Stanford Economic Review

Stanford Undergraduate Economics Journal

Volume 12

Editors and Staff

Note from the Editors

On behalf of the *Stanford Economic Review* Editorial Board, we are honored to present the twelfth volume of Stanford University's undergraduate economics journal.

Our publication has continued to benefit from robust growth in both readership and volume of submissions, allowing us to reach new audiences across the globe. During the 2023-2024 academic year, we have also made it a priority to strengthen our presence in the Stanford community. Most notably, in an effort to bridge the gap between Stanford faculty and students, we started organizing ”Econ Conversations”—informal meetings between professors and small groups of undergraduates.

This journal issue features undergraduate work on topics ranging from the effects of geopolitical risk on cryptocurrency markets to the dynamics between pitchers and batters in Major League Baseball. Commentary pieces written for our publication over the course of this year have contributed to the dialogue surrounding several key issues, opportunities, and policies impacting the United States and the world at large, including the rise of antimicrobial-resistant infections, the role played by voluntary carbon offsets in the fight against climate change, and the future of the US dollar in the global financial system.

We would like to take this opportunity to once again extend our gratitude to the authors whose writing appears in this journal edition and on the commentaries section of our website. Finally, we would like to thank the Stanford Economics Association (SEA) and the Stanford Economics Department for their thoughtful support of our publication and its mission.

Karthick Arunachalam and Eric Gao
2023-24 Editors in Chief

Contents

Analyzing the Effects of Geopolitical Shocks on the Cryptocurrency Market

Tucker Saland

Department of Economics, Princeton University

***Abstract*— The Russo-Ukrainian war had an asymmetric effect on cryptocurrency prices, with some cryptocurrencies appreciating and functioning as hedges against geopolitical risk, even as the rest of the market experienced a downturn. In the aftermath of the war, exchange-level data indicates that individuals in Russia and Ukraine purchased increased amounts of Tether, a stablecoin pegged to the US dollar. This paper has two objectives: (1) examine whether there exists a geopolitical risk premium in the cross-section of cryptocurrency returns using a rolling regression of excess weekly returns on the geopolitical risk index by Caldara, Iacoviello (2018) and control factors, and (2) understand the role of stablecoins during times of geopolitical turmoil. Exposure to geopolitical risk is not constant over time, and investors are willing to pay a premium for assets with predictable returns in response to geopolitical shocks. In addition, there is evidence that stablecoins are used as stores of value during geopolitical events and were potentially used as mediums of exchange in Russia to bypass economic sanctions. This research contributes to the growing body of cryptocurrency asset pricing literature. It underscores the need for regulators to carefully consider cryptocurrencies when making policy decisions, given their potential for misuse.**

I. Introduction

Since its inception in 2009 as the first decentralized digital currency, Bitcoin has inspired the creation of numerous other cryptocurrencies, with CoinMarketCap (2022) listing 18,000 coins on the site as of March 2022. With their extreme volatility and high average returns relative to traditional currencies (Brauneis & Mestel, 2018), cryptocurrencies are viewed as a new asset class. As a result, recent economic literature has focused on developing an appropriate asset pricing model for cryptocurrencies to complement modern investment theory by identifying risk factors that drive excess returns.

Economic debate has centered around whether these cryptocurrencies should be considered currencies, investment vehicles, or speculative assets. Since Bitcoin returns follow closely with financial conventions (Fama et al., 2019), some economists argue that Bitcoin and other cryptocurrencies merely function as speculative assets. However, a key argument among cryptocurrency enthusiasts is that these digital currencies function as stores of value, given their use as mediums of exchange. Empirical evidence has shown that Bitcoin appreciates during inflation shocks, supporting the idea that it may be a hedge against inflation (Choi & Shin, 2022). While recent research has primarily focused on the effects of monetary policy and economic shocks on the price of cryptocurrencies, the impact of geopolitical tension on cryptocurrencies has yet to be explored.

During times of political uncertainty, investors tend to reallocate their funds from risky investments to safe-haven assets, such as the US dollar, treasury bills, or gold, which tend to retain their value even when the overall market experiences a downturn (Baur & McDermott, 2010). Since cryptocurrencies are largely unregulated and highly volatile, one might expect that these currencies would not be safe hedges against geopolitical risk. However, given their ease of access and decentralized nature, people will turn to these currencies as a means of exchange to bypass economic sanctions and governmental restrictions during time of political turmoil. Stablecoins, which peg their value to safe-haven assets, such as the US dollar or Japanese yen, will be especially attractive to individuals in affected regions due to their price stability and utility as stores of value. Data from the Kaiko, a Paris-based cryptocurrency research provider, lends credence to this theory, with ruble-denominated bitcoin volume surging to a 9-month high of 1.5 billion RUB on March 22, 2022, following the Russian invasion of Ukraine (Godbole, 2022). Similar increases in ruble-denominated trading volume were observed for Tether (Figure 1), a reserve-backed stablecoin whose price is anchored to the US dollar. Following the invasion of Ukraine, Russians were willing to pay a premium relative to its underlying value to protect against the depreciating value of the ruble (Kharif, 2022).

With the rapid emergence of cryptocurrencies and recent case studies, such as the Russo-Ukrainian War, indicating that stablecoins may function as safe-haven assets in response to geopolitical shocks, it is essential to understand how the dynamics of the cryptocurrency market change in response to geopolitical shocks. The focus of this paper is twofold. The first objective is to examine whether a geopolitical risk premium exists in

I would like to thank Professor Jonathan Payne and Yinuo Zhang for their continued advice and support throughout this project. I am also grateful for Oscar Torres-Reyna, Narek Alexanian, Filippo Palomba, and Benjamin Cai for their invaluable feedback and critiques of my work. Lastly, I would like to thank the entire ESS department for their countless hours spent helping me to debug and refine my models.

