Экзистенциальный риск и

Рост

Леопольд Ашенбреннер и Филип Траммелл (Институт глобальных приоритетов и факультет экономики Оксфордского университета)

Институт глобальных приоритетов | Май 2024 г.

Рабочий документ GPI № 13-2024

Пожалуйста, ссылайтесь на этот рабочий документ следующим образом: Ашенбреннер, Л. и Траммелл, П. Экзистенциальный риск и рост.*Серия рабочих документов Института глобальных приоритетов, № 13-2024.*Доступно по адресу https://globalprioritiesinstitute.org/existential-risk-and-growthaschenbrenner-and-trammell

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

Экзистенциальный риск и рост

Леопольд Ашенбреннер* и Филип Траммелл†

9 июня 2024 г.

Абстрактный

Технологии могут представлять экзистенциальные риски для цивилизации. Хотя ускорение техно-
нологическое развитие может увеличить уровень опасности (риск экзистенциальной катастрофы)
фе за период) в краткосрочной перспективе, два соображения предполагают, что ускорение
снижает риск возникновения такой катастрофы *всегда* происходит. Сначала происходит ускорение де-
увеличивает время, проведенное на каждом уровне технологии. Во-вторых, поскольку более богатое общество
готов пожертвовать большим ради безопасности, оптимальная политика может привести к «экзистенциальному
Кривая риска Кузнеца», в которой ускорение тянет вперед периоды, когда риск
низкий. Ускорение обычно увеличивает риск только при достаточно экстремальной политике
неудачи или прямое влияние ускорения на риск.

---

*Я благодарен Центру эффективного альтруизма, проекту «Будущее человечества» Оксфордского университета.
Институт и Колумбийский институт исследований и разработок (ISERP) за их поддержку.

†Институт глобальных приоритетов и факультет экономики Оксфордского университета. Контакты:
philip.trammell@ Economics.ox.ac.uk . Мы благодарим Дэнни Бресслера, Леннарта Стерна и Майкла Вибе.
за предложение идеи; многим людям в GPI, Бену Снодину, Луису Моте, Тайлеру Коуэну, Рику ван дер Плугу,
Питу Кленову и особенно Чаду Джонсу за полезные комментарии; и Алексу Холнесс-Тофтсу и
Арво Муньосу за помощь в исследовании более раннего черновика.

1 Введение

Технология приносит процветание. Ее влияние на *экзистенциальный риск* — риск человеческого экспомола или, что эквивалентно для целей принятия решения, столь же полного и постоянного потеря человеческого благополучия — многим кажется в лучшем случае двусмысленным.[1] Достижения в области вакцинных технологий нологий делает нас менее уязвимыми для разрушительных эпидемий; достижения в области приобретения функций вирусология, возможно, делает их более вероятными (Миллетт и Снайдер-Битти, 2017).

Это повышает вероятность компромисса↲: забота о выживании цивилизации может мотивировать замедление развития, по крайней мере за пределами некоторых узких областей. Подобные менталистские настроения восходят, по крайней мере, к Римскому клубу 1972 года отчет о «Пределах роста», и недавно вновь выступили с призывами сделать паузу Развитие ИИ (Институт будущего жизни, 2023). Джонс (2024) исследует, как сделать компромисс между разработкой ИИ и рисками ИИ, при условии, что компромисс существует.

Снизит ли замедление технологического развития экзистенциальный риск?

Чтобы пролить свет на этот вопрос, мы начнем в Разделе 2 с простой модели, в которой Коэффициент опасности — вероятность катастрофы за период — является положительной функцией уровень технологий. Здесь экзистенциальная катастрофа должна произойти в конце концов, если только В долгосрочной перспективе более высокие уровни технологий приводят к снижению уровня опасности до нуля.

Это оставляет две возможности. Если передовые технологии в конечном итоге не приведут к уровень опасности стремится к нулю, то катастрофа неизбежна, поэтому ускорение технологического Развитие не может увеличить его вероятность. В противном случае, однако, катастрофа избежать, а ускорение может снизить ее вероятность, ускорив прибытие в безопасное место.

Эта простая модель формализует два наблюдения. Первое заключается в том, что если мы считаем, что уровень опасности в настоящее время высок, наша единственная надежда на долгое и полноценное будущее есть надежда, что мы переживаем временное «время опасностей». Эта точка зрения была

---

[1] См., например, Bostrom (2002), Posner (2004), Farquhar et al. (2017), Ord (2020) и Jones (2024). Мы будем ссылаться на событие, когда человечество немедленно вымирает или страдает от аналогичного полного и постоянная потеря благосостояния как «экзистенциальная катастрофа» или просто «катастрофа».

как известно, выразил Саган (1997), который придумал эту фразу, и ее последствия для тех, кто особенно обеспокоен долгосрочным будущим, Парфит подчеркивает (1984), Ord (2020) и др. Второе менее широко распространено: если мы находимся в опасном времени, где уровень опасности является положительной функцией технологии уровень, тогда, хотя технологическое развитие увеличило уровень опасности до сих пор, Замедление ради долгосрочного выживания ошибочно. Ускорение технологического Развитие может быть временно рискованным, но в долгосрочной перспективе оно безопаснее.[2]

Модель раздела 2 не является «экономической». Она изучает влияние на риск быстро избегание рискованных состояний, а не оптимальная политика в условиях ограничений. Таким образом, это оставляет открытым возможность того, что когда компромиссы между потреблением и риском регулируются политиком при малой заботе о долгосрочном выживании технологическое ускорение может увеличить риск В конце концов. Раздел 2 также не предлагает никаких оснований полагать, что будущие государства*воля*будьте в безопасности. Если кто-то считает, что технология исторически увеличила уровень опасности, и есть надежда, что эта связь изменится в будущем, может показаться наивной.

Поэтому в разделе 3 мы представляем среду, в которой развиваются технологии. экзогенно и его риски могут быть смягчены политикой. Как новый потенциально опасный внедряются технологии, планировщик, дисконтирующий будущее по произвольной ставке, решает, каким объемом потребления следует пожертвовать, чтобы снизить уровень опасности.

Мы показываем, что даже если технологические достижения сами по себе всегда повышают уровень опасности, оптимальная политика может генерировать «кривую Кузнеца экзистенциального риска», с уровень опасности растет, а затем падает по мере развития технологий. Ранние, когда Ожидаемая дисконтированная стоимость будущего цивилизации низка, а предельная полезность потребление высокое, стоит перенимать рискованные технологии по мере их появления. Позже, когда дисконтированное будущее более ценно и предельная полезность потребления упал, существенное снижение риска становится целесообразным.

Таким образом, возможность политического ответа дает экономическое обоснование

---

[2] Однако этот момент неофициально отмечен Бостромом (2014), стр. 234, и недавно Ордом (2024).

мнение, что мы действительно можем переживать уникальное время опасностей. Безопасность - это предмет роскоши, а технологическое развитие создает эффект богатства. Если богатство е↲эффект достаточно силен, то оптимальная политика в конечном итоге быстро снижает опасность достаточно того, чтобы вероятность избежания опасного времени была положительной.

Это понимание отражает логику Стоуки (1998) и Брока и Тейлора (2005) какой ущерб окружающей среде увеличивается, а затем уменьшается с экономическим развитием, и Джонс (2016, 2024), по которому рост дает увеличение стоимости жизни по отношению к предельное потребление. Как и представленный здесь анализ, эти работы показывают, что, учитывая достаточно вогнутые функции полезности, увеличение богатства мотивирует крупные перераспределения от потребления к безопасности. Ни один из этих источников не решает оптимальный путь Однако уровень опасности с течением времени или характеризуют условия, при которых вероятность бинарного события (в данном случае экзистенциальной катастрофы) при оптимальной политике меньше 1.

Наша модель катастрофического риска отличается от модели Мартина более существенно. и Пиндик (2015, 2021 г.) и Аурланд-Бредесен (2019 г.). Эта литература изучает готовность общества платить за снижение риска катастроф, которые произошли или произошли эквивалентно пропорциональному сокращению потребления. В таком контексте нет богатства эф↲фекты: доля потребления, которой человек готов пожертвовать, чтобы избежать пропорционального Сокращение потребления по определению не зависит от уровня потребления.

По причинам, по которым политика способствует выживанию на*данный*(увеличение) технологии путь, оптимальная политика имеет тенденцию увеличивать степень, в которой технологическое ускорение снижает долгосрочный риск. Как и в модели без политики, ускорение снижает время, проведенное в любом данном рискованном состоянии. При оптимальной политике, однако, более богатые будущие состояния, движимые ускорением, систематически склонны быть более безопасными, из-за эффекта богатства. Кроме того, учитывая увеличение будущих темпов роста, даже до того, как фактическая производственная мощность еще увеличится, ожидание более ценное будущее мотивирует более строгую политику безопасности в настоящем.

Этот анализ можно сравнить с анализом Баранзини и Бургиньона (1995).

Баранзини и Бургиньон находят условия, при которых путь роста, который максимизирует ожидаемую дисконтированную полезность, а также минимизирует вероятность существования катастрофа. В нашей модели эти цели никогда не совпадают идеально, но мы исследуем как технический прогресс, когда он регулируется с целью максимизации ожидаемого дисконтированная полезность может снизить вероятность экзистенциальной катастрофы.

В разделах 2 и 3 изучаются модели, в которых*состояние*технологий в определенное время способствует уровню опасности. Раздел 4 рассматривает возможность того, что риск является «транс- сициональный», увеличение скорости технологического*разработка*.

При отсутствии политики влияние ускорения на долгосрочный переходный риск неоднозначно. В частности, ускорение не оказывает никакого влияния на долгосрочный риск при условии, что «эксперимент», связанный с разработкой данной технологии, представляет собой риск, который независимо от того, сколько экспериментов происходит одновременно. Это предположение например, модель «русской рулетки» Джонса (2016) рискованного технологического развития. Если будущее содержит последовательность экспериментов, каждый из которых вызовет некоторые вероятность экзистенциальной катастрофы, тогда стагнация может снизить риск, избегая продвинутые эксперименты в целом, но технологическое ускорение, которое только тянет перенос даты на более поздний срок оставляет вероятность катастрофы неизменной.

Как и в Разделе 3, внедрение оптимального политического ответа способствует выживанию благодаря эффекты богатства, потенциально заменяющие постоянно растущий уровень опасности на уровень Кузнеца кривая. Кроме того, хотя эффект ускорения на долгосрочный переходный риск остается неоднозначная данная политика, политика может изменить условия, при которых ускорение имеет место заданный эффект на риск. По крайней мере, в конкретной модели переходного риска, изученной в Раздел 4, существование политического ответа значительно расширяет условия какое ускорение снижает риск перехода.

В разделе 5 обобщены результаты этих анализов и их ограничения.

2. Государственный риск без смягчения

2.1 Модель

Степень опасности—«Степень опасности»$_t$вероятность потока приtантропогенного экзистенциальная катастрофа. Предположим, что это непрерывная функция технологии уровеньA_t:

$$_t=(A_t).$$

Предположим, чтоA_tэкзогенно и строго возрастает без ограниченийt. Шерсть-Другие предположения о технологическом пути могут быть сделаны без потери общности, как они просто сводятся к переиндексации уровней технологий без изменения их порядка. Поэтому, не теряя общности, предположим, что$A(\cdot)$дифференцируемо и что его производная везде положительна.

Наконец, предположим, что$(A)>0$ для всехA.

Выживание—Вероятность того, что цивилизация сохранится до настоящего времениtдается

$$C_t ⌘ e^{-\int_0^t{}_s ds} \qquad () \; \dot{S}_t = -{}_t C_t, \; C_0=1.$$

Вероятность того, что человеческая цивилизация избежит [антропогенной] экзистенциальной катастрофы, трофей и, по крайней мере, в ожиданиях, наслаждается долгим и процветающим будущим[3]является

$$C_1 ⌘ \lim_{t\to 1} C_t = e^{-\int_0^1 {}_s ds}. \qquad (1)$$

Мы будем ссылаться на$\{_t\}_{t=0}^1$каккривая опасности , в область под кривой опасности ($\int_0^1 {}_t dt$) как кумулятивный риск, и кC_1как вероятность выживания. Обратите внимание, что

[3]Перед лицом естественного экзистенциального риска это повлечет за собой податливость естественной экзистенциальной катастрофе. трофе вместо этого. Из очень-долгосрочных исторических данных о крупномасштабных природных катастрофах и типичный уровень выживаемости других видов млекопитающих, Снайдер-Битти и др. (2019) подсчитали, что человек-уровень естественной экзистенциальной опасности для ity ниже одного на 870 000 в год. В этой статье мы игнорируем возможность того, что технологические достижения могут смягчить естественные экзистенциальные риски. Бухгалтерский учет поскольку эта возможность только усилит основные результаты.

вероятность выживания уменьшается при совокупном риске, и что выживание возможно (C_1>0) i↵ кумулятивный риск конечен.

2.2 Как ускорение влияет на риск?

При отсутствии негативного шока, достаточно сильного, чтобы вызвать стагнацию или рецессию, технология пересекает каждое значение изA_0к 1ровно один раз. Таким образом, площадь под кривой опасности можно определить путем интегрирования по отношению к технологии, а не по времени:

$$З_0^1 (A_T) дт= З_{A_0}^1 (A) \left(\frac{дA}{дт}\right)^{-1} дA= З_{A_0}^1 (A)\dot{A}^{-1}_A дA, \quad (2)$$

где, несколько злоупотребляя обозначениями,$\dot{A}_A$обозначает значение$\dot{A}$когда технология уровень равен индексуA. Эта смена переменных позволяет легче увидеть, как различные потрясения на пути роста влияют на кумулятивный риск.

Мгновенный уровень эф↵фектов—Рассмотрим кратковременный шок на уровне технологий. период, начинающийся в$т$, так что уровень технологий за этот период приблизительно $\tilde{A}$скорее, чемA_T(и последующий технологический путь неизменен). Знак Влияние этого шока на кумулятивный риск зависит от того,($\tilde{A}$) больше или меньше чем(A_T). Из самого левого интеграла (2) мы видим, что влияние на кумулятивный риск*за единицу времени*мгновенного шока на уровне технологий$т$, отA_Tк $\tilde{A}$, равно

$$(\tilde{A})\ (A_T).$$

Мгновенные ускорения—Рассмотрите влияние на кумулятивный риск за единицу времени. мгновенного шока для технологий*рост*в$т$, так что рост технологий ставка в$т$является$\dot{\tilde{A}}$скорее, чем$\dot{A}_T$, и последующие темпы роста технологий на каждом этапе уровень технологии не изменился. Из самого правого интеграла (2) мы видим, что влияние этого шока на совокупный риск*на единицу прироста к уровню технологии* во время ускорения$(A_T)(\dot{\tilde{A}}^{-1} \dot{A}^{-1}_T)$.Умножая это на новую ставку рост технологий в единицу времени, влияние на кумулятивный риск в единицу времени

$$(A_T)\dot{A}/\dot{A}_T\ 1.$$

Ускорения—Выберите уровень технологии $A > \bar{A}$. Поскольку базовый технологический путь увеличивается непрерывно и безгранично, $\bar{A}=A_T$ для некоторых $T > \tau$.

Рассмотрим эффект повышения уровня технологии на τ от A к $\bar{A}$, и субпоследовательное поддержание технологического пути $A_c=A_{c+(T\tau)}(c\ \tau)$. Этот шок для Технологический путь представляет собой «скачок вперед во времени». Влияние этого шока на Таким образом, кумулятивный риск — это отсечение части кривой опасности. риск падает с (2) до

$$\int_{A_0}^{A_T} (A)\dot{A}^{1}\ {}_A дA + \int_{A_T}^{1} (A)\dot{A}^{1}\ {}_A дA.$$

То есть, он падает на

$$\int_{A_T}^{\bar{A}} (A)\dot{A}^{1}\ {}_A дA.$$

В более общем плане, определите <u>временное ускорение</u> как увеличение до $\dot{A}$ в некотором диапазоне уровней технологий: скажем, от A_τ к A_T. Потому что показатель степени $\dot{A}$ в интеграле отрицательно, ускорение снижает риск, переносимый в данном диапазоне технологий уровни. Резкий скачок в уровне технологий означает повышение $\dot{A}_A$ к 1, и таким образом снижая $\dot{A}_A$ до 0, от $A=A_\tau$ к A_T.

Скачок в уровне технологий от A_τ к A_T временно увеличивает опасность оцените, если $(A_T)>(A_\tau)$. Аналогично, ускорение роста технологий ускоряет увеличить до уровня опасности, если $(\cdot)$ увеличивается вокруг A_τ. Поэтому может показаться, современникам, что данный постоянный уровень эффекта уменьшает вероятность выживание. Здесь это было бы неверно. Если (2) бесконечно, вероятность выживания равна ноль с постоянным уровнем эффекта или без него.[4] Если (2) конечно, то постоянный уровень effect увеличивает, снижает кумулятивный риск и увеличивает вероятность выживания.

[4] $\int_{A_T} (A_\tau) д\tau$ конечен непрерывностью в A и из A в τ.

Определите постоянное ускорение быть постоянным увеличением $\dot{A}$ с некоторого времени T— или, что то же самое, некоторый уровень технологий A_T—вперед. Как ясно из (2), постоянный ускорение, как временное ускорение, должно снижать кумулятивный риск, если кумулятивный риск конечен на базовом технологическом пути.

Однако в отличие от временных ускорений постоянные ускорения могут оказывать су- выживание возможно, когда в противном случае оно было бы невозможно. Сокращение кривой с тяжелым хвостом с бесконечным интегралом может дать тонкохвостую кривую с конечным интегралом.

Чтобы сформулировать этот урок наоборот, рассмотрим застой: постоянное «отрицательное ускорение». «эрация» установка $A_t=A_T$ для всех $t \geq T$. Тогда уровень опасности постоянно положительный, и выживание невозможно, даже если бы оно было возможно при любой положительной технологии темпы роста. Более конкретно, рассмотрим последствия большого отрицательного технологического шок сегодня, который вернул мир в состояние невежества относительно всех технологий огия, разработанная с 1924 года. Возможно, уровень опасности был намного ниже в 1924 году, чем сегодня, но даже если это так, эта перезагрузка во многом обречет нас на повторное переживание ядерного противостояния, индустриализации с интенсивными выбросами и биотехнологические опасности прошлого. Если бы повторилось достаточно событий прошлого столетия, катастрофа, по-видимому, была бы неизбежна.

3 Состояние риска с возможностью его смягчения

3.1 Мотивация

Если технический прогресс исторически увеличил уровень опасности, то сообщение предыдущий раздел заключается в том, что те, кто хочет уменьшить экзистенциальный риск, должны ускорить технический прогресс в надежде, что связь между риском и технологией в конечном итоге меняет направление. Это может показаться наивным. Возможно, более естественным предположением будет что при прочих равных условиях технический прогресс только увеличит уровень опасности, в результате чего неминуемая катастрофа наступит раньше.

Но уровень опасности, по-видимому, зависит также от политики. Если уровень опасности имеет увеличилось исторически, это свидетельствует о неспособности политики идти в ногу с новыми рисками, поскольку они возникли. В свете взаимодействия между технологией и политикой, может

можно ли снизить экзистенциальный риск за счет более медленного развития технологий?

Если путь политики не оптимален, то да. Например, если политика экзогенна, кумулятивный риск ниже, когда периоды особенно рискованной технологии совпадают с периодами особенно строгой политики. Для иллюстрации предположим, что

$$_t = A_t x_t,\ x_t = (1+t)^2,$$

где x обозначает переменную политики. Затем рассмотрим ускорение от технологии путь $A_t = (1+t)^\kappa$ на технологический путь $A_t = (1+t)^{\tilde{\kappa}}$, где $\kappa < 1 < \tilde{\kappa}$. Этот ускорение увеличивает кумулятивный риск от

$$\int_0^1 (1+t)^{\kappa 2} dt \quad \text{к} \quad \int_0^1 (1+t)^{\tilde{\kappa} 2} dt.$$

Первое конечно, потому что $\kappa\ 2 < 1$. Последнее бесконечно, потому что $\tilde{\kappa}\ 2 > 1$. В этом случае ускорение снижает вероятность выживания до нуля.

Менее очевидно, может ли ускорение увеличить кумулятивный риск, когда политика возвращается sponse является оптимальным, в рамках правдоподобной модели осуществимого набора политик. Можно было бы беспокоиться что в течение периода, когда более продвинутая технология несет в себе большую опасность, планировщик будет адаптировать политику к степени риска, но слишком слабо для ускорения более низкий совокупный риск в целом — возможно, отчасти потому, что она слишком мало заботится о будущее, приносящее большую часть нынешнего потребления в жертву безопасности.

Чтобы оценить эту возможность, в этом разделе представлен политический канал, через который планировщик, дисконтируя будущее по произвольной ставке, может пожертвовать потреблением снизить уровень опасности. Как мы увидим, когда политика установлена оптимально — по отношению к любая ставка дисконтирования — вывод о том, что ускорение снижает кумулятивный риск, как правило, не только сохранены, но и укреплены.

Как и в технологической модели Раздела 2, выжить можно, только вытащив вперед будущее, которое асимптотически приближается к полной безопасности. В то время как более ранние модель не делает выводов о том, что более продвинутые технологии на самом деле будут нести более низкий уровень опасности, однако, политический ответ вводит тенденцию к более быстрому технологическому развитие нести меньший риск в долгосрочной перспективе. Это потому, что технология увеличивает потребление, которое одновременно снижает стоимость полезности предельной жертвы потребления и увеличивает ценность жизни. Кроме того, перспектива *будущее* ускорение сейчас снижает *подарок* уровень риска, потому что когда ценность будущего больше, это стоит пожертвовать сегодня большим, чтобы предотвратить его крах.

Эта динамика проиллюстрирована в простой модели технологии и оптимальной политики. в остальной части этого раздела. Обобщенные результаты приведены в Приложении А.3.

3.2 Экономическая среда

3.2.1 Технология

Максимально возможный уровень потребления при t соответствует уровню технологии A_t. Действительный потребление - это A_t умноженное на выбор политики $x_t \in [0,1]$:

$$C_t = A_t x_t. \tag{3}$$

Суть этого раздела заключается в том, что технологически развитая цивилизация может рисковать самоуничтожением, но этот риск может быть снижен за счет потребления ция, представленная здесь выбором x ниже 1. (Мы обозначаем переменную выбора x напомнить читателю, что более высокий выбор x (Они сопряжены с более высоким экзистенциальным риском.) Выбор x ниже 1 может означать запрет на принятие мер, увеличивающих потребление но рискованные производственные процессы и/или распределение ресурсов для производства товары и услуги, повышающие безопасность, такие как мониторинг пандемии.

Граница технологий A растет с постоянной скоростью g:

$$\dot{A}_t = A_t g, \; g > 0, \; A_0 > 1. \tag{4}$$

3.2.2 Степень опасности

Степень опасности $_t$ теперь является функцией уровня технологии A_t и выбор политики $x_t \in [0,1]$ и увеличивается в x_t. В этой простой модели эластичности опасности ставка в A и в x постоянны, так что

$$(A_t, x_t) = \bar{} A_t^{\alpha} x_t^{}, \quad \bar{} > 0, \quad > \alpha > 0, \quad > 1. \tag{5}$$

Мы накладываем три неравенства $> \alpha > 0, > 1$, чтобы удовлетворить три желания.[5]

Первое заключается в том, что исправление $x_t > 0$, $_t$ увеличение в A_t. Это налагает $\alpha > 0$. предположение, что $_t$ увеличение в A_t необходимо, если мы хотим признать, что технологический Развитие сделало экзистенциальную катастрофу более вероятной сейчас, чем когда-либо назад, и что эта тенденция будет продолжаться, если не произойдет изменение политики. Доля $1 \; x$ потенциального потребления, принесенного в жертву ради экзистенциальной безопасности, имеет только увеличилось вместе с развитием технологий: когда-то оно было нулевым, теперь оно небольшое но позитивная доля сегодня.[6] Если бы он оставался неизменным, то, по-видимому, уровень опасности был бы пошли по более слабому пути.

