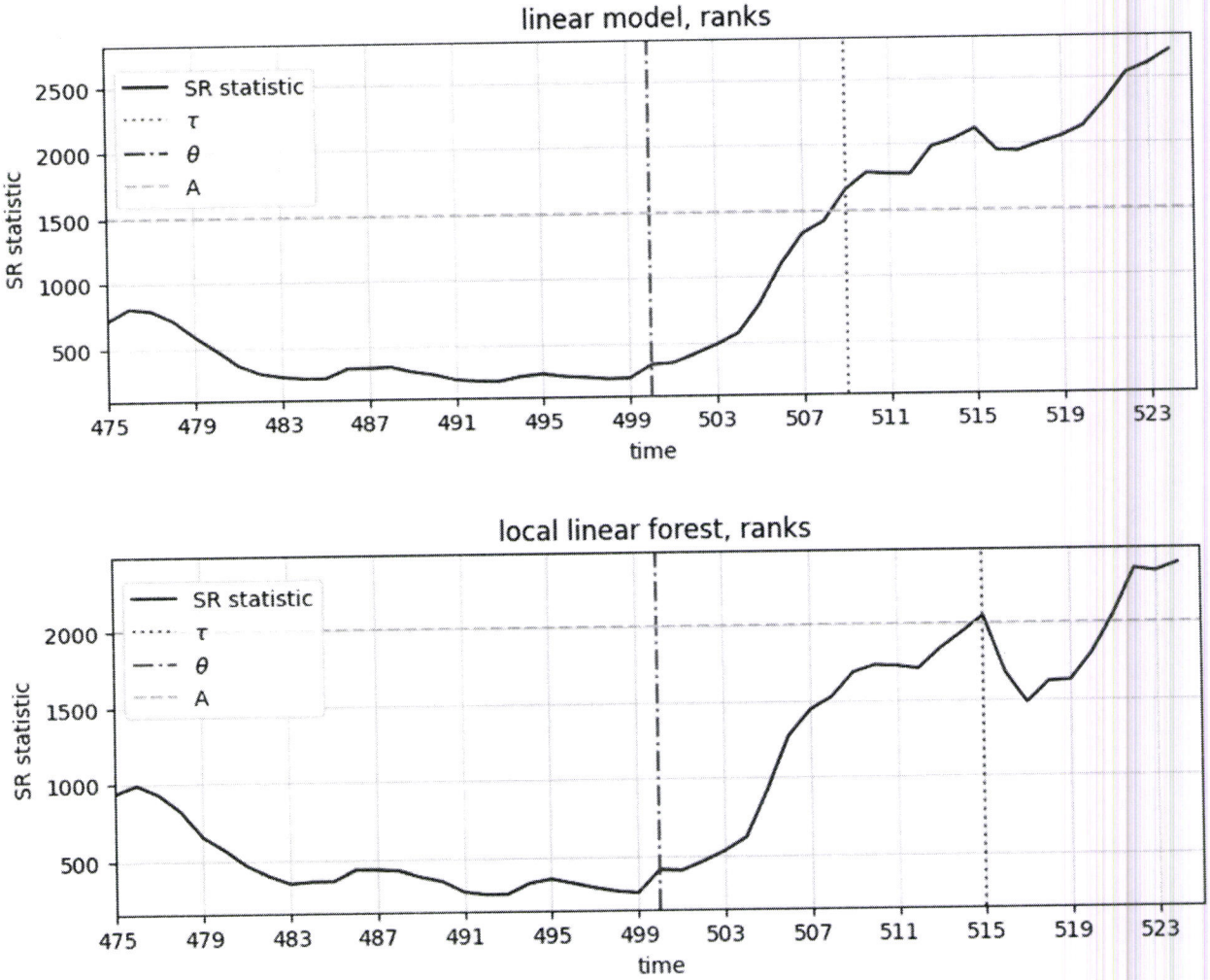
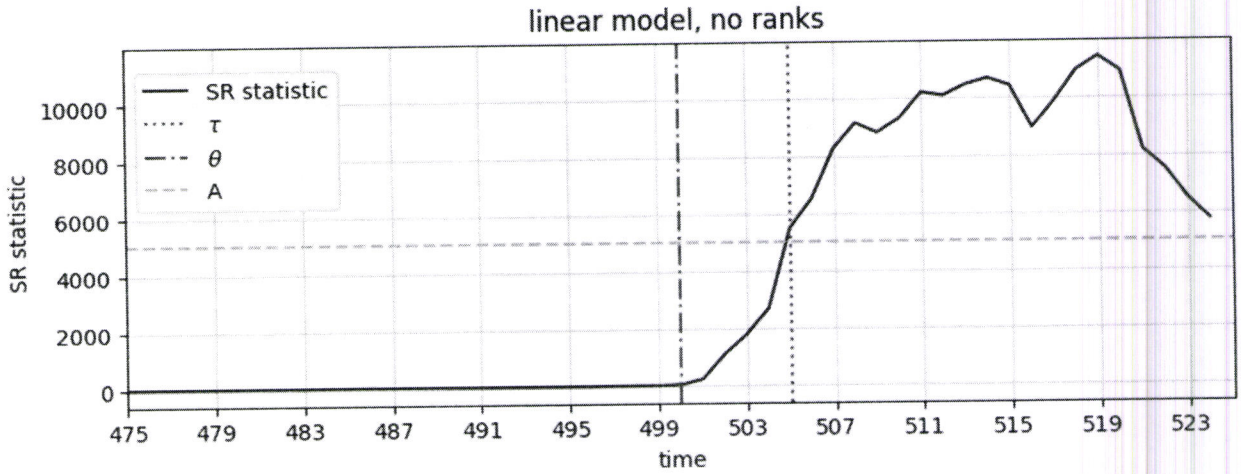


Рисунок 2: Статистика Ширяева-Робертса по прекращению передачи информации



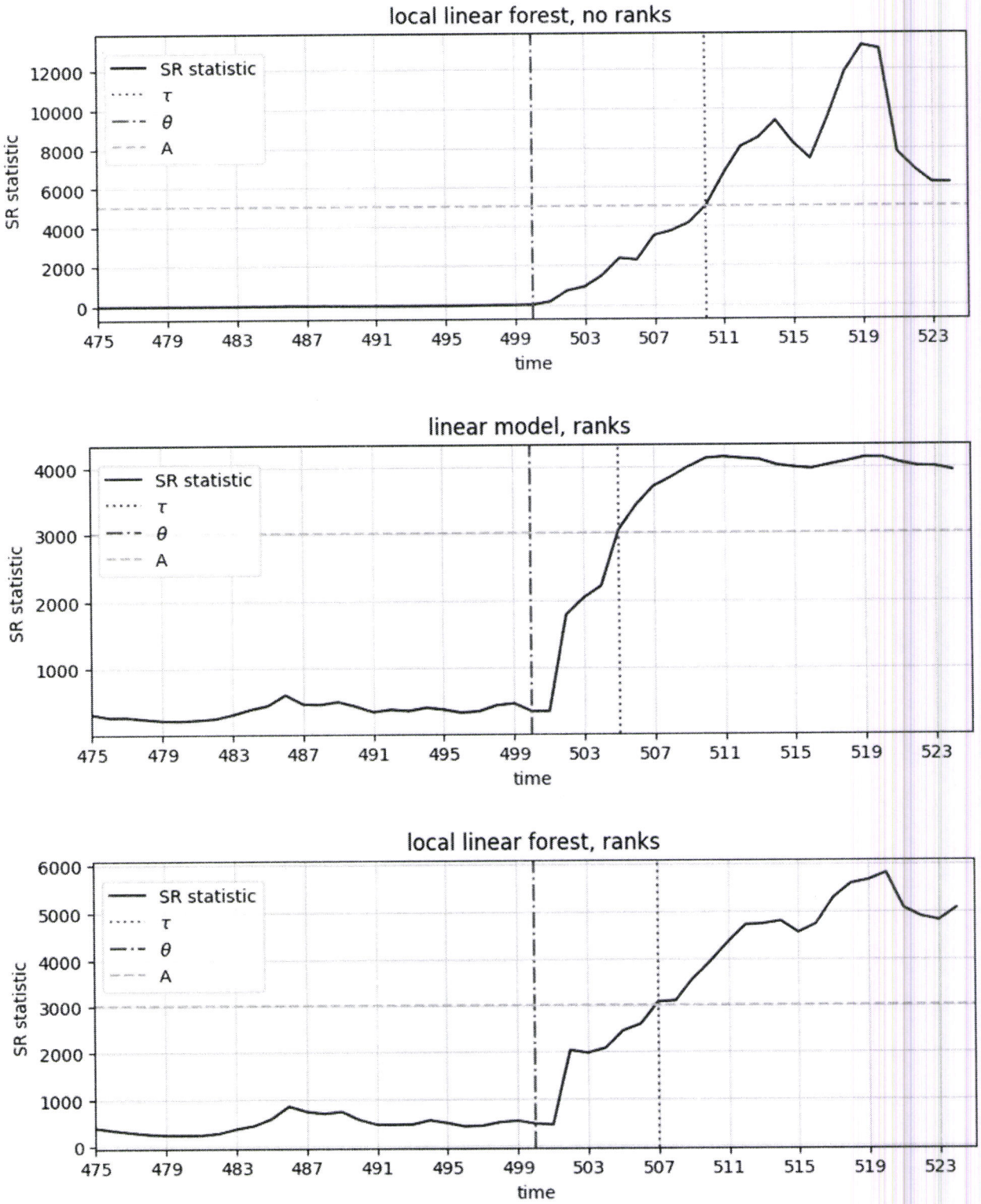


Рисунок 3: Статистика Ширяева-Робертса для инверсии передачи информации

Интересно, что сравнение между нижней и верхней панелями рисунков показывает, что PFA для рангов выше, чем для уровней, о чем свидетельствуют ненулевые значения SR перед разрывом на нижних панелях (и нулевые SR на верхних панелях). Однако это не позволяет нам четко определить случаи, когда меры, основанные на рангах, могут быть предпочтительными, а могут и не быть, что рассматривается в следующем подразделе.

Таблица 2: Задержка обнаружения точки изменения ($\tau-\theta$)

Модель	Termination	Inversion
linear	10	5
LLF	9	10
linear + ranks	9	5
LLF + ranks	15	7

5.4 Тяжелые хвосты и зависимость от хвостов

Для дальнейшего изучения преимуществ ранговых показателей мы рассматриваем ситуацию, в которой свойство с тяжелохвостности отдельных распределений препятствует продуктивному использованию таких характеристик, как среднее значение, дисперсия или корреляции. Этот параметр особенно актуален для финансовых данных с развивающихся рынков, которые, как известно, чрезвычайно волатильны.

Мы модифицируем описанную схему моделирования, включив компонент с данными с тяжелыми хвостами и зависимостью от хвостов. С этой целью мы добавляем штраф λ к упрощенной версии прекращения информационного потока следующим образом:

$$Y_t = \begin{cases} X_{t-1} + (1 - \sqrt{\lambda_{t-1}})\xi_t, & t \leq 500, \\ 0.9 + \xi_t, & t > 500, \end{cases}$$

где

$$\xi_1, \ldots, \xi_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.5), X_1, \ldots, X_{500} \overset{\text{i.i.d.}}{\sim} WeibMin(1.5)$$