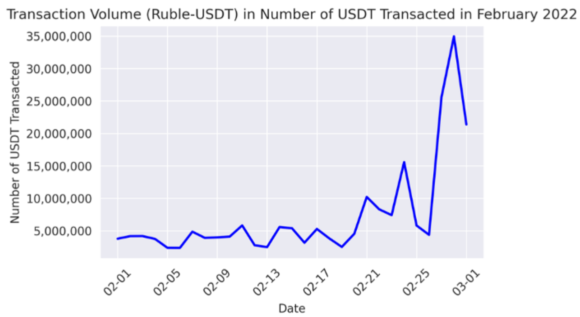


Fig. 1: Ruble-Denominated Tether Volume in February 2022 Note. Following the Russian invasion of Ukraine on February 24, 2022, there was a marked increase in trading volume. Data aggregated from Binance (2022).

the cross-section of cryptocurrency returns by calculating β^{GPR} – a measure of a given coin's exposure to changes in geopolitical risk. The second objective is to understand the role stablecoins play during geopolitical turmoil by first considering the global cryptocurrency market before focusing more specifically on the Russo-Ukrainian War.

I hypothesized that since assets with the most positive geopolitical betas act as risk hedges, with spikes in geopolitical risk coinciding with greater returns, risk-averse investors should be willing to pay a premium for these high-β^{GPR} assets. During political turmoil, when other assets perform poorly, the marginal utility of additional income increases. Assets with the most negative geopolitical betas have lower payoffs when geopolitical risk is high. Thus, investors holding low-β^{GPR} assets should be compensated with a premium relative to high-β^{GPR} assets since they are exposed to more systematic risk, consistent with findings from Long et al. (2022). However, contrary to this initial hypothesis, there is no statistically significant difference in returns between low-β^{GPR} and high-β^{GPR} assets when accounting for other known risk factors in cryptocurrency returns.

Unlike existing literature, I examine the variation in geopolitical risk exposure among cryptocurrencies over time and find that most cryptocurrencies have geopolitical betas that are highly volatile. Cryptocurrencies with the greatest variance in their geopolitical betas are the least predictable in response to changes in geopolitical risk and markedly outperform cryptocurrencies with the most stable geopolitical betas in returns, even when adjusting for other risk factors. These excess returns cannot be fully explained by current asset pricing models, suggesting that there exists a safe asset premium in the cross-section of cryptocurrency returns. One potential rationale for this safe asset premium is that when β^{GPR} is constant, investors can anticipate how the asset will react to changes in geopolitical risk and can diversify their portfolios accordingly. Investors holding assets with volatile geopolitical betas are compensated for this additional risk, resulting in abnormal returns.

On a macroscopic scale, changes in geopolitical risk are not a significant determinant of stablecoin use. VAR estimates indicate that a shock in geopolitical risk has a negligible effect on Tether volume relative to an equivalent shock in implied cryptocurrency market volatility. This result suggests that in the global cryptocurrency market, investors reallocate their funds into stablecoins during times of market uncertainty, avoiding the time and monetary costs of withdrawing entirely while readjusting their portfolios to better align with their risk tolerance. In Russia and Ukraine, there is evidence that stablecoins were used as stores of value and mediums of exchange during the war. Stablecoins allow individuals to protect against depreciating currency value in their home country, benefiting those affected by conflict. However, if used to circumvent economic sanctions, as suggested by the findings in this paper, regulators should carefully consider these cryptocurrencies when making policy decisions.

II. Literature Review

There is substantial economic literature on asset pricing in traditional markets, which serves as the basis for recent research on cryptocurrency pricing. The cross-section of returns among stocks has been extensively studied in academic literature since Sharpe introduced the Capital Asset Pricing Model (CAPM) in 1964, postulating a linear relationship between market risk and expected stock returns (Fama & French, 2004). In the subsequent years, numerous anomalies were discovered in stock returns that could not be adequately explained by the CAPM model, such as size (Banz, 1981), value (Basu, 1977), and momentum (Jegadeesh & Titman, 1993). Empirically, it has been shown that small-cap stocks, on average, outperform large-cap stocks, value stocks tend to outperform growth stocks, and winning stocks (those that have performed well previously) tend to outperform the market. These findings prompted Fama and French (1993) to develop a three-factor model which combines the value premium with size and the excess market risk factor of the CAPM model, which was later extended by Carhart (1997) to include an additional momentum factor.

These pricing models, however, cannot be directly applied to crypto assets as they fundamentally differ from traditional financial instruments, such as stocks, which are priced based on the present value of expected dividends and cash flows (Campbell & Shiller, 1988). As a result, economic literature has started to examine crypto-specific characteristics to explain the cross-section of cryptocurrency returns. Liu et al. (2020) constructed three common risk factors that are specific to the cryptocurrency market, adopting a similar approach to Fama and French (1993). They discovered that the combined effect of size and momentum can explain most of the cross-sectional variation in cryptocurrency returns. Liebi (2022) expanded on this research, finding a value anomaly in cryptocurrency returns using the active addresses-to-network ratio as a valuation metric.