Во-вторых, эластичность $_t$ в отношении x_t предполагается, что превышает эластичность из $_t$ в отношении A_t; т.е., $> \alpha$. Это эквивалентно условию, что при С развитием технологий всегда можно снизить уровень риска, сохранив прежний уровень потребления, распределяющий всю предельную производительную мощность на экзистенциальные

---

[5] Функция опасности (5) во многом аналогична функции экологического ущерба Стоки (1998). В то время как Стоки фокусируется на последствиях функции ущерба для выбранного пути x (или "z" в ее обозначениях), мы будем изучать, как ускорения на пути A а α влияет на вероятность двоичного событие: возникновение антропогенной экзистенциальной катастрофы в любое время.

[6] Орд (2020, стр. 313) подсчитал, что по состоянию на 2020 год приблизительно 100 млн долларов США в год было потрачено специально по снижению экзистенциального риска. Это, вероятно, сильно недооценивает расходы на экзистенциальную безопасность в смысле, уместном здесь, по двум причинам. Во-первых, явные расходы не включают в себя упущенные кон- потребление из-за нормативных барьеров. Во-вторых, многие катастрофические усилия по снижению риска мотивированы как желанием снизить экзистенциальные риски, так и желанием уменьшить ущерб меньшего масштаба. Напротив, Мойнихан (2020) утверждает, что сама концепция антропогенной экзистенциальной катастрофы по сути не существовало 300 лет назад; похоже, тогда не было предпринято никаких усилий для его предотвращения.

меры безопасности. Это можно увидеть, заменив $x_t = C_t/A_t$ (из (3)) в

функция опасности (5), дающая

$$_t = \bar{A} C_t \quad _t.$$

Фиксация C, уровень опасности падает со временем i > . Если это (неопределенно) неосуществимо

снизить уровень опасности при фиксировании потребления, как в этой модели, если - ,

тогда экзистенциальная катастрофа неизбежна, если потребление не упадет до нуля. Это

замедление роста будет равносильно уничтожению развитой цивилизации другими способами.

Если - , поэтому ускорение или замедление роста не может оказать никакого влияния на вероятность

экзистенциальной катастрофы в широком смысле.

Третье, исправление $A_t > 0$, $_t$ предполагается строго выпуклым в x_t. Это налагает

>1. Выпуклость подразумевает убывающую отдачу от смягчения экзистенциального риска.

усилия. Мы считаем это разумным предположением как из первых принципов, так и

из недавних оценок Шульмана и Торнли (2024) экономической эффективности

Меры по снижению экзистенциального риска (Приложение A.1).

Связь между кривой опасности и соответствующей вероятностью выживания

C_1 описано в разделе 2.1.

3.2.3 Настройки

Планировщик стремится максимизировать

$$З_1 \int_0 e^{-t} C_t ты(C_t) дт; \qquad ты(C_t) = \frac{C_t^1 \quad 1}{1}, >1. \qquad (6)$$

То есть, поток полезности $ты(\cdot)$ — это CRRA в потреблении для некоторого коэффициента относительного риска

отвращение >1. Полезность потока дисконтируется по экспоненциальной ставке $\to > 0$, представляющий

сумма некоторой ставки чистого временного предпочтения, если таковая имеется, и некоторой ставки естественного и

неизбежный экзистенциальный риск.[7]

[7] Одной из допустимых интерпретаций этих предпочтений было бы то, что популяция фиксирована и (6) ожидаемая полезность репрезентативного домохозяйства. Другой пример: население растет

Полезность смерти неявно нормализована до 0, а эквивалент смерти - уровень потребления к 1. Эквивалентно, мы нормализуем к 1 уровень технологии на когда потребление максимально, полезность потока равна 0.

Планировщик выбирает путь x максимизировать (6) при условии (3)–(5).

Как и Мартин и Пиндайк (2015, 2021), мы предполагаем, что >1 на протяжении всей остальной части в статье, за исключением раздела 3.3.4. Мы предполагаем это отчасти потому, что это кажется верно, как это задокументировано Холлом (1988), Лукасом (1994), Четти (2006) и другими. Также, Однако результаты в остальном относительно неинтересны. Этому есть две причины.

Во-первых, обратите внимание, что когда >1, полезность потока ограничена сверху $\gamma \to 0$. Ускоритель-стабилизация роста потребления, от исходного уровня положительного роста потребления, следовательно, дает поток выгод от полезности, который со временем уменьшается. Эта динамика производит ключевой трейд↵: беспокойство о будущем может поставить под сомнение ценность ускорение технологического развития, поскольку это приносит пользу потреблению в первую очередь, возникают в краткосрочной перспективе, тогда как издержки экзистенциальной катастрофы вечный. Напротив, когда -1, поток полезности может расти без ограничений, поэтому ускорение воздействия на рост потребления и снижение экзистенциального риска могут иметь сопоставимые долгосрочные выгоды.

Во-вторых, и это связано с тем, что когда -1, предельная полезность потребления не снижаться достаточно быстро (по сравнению с растущей ценностью цивилизации), чтобы мотивировать быстрое увеличение потребления жертвует ради безопасности. В результате вероятность долгосрочного выживания всегда равна нулю на выбранном планировщиком пути, а ускорения или Замедление технологического развития не влияет на вероятность. Это подробно описано в разделе 3.3.4.

экспоненциально по ставке $н <$ ⇢, что ставка чистого временного предпочтения и экзогенного риска на самом деле ⇢$+н$, и что планировщик использует общую утилитарную функцию общественного благосостояния.

3.3 Кривая Кузнеца экзистенциального риска

3.3.1 Оптимальность

Подводя итоги раздела 3.2, планировщик выбирает $\{x_т\}_{т=0}^{1}$ максимизировать

$$\text{З}_0^{1} e^{\cdots т} С_т ты(С_т) дт, \tag{7}$$

$$ты(С_т) ⌘ \frac{С_т^{1} \quad 1}{1}, >1 \tag{8}$$

при условии

$$А_0 > 1,\ \dot{А}_т = гА_т (г > 0),$$

$$С_т = А_т х_т,$$

$$С_0 = 1,\ \dot{S}_т = {}_т С_т,$$

$$_т = \bar{А} \quad _т х\ _т (\bar{} > 0, > \quad > 0, > 1). \tag{9}$$

В этом разделе находит путь уровня опасности в решении планировщика, наблюдая что он растет, а затем падает со временем. В следующем разделе мы рассмотрим, что это подразумевает влияние ускорения на кумулятивный риск.

Планировщик сталкивается с одной переменной выбора, $х_т$, и одна переменная состояния, $С_т$. Ее (ожидается) поток рауо в $т$ является $С_т ты(С_т)$. Ее проблема может быть представлена следующим текущим значение Лагранжиана:

$$Л_т = С_т ты(С_т) + в_т \dot{S}_т + \mu_т (1 \quad х_т)$$

$$= С_т \frac{(А\ х)_т^{1} \qquad 1}{1} \quad в_т \bar{А} \quad _т х\ _т С_т + \mu_т (1 \quad х_т). \tag{10}$$

$\mu_т$ является множителем Лагранжа на $х$, положительный і $х_т$-1 ограничение связывает.

$$в_т = \text{З}_т^{1} e^{\cdots(ст)} С_с \frac{ты(С}{С_т} {}_с)\, дс \tag{11}$$

является переменной, влияющей на выживание: ожидаемое значение цивилизации по состоянию на $т$.[8]

[8] Тот факт, что косвенная переменная по выживанию должна быть равна (11), можно сразу увидеть, если задуматься о том, что, по сути, ценность спасения мира должна быть равна ценности мира. Он также формально выведен в Приложении В.1.

На оптимальном пути условие первого порядка по (10) относительно выбора переменная x_T удовлетворяется. Дифференцируя (10) по x_T, у нас есть

$$C_T A_1 x_T \quad _T \bar{A} \quad _T x_1 B_T C_T \; 0, \tag{12}$$

при неравенстве i левая часть положительна при x_T=1, в этом случае x_T=1 есть оптимальным.[9] Таким образом,

- Пока (12) неотрицательно при x_T=1, оптимальный x_T 2[0,1] равно 1. Любой Потери в потреблении приведут к большим расходам, чем ожидаемые выгоды.
- Когда (12) отрицательно при x_T=1, оптимальный выбор x_T является внутренним. Он устанавливает (12) равным нулю, сохраняя условие, что предельная потеря полезности потока от снижения потребления равняется ожидаемой выгоде за счет снижения риска.[10]

На самом деле есть уникальный[11] оптимальный путь, характеризующийся условием первого порядка (12), условие первого порядка, соответствующее переменной состояния C_T, и идентичность (11). Это показано в Приложении B.1. На данный момент наше обсуждение будет опираться только на наблюдения что (12) выполняется на любом оптимальном пути, и что B_T ограничен сверху

$$\bar{V} \quad \frac{1}{(\ 1)}. \tag{13}$$

3.3.2 Первоначальный риск увеличивается

Условие, что (12) неотрицательно при x_T=1 эквивалентно условию, что

$$A_T^{(\ +1)} \quad \bar{B}_T. \tag{14}$$

Ценность продолжения цивилизации на T учитывая выживание T, B_T, всегда строго поднимается со временем. Это следует из того, что при наличии оптимальных путей $\{C_C\}_{CT}$ и

---

[9] Вторая производная по x_T является отрицательным, если предположить, что $>$1.

[10] Мы можем игнорировать возможность того, что оптимум x_T равно 0, поскольку это приводит к бесконечной отрицательной полезности потока.

[11] При условии кусочной непрерывности. Если путь x является оптимальным, мера - нулевые отклонения от x конечно, также оптимально.

$\{c\}_{ct}$ достижимо при заданном начальном уровне технологии A_T, более высокая начальная технология уровень допускает путь с равным уровнем опасности, но большим потреблением на каждом будущий период, предполагая, что $>$ ↵. Более высокий начальный уровень технологии всегда позволяет планировщику реализовать предпочтительное будущее.

Предположим, что неравенство (14) выполняется строго при T=0. Тогда рано, когда A_T низкий, оптимальный выбор политики - x=1, а уровень опасности возрастает со скоростью

$$\Gamma_T = ↵\Gamma.$$

3.3.3 Возможное снижение риска и выживание

Поскольку левая часть (14) экспоненциально падает с A_T и правая сторона поднимается, есть уникальное время $T_\leftarrow$ при котором (14) выполняется с равенством. После $T_\leftarrow$, оптимальный выбор x_T является внутренним и устанавливает (12) равным нулю.

Приравнивая (12) к нулю и переставляя, получаем оптимальный выбор x_T после $T_\leftarrow$, и, таким образом, оптимальный выбор x_T в общем:

$$x_T = \begin{cases} 1, & T - T_\leftarrow; \\ \bar{A}_T^{↵+1} \quad B_T^{\frac{1}{+1}}, & T > T_\leftarrow. \end{cases} \tag{15}$$

Взяв темпы роста каждой стороны, мы можем найти темпы роста выбранной политики переменная после $T_\leftarrow$:

$$\Gamma_{xT} = \frac{↵+1}{+1}\Gamma \quad \frac{1}{+1}\Gamma_{BT}, \tag{16}$$

где, учитывая зависящую от времени переменную y, Γ_{yt} ⌘ $\dot{y}_T/\Gamma_T$ обозначает его пропорциональный рост ставка в T. Уровень опасности в свою очередь растет по мере

$$\Gamma_T = ↵\Gamma + \Gamma_{xT} = \frac{(\ ↵)(\ 1)}{+1}\Gamma \quad \frac{}{+1}\Gamma_{BT}. \tag{17}$$

Потому что $>$ ↵ и $>$1, (17) отрицательно.

Кроме того, хотя Γ_{BT} всегда позитивен, Γ_{BT} ! 0. Это примерно следует из тот факт, что ожидаемая стоимость будущего $в_т$ ограничено сверху $\bar{v}$.[12] Это дает нас асимптотические долгосрочные отрицательные темпы роста $г_x$ и $г$.

Наконец, поскольку $C_т = A_т x_т$, у нас есть

$$г_{Кт} = г + г_{хт} = \frac{↵}{+1} г \quad \frac{1}{+1_{вт}} г.$$

Потому что $> ↵$, долгосрочный рост потребления положительный: x снижается до 0, но A растет быстрее, чем x снижается. Действительно, рост потребления является ключом к рост жертв ради безопасности. С уменьшением предельной полезности *потребление* и уменьшение предельной отдачи *жертвы ради безопасности*, потенциальное потребление разделено между первым и вторым, так что предельная стоимость каждого остается одинаковой.

Подводя итог:

Предложение 1. *Кривая Кузнеца экзистенциального риска*

На пути, определяемом соотношениями (7)–(9), есть время $т_↼ ≥ 0$ таким образом, что для $т < т_↼$,

$$x_т = 1, \; г_{Кт} = г > 0, \qquad г_т = ↵ г > 0$$

и для $т ≥ т_↼$,

$$\lim_{т!1} г_{хт} = \frac{↵ + 1}{+1} г < 0, \tag{18}$$

$$\lim_{т!1} г_{Кт} = \frac{↵}{+1} г > 0, \tag{19}$$

$$\lim_{т!1} г_т = \frac{(↵)(1)}{+1} г < 0 \tag{20}$$

при этом все три предела достигаются снизу.

Следствие 1.1. *Выживание*

На пути, определенном (7)–(9), $C_1 > 0$.

Следствие вытекает из (20) и определения C_1. Потому что $_т$ в конечном итоге падает экспоненциально, $\sum_0^1 {_т} дт < 1$, так $C_1 ⌘ e^{-\sum_0^1 {_т} дт} > 0$.

Обратите внимание, что $_т ! 0$ недостаточно для выживания. Если $_т$ упал до 0 слишком медленно, интеграл расходились бы, и мы бы имели $C_1 = 0$.

[12] The $г_{вт} ! 0$ предел формально показан в Приложении B.2.

3.3.4 Никакого выживания с-1

Как отмечено в разделе 3.2.3, одна из причин для предположения>1 заключается в том, что когда маргинальный Полезность потребления снижается слишком медленно, быстрый переход от потребления к безопасности усилия не предпринимаются, и вероятность долгосрочного выживания всегда равна нулю.

Этот результат напоминает модель «русской рулетки» Джонса (2016). Там это обнаружили, что планировщик выбирает пожертвовать достаточным потреблением ради безопасности, технологически вызванная катастрофа не является неизбежной i↲ 1. В этой модели Однако риск представляет собой развитие, а не существование передовых технологии. Таким образом, это более близко аналогично (действительно, по сути, частный случай) из) нашей модели «переходного риска» Раздела 4, и обсуждается там более подробно.

Результат также напоминает и заостряет внимание на наблюдении Джонса (2024) о важности Коэффициент относительного неприятия риска для готовности избегать экзистенциальных рисков риск. Джонс обнаруживает в условиях одного периода, что когда низкий, планировщик готов терпеть высокий риск экзистенциальной катастрофы в обмен на скачок потребления рост. В условиях одного периода допустимый риск непрерывен в; никаких разрывов нюити наблюдается в =1. Однако в динамической обстановке, изучаемой здесь, планировщик эффективно выбирает, какой риск терпеть период за периодом. Когда-1, достаточно риск допускается каждый раз, когда гарантирована возможная катастрофа.

Предложение 2.*Выбор политики и риск с*-1

Предположим, что планировщик сталкивается с проблемой (7)–(9), но с функцией полезности (8), замененной на

$$ты(C_т) = \begin{cases} \text{журнал}(C_т), & =1; \\ \frac{C_т^{1} \quad 1}{1}, & <1 \end{cases} \qquad (21)$$

для некоторых-1*, и, чтобы обеспечить существование оптимальной политики,*

$$\dashrightarrow > \underline{\dashrightarrow} ⌘ \frac{(\quad ↲)(1\ )}{\quad} г. \qquad (22)$$

Тогда наступает время$т_←$ 0*таким образом, что для*$т < т_←$,

$$х_т=1,\ г_{Кт}=г>0,\ г_т=↲г>0$$

и для т $t\leftarrow$,

$$\lim_{t\to 1} r_{xt} = \frac{}{} r < 0, \quad (23)$$

$$\lim_{t\to 1} r_{Kt} = \frac{}{} r > 0,$$

$$\lim_{t\to 1} \tau t = \frac{}{(\ )r} > 0, \quad =1; \quad (24)$$

$$\lim_{t\to 1} t = \frac{(\ )(1\ )}{+\ 1} > 0, <1. \quad (25)$$

Доказательство. См. Приложение Б.2. □

Следствие 2.1. *Никакого выживания с* -1

На пути, определенном (7)–(9), с заменой функции полезности (8) на (21), $C_1=0$.

Доказательство. Результат следует из (24)–(25) и определения C_1. Когда t асимптотически-тотически постоянный или пропорциональный $1/t$, $\sum_0^1 {}_{1t}dt=1$, так C $e^{\sum_0^1} {}_{0t}dt=0$. □

Случай, в котором t уменьшается пропорционально $1/t$, полученный =1, это край случай, когда ожидаемая продолжительность времени до катастрофы бесконечна, хотя вероятность катастрофы равна 1.

Хотя катастрофа здесь неизбежна на избранном пути, она видна из (25) что более быстрый рост технологий r снижает асимптотическую степень опасности когда <1. Это по сути потому, что когда <1, потребление и, следовательно, поток полезности расти более высокими экспоненциальными темпами в долгосрочной перспективе, когда r выше, поэтому эффект повышение r похож на эффект снижения учетной ставки.

Понимание пути выбора политики и риска становится несколько сложнее, когда -1 чем когда >1, потому что у нас нет результата, который в t асимптотически константа, но схема выглядит следующим образом.

Как и в >1 настройка, на раннем этапе неравенство (14) выполняется и оптимально набор x_t=1. Аналогично, позднее, оптимальность требует настройки x_t<1 для поддержания

$$A_t ты_0 (C_t) = \frac{@}{@x} \cdot в_t$$
$$=) A_t x_t C_t = \bar{A}_{\leftarrow t} x_t в_t$$
$$=)_t = \frac{C_t^1}{в_t}. \tag{26}$$

Заметим из (11), что $в_t$ растет примерно с потоком полезности $ты(C_t)$. Поток полезности, для больших C_t, затем растет примерно как C^1_t когда <1. Итак, хотя потребление растет экспоненциально в долгосрочной перспективе для любого, асимптотически постоянна, когда <1.

Интуитивно понятно, что для того, чтобы политический путь был оптимальным, он должен поддерживать

а) полезность потока к пропорционально увеличивающемуся потреблению, $C_t \cdot C_t$
=

б) ущерб, нанесенный путем пропорционального повышения уровня опасности,

что равно уровень опасности → ценность цивилизации .

Когда ценность цивилизации также растет, как C^1_t, как это происходит, когда <1, опасность Скорость должна быть постоянной, чтобы (a) и (b) росли с одинаковой скоростью. Когда >1, ценность цивилизации асимптотически постоянна, поэтому уровень опасности падает как C^1_t.

Когда =1, учитывая, что потребление растет экспоненциально, log(C_t) и таким образом $в_t$ растут линейно. Затем уровень опасности падает пропорционально 1/t.

3.3.5 Моделирование

Ниже моделируются пути выбора политики и уровень опасности для параметра Значения, указанные в Таблице 1. Значения ⇢,, и r были выбраны в качестве центральных оценок

⇢ 0,02	1.5	r 0,02	A_0 2	↵ 1	2	0,00012

Таблица 1: Параметры моделирования для рисунка 1

из макроэкономической литературы. A_0=2 выбирается таким образом, чтобы значение статистической год жизни в t=75 — это в четыре раза больше потребления на душу населения, что примерно соответствует оценкам из Кленова и др. (2023).[13] То есть, первый год моделирования можно было бы отнести к обозначить 1949 год, когда впервые стала возможной ядерная война между сверхдержавами, в которой в данном случае 75-й год обозначает настоящее время.,̄ ↵, и выбираются таким образом, чтобы уровень опасности сегодня составляет приблизительно 0,1%, что соответствует часто цитируемой цифре Стерна (2007); так что Уровень опасности начинает падать примерно при τ=100; и так, чтобы темпы роста и тогда скорость снижения уровня опасности не является пренебрежимо малой для ясности иллюстрации.

Вероятность выживания C_1 При этих параметрах, из τ=75 и далее, это примерно 65%.

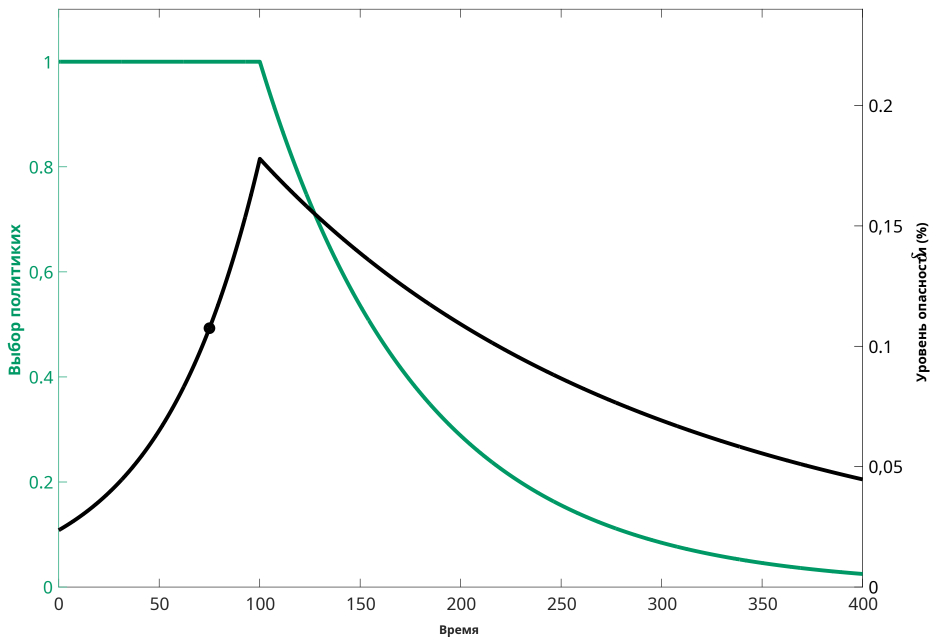


Рисунок 1: Эволюция выбора политики и уровня опасности вдоль оптимального пути

Расчеты и код для воспроизведения моделирования и соответствующей вероятности выживаемости можно найти в Приложении C.

Как показано на рисунке 1, одна потенциально непривлекательная особенность этой простой модели

---

[13] По их оценкам, в 2019 году в США это соотношение составляло примерно 5. Этот показатель необходимо скорректировать в сторону увеличения с учетом экономического роста с 2019 года, но в сторону уменьшения, если мы рассматриваем оптимальные политика во всех странах достаточно продвинута, чтобы внедрять экзистенциально опасные технологии.

заключается в том, что это означает, что на оптимальном пути уровень опасности только возрастает, в то время как Жертвы, какие бы они ни были, приносятся ради экзистенциальной безопасности. В этом он напоминает Стоки (1998) «экологическая кривая Кузнеца», ущерб от которой также растет экспоненциально рост, а затем резкое падение, как только наступает оптимальная ситуация для принятия мер.

Как и в работе Стоки (1998), эта динамика обусловлена отсутствием нижнего состояния Инады. на 1 x. Если предельные «затраты на безопасность» бесконечно снижают уровень опасности на единицу провел в $x=1$, то до тех пор, пока $\delta_T>0$ оптимально установить $x_T<1$, даже если сначала уровень опасности может расти. таким образом, можно найти рядом с падающим x к настройка функции опасности вокруг $x=1$. Такие изменения не влияют на долгосрочную перспективу. поведение политики или риска, как указано в (18)–(20), которые устанавливаются формой функция опасности вокруг $x=0$. Это более подробно обсуждается в Приложении А.4.1.

3.4 Ускорение и государственный риск

Как и в технологической модели Раздела 2, влияние на совокупный риск временного шок неоднозначен, но воздействие ускорения — например, постоянного уровня или эффект роста — всегда заключается в снижении совокупного риска.

3.4.1 Предварительные данные

Позволять $A(\cdot)$ обозначим базовый технологический путь, заданный формулой (4). Пусть $A^* \equiv A_{T^*}$, где T^* является определяется как в предложении 1.