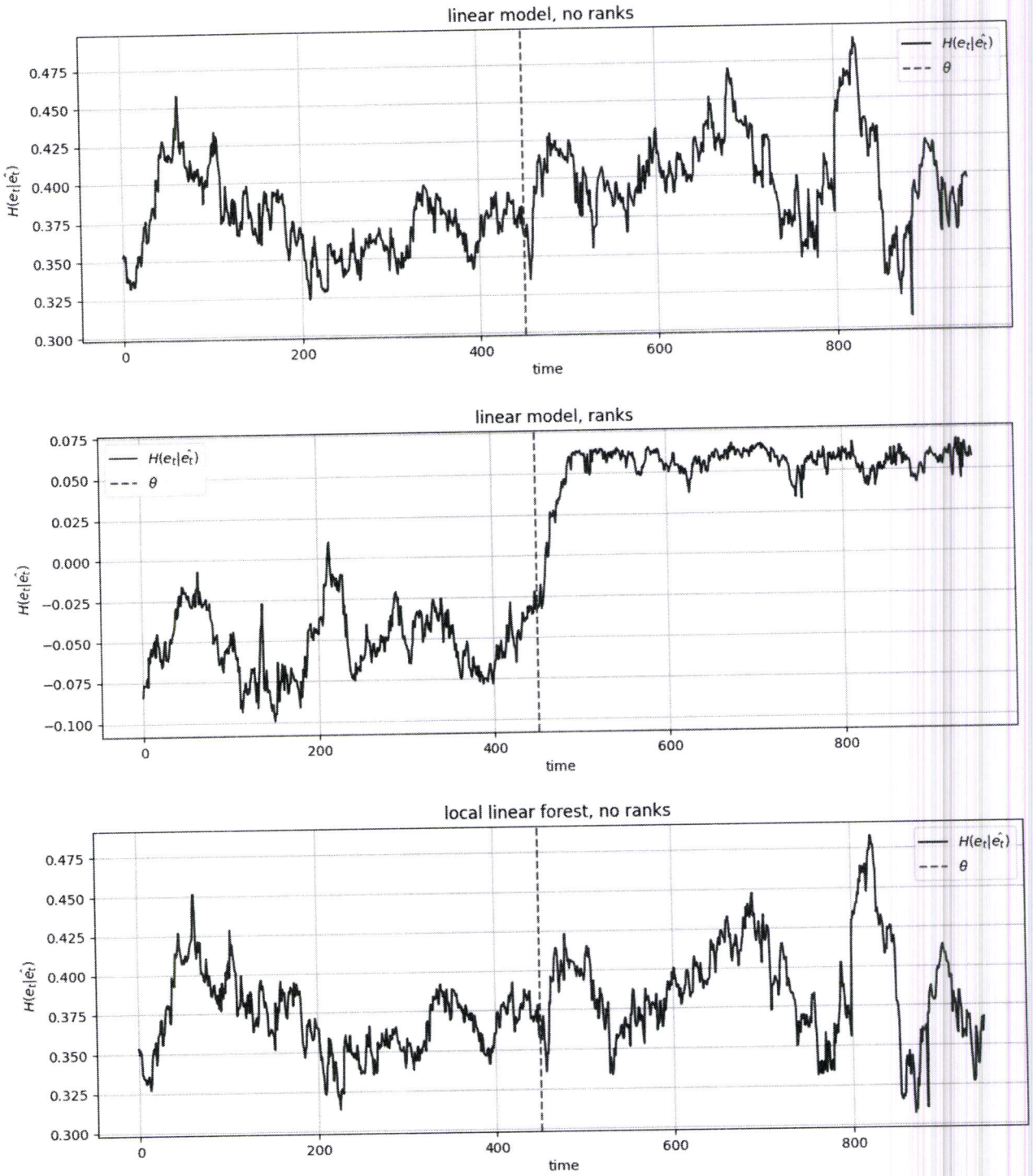
Роль λ заключается в контроле уровня шума в зависимости от того, насколько близко наблюдение к медиане выборки. Пусть $d_t = X_t - \mathrm{med}(X_t)$ обозначим такое расстояние. Мы определяем λ следующим образом:

$$\lambda_t = \frac{d_t - \min_t\{d_t\}}{\max_t\{d_t\} - \min_t\{d_t\}} \in [0, 1].$$

Очевидно, что чем ближе наблюдение к медиане, тем больше шума Y_t относительно X_{t-1}. Что касается наблюдений, близких к экстремумам выборки, то они Y_t ближе к тому, чтобы полностью совпадать с X_{t-1}. Это приводит к ненулевой хвостовой зависимости, то есть к ненулевой корреляции между X_t и Y_t ближе к экстремумам этих переменных. Кроме того, мы используем распределение минимальных экстремальных значений Вейбулла для X_t, чтобы вызвать свойство с тяжелохвосности.

На рисунке 4 показаны условные энтропийные процессы, H полученные с использованием четырех рассмотренных нами методов: (а) базовой линейной модели, основанной на уровнях, (б) локальной линейной модели леса, основанной на уровнях, (в) линейной

модели, основанной на рангах, (г) локальной линейной модели леса, основанной на рангах. Вертикальная пунктирная линия указывает точку изменения (θ).



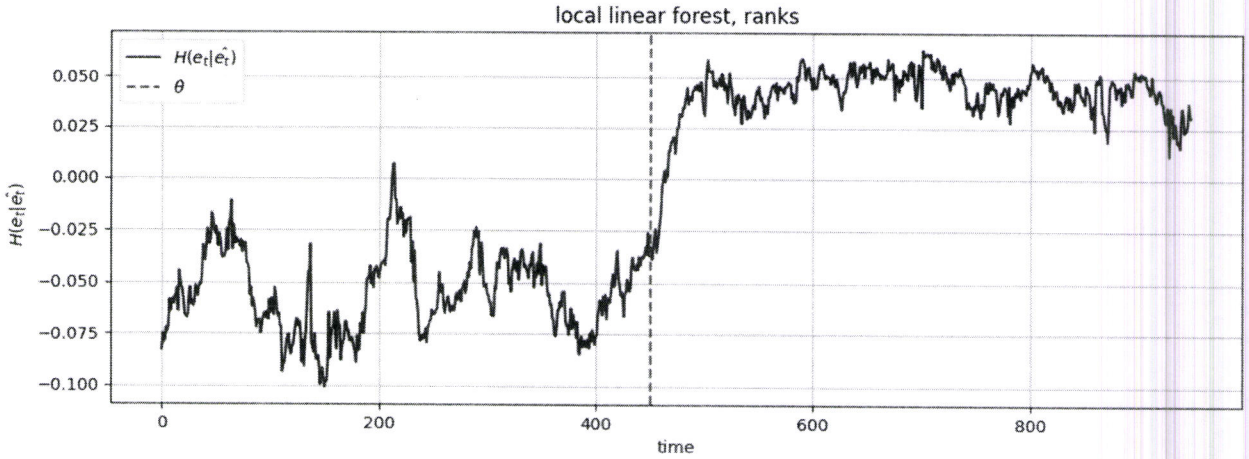


Рисунок 4: Оценки энтропии для данных с тяжёлыми хвостами и зависимостью от хвоста

Из рисунка 4 видно, что даже без вычисления статистики SR, основанной на энтропии, этот подход не позволяет обнаружить сдвиг режима на основе исходных данных (верхняя и нижняя левые панели не показывают заметного изменения H), в то время как EWS, основанная на ранжировании, дает согласованные результаты и быстро определяет режим сдвиг (верхняя и нижняя правые панели показывают четкий переход к новому среднему значению H после точки изменения). Это говорит о том, что новая методология, основанная на ранжировании, привлекательна в тех случаях, когда финансовые данные являются статистически значимыми и зависят от хвоста.

6 Апробация подхода на эмпирических данных

В разделе приводится эмпирическая иллюстрация того, как подход может быть использован при построении системы раннего предупреждения на основе макроэкономических данных.

6.1 Узбекистан: влияние сырьевых рынков

Изучение динамики фондового рынка Узбекистана, измеряемой с помощью Сводного индекса Узбекистана (UCI). UCI - средневзвешенное значение цен акций, торгуемых на Ташкентской фондовой бирже.

Источники данных:

The following sources were used to obtain the data for Section 6:

Uzbekistan Composite Index: https://uzse.uz/price_indices/

Brent Crude Oil Continuous Contract: https://www.marketwatch.com/investing/future/brn00

Wheat ($/bsh): https://www.macrotrends.net/2534/wheat-prices-historical-chart-data

Gold spot: https://www.gold.org/goldhub/data/gold-prices

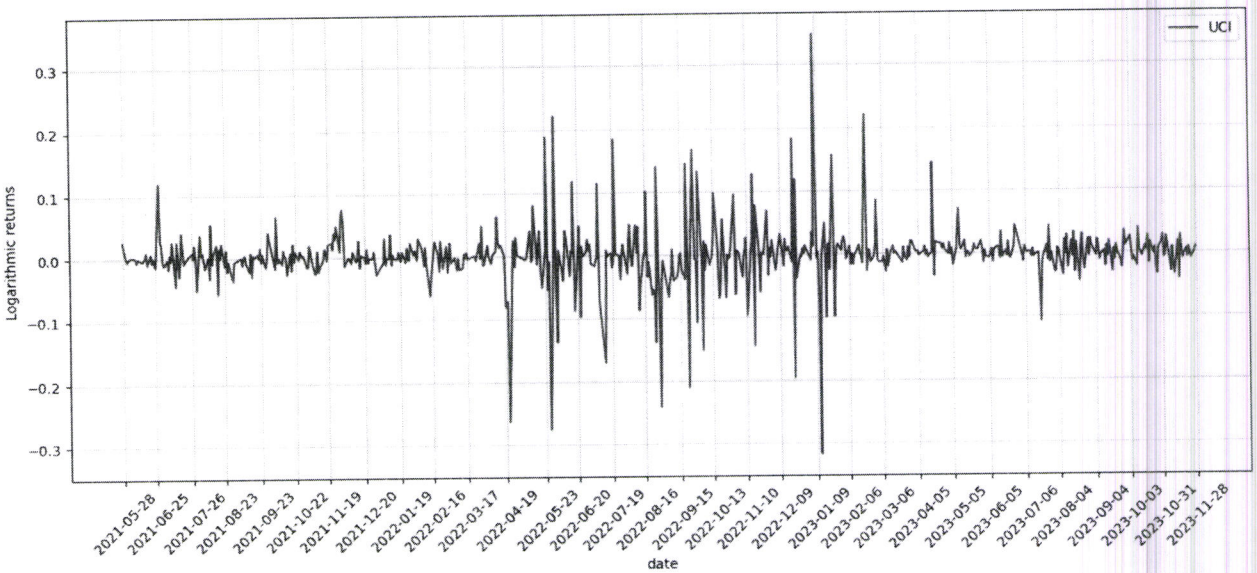
Cotton spot: https://finance.yahoo.com/quote

Согласно отчету Всемирного банка за 2022 год, сырьевой рынок играет ключевую роль в экономике этой страны, например, золото и хлопок составляют 30,8% и 11% экспорта страны соответственно. Структура импорта более однородна, при этом наибольшие доли импорта приходится на нефтепродукты (2,5% импорта страны) и пшеницу (2,2% импорта). Естественно предположить, что фондовый рынок этой открытой экономики подвержен шокам со стороны мировых товарных рынков. Поэтому мы изучаем передачу информации от цен на сырьевые товары в UCI, где список сырьевых товаров включает золото, пшеницу, хлопок и сырую нефть. Мы рассматриваем ежедневные данные с 1 января 2021 года по 30 ноября 2022 года. В таблице 4 представлена сводная статистика за весь период.

Таблица 4:Сводные статистические данные по Узбекистану

	среднее	**std**	**асимметрия**	**эксцесс**	$corr(Y_t;Y_{t-1})$	**rank** $corr(Y_t;Y_{t-1})$
UCI	0.000	0.047	-0.226	14.259	-0.075	-0.049
Crude oil	0.000	0.027	-0.523	3.636	-0.013	0.028
Gold	0.000	0.009	-0.238	1.844	-0.023	-0.009
Wheat	0.000	0.022	0.309	3.413	-0.025	-0.024
Cotton	0.000	0.023	-2.202	23.497	0.018	-0.012

Мы используем скользящее окно в 100 дней для оценки процесса условной энтропии $\widehat{H}$ как описано в разделе 4.1. Порядок процесса AR для UCI выбирается BIC и отличается для каждого скользящего окна; наибольшее запаздывание объясняющих переменных в уравнении (7) равно трем, но мы провели анализ чувствительности к этим параметрам, варьирующимся от 1 до 10, без существенной разницы в результатах. Количество сдвигов m=80, параметры экспоненциального сглаживания равны α=0.5 и β=0.9, как и в симуляциях.



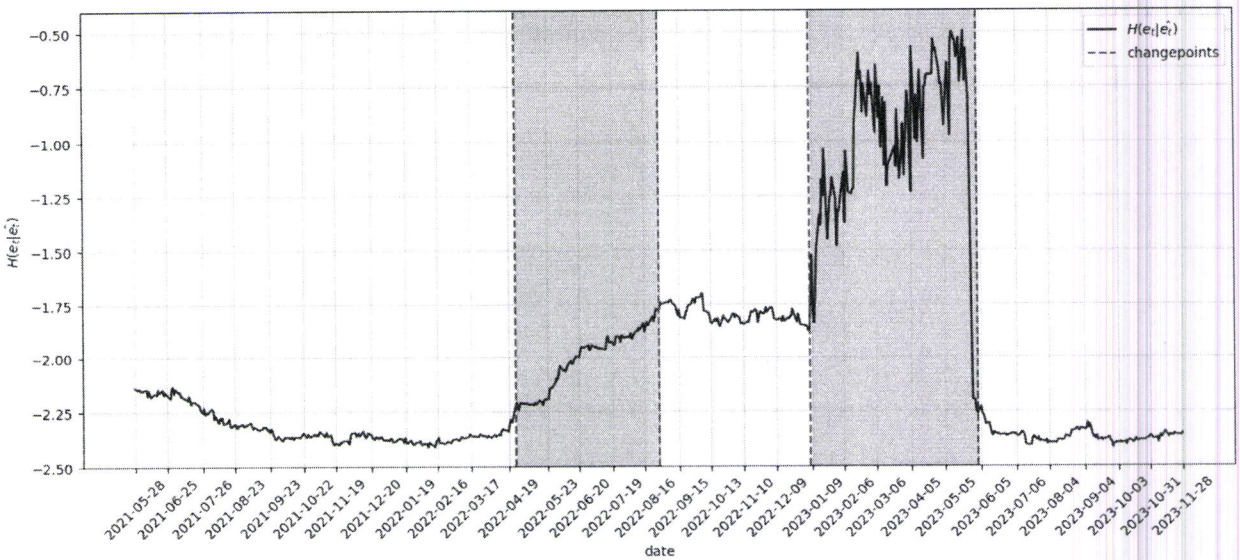


Рисунок 5: Логарифмическая доходность и условная энтропия для фондового рынка Узбекистана.