In 2018, Caldara and Iacoviello developed the geopolitical risk index (GPR), which measures geopolitical risk through the algorithmic tracking of news headlines. Within the existing literature, research into the effects of geopolitical risk on cryptocurrencies has remained relatively sparse and has primarily focused on Bitcoin. Using a Bayesian Graphical Structural Vector Autoregressive (BSGVAR) approach, Aysan et al. (2019) found that GPR has predictive power on both returns and volatility of Bitcoin. Long et al. (2022) built upon this research by identifying variations in geopolitical risk exposure among cryptocurrencies.

My research is unique in that, unlike the aforementioned studies, I explore how exposure to geopolitical risk evolves over time. Long et al. (2022) construct long-short portfolios to determine whether there is a difference in returns between cryptocurrencies with differing levels of exposure to geopolitical risk at a fixed point in time. However, their analysis does not consider that a coin's sensitivity to geopolitical risk may not be stable over time. In addition, my research focuses on stablecoin adoption following geopolitical events. This area has not been broadly explored in extant economic literature and has significant implications for monetary policy and the effectiveness of economic sanctions. However, this analysis is not without its limitations. This paper employs time series forecasting to examine the relationship between stablecoin adoption and geopolitical risk, which, while helpful in exploring patterns in the data, cannot directly establish causality. Moreover, while Tether is used in this analysis due to its long-standing and complete volume history and its position as the largest stablecoin by market capitalization (CoinMarketCap, 2022), adoption rates for Tether may not necessarily reflect stablecoins as a whole.

III. Data

Daily cryptocurrency data was collected from CoinMarketCap (Fig 2), which contains pricing information and trading volume aggregated from all major exchanges. CoinMarketCap is a reliable source for cryptocurrency-related data and has been cited exhaustively in economic literature (Liu et al., 2022). The data ranges from 01/01/2018 to 10/31/2022, and numerous filters were applied to mitigate potential inaccuracies. The subset of data used for cross-sectional analysis excludes stablecoins. It consists of all coins with a market capitalization greater than 1 million dollars in all observations and a trading history longer than six months. As a proxy for the risk-free return, I use the one-month treasury bill rate gathered from FRED (2022), similar to the approach used by Fama & French (2015).

Geopolitical risk (GPR) is defined as "the risk associated with wars, terrorist acts, and tensions between states that affect the normal and peaceful course of international relations." (Caldara & Iacoviello, 2018). I account for this using the novel GPR index pioneered by Caldara & Iacoviello. The index counts the frequency of articles related to geopolitical risks in leading international newspapers and spikes around major geopolitical events, such as the Gulf War, 9/11, and the onset of the COVID-19 pandemic (fig3), suggesting that it accurately captures political instability. The data contains measures of geopolitical threats and acts in addition to the composite index.

The Cryptocurrency Volatility Index (CVI) is used in this paper to measure overall volatility in the cryptocurrency market. The CVI is the cryptocurrency industry's analog of VIX, an index that tracks the market's expectation of 30-day volatility. Kim et al. (2021) used similar measures of cryptocurrency volatility, and I will be using daily CVI pricing data from 3/31/2019 to 10/31/2022 (Figure 5) as a proxy for overall volatility in the cryptocurrency market.

In order to examine the effects of the Russo-Ukrainian War more precisely, ruble-denominated and hryvnia-denominated Tether volumes were obtained through Binance, one of the largest and most well-established cryptocurrency exchanges (Sun et al., 2019). Binance was used due to its easily accessible and segmented

Variable	Mean	Standard Deviation	Standard Deviation Between	Standard Deviation Within
Closing Price ($)	236.61	3462.55	3023.98	1678.11
Market Capitalization ($)	8.76×10^{8}	1.83×10^{10}	9.12×10^{9}	1.22×10^{10}
24-Hour Trading Volume ($)	8.28×10^{7}	1.40×10^{9}	6.80×10^{8}	9.60×10^{8}

Fig. 2: Summary of Daily Cryptocurrency Data
Note. 1,560,526 total observations, 1,984 cryptocurrencies in the sample, Mean Number of Days for cryptocurrencies=786.56 days

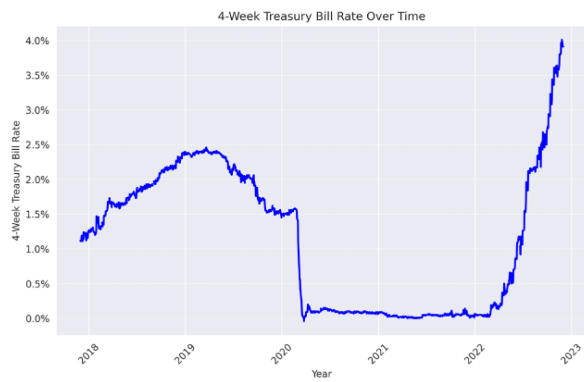


Fig. 3: 4-Week Treasury Bill Rate Over Time
Note. The Federal Reserve slashing the Fed Funds Rate in response to the COVID-19 pandemic resulted in a precipitous drop in the T-Bill rate during 2020.

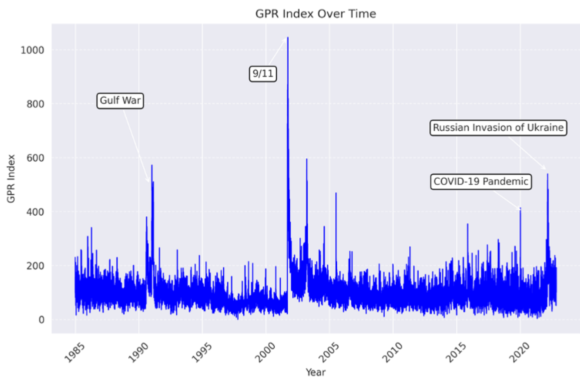


Fig. 4: GPR Index Over Time
Note. The GPR index constructed by Caldara & Iacoviello (2018) spikes around major geopolitical events, such as 9/11 and the Russo-Ukrainian War.

Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
CVI Price	1346	85.886	18.921	50.325	170.551

Fig. 5: Price of CVI Over Time
Note. Data was aggregated from Investing.com (2022) and spans 3/31/2019 to 10/31/2022.

data, though it may not be fully representative of the overall cryptocurrency market. In addition, there has been evidence of manipulated trading volumes on some exchanges in recent years (Fratrič et al., 2022). From this data, there is no way to definitively determine that this trading activity originated from Russia and Ukraine, respectively. Despite these shortcomings, this data should provide a more detailed look at stablecoin adoption in Russia and Ukraine due to the war.

In order to capture attention towards the Russo-Ukrainian War among individuals in affected regions, I gathered Google Trends data from 10/03/2021 to 01/01/2023. Using keyword data from Google, two indices were constructed to capture interest in the war in Russia and Ukraine. I defined the *RussianWarIndex* to be $\frac{\textbf{(War+Ukraine)}}{2}$, where *War* is the keyword score for (the Russian word for war) among Russian users, and Ukraine was similarly defined as the keyword score for ´ (the Russian word for Ukraine) among Russian users. An analogous measure was defined for Ukraine, with the keyword interest in Russia used in place of Ukraine. Google provides weekly scaled scores between 0 and 100 for these keywords, with 100 indicating that a given week had the highest number of keyword searches in the sample. Daily data was extrapolated from these weekly scores using the Chow-Lin methodology (Chow & Lin, 1971).

USD-RUB and USD-UAH exchange rates from 10/03/2021 to 01/01/2023 were aggregated from Yahoo! Finance (2023) to investigate the impact of exchange rate fluctuations on stablecoin adoption. Linear interpolation was used for missing values, such as weekends and holidays, to align with the available cryptocurrency data.

IV. Methodology

A. Asset Pricing

The methodology for this study is inspired by Carhart (1997), who developed a four-factor model for asset pricing to account for pricing anomalies that the CAPM model did not capture. Using a similar approach to Long et al. (2022), I calculate β^{GPR}, a given coin's exposure to geopolitical risk, using a rolling regression of excess weekly returns on change in GPR and control factors. Weekly returns were used to reduce the impact of short-term noise and volatility while maintaining sufficient data points for analysis.

The results from Liu et al. (2020) show that size and momentum effect are strong in the cryptocurrency market. One factor that Fama & French used in their model is HML (High Minus Low), which is the difference between the return on a portfolio of high book-to-market ratio stocks and a portfolio of low book-to-market stocks. Since there is no book-to-market ratio for cryptocurrencies, the Network Value to Transactions (NVT) ratio is used as an analogous measure for the cryptocurrency market. The NVT ratio is defined as $\frac{\textbf{Market Cap}}{\textbf{Transaction Volume}}$, and if the NVT ratio for a given currency is high, then the market cap is outpacing the on-chain transaction volume. Thus, the currency may be overpriced relative to its value as a medium of exchange, resulting in lower future returns. The resulting regression equation is given as follows:

$$R_{i,t} - R_{f,t} = \alpha_{i,\tau} + \beta_{i,\tau}^{GPR}\Delta GPR_t + \beta_{i,\tau}^{M}(R_{m,t} - R_{f,t}) + \beta_{i,\tau}^{SMB}SMB_t + \beta_{i,\tau}^{HML}HML_t + \beta_{i,\tau}^{WML}WML_t + \epsilon_{i,t}.$$

$R_{i,t}$ is the return on a particular coin, $R_{f,t}$ is the riskless return measured as the one-month treasury bill date expressed as a weekly return, $R_{m,t}$ is the market return measured as the value-weighted return of all available coins, and ΔGPR_t is the percentage change in the weekly moving average of the GPR index. The τ indicates that the β's are fixed for a year using a rolling-window approach. The other regressors are calculated using an approach similar to Carhart (1997). SMB_t, HML_t, and WML_t are the returns on zero-investment, factor-mimicking portfolios for size, NVT ratio, and daily momentum.

More precisely, each week, the coins are split into deciles based on the moving average of each characteristic for the previous week. The size factor (*SMB*) is the difference in returns between equal-weighted portfolios consisting of the smallest and largest coins based on market capitalization. One prevailing theory in the literature is that this size factor functions as an illiquidity premium since small-cap coins carry more liquidity risk, which is compensated with greater returns (Liu et al., 2022). Similarly, each week, the coins are divided into deciles