При отсутствии негативного шока, достаточно сильного, чтобы вызвать стагнацию или рецессию, A кресты каждое значение из A_0 к ∞ ровно один раз, поэтому площадь под кривой опасности может быть определяется путем интегрирования по отношению к A вместо T. Мы позволим X обозначают кумулятивный риск, учитывая, что технологический путь $A(\cdot)$ и политический путь x является оптимальным, учитывая $A(\cdot)$:

$$X \equiv \int_0^{\infty} \bar{\delta}_{A_T} x_T \, dT = \int_{A_0}^{\infty} \bar{\delta}_A \, x_A \left(\frac{dA}{dT} \right)^{-1} dA$$
$$= \int_{A_0}^{\infty} \bar{\delta}_A \, x_A \, \dot{A}_A^{-1} \, dA, \qquad (27)$$

где мы снова немного злоупотребим обозначениями, позволив x_A и $\dot{A}_A$ обозначать, соответственно-соответственно, оптимальное значение x (данный технологический путь $A(\cdot)$) и значение $\dot{A}$ когда уровень технологии равен индексу A.

Мы определим B_A и A аналогично. Обратите внимание, что $A ⌘ \bar{A} \cdot x \quad A$, не разделяя это выражение по $\dot{A}_A$. То есть, это все еще показатель опасности: он представляет собой вероятность катастрофа в единицу времени на технологическом уровне A, а не вероятность катастрофы на единицу технологического развития.

$\tilde{A}(\cdot)$ это ускорение от $\underline{A}$ $2[A_0, 1)$ к $A 2(A_T, 1]$ если $\tilde{A}_0 = A_0$ и

$$\dot{\tilde{A}}_A = \dot{A}_A, \quad A < \underline{A};$$
$$\dot{A}_A, \quad A = \underline{A};$$
$$> \dot{A}_A, \quad A 2(\underline{A}, \bar{A});$$
$$= \dot{A}_A, \quad A \; \bar{A}.$$

Ускорение равно постоянный если $\bar{A} = 1$ и временный в противном случае.

Позволять $\tilde{A}(\cdot)$ быть ускорением от A. Определять $\tilde{v}_A$ таким образом, что в $\underline{A} < A$, $\tilde{v}_A = B_A$, и в $A \; \underline{A}$, $\tilde{v}_A$ является переменной, влияющей на выживаемость A учитывая, что последующее технологический путь - это $\tilde{A}(\cdot)$. Затем $\tilde{x}_A$ определяется как равный (15) с A, $\tilde{v}_A$ вместо A_T, B_T;

$$\tilde{}_A ⌘ (A, \tilde{x}_A); \text{ и } X \quad \tilde{} ⌘ \int_{A_0}^{1} \tilde{}_A \dot{\tilde{A}}_A^{-1} дA.$$

Учитывая базовый уровень технологий $\underline{A}$ и темпы роста технологий $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$, обозначают к $\tilde{A}(\cdot)[✏]$ ускорение от A к A+✏ с

$$\dot{\tilde{A}}_A = \dot{\tilde{A}}, A 2[\underline{A}, \underline{A}+✏).$$

Тогда эффект на кумулятивный риск, на единицу технологического развития, моментально-стремительно ускоряясь $\dot{\tilde{A}}$ в A определяется как

$$_{\underline{A}, \dot{\tilde{A}}} ⌘ \lim_{✏ \to 0} \tilde{X}[✏] \; X / ✏,$$

где $\tilde{X}[✏]$ — это кумулятивный риск X, как определено выше, учитывая ускорение $\tilde{A}(\cdot)[✏]$.[14]

[14] Влияние мгновенного ускорения на кумулятивный риск *за единицу времени* является $_{\underline{A}, \dot{\tilde{A}}} \; \dot{\tilde{A}}$, с

3.4.2 Три шока

Мгновенный уровень эффектов—Эффект за единицу времени положительного шока для техно-

уровень нологий A_τ, позволяя политике мгновенно корректироваться, зависит от того, будет ли шок происходит до или после смены режима $\tau_\leftarrow$. В $\tau < \tau_\leftarrow$, временно размножаясь уровень технологии по $m > 1$ не влияет на оптимальный выбор x.[15] Опасность

Таким образом, ставка повышается. Будущая ставка опасности не влияет, поэтому кумулятивный риск увеличивается на

$$_\tau(m - 1) > 0$$

за единицу времени, за которое повышается уровень технологии.

В $\tau\ \tau_\leftarrow$, временно умножая уровень технологии на $m > 1$ умножает переменная политики по m(15). В сочетании, положительный шок для технологий-нология и отрицательное влияние на переменную политики умножают коэффициент опасности на $M^{+1} \overline{{}^{+1}} = M_{(\cdot)(1)} \overline{{}^{+1}} < 1$. Это результирующее изменение совокупного риска

$$\tau \quad M_{(\cdot)(1)} \overline{{}^{+1}} \quad 1 < 0$$

за единицу времени, за которое повышается уровень технологии.

Мгновенные ускорения—Умножение темпов роста технологий на τ к $m > 1$ снижает кумулятивный риск (за единицу времени, в течение которого длится шок) независимо от τ. Это делает так только потому, что шок уменьшает время, проведенное на технологических уровнях вокруг A_τ. Шок не оказывает никакого влияния на политику, связанную с каким-либо уровнем технологий.

Таким образом, как и в модели, учитывающей только технологии, мы видим, что влияние этого шока на кумулятивный риск *на единицу прироста к уровню технологии* во время ускорения

$$_\tau((m\dot{A}_\tau)^{1}\dot{A}^{1}\quad{}_\tau) < 0.$$

Итак, влияние на совокупный риск *за единицу времени* что ускорение длится - это

во время ускорения, $\dot{A}$ Единицы техники разрабатываются в единицу времени. Это одного знака.

[15] Пока не m достаточно велик, чтобы обратить неравенство (14), случай, который мы проигнорируем для простоты.

выше, умноженное на темпы роста новых технологий $м\dot{A}_T$:

$$_T(1\ м)<0.$$

Ускорения—Рассмотрим резкое временное ускорение, при котором технология уровень прыгает на T от A_T к $\bar{A} > A_T$ и экспоненциальный рост технологий впоследствии поддерживается. Поскольку в этой модели оптимальная политика не зависит от истории, эта технология ogу шок представляет собой «скачок вперед во времени». Результирующее воздействие на кумулятивный риск есть

$$З_{A_T}^{\bar{A}} \dot{A}_A^1 дA.$$

В более общем смысле, ускорение от $\underline{A}$ к $\bar{A}$ может снизить риск, которому подвергаются данный диапазон уровней технологий по двум причинам.

1. Как и в модели, ориентированной только на технологии, увеличение темпов роста технологий на A всегда снижает кумулятивный риск напрямую, поскольку показатель степени $\dot{A}_A$ в интеграле (27) есть отрицательный: $\dot{\tilde{A}}_A^1 < \dot{A}_A^1$.

2. Выход за рамки модели, основанной только на технологиях, учитывая A 2 $[A_T, \bar{A})$, ценность будущего в A выше, учитывая более быстрый рост технологий в будущем: $\tilde{v}_A > в_A$. По (15), это слабо мотивирует более жесткую политику $\tilde{x}_A$ - x_A и, таким образом, слабо более низкая опасность ставка $\tilde{A}$-A. Если $\dot{\tilde{A}}_A<1$, более низкий уровень опасности при A снижает совокупный риск.

Только через первый канал изменение совокупного риска, достигнутое за счет ускорения, равно интеграл, на всех технологических уровнях, снижения риска, достигаемого мгновенным ускорения на каждом технологическом уровне:

$$З_{A_T}^{\bar{A}} {}_A(\dot{\tilde{A}}_A^1 \quad \dot{A}_A^1)дA <0.$$

Учитывая влияние политики, совокупное снижение риска достигается более значительным.

Подводя итог:

Предложение 3.*Ускорение и государственный риск*

Мгновенное ускорение в$\underline{A}$ к $\dot{A} > \dot{A}_{\underline{A}}$ снижает кумулятивный риск на единицу

технологическое развитие, в ходе которого он переживает:

а) $_{\underline{A},\tilde{A}} = (\dot{\tilde{A}}_{\underline{A}} \dot{A}_1) < \quad _{\underline{A}} \quad 0.$

Ускорение $\tilde{A}(\cdot)$отАкАснижает кумулятивный риск (и делает это слабо больше соответствующего интеграла мгновенных ускорений):

б)$\tilde{X}$ -X+$_A$ $\sum_{\underline{A}}$ $_{A,A_A}$ *дА < Х,со строгим равенством только если$\bar{A}$ - А.*

Влияние шока на рост на вероятность выживания рассматривается в строго более общая модель Приложения А.3. Обобщенные результаты даны и доказано там в предложении 8.

3.4.3 Моделирование

Эффекты резкого временного ускорения показаны на рисунке 2. Па-Значения параметров, используемые для иллюстрации базового пути, такие же, как и те, которые используются для Имитация рисунка 1. Ускорение происходит «сегодня», вτ=75, и умножаетA к$e^{0.2}$ $\approx$1.22, так что вr=0.02, это будет скачок вперед на 10 лет.

Вспомним из раздела 3.3.5, что вероятность выживания (отτ=75 и далее) на базовый путь составляет примерно 65%. Пропорциональное увеличение вероятности выживаемости можно найти аналитически. Кумулятивный рискXснижается именно на площадь под базовой кривой опасности отτ=75-85; и с тех пор$_{75}$=0.1%,r=0.02, и ε=1, эта разность равна

$$X = 0.001 \int_0^{10} e^{0.02\tau} d\tau = 0.05(e^{0.2} \; 1).$$

C_1=e^{X}затем умножается на$e^{X}$$\approx$1.011, так что в абсолютном выраженииC_1повышается на приблизительно 0.65$\cdot$0.011$\approx$0.7%.

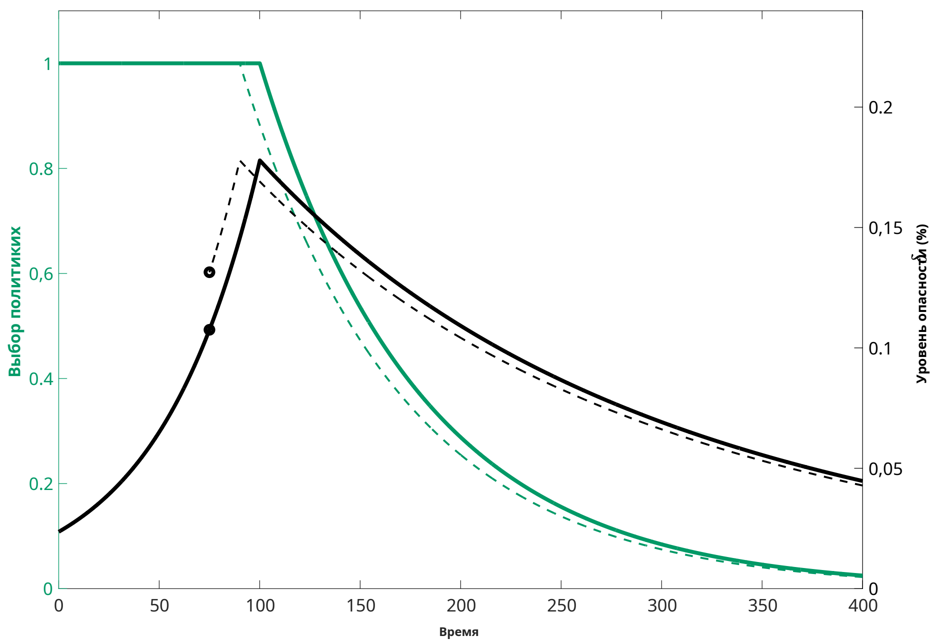


Рисунок 2: Ускорение снижает кумулятивный риск

Расчеты и код для воспроизведения моделирования можно найти в Приложении C.

3.4.4 Обсуждение

Медленный рост делает катастрофу неизбежной—Как отмечено в разделе 2.2, постоянный Отрицательное ускорение, или «замедление», может сделать выживание невозможным: например, если оно вызывает застой.

В этой простой обстановке технологические условия, необходимые для выживания, могут быть точнее. Рассмотрим постоянное замедление, после которого технология растет мощностно-функционально, так что $\tilde{A}_T = T^K$ для некоторых $K > 0$. Экспоненциальный рост скорость $\tilde{A}$, обозначается $\tilde{r}$, тогда не является постоянным при l но изменяющиеся во времени, с $\tilde{r}_T = K/T$. К (17) и с тех пор $\tilde{r}_B / 0$, $_T$ затем падает до 0 как $T_{(\downarrow)(1)K}$. Поскольку кумулятивный риск конечный для $_T/T$ я $\downarrow$ - >1, вероятность выживания положительна i $\downarrow$

$$K > \frac{+\ 1}{(\downarrow)(\ \ 1)}. \tag{28}$$

Рост против терпения—Более быстрый рост увеличивает готовность платить за безопасность. Напротив, те, кто обеспокоен безопасностью далекого будущего, часто пытаются увеличить готовность других платить за безопасность с помощью этических аргументов за низкую ставку Временное предпочтение. Рассмотрим, например, спор Штерна-Нордхауса (и долгие дебаты с тех пор) по поводу ставки дисконтирования, которую следует использовать в климатической политике, или аргументы в пользу беспокойства по поводу будущее, созданное такими философами, как Парфит (1984), Коуэн и Парфит (1992), Орд (2020) и Макаскилл (2022).

Оба вмешательства оказывают одинаковое влияние на вероятность выживания, поскольку подробно в Приложении А.2. Интуиция такова. Если A постоянно умножается к $м$, то, фиксируя политику, предельная полезность потребления умножается на $м$, а стоимость будущего неизменна в пределе при $\bar{v}=$ $\frac{}{\dashrightarrow(1)}$. Если $\dashrightarrow$ является постоянно разделенный $м$, то, фиксируя политику, ценность будущего умножается к $м$; предельная полезность потребления неизменна. Политика уравнивает предельную полезность полезность потребления с предельной стоимостью расходов на безопасность, которая равна пропорционально $в$. В результате, рост терпения и технологий вызывают ком-иносказательные политические ответы, и в конечном итоге сопоставимо влияют на вероятность выживания.

Ускорение может снизить продолжительность жизни—Как мы видели, ускорения всегда... увеличить вероятность выживания в соответствии с предположениями Раздела 3.2. Можно было бы поэтому ожидайте, что в <1 случай из раздела 3.3.4, в котором катастрофа неизбежна и цивилизационная «продолжительность жизни» конечна, ускорения увеличивают эту продолжительность жизни. Однако это не обязательно так.

При любом значении, застой на низком технологическом уровне A создает постоянную опасность скорость $\bar{A}_{\leftarrow}$. Это может быть произвольно низким, поэтому ожидаемая продолжительность до катастрофы $1/(\bar{A}_{\leftarrow})$ может быть произвольно высоким. Когда <1, ускорение может быстро привести к уровень опасности, который постоянно приближается $\leftarrow$(25). Таким образом, ускорение может снизить ожидаемая продолжительность жизни цивилизации примерно до 1 года$/\leftarrow$.

4. Риск перехода

4.1 Мотивация

Функция опасности вида(A_T, x_T) отражает то, что мы назвали «государственным риском»: зависит от*уровень*технологии. В этом контексте, возможно, неудивительно, что Более быстрый выход из рискованных состояний снижает кумулятивный риск.

Но риск может быть «переходным»: вызванным*технологическое развитие*. Этот интуиция, отраженная в модели «русской рулетки» Джонса (2016) в технологическом разработка и (2024) модель риска ИИ, а также по аналогии Бострома (2019) с рисунком потенциально разрушительные шары из урны. Возможно, застой на данном уровне технологий-нология по сути безопасна, а риск возникает в процессе обнаружения и развертывания новые технологии с неизвестными последствиями. Если так, то при положительном росте базовой линии, приводит ли ускорение технологического развития к дальнейшему увеличению совокупного риска?

4.2 Функция опасности, основанная на риске перехода

Чтобы исследовать эту возможность, предположим, увеличивается в$\dot{A}_T$вместо или также как и вA_T. Мы снова ограничим наше рассмотрение функцией риска постоянной эластичности:

$$_T = \bar{A}^{↲\dot{}} \quad {}_T A^{⇣}{}_T x_T, \ \bar{} > 0, \ ⇣\ 0, \ > 1. \tag{29}$$

Наша исходная функция риска (5) является частным случаем (29) с⇣=0 (и> ↲ >0). Таким образом, эта модель является обобщением функции риска (5), дополнительной к Приложению A.3.

Если⇣ >0, однако, модель наиболее естественно интерпретировать как такую, в которой риск вызванные внедрением новых технологий — новых «извлечений из урны Бострома» — которые состоят из абсолютных увеличенийA. Исправление политики, внедрение множественных технологий действия могут представлять больший, меньший или равный риск, если они выполняются одновременно, чем если они выполняются последовательно, в зависимости от знака⇣ 1. Внедрение более передовых технологий может создать больший, меньший или равный риск, чем менее продвинутый, в зависимости от знака↲.

В качестве альтернативы можно интерпретировать одну «новую технологию» как *пропорциональный* увеличить до A,

просто перепишите функцию опасности как

$$_T = ^{-} A^{\alpha+\beta}_T \left(\frac{\dot{A}^{\beta}_T}{A_T}\right) x_T.$$

В этой интерпретации, $\alpha+\beta>0$ - это состояние, при котором развивается больше передовые технологии несут в себе больший риск, чем разработка менее передовых технологий. Потому что $\dot{A}/A$ была примерно постоянной на протяжении последнего столетия, мнение о том, что уровень опасности возрос, что следует отнести к растущей опасности каждого «технологически-«физиологическое развитие» в этом смысле.

Наконец, рассмотрим случай $\alpha=1$, $\beta=1$, так что

$$_T = \bar{}_T x \frac{\dot{A}}{A_T}$$

Здесь, исправление x, каждое пропорциональное увеличение до A вызывает постоянную опасность, независимую в зависимости от того, как быстро происходит увеличение. При отсутствии политики — с $x=1$ (или любая другая константа) постоянно — эта модель по сути эквивалентна «русской «рулетка» модель Джонса (2016)[16] и модель риска ИИ Джонса (2024).

4.3 Ускорение и переходный риск

4.3.1 Без смягчения

Предположим, что базовый технологический путь $A(\cdot)$ непрерывно дифференцируема, с Положительная производная. Пусть $\hat{A} \equiv \lim_{T\to\infty} A_T$ быть конечным или бесконечным.

Как подразумевается выше, фиксирование политики, будь то ускорение увеличивает или уменьшает кумулятивный относительный риск зависит от того, β больше или меньше 1. Это, как обычно, можно увидеть наиболее наглядно путем интегрирования кривой опасности относительно A:

$$X = \int_0^{T} \bar{} A^{\alpha} \dot{}_T A^{\beta}_T \, dT = \int_{A_0}^{\hat{A}} \bar{} A^{\alpha} \dot{A}^{\beta-1} A \, dA.$$

[16] Наш $\bar{}$ это переменная, обозначенная там $\bar{}$.

При заданном ускорении $\tilde{A}(\cdot)$ от $A \in [A_0, A)$ к $A \in (\bar{A}, B]$, кумулятивный риск равен

$$\tilde{X} = X + \int_{A}^{A} \bar{A}^{\varepsilon} \dot{A}^{\delta-1} \dot{A}^{\delta-1} dA.$$

Интеграл отрицателен, если $\delta < 1$, ноль, если $\delta = 1$, и положительно, если $\delta > 1$.

В модели русской рулетки, например, хотя и есть уровень технологий $\hat{A} < 1$ при котором оптимально остановить технологическое развитие (Приложение А.5), ускорение технологического развития до $\hat{A}$ не влияет на кумулятивный риск.

4.3.2 С учетом смягчения

В разделе 2 мы увидели, что ускорение слабо снижает кумулятивный государственный риск отсутствует политики. В разделе 3 мы увидели, что тенденция ускорения к снижению кумулятивного государственный риск был усилен наличием оптимальной политики. Здесь мы увидели, что влияние ускорения на совокупный переходный риск неоднозначно при отсутствии политики. Теперь мы увидим, что это остается неоднозначным при оптимальной политике, но эта политика может вновь ввести тенденцию к ускорению для снижения совокупного риска.

Для простоты мы теперь снова предположим, что A растет с постоянной скоростью экспоненциальная скорость r. Кроме того, поскольку при экспоненциальном росте $r_A = r$, мы навяжем

$$> \varepsilon + \delta, \tag{30}$$

который, вместо того, чтобы $> \varepsilon$, теперь является условием, необходимым для выживания без $C_T = A_T x_T$!0. При этих условиях, поскольку $\dot{A}$ пропорционален A, проблема планировщика точно такой, как описано в разделе 3.3, с $\varepsilon + \delta$ заняв место ε (до коэффициент r^{δ} которые можно свернуть в). Исходный уровень x и пути и C_1, являются Неизмененной. Кривая Кузнеца экзистенциального риска остается.

Позволять A обозначить высший уровень технологии, на котором оптимально установить $x=1$ на базовом технологическом пути. Поскольку условие первого порядка

$$\frac{\partial u}{\partial x_T}(A_T, x_T) \quad \frac{\partial}{\partial x_T}(A_T, x)B_T \quad =) \quad A_T^1 \quad x_T \quad \bar{A}^{\varepsilon} \quad {}_T x_T^{\ 1} B_T$$

должны быть удовлетворены везде и соблюдаться с равенством для $x < 1$, у нас есть

$$x_A = \begin{cases} 1 & A - A_{\leftarrow}, \\ \bar{A}^{+1} \dot{A} \quad \dots \; \frac{1}{+1} & A > A. \end{cases} \tag{31}$$

Подставляя (31) в выражение для кумулятивного риска

$$X = \int_{A_0}^{1} \bar{A} \dot{A}^{1}_{A} \, x_A \, dA, \tag{32}$$

у нас есть

$$X = \int_{A_0}^{A_{\leftarrow}} \bar{A} \dot{A}^{1}_{A} \, dA + \int_{A_{\leftarrow}}^{1} \bar{A}^{(\,)(} \; {}^{1)} B_A^{\frac{1}{+1}} \dot{A}_A^{\frac{1}{+1}} \; {}^{1} dA. \tag{33}$$

Напомним, что технологический путь $\tilde{A}(\cdot)$ ускорение, если $\dot{A}_A > A$ для технологий уровни $A \in [A_0, 1)$ к $A \in (\bar{A}-1]$. С политикой или без нее ускорение влияет кумулятивный риск напрямую, изменяя темпы роста технологий с A к $\bar{A}$. С политика, ускорение также влияет на кумулятивный риск косвенно, влияя B_A для $A \in [\bar{A}, A)$, который влияет на политику в этом диапазоне технологических уровней.

В соответствии с функциями опасности предыдущих разделов, как мы видели, более быстрые технологии Рост нологий всегда слабо предпочтителен. Это следует из того, что он осуществим для компенсации более высоких значений A_t с меньшим выбором x_t, таким образом, что первоначальное потребление сохраняется путь ции, и из предположения от $>$ что учитывая эту политику ответ, кривая опасности слабо снижается. Поскольку будущее с более быстрым ростом более ценным, ускорение от A к $\bar{A}$ поднимает B_A для $A \in [A, \bar{A})$.

При функции опасности (29) этот аргумент больше недействителен. Это потому, что, не- как увеличение до A_T, увеличение до $\dot{A}_T$ не приносит никакой одновременной выгоды, хотя это создает риски, которые все еще можно смягчить только за счет менее одновременного потребления. tion. И действительно, при функции риска (29) более быстрый рост технологий больше не всегда предпочтительнее. Мы можем увидеть это наиболее наглядно в случае $=1$, $=1$: опять же, это модель русской рулетки Джонса (2016), и как обнаруживает Джонс, с>1, оптимально, чтобы технология росла только до конечного уровня. В более общая модель здесь, результат, что стагнация является оптимальной, является острым, как обсуждалось в Приложении А.5. Тем не менее, результат, что ускорение от $\underline{A}$ к $\bar{A}$ не делает обязательно уступить $\tilde{v}_A$ B_A для $A 2 [\underline{A}, \bar{A})$ справедливо в более общем смысле.

Этих сложностей можно избежать, если сосредоточиться на мгновенных ускорениях. воздействие ускорения от $\underline{A}$ к A на $\bar{B}_A$, для $A 2 [\underline{A}, \bar{A})$, падает до нуля, как $\bar{A}A$! 0. Воздействие *краткий* ускорение совокупного риска, таким образом, приблизительно равно воздействие обнаружено, когда мы игнорируем воздействие на B_A.