Рисунок 5 отображает результирующую условную энтропию вместе с логарифмической доходностью UCI. На рисунке представлены пять различных режимов, обнаруженных этим EWS. Они обозначены вертикальными пунктирными линиями и заштрихованными областями на рисунке. Даты и сводная статистика для каждого режима приведены в таблице 5.

Таблица 5: Сводная статистика UCI по режимам

режим	**среднее**	**std**	**асимметрия**	**эксцесс**	$corr(Y_t;Y_{t-1})$	**rank** $corr(Y_t;Y_{t-1})$
2021.05.28 - 2022.04.25	0.002	0.022	-3.051	30.794	0.229	-0.021
2022.04.26 - 2022.08.29	-0.008	0.077	-0.346	3.586	-0.165	-0.107
2022.08.30 - 2023.01.07	0.001	0.073	1.010	5.782	-0.273	-0.194
2023.01.08 - 2023.06.02	0.002	0.050	-1.404	19.710	-0.036	-0.0620
2023.06.03 - 2023.11.28	-0.001	0.020	-1.178	5.380	-0.147	-0.255

Обнаруженные режимы не видны в первоначальном ряду показателей доходности UCI, за исключением, возможно, первого момента изменения, который совпадает с заметным увеличением волатильности UCI. Это неудивительно, учитывая, что условная энтропия отражает совокупный частичный эффект товарных рынков. Существуют следующие пять режимов: (а) период низкой энтропии, (б) период медленного роста энтропии, (в) период стабилизации, (г) период повышения энтропии и (д) период возврата к нормальному состоянию. Режим (d), возможно, является неожиданным и, должно быть, отражает разрыв между фондовым и товарным рынками в тот период.

7 Заключение

Мы разработали новую систему раннего предупреждения для обнаружения концептуального дрейфа с помощью оценки условной энтропии. Это оригинальный подход, который сочетает в себе преимущества теории информации и машинного обучения, чтобы помочь отслеживать уровень неопределенности, остающийся после учета задержек целевых и объясняющих переменных. Этот подход особенно актуален для моделирования информационных потоков на развивающихся рынках, которые характеризуются повторяющимися структурными разрывами и тяжелыми хвостами в распределении данных.

Мы предложили базовые методы, основанные на новой оценке условной плотности с использованием весов случайных лесов, и мы разработали два расширения базового уровня, которые лучше отражают нелинейные взаимодействия и учитывают распределения с большим хвостом. Расширения используют локальную линейную оценку условного среднего и ранговую оценку условной плотности связки соответственно.

По сути, предлагаемые методы являются первым шагом к использованию последних достижений в области машинного обучения и статистики для автоматизированного мониторинга рынка, которые могут использоваться для многих целей, но имеют особое значение для рынков с выраженной волатильностью, структурными разрывами, тяжелыми хвостами и зависимостью от хвостов.

Список литературы

1. Ahmad, W., A. V. Mishra, and K. J. Daly (2018): "Financial connectedness of BRICS and global sovereign bond markets," *Emerging Markets Review*, 37, 1–16.
2. Alanqary, A., A. Alomar, and D. Shah (2021): "Change Point Detection via Multivariate Singular Spectrum Analysis," in *Advances in Neural Information Processing Systems*, ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Curran Associates, Inc., vol. 34, 23218–23230.
3. Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.
4. Bai, J., & Perron, P. (2003): "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1–22.
5. Baidari, I. and H. Nagaraj (2021): "Bhattacharyya Distance Based Concept Drift Detection Method for Evolving Data Stream," *Expert Systems with Applications*, 183.
6. Barros, R. S. M. d. e. a. (2018): "Wilcoxon Rank Sum Test Drift Detector," *Neurocomputing*, 275, 1954–1963.
7. Bhattacharyya, A. K. (1943): "On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions," *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.
8. Bifet, A. and R. Gavalda (2007): "Learning from Time-Changing Data with Adaptive Windowing," *Proceedings of the Seventh SIAM International Conference on Data Mining*.

9. Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

10. Caporale, G. M., L. A. Gil-Alana, and C. Poza (2020): "Persistence, non-linearities and structural breaks in European stock market indices," *The Quarterly Review of Economics and Finance*, 77, 50–61.

11. Chaudhuri, K. and Y. Wu (2003): "Random walk versus breaking trend in stock prices: Evidence from emerging markets," *Journal of Banking & Finance*, 27, 575–592.

12. Chen, H. (2019): "Sequential change-point detection based on nearest neighbors," *Annals of Statistics*, 47, 1381 – 1407.

13. Chen, Z., M. Han, H. Wu, M. Li, and X. Zhang (2022): "A multi-level weighted concept drift detection method," *Journal of Supercomputing*, 79, 5154–5180.

14. Chen, Z. and R. Ibragimov (2019): "One country, two systems? The heavy-tailedness of Chinese A- and H- share markets," *Emerging Markets Review*, 38, 115–141.

15. Cherubini, U., S. Mulinacci, F. Gobbi, and S. Romagnoli (2011): *Dynamic Copula Methods in Finance*, Wiley Finance Series, John Wiley and Sons Ltd.

16. Chu, L. and H. Chen (2019): "Asymptotic distribution-free change-point detection for multivariate and non-euclidean data," *Annals of Statistics*, 47, 382 – 414.

17. Ditzler, G. and R. Polikar (2011): "Hellinger Distance Based Drift Detection for Nonstationary Environments," *IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (CIDUE)*.

18. Friedberg, R., J. Tibshirani, S. Athey, and S. Wager (2021): "Local Linear Forests," *Journal of Computational and Graphical Statistics*, 30, 503–517.

19. Granger, C. W. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 424–438.

20. Guo, Y., M. Gao, and X. Lu (2022): "Multivariate change point detection for heterogeneous series," *Neurocomputing*, 510, 122–134.

21. Hamilton, J. (2016): "Macroeconomic Regimes and Regime Shifts," Elsevier, vol. 2 of *Handbook of Macroeconomics*, 163–201.

22. Hartley, R. (1928): "Transmission of information," *Bell System Technical Journal*, 535–563.

23. Hoens, R. e. a. (2012): "Learning from Streaming Data with Concept Drift and Imbalance: An Overview," *Progress in Artificial Intelligence*, 1, 89–101.

24. Ibragimov, R. and A. Prokhorov (2017): *Heavy tails and copulas: Topics in dependence modelling in economics and finance*, World Scientific Publishing: Singapore.

25. Khalfaoui, R., S. Hammoudeh, and M. Z. Rehman (2023): "Spillovers and connectedness among BRICS stock markets, cryptocurrencies, and uncertainty: Evidence from the quantile vector autoregression network," *Emerging Markets Review*, 54, 101002.

26. Li, Q. and J. S. Racine (2023): *Nonparametric econometrics: theory and practice*, Princeton University Press.

27. Neaime, S. (2016): "Financial crises and contagion vulnerability of MENA stock markets," *Emerging Markets Review*, 27, 14–35.

28. Orlitsky, A. (2003): "Information Theory," in *Encyclopedia of Physical Science and Technology (Third Edition)*, ed. by R. A. Meyers, New York: Academic Press, 751–769, third edition ed.

29. Pepelyshev, A. and A. S. Polunchenko (2017): "Real-time financial surveillance via quickest change-point detection methods," *Statistics and its Interface*, 93–106.

30. Pergamenchtchikov, S. and A. G. Tartakovsky (2018): "Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data," *Statistical Inference for Stochastic Processes*, 21, 217 – 259.

31. Pergamenchtchikov, S., Tartakovsky, A. G. (2019): "Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis," *Journal of Multivariate Analysis*, 174.

32. Polunchenko, A. S. and A. G. Tartakovsky (2010): "On Optimality of the Shiryaev-Roberts Procedure for Detecting a Change in Distribution," *The Annals of Statistics*, 38, 3445–3457.

33. Pospisil, T. and A. B. Lee (2018): "RFCDE: Random Forests for Conditional Density Estimation," *arXiv preprint arXiv:1804.05753*.

34. Probst, P., M. N. Wright, and A.-L. Boulesteix (2019): "Hyperparameters and tuning strategies for random forest," *WIREs Data Mining and Knowledge Discovery*, 9, e1301.

35. Raab, C., M. Heusinger, and F.-M. Schleif (2020): "Reactive soft prototype computing for concept drift streams," *Neurocomputing*, 340–351.

36. Roberts, S. (1966): "A Comparison of Some Control Chart Procedures," *Technometrics*, 411–430.

37. Shannon, C. (1948): "A mathematical theory of communication," *Bell System TechnicalJournal*, 27, 379–423.

38. Shiryaev, A. (1963): "On Optimum Methods in Quickest Detection Problems," *Probability Theory Application*, 8, 22–46.

39. Shiryaev, A. N. (1961): "The problem of the most rapid detection of a disturbance in a stationary process," in *Soviet Math. Dokl*, vol. 2, 103.

40. Smimou, K. and W. Khallouli (2015): "Does the Euro affect the dynamic relation between stock market liquidity and the business cycle?" *Emerging Markets Review*, 25, 125–153.

41. Tajeuna, E. G., M. Bouguessa, and S. Wang (2023): "Modeling Regime Shifts in Multiple Time Series," *ACM Transactions on Knowledge Discovery from Data*, 17.

42. Tartakovsky, A. and V. Spivak (2021): "Analysis of the Weighted Shiryaev–Roberts Procedure in the Problem of Changepoint Detection for Models with Unknown Post Change Parameters." *Works of MIPT, Computer Science and Management*, 13.

43. Tartakovsky, A. G. (2016): "On Asymptotic Optimality in Sequential Changepoint Detection: Non-iid Case," ArXiv:1510.03827.

44. Wang, Shusen, e. a. (2023): "Unsupervised Concept Drift Detection for Time Series on Riemannian Manifolds," *Journal of the Franklin Institute*, 13186–13204.

45. Yang, Y., L. Zhao, Y. Zhu, L. Chen, G. Wang, and C. Wang (2023): "Spillovers from the Russia-Ukraine conflict," *Research in International Business and Finance*, 66, 102006.

46. Zhao, H., J. Li, Y. Lei, and M. Zhou (2022): "Risk spillover of banking across regions: Evidence from the belt and road countries," *Emerging Markets Review*, 52, 100919.

Early warning systems for financial markets of emerging economies*

Artem Kraevskiy[†] Artem Prokhorov[‡§¶] Evgeniy Sokolovskiy[‖]

August 2024

Abstract

Financial markets of emerging economies are particularly vulnerable to extreme and cascading information spillovers. We develop and apply a new online early warning system (EWS) for what is known in machine learning as concept drift, in economics as a regime shift and in statistics as a change-point in information transfer. The system is adapted to capturing the specifics of financial information flows in emerging markets. It goes beyond linearity assumed by many conventional methods, and it is robust to heavy tails and tail-dependence in the data, making it particularly suitable for volatile financial markets. The key component is an effective change-point detection mechanism based on the conditional entropy as a measure of information transmission, rather than on a particular feature of the data such as the conditional mean or variance. Combined with recent advances in machine learning techniques for high-dimensional random forests, the mechanism is capable of finding significant shifts in information transfer between interdependent time series when traditional methods fail. We explore when this happens using simulations and we provide two illustrations when the method generates early warnings in emerging financial markets.

Keywords: emerging financial markets, early warning, copula, local linear forest, machine learning, Shiryaev-Roberts statistic, spillover effects

*An early version of the paper was presented at iCEBA23 and META23 and the authors are grateful for feedback received from the participants. Prokhorov's research was supported by a grant from the Russian Science Foundation (№ 20-18-00365).

[†]National Research University Higher School of Economics, Moscow, Russia
[‡]The University of Sydney Business School, Sydney, Australia
[§]CEBA, St.Petersburg State University, Russia
[¶]CIREQ, University of Montreal, Canada
[‖]National Research University Higher School of Economics, Moscow, Russia

1 Introduction

Financial markets of emerging economies are characterized by unusually high volatility, structural breaks, non-linear dependence, asymmetry and strong correlation in extremes also known as tail-dependence (see, e.g., Chaudhuri and Wu, 2003; Chen and Ibragimov, 2019; Mamonov et al., 2024). Moreover, modern financial markets generally have a complicated structure of information flows often represented by various global and local spillover and contagion effects, particularly strongly affecting emerging markets and manifesting themselves with significant lags and frictions in information transfer (see, e.g., Zhao et al., 2022; Khalfaoui et al., 2023; Ahmad et al., 2018). In this setting, rapid and reliable detection of shifts in information channels at work in emerging markets has been found to be crucial for ensuring stability of the markets, effective economic policy and prevention of financial contagion (see, e.g., Neaime, 2016; Smimou and Khallouli, 2015; Yang et al., 2023).