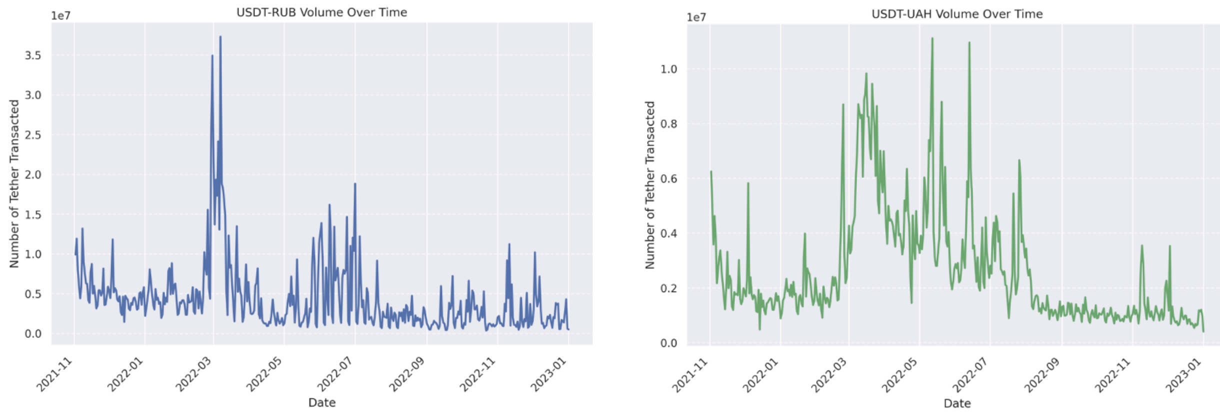


Fig. 6: USDT-RUB and USDT-UAH Trading Volume
Note. There was a significant increase in USDT-RUB and USDT-UAH trading volumes around the onset of the war in February 2022. There is a similar rise in USDT-UAH trading volume in July of 2022, which coincides with a 25% devaluation of the hryvnia relative to the dollar as a result of the war (Gorodnichenko, 2022).

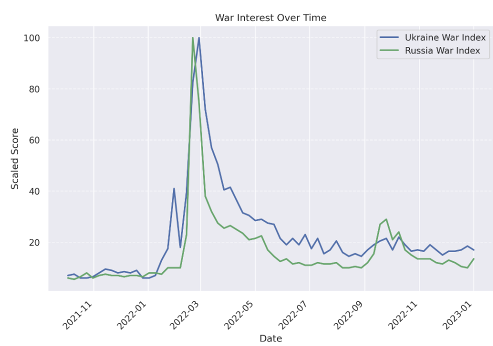


Fig. 7: War Indices Over Time
Note. Data aggregated from Google Trends (n.d.). Both indices reach their apex around the onset of the war in February 2022, indicating that the indices likely proxy war interest.

based on the average $\frac{\textbf{Transaction Volume}}{\textbf{Market Cap}}$ over the previous week, with *HML* equal to the difference in returns between portfolios constructed from the highest and lowest deciles. This factor accounts for the value anomaly observed in cryptocurrency returns, with research (Liebi, 2022) showing that crypto assets with a high active addresses-to-network value ratio yield higher returns on average than crypto assets with a low active addresses-to-network value ratio and comparable size. The last factor included in the regression is *WML* (Winners Minus Losers), which accounts for the momentum anomaly in cryptocurrency returns. Coins are separated into deciles based on the three-week trailing returns, and the spread in returns between portfolios formed from the top and bottom deciles is equal to *WML*.

In this regression, β^{GPR} denotes a given coin's exposure to geopolitical risk. A positive β^{GPR} means that spikes in GPR are associated with greater returns and suggests that the coin is a hedge against increases in GPR. Conversely, a negative β^{GPR} suggests that a rise in GPR results in lower returns for a given coin. It is plausible that investors place more value on assets that have higher returns during periods of political turmoil since, during these times, when other investments are performing poorly, additional income is at a premium. Assuming that state prices increase with GPR, economic theory suggests that investors holding low-B^{GPR} assets should be compensated with a premium relative to higher-β^{GPR} assets since they bear additional systematic risk. As a result, it is expected that β^{GPR} will be negatively correlated with future returns, which would align with findings from Long et al. (2022).

To test this hypothesis, all cryptocurrencies are sorted into deciles on β^{GPR} weekly, and a long-short portfolio is constructed that longs the lowest decile and shorts the highest decile. Similar long-short portfolios are constructed based on the standard deviation of β^{GPR} over time to consider alternative hypotheses. These long-short portfolios are evaluated using two models: (i) a single-factor market model and (ii) a three-factor model using the size, value, and momentum factors presented by Liu et al. (2020). A statistically significant alpha provides evidence of excess returns beyond what would be expected from the factor model and indicates that there exists a geopolitical premium in the cryptocurrency market.

$$R_{p,t} - R_{f,t} = \alpha_{p,\tau} + \beta^{M}_{p,\tau}(R_{m,t} - R_{f,t}) + \epsilon_{i,t}$$

$$R_{p,t} - R_{f,t} = \alpha_{p,\tau} + \beta^{M}_{p,\tau}(R_{m,t} - R_{f,t}) + \beta^{SMB}_{p,\tau} SMB_t + \beta^{HML}_{p,\tau} HML_t + \beta^{WML}_{p,\tau} WML_t + \epsilon_{i,t}$$

B. Stablecoin Adoption

In order to examine the effects of geopolitical risk on stablecoin adoption, I employ time series forecasting