Предложение 4. *Мгновенное ускорение и риск перехода*

Учитывая функцию опасности (29) и технологический путь (4), выберите уровень технологии $\underline{A}$ >1 *и темпы роста* $\dot{A} > \dot{A}_{\underline{A}}$ *. Если*

а. $\underline{A}$ $A_\leftarrow$ и $\downarrow$ <(=, >) 1 + , *или* $\frac{}{1}$

б. $\underline{A}$ < $A_\leftarrow$, $\downarrow$ <(=, >) 1, *и* $\dot{A}$ *поддерживает (31)* =1 в A=$\underline{A}$,

затем $_{\underline{A},\dot{A}}$ <(=, >) 0.

Доказательство. См. Приложение В.3. □

Результат следует по существу сразу из показателя степени $\dot{A}_A$ в (32). В частности, мгновенное ускорение после $A_\leftarrow$ снижает кумулятивный риск до тех пор, пока

$$\downarrow \frac{1}{+\ 1}\ 1 < 0 =) \downarrow < 1 + \frac{}{1}. \quad (34)$$

Достаточно, хотя и не необходимо, для (34), чтобы

$$\downarrow \text{-1 или } \leftarrow 1, \text{-2}.$$

The $\downarrow$ -1 случай следует из того, что $_1$ <1. $\leftarrow$ 1, -2 случая следует из того, что если $\leftarrow$ 1, тогда, по (30), $\downarrow$ < +1, так что $\downarrow$ <1. Так как макроэкономические оценки -2 являются стандартными, этот результат предполагает, что ускорения более низкий кумулятивный риск на оптимальном пути в контексте риска перехода, по крайней мере, если они происходят достаточно поздно, и меры по смягчению последствий уже принимаются.

Более того, это, опять же, даже без учета того факта, что увеличение к будущему росту может изменить ценность будущего. Хотя направление этого изменение в принципе неоднозначно, большинство наблюдателей сегодня считают само собой разумеющимся, что, по крайней мере, по крайней мере, по обычной ставке дисконтирования, более быстрый рост технологий был бы преимуществом текущая маржа. Это тогда будет еще один канал, через который (положительно-длительность) ускорения вызвало бы сегодня большую озабоченность по поводу безопасности.

Может показаться нелогичным, что мгновенное ускорение снижает риск только тогда, когда ложь*ниже*граница, потому что когда чем выше предельная полезность потребления ниже, и оптимально быстрее перераспределять ресурсы из потребления в безопасность. Результат вытекает из того факта, что когда высока, предельная полезность потребления-ция быстро растет, какxразрезается, поэтому после ускорения, небольшой разрезxпреуспевает уравнять предельную полезность потребления с предельной стоимостью расходов на безопасность ing. Чем выше есть, тем быстрееxпадает какAподнимается, но*меньше*чувствительныйxэто изменение в*@/@x*— например, увеличение из-за более высокого$\dot{A}$—при заданном значенииA.

4.3.3 Обсуждение

Неконкурентность усилий по обеспечению безопасности—Функция опасности (29) изучается здесь в основном для ее простота и сходство с (5). Одна из обоснованных критических замечаний к этой функциональной форме заключается в том, что она преувеличивает значение канала, через который распространяются риски, связанные с рядом технологических смягчить последствия событий может быть дешевле, если они происходят сразу, чем если они происходят постепенно последовательность. Предположим, что ↑1, что↓=1, и что два небольших увеличения до$\dot{A}$-Давайте назовите их двумя «экспериментами» — могут происходить последовательно или одновременно. Если они происходят последовательно, уменьшение вдвое риска, создаваемого каждым, требует уменьшения вдвоеxи таким образом потребление для двух периодов подряд. Если эксперименты происходят одновременно, то же самое сокращение при совокупном риске требуется сократить потребление вдвое только один раз.

Для некоторых видов экспериментов и некоторых видов инфраструктуры безопасности, как-Предположение, что усилия по обеспечению безопасности являются «неконкурентными» в этом смысле, является обоснованным. Сточные воды Мониторинг с целью раннего выявления пандемии снижает риск, связанный с потенциальными биологические эксперименты, вызывающие пандемию, в пропорции, не зависящей от того, как

В настоящее время проводится множество экспериментов.

В других случаях, однако, это не разумно. Это не относится, например, к

расходы на оборудование для обеспечения безопасности, которое должно использоваться в каждой лаборатории. Меры по обеспечению безопасности этого

Этот тип может быть лучше смоделирован с помощью модифицированной версии «безопасности в избыточности»

модель Приложения A.4.2.

Тщательная попытка пролить свет на взаимосвязь между ростом и

Риск перехода потребует дальнейшего изучения. Тем не менее, базовая модель исследовалась

здесь можно извлечь два урока. Во-первых, при отсутствии политики, эффект ускорения на

Риск перехода неоднозначен, и нет никакого эффекта в, возможно, центральном «⇣=1”

случай, принятый Джонсом (2016, 2024). Во-вторых, наличие оптимальной политики

ответ может сместить соответствующий «⇣«порог, в частности, значительно сдвинув его

в той степени, в которой усилия по обеспечению безопасности не являются конкурирующими по отношению к одновременным рискам.

Стагнация против замедления—Когда⇣ >0, полная стагнация ($\dot{A}$=0) всегда

самый безопасный путь из всех. Тем не менее, мы видели с политикой и без нее, что учитывая

положительный темп роста, более быстрый рост может снизить риск.

Ключ к этой головоломке в том, что, учитывая стагнацию на$\hat{A}$, уровни$A > B$никогда не бывают

достигнуто. Кумулятивный риск, следовательно, не (32), а (32) с1заменено на$\hat{A}$.

Отсутствие стагнации, однако, как бы медленны ни были темпы роста, все уровниAдостигнуты.

Темпы роста определяют только риск, который приходится переносить в каждом из них. Прямая стоимость более быстрого

прогресс в течение заданного диапазонаA-значения (более высокий риск за единицу времени, в той степени, в которой

что⇣ >0) частично, а может быть и более чем полностью, перевешивается тем фактом, что

более быстрый прогресс мотивирует к большему смягчению последствий в каждый момент времени в сочетании с

теперь уже известный факт, что когда прогресс идет быстрее, мы не задерживаемся в определенном диапазоне

изA-значения такой же длины.

5 Заключение

Человеческая деятельность может создавать или смягчать экзистенциальные риски. Представленная структура здесь показано, что при различных условиях следует ожидать экзистенциального риска демонстрируют кривую Кузнеца. Это наблюдение предлагает потенциальное экономическое объяснение утверждение некоторых выдающихся мыслителей о том, что человечество переживает критическое «время опасностей». Мы можем быть достаточно развиты, чтобы уничтожить себя, но еще не настолько, чтобы мы готовы пойти на большие жертвы ради безопасности. Если мы действительно живем в период опасностей, сокращение экзистенциального риска сегодня имеет огромные ожидаемые долгосрочные последствия.

В то же время эта структура выделяет канал, через который достигаются некоторые эф- форты, предназначенные для снижения экзистенциального риска, могут иметь обратный эффект. При отсутствии политики, когда риск представляет собой существование передовых технологий, широкомасштабное замедление технологическое развитие обычно ухудшает или не влияет на шансы долгосрочного выживание. При оптимальном политическом ответе, даже если политик мало заботится о В долгосрочной перспективе это влияние усиливается. Влияние может быть значительным, с пропорциональное снижение потребления, имеющее сопоставимые последствия с пропорциональным увеличением увеличение в норме предпочтения времени планировщика. В крайнем случае, постоянная стагнация может сделать катастрофу неизбежной, которую в противном случае можно было бы избежать.

Этот урок сопровождается тремя оговорками. Во-первых, это не аргумент против регули- lating использование рискованных технологий. Действительно, основной канал исследовался здесь через технологическое развитие снижает риск, так как оно приближает день, когда регулирование tion является серьезным. Некоторые недавние реакции на призывы к жесткому регулированию ИИ, например, Андрессен (2023) можно интерпретировать как выражение мнения о том, что наши «x«никогда не следует быть установлено намного ниже единицы. Если это так, то не по причинам, представленным в этой статье.

Во-вторых, когда риск исходит от*разработка*передовых технологий, эффект ускорения на риск более неоднозначно. В моделях «переходного риска» Джонса (2016, 2024), ускорение не влияет на кумулятивный риск при отсутствии политики. При небольших изменениях этих моделей влияние может быть как положительным, так и отрицательным.

Политика может усилить тенденцию к ускорению, чтобы слабо уменьшить кумулятивный риск, как иллюстрирует модель Раздела 4. Однако, кажется вероятным, что при некоторых правдоподобные модели, политика ослабила бы или опровергла бы эту тенденцию.

В-третьих, мы обнаружили, что политика усиливает отрицательную связь между ускорением ation и кумулятивный риск, мы предположили, что политика оптимальна. Если это не так, то влияние ускорения на кумулятивный риск может быть уменьшено или даже отменено, как показано в разделе 3.1. Фактически, Шульман и Торнли (2024) утверждают, что Политический ответ на опасные технологии на сегодняшний день далек от оптимального, даже относительно обычной учетной ставки. Соответствующий урок, который следует извлечь о влияние политики на связь между ускорением и риском заключается только в том, чтобы в какой степени режим регулирования уравнивает или в конечном итоге уравняет предельные полезность потребления к предельной ожидаемой дисконтированной полезности расходов на безопасность дитура, рост потребления, технологическое развитие сегодня имеет невидимое, но потенциально большая выгода от ускорения будущих усилий по обеспечению безопасности. Для замедления технологических для снижения совокупного риска, рассматриваемый провал политики должен быть серьезным и достаточно продолжительный, чтобы перевесить эту выгоду.

В этом свете необходимы дальнейшие исследования характера политических искажений вокруг рег- интерпретация опасных технологий была бы ценной. Изучение долгосрочных последствий катионы других моделей антропогенного экзистенциального риска и оптимальной политики в на первый взгляд, это также может быть ценным, чтобы лучше охарактеризовать масштаб результата, который Оптимально регулируемое ускорение слабо снижает кумулятивный риск. Если правдоподобные модели найдены, при которых результат отменяется, это, естественно, будет представлять важную вопросы, на которые можно ответить только эмпирически. Однако на данный момент результаты представленные здесь данные предполагают, что даже те, кто озабочен исключительно снижением кумулятивного жизненный риск должен часто приветствовать технологические достижения, несмотря на их краткосрочность опасности и выступать за меры по снижению риска сегодня только тогда, когда они достаточно целенаправленны, а затраты на технологическое развитие достаточно малы.

Ссылки

Андрессен, Марк, «Манифест технооптимиста», 2023.

Аурланд-Бредесен, Кине Жозефина, «Оптимальное экономическое управление Катастрофический риск». Кандидатская диссертация, Норвежский университет естественных наук, факультет экономики и бизнеса 2019.

Баранзини, Андреа и Франсуа Бургенион, «Является ли устойчивый рост оптимальным? мал?,"*Международное налогообложение и государственные финансы*, 1995,*2*, 341–356.

Бостром, Ник, «Экзистенциальные риски: анализ сценариев вымирания человечества»,*Журнал-финал эволюции и технологий*, март 2002 г.,*9*(1), 1–35.

__ , *Сверхразум: пути, опасности, стратегии*, Издательство Оксфордского университета, 2014.

__ , «Гипотеза уязвимого мира»,*Глобальная политика*, 2019,*10*(4), 455–476.

Брок, Уильям А. и М. Скотт Тейлор, «Экономический рост и окружающая среда «Окружение: обзор теории и эмпирики» в книге Филиппа Агиона и Стивена Н. Дурлауф, ред.,*Справочник экономического роста*, Том 1, Elsevier, 2005, гл. тер 28, стр. 1749–1821.

Капуто, Майкл Р.,*Основы динамического экономического анализа: оптимальное управление Теория и приложения*, Издательство Кембриджского университета, 2005.

Четти, Радж, ««Ограничение неприятия риска с использованием эластичности предложения рабочей силы»*Амер-ican экономический обзор*, 2006,*96*(5).

Коуэн, Тайлер и Дерек Парфит, «Против социальной учетной ставки», в книге Питера Ласлетт и Джеймс С. Фишкин, ред.,*Справедливость между возрастными группами и поколениями*, Нью-Хейвен: Издательство Йельского университета, 1992, стр. 144–161.

Фаркуар, Себастьян, Джон Холстед и Оуэн Коттон-Барратт, «Экзистенциальный Риск: Дипломатия и управление», Технический отчет 2017 г.

Институт будущего жизни, «Приостановка гигантских экспериментов с ИИ: Открытое Письмо», март 2023 г. https://futureoflife.org/open-letter/

pause-giant-ai-experiments/.

Холл, Роберт, ««Интервременное замещение в потреблении», *Журнал Политических Экономика*, 1988, *96*(2), 339–357.

Джонс, Чарльз И., «Жизнь и рост», *Журнал политической экономии*, 2016, *124*(2), 539–578.

__ , «Дилемма ИИ: рост против экзистенциального риска», 2024. Рабочий документ.

Кленов, Питер Дж., Чарльз И. Джонс, Марк Билс и Мохамад Адхами, «Население и благосостояние: величайшее благо для наибольшего числа людей», декабрь 2023.

Лукас, Дебора, «Оценка активов с недиверсифицируемым риском и короткими продажами «Стремления: углубление головоломки премии за акции», *Журнал денежной экономики*, 1994, *34*(3), 325–342.

Макаскилл, Уильям, *Чем мы обязаны будущему: взгляд на миллион лет вперед*, Один мир Публикации, 2022.

Мартин, Ян В.Р. и Роберт С. Пиндайк, «Предотвращение катастроф: Странная экономика Сциллы и Харибды», *Американский экономический обзор*, 2015, *105*(10), 2947–2985.

__ и __ , «Издержки катастроф на благосостояние: потерянное потребление и потерянные жизни», *The Экономический Журнал*, 2021, *131*(634), 946–969.

Медоуз, Донелла Х., Деннис Л. Медоуз, Йорген Рандерс и Уильям В. Беренс III, *Пределы роста: отчет для проекта Римского клуба о затруднительном положении человечества*, Нью-Йорк: Universe Books, 1972.

Миллетт, Пирс и Эндрю Снайдер-Битти, «Экзистенциальный риск и экономическая эффективность Биобезопасность», *Безопасность здоровья*, 2017, *15*, 373–383.

Мойнихан, Томас, *X-Риск: как человечество обнаружило собственное вымирание*, Урбаномический, 2020.

Нордхаус, Уильям, «Обзор*Обзор Стерна по экономике климата*

Изменять," *Журнал экономической литературы*, 2007, *45*(3), 686–702.

Орд, Тоби, *Пропасть: экзистенциальный риск и будущее человечества*, Нью-Йорк: Блумсбери, 2020.

— , «Надежные долгосрочные сравнения», 2024. Запись в блоге.

Парфит, Дерек, *Причины и лица*, Издательство Оксфордского университета, 1984.

Познер, Ричард А., *Катастрофа: риск и реагирование*, Нью-Йорк: Оксфордский университет Пресс, 2004.

Саган, Карл, *Бледно-голубая точка: видение будущего человечества в космосе*, Баллантайн Книги, 1997.

Шульман, Карл и Эллиот Торнли, «Сколько правительства должны платить Предотвратить катастрофы? Ограниченная роль долгосрочного прогнозирования», в книге Джейкоба Барретта, Хилари Гривз и Дэвид Торстад, ред., *Эссе о долгосрочном прогнозировании*, Оксфорд: Оксфордский универ- Издательство «Университи», 2024.

Снайдер-Битти, Эндрю Э., Тоби Орд и Майкл Б. Бонсалл, «Верхний «Связано с фоновой скоростью вымирания человечества», *Научные отчеты*, Декем- январь 2019, *9*(1), 11054.

Стерн, Николас, *Экономика изменения климата: обзор Стерна*, Кембридж и Нью-Йорк: Издательство Кембриджского университета, 2007.

Стоуки, Нэнси, «Есть ли пределы росту?» *Международное Экономическое Обозрение*, 1998, *39*(1), 1–31.

Всемирный банк, «Данные национальных счетов Всемирного банка и Национальные счета ОЭСР файлы данных подсчета: расходы на конечное потребление (текущие доллары США)», 2022. попробовал отhttps://data.worldbank.org/indicator/NE.CON.TOTL.CD?end= 2022&старт=2022,доступ получен 20 мая 2024 г.

Приложения

Дополнительные материалы

A.1 Калибровка эластичности уровня опасности по отношению к безопасности расходы

Шульман и Торнли (2024) подсчитали, что целенаправленные расходы в размере 400 млрд долларов США следующее десятилетие снизит вероятность экзистенциальной катастрофы в течение следующего десятилетие по крайней мере на 0,1% в абсолютном выражении по сравнению с исходным уровнем 1,85%.

Шкала величины риска взята из работы Орда (2020, стр. 167) догадки и могут быть оспорены. Однако оценка зависит только от*пропорциональ-ция*по которому данная жертва потребления снизит уровень опасности. Мы будем полагаться по оценке Шульмана и Торнли, расходы в размере 400 млрд долларов увеличатся многократно вероятность экзистенциальной катастрофы в течение следующего десятилетия не более

$$1 \quad \frac{0.1\%}{1.85\%} \approx 0.946, \tag{35}$$

оставаясь агностиком относительно величины вероятности. Например, мы доверяют своим оценкам того, в какой степени расходы на мониторинг заболеваний сможет предотвратить экзистенциально опасные антропогенные пандемии, помогая власти должны сдерживать их на ранней стадии, оставаясь при этом агностиком относительно вероятности год возникновения такой пандемии.

В настоящее время мировое потребление в год составляет около 72,5 трлн долларов США (Всемирный банк, 2022). Если реальное потребление растет на 2% в год, а соответствующая процентная ставка составляет 5% в год, текущая стоимость мирового потребления в течение следующих десяти лет составит приблизительно 72 доллара.5T $\cdot (1\ e^{10(0.05\,0.02)})/(0.05\ 0.02) \approx$ \$626.4T. Жертва \$400 млрд = \$0,4 трлн в сегодняшних долларах в течение следующего десятилетия — это жертва, которая умножается потребление фракцией

$$1 \quad \frac{0.4}{626.4} \approx 0.99936. \tag{36}$$

Данный $x < 0.946$ в $x ≈ 0.99936$, следует, что

$$> \frac{\text{журнал}(0.946)}{\text{журнал}(0.99936)} ≈ 86.7.$$

Это упражнение, конечно, ничего не говорит нам о том, разумно ли предполагать Функция риска постоянной эластичности в целом. Если оценка Шульмана и Торнли верна в пределах трех порядков, однако она доказывает, что опасность В настоящее время функция является выпуклой, по крайней мере, в некотором диапазоне возможных уровней потребления. Это следует непосредственно из того факта, что (36) > (35) и что уровень опасности не может быть сокращено в пропорции, большей, чем единица.

A.2 Государственный риск с его смягчением: рост против терпения

Резкий и постоянный уровень эффекта на T, посредством чего A умножается на M немного больше чем 1, составляет скачок вперед примерно на $M/Г$ лет. Это уменьшает cu-мулятивный риск приблизительно $_T M/Г$.

До $T_←$, следовательно, влияние уровня эффекта на кумулятивный риск возрастает экспоненциально. точно с $_T$. Рано по времени $_T$ может быть произвольно низким, поэтому влияние уровня эф↵фекта на кумулятивный риск может также. Влияние снижения на ⇢ по совокупному риску, с другой стороны, не меняется со временем до $T_←$. Уменьшение до ⇢ не делает а↵ влияет на уровень опасности немедленно, но уменьшает его в будущем, подтягивая вперед время смены режима и изменение траектории скорости опасности после этого. Эти воздействия не зависят от *когда* (до $T_←$) ⇢ снижен.

Напротив, рассмотрим, что происходит, как $в_T ! V$. По (15), в пределе,

$$x_T ≈ \bar{V} \quad \frac{1}{+1} A_T \quad \frac{↵+1}{+1}. \tag{37}$$

На свободе T, постоянно размножаясь A к $M > 1$ умножается x_c, на каждом $c\ T$, к примерно $M_{↵+1}$ $\overline{\quad}_{+1}$. В связи с этим увеличение до A_c и пропорциональный уменьшить до x_c умножить $_c$ к $M_{(↵)(1)}^{+1}$ для $c\ T$. Аналогично, постоянно разделяя ⇢ к $M > 1$ умножается x_c($c\ T$) примерно M $\frac{1}{+1}$, который умножается $_c$($c\ T$)

примерно*м* $_{+1}$.Удары равны i↵

$$(\ \hookleftarrow)(\ 1) =$$

$$()=2+\frac{\hookleftarrow}{\hookleftarrow}, \tag{38}$$

с уровнем эффект более эффективен, если левая сторона больше и уменьшение до ⇢более эффективно, если правая сторона больше. Вмешательство, основанное на росте, более эффективен, когда выше, потому что более высокие значения мотивировать более быстрые переходы от потребления к снижению риска.

С> ↵ >0, выражение (38) показывает, что уровень е↵effect может быть только больше эффективно в этой модели, если>2. Тем не менее, примечательно, что простой уровень влияет на рост может в конечном итоге повлиять на вероятность выживания в сопоставимом масштабе до постоянного, равномерно пропорционально уменьшается социальная норма чистого временного предпочтения (удерживание технологический рост зафиксирован). Другими словами, даже временная стагнация может нести долгосрочные издержки, аналогичные издержкам постоянного удаления этических установок из забота о будущем.

А.3 Состояние риска с учетом смягчения: обобщенные результаты

Разделы 3.3–3.4 устанавливаются в среде Раздела 3.2. Три компонента эта среда - это технологический путь, функция от технологии и политики к уровню риска и функции полезности. Для каждого предполагается функциональная форма.

Здесь мы будем поддерживать полезность CRRA с>1. Однако мы значительно смягчим наши предположения о технологическом пути и уровне опасности для определения условий при котором сохраняются уроки разделов 3.3–3.4.

В разделах А.3.2–А.3.3, обобщая предложение 1 из раздела 3.3, мы находим, что рост мотивирует растущую озабоченность безопасностью: часто бывает оптимальным установить*x*=1 ранний вовремя и*x*!0 поздно по времени. Главный результат заключается в том, что, если снижение риска не так трудно, что это не достигается даже с*стагнация в потреблении*, уровень опасности также доведен до 0.

В разделе А.3.4, обобщающем раздел 3.4, мы также обнаруживаем, что когда возникает опасность функция совместима с выживанием, более быстрый рост технологий обычно увеличивает вероятность выживания. Результаты подтверждают надежность извлеченных уроков из функции риска (5): что выживание, скорее всего, возможно на оптимальном пути, и что Более быстрый рост потребительских технологий, если он будет оптимально регулироваться, увеличит его вероятность.

А.3.1 Предположения

Предположения о росте технологий—Мы будем предполагать, что технология путь нологий$A_{(\cdot)}$удовлетворяет некоторым или всем из следующих условий:

А1. непрерывная справа дифференцируемость с$\dot{A}_T$>0 для всехT;

А2.A_0>1;

А3.лим$_{T!1}A_T$=0; и

А4.лим$_{T!1}A_T$= *1*.

Мы назовем технологический путь$A_{(\cdot)}$допустимый если он удовлетворяет А1–А4.

Предположения о степени опасности—Мы также рассмотрим более широкий класс опасности. функции. Среди них мы найдем относительно простые условия, при которых заданное Функция риска и заданный путь роста технологий совместимы с выживанием политика планировщика.

Вернемся к трем пожеланиям, предшествующим введению функции опасности (5). Мы предположим ослабление двух из этих желательных условий напрямую, и определенные результаты будут требуют ослабления третьего. В частности, будем считать, что уровень опасности увеличивается вxне менее быстро, чем вAи слабо выпуклый вx. Для определенных результатов мы предположим, что уровень опасности не уменьшается слишком быстро вA.

К этому мы добавим предварительные условия, которые$(\cdot)$ непрерывно отличается tiable; что, когда потребление равно нулю, так что вся производственная мощность общество стремится к снижению экзистенциального риска, =0; и что в противном случае >0.[17]

Формально мы будем предполагать максимум, что уровень опасности является функцией $A > 0$ и $x \in [0,1]$ удовлетворяющий следующим условиям:

Д1. $(A, x) > 0$,

D2. лим$_{x\to0}(A, x)$ = лим$_{A\to0}(A, x) = 0$,

D3. дважды непрерывная дифференцируемость,[18]

Д4. ⌘$_x(A, x)$ ⌘$_A(A, x)$, и

D5. слабая вогнутость в x,

где ⌘$_y$ обозначает эластичность в отношении $y \in \{A, x\}$. Мы назовем опасность функция допустимый если он удовлетворяет D1–D5.