In statistics literature, the area that develops principled ways of constructing early warning systems (EWS) which react to changes in a stochastic process with the shortest delay, is known as change-point detection. This literature tracks back to Shiryaev (1961, 1963) and Roberts (1966), among others, and has by now accumulated an impressive set of optimality results (see, e.g., Chen, 2019; Chu and Chen, 2019; Pergamenchtchikov and Tartakovsky, 2018, 2019). Importantly, various versions of the so called Shiryaev-Roberts (SR) statistic have been shown to be optimal in terms of minimizing the detection delay as the probability of false alarm approaches zero, for a wide range of stochastic processes including autocorrelated time series of financial returns with unknown pre- and post-change distribution. This has spurred the development of online financial surveillance systems based on the SR statistic where "online" means real-time (see, e.g., Pepelyshev and Polunchenko, 2017).

Adapting these methodologies for use in emerging markets has been lagging. In econometrics, most of what is known as structural breaks methodology has been "offline," that is, it uses historical time series to find regime shifts (see, e.g., Hamilton, 2016; Bai and Perron, 1998, 2003). The focus has been on

obtaining consistent tests for the number and location of regime shifts in the available data, rather than on sequential monitoring for change points. The conventional econometric methodology often assumes the number of regime shifts to be known or to be within a fixed region. It also imposes restrictions on the minimum duration of regimes and on distributional characteristics of time series such as finite higher-order moments. Such restrictions are particularly problematic when applied to volatile indicators of emerging markets.

Moreover, traditionally macroeconomic analysis has focused on linear models of information transfer, represented in such concepts and models as Granger (1969) causality, vector autoregression (VAR) and dynamic conditional covariance (DCC) models, which use primarily parametric specifications and may suffer from large misspecification biases. Combined with the offline nature of such methods, their applicability in the setting of emerging economies seems questionable. At the same time, comparisons of performance between online and offline methods requires re-designing the offline methods to make them applicable in an online environment for which they were not intended. Therefore, there is little guidance in the literature on choosing between these methodologies.

Meanwhile, the field of machine learning has seen remarkable advances in areas related to change-point detection. Known as concept drift or data shift detection, this field of machine learning recognizes that the entire conditional distribution of the target variable given explanatory variables may change, drastically affecting predictive capacity of a machine learning engine trained on data from that distribution (see, e.g., Hoens et al., 2012; Chen et al., 2022). This brings to the fore the recent successes of such advanced machine learning methods as random forests and artificial neural networks in estimating flexible nonparametric conditional distributions and conditional expectation functions with many variable in the conditioning set (see, e.g., Breiman, 2001; Pospisil and Lee, 2018; Friedberg et al., 2021).

Much relevant work is now happening in the intersection of machine learning and statistics. For example, Barros et al. (2018) develop a statistical estimator for concept drift detection based on Wilcoxon

rank sum test. Baidari and Nagaraj (2021) apply the Bhattacharyya (1943) test to online concept drift detection. Wang et al. (2023) uses shifts in estimated covariance matrices to detect change points, focusing attention on linear dependence; Ditzler and Polikar (2011) use shifts in the target variable distribution over time. Bifet and Gavalda (2007) do the same but focus on developing a method of adaptive windowing (ADWIN) for high-frequency data stream monitoring. Raab et al. (2020) extend ADWIN to use the Kolmogorov-Smirnov statistic (KSWIN) and provide a more stable and precise estimation method. Tajeuna et al. (2023) develop methods for finding regime shifts in multivariate time series via analyzing the co-evolving time series representation called mapping grid, while Guo et al. (2022) and Alanqary et al. (2021) develop new ways of measuring changes in statistical properties of multivariate distributions. James et al. (2023) develop a distribution-free market surveillance methodology based on dynamic time warping. In spite of these impressive developments, none of the methodologies have been adapted to the settings commonly encountered in empirical analysis of emerging markets.

The goal of this paper is to advance the early warning methodology by combining advances in statistics and machine learning with application in economics and finance, in particular in settings with highly volatile and tail-dependent data. Specifically, we design a new version of the change-point detection methodology based on conditional entropy. The novelty here is that information transfer is no longer assumed to be linear and it incorporates regime shifts in any feature of the distribution, not only the conditional mean or variance. In fact, the mean and variance of the underlying series may not even be well-defined, which is not an uncommon setting in the analysis of extremely heavy-tailed financial time series, e.g., during financial crises and contagion periods. The EWS we design is sensitive to any shift in the conditional entropy estimated using state-of-the-art methods from machine learning. The conditional entropy relates the present state of the target variable to the information in the explanatory variables not captured by the lags of the target variable. Therefore, we directly measure the amount of uncertainty that cannot be explained either by the lags of the target or by explanatory variables without restricting

the relationships between the variable to be linear and parametric, and without focusing on a specific feature of the target distribution.

As a starting point we take the conventional linear specification of the kind commonly used in empirical economics for Granger causality testing, and we convert it into a sequence of conditional entropy estimates. This permits an online sequential SR test based on entropy. To our knowledge, the use of conditional entropy in the SR statistic is novel and of interest in its own right. We show how this approach helps model the stylized facts characterizing emerging markets data such as intensive information transmission, transfer termination and inversion (see, e.g., Sahiner, 2024; Hassan et al., 2024).

Then, we develop extensions of this baseline setting that take advantage of local linear forests and of rank-based estimation. These extensions acknowledge the presence of non-linearities and tail-dependence in emerging market data, and they permit infinite moments in the distribution of financial variables. By using ranks of the data, rather than the data itself, we introduce the concept of statistical copulas in the study of change-points, which, to our knowledge, has not been done before. Copulas capture rank-invariant non-linear dependence in the data even if the marginal distributions are extremely heavy tailed and correlations and covariances are inappropriate dependence measures — a feature that has prompted the growing popularity of copulas in modeling emerging markets data (see, e.g., Mamonov et al., 2024; Tian et al., 2023).

We present simulation results demonstrating how the new approach dominates a direct application of the SR test to the target variable or to AR process residuals in settings typical for emerging market studies. Specifically, we outline the cases when using ranks allows us to uncover regime shifts that cannot be detected by the conventional SR statistic and when the other methods of change-point detection fail or perform suboptimally. Additionally, the empirical section provides illustrations of how the proposed EWS could have issued alarms for the Russian equity market ahead of the Ukraine crisis and for the South African equity markets ahead of the COVID-19 recession.

The paper is organized as follows. Section 2 sets the stage by defining notation and concepts such as entropy, random forests and change-points. Section 3 describes the baseline version of the proposed procedure. Section 4 describes extensions of the baseline method that account for non-linearity and use ranks. Simulations are provided in Section 5, and Section 6 contains real data examples. Section 7 concludes. All the codes and data for this paper are available at `https://github.com/kraevskiyAA/EWS_CondEnt`.

2 Preliminaries and related literature

2.1 Entropy and information transfer

Suppose we have a target time series Y_t and a history of explanatory variables available at time t, $\mathcal{X}_t = \{X_{t-1}, X_{t-2}, \ldots, X_{t-l}\}$, where each X_s, $s = t-1, \ldots, t-l$, is a matrix containing d explanatory variables (or their lags). As an example, Y_t can be a composite stock index of an emerging market and $\mathcal{X}_t$ can contain histories of commodity prices the country trades in. Our goal is to detect shifts in information transfer between the target and explanatory variables, causing an early warning about the state of the market described by the target variable.

It is common in information theory literature to evaluate the intensity of information transfer from $\mathcal{X}$ to Y by *conditional entropy of* Y *given* $\mathcal{X}$ (see, e.g., Orlitsky, 2003) which can be written as follows:

$$H = -\int_{\mathcal{X},\mathcal{Y}} f(y,x)\cdot\log f(y|x)dydx = -\int_{\mathcal{X}} f(x)\int_{\mathcal{Y}} f(y|x)\cdot\log f(y|x)dydx = \mathbb{E}_X\mathbb{E}\log f(y|x), \tag{1}$$

where $f(y|x)$ is the conditional density of Y given $\mathcal{X}$, $f(y,x)$ is the joint density of $(Y,\mathcal{X})$, and $f(x)$ is the marginal density of $\mathcal{X}$. The last equality follows from the definition of expectations and the law of interated expectations.

Fundamentally, the concept of entropy measures the uncertainty about Y_t given the information $\mathcal{X}_t$. In

the context of signal transmission and unconditional distributions, Hartley (1928) is credited for observing that the information to be gained from n outcomes each having a different probability p_i can be defined as

$$-\sum_{i=1}^{n} p_i \log_2 p_i.$$

This has become known as the Shannon entropy (Shannon, 1948). The higher this value, the higher is the uncertainty about the n realizations.[1] The largest amount of uncertainty corresponds to uniformly distributed outcomes, i.e., the same p_i for all i. Eq. (1) can be viewed as a generalization of this concept to continuous conditional distributions, showing the average amount of information to be supplied before one can communicate the information in Y given that the other party knows the information in $\mathcal{X}$. Yang (2018) provides an intuitive introduction to information theory for economists with a focus on entropy.

2.2 Nonparametrics and random forests

An important component of computing H is estimation of the conditional density. A flexible way of estimating $f(y|x)$ without making strong distributional assumptions is to use the tools from nonparametric econometrics (see, e.g., Li and Racine, 2023). However, most of the standard methods will not work with conventional sample sizes when the dimension of $\mathcal{X}_t$ is even moderately large. This is known as the curse of dimensionality. Additionally, the nonparametric methods are known to perform poorly in the extremes of the distribution, which are of particular interest when dealing with data from emerging markets.

An alternative approach that has been gaining popularity but has not been explored in the context of entropy-based EWS is to use machine learning techniques. As a prominent example, Pospisil and Lee (2018) recently proposed to use random forests for conditional density estimation (RFCDE). The

[1] A famous and straightforward example is a coin flip: the uncertainty is limited to two outcomes, head or tail. The information to be gained from a flip, known as a bit, is binary and can be defined by $\log_2(2) = 1$. Then, the information to be gained from n flips is $\log_2(2)^n = n$. Intuitively, entropy is the level of uncertainty, or disorder, inherent in random outcomes.

estimator improves on a standard nonparametric method of kernel density estimation (KDE) by employing adaptive weights obtained using random forests of Breiman (2001). The main advantage here is that random forests break the curse of dimensionality by using random subsamples of the entire set of conditioning variables. This allows RFCDE to achieve consistent conditional density estimation with low computational costs and high accuracy.

We use the RFCDE to obtain consistent estimates of $f(y|x)$. It is then intuitive that a spike in H evaluated using the RFCDE of $f(y|x)$ should indicate an increase in uncertainty (a decrease in the information transfer from $\mathcal{X}$ to Y) and the remaining question is how to identify significant spikes. We note that the RFCDE has not been used to obtain conditional entropy estimates before and offers a new way of evaluating information transmission in high-dimensional settings.

2.3 Change-point detection

Traditional change-point detection literature, going back more than 60 years (Page, 1954; Shiryaev, 1963; Roberts, 1966), looks at an aggregated sequence of likelihood ratios based on the likelihood of a change at time t versus the likelihood of no change, over a history of observations, where the densities before and after the change are known but the change-point is not. A recent semiparametric application to spillovers in conditional correlations between developed and emerging markets is Barassi et al. (2020). Now there exist important relaxations of the assumptions of known densities and operational methods that look at a sliding window of data rather than the entire history. However, this literature has not looked at entropy- or rank-based evaluation of information transfer, not to mention the settings of emerging markets.

Define the Shiryaev-Roberts statistic as follows:

$$SR_t = \sum_{k=1}^{t}\prod_{i=k}^{t}\frac{f_0(\xi_i)}{f_\infty(\xi_i)}, \tag{2}$$

where ξ_i, $i = 1, \ldots, t$, are generic independent random variables or vectors, f_0 is the density of ξ_i before

the change and f_∞ is the density of ξ_t after the change. Then, it is easy to derive the recursive version

$$SR_t = (1 + SR_{t-1})\frac{f_0(\xi_t)}{f_\infty(\xi_t)}. \tag{3}$$

We use a version of the weighted SR statistic proposed by Tartakovsky and Spivak (2021). Their procedure permits randomness in the drift of ξ_t in the sense that at each t the change may take one of m values and the weighted statistic has the form

$$SR_t^w = \frac{1}{m}\sum_{i=1}^{m} SR_t(\Delta\mu_i), \tag{4}$$

where $SR_t(\Delta\mu_i)$ is the value of SR_t when the difference between the mean of ξ_t before and after the change is $\Delta\mu_i$.[2] This affords a much better performance in cases when distribution parameters are unknown. Other than that, the construction of the SR_t statistic is standard (see, e.g., Pepelyshev and Polunchenko, 2017).

An alarm will sound when SR_t exceeds a threshold value that controls the false alarm rate and, according to a series of important results in statistics, the SR procedure is optimal in the sense of minimizing the average detection delay (see, e.g., Polunchenko and Tartakovsky, 2010).