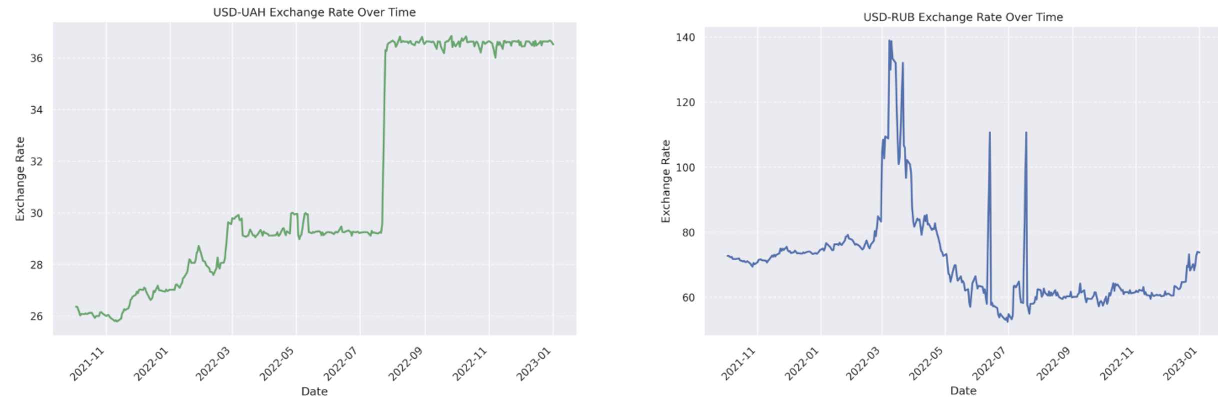


Fig. 8: Exchange Rates Over Time
Note. Russia experienced substantial currency depreciation following the onset of the war, as shown by the peak in the USD-RUB exchange rate during February 2022. Unlike in Russia, the USD-UAH exchange rate did not remain floating after the onset of the war. The National Bank of Ukraine (NBU) fixed the exchange to prevent widespread panic. In July 2022, the NBU had to devalue the hryvnia by roughly 25% to relieve pressure on foreign exchange reserves, improve the competitiveness of Ukrainian exports, raise more fiscal revenue, and align the official and cash-based exchange rates, as seen on the graph (Gorodnichenko, 2022).

techniques. First, to determine whether geopolitical risk helps forecast Tether volume, I construct a Vector Autoregressive (VAR) model, which fits a multivariate time-series regression of each of the variables on lags of itself and lags of all other variables. A VAR model captures the interdependence and dynamic relationships among the endogenous variables in the system. I include the Crypto Volatility Index (CVI) as a proxy for overall volatility in the cryptocurrency market, along with the GPR index, and daily Tether volume. The number of lags in the model was selected using the Akaike Information Criteria (AIC). Each variable was checked for stationarity using the Augmented Dickey-Fuller and Phillips–Perron tests, and the model residuals were examined for serial correlation using a Lagrange multiplier test. Log transformations were applied to all endogenous variables to measure impulse responses in terms of elasticities, and robustness checks were conducted in levels to ensure the choice of transformation did not drive the findings. The VAR model has three equations, one for each endogenous variable.

$$\ln(\mathrm{Vol}_t) = \alpha_1 + \sum_{i=1}^{p} \beta_{1i} \ln(\mathrm{Vol}_{t-i}) + \sum_{i=1}^{p} \gamma_{1i} \ln(\mathrm{GPR}_{t-i}) + \sum_{i=1}^{p} \theta_{1i} \ln(\mathrm{CVI}_{t-i}) + \epsilon_1$$

$$\ln(\mathrm{GPR}_t) = \alpha_2 + \sum_{i=1}^{p} \beta_{2i} \ln(\mathrm{Vol}_{t-i}) + \sum_{i=1}^{p} \gamma_{2i} \ln(\mathrm{GPR}_{t-i}) + \sum_{i=1}^{p} \theta_{2i} \ln(\mathrm{CVI}_{t-i}) + \epsilon_2$$

$$\ln(\mathrm{CVI}_t) = \alpha_3 + \sum_{i=1}^{p} \beta_{3i} \ln(\mathrm{Vol}_{t-i}) + \sum_{i=1}^{p} \gamma_{3i} \ln(\mathrm{GPR}_{t-i}) + \sum_{i=1}^{p} \theta_{3i} \ln(\mathrm{CVI}_{t-i}) + \epsilon_3$$

Impulse response analysis and Granger causality tests were then conducted to determine whether there is independence, unilateral Granger causality, or bilateral Granger causality between the variables. This process was then repeated using the subcomponents of the composite GPR index. *GPRActs* and *GPRThreats* were used as endogenous variables in the model in place of *GPR* to determine whether investors react differently to the realization of a geopolitical event compared to the potential threat of one. I hypothesize that geopolitical risk will help forecast Tether volume. I anticipate that both *GPRActs* and *GPRThreats* will have predictive power for future volume. However, I expect a shock in *GPRActs* to have a more significant and longer-lasting impact on volume since it quantifies geopolitical events that have tangible effects on individuals. While this time series analysis cannot establish causality between GPR and stablecoin adoption, if GPR is found to be a meaningful predictor of Tether volume, this may suggest that individuals turn to stablecoins when geopolitical risk is high.

The final part of my analysis uses the Russo-Ukrainian War as a case study to determine whether geopolitical events lead to increased stablecoin adoption in impacted regions. Two VAR models were estimated to identify whether ruble-denominated and hryvnia-denominated Tether volumes change in response to interest in the war. For Russia and Ukraine, a VAR model was estimated using currency-denominated Tether volume and log transformations of each country's war index and exchange rate as endogenous variables. The number of lags in the models was selected using the HQIC and AIC. All variables were checked for stationarity, and model residuals were examined for autocorrelation. The difference in logarithms was taken for the USD-UAH exchange rate to ensure stationarity. The Johansen cointegration test was used to test for cointegrating variables before estimating the VAR models. For Russia and Ukraine, respectively, the VAR model is a set of equations of the form:

$$\text{Vol}_t = \alpha_1 + \sum_{i=1}^{p} \beta_{1i}\text{Vol}_{t-i} + \sum_{i=1}^{p} \gamma_{1i} \ln(\text{USD-RUB}_{t-i}) + \sum_{i=1}^{p} \theta_{1i} \ln(\text{War}_{t-i}) + \epsilon_1$$

$$\text{Vol}_t = \alpha_1 + \sum_{i=1}^{p} \beta_{1i}\text{Vol}_{t-i} + \sum_{i=1}^{p} \gamma_{1i}(\ln(\text{USD-UAH}_{t-i}) - \ln(\text{USD-UAH}_{t-i-1})) + \sum_{i=1}^{p} \theta_{1i} \ln(\text{War}_{t-i}) + \epsilon_1$$