Обратите внимание, что постоянная эластичная функция опасности разделов 3.2–3.4 допустима, с ⌘$_A$ = ↵ и ⌘$_x$ = независимый от A и x. Также обратите внимание, что мы не требуем ⌘$_A(A, x)$ всегда быть позитивным: мы позволяем новым технологиям снизить уровень опасности заданная степень отказа от потребления.

A.3.2 Конец роста потребления

Позволять C_∞ ⌘ лим$_{t\to\infty} A_t x_t$, когда этот предел определен.

Учитывая функцию опасности (5), $C_\infty = 1$, по (19) из Предложения 1. Однако некоторые допустимые функции опасности мотивируют уменьшение x достаточно быстро, что у нас нет $C_\infty = 1$. C_∞ может быть конечным или C_t может колебаться бесконечно.

Предложение 5. *Конец роста потребления*

---

[17] Напомним, что уровень опасности обозначает вероятность потока *антропогенный* экзистенциальная катастрофа.

[18] Мы определим @⌘$_y(A,1)$ ⌘ лим$_{x\to1}$ @⌘$_y(A, x)/$@x для $y \in \{A, x\}$, и позволяют этим производным быть бесконечный.

При заданной допустимой функции опасности$(\cdot)$*, определять*

$$P(C) ⌘ \text{лим}_{A!1} \frac{@\downarrow}{@x}\left(A, \frac{C ⌘ C}{AA}\right) в, \tag{39}$$

$$P_{\leftarrow} ⌘ \text{лим}_{C!1} P(C).$$

Учитывая допустимый технологический путь и функцию опасности,

а) *Если* $P_{\leftarrow}$ -1, *затем* $C_{\leftarrow} = 1$.

б) *Если* $P_{\leftarrow} > 1$, *затем* $C_{\leftarrow} 6 = 1$.

Доказательство. См. Приложение B.4. □

Чтобы интерпретировать результат, напомним, что $x = C/A$. (39) характеризует, если C фиксировано, даже если A растет, что происходит с отношением предельной стоимости снижения x через увеличенный безопасность ($\frac{@}{@x} \cdot в$) к предельной полезности повышения x через повышенное потребление (AC). Если соотношение приближается к 1, то оптимально, чтобы потребление в долгосрочной перспективе стагнировало в C. Если отношение больше 1 для достаточно большого C, следовательно, тогда застой в какой-то конечный момент C является оптимальным.

Напомним из (13), что $v ⌘ \frac{1}{\to(1)}$. Когда $P(C) > 0$, следовательно, $P(C)$ уменьшается в $\to$. Таким образом, более низкая ставка дисконтирования может сместить $P_{\leftarrow}$ снизу вверх 1, что приводит к стагнации когда в противном случае был бы долгосрочный рост потребления, но никогда обратный. Стагнация потребления в целом нежелательна, когда $\to$ достаточно низкий или нежелательный, когда $\to$ достаточно велик: для многих функций опасности, как показано в конце следующего подраздела, $P_{\leftarrow}$ больше 1 (даже бесконечность) или меньше 1 (даже 0) для любого $\to > 0$. Тем не менее, Предложение 5 иллюстрирует, как призывы к «прекращению роста» может быть совместим с этой моделью. Забота о будущем может мотивировать контроль технологическое развертывание достаточно строгое, чтобы остановить рост *потребление*, несмотря на тенденция к ускорению *технологическое развитие* для снижения совокупного риска.

A.3.3 Обобщенная кривая Кузнеца

Предложение 6.*Обобщенная кривая Кузнеца*

Учитывая допустимый технологический путь и функцию опасности,

а) лим$_{т!1}$$x_т$=1*.*

Если⌘$_A$*ограничено сверху*1 *,затем*лим$_{т!1}$$x_т$=0*.*

б) лим$_{т!1т}$=0*.*

ЕслиС$_{\leftarrow}$*= 1,затем*лим$_{т!1т}$=0*.*

ЕслиС$_{\leftarrow}$*6= 1,*⌘$_A$*ограничено сверху*1 *,и*⌘$_x$*ограничено сверху, то* лим$_{т!1т}$*6*= 0*.*

*Доказательство.*Доказательство (а) приведено в Приложении B.5. Доказательство (б) следующее.

По D1, D2 и D5,(*A, x*) не убывает по*x*. Так что для всех*т*,$_т$-(*A*$_т$*,*1). По D2, предел$_{A!0}$(*A,*1) = 0. Таким образом, по A3, lim$_{т!1т}$=0.

Для положительного предела начните со слабого условия первого порядка, что маргинальное поток полезности увеличения*x*должны слабо превышать предельные издержки за счет увеличения Коэффициент опасности. Затем умножьте обе стороны на*x*$_т$:

$$A_{1}x_{тт} \qquad \frac{@}{@x}(A_т, x_т)в_т$$

$$=)(A_т x_т)_1 \qquad \frac{@}{@x}(A_т, x)_т x_т в_т. \tag{40}$$

Если*С*$_{\leftarrow}$*= 1*, левая часть (40) стремится к 0. Так как*в*(в конечном итоге) положительный и не падает на D4,$\frac{@x}{@x}$!0. Так как$\frac{@x}{@x}$ по D1 и D5,!0.

Если⌘$_A$ограничено выше 1 , Лим$_{т!1}$$x_т$=0 по (а). Так как в конечном итоге$x_т$<1, в конечном итоге (40) выполняется с равенством. Если*С*$_{\leftarrow}$*6= 1*, левая сторона не имеет тенденции к 0 в пределе. Потому что*в*$_т$ограничен сверху,$\frac{@x}{@x}$также не стремится к нулю. Так что если ⌘$_x$⌘$_{@x}$ограничен сверху, *6!*0. □

Часть (б) предложения вытекает из того факта, что до тех пор, пока потребление растет без ограничений его предельная полезность падает до нуля. Если уровень опасности также не падают до нуля, предельная ценность принесения в жертву потребления для его дальнейшего снижения остается

Положительный. Поэтому уровень опасности должен упасть до нуля.

Тем не менее, неограниченный рост потребления не обязательно совпадает с Положительная вероятность выживания. Достичь $C_1>0$, $_T$не должен просто упасть до 0 но падают достаточно быстро. Это в свою очередь гарантируется всякий раз, когда потребление растет достаточно быстро, что выполняется при усилении условия неограниченности рост потребления из Предложения 5.

Предложение 7. *Выживание обобщенное*

При заданной допустимой функции опасности$(\cdot)$*и допустимый технологический путь*$A(\cdot)$*такой что для некоторых* $\kappa>1$ *и некоторые* $\underline{t}$ *у нас есть*

$$A_t\ t^{\frac{\kappa}{1}}\ 8t>t, \ \_ \tag{41}$$

определять

$$P(\kappa)⌘\text{лим}_{t!1}\frac{@\downarrow}{@x}\ A_t,\ \frac{t^{⌘}_1}{A_t}\ \frac{t^{\kappa}_1}{A_t}\ в.$$

а) *Если*$\text{Лим}_{\kappa\#}P(\kappa)<1$*, затем*$9t.\underline{C}_t>t^{\frac{1}{1}}$ $8t>\underline{t}$*и*$C_1>0$*.*

б) *Если*$\text{лим}_{\kappa"1}P(\kappa)>1$*, затем*$9\underline{t}.C_t<t_1^{\overline{1}}$ $8t>\underline{t}$*.*

Если в дополнение$⌘_x$*ограничено сверху, то*$C_1=0$*.*

*Доказательство.*См. Приложение B.6. □

Обратите внимание, что, подобно$P(\cdot)$,$P(\kappa)$ — это долгосрочное отношение предельной стоимости снижение риска до предельной стоимости увеличения потребления, когда

$$C_t/t^{\frac{\kappa}{1}}. \tag{42}$$

Если$P(\kappa)<1$ на этом пути потребления, для некоторых$\kappa>1$, то на этом пути потребление-ция растет слишком медленно. В конечном итоге предпочтительнее поднятьx_tвыше предполагаемого уровня приблизительноt_1^{κ}/A_t. Так что если $\lim_{\kappa\#1}P(\kappa)<1$,$C_t$в конечном итоге растет быстрее чем (42) для некоторых $\kappa>1$ на оптимальном пути. Наоборот, если $\lim_{\kappa\to 1} P(\kappa)$ 1, $C_т$ в конечном итоге растет медленнее, чем (42) для κ=1.

Если $C_т$ растет быстрее, чем (42) для некоторых $\kappa>1$, затем левая часть (40) падает быстрее, чем $т_\kappa$ для некоторых $\kappa>1$. Итак $@x$ делает также. Вспоминая, что $<@x$, это обеспечивает положительную вероятность выживания.

Если $C_т$ растет медленнее, чем (42) для κ=1, то левая часть (40) падает медленнее, чем $1/т$. Правая часть равна $@x\cdot в=⌘_x/\cdot в$. Если $⌘_x$ является ограниченный сверху, падает медленнее, чем $1/т$. Таким образом, кумулятивный риск бесконечен, и выживание невозможно.

Для иллюстрации давайте оценим функцию риска постоянной эластичности из раздела 3.2. для случая экспоненциального роста со скоростью $г$.

$$P(\kappa) = \text{лим}_{т\to 1} \bar{e}^{\leftarrow гт} \frac{т_\kappa ⌘1\, т}{e^{гт}} \frac{\frac{\kappa}{1}}{e^{гт}} \bar{v}$$
$$= \bar{v} \quad \text{лим}\, e^{(\leftarrow)гт\, т\to 1}\, т^{\frac{+\;1\kappa}{1}} = 0 \qquad (43)$$

для любого κ, с $>\leftarrow$. Так что $\lim_{\kappa\to 1} P(\kappa) = 0<1$. Часть (а) Предложения 7, таким образом, обобщает вывод 28 о том, что с функцией риска (5) потребление растет по крайней мере так же быстро, как и достаточная степенная функция (на самом деле она растет экспоненциально) и что существует положительная вероятность выживания.

Напротив, рассмотрим функцию риска постоянной эластичности, но с $\leftarrow$=. В в этом случае, (43) = 1 для любого κ, так что $\lim_{\kappa\to 1} P(\kappa) = 1 > 1$. Также, $⌘_x$ постоянна при , и поэтому ограничены сверху.$(A, x) =$ *Топор* Таким образом, это пример функции опасности удовлетворяющий D1–D5, для которого вероятность выживания на оптимальном пути равна нулю учитывая экспоненциальный рост технологий (и, конечно, учитывая любой $A(\cdot)$ это в конечном итоге ограничено выше нуля).

A.3.4 Ускорение и государственный риск обобщены

Для любой допустимой функции опасности уроки раздела 3.4 по существу являются основными. Поддерживается. Влияние временного уровня на вероятность выживания неоднозначно. ous. Однако, если вероятность выживания положительна при оптимальной для планировщика политике путь, учитывая базовый технологический путь, затем ускорение технологического развития-opment увеличивает вероятность выживания. Если вероятность выживания равна нулю оптимальный для планировщика путь политики, затем ускорение технологического развития может увеличить вероятность выживания или не иметь никакого эффекта.

Предложение 8.*Ускорение и государственный риск обобщены*

Выберите допустимый технологический путь$A(\cdot)$и функция опасности$(\cdot)$.

Данный$\underline{A}$, $\dot{A}$ с $\dot{A} > \dot{A}_{\underline{A}}$,

а) $_{\underline{A},\tilde{A}} = (\dot{A}_{\underline{A}} \dot{A}_1 \quad {}_{\underline{A}}) < 0.$

При заданном ускорении $\tilde{A}(\cdot)$ отAк$\overline{A}$,

б)*Если$X < 1$,тогда $\tilde{X}$ -X+$_A$ $\overset{\text{P}}{\underline{A}} \quad {}_{A,A_A}$ д$A < X$.*

в)*ЕслиX=1и ускорение временное, то $\tilde{X}$=1.*

ЕслиX=1и ускорение постоянно, то $\tilde{X}$ может быть конечным или бесконечным.

*Доказательство.*См. Приложение B.7. □

Интуиция та же, что проиллюстрирована в разделе 3.4. Ускорение в эффекте горизонт-зонально изменяет масштаб всей или части кривой опасности, оставляя меньше времени, затрачиваемого на каждую состояние. Это также может вызвать более жесткую политику в каждом штате, в этом случае слабый Неравенство части (б) является строгим.

A.3.5 Обсуждение

Ускорения против эффектов уровня—Учитывая технологический путь$A(\cdot)$удовлетворяющий A1 и A4, говорят, что дифференцируемый технологический путь$\tilde{A}(\cdot)$это<u>эффект уровня</u> к$A(\cdot)$(в момент времени 0) если

$$9м > 1 : \tilde{A}_т = мA_т 8т.$$

Когда рост технологий происходит экспоненциально, эффекты уровня представляют собой (резкое) временное ускорение. В противном случае они могут быть различны.

В отличие от временных ускорений, эффекты уровня не всегда уменьшают кумулятивный риск вне контекста экспоненциального роста. Рассмотрим, например, функцию опасности (5) с технологическим путем $A_{(\cdot)}$ который практически стагнирует в течение произвольно длительного периода, сказать для T -99; который растет экспоненциально T >99; и для которых подразумеваемый режим-время изменения $T_{\leftarrow}$=100. Уровень↲эффект — скачок в уровне технологий на T=0—тогда повышает уровень технологий в течение сколь угодно длительного периода застоя, который существенно повышает кумулятивный риск, а снижение кумулятивного риска незначительно путем вырезания вертикального среза из кривой опасности, следуя T=99.

Направление технических изменений—Это модель, в которой есть одно ди-Упоминание о технологическом развитии. Изобретения просто происходят в последовательности, каждое из что увеличивает потенциальное потребление и имеет некоторое влияние на уровень опасности в любой заданный уровень потребления. На практике, однако, технологическое развитие, безусловно, по крайней мере, немного *направленный*: компромиссы между потреблением и риском в более поздние периоды зависят от того, в какой степени политики и фирмы в более ранние периоды разработали различные типы технологий. Рассмотрим, например, «более богатую модель» Джонс (2016), в котором рост ценности жизни по отношению к потреблению мотивирует увеличение не только расходов на здравоохранение, но и на медицинские НИОКР.

При установлении базовой последовательности максимальных потенциальных уровней потребления $\{A_T\}$ и функция опасности$(A, \cdot)$, мы просто описываем путь границ возможностей с течением времени, не встраивая никаких предположений о том, как этот путь генерируется. в частности, мы не предполагаем, что существует только один возможный путь для технологии разворачиваться. Если мы предположим более широкое пространство возможных производственных технологий, чем последовательность, принятая на базовом пути, мы должны просто уточнить, что наши результаты только относятся к «ускорениям» в смысле увеличения скорости движения вдоль Базовый путь. Субсидирование разработки рискованных технологий, которые не будут иначе были изобретены, или выбрали технологический путь, на котором они находятся изобретены раньше, чем могли бы быть, но технологии снижения риска не изобретены, не обязательно снижает совокупный риск.[19]

В следующем разделе (Приложение A.4) уроки этой обобщенной модели используются для исследуйте две конкретные функции опасности, которые могут представлять интерес. Первая иллюстрирует что на раннем этапе уровень опасности может увеличиваться наряду с плавным снижением*x*. во-вторых, «микрообоснован» предположением, что рост расходов на безопасность ниже риск из-за избыточных мер безопасности.

A.4 Состояние риска с его смягчением: еще две функции опасности интерес

Мы предположим, что технологии развиваются с постоянной скоростью.$r > 0$.

A.4.1 Более низкое условие Инады по безопасности

Как показано в разделе 3.3, при постоянной эластичности функции риска, поднимается до тех пор, пока остается оптимальным для максимизации потребления и сразу же падает, как только становится

[19] В дополнение к моделированию политического выбора относительно того, какой уровень потребления следует пожертвовать ради мгновенное снижение уровня опасности, более ранняя версия этой статьи моделирует технологический путь, как предписано политикой. Модель роста является полуэндогенной, поэтому общий потенциал технологии рост обусловлен экзогенным ростом населения, но исследования оптимально распределены между рисками увеличение «технологии потребления» и снижение риска «технологии безопасности». Концептуально, что модель проливает свет на тот же вопрос, что и этот — как ускорение влияет на кумулятивный риск, учитывая эндогенный ответ политики — но объектами исследования являются ускорения населения, а не к самой технологии. Численная оценка показывает, что ускорение слабо уменьшает кумулятивный риск в этом контексте, по тем же причинам, что и здесь. Когда рост населения ускоряется-ated, и труд оптимально распределяется по полям, цивилизация проходит примерно тот же технологический путь, но быстрее. Когда ожидается, что будущий рост населения будет более быстрым, стоимость будущее выше (из-за более быстрого будущего технологического развития, даже если большее население не ценится больше по своей сути), поэтому оптимальная политика сдвигает технологический путь в более безопасном направлении.

оптимально начать выбирать субмаксимальное потребление из соображений безопасности. Это Результат, возможно, противоречит опыту прошлого столетия, в течение которого Уровень опасности, вероятно, возрос, в то время как расходы на обеспечение безопасности существования возросли (с по сути 0). Поэтому здесь мы рассмотрим, как настроить функцию опасности так, что кривая Кузнеца сглаживается, а переменная выбора политики падает даже на ранних стадиях время, пока уровень опасности продолжает расти.

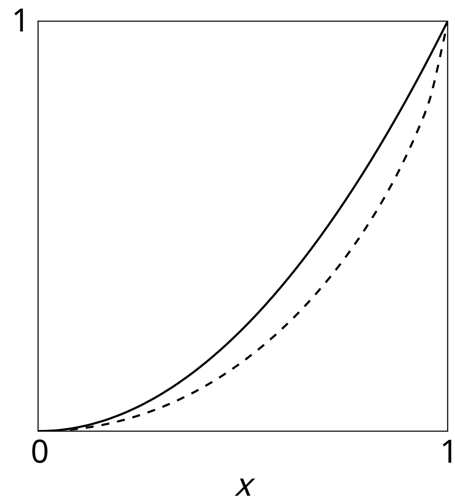
Функция риска постоянной эластичности генерирует отдельную пару режимов для та же причина здесь, что и у Стоки (1998): потому что, когда x=1, предельная «безопасность» расходы» — уменьшается до x— производят только конечные предельные выгоды. То есть, есть нет «нижнего условия Инады по безопасности». Мы скажем, что функция опасности демонстрирует нижнее состояние Инады по безопасности, если предел $x \to 1$ $\frac{\partial}{\partial x} = 1$. При этом условии оптимально установить x_T<1 до тех пор, пока B_T>0: пока цивилизация вообще заслуживает сохранения, некоторые Расходы на снижение экзистенциального риска оправданы.

Не каждая функция опасности с более низким условием Инады по безопасности ведет себя как Сглаженная версия функции риска постоянной эластичности. Если обратная функция риска функция слишком вогнута вокруг x=1 (когда A низкий), то x может быстро упасть, скорее чем мягко, с самого начала, не давая никакого раннего периода, в течение которого $x \uparrow 1$. Если это не так достаточно вогнутый вокруг x=1, с другой стороны, затем рано уменьшается до x производить значительное снижение до, так что уровень опасности снижается даже на ранних стадиях.

Один класс функций опасности с желаемыми характеристиками — это

$$_T = \bar{A}_{\downarrow} \quad _T x_T \frac{1\ (1 \quad x)_{✏}}{x_T}\ _T, ✏ 2 \quad \frac{1}{2}, 1, \tag{44}$$

где условия по параметрам, отличным от ✏ как в (5). Различие между функции опасности проиллюстрированы ниже для случая $\bar{A}$=1, ✏=0.6, =2. Сплошная кривая представляет старую функцию риска; пунктирная кривая представляет новую функция опасности, вертикальная на x=1.



Обратите внимание, что

$$\lim_{x \to 0} \frac{1-(1-x^{\tau})^{✏}}{x^{\tau}} = ✏,$$

поэтому асимптотика в этом случае идентична асимптотике в случае постоянного упругого функция опасности тильности (за исключением того, что уровень опасности умножается на ✏). Однако, Динамика перехода разная. Хотя сейчас оптимально установить $x < 1$ до тех пор, пока $в > 0$, x теперь падает плавно и плавно поднимается и опускается. Пути риска и политика проиллюстрирована ниже для ✏ $=0.6$, $A_0=2.03$, а в остальном тот же параметр значения как в таблице 1.[20]

[20] A_0 слегка повышена для того, чтобы сохранить ценность статистического года жизни «сегодня» (в $\tau=75$) в четыре раза превышает потребление на душу населения, а уровень опасности составляет приблизительно 0,1%, несмотря на то, что в этой модели потребление и уровень опасности даже немного меньше максимальных рано по времени.

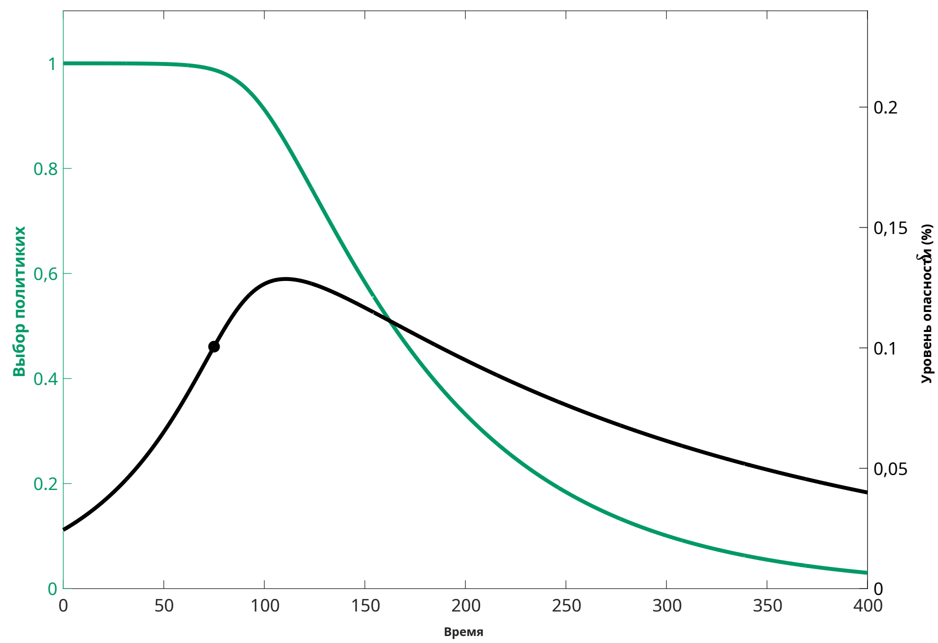


Рисунок 3: Эволюция выбора политики и уровня опасности вдоль оптимального пути

учитывая более низкое условие Инады по расходам на безопасность

Выводы и код для воспроизведения моделирования можно найти в Приложении C.

A.4.2 Безопасность при избыточности

Постоянная эластичная функция риска из разделов 3.2–3.4 и ее корректировка выше, были выбраны для ясности. Однако мы могли бы быть заинтересованы в более обоснованной истории о форме функции опасности, в которой уровень опасности определяется производство потребительских товаров и товаров безопасности. Для иллюстрации, один относительно Проще говоря, история выглядит следующим образом.

- Каждая единица потребления (все еще производимая как $C_T=A_T x_T$) представляет определенный риск $п$ из катастрофа за период при отсутствии каких-либо мер безопасности.
- Для каждой единицы потребительского товара, если одна единица товара безопасности (произведена как $ЧАС_T=A_T(1\ x_T)$) направлено на предотвращение производственного процесса

вызывая катастрофу, это не предотвращает катастрофу с вероятностью$б$ <1. То есть, одна единица$ЧАС$за единицу$С$умножает риск, создаваемый каждой единицей $С$к$б$, от исходного уровня$п$.

- Вероятность того, что производство данной единицы потребления приведет к катастрофа - это вероятность того, что (а) произошла бы катастрофа при отсутствии каких-либо мер безопасности и (б) все$Н/С$меры безопасности не срабатывают независимо:$пб_{Н/С}$.