3 The new procedure: baseline version

The core of the new procedure is application of the Shiryaev-Roberts statistic to a sequence of entropy estimates of H assessed on a moving window of length Δ. The window length needs to be short enough so that the regime shift is noticeable but long enough not to pick noise (we use window lengths of 50 to 100 observations in practice). An EWS signal is triggered when SR_t^w evaluated using the entropy process H_t surpasses a threshold $A > 0$, indicating a significant shift in information transmission from $\mathcal{X}_t$ to Y_t. In this section we discuss how we construct the sequence, statistic, and threshold.

[2] We use $m = 6$ shifts in the mean with a fixed step $\Delta\mu$ equal to the range of data divided by m.

We start by fitting an autoregressive model for the target variable Y_t on a moving window of length Δ:

$$Y_t = \hat{\alpha}_0 + \sum_{s=1}^{k} \hat{\alpha}_s Y_{t-s}, \tag{5}$$

where we use the Bayesian Information Criterion (BIC) to determine the optimal AR order k. This gives the part of Y_t orthogonal to the history of the target variable. This is the unexpected component of Y_t we are interested in, which we write as follows:

$$e_t = Y_t - \hat{Y}_t. \tag{6}$$

The AR(k) residuals e_t capture any remaining news in the target variable Y_t and we wish to explore how the distribution of this variable changes depending on the explanatory variables $\mathcal{X}$. We wish to remain agnostic about the specific moments of the distribution that change and about the form of the distribution.

As the next step, we obtain the fitted value from the regresion of e_t on lags of the explanatory variables contained in $\mathcal{X}$ using the same moving window:

$$\hat{e}_t = \hat{\beta}_0 + \sum_{j=1}^{d}\sum_{s=1}^{l} \hat{\beta}_s^{(j)} X_{t-s}^{(j)}, \tag{7}$$

where $X_{t-s}^{(j)}$ is one of the d available explanatory variables, lagged s times, $s = 1, \ldots, l, j = 1, \ldots, d$. Again, the optimal lag l is selected by BIC. Clearly, these two steps are equivalent to a one-step regression of Y_t on both its own history and history of the covariates. This corresponds to the familiar concept of (linear) Granger causality, which could be detected by a joint significance test of $\hat{\beta}_s^{(j)}$ (see, e.g., Balcilar et al., 2022, for a recent example of Granger causality analysis for emerging markets).

In empirical finance, it has been noted at least since Hiemstra and Jones (1994), that linear causality is too restrictive. A standard response was to include nonlinear functions of the explanatory variables. This limits the nonlinearity to the specific functional forms of X. We go beyond the Granger causality

by computing the conditional entropy of e_t given $\hat{e}_t$. This recognizes the fact that change-point detection should be based on the conditional distribution $f(e_t|\hat{e}_t)$ rather than the linear regression $\mathbb{E}(e_t|\hat{e}_t)$, thus accommodating more complex (than linear in mean) patterns of dependence.

Let $\hat{f}$ denote the RFCDE of the conditional density of e given $\hat{e}$ obtained using the moving window of size Δ. Let $\hat{\mathcal{E}}_{[t_0,t_1]} = \{\hat{e}_{t_0}, \hat{e}_{t_0+1}, \ldots, \hat{e}_{t_1}\}$ denote the set of $\hat{e}$'s that fall into the window. Then, the estimator of aggregate information transfer over the window can be written as a sum of the entropy contributions from the values of $\hat{e}$ that fall into the window:

$$\hat{H}_{[t_0,t_1]} = \sum_{\hat{e}\in\hat{\mathcal{E}}_{[t_0,t_1]}} \left[\int_{\mathbb{R}} \hat{f}(\varepsilon|\hat{e}) \cdot \log \hat{f}(\varepsilon|\hat{e}) d\varepsilon\right]. \tag{8}$$

We can approximate the integrals in (8) using stochastic or deterministic integral approximation, e.g., Simpson's rule, which is a deterministic method and our preference in simulations and applications. We are interested in early identification of spikes in $\hat{H}_{[t_0,t_1]}$ as we move across time. Denote the sequence of entropy estimates over a sliding window of size Δ as follows

$$\hat{H}_{[t_0,t_0+\Delta]}, \hat{H}_{[(t_0+1),(t_0+1)+\Delta]}, \cdots$$

We note that the step between consecutive values of H in the sequence can be larger than one time period but this will delay detection while expediting computation. The sequence of H is the time series to which we apply the SR statistic in the baseline version of our approach.

Algorithm 1 outlines the Shiryaev-Roberts procedure for entropy-based change-point detection described above. The algorithm uses several inputs besides the sequence of H, namely, the set of shift values $M_\Delta = \{\Delta\mu_1, \ldots, \Delta\mu_m\}$, the threshold A and the smoothing parameters α and β. Although there is no strict rule for selecting the value of threshold A, it is customary to consider two minimization criteria, namely, average detection delay (ADD) defined as $\mathbb{E}_\theta[\tau - \theta|\tau > \theta]$ and probability of false alarm (PFA) defined as $\mathbb{P}_\theta(\tau < \theta)$, where θ is the actual moment of the change and $\tau = \arg\min_t\{t|SR_t > A\}$

is the moment of time it is detected.

Pergamenchtchikov and Tartakovsky (2018, 2019) show that the weighted SR procedure approximately minimizes ADD among all change-point detection procedures with a given PFA within a window of a fixed length for non-i.i.d. data, including autoregressions and conditional heteroskedasticity GARCH-type models. However, these quantities are generally unavailable to the researcher in practice, unless one deals with synthetic data and is able to generate variables using a given constellation of these parameters. For instance, in Section 5 where we report simulation results, we are able to set A by choosing the values of ADD and PFA via the numerous relationships derived in the literature (see, e.g., Tartakovsky, 2017). In applications, we can use a grid of values for A.

Smoothing parameters α and β are used at the end of each iteration of Algorithm 1 in order to produce an exponential smoothing update of the conditional expectation and variance of H, balancing the past values and the new information. The natural value for α is 0.5, reflecting equal weights on the old value of the mean $\hat{\mu}_{t-1}$ and the current H_t, while for the variance update, we choose $\beta = 0.9$, putting more weight on the estimate of $\hat{\sigma}$ from the previous period and thus dampening the variance update. Larger values of α and β result in smoother trajectories of the moments of H.

Algorithm 1: Shiryaev-Roberts procedure for change-point detection in entropy

Data: Time series $H_1, H_2, \ldots, H_n$; set of possible shifts in mean M_Δ; threshold A

$SR_0 = 0; \hat{\mu}_1 = H_1; \hat{\sigma}_1^2 = 1;$

while $2 \leq t \leq n$ **do**

  $\xi_t = \frac{H_t - \hat{\mu}_{t-1}}{\hat{\sigma}_{t-1}};$

  for $\Delta\mu_i \in M_\Delta$ **do**

    $SR_t^{(i)} = (1 + SR_{t-1}^{(i)}) \cdot \exp\{\Delta\mu_i(\xi_t - \frac{\Delta\mu_i}{2})\};$

  end

  $SR_t^w = \frac{1}{m}\sum_{i=1}^{m} SR_t^{(i)}$

  if $SR_t^w > A$ **then**

    $\tau = t;$

    break

  end

  $\hat{\mu}_t = \alpha\hat{\mu}_{t-1} + (1-\alpha)H_t;$

  $\hat{\sigma}_t^2 = \beta\sigma_{t-1}^2 + (1-\beta)(H_t - \hat{\mu}_t)^2;$

  $t = t + 1;$

end

Result: change-point τ

4 Extensions to non-linear and rank-based estimation

The baseline procedure described in the previous section uses a conditional density estimator of the AR process residuals e_t given $\hat{e}_t$, where $\hat{e}_t$ represents the linear span of the lagged explanatory variables $X_t^{(j)}$ in Eq. (7). This aligns with the traditional view of the Granger causality, which identifies information transfer by testing joint significance of the lagged regressors in a linear autoregressive model. We went

beyond the traditional approach by constructing a flexible, kernel-based estimator $\hat{f}(e_t|\hat{e}_t)$ and by applying the SR statistic to a moving window of entropy estimates $\hat{H}_{[t_0,t_1]}$ obtained using $\hat{f}(e_t|\hat{e}_t)$. The goal of this section is to adapt this new method to an even more flexible non-linear specification, recognizing the well documented presence of non-linearities in financial market data (see, e.g., Caporale et al., 2020).

4.1 Local linear forest

The univariate conditioning in $\hat{f}(e|\hat{e})$ substantially simplified the task of nonparametric density estimation. The alternative of conditioning on the entire set of covariates $X_{t-s}^{(j)}$, without obtaining $\hat{e}_t$, leads to a nonparametric estimator $\hat{f}(e_t|\mathcal{X}_t)$ which suffers from a severe curse of dimensionality due to the high dimensionality of $\mathcal{X}_t$, making the estimate very imprecise. The problem is further exacerbated by the trade-off between quick detection delay and accuracy of density estimation. It is well known in the change-point detection literature that if we increase the rolling window, making the density estimation more precise, the time before change-point detection increases. Thus, the standard multivariate nonparametric detection procedures are either inaccurate or slow.

To overcome this issue without increasing the dimensionality of density estimation, we use a recently developed method of local linear forest (LLF) proposed by Friedberg et al. (2021). LLF constructs an estimate of the conditional mean $\hat{e}_t = \mathbb{E}[e_t|X_{t-1} = x_{t-1}, \ldots, X_{t-d} = x_{t-d}]$ achieving a faster convergence rate than traditional nonparametric estimators of the same dimensionality, and preventing the overfitting. The standard (non-local) random forest estimator (see, e.g., Breiman, 2001; Probst et al., 2019) is sensitive to hyperparameter selection and tends to overfit. The use of LLF improves upon the linear Granger causality regression (7) and is new to financial econometrics.

Specifically, our first extension of the baseline methodology is to use the flexible LLF estimator $\hat{e}_t$ instead of the fitted value $\hat{e}_t$ in constructing $\hat{f}(e_t|\hat{e}_t)$. The SR statistic is now based on a sequence of the

new estimates of H on a sliding window $[t_0, t_1]$:

$$\hat{H}_{[t_0,t_1]}^{LLF} = \sum_{\tilde{e} \in \Psi_{[t_0,t_1]}} \left[\int_{\mathbb{R}} \hat{f}(\varepsilon|\tilde{e}) \cdot \log \hat{f}(\varepsilon|\tilde{e}) d\varepsilon \right], \tag{9}$$

where $\Psi_{[t_0,t_1]} = \{\tilde{e}_{t_0}; \tilde{e}_{t_0+1}; \ldots; \tilde{e}_{t_1}\}$. The rest of the methodology remains unchanged.

We note that both $\hat{H}_{[t_0,t_1]}$ and $\hat{H}_{[t_0,t_1]}^{LLF}$ permit non-linearity in the relationship of Y_t on $\mathcal{X}_t$ because both use RFCDE to construct the conditional density of e_t. The advantage of using the new estimates $\tilde{e}_t$ is that it implies a non-linear conditional mean of Y_t in terms of $\mathcal{X}_t$, given lags of Y_t. Therefore, it aligns with the notion of non-linear Granger causality where $\mathcal{X}_t$ Granger-causes Y_t if the nonparametric estimator $\tilde{e}_t$ is indistinguishable from a constant in each dimension of $\mathcal{X}_t$. Unfortunately, powerful statistical tests of this null hypothesis in a setting with many covariates in $\mathcal{X}_t$ are not available. Importantly, the procedure we propose does not require such a testing.

4.2 Copulas

Our second extension addresses the problem of heavy tails and tail-dependence, in addition to the problem of non-linearity. It is well documented that financial markets of emerging economies are characterized by asymmetric heavy-tailed distribution, which make application of standard econometric methods problematic (see, e.g., Chen and Ibragimov, 2019). As an alternative we follow the growing literature on copula-based estimation in finance (see, e.g., Ibragimov and Prokhorov, 2017) and re-formulate the problem of density estimation using ranks instead of actual values. It is important to note that for any set of random variables, their copula function contains all information about the dependence structure between the random variables as well as between any rank-preserving transformations of the random variables (see, e.g., Cherubini et al., 2011, for a survey of copula methods in finance). Therefore, if we use the copula instead of the original data distribution in the baseline method, the structure of information transfer will remain unchanged. Meanwhile, the estimation is now over ranks and H uses conditional

copula density, rather than conditional density. This makes the procedure robust to heavy tails in the marginal distributions and well-defined even when moments of the original distributions are infinite.