V. Results

A. GPR in the Cross Section of Cryptocurrency Returns

In examining whether geopolitical risk is priced into the cross-section of cryptocurrency returns, it is essential to first determine the persistence of geopolitical risk sensitivity over time and how it is distributed across cryptocurrencies. A definitive pattern emerged in the data when analyzing the rolling geopolitical betas. For the vast majority of cryptocurrencies in the sample, B^{GPR} was not constant over time. While large market-cap, well-established cryptocurrencies, such as Bitcoin, had geopolitical betas that tended to remain stable over time, smaller cryptocurrencies generally had highly volatile geopolitical betas. This trend is illustrated in Figure 7, with Dogecoin, a smaller market cap cryptocurrency known for its high price volatility, exhibiting high variance in geopolitical sensitivity over time. This observation was supported empirically with a regression of the logarithm of market cap on the logarithm of the standard deviation of geopolitical betas over the preceding six months and a constant. The coefficient on $ln(\sigma(B^{GPR}))$ was statistically significant at the 1% level and indicates that a 1% increase in the standard deviation of past geopolitical betas is associated with a 0.41% decrease in market cap (Figure 20). For most cryptocurrencies, geopolitical betas were close to zero, with a plurality of cryptocurrencies exhibiting slightly positive betas.

Smaller cryptocurrencies have greater geopolitical betas in absolute value on average, which suggests that they are comparatively more sensitive to geopolitical risk than larger, more established cryptocurrencies. One potential explanation for this phenomenon is that smaller cryptocurrencies are inherently more volatile during periods of geopolitical uncertainty since they do not have the same user base and infrastructure as more broadly accepted cryptocurrencies. For instance, Ethereum has achieved widespread adoption and has numerous use cases beyond merely facilitating payments. There is a network of decentralized finance applications on the Ethereum blockchain, with various financial services, such as lending, borrowing, and trading, being offered in a decentralized, trustless manner (Leonhard, 2019). When faced with geopolitical uncertainty, larger cryptocurrencies with greater real-world utility are less likely to experience demand shocks, resulting in smaller geopolitical betas in absolute value. This insight suggests that major geopolitical events could lead to further market concentration among large, established cryptocurrencies, which could fundamentally change the dynamics of the cryptocurrency market.

Contrary to my initial hypothesis, the long-short portfolio constructed on B^{GPR} did not have an alpha statistically different from zero. However, constructing the long-short portfolio on the standard deviation of B^{GPR} over the preceding six months resulted in a statistically significant alpha, even when accounting for known factors in the cross-section of cryptocurrency returns (Figure11). These results were robust to changes in the window size used to calculate standard deviation and indicate that a safe asset premium exists in the cryptocurrency market. Investors are willing to pay a premium for an asset that moves predictably in response to fluctuations in geopolitical risk, even in the highly volatile and speculative cryptocurrency market.

B. Global Stablecoin Adoption

The following outcome of interest in this study is stablecoin adoption in response to changes in geopolitical risk. A VAR model was estimated with endogenous variables $ln(TetherVolume)$, $ln(CVI)$, and $ln(GPR)$ using a lag length of 15 days (Figure 22). There is bilateral Granger causality between $ln(TetherVolume)$ and $ln(GPR)$ (Figure 12), which supports the hypothesis that geopolitical risk is useful in forecasting future stablecoin volume. There is also bilateral Granger causality between $ln(TetherVolume)$ and $ln(CVI)$, which lends credence to the theory that increased volatility in the cryptocurrency market results in increased stablecoin since existing investors reallocate their funds into stablecoins, avoiding the time and monetary costs of withdrawing entirely while readjusting their portfolio to better align with their risk tolerance. Impulse response analysis (Figure 8) shows that a shock in the logarithm of market volatility results in a significant increase in Tether volume the following day, and this volume remains elevated over the following week. The effect of a shock in geopolitical risk is negligible in comparison, suggesting that in the global cryptocurrency market, investors use stablecoins as a means to balance portfolio risk during times of uncertainty and that on a macroscopic scale, changes in geopolitical risk are not a significant determinant of stablecoin use. A VAR model was then estimated using a lag length of 8 days with $ln(GPRActs)$ and $ln(GPRThreats)$ as additional endogenous variables in place of $ln(GPR)$ (Figure 23). GPR Threats were found to Granger cause Tether volume, while GPR Acts did not have the same predictive power (Figure 24).