- Вероятность того, что мир выживет в течение определенного периода, равна вероятности того, что все$С$единицы потребления, независимо, делают$нет$сгенерировать катастрофу: (1 $пб_{Н/С})^{с}$.

В дискретном времени история выше будет соответствовать функции опасности

$$(А_т, х_т) = 1 \quad 1 \quad пб \quad \frac{1 \quad х}{х_т} {}_т А_т х_т, \quad б2(0,1). \tag{45}$$

Аналог (45) в непрерывном времени:

$$(А_т, х_т) = А_т х_т е^{б \frac{1 \quad х_т}{х_т}}, \quad б > 0 \tag{46}$$

(см. Приложение B.8.1).

Поскольку функция опасности (46) не имеет какого-либо нижнего условия Инады на 1 $х$,$х$является фиксировано на 1, и поднимается, рано по времени, пока$в$ >0. После соответствующих расчетов, Предложения 5–7 говорят нам, что (46) дает кривую Кузнеца, при этом в конечном итоге падение достаточно быстро, чтобы обеспечить выживание.

Предложение 9.*Долгосрочная политика и риск, учитывая безопасность при избыточности*

Учитывая функцию опасности (46), оптимальный путь характеризуется

$$\lim_{т!1} х_т т = \frac{б}{г}, \tag{47}$$

$$\lim_{т!1} г_т = г(\quad 1). \tag{48}$$

*Доказательство.*См. Приложение B.8.2. □

Таким образом, снижение выбора политики здесь происходит медленнее, чем в случае постоянной эластичности: xуменьшается пропорционально $1/t$, а не экспоненциально. Это потому, что избыточность-основанная модель дает показатель опасности, который быстро падает в переменной выбора политики: единица уменьшается в$A_t x_t$, а не просто пропорциональное увеличение, генерируют пропорциональное уменьшается до. В обоих случаях, однако,x_t!0. И в обоих случаях,$_t$снижается экспоненциально, и достаточно быстро, чтобы обеспечить выживание.

Сравнивая (48) с предельным выражением дляrИз предложения 1 мы видим, что, в пределе уровень опасности снижается быстрее в модели, основанной на избыточности чем в исходной модели. Это следует из того факта, что дополнительный коэффициент на r(1) в предельном выражении дляrиз Предложения 1 меньше единицы:

$$>0, >1 =) \qquad \frac{}{+\ 1} <1.$$

Интуитивно это объясняется тем, что в модели, основанной на избыточности, меньшие потери в потреблении (линейные, а не пропорциональные) необходимы для пропорционального уменьшения опасности ставка. Ответ планировщика на эту расширенную границу возможностей приходит частично в форме более медленного увеличения отказа от потребления, как описано в (47), и частично в форме более быстрого снижения уровня опасности, как описано в (48).

A.5 Риск перехода: Оптимальный рост технологий

A.5.1 Без смягчения оптимальность стагнации дана=1

Предположим сначала, что=1 и=1, так что

$$_t = -\frac{\dot{A}_t}{A_t}.$$

Как отмечено в тексте, эта модель в точности соответствует модели русской рулетки Джонса. (2016), с̄представляющая переменную, обозначенную там.

Джонс обнаруживает в своей обстановке, что с>1, оптимально, чтобы технологии развивались только до конечного уровня$\hat{A}$. В наших обозначениях это происходит из-за застоя в какой-то момент$\hat{A}$, с никакого риска, приносит постоянную полезность потока$ты(\hat{A})$ и постоянная стоимость будущего

$B(\hat{A})$ ⌘ты$(\hat{A})/$⇢. Таким образом, оптимальным является прекращение роста на том технологическом уровне, на котором будущие выгоды от стагнации на немного более высоком уровне A выравнивание затрат через временно вызывая положительный коэффициент опасности:

$$B_0(\hat{A}) = \frac{@}{@\hat{A}} \cdot B(\hat{A})$$

$$=) \quad \frac{ты_0(\hat{A})}{⇢} = \frac{\bar{y}(\hat{A})}{\hat{A}} \frac{}{⇢} \tag{49}$$

$$=) \; \hat{A} = \frac{\bar{+} \; 1⌘_1}{-} \; \frac{1}{} . \tag{50}$$

Когда ↵=1, мы можем вывести аналитическое решение для оптимального уровня технологии (50) при котором застаиваться. Хотя это невозможно для других значений ↵, это легко чтобы проверить, что для любого ↵ , этот результат качественно не меняется. Равенство (49) затем модифицируется в

$$\frac{ты_0(\hat{A})}{⇢} = \bar{\hat{A}}_{↵} y\, A) \frac{(\hat{\ }}{⇢}$$

$$=) \; \hat{A}_{(↵+)} = \bar{т}ы(\hat{A}). \tag{51}$$

Данный ↵+ >0, левая часть строго монотонно падает от 1 до 0 как $\hat{A}$ поднимается от 1 до 1. Правая часть строго монотонно возрастает от 0 до $7(1) > 0$ как $\hat{A}$ увеличивается от 1 до 1. Таким образом, существует уникальный $\hat{A} > 1$, при котором (51) выполняется: то есть есть, при котором рост технологий предпочтительнее стагнации i↵ $A < B$.

A.5.2 Без смягчения последствий невозможна оптимальная стагнация ⇣ 6= 1

Если мы сделаем дальнейшее обобщение ⇣=1 к произвольному ⇣ Однако мы обнаруживаем, что результат что стагнация является оптимальной, когда ⇣=1 — острый.

Позволять $B_T(A(\cdot))$ обозначают стоимость будущего в T 0 заданный технологический путь $A(\cdot)$. Как базовый уровень, выбор технологического пути $A(\cdot)$ удовлетворяющий A1 и A2.

Если ⇣ <1, затем через каждые T, и для каждого уровня технологии $\bar{A} > A_T$, есть дифференцируемый и слабо возрастающий технологический путь $\tilde{A}(\cdot)$ с $\tilde{A}_c = A_c$ для всех c - T, $\tilde{A}_{T'} = \bar{A}$ для некоторых $T' > T$, и $B_T(\tilde{A}(\cdot)) > B_T(A(\cdot))$.

Чтобы построить такой путь, выберите τ и $\bar{A} > A_\tau$. Заметьте, что если $\dot{A}$ равно а постоянное значение $\dot{A}$ для $A \in (A_\tau, \bar{A})$, кумулятивный риск, перенесенный на пути $\tilde{A}(\cdot)$ от A_τ к $\bar{A}$ равно

$$\int_{A_\tau}^{\bar{A}} \bar{A}^{\epsilon} \dot{A}^{-1} dA,$$

который $\to 0$ как $\dot{A} \to \infty$. С $\epsilon < 1$, поэтому достаточно быстрый рост от A_τ к $\bar{A}$ приближает немедленный, безрисковый прыжок из A_τ к $\bar{A}$, как в государственном риске « $\epsilon = 0$" случай.

Теперь пусть

$$\underline{\tau} \equiv \text{мин}\{\tau : \tilde{A}_\tau = \bar{A}\} = \frac{\bar{A} - A_\tau}{\dot{A}},$$

$$\bar{\tau} \equiv \text{Как дела}\{\tau : A_\tau < \bar{A}\},$$

отмечая, что $\bar{\tau}$ может быть бесконечным, и выбрать $\bar{A}$ и $\dot{A}$ так что $\dot{A} > \dot{A}_c$ для всех $c \in [\tau, \underline{\tau})$. Это возможно для некоторых достаточно высоких $\dot{A}$ по непрерывной справа дифференцируемости из $A(\cdot)$, и гарантирует, что $\tilde{A}_c > A_c$ на протяжении этого интервала. Предположим, что $\tilde{A}_\tau = \bar{A}$ для $\tau \in [\underline{\tau}, \bar{\tau}]$ и $\tilde{A}_\tau = A_\tau$ для $\tau > \bar{\tau}$—т.е. новый путь останавливает рост на $\bar{A}$ до старый путь догнал, если вообще догнал, после чего пути стали идентичными. Тогда $\tilde{A}(\cdot)$ предлагает строго более высокое потребление, чем $A(\cdot)$ через $(\tau, \bar{\tau})$ в обмен на произвольную мелочь изначальный риск и отсутствие последующего увеличения степени опасности.

Кстати, эта структура ясно показывает, что при отсутствии каких-либо затрат технологическое развитие помимо переходного экзистенциального риска, с $\epsilon < 1$ там нет оптимального непрерывного технологического пути. Немедленный скачок в технологии уровень всегда желателен, и больший прыжок всегда предпочтительнее меньшего. Более того, если ввести в модель затраты на НИОКР, то будет существовать оптимальный путь. только если издержки достаточно выпуклы по отношению к скорости технологического развития. В противном случае попытки определить оптимальный технологический путь столкнутся с Проблема «болтания»: быстрые смены медленного и быстрого роста будут предпочтительнее непрерывного роста, потому что они могут достичь заданного количества технологий

логический прогресс в течение определенного интервала времени, внося при этом меньший вклад в кумулятивный риск.

Застой не является оптимальным, учитывая $\downarrow < 1$ потому что из-за «верхнего состояния Инады» на $/\dot{A}_t$с$\downarrow < 1$, достаточно быстрое технологическое развитие несет в себе произвольно мало риск на единицу новой технологии. Застой не является оптимальным, учитывая $\downarrow > 1$ потому что, так как лим$_{\dot{A}/0}$ $@\bar{A}$ $=0$ когда $/\dot{A}_t$с$\downarrow > 1$, достаточно медленное технологическое развитие несет в себе произвольно малый риск на единицу новой технологии.

Чтобы увидеть это, рассмотрим оптимальную скорость роста технологий T учитывая технологию путь $A(\cdot)$ с $A_T = \hat{A} > 1$ и $\dot{A}_c = 0$ для $c > T$. В отличие от $\downarrow < 1$ случай, есть оптимальная скорость роста технологий для принятия T: ставка $\dot{A}_\leftarrow$ который устанавливает предельный ожидаемая выгода от полезности (за счет увеличения будущего потребления) от незначительного увеличения $\dot{A}$, за единицу времени, что $\dot{A}$ увеличивается, равную предельной ожидаемой стоимости полезности за единицу времени (через повышенный уровень опасности при T):

$$B_0(\hat{A}) = \bar{\hat{A}}_{\downarrow}\downarrow\dot{A}_{\leftarrow\downarrow 1}\, B(\hat{A})$$

$$=)A_\leftarrow^{\cdot} = \frac{1}{-} \cdot \left(\frac{\hat{A}_{(\downarrow+)}}{1\hat{A}_1} \right)^{\frac{1}{\downarrow 1}} > 0.$$

Аналогично, учитывая технологический путь $A(\cdot)$ с лим$_{T!1} A_T = \hat{A} < 1$, оптимальный Темпы роста технологий должны удовлетворять равенству, указанному выше в пределе. Поскольку $A(\cdot)$ не может приближаться к конечной верхней асимптоте, если $\dot{A}$ ограничено выше нуля, такой технологии нет путь оптимальный.

A.5.3 При смягчении последствий аналогичные результаты для $\downarrow$-порог $1 + \frac{}{1}$

В этом разделе мы будем предполагать функцию опасности (29) с $\downarrow > 0$:

$$_T = (A_T, \dot{A}_T, x_T) = \bar{A}_{\downarrow}^{\cdot} \quad {}_T A_{\downarrow T} x_T^{\bar{}} > 0, \ \downarrow > 0, \ > 1.$$

Для простоты мы также предположим, что базовый технологический путь характеризуется стадийностью: нация на технологическом уровне $\hat{A}$. Затем мы рассмотрим воздействие за единицу времени мгновенное предельное увеличение темпов роста технологий$\dot{A}_t$.

Мы увидим, что в $\downarrow <1$ + случай, как в $\downarrow <1$ случай без смягчения, оптимального темпа роста не существует: достаточно быстрый рост всегда предпочтительнее, чем застойный. нация. В $\downarrow >1$ + случай, как в $\downarrow >1$ случай без смягчения, рост может быть «слишком быстрым», но пока еще не существует такого уровня технологий, при котором было бы оптимально остановиться на достигнутом.

Однако, $\downarrow=1$ + случай есть *нет* очень похоже на $\downarrow=1$ случай без смягчение. Вместо этого, для низких значений $\hat{A}$ это похоже на $\downarrow <1$ + случай, без оптимальные темпы роста технологий, а также для высоких значений $\hat{A}$ это похоже на $\downarrow <1$+ случай, когда медленный рост предпочтительнее как быстрого роста, так и стагнации. Интуитивно понятно, что это происходит потому, что $\downarrow=1$ + подразумевает $\downarrow >1$. Поскольку медленный рост без смягчение предпочтительнее стагнации, учитывая $\downarrow >1$, и с момента введения опции для снижения риска с помощью $x_t<1$ не исключает возможность медленного роста без смягчение последствий, введение опции политики не может сделать стагнацию оптимальной.

В этой ситуации есть две переменные состояния: вероятность выживания C и уровень технологии A. Есть две переменные выбора: политика x_t и технология темпы роста $\dot{A}_t$.[21] Данный $C_t=1$, предельное чистое воздействие на ожидаемую полезность незначительное увеличение $\dot{A}_t$, в единицу времени, определяется соответствующей производной Гамильтоново выражение

$$ты(A, x_t)\ в_t(\hat{A}, \dot{A}_t, x_t) + \dot{a} \qquad {}_tA_t \tag{52}$$

(адаптировано из Приложения B.1 ниже), где a является переменной стоимостью технологии.

Под x_t-1 ограничение, оптимальный выбор x_t данный $\dot{A}_t$ дается первым условия заказа $@Л/@x_t=0$, $@Л/@\mu_t\ 0$, $\mu_t @Л/@\mu_t=0$ на лагранжиане

$$Л = ты(A, x_t)\ в_t(\hat{A}, \dot{A}_t, x_t) + \dot{a} \qquad {}_tA_t + \mu(1\ x_t). \tag{53}$$

[21] Было бы эквивалентно и более стандартно, но в данном случае и более сложно определить новый выбор. переменная $_t$ таким образом, что технологический закон движения $\dot{A}_t = {}_t$.

Это сводится к

$$x_T = \text{мин}\ 1,\ \quad \bar{\hat{A}}^{+1}\dot{A} \quad {}_T B(\hat{A}) \quad \frac{1}{+1}\ ,\tag{54}$$

с $\mu_T > 0$ i второй член вышеуказанного минимума — неограниченный оптимум выбор x_T — больше 1. (Это адаптировано из (62)–(63) ниже.)

Чтобы найти предельное чистое влияние на ожидаемую полезность предельного увеличения $\dot{A}_T$ за единицу времени, учитывая, что x_T настроен оптимально в ответ, мы можем взять первый производная (53) по $\dot{A}_T$ и оцените его на x_T = (54). Потому что (52) и (53) непрерывно дифференцируемы по $\dot{A}_T$, x_T, и μ_T, по теореме о конверте мы можно дифференцировать (53) относительно $\dot{A}_T$ а затем заменитe x_T = (54), а не учет влияния изменения $\dot{A}_T$ по выбору x_T подставив (54) в (52) и дифференцируя результат по $\dot{A}_T$.

Наконец, учитывая уровень технологий $A_T = \hat{A}$ и перманентный застой после T, значение переменных costate в T являются простыми. Ценность [спасения] цивилизации в T является $B(\hat{A})$, а стоимость предельного повышения уровня технологии — это стоимость равное предельное увеличение потребления во все будущие периоды:

$$B_T = B(\hat{A}) = \frac{1\hat{A}^1}{} \cdot \frac{1}{1},$$

$$a_T = B_0(\hat{A}) = \frac{\hat{A}}{}.$$

Предельное чистое влияние на ожидаемую полезность предельного увеличения $\dot{A}_T$ за единицу время поэтому

$$r(\dot{A}_T) \frac{\hat{A}}{} B(\hat{A})\bar{\hat{A}}A_T \quad {}^{\cdot 1} \quad x_T \tag{55}$$

$$= \frac{\hat{A}}{} B(\hat{A})\bar{\hat{A}}\dot{A}^{1} \quad {}_T\ , \qquad \dot{A}_T < \underline{\dot{A}_T};$$

$$= \frac{\hat{A}}{} \ \bar{1} B(\hat{A})^1 \hat{A}^{(\ )(1)} \qquad \frac{1}{+1} \quad \dot{A}_T^{\frac{1}{+1}\ 1}\ ,\ \dot{A}_T\ A_T, \underline{\quad}$$

где

$$\underline{\dot{A}_T}\ \ \bar{\hat{A}}^{+1} B(\hat{A}) \quad \frac{1}{}$$

это максимальная скорость роста, при которой оптимально установить $x_T = 1$, и $в(\hat{A})$ как определено выше.

Если ⇣ $< 1 + \frac{}{1}$, тогда показатель степени на $\dot{A}_T$ в (55) отрицательно для $\dot{A}_T \dot{A}_T$, так

$$\lim_{\dot{A}_T \to 1} r(\dot{A}_T) = \hat{A}/⇢ > 0.$$

Как и в ⇣ < 1 случай без политики, это гарантирует, что достаточно быстрая технология Рост всегда предпочтительнее стагнации.

Если ⇣ $> 1 + \frac{}{1}$, тогда показатель степени на $\dot{A}_T$ в (55) всегда положительно . Таким образом, есть уникальная и позитивная ценность $\dot{A}_T$ что устанавливает $r(\dot{A}_T) = 0$, и это оптимальный выбор из $\dot{A}_T$ Достаточно медленный рост технологий всегда предпочтительнее стагнации.

Если ⇣ $= 1 + \frac{}{1}$, тогда показатель степени на $\dot{A}_T$ в (55) положительно для $\dot{A}_T < \overline{\dot{A}_T}$ и ноль для $\dot{A}_T \underline{\dot{A}_T}$. Так что если $r(\underline{\dot{A}_T}) > 0$, оптимальной скорости роста не существует: от $A_T = \hat{A}$ допуск, желательно, хотя бы на короткое время, чтобы технологии развивались как можно быстрее. Если $r(\underline{\dot{A}_T}) < 0$, существует уникальное значение $\dot{A}_T$ что устанавливает $r(\dot{A}_T) = 0$, он лежит в $(0, \underline{\dot{A}_T})$, и это оптимально.

Технически, если $r(\underline{\dot{A}_T}) = 0$, то любой $\dot{A}_T \underline{\dot{A}_T}$ оптимален при $A_T = \hat{A}$; но однажды $A_T > \hat{A}$, у нас будет $r(\underline{\dot{A}_T}) < 0$ и уникальный оптимальный темп роста, который является положительным, но конечным.

Б Доказательства

Б.1 Существование и единственность оптимальной политики

Б.1.1 Необходимые и достаточные условия

Все динамические задачи оптимизации, проанализированные в разделах 3–4, имеют один выбор переменная x и одна переменная состояния C. Ожидаемая полезность потока при t является $C_t ты(A_t, x_t)$ для дважды непрерывно дифференцируемая функция $ты(\cdot)$, строго вогнутый в x, с более низким Состояние Инада на x. Закон движения для C дается $C_t(A_t, \dot{A}_t, x_t)$ для дважды непрерывно дифференцируемая функция $(\cdot)$. A и $\dot{A}$ независимы от x, так что работайте просто как функции t.

Сдача в аренду*в*обозначим переменную costate на*C*, текущее значение Лагранжа соответствует

Тогда реагирование на проблему

$$Л(C_т, x_т, в_т, мк_т, т) = C_т ты(x_т, т)\ в_т C_т(x_т, т) + \mu_т(1\ x_т) \tag{56}$$

(немного злоупотребляя обозначениями, повторно используя*ты*(·) и(·) как функции времени), где$\mu_т$ представляет собой множитель Лагранжа на$x_т$. Мы навязываем$x_т$-1 ограничение, но не тот$x_т$ 0 ограничений, потому что последнее никогда не может связывать, по нижней Инаде состояние на*ты*(·).

(56) удовлетворяет условию вогнутости Мангасариана, что*Л*(·) везде вогнутый в*C*и*x*. Итак, применяя Капуто (2005), теоремы 14.3-4 и лемму 14.1,[22]данный непрерывные путиx2[0,1] иC2[0,1] сC_0=1 и$\dot{S}_т = C_т(x_т, т)$, мы

есть, что*x, C*путь является оптимальным, если — и, учитывая кусочную непрерывность*x*и*C*, только если — для некоторого кусочно-дифференцируемого пути*в*и некоторый кусочно-непрерывный путь μ 0, вообще*т*выполняются следующие условия первого порядка

$$\frac{@Л}{@x_т}(C_т, x_т, в_т, мк_т, т) = 0, \tag{57}$$

$$\frac{@Л}{@\mu_т}(C_т, x_т, в_т, мк_т, т)\ 0, \tag{58}$$

$$\mu_т \frac{@Л}{@\mu_т}(C_т, x_т, в_т, мк_т, т) = 0 \tag{59}$$

а также условие трансверсальности, что

$$\lim_{т!1} e^{\cdots т} в_т = \lim_{т!1} e^{\cdots т} в_т C_т = 0. \tag{60}$$

Кроме того, учитывая оптимальные пути*x*и*C*и соответствующие пути*в*иμ,*в* удовлетворит

$$\begin{aligned} \dot{v}_т = \cdots в_т \quad \frac{@Л}{@C_т} &= \cdots в_т\ ты(x_т, т)\ в_т \dot{S}_т \\ &= (\cdots + (x_т, т))в_т\ ты(x_т, т) \end{aligned} \tag{61}$$

---

[22]Caputo (2005) использует более общую нотацию текущей стоимости. Поскольку проблема управления в рука экспоненциально дисконтирована, здесь мы используем более простую запись текущего значения.

за исключением любых точек разрыва x, при котором $в$ будут иметь разные правая и левая стороны производные.

B.1.2 Интерпретация условия трансверсальности

Учитывая непрерывность $в$ путь, только пути x и μ определяется

$$x_T = \begin{cases} 1, & \frac{@u}{@x}(1, T) \quad \frac{@}{@x}(1, T)в \geq 0;_T \\ x_T: \frac{@u}{@x}(x_T, T) \quad \frac{@}{@x}(x_T, T)в_T = 0, & \text{в противном случае} \end{cases} \quad (62)$$

$$\mu_T = \frac{@u}{@x_T}(x_T, T) \quad \frac{@}{@x_T}(x_T, T)в_T \quad (63)$$

удовлетворяют (57)–(59) для всех T. Любой такой x путь хорошо определен, благодаря непрерывной дифференциации целостность $ты(\cdot)$ и$(\cdot)$ в x и тот факт, что $ты(\cdot)$ и$(\cdot)$ строго увеличиваются в x. Любой такой x путь также непрерывен справа во времени,

- дважды непрерывная дифференцируемость $ты(\cdot)$ и$(\cdot)$ (выражено как функции x, A, и, возможно, в случае$(\cdot)$, $\dot{A}$);
- непрерывная справа дифференцируемость $A(\cdot)$ во времени;
- непрерывная справа дифференцируемость $\dot{A}(\cdot)$ предполагается в сочетании с функции опасности, рассмотренные в разделе 4;

и теорема о неявной функции. Любой такой μ Тогда путь также непрерывен справа в время по композиции непрерывных функций. Чтобы показать, что существует оптимальное путь, и что только один такой путь является кусочно-непрерывным, теперь достаточно показать, что есть уникальный $в$ путь, для которого (60)–(61) удовлетворяются при соответствующем x путь (62) и его подразумеваемый C путь, и что соответствующий x путь кусочный непрерывен (фактически непрерывен справа).

Решение дифференциального уравнения (61) имеет вид

$$в_T = e^{\sum_0^T (\cdots + c) г} \left(в_0 \int_0^T e^{c \sum_0 (\cdots + д) дк} ты(x_c, c) дс \right) \quad (64)$$

$$=) в_0 = \int_0^T e^{\cdots c} C_c ты(x_c, c) дс + e^{\cdots T} C_T в_T. \quad (65)$$

Поскольку (65) непрерывно по t (ограниченностью $u(\cdot)$ и непрерывная эволюция из C) и справедливо для всех t, v удовлетворяет (60)–(61) i↵

$$v_0 = \int_0^1 e^{-\rho t} C_t u(x_t, t)\, dt. \qquad (66)$$

То есть, величина уменьшения вероятности катастрофы в момент времени 0 (по состоянию на момент времени 0) должна равняться ожидаемой полезности будущего (на момент времени 0, учитывая выживание до время 0).