In the proposed modification, the baseline RFCDE estimator $\hat{f}(e_t|\hat{e}_t)$ is replaced by the rank-based RFCDE estimator $\hat{f}(u_t|\hat{u}_t)$, where

$$u_t = \frac{1}{\Delta}\sum_{s=1}^{\Delta} \mathbb{1}\{e_{t-s} \leq e_t\}, \qquad \hat{u}_t = \frac{1}{\Delta}\sum_{s=1}^{\Delta} \mathbb{1}\{\hat{e}_{t-s} \leq \hat{e}_t\},$$

where, as before, Δ is the length of the sliding window over which we estimate the densities. The values of u_t and $\hat{u}_t$ are known as pseudo-observations; they are values between $[0, 1]$ corresponding to the ranks of observations e_t and $\hat{e}_t$, respectively, in the sliding window, divided by the number of observations in the window. The corresponding entropy estimates can be written as follows

$$H_{[t_0,t_1]}^{rank} = \sum_{u \in \mathcal{U}_{[t_0,t_1]}} \left[\int_0^1 \hat{f}(w|\hat{u}) \cdot \log \hat{f}(w|\hat{u}) dw \right], \tag{10}$$

where $\mathcal{U}_{[t_0,t_1]} = \{\hat{u}_{t_0}; \hat{u}_{t_0+1}; \ldots; \hat{u}_{t_1}\}$ and $\hat{f}(u|\hat{u})$ can be interpreted as the conditional copula density of e_t given $\hat{e}_t$. The remaining parts of the procedure are unchanged.[3]

We note that rank-based estimates such as (10) are invariant to rank-preserving transformations of variables and thus make heavy tails in marginal distributions irrelevant. This permits additional focus on shifts in conditional distributions when the first two moments of the data are not even estimable as would be the case during extreme market conditions such as financial crises and contagion episodes.

5 Simulation experiments

5.1 Simulation designs

In this section we compare the performance of the baseline and extension methods with alternative procedures and among themselves, using synthetic data, representative of real-life market data from

[3] We also tried the LLF estimates $\tilde{e}_t$ instead of $\hat{e}_t$ but do not report this case for brevity.

emerging economies. We use two common scenarios for the information transmission literature, namely, termination of transfer and inversion of transfer. The two scenarios correspond to the situations on emerging markets known as sudden stops (see, e.g., Hassan et al., 2024) and spillover direction changes (see, e.g., AuYong et al., 2004). Examples of the two designs we use are as follows:

(a) Termination of information transfer: at a change-point, the information transfer from X and Z to Y stops. We consider the following data generating process (DGP):

$$Y_t = \begin{cases} 0.5 \ln X_{t-1} + 0.2 Z_{t-1}^2 + 0.3\nu_t, & t < 500 \\ \xi_t, & t > 500 \end{cases}$$

where $\nu_1, \ldots, \nu_{500} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\xi_{501}, \ldots, \xi_T \overset{i.i.d.}{\sim} \mathcal{N}(5, 36)$ $X_1, \ldots, X_T \overset{i.i.d.}{\sim} Exp(3)$, and $Z_1, \ldots, Z_T \overset{i.i.d.}{\sim} \Gamma(3,1)$.

(b) Inversion of information transfer: at a change-point, the information flow from X and Z to Y is reversed. The DGP we use is as follows

$$Y_t = \begin{cases} 0.5 \ln X_{t-1} + 0.2 Z_{t-1}^2 + 0.3\nu_t, & t < 500 \\ \xi_t, & t > 500 \end{cases}$$

where

$$X_t = \begin{cases} v_t, & t \leq 500 \\ 0.3 Y_{t-1} + 0.1\psi_t, & t > 500 \end{cases}$$

$$Z_t = \begin{cases} \delta_t, & t \leq 500 \\ 0.6 Y_{t-1}^2 + 0.1\zeta_t, & t > 500 \end{cases}$$

and where $\nu_1, \ldots, \nu_{500}, \psi_{501}, \ldots, \psi_T, \zeta_{501}, \ldots, \zeta_T \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\xi_{501}, \ldots, \xi_T \overset{i.i.d.}{\sim} \mathcal{N}(5, 9)$ and $v_1, \ldots, v_{500} \overset{i.i.d.}{\sim} Exp(3)$, $\delta_1, \ldots, \delta_{500} \overset{i.i.d.}{\sim} \Gamma(3,1)$.

The summary statistics of simulated data using the two DGPs are presented in Table 1. The data are

leptokurtic, asymmetric and weakly dependent both in levels and in ranks, which are generally typical characteristics of financial return data (see, e.g., Cont, 2001).

regime	mean	std	skewness	kurtosis	corr(Y_t; Y_{t-1})	rank corr
Information transfer termination						
before change	2.541	2.768	-2.106	5.812	-0.062	-0.040
after change	5.126	5.968	0.110	0.241	-0.076	-0.080
Information transfer inversion						
before change	2.541	2.768	-1.527	2.973	0.040	-0.040
after change	5.126	2.99	-0.192	0.137	0.007	-0.064

Table 1: Summary statistics of synthetic data

5.2 Comparison with traditional procedures

We start by benchmarking the new approach against the conventional applications of the SR procedure, that is, without the use of entropy. The traditional alternatives we consider are as follows: (a) application of the SR statistic to detect a change in the mean of the target variable Y_t directly; (b) application of the SR statistic to detect a change in the mean of the residuals e_t from the AR model in Eq. (6); (c) application of the SR statistic to detect a change in the mean of the fitted values $\hat{e}_t$ from Eq. (7). We focus on termination of information transfer.[4]

The specific DGP we consider in this section is as follows:

$$Y_t = \begin{cases} 0.5 \ln X_{t-1} + 0.2 Z_{t-1}^2 + 0.1\nu_t, & t \leq 500 \\ \xi_t, & t > 500 \end{cases}$$

where $\nu_1, \ldots, \nu_{500} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\xi_{501}, \ldots \xi_T \overset{i.i.d.}{\sim} \mathcal{N}(2.35, 1.5)$, $X_1, \ldots, X_T \overset{i.i.d.}{\sim} Exp(3)$, and $Z_1, \ldots, Z_T \overset{i.i.d.}{\sim} \Gamma(3,1)$. The process is similar to termination of information flow in Section 5.1 except the mean shifts by a smaller magnitude, too small for the traditional procedure to capture. All parameters for computing SR statistics are the same for all implementations of the traditional SR statistics. The window length is $\Delta = 50$.

[4]The inversion of information transfer scenario is available upon request.

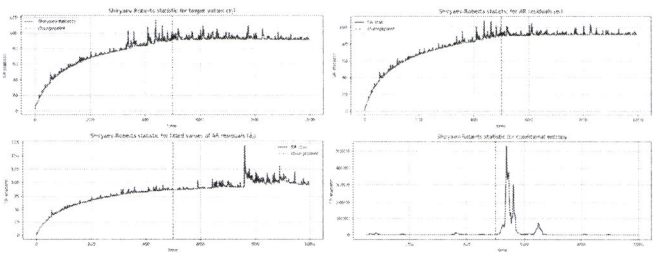


Figure 1: Shiryaev-Roberts statistics for termination of information transfer

Figure 1 shows the values of respective SR statistics and the location of the change-point (vertical dashed line). It is clear from the figure that the traditional ways of applying the SR procedure fail to detect the change point. The SR statistics for three of the four figures grows with time and flattens out near the change-point but no spikes are large enough to trigger an alarm. On the other hand, the entropy-based procedure shown on the lower right panel produces an immediate and strong alarm.

This difference in performance is due to the limitation of traditional approaches. When applied to specific variables, as opposed to the entropy, they seem to target shifts of specific nature, e.g., shifts in mean, variance, autocorrelation. In this case, the shift in mean is too small. A key advantage of our approach is that it captures jumps in *information transfer*. In this sense, instead of analyzing various shifts in SR statistics, each verifying the presence of a specific drift, the entropy-based measure directly addresses stability of the dependence structure.

5.3 Non-linearity and ranks

We now turn to detection of termination and inversion of information transfer using the extensions of the entropy-based approach, that is, using local linear forests and ranks. The simulation designs are

as described in Section 5.1. Simulation results are summarized in Figures (2)-(3), where θ denotes the moment of change, τ denotes the moment of change detection and A is the threshold value used for the Shiryaev-Roberts statistic. The window length is $\Delta = 50$.

It follows from Figures (2)-(3) that all extensions to the baseline entropy-based methods work very well in all four setups. There is no clear ordering in performance among the four setups: (a) baseline version (linear model, no ranks); (b) extension with LLF (local linear forest, no ranks); (c) extension with ranks of $\hat{e}_t$ (linear model, ranks); (d) extension with ranks of $\hat{e}_t$ (local linear forest, ranks). For both termination and inversion, all four are able to trigger an alarm within 7-15 time periods, using the respective thresholds; Table 2 summarizes this finding.

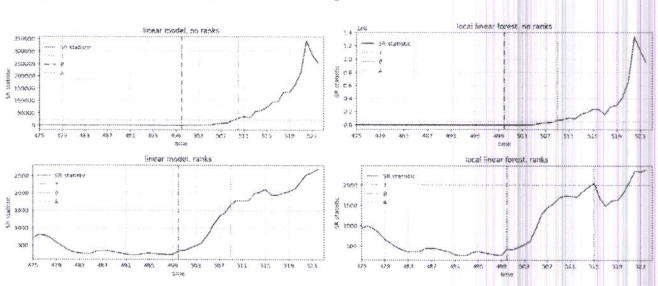


Figure 2: Shiryaev-Roberts statistic for termination of information transfer

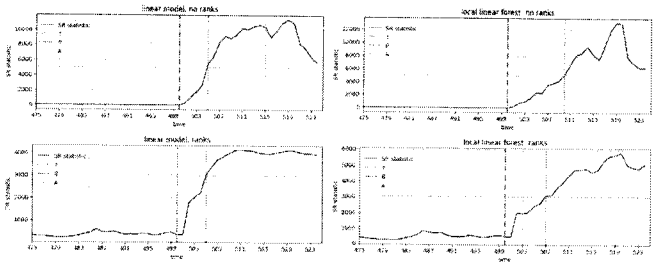


Figure 3: Shiryaev-Roberts statistic for inversion of information transfer

Interestingly, a comparison between the lower and upper panels of the figures suggests that PFA for ranks is higher than for levels, as can be seen by the non-zero values of SR before the break in the lower panels (and zero SR in the upper panels). However, this does not allow us to clearly identify cases when rank-based measures may or may not be preferred – something we address in the next subsection.

Model	Termination	Inversion
linear	10	5
LLF	9	10
linear + ranks	9	5
LLF + ranks	15	7

Table 2: Change-point detection delay ($\tau - \theta$)

5.4 Heavy tails and tail-dependence

To further investigate the benefits of rank-based measures we consider a setting where the heavy-tailed property of individual distributions prevents a productive use of such characteristics as the mean, variance or correlations. This setting is of particular relevance for financial data from emerging markets which are known to be extremely volatile (see, e.g., Chaudhuri and Wu, 2003; Chen and Ibragimov, 2019).

We modify the simulation design of the previous sections by including a component with heavy-tailed

data and tail dependence. For this purpose we add a penalty λ to a simplified version of termination of information flow as follows:

$$Y_t = \begin{cases} X_{t-1} + (1 - \sqrt{\lambda_{t-1}})\xi_t, & t \leq 500, \\ 0.9 + \xi_t, & t > 500, \end{cases}$$

where $\xi_1, ..., \xi_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.5)$, $X_1, ..., X_{500} \overset{\text{i.i.d.}}{\sim} WeibMin(1.5)$. The role of λ is to control the noise level depending on how close the observation is to the sample median. Let $d_t = X_t - \text{med}(X_t)$ denote such a distance. We define λ as follows:

$$\lambda_t = \frac{d_t - \min_t\{d_t\}}{\max_t\{d_t\} - \min_t\{d_t\}} \in [0, 1].$$

Clearly, the closer an observation is to the median the more noise Y_t has over X_{t-1}. For the observations closer to the sample extremes, Y_t is closer to being perfectly collinear with X_{t-1}. This produces non-zero tail dependence, that is, non-zero correlation between X_t and Y_t when they get closer to the extremes of their distributions. Additionally, we use the Weibull minimum extreme value distribution for X_t to induce the heavy-tailed property.

Figure 4 plots the conditional entropy processes H obtained using the four methods we consider: (a) the baseline linear model based on levels, (b) the local linear forest model based on levels, (c) the linear model based on ranks, (d) the local linear forest model based on ranks. The vertical dashed line indicates the change point θ.

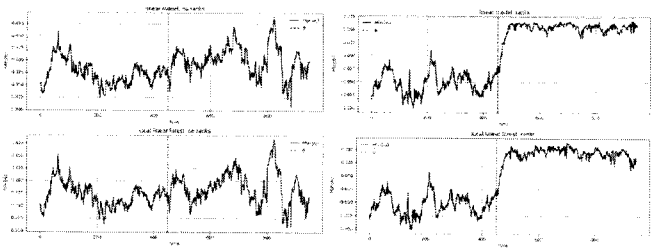


Figure 4: Estimates of entropy for heavy-tailed data with tail dependence

It is evident from Figure 4 even without computing the entropy-based SR statistic that this approach fails to detect the regime shift based on the original data (the upper and lower left panels do not show a visually significant change in H), while the rank-based EWS provides consistent results and rapidly detects the regime shift (the upper and lower right panels show a clear jump to a new mean of H after the change-point). This suggests that the new rank-based methodology is attractive in cases when the financial data are heavily leptokurtic and tail-dependent – a particularly relevant setting for emerging markets.

5.5 Comparison with other detection methods

Finally, we compare the proposed methods with non-SR approaches to concept drift detection from computer sciences such as Early Drift Detection Method (EDDM) and Adaptive Windowing Method (ADWIN) of Bifet and Gavalda (2007), and Kolmogorov-Smirnov Windowing Method (KSWIN) of Raab et al. (2020) and from economics such as an online version of the structural break detection method of Bai and Perron (1998). We focus on information transfer termination from Section 5.1.

Since EDDM, ADWIN and KSWIN are designed to operate on univariate time series they needed

to be adapted to be comparable with our methods. Similarly, the Bai-Perron test, being an offline test, needed to be turned into an online procedure. A perfectly fair comparison is impossible here. The compromise we use is as follows:

(a) For ADWIN, EDDM, and KSWIN, we split time series Y_t into non-overlapping segments of length 50, i.e., $[0, 50], [51, 100], \ldots$. For each segment we obtain residuals $e_t - \hat{e}_t$, where e_t and $\hat{e}_t$ are computed according to Eqs. (6) and (7), respectively. We then concatenate these segments and apply the relevant algorithms to it. We repeat this for 100 replications and compute PFA (probability of false alarm), ADD (average detection delay) and ND (fraction of non-detections).

(b) For the Bai-Perron test, we apply the test at the 1% significance level to an expanding window of data starting with $[0, 400]$ and expanding it to $[0, 401], [0, 402], \ldots$. We stop when the test rejects the null. We repeat this over 100 replications and compute PFA, ADD and ND.

(c) For our proposed methods, we apply them as described in Section 3 (baseline) and 4.1 (LLF), with window length 50, $\alpha = 0.5, \beta = 0.9, m = 1$. We repeat this over 100 replications and compute PFA, ADD and ND.

Metric	ADWIN	EDDM	KSWIN	Bai-Perron	baseline	LLF
PFA	-	-	0.89	0.66	0.1	0.09
ADD	-	-	8.5	23.6	10.1	10.4
ND	1	1	0.09	0	0.05	0.04

Table 3: Performance comparison with non-SR methods

The results are summarized in Table 3. Using this simulation design, both ADWIN and EDDM failed to detect the break, while KSWIN detected it in 91% of replications and produced a false alarm in 89% of replications. This implies that there are only 2% of replications where KSWIN detected the drift correctly. The Bai-Perron test detected all drifts, which is perhaps expected given the repeated testing problem implied by algorithm (b). It also produced 66% false positives. The proposed methods strongly dominate the alternatives in this exercise both in terms of PFA and ADD.

6 Real data examples

In this section we provide two examples of applying our approach to produce early warnings using financial data.

6.1 Russia: market uncertainty and the conflict in Ukraine

Many recent studies have found a significant impact of the Russian-Ukrainian conflict on global financial markets (see, e.g., Izzeldin et al., 2023). Less is known about information transmission in the time before the conflict broke out in February 2022. In this subsection, we explore whether the Russian stock market showed early warning signals that could have been detected using our approach. We consider the level of uncertainty captured by the conditional entropy – and the corresponding SR statistic – of RTSI, the main indicator of the Russian equity market, which is a capitalization-weighted average of the 50 most liquid Russian stocks traded on the Moscow Stock Exchange. RTSI is an U.S. dollar nominated index, known to attract foreign investors. The entropy is conditional on the information contained in key indicators of various international markets (commodity, equity, fixed-income) as well as key internal indicators. The list of variables is typical for models studying volatility spillover across financial markets of emerging economies and includes a stock market index of the emerging markets, a stock market index of a mature economy, an export-oriented exchange rate, a risk free rate and a government bond spread as well as an oil price and a few key domestic stocks dominated by internal investors (see, e.g., Assaf et al., 2023; Lin et al., 2024). The list is inevitably incomplete but sufficient to illustrate the main points. We provide details of the conditioning variables, their sources and summary statistics in the Appendices A.1-A.2.

We use a 100-day sliding window to compute the conditional entropy process H as described in Section 4.1, that is, we use the local linear forest in the estimation of the conditional mean $\hat{c}$ to allow for non-linearity and to mitigate the curse of dimensionality. The order of the autoregressions in Eq. (7) is picked using BIC and is allowed to vary for each sliding window. We conducted sensitivity analysis, running

the procedure for different number of the autoregressive order between 1 and 10 with no significance difference in the results. The number of shifts $m = 100$, the exponential smoothing parameters are $\alpha = 0.95$ and $\beta = 0.95$. The period for which we compute $\hat{H}$ is May 27, 2021 to April 12, 2022. It contains 273 daily observations prior to the start of the military conflict on February 24, 2022 and 47 observations thereafter.

Figure 5 plots the estimated conditional entropy and the corresponding SR statistic (on the log scale). Several observations from the figure are noteworthy. First, the conditional entropy reaches its peak precisely on February 24. Such precision is perhaps surprising, especially since the entropy measures the degree of disorder in RTSI after accounting for the information in a fairly rich conditioning set. Second, there is a rising trend in the SR statistic throughout the period, indicating a general increase in uncertainty. Third, the proposed method detects two change-points where the information flow changed significantly, at three months and at one month before February 24. Given the upward trend in SR in this example, we can think of the two change-points as specifying two levels of alarm intensity, more and less conservative. We stress that nothing in the procedure targets the detection of the start of the conflict. In fact, the period following the second change-point extends to the time of the first waves of sanctions, the interest rate hikes and capital controls that followed after February 24. Yet, the RTSI-based entropy captures the accumulated market uncertainty of a preceding period, with the peak on February 24 and two warnings, starting more than a quarter earlier.

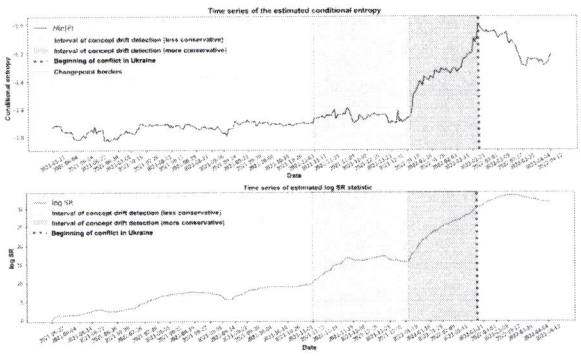


Figure 5: RTSI-based conditional entropy and log SR statistic

6.2 South Africa: financial spillover during COVID-19 pandemic

Global financial spillovers that happened around the time of the COVID-19 pandemic have been widely documented (see, e.g, McKibbin and Roshen, 2023; Gagnon et al., 2023). With regard to the BRICS economies, Khalfaoui et al. (2023) explore the spillover patterns using a large number of financial indicators, including market indices, exchange rates, cryptocurrencies prices, evaluated at several quantiles of their distributions. As a result, Khalfaoui et al. (2023) offer a number of quantile- and country-specific directional networks reflecting the interconnectedness between conventional and Islamic BRICS stock markets, crypto markets (Bitcoin, Ethereum, Litecoin) and commodity markets (oil and gold). In this subsection, we draw on their results to further study information transfer around the time of the pandemic affecting the South African financial market.

Specifically, when looking at the middle quantile, Khalfaoui et al. (2023) find evidence of information transfer between VIX (global volatility index), OVX (oil volatility), GVZ (gold volatility), cryptos (BTC,

ETH, LTC), as well as conventional and Islamic indexes; see Appendix A.1 for details and sources of the data. These connections become stronger at extreme quantiles and may change direction. For example, they note that MXSAI (MSCI Islamic South Africa), MXSA (MSCI South Africa), and TEU (Twitter Economic Uncertainty) transmitted stronger shocks to VIX when the markets were bearish or bullish than when they were normal.

We expand on this result by applying the proposed change-point detection method to South Africa's MXSA index, an equity index measuring the performance of 32 large and mid-cap South African stocks. Its behavior conditional on other indicators reflects spillover effects in the period preceding the COVID-19 pandemic. We evaluate the MXSA-based entropy using the methods of Section 4.1, conditional on those variables from the original list presented in Appendix A.1 that were found to be important transmitters by Khalfaoui et al. (2023, Figure 7). Summary statistics of the variables used in the conditioning set are presented in Appendix A.2. The data covers the period June 2, 2019 to May 28, 2021.[5] The exponential smoothing parameters are $\alpha = 0.75$ and $\beta = 0.75$, the number of shifts $m = 100$ and the size of sliding window is 50 observations.

Figure 6 shows the estimated conditional entropy and log SR statistic as well as a change-point detected on February 23, 2020 and the official start of the pandemic on March 11, 2020. Again, we observe a strong upward trend in the MXSA-based entropy as more uncertainty spills over into the South African stock market. The change-point can be detected about a month ahead of the highest entropy point, which is reached in late March - early April of 2020, and a couple of weeks ahead of the official start of the pandemic. This detection is a reflection of a concept drift that may have started with the first news about COVID in the late 2019. The time pattern of the conditional entropy has a rapid increase right before the crisis and a slow mean reversion after the onset of recession in late 2020. The mean reversion is noteworthy and it suggests that our approach can be used for impulse response analysis –

[5] We are grateful to Professor Rabeh Khalfaoui for providing their data to us.

something that we leave for future work.

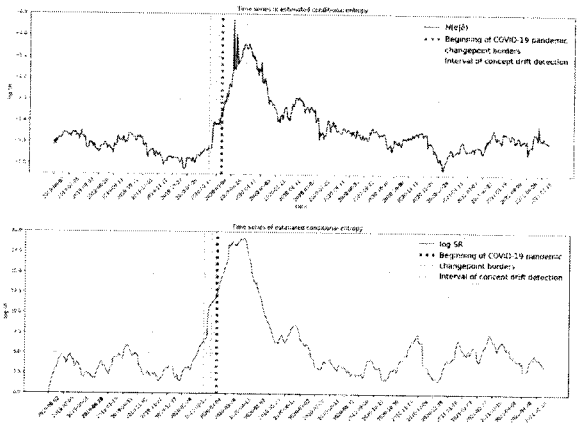


Figure 6: MXSA-based conditional entropy and log SR statistic

7 Concluding remarks

We developed a new framework for the construction of early warning systems, designed to detect concept drift via conditional entropy estimation. This is an original approach that combines benefit of information theory and machine learning to help monitor the level of uncertainty, remaining after accounting for lags of the target and explanatory variables. This approach is particularly relevant for modeling information flows in emerging markets which are characterized by repeated structural breaks, tail dependence and heavy tails in the data distribution.

We suggested a baseline method based on a new conditional density estimator using random forest

weights, and we worked out two extensions to the baseline that better capture non-linear interactions and accommodate heavy-tailed distributions. The extensions use a local linear forest estimator of the conditional mean and a rank-based estimator of the conditional copula density, respectively.

We illustrate the use of the new framework on synthetic and real-life data. The simulation results show when the conventional SR statistic and alternative approaches from machine learning and economics fail while the proposed approach provides robust performance. Fundamentally, the proposed methods constitute a first step towards leveraging recent advances in machine learning and statistics for automated market monitoring that has particular relevance for markets with pronounced volatility, structural breaks, heavy tails and tail-dependence. We leave further extensions of the proposed framework, such as impulse response analysis using conditional entropy, for future work.

References

Ahmad, W., A. V. Mishra, and K. J. Daly (2018): "Financial connectedness of BRICS and global sovereign bond markets," *Emerging Markets Review*, 37, 1–16.

Alanqary, A., A. Alomar, and D. Shah (2021): "Change Point Detection via Multivariate Singular Spectrum Analysis," in *Advances in Neural Information Processing Systems*, ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Curran Associates, Inc., vol. 34, 23218–23230.

Assaf, R., G. Deeksha, and R. Kumal (2023): "The price of war: Effect of the Russia-Ukraine war on the global financial market," *The Journal of Economic Asymmetries*, 28, e00328.

AuYong, H. H., C. Gan, and S. Treepongkaruna (2004): "Cointegration and causality in the Asian and emerging foreign exchange markets: Evidence from the 1990s financial crises," *International Review of Financial Analysis*, 13, 479–515.

Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

——— (2003): "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1–22.

Baidari, I. and H. Nagaraj (2021): "Bhattacharyya Distance Based Concept Drift Detection Method for Evolving Data Stream," *Expert Systems with Applications*, 183.

Balcilar, M., G. Ike, and R. Gupta (2022): "The Role of Economic Policy Uncertainty in Predicting Output Growth in Emerging Markets: A Mixed-Frequency Granger Causality Approach," *Emerging Markets Finance and Trade*, 58, 1008–1026.

Barassi, M., L. Horváth, and Y. Zhao (2020): "Change-Point Detection in the Conditional Correlation Structure of Multivariate Volatility Models," *Journal of Business and Economic Statistics*, 38, 340–349.

Barros, R. S. M. d., J. I. G. Hidalgo, and D. R. d. Lima Cabral (2018): "Wilcoxon Rank Sum Test Drift Detector," *Neurocomputing*, 275, 1954–1963.

Bhattacharyya, A. K. (1943): "On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions," *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.

Bifet, A. and R. Gavalda (2007): "Learning from Time-Changing Data with Adaptive Windowing," *Proceedings of the Seventh SIAM International Conference on Data Mining*.

Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

Caporale, G. M., L. A. Gil-Alana, and C. Poza (2020): "Persistence, non-linearities and structural breaks in European stock market indices," *The Quarterly Review of Economics and Finance*, 77, 50–61.

Chaudhuri, K. and Y. Wu (2003): "Random walk versus breaking trend in stock prices: Evidence from emerging markets," *Journal of Banking & Finance*, 27, 575–592.

Chen, H. (2019): "Sequential change-point detection based on nearest neighbors," *Annals of Statistics*, 47, 1381–1407.

Chen, Z., M. Han, H. Wu, M. Li, and X. Zhang (2022): "A multi-level weighted concept drift detection method," *Journal of Supercomputing*, 79, 5154–5180.

Chen, Z. and R. Ibragimov (2019): "One country, two systems? The heavy-tailedness of Chinese A- and H- share markets," *Emerging Markets Review*, 38, 115–141.

Cherubini, U., S. Mulinacci, F. Gobbi, and S. Romagnoli (2011): *Dynamic Copula Methods in Finance*, Wiley Finance Series, John Wiley and Sons Ltd.

Chu, L. and H. Chen (2019): "Asymptotic distribution-free change-point detection for multivariate and non-euclidean data," *Annals of Statistics*, 47, 382–414.

Cont, R. (2001): "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, 1, 223–236.

Ditzler, G. and R. Polikar (2011): "Hellinger Distance Based Drift Detection for Nonstationary Environments," *IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (CIDUE)*.

Friedberg, R., J. Tibshirani, S. Athey, and S. Wager (2021): "Local Linear Forests," *Journal of Computational and Graphical Statistics*, 30, 503–517.

Gagnon, J. E., S. Kamin, and J. Kearns (2023): "The impact of the COVID-19 pandemic on global GDP growth," *Journal of the Japanese and International Economics*, 68, 101258.

Granger, C. W. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 424–438.

Guo, Y., M. Gao, and X. Lu (2022): "Multivariate change point detection for heterogeneous series," *Neurocomputing*, 510, 122–134.

Hamilton, J. (2016): "Macroeconomic Regimes and Regime Shifts," Elsevier, vol. 2 of *Handbook of Macroeconomics*, 163–201.

Hartley, R. (1928): "Transmission of information," *Bell System Technical Journal*, 535–563.

Hassan, T. A., J. Schreger, M. Schwedeler, and A. Tahoun (2024): "Sources and Transmission of Country Risk," *Review of Economic Studies*, 91, 2307 – 2346.

Hiemstra, C. and J. D. Jones (1994): "Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation," *The Journal of Finance*, 49, 1639–1664.

Hoens, R., R. Polikar, and N. V. Chawla (2012): "Learning from Streaming Data with Concept Drift and Imbalance: An Overview," *Progress in Artificial Intelligence*, 1, 89–101.

Ibragimov, R. and A. Prokhorov (2017): *Heavy tails and copulas: Topics in dependence modelling in economics and finance*, World Scientific Publishing: Singapore.

Izzeldin, M., Y. G. Muradoglu, V. Pappas, A. Petropoulou, and S. Sivaprasad (2023): "The impact of the Russian-Ukrainian war on global financial markets," *International Review of Financial Analysis*, 18, 102598.

James, R., H. Leung, and A. Prokhorov (2023): "A machine learning attack on illegal trading," *Journal of Banking & Finance*, 148, 106735.

Khalfaoui, R., S. Hammoudeh, and M. Z. Rehman (2023): "Spillovers and connectedness among BRICS stock markets, cryptocurrencies, and uncertainty: Evidence from the quantile vector autoregression network," *Emerging Markets Review*, 54, 101002.

Li, Q. and J. S. Racine (2023): *Nonparametric econometrics: theory and practice*, Princeton University Press.

Lin, Y., Y. Wang, and T. S. C. U (2024): "The impact of the Russia–Ukraine war on volatility spillovers," *International Review of Financial Analysis*, 93, 103194.

Mamonov, M., C. F. Parmeter, and A. B. Prokhorov (2024): "Bank Cost Efficiency and Credit Market Structure Under a Volatile Exchange Rate," *Journal of Banking and Finance*, 107285.

McKibbin, W. and F. Roshen (2023): "The global economic impacts of the COVID-19 pandemic," *Economic Modelling*, 129, 106551.

Neaime, S. (2016): "Financial crises and contagion vulnerability of MENA stock markets," *Emerging Markets Review*, 27, 14–35.

Orlitsky, A. (2003): "Information Theory," in *Encyclopedia of Physical Science and Technology (Third Edition)*, ed. by R. A. Meyers, New York: Academic Press, 751–769, third edition ed.

Page, E. S. (1954): "Continuous Inspection Schemes," *Biometrika*, 41, 100–115.

Pepelyshev, A. and A. S. Polunchenko (2017): "Real-time financial surveillance via quickest change-point detection methods," *Statistics and its Interface*, 93–106.

Pergamenchtchikov, S. and A. G. Tartakovsky (2018): "Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data," *Statistical Inference for Stochastic Processes*, 21, 217 – 259.

——— (2019): "Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis," *Journal of Multivariate Analysis*, 174.

Polunchenko, A. S. and A. G. Tartakovsky (2010): "On Optimality of the Shiryaev-Roberts Procedure for Detecting a Change in Distribution," *The Annals of Statistics*, 38, 3445–3457.

Pospisil, T. and A. B. Lee (2018): "RFCDE: Random Forests for Conditional Density Estimation," *arXiv preprint arXiv:1804.05753*.

Probst, P., M. N. Wright, and A.-L. Boulesteix (2019): "Hyperparameters and tuning strategies for random forest," *WIREs Data Mining and Knowledge Discovery*, 9, e1301.

Raab, C., M. Heusinger, and F.-M. Schleif (2020): "Reactive soft prototype computing for concept drift streams," *Neurocomputing*, 340–351.

Roberts, S. (1966): "A Comparison of Some Control Chart Procedures," *Technometrics*, 411–430.

Sahiner, M. (2024): "Volatility Spillovers and Contagion During Major Crises: An Early Warning Approach Based on a Deep Learning Model," *Computational Economics*, 63, 2435 – 2499.

Shannon, C. (1948): "A mathematical theory of communication," *Bell System TechnicalJournal*, 27, 379–423.

Shiryaev, A. (1963): "On Optimum Methods in Quickest Detection Problems," *Probability Theory Application*, 8, 22–46.

Shiryaev, A. N. (1961): "The problem of the most rapid detection of a disturbance in a stationary process," in *Soviet Math. Dokl*, vol. 2, 103.

Smimou, K. and W. Khallouli (2015): "Does the Euro affect the dynamic relation between stock market liquidity and the business cycle?" *Emerging Markets Review*, 25, 125–153.

Tajeuna, E. G., M. Bouguessa, and S. Wang (2023): "Modeling Regime Shifts in Multiple Time Series," *ACM Transactions on Knowledge Discovery from Data*, 17.

Tartakovsky, A. and V. Spivak (2021): "Analysis of the Weighted Shiryaev–Roberts Procedure in the Problem of Changepoint Detection for Models with Unknown Post Change Parameters," *Works of MIPT. Computer Science and Management*, 13.

Tartakovsky, A. G. (2017): "On Asymptotic Optimality in Sequential Changepoint Detection: Non-iid Case," *IEEE Transactions on Information Theory*, 63, 3433–3450.

Tian, M., R. El Khoury, and M. M. Alshater (2023): "The nonlinear and negative tail dependence and risk spillovers between foreign exchange and stock markets in emerging economies," *Journal of International Financial Markets, Institutions and Money*, 82.

Wang, S., C. Luo, and S. Rui (2023): "Unsupervised Concept Drift Detection for Time Series on Riemannian Manifolds," *Journal of the Franklin Institute*, 13186–13204.

Yang, J. (2018): "Information Theoretic Approaches in Economics," *Journal of Economic Surveys*, 32, 940–960.

Yang, Y., L. Zhao, Y. Zhu, L. Chen, G. Wang, and C. Wang (2023): "Spillovers from the Russia-Ukraine conflict," *Research in International Business and Finance*, 66, 102006.

Zhao, H., J. Li, Y. Lei, and M. Zhou (2022): "Risk spillover of banking across regions: Evidence from the belt and road countries," *Emerging Markets Review*, 52, 100919.

A Data description

A.1 Variables and sources

Russia:

- Russian Trading System (RTS) Index: `https://www.moex.com/en/index/RTSI/archive`
- S&P500: `https://www.spglobal.com`
- USD/CNY exchange rate: `https://finance.yahoo.com/quote/CNY%3DX/history`
- Two-year Russian government bonds prices: `https://cbonds.ru/country/Russia_bond/`
- Sberbank shares: `https://www.tradingview.com/symbols/ALOR-SBER`
- Brent crude oil futures: `https://www.investing.com/commodities/brent-oil-historical-data`
- Ruble Overnight Index Average (RUONIA): `https://www.cbr.ru/hd_base/ruonia/dynamics/`
- MSCI for emerging markets: `https://www.investing.com/indices/msci-emerging-markets-historical-data`
- MOEX index: `https://finance.yahoo.com/quote/IMOEX.ME/history/`
- Spread between two- and ten-year Russian government bonds: `https://cbonds.ru/country/Russia_bond/`

South Africa (see Khalfaoui et al., 2023, Table 1):

- FTSE/JSE Africa All Share index [JALSH]: Bloomberg terminal
- MSCI South Africa index [MXSA]: Bloomberg terminal

- MSCI South Africa Islamic index [MXSAI]: Bloomberg terminal
- Coinbase Ethereum, U.S. Dollars [ETH]: Federal Reserve Bank of St. Louis
- Coinbase Litecoin, U.S. Dollars [LTC] Federal Reserve Bank of St. Louis
- Coinbase Bitcoin, U.S. Dollars [BTC] Federal Reserve Bank of St. Louis
- 1-Year Treasury Bill Secondary Market Rate [TB1YR]: Federal Reserve Bank of St. Louis
- Infectious Disease Equity Market Volatility [FDEMV]: Economic Policy Uncertainty
- Twitter Economic Uncertainty [TEU]: Economic Policy Uncertainty
- CBOE Volatility Index [VIX]: Bloomberg terminal
- CBOE Gold ETF Volatility Index [GVZ]: Bloomberg terminal
- Crude Oil ETF Volatility Index [OVX]: Bloomberg terminal

A.2 Summary statistics

Russia

variable	mean	std	skewness	kurtosis	corr(Y_t, Y_{t-1})	rank corr
RTS	0.005	0.071	1.543	6.251	-0.037	-0.046
SnP500	0.001	0.082	0.778	2.902	0.026	0.050
USD/CNY	0	0.002	0.547	4.142	-0.168	-0.192
2Y bonds	0.003	0.031	4.418	53.737	-0.234	0.229
Sber shares	-0.004	0.041	-6.305	63.399	-0.028	0.010
Brent	0.002	0.026	-0.858	6.655	-0.034	-0.056
RUONIA	0.005	0.053	9.208	117.301	0.059	0.032
MSCI EmMar	0	0.011	0.140	3.627	0.180	0.176
MOEX	-0.002	0.034	-6.759	88.344	0.252	0.010
Spread 2Y-10Y bonds	-0.002	0.322	0.667	15.863	-0.135	-0.084

South Africa

variable	mean	std	skewness	kurtosis	corr(Y_t, Y_{t-1})	rank corr
MXSA	0	0.015	-0.872	9.831	0.097	0.167
JALSH	0	0.013	-0.945	10.972	0.089	0.155
BTC	0.002	0.043	-1.164	13.477	-0.081	-0.013
LTC	0.001	0.063	0.706	14.844	-0.054	-0.052
GVZ	0.002	0.042	0.476	7.363	-0.007	0.093
OVX	0	0.054	0.361	23.076	0.106	0.180

Приложение № 1
к Договору на выполнение научно-исследовательских работ
от «29» 07 2024 г. № ЕД-135482/Ф

Техническое задание и разбивка цены

Проведение исследовательских работ по тематике «Новые подходы к экономико-математическому моделированию современных рынков и производств» (соглашение от 10.05.2023 № 20-18-00365-П):

№ п/п	Наименование работ	Результат работ	Количество, ед.	Цена, за ед., руб.	Общая стоимость (вознаграждение Исполнителя), руб.
1.	Исследование по выявлению возможностей усовершенствования методологии раннего предупреждения для финансовых рынков с апробацией на финансовых рынках стран с формирующейся рыночной экономикой	Отчет по ГОСТ в формате MS Word объемом не менее 7 страницах и текст статьи на английском языке.	1	300 000,00	300 000,00
	Страховые взносы 30 %	1 усл. ед.			90 000,00
				ВСЕГО	**390 000,00**

ЗАКАЗЧИК

Проректор по научной работе СПбГУ

_____ С. В. Микушев

ИСПОЛНИТЕЛЬ

_____ А. Б. Прохоров

Согласовано
Руководитель проекта
_____/Л. В. Гадасина