Учитывая (62), v_t определяет x_t для всех t, и учитывая (61), v_t и x_t определять $\dot{v}_t$ для все t. Для данного v_0, поэтому существует уникальный путь v—и таким образом x, и таким образом из C—совместимо с (61)–(62). Теперь покажем, что существует по крайней мере одно значение из v_0 для которого (66) выполняется, учитывая соответствующее x и C пути. Для такого v_0, соответствующие переменные пути будут по построению удовлетворять (57)–(60), и, таким образом, представляют собой оптимум.

Б.1.3 Существование

Позволять $v(v_0)$ и $x(v_0)$ обозначают уникальные пути v и x совместимо с (61)–(62) для который $v_0(v_0) = v_0$. По (64), предел$_{v_0 \to 1} v_t(v_0) = 1$ для всех t 0. По (62), следовательно, для каждого t 0, есть $\tilde{v}_0$ такой что $x_t(v_0) = 1$ для всех $v_0 < \tilde{v}_0$. Позволять c 0 обозначает а время, в которое A_c 1, и выберите $\tilde{v}_0$ достаточно низко, что $\tilde{v}_c < 0$ и таким образом $x_c(\tilde{v}_0) = 1$. По (61), потому что $u(1, c)$ 0, $\dot{\tilde{v}}_t < 0$. Таким образом, мы имеем $\tilde{v}_t < 0$, и таким образом $x_t = 1$, для всех t c.

Теперь заметьте, что если $v_0 < \tilde{v}_0$, $v_t(v_0) < v_t(\tilde{v}_0)$ для всех t. В противном случае, по непрерывности из v по отношению ко времени, будет t с $v_t(v_0) = v_t(\tilde{v}_0)$, и интегрируя (61), где (62) заменено на x_t, позволит нам идентифицировать $v_0 = \tilde{v}_0$. Таким образом, если $v_0 < \tilde{v}_0$, $x_t(v_0)$ $x_t(\tilde{v}_0)$ для всех t 0. Из этого следует, что для некоторого достаточно низкого v_0, право-левая часть (66) превышает левую часть.

Для каждой рассматриваемой задачи оптимизации существует некоторая $\overline{U}$ по которому Допустимые значения правой части (66) ограничены сверху. Так, для $v_0 > \overline{U}$, левая часть (66) превышает правую часть.

Согласно (62), теорема о неявной функции дает нам, что x_T является непрерывным (действительно, непрерывно дифференцируемо, за исключением одной точки) в $в_T$ для всех T. (61) тогда подразумевает, что $\dot{v}_T$ непрерывен в $в_T$ для всех T, и таким образом, что $в_T(в_0)$, затем $x_T(в_0)$, а затем в конечном итоге правая часть (66) непрерывна по $в_0$ для всех T. Из этого следует теорема о промежуточном значении, что существует $в_0 2(в_0, в_0)$, для которого справедливо (66).

Б.1.4 Уникальность

Результат единственности Капуто (2005), теорема 14.4 (цитированная выше) не подразумевает непосредственно применим здесь, поскольку лагранжиан является линейным, а не строго вогнутым, в состоянии переменная C. К счастью, это можно легко исправить, определив переменную состояния как быть например C_2 не затрагивая никаких условий, необходимых для других результатов.

Уникальность (среди кусочно-непрерывных x пути) также следует непосредственно из наблюдения, что путь является оптимальным i↲ $в_0$ достигает своего максимально возможного значения и что, учитывая (57)–(60), $в_0$ определяет уникальный путь для каждой переменной.

Б.2 Долгосрочный $г_в$ и доказательство предложения 2

Б.2.1 Долгосрочное постоянство $г_в$ для всех

Из (61), потому что $в$ является переменной costate на C, он должен следовать закону движения

$$\dot{v}_T = (\dashrightarrow + _T) в_T \; ты(C_T)$$

$$=) г_{вT} = \dashrightarrow + (A_T, x_T) \qquad \frac{ты(A_T x_T)}{в_T}. \tag{67}$$

Позволять

$$\tilde{} ⌘ + 1.$$

Из (15), один раз x_T есть ли у нас интерьер

$$x_T = A_T^{\frac{\lrcorner + 1}{\tilde{}}} \; \bar{в}_T^{\tilde{} \frac{1}{}}. \tag{68}$$

Подстановка (68) в (67) дает

$$г_{вт} = г_в(в_т, т) ⌘ \begin{cases} \dashrightarrow + KA_т \frac{(\hookleftarrow)(1)}{\sim} в_т^{\bar{\sim}} + \frac{1}{1} в^{т1}, & б = 1; \\ \dashrightarrow + \text{бревно} \; A_т^{\hookleftarrow} \; \bar{v}_т \; \frac{1}{}в1_т, & =1, \end{cases} \quad (69)$$

где

$$K ⌘ \frac{1}{\sim} \quad \frac{1}{1} \quad \frac{1}{\sim}.$$

Если >1, напоминая, что $в_т$ монотонно возрастает и что $A_т !1$, центральный термин (69) исчезает. Также в этом случае $в$ ограничен сверху, поэтому он приближается к верхнему граница $в_\leftarrow$ по теореме о монотонной сходимости. Так что $\lim_{т!1} г_{вт}$ определяется, с

$$\lim_{т!1} г_{вт} = \dashrightarrow + \frac{1}{в_\leftarrow(1\;)}. \quad (70)$$

Этот предел не может быть положительным, потому что $в$ ограничен сверху и не может быть отрицательным, потому что $в$ увеличивается со временем. Так что $\lim_{т!1} г_{вт} = 0$, и $в_\leftarrow = \frac{1}{\dashrightarrow(1)}$.

Если <1, тогда $K<0$, а центральный член (69) растет по величине без связанный, фиксирующий $в$. $в$ поэтому также должен расти без ограничений, иначе $г_{вт}$ в конечном итоге отрицательно.

Теперь заметьте, что

$$\begin{aligned} г_{вт} &= KA_т \frac{(\hookleftarrow)(1)}{\sim} в_т^{\bar{\sim}} \; \frac{(\hookleftarrow)(1\;)}{\sim} г - г_{вт} ⌘ \frac{1/в_т г}{1 \; вт} \\ &= г_{вт} \dashrightarrow \frac{1/в_т}{1} ⌘ \frac{(\hookleftarrow)(1\;)}{\sim} г - г_{вт} ⌘ \frac{1/в_т г}{1 \; вт} \\ &= -г^2_{вт} + \frac{(\hookleftarrow)(1\;)}{\sim} г + - \dashrightarrow + \frac{1 ⌘}{\sim в_т} г_{вт} \dashrightarrow + \frac{1/в_т ⌘}{1} \frac{(\hookleftarrow)(1\;)}{\sim} г. \end{aligned}$$

Это дифференциальное уравнение имеет два устойчивых состояния, оба положительные. Так как $1/в_т!0$, квадратичная формула говорит нам, что эти устойчивые состояния приближаются $\dashrightarrow$ и $г(\hookleftarrow)(1)/$, причем первый является привлекательным, а второй — отталкивающим. Согласно (22), $\dashrightarrow$ выше и управляется как устойчивое состояние по условию трансверсальности (60). Тогда, поскольку пределы

$$\underset{т!1}{лим} г_в(г_в, т)>0 8г_в 2 \quad \frac{(\ \ )(1\ )}{\quad} г, ,$$

$$\underset{т!1}{лим} г_в(г_в, т)<0 8г_в< \quad \frac{(\ \ )(1\ )}{\quad}$$

определены и непрерывны в$г_в$, мы должны иметь

$$\underset{т!1}{лим} г_{вт}= \frac{(\ \ )(1\ )}{\quad} г. \tag{71}$$

В противном случае мы бы имели$г_в!1$, исключено монотонностью$в$, или$г_в!$, исключено выше.

The =1 случай аналогичен>1 случай. Дифференцируем (69) по время дает$г_{vt}$строго и непрерывно увеличивается в$г_{вт}$от1в$в_т$=0 к в$в_т$= 1. Таким образом, существует уникальный, положительный и отталкивающий «зависящий от времени устойчивый состояние» значение$г_в$(то есть$г_в$для которого$г_в(г_в, т)$ = 0), который уменьшается до нуля, как$т !1$. Пределы

$$\underset{т!1}{лим} г_в(г_в, т)>0 8г_в>0,$$

$$\underset{т!1}{лим} г_в(г_в, т)<0 8г_в<0$$

определены и непрерывны в$г_в$, и мы должны иметь

$$\underset{т!1}{лим} г_{вт}=0$$

избегать$г_в!1$или$г_в!1$.

B.2.2 Доказательство предложения 2

С ограничивающим поведением$г_в$зафиксировано асимптотическое поведение другого переменные следует прямолинейно. Подставляя (71) вместо$г_{вт}$в выражение (16) для $г_{хт}$(и замечая, что выражение охватывает все-1) производит

$$\underset{т!1}{лим} г_{хт}= \quad \frac{}{} г,$$

и добавление ↵ r затем производит предел $r_{Топор} = r_C$:

$$\lim_{т!1} r_{Kт} = \frac{↵}{} r. \tag{72}$$

Для коэффициента опасности переставьте (69), чтобы получить

$$в_т = \frac{ты(c)}{\dashrightarrow + \; т \; r_{вт}}, \tag{73}$$

и подставьте (73) в (26), чтобы получить

$$т = \begin{cases} \frac{\dashrightarrow + т r_{вт}}{} \frac{1}{1 C_{1}, т} & <1; \\ \frac{\dashrightarrow + т r_{вт}}{бревно(C_т)}, & =1. \end{cases}$$

Решение для $т$,

$$т = \begin{cases} \frac{(\dashrightarrow r_{вт})(1)}{1 C_1 \; т \; 1+}, & <1; \\ \frac{\dashrightarrow r_{вт}}{бревно(C_т) 1}, & =1. \end{cases}$$

В <1 случай, предел $r_в$ (71) и $C\ !1$ из (72) следует

$$\lim_{т!1} т = \frac{(\dashrightarrow (\; ↵)(1\;) r/)(1\;)}{+ \; 1}.$$

В =1 случай, замените 0 на $r_{вт}$ и заметим, что по (72)

$$\lim_{т!1} \frac{C_т}{e^{\frac{↵}{} rт}} = \underline{C}$$

для некоторых $\underline{C} > 0$, так что

$$\begin{aligned} \lim_{т!1} т т &= \lim_{т!1} \frac{\dashrightarrow r_{вт}}{бревно(C_т/e^{\frac{↵}{} rт}) + лог(e^{\frac{↵}{} rт}) \;\; /т \; 1/т} \\ &= \lim_{т!1} \frac{\dashrightarrow}{бревно(\underline{C}\;)/т + (\;↵) r \; 1/т} \\ &= \frac{\dashrightarrow}{(\;↵) r}. \end{aligned}$$

В.3 Доказательство предложения 4

Доказательство аналогично доказательству предложения 8а (Приложение В.7.1), которое обобщает

Предложение 3. Как там, $в_{\underline{A}+✏}$ непрерывен в ✏ и $\tilde{v}_{\underline{A}+✏}[✏] = в_{\underline{A}+✏}$ для всех ✏. В

Однако в этой настройке мы не можем предположить, что $\tilde{v}_A[\epsilon]$ слабо увеличивается в A или что $\tilde{v}_{\underline{A}}[\epsilon]$ $в_{\underline{A}}$ для всех ϵ. Поэтому мы будем использовать другую стратегию для равномерного связывания $\tilde{v}_A[\epsilon]$, для $A \in [A, \underline{A} + \underline{\epsilon}]$, в интервале, максимум и минимум которого сходятся к $в_{\underline{A}}$ как $\epsilon \to 0$.

Позволять $\underline{T}$ обозначить время, в которое $\underline{A}_T = A$. Ускорение $\tilde{A}_{(\cdot)}$, показывая технологию темпы роста $\dot{A} > \dot{A}_{\underline{A}}$ до уровня технологии $\underline{A} + \epsilon$, показывает рост технологий со скоростью $\dot{A}$ сквозь времена

$$(\underline{T}, \underline{T} + \epsilon/\dot{A}).$$

В более общем плане, путь ускорения достигает уровня технологий. $A \in [\underline{A}, \underline{A} + \epsilon]$ во время

$$\tilde{t}(A) ⌘ \underline{T} + (A \quad \underline{A})/\dot{A}.$$

$\tilde{v}_A[\epsilon]$ — максимальное значение выживаемости $\tilde{v}_{\tilde{t}(A)}$, через возможные пути политики, достижимые в $\tilde{t}(A)$ данный технологический путь $\tilde{A}_{(\cdot)}[\epsilon]$. Таким образом, его можно ограничить снизу одним таким достижимая ценность выживания, такая как достигнутая при заданном $x_T = 1$ для $T \in [\tilde{t}(A), T + \epsilon/\dot{\underline{A}})$. С $\tilde{A}_T > 1$ на протяжении всего этого интервала эта нижняя граница в свою очередь строго больше, чем ценность выживания в $\tilde{t}(A)$ данный *нет* флнизкая полезность, полученная в течение всего интервала.

Помня, что $\tilde{v}_{\underline{A}+\epsilon}[\epsilon] = в_{\underline{A}+\epsilon} > 0$ для любого ϵ, таким образом, мы имеем

$$\begin{aligned}
\tilde{v}_A[\epsilon] & \quad \int_{\tilde{t}(A)}^{\underline{T}+\epsilon/\dot{A}} e^{\cdots(T \tilde{t}(A))} \quad e^{P_{\tilde{t}(A)}^{T} \tilde{A}_{\leftarrow} \quad c \dot{A} \downarrow дс} ты(\tilde{A}_T) дт \\
& + e^{\cdots(T + \epsilon/\dot{A} \tilde{t}(A))} \quad e^{P_{\tilde{t}(A)}^{\underline{T}+\epsilon/\dot{A}} - \tilde{A}_{\in} \dot{A} \downarrow дс} В_{\underline{A}+\epsilon} \\
& > \underline{В}[\epsilon] ⌘ e^{\cdots(T + \epsilon/\dot{A} \tilde{t}(A)) \, A} \quad e^{P_{\tilde{t}(A)}^{\underline{T}+\epsilon/\dot{A}} \tilde{A}_{\leftarrow} \dot{A}_{\in} \downarrow дс} В_{\underline{A}+\epsilon}.
\end{aligned} \tag{74}$$

Потому что $\tilde{t}(A)$ увеличивается в A, $\underline{B}_A[\epsilon]$ увеличивается в A, так $в_A[\epsilon]$ $\underline{в}_{\underline{A}}[\epsilon]$ для всех $A \in [\underline{A}, \underline{A} + \epsilon]$.

$\tilde{v}_A[\epsilon]$ может быть ограничена сверху (недопустимым) значением выживания, достигнутым при $\tilde{t}(A)$ учитывая, что, в $T \in [\tilde{t}(A), \underline{T} + \epsilon/\dot{A})$, поток полезности равен его супремуму $1/(\quad 1)$ и коэффициент опасности равен 0:

$$\tilde{v}_A[✏] < \frac{1}{1} \int_{\tilde{\tau}(A)}^{T+✏/\dot{A}} e^{\cdots(T\tilde{\tau}(A))} ДT + e^{\cdots(T+✏/\dot{A}\tilde{\tau}(A))} B \quad A+✏$$

$$< \bar{B}_A[✏] ⌘ \frac{1}{1} \int_{\tilde{\tau}(A)}^{T+✏/Г} e^{\cdots(T\tilde{\tau}(A))} ДT + B_{A+✏}. \qquad (75)$$

Потому что $\tilde{\tau}(A)$ увеличивается в A, $B_A[✏]$ уменьшается в A, так $B_A[✏]$ $B_{A+✏}[✏]$ для всех $A 2$ $[A, A+✏]$.

Из (74), (75) следует непрерывность $B_{A+✏}$ в ✏, и тот факт, что $\tilde{v}_{A+✏}[✏] = B_{A+✏}$ для все ✏,

$$лим_{✏!0} B_A[✏] = лим_{✏!0} \bar{B}_{A+✏}[✏] = B_A.$$

Далее доказательство продолжается по принципу доказательства предложения 8 после (96), с

$$\tilde{x}_A[✏] = мин\ 1, \quad \bar{A}^{+1}\dot{A}\tilde{v} \quad A[✏] ⌘ \frac{1}{+\ 1} ⌘$$

вместо (97), в конечном итоге получая

$$_{A,\dot{A}} = (A, \dot{A}, \tilde{x}_A) \tilde{A}_1 (A, A_A, x_A) \dot{A}_1 \quad A, \qquad (76)$$

где $\tilde{x}_A$ дается формулой (31), при $A = A$, с $\dot{A}$ вместо $\dot{A}_A$.

Если $A\ A$, (76) сводится к

$$\bar{1} A^{(1)(1)} B \quad A \quad \frac{1}{+\ 1} \dot{A} \frac{11}{+1} \quad \dot{A}_A \frac{1}{+1} 1 ⌘.$$

С $\dot{A} > \dot{A}_A$, это отрицательно, если $< 1 + \frac{}{1}$, ноль, если $= 1 + \frac{}{1}$, и положительно, если $> 1 + \frac{}{1}$.

Если $A < A$, так что $x_A = 1$, и $\dot{A}$ достаточно мал, чтобы поддерживать $\tilde{x}_A = 1$, затем (76) сводится к

$$\bar{A}\ \dot{A}_1 \dot{A}_1. \quad A$$

С $\dot{A} > \dot{A}_A$, это отрицательно, если < 1, ноль, если $= 1$, и положительно, если > 1.

В.4 Доказательство предложения 5

Предположим, что $P_{\leftarrow}$-1, и, от противного, что у нас нет $C_{\leftarrow}= 1$.

Из-за провала $C_{\leftarrow}= 1$, существует возрастающая и неограниченная последовательность времен, $т_н!1$, такой что $C_{т_н}$-$C8н$ 1.

Рассмотрим последовательность уровней потребления $нC8н$ 1. Так как $нC!1$, к $P_{\leftarrow}$-1 у нас есть

$$\lim_{н!1} P(н\overline{C}) = \lim_{н!1} \lim_{A!1} \frac{@}{@x} \quad A, \frac{н\overline{C}}{A} \quad \frac{(н\overline{C})}{A \dashrightarrow (\ 1)} \text{-}1. \tag{77}$$

По D5, $\frac{@}{@x}(A, x)$ слабо увеличивается в x для любого A. Так

$$P(C_{тн})\text{-}P(нC) \quad — \quad \frac{C_{тн}}{н\overline{C}} \quad \text{-}P(нC)\overline{н} \qquad 8н, \tag{78}$$

где первое неравенство следует из того, что $нC\ C_{тн}$ для каждого $н$, и второе следует из $C\ C_{тн}$ для каждого $н$. По (77), $P(нC)н <1$ для достаточно большого $н$, поэтому согласно (78) и А4 существует $\underline{н}$ такой что

$$\frac{@}{@x} A_{тн}, \frac{C_{тн}}{A_{тн}} \quad \frac{C_{тн}}{A_{тн} \dashrightarrow (\ 1)} <1\ 8н > н. \ \_$$

С $в_т$ не может превышать $\frac{1}{\dashrightarrow(1)}$,

$$\frac{@}{@x} A_{тн}, \frac{C_{тн}}{A_{тн}} \quad в_{тн} < A_{тн} \quad C_{тн}\ 8н > н. \quad \_$$

Это совместимо с оптимальностью только если $x_{тн}$ = 1. Но это невозможно для достаточно большой $н$, с $C_{тн}$=$A_{тн}x_{тн}$-C и $\overline{Л}$им$_{н!1}A_{тн}$= 1.

Предположим, что $P_{\leftarrow}$>1 и, от противного, что $C_{\leftarrow}= 1$. Затем есть некоторые $\underline{C}$ такой что $P(\underline{C})$>1:

$$\lim_{A!1} \frac{@}{@x} \quad A, \frac{\underline{C}}{A} \quad \frac{\underline{C}}{A \dashrightarrow (\ 1)} > 1.$$

Итак, есть $\underline{A}$ такой что

$$\frac{@}{@x} A, \frac{\underline{C}}{A} \quad \frac{1}{\dashrightarrow (\ 1)} > A\underline{C} \tag{79}$$

для всех A $\underline{A}$. Кроме того, поскольку левая часть слабо увеличивается в $\underline{C}$ от D5 а правая часть строго убывает по $\underline{C}$, (79) справедливо для всех A $\underline{A}$ и C $\underline{C}$. По А4 и предположению, что $C_{\leftarrow} = 1$, есть $\underline{T}$ такой что

$$\frac{@}{@x} A_{T,T} \frac{C}{A_T} \quad \frac{1}{\cdots(\ 1)} > AC_{TT} \qquad 8T\ T. \_ \tag{80}$$

Наконец, оптимальность требует

$$A_{TT}^{1} x \qquad \frac{@}{@x_T} A_{T,x_T} B \quad T \quad 8T$$

$$=)(A_T x \quad T)^{1} \ / B_T \qquad \frac{@}{@x_T} A_{T,x_T} (\quad T \quad A_T \quad , x)_T,$$

с окончательным неравенством, поскольку, согласно D5, $\frac{@}{@x} x$. Данный $C_{\leftarrow} = 1$, с B является ограниченный сверху, следует, что $T \,!\, 0$. $C_T \,!\, 0$ и $C_T \,!\, 1$, B_T приближается к своему верхняя граница $\frac{1}{\cdots(1)}$.

Поэтому из (80) следует, что при достаточно больших T,

$$\frac{@}{@x} A_{T,T} \frac{C}{A_T} B > A_T \qquad {}_T C_T.$$

Это несовместимо с оптимальностью. Таким образом, если $P_{\leftarrow} > 1$, невозможно, чтобы $C_{\leftarrow} = 1$.

Б.5 Доказательство предложения 6а

Б.5.1 Предварительные

Оптимально установить x_T=1 до тех пор, пока, в x=1, предельная отрицательная полезность потока уменьшение x слабо превышает предельную ожидаемую полезность этого действия путем уменьшения уровень опасности:

$$A_T^{1} \qquad \frac{@}{@x}(A_{T},1) B\ ._T \tag{81}$$

Оптимально установить x_T<1 до тех пор, пока (81) не выйдет из строя, сохраняя

$$A_T^{1} \quad x_T \quad = \frac{@}{@x}(A_{T}, x_T) B_T \tag{82}$$

$$=) x_T = A_T^{\ 1} \qquad \frac{@}{@x}(A_{T}, x)_T \ B_T^{\ \text{⌘} 1}. \tag{83}$$

Единственность оптимального пути показана в Приложении Б.1.

Б.5.2 Доказательство того, что $лим_{т!1} х_т = 1$

Мы покажем, что существует время $\underline{т}$ такой что $в_{\underline{т}}$ -0. Затем следует немедленно что $х_т$=1 для $т$ - $т$. _

Позволять

$$Т ⌘ А_1 (1) \frac{1}{1}$$

обозначить время, в которое $А_Т$= (1) $\frac{1}{1}$, и в котором поэтому $ты$($А_Т$) =1. Если $в_Т$-0, результат следует немедленно. Поэтому предположим, что $в_Т$>0.

Для $т < Т$,

$$в_т = З_1 е^{\cdots(ст)} \; Р_с \; _{тдк} ты(С_с)дс$$

$$= З_т^Т е^{\cdots(ст)} \; Р_с \; _{тдк} ты(С_с)дс + е^{\cdots(Тт)} \; _Т \; Р \; _{тдк} в_Т. \quad (84)$$

С $ты$($С_с$)-$ты$($А_с$)-1 для $с$ - $Т$, первый член (84) отрицателен — действительно, интеграл по $с$ значений, которые являются отрицательными для всех $с$. Интеграл сокращается. по величине, когда для всех $с$, $ты$($С_с$) заменяется на 1 и коэффициент дисконтирования $е^{\cdots(ст)}$ $_с$ Р $_{тдк}$ заменено его минимальным значением в диапазоне, а именно скидкой фактор в $Т$. Так

$$в_т <(т \; Т + в_Т) е^{\cdots(Тт)} \; _Т \; Р \; _{тдк}$$

$$=) в_{Тв} < 0. \; _Т$$

Это доказательство, по общему признанию, «слишком буквально воспринимает модель», предполагая, что технология рост всегда был экспоненциальным и поэтому жизнь не стоила того, чтобы ее прожить до некоторого момента в прошлом. Тем не менее, динамика, которую он прямо иллюстрирует, не должна быть спорным. Когда >1, пропорциональные жертвы в потреблении — уменьшаются до $х$— нести тем большие затраты на коммунальные услуги, чем ниже базовый уровень потребления. В начале время, обесцененная ценность цивилизации $в$ и базовый уровень потребления $А$ были оба показателя низкие, поэтому большие жертвы ради безопасности были бы неоптимальными.

Б.5.3 Доказательство того, что $\lim_{T\to\infty} x_T = 0$ если η_A ограничено сверху 1

Обобщая (83), независимо от того, x_T-1 ограничение связывает нас

$$x_T \geq A_T^{\frac{1}{\eta}} \left(\frac{\partial}{\partial x}(A_T, x)_T \; B_T \right)^{\frac{1}{\eta-1}}. \tag{85}$$

Мы покажем, что если $\eta_A(\cdot)$ ограничено выше 1 , правая сторона имеет верхнюю граница, которая падает до 0, как (по A4) $A_T\to\infty$.

Потому что D1 положительно, по D2 и D5 имеем $\frac{\partial}{\partial x}(A_T, x_T)$ (A_T, x_T).

правая часть, таким образом, ограничена сверху

$$A_T^{\frac{1}{\eta}} \quad (A_T, x_T) B_T^{\frac{1}{\eta-1}}. \tag{86}$$

Фиксация x и B, эластичность этой верхней границы по отношению к A есть (1 $\eta_A(A, x))/\eta$. Поскольку здесь это ограничено ниже 0, (86) стремится к 0 как $A\to\infty$. Окончательно, B_T положительно для всех T 0, потому что по A1 и A2 $A_T>1$ для всех T 0 (рендеринг $B_T>0$ осуществимо с $x=1$ постоянно), и B_T не падает, потому что достаточный меры предосторожности в отношении новых технологий, например, запрет на их использование, позволяют потребителю поддерживаться без увеличения риска, к D4. Поэтому, если $\eta_A(\cdot)$ ограничено сверху 1 , сохраняя условие оптимальности (85) как $A_T\to\infty$ требует $x_T\to 0$.

В.6 Доказательство предложения 7

Если $\lim_{\kappa\downarrow 1} P(\kappa)<1$, есть $\bar{\kappa} > 1$ такой, что

$$\lim_{T\to\infty} \frac{\partial}{\partial x}\left(A_T, \frac{T^{\frac{\bar{\kappa}}{\eta}}}{A_T}\right) \frac{T^{\frac{\bar{\kappa}}{\eta-1}}}{A_T(\eta-1)} < 1. \tag{87}$$

Выбирать $\kappa\in(1,\bar{\kappa})$. Предположим, что $\exists \underline{T}: C_T > T^{\frac{\kappa}{\eta-1}} \; \forall T > \underline{T}$. Затем происходит увеличение и неограниченная последовательность времен, $\{T_n\}\to\infty$, такой что

$$C_{T_n} \leq T_n^{\frac{\kappa}{\eta-1}} \quad \forall n\; 1. \tag{88}$$

Обратите внимание, что

$$\text{лим}_{н!1} \frac{@}{@x} A_{т_н}, \frac{т_н^{\frac{к}{1}}}{A_{тн}} \quad \frac{т_н^{\frac{к}{1}}}{A_{тн}(1)}$$

$$-\text{лим}_{т!1} \frac{@}{@x} A_т, \frac{т^{\frac{к}{1}}}{A_т} \quad \frac{т^{\frac{к}{1}}}{A_т(1)} \cdot т_{кк}^{\frac{}{1}} = 0, \tag{89}$$

где неравенство следует из того, что по D5 $\frac{@}{@x}(A, x)$ слабо увеличивается в x, и предел перед $т_{кк}^{\frac{}{1}}$ член меньше 1 по (87).

По (88), (89) и тому факту, что $в_т < \frac{1}{(1)}$ для всех $т$, есть $н$ таким образом, что для всех $н \underline{н}$,

$$\frac{@}{@x} A_{тн}, \frac{C_{тн}}{A_{тн}} в_{тн} < A_{тн} C_{тн} \cdot$$

Это совместимо с оптимальностью только если $x_{тн} = A_{тн} x_{тн} = 1$. Но это невозможно. для достаточно большого $н$, по (41) и (88).

Так что для некоторых $к > 1$,

$$9\underline{т} : C_т > т^{\frac{к}{1}} \; 8т > т. \tag{90}$$

Итак, (90) справедливо для $к = 1$ тоже.

Дано (90) для некоторых $к > 1$, у нас есть, для некоторых $\underline{т}$ и некоторые $\underline{к} 2 (1, к)$, что для всех $т > \underline{т}$

$$(A_т x_т)_1 < т_к$$

$$=) \quad \frac{@}{@x}(A_т, x_т) x_т в_т < т_к$$

$$=) _т в_т < т_к$$

$$=) _т < т_к. \tag{91}$$

Первое следствие следует из того, что $A_1 x_т \; _т \frac{@}{@x}(A_т, x_т) в_т$ или нет x является внутренним. Второе следствие следует из того, что $\frac{@}{@x}$ x D1 и D5.

Третье следует из того, что $в_т$ в конечном итоге оказывается положительным и не падает до нуля.

$_{T}$равномерно ограничена от 0 до Tпо макс.$_{A2[A_0,A_{\bar{T}}]}(A,1)$, который существует и конечно в силу непрерывности$(\cdot)$ (D3). Из этого и из (91) следует, что $C_1>0$.

Если Лим$_{k''1}$ $P(k)>1$, есть $k<1$ и c такой что

$$\frac{@\downarrow}{@x}A_{T}, \frac{T^{1}⌘}{A_T}\quad\frac{T^{\frac{k}{1}}}{A_T\dashrightarrow(1)}>1\ 8T>\underline{c}. \tag{92}$$

Предположим от противного, что $@\underline{T}\ :C_T<T^{\frac{1}{1}}\ 8T>T.$ Затем происходит увеличение и неограниченная последовательность времен, $\{T_H\}\ !\ 1$, такой что

$$C_{T_H}\ T_1\ {}_H^{1}\quad 8H\ 1. \tag{93}$$

Обратите внимание, что

$$\text{лим}_{H!1}\frac{@\downarrow}{@x}A_{T_H}, \frac{T_H^{1}⌘}{A_{T_H}}\quad\frac{T_H^{1}}{A_{T_H}\dashrightarrow(1)}$$

$$\text{лим}_{T!1}\frac{@\downarrow}{@x}A_T, \frac{T^{1}⌘}{A_T}\quad\frac{T^{\frac{k}{1}}}{A_T\dashrightarrow(1)}\cdot T^{\frac{1k}{1}}=1, \tag{94}$$

где неравенство следует из того, что по D5$\frac{@}{@x}(A, x)$ слабо увеличивается в x, и предел перед $T^{\frac{1k}{1}}$ член больше 1 по (92).

По (93), (94) и тому факту, что $B_T6!0$, есть H такой что

$$\frac{@\downarrow}{@x}A_{T_H}, \frac{C⌘_{T_H}}{T_H}B_{T_H}>A_{T_H}KA\ \dot{}_{T_H}$$

Это несовместимо с оптимальностью. Так что

$$9\underline{T}:C_T<T^{\frac{1}{1}}8T>T. \tag{95}$$

По (95) и (41), $x_T!0$. Итак, существует $T\ \bar{T}$ таким образом, что для всех $T>\bar{T}$, выбор из x это интерьер

$$\frac{@}{@x}(A_T, x_T)B_T=A_1x_T\quad T$$

и поэтому, по (95),

$$\frac{@}{@x}(A_T, x_T)x_T B_T = \frac{C_1}{T} > 1/T.$$

С⌘x⌘@x, $\overline{@x}$ –

$$⌘_x(A_T, x_T)(A_T, x_T)B_T > 1/T 8T\ T.$$

Напомним, что выбор интерьера x_T подразумевает, что $B_T > 0$, что B ограничен сверху $\frac{1}{(1)}$, и что $T > 0$ по D1. Так что $⌘_x > 0 8T\ T$. Так что если $⌘_x$ ограничен сверху $⌘ \overline{x}$,

$$(A_T, x_T) > \frac{(1)}{⌘_x} \cdot \frac{1}{T} 8T\ T.$$

Так $C_1 = 0$.

B.7 Доказательство предложения 8

Выберите допустимый технологический путь $A(\cdot)$ и функция опасности $(\cdot)$.

B.7.1 Доказательство части а

Выбирать $\underline{A}$, $\dot{A}$ с $\dot{A} > \dot{A}_{\underline{A}}$. Определять $\tilde{x}_A[✏]$ как $\tilde{x}_A$ заданное ускорение $\tilde{A}(\cdot)[✏]$, и т. д.

B_T слабо возрастает и непрерывна (действительно дифференцируема; см. Приложение B.1) по T. С A_T является непрерывным, возрастающим и обратимым в T, B_A непрерывен и слабо увеличивается в A. $B_{\underline{A}+✏}$ поэтому непрерывен и слабо возрастает в ✏.

С технологического уровня $\underline{A}+✏$ далее, технологические пути, и, таким образом, пути как потребление, так и уровень опасности, идентичны при $A(\cdot)$ и $\tilde{A}(\cdot)$. Так что для любого ✏ (включая 0), $\tilde{v}_{\underline{A}+✏}[✏] = B_{\underline{A}+✏}$. Из этого следует тот факт, что $\tilde{v}_A[✏]$ слабо увеличивается в A, и тот факт, что $\tilde{v}_{\underline{A}}[✏]\ B_{\underline{A}}$ для всех ✏, у нас есть это для всех ✏

$$\tilde{v}_A[✏] \in [\tilde{v}_{\underline{A}}, \tilde{v}_{\underline{A}+✏}] \checkmark [B_{\underline{A}}, B_{\underline{A}+✏}]\ 8A \in [\underline{A}, \underline{A}+✏]. \tag{96}$$

Тогда по непрерывности $B_{\underline{A}+✏}$ в ✏, для любого $✏_1$ есть $\bar{✏}$ такой что $|B_{\underline{A}+✏} - B_{\underline{A}}| < ✏_1 8✏ < \bar{✏}$.

Адаптация (62),

$$\tilde{x}_A[✏] = \text{мин} \quad 1, x: \quad \frac{@}{@x}(A, x)A_1 \quad x = \frac{1}{\tilde{v}_A[✏]} \quad ⌘ . \qquad (97)$$

По (97) и А2, $\tilde{v}_A[✏]$ $в_{\underline{A}}$ >0 для всех ✏ 0, $A2[\underline{A}, \underline{A} + ✏]$. По D3, неявное теорема о функции и непрерывность min(·), $\tilde{x}_A[✏]$ непрерывен в $\tilde{v}_A[✏]$. Так что по (96) и предложение, следующее за ним, для любого ✏₂ есть ✏ таким образом, что для всех ✏ < ✏,

$$\tilde{x}_A[✏] \text{ мин } 1, x: \quad \frac{@}{@x}(A, x)A_1 x \quad = \quad \frac{1⌘}{} \quad < ✏_2 \ 8A2[A, A + ✏]. \ в_{\underline{A}}$$

Снова по D3, теореме о неявной функции и непрерывности min(·), второй член в абсолютном значении непрерывен по A. Так что для любого ✏₃ есть ✏ такой что для всех ✏ < ✏,

$$\tilde{x}_A[✏] \text{ мин } 1, x: \quad \frac{@}{@x}(\underline{A}, x)\underline{A_1} x \quad = \quad \frac{1⌘}{в_{\underline{A}}}$$

$$= \tilde{x}_A[✏] \ x \quad _{\underline{A}} < ✏_3 \ 8A2[\underline{A}, \underline{A} + ✏].$$

При этой равномерной сходимостью, поскольку

$$X[✏] \ X = \int_{\underline{A}}^{\underline{A}+✏} A, \tilde{x}_A[✏]\tilde{A}_1 дA \quad \int_{\underline{A}}^{\underline{A}+✏} (A, x \quad \dot{A}_1 \quad _A) \dot{A} дA,$$

c(·) непрерывна по обоим аргументам, так как x_A непрерывен в A, и так как $\dot{A}_1 \quad _A$ непрерывна справа во времени и, таким образом (в силу непрерывности и монотонности $A(\cdot)$) в A,

$$_{\underline{A}, \tilde{A}} \ ⌘\text{лим}_{✏ \to 0} \frac{X[✏] \ X}{✏} = (\underline{A}, x)\tilde{A}_{\underline{A}} \ (\underline{A}, x_{\underline{A}})\dot{A}_A \qquad \frac{1}{}$$

$$= _{\underline{A}} \quad \dot{A}_1 \quad _{\underline{A}} . \quad \cdot_1 \quad A$$

Это доказывает (а).

B.7.2 Доказательство части b

Позволять $\tilde{A}(\cdot)$ быть ускорением $A(\cdot)$ от A к A. По определению ускорения и определение кумулятивного риска,

$$X = X + \int_{\underline{A}}^{A\ddagger} A, \tilde{x} \ \dot{A}_{1A} \quad _A \ (A, x_A)\dot{A}_1 \quad _A \ дA. \quad ⌘ \qquad (98)$$

Для всех $A 2 [\underline{A}, \overline{A})$, у нас есть $\tilde{\gamma}_A$ $в_A$, и таким образом, по (97) (опуская «[✏]” аргументы) и Д5, $\tilde{x}_A$ - x_A. D1, D2 и D5 подразумевают, что($\cdot$) слабо увеличивается в x, так $(A, \tilde{x}_A)$ - (A, x_A). Так

$$A, \tilde{x}\ \dot{A}^1_A \quad _A\ (A, x_A)\dot{A}^1 \quad _A - \quad _{A,A_A} \quad 8A 2 [\underline{A}, \overline{A}].$$

Это доказывает (б).

В.7.3 Доказательство части с

Если $\overline{A} < 1$, интеграл (98) конечен. Так что дан технологический путь $A(\cdot)$ для которого $X = 1$ и ускорение до $A < 1$, $\overline{X} = 1$. Это доказывает первую часть (с).

Для доказательства второй части (с) достаточно найти функцию опасности($\cdot$) и технологический путь $A(\cdot)$ для которого $X = 1$ и пара ускорений $\tilde{A}(\cdot)$ к $\overline{A} = 1$, для один из которых $\tilde{X}$ конечен и для другого из которых $\tilde{X}$ бесконечно. Мы уже сталкивался с обоими.

В первом случае рассмотрим функцию опасности $(A_T, x_T) = A_T x_T$, обсуждается следуя предложению 7. Как там обсуждалось, кумулятивный риск при оптимальной политике равен затем бесконечность для любого технологического пути, в конечном итоге ограниченного выше нуля.

В последнем случае рассмотрим функцию опасности (5) — $(A_T, x_T) = \bar{A}^{\dashv}\ _T x\ _T$ — с базовый технологический путь $A_T = (T\ 1)_\kappa (T\ 0)$ и ускорение $\tilde{A}_T = (T\ 1)_{\tilde{\kappa}} (T\ 0)$, где

$$\kappa - \frac{+\ 1}{(\dashv)(\ 1)} < \tilde{\kappa}.$$

Чтобы убедиться, что это ускорение, $A_T = (T\ 1)_\kappa \Rightarrow T = 1 + A_T^{\frac{1}{\kappa}}$, так $\dot{A}_T = \kappa(T\ 1)_{\kappa 1} \Rightarrow \dot{A}_A = \kappa A_{\frac{\kappa\ \ 1}{\kappa}}$, который увеличивается в κ данный $A > 1$ (что справедливо для $T > 0$).

Как показано в (28), здесь $X = 1$ и $\tilde{X} < 1$.

Б.8 Безопасность в избыточности

Б.8.1 От дискретного к непрерывному

Предположим, что единица продукции несет постоянную вероятность потока $\bar{}$запуска экзистенциальная катастрофа, так что при отсутствии каких-либо гарантий вероятность того, что это не вызывает катастрофу после cЕдиницы времени - это e^{c}. Чтобы быть последовательным с спецификация дискретного времени, согласно которой вероятность того, что это вызовет катастрофу после 1 единица времени равна $п$, у нас есть 1 $e = п$ и таким образом $\bar{=}$ журнал(1 $п$).

С $\frac{1}{x_т}$ единицы охраны поддерживаются вокруг $т$, так как каждая единица умножает вероятность катастрофического отказа за единицу времени в раз $б2(0,1)$, у нас есть это вероятность того, что катастрофа будет предотвращена до тех пор, пока $т+с$ равно $e^{б^{\frac{1}{x}}хс}$.

Вероятность того, что $А_тх_т$ одинаково защищенные единицы продукции все избегают катастроф-трофей до $т+с$ таким образом

$$e^{б^{\frac{1 x т}{}} х т с \, А_т х_т} = e^{б^{\frac{1 x т}{}} х т А_т х_т с}. \tag{99}$$

Таким образом, вероятность катастрофы $с$ задано локально постоянное $А, х$ равно 1-(99), и уровень опасности — вероятность катастрофы в единицу времени — в момент времени $т$ именно это

$$_т \text{⌘предел}_{с!0} \; 1 \; e^{б^{\frac{1 x т}{x т}} А_т х_т с} / с = \bar{А}\tilde{} \quad _т х_т б^{\frac{1 x т}{x т}}.$$

Сдача в аренду $б$⌘ бревно($б$)>0 урожайность

$$_т = \bar{А}_т х_т e^{б^{\frac{1 x т}{x т}}}.$$

В.8.2 Доказательство предложения 9

Согласно Приложению В.1, существует единственный оптимальный путь. По рассуждениям, следующим за (10), оптимальный выбор x равно 1 до (уникального) времени, в которое

$$\frac{@u}{@x_т}(А_т, x_т) = \frac{@}{@x_т}(А_т, x_т) в_т \tag{100}$$

в $x_т$=1, после чего оптимальный выбор $x_т$ является внутренним и сохраняет равенство (100).

Дифференцируя функцию полезности и функцию риска (46), имеем

$$A_1 x_{TT} = \bar{A}_T e^{б_1 x_T} \overline{x_T} \; 1 + \frac{б}{x_T} \; в_T$$

$$=) \quad \frac{1}{в_T} = \bar{A} \quad _T e^{б_1 x_T} \overline{x_T} \quad x_T + б x_T^{1} . \tag{101}$$

Потому что $в_T$ монотонно возрастает и ограничена сверху, она асимптотически положительна и постоянна, по теореме о монотонной сходимости.

Мы должны иметь $C_T!1$. Если мы этого не сделаем, то существует неограниченная последовательность времен T_H и уровень потребления $\bar{C}$ такой что

$$x_{T_H} - C/\bar{A}_{T_H} 8H. \tag{102}$$

Подставляя (102) в (101) и вспоминая, что $A_{T_H}!1$, это будет означать, что правая часть (101) стремится к 0 по горизонтали $\{T_H\}$, и таким образом, что это не асимптотически положительный.

Из (101),

$$\frac{1}{в_T} = _T C_1 (1_T + б/x_T).$$

С $C_1!1$, x_T не может быть отрицательным, и $1/в_T б! 1$, следует, что $_T!0$.

С $C_T!1$ и $_T!0$, $в_T! \bar{v}$.

Разделите обе части (101) на $\bar{A}_0$, и берем логарифм, а затем предел. С

$$-бревно \; A_0 \; \frac{1}{(+1)^{-}} \; ,$$

у нас есть

$$\lim_{T!1} гт \; б \; \frac{1 \; x_T + бревно \; x}{x_T} \; _T + б_T \; x_1 \quad = -$$

$$=) \lim_{T!1} 1 \; \frac{x_T}{x_T} \; T = \lim_{T!1} \frac{б}{г - /T + бревно \quad x_T + б x_T^{1} \; /T}$$

Кроме $г$, члены в знаменателе с правой стороны должны сходиться к 0. Этого можно было бы избежать только в том случае, если бы существовала неограниченная последовательность времен T_H

через которыйx_t росли по крайней мере экспоненциально со временем, что невозможно, или сокращались по крайней мере экспоненциально со временем, что привело бы к тому, что правая часть (101) стала бы ноль. Так что

$$\lim_{t!1} \frac{x_t}{x_t} t = \frac{б}{г}$$

$$=) \lim_{t!1} x_t t = \lim_{t!1} (1 \; x_t) \quad \frac{б}{г} = \frac{б}{г}$$

$$=) \quad \lim_{t!1} x_t \frac{г}{б} t = 1,$$

сx_t!0. Тогда из функции риска следует, что в пределе падает до 0

в геометрической прогрессии$г$(1)<0.

С Динамика перехода для моделирования

Для моделирования динамики перехода полезно найти$\dot{x}_t$и$\dot{r}_t$как функции tиx_tв режиме, гдеxэто интерьер.

Функция опасности (5), используемая в разделах 3.2–3.4 и применяемая для моделирования Рисунки 1 и 2 представляют собой частный случай функции опасности (44), используемый для моделирования рисунка 3, с✏=1. Таким образом, приведенные ниже расчеты применимы ко всем симуляциям.

ВОК:

$$\frac{@u}{@x_t}(A_t, x_t) = \frac{@}{@x_t}(A_t, x) в_t$$

$$=) A_1 x_{tt} \qquad = A_{\leftarrow} \; _t x_t 2(\; 1) \; 1 \; (1 \; x)_{✏} \quad _t \; + \; ✏x(1 \; x \quad _t)^{✏ \; 1} \quad в_t. \quad ⌘$$

Перестановка и дифференциация дает

$$в_t = \frac{1}{=} \frac{A_1 \leftarrow x_t^2}{(\; 1) \; 1 \; (1 \; x_{✏} \quad _t) \; + \; ✏x_t(1 \; x \; _t)_{✏} 1} \tag{103}$$

$$=) \; v_t = в_t(1 \quad \leftarrow) г + (2 \quad) \frac{\dot{x}_t}{x_t} \tag{104}$$

$$✏ \frac{(✏ + \; 1)x \quad _t}{(\; 1)(1 \; x_t)_{1 \; ✏} + 1 \; + (✏ + \; 1) x_t 1 \; x_t} \quad \frac{\dot{x}_t}{} \; ⌘ \; .$$

Из условия первого порядка относительно переменной состояния C_T,

$$\dot{v}_T = в_T(\cdots + _T)\ ты(c_T)$$
$$= в_T \cdots + \bar{A}_{\hookleftarrow 1}\quad _T x_T \quad (1\ (1\ x_T)\ )\quad ✏\ ⌘\ \frac{(A_T x_T)^1 \quad 1}{1}. \qquad (105)$$

Подставляем (103) в (104) и (105), приравниваем результаты и решаем для $\dot{x}_T$

урожайность

$$\dot{x}_T = x_T(\ 1)(1\ x_T)_{1✏}{+1}\ + (✏+\ 1)x_T \qquad (1\ x_T)$$
$$(2\ \ )\ (\ 1)(1\ x_T)_{1✏}{+1}$$
$$+ (✏+\ 1)x_T \quad (1\ x_T)\ ✏(\ (✏+\ 1)x_T)x_T \qquad ⌘1$$
$$\cdots + \bar{A}_{\hookleftarrow}\ _T x\ _T^{\ 1}\ (1\ (1\ x_T)✏)\ r(1\ \hookleftarrow) \qquad (106)$$
$$\frac{(A_T x_T)_1\ 1}{1} \bar{A}_T^{\hookleftarrow +1}\quad x_+^{T\ \ 2}\quad (\ 1)\quad (1\ (1\ x_T)✏) + ✏x_T(1\ x_T)_{✏ 1} \quad ⌘ .$$

Дифференцируя функцию опасности (44) по отношению к T урожайность

$$\dot{}_T = \bar{A}_{\hookleftarrow}\quad _T x^T \frac{1\ (1\quad x_{✏T})}{x_T}\quad \hookleftarrow r + (\ 1)_T + ✏\frac{\dot{x}}{x_T}\quad \frac{(1\ x)_{✏T}}{1\ (1\ x_T)_{✏}1\ x_T}\ \frac{\dot{x}_T ⌘}{} . \qquad (107)$$

Скрипты для воспроизведения рисунков 1, 2 и 3 с использованием (106) и (107),

и оценка C_1 Ниже, на Рисунке 1, представлены:

https://philiptrammell.com/static/ERAG code.zip.