This finding contradicts my initial hypothesis and suggests that investors are more reactive to the possibility of geopolitical events and economic sanctions than realizing these events themselves. One rationale for this behavior is that investors are loss averse and increase

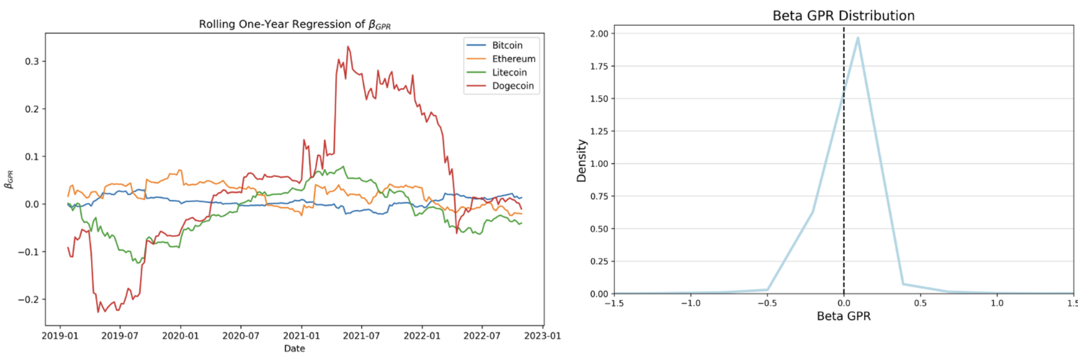


Fig. 9: Persistence and Distribution of B^{GPR}
Note. Large market-cap cryptocurrencies, such as Bitcoin and Ethereum, have geopolitical betas that remain approximately constant over time. In contrast, Dogecoin, which has a comparatively less well-established ecosystem, exhibits high volatility in its geopolitical beta over time. For the majority of cryptocurrencies, geopolitical betas are close to zero, with small, positive betas.

Long Low B^{GPR} Decile and Short High B^{GPR} Decile

	(1)	(2)	(3)	(4)
VARIABLES	Excess Returns	Excess Returns	Excess Returns	Excess Returns
Excess Market Returns	0.0728	0.668**	0.676**	0.681**
	(0.212)	(0.325)	(0.330)	(0.337)
SMB		1.353	1.375	1.386
		(0.990)	(0.869)	(0.866)
HML			-0.0208	-0.0264
			(0.296)	(0.307)
WML				0.00432
				(0.0120)
Constant	-0.00924	0.00138	0.00142	0.000133
	(0.0217)	(0.0137)	(0.0136)	(0.0136)
R-squared	0.001	0.313	0.313	0.313

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fig. 10

their holdings of stablecoins to mitigate potential losses when geopolitical uncertainty is high. Investors need more information about the future course of geopolitical events, leading to greater uncertainty and a tendency to overestimate the possibility of adverse outcomes. When the realization of these events does not have as significant of an impact as anticipated, and investors realize that the broader markets are unlikely to be affected, this increased stablecoin volume is not sustained.

C. The Russo-Ukrainian War

For the next part of my analysis, I consider the effects of exchange rate fluctuations and interest in the Russo-Ukrainian War on currency-denominated stablecoin volume in impacted regions. For Russia, I estimated a VAR model with ruble-denominated Tether volume, the logarithm of the USD-RUB exchange range, and the logarithm of the Google Trends index tracking interest in the war among individuals located in Russia as

Long High $\sigma(B^{GPR})$ Decile and Short Low $\sigma(B^{GPR})$ Decile

	(1)	(2)	(3)	(4)
VARIABLES	Excess Returns	Excess Returns	Excess Returns	Excess Returns
Excess Market Returns	0.381*	-0.389	-0.418	-0.415
	(0.211)	(0.372)	(0.384)	(0.392)
SMB		-1.658	-1.731	-1.725
		(1.259)	(1.084)	(1.078)
HML			0.0686	0.0650
			(0.275)	(0.290)
WML				0.00257
				(0.0175)
Constant	0.0547**	0.0377***	0.0377***	0.0369***
	(0.0274)	(0.0109)	(0.0110)	(0.0123)
R-squared	0.011	0.394	0.394	0.394

Fig. 11

Equation	Excluded	chi2	df	Prob>Chi2
ln(volume)	ln(GPR)	51.631	15	0.000
ln(volume)	ln(CVI)	321.080	15	0.000
ln(volume)	ALL	375.770	30	0.000
ln(GPR)	ln(volume)	54.471	15	0.000
ln(GPR)	ln(CVI)	22.315	15	0.100
ln(GPR)	ALL	78.257	30	0.000
ln(CVI)	ln(volume)	26.912	15	0.029
ln(CVI)	ln(GPR)	26.464	15	0.033
ln(CVI)	ALL	65.141	30	0.000

Fig. 12: Granger Causality Wald Tests
Note. There is bilateral Granger causality between $ln(TetherVolume)$ and $ln(GPR)$ at the 1% level and bilateral Granger causality between $ln(TetherVolume)$ and $ln(CVI)$ at the 5% level.

$$\ln(\text{Vol}_t) = \alpha_1 + \sum_{i=1}^{15} \beta_{1i} \ln(\text{Vol}_{t-i}) + \sum_{i=1}^{15} \gamma_{1i} \ln(\text{GPR}_{t-i}) + \sum_{i=1}^{15} \theta_{1i} \ln(\text{CVI}_{t-i}) + \epsilon_1$$

$$\ln(\text{GPR}_t) = \alpha_2 + \sum_{i=1}^{15} \beta_{2i} \ln(\text{Vol}_{t-i}) + \sum_{i=1}^{15} \gamma_{2i} \ln(\text{GPR}_{t-i}) + \sum_{i=1}^{15} \theta_{2i} \ln(\text{CVI}_{t-i}) + \epsilon_2$$

$$\ln(\text{CVI}_t) = \alpha_3 + \sum_{i=1}^{15} \beta_{3i} \ln(\text{Vol}_{t-i}) + \sum_{i=1}^{15} \gamma_{3i} \ln(\text{GPR}_{t-i}) + \sum_{i=1}^{15} \theta_{3i} \ln(\text{CVI}_{t-i}) + \epsilon_3$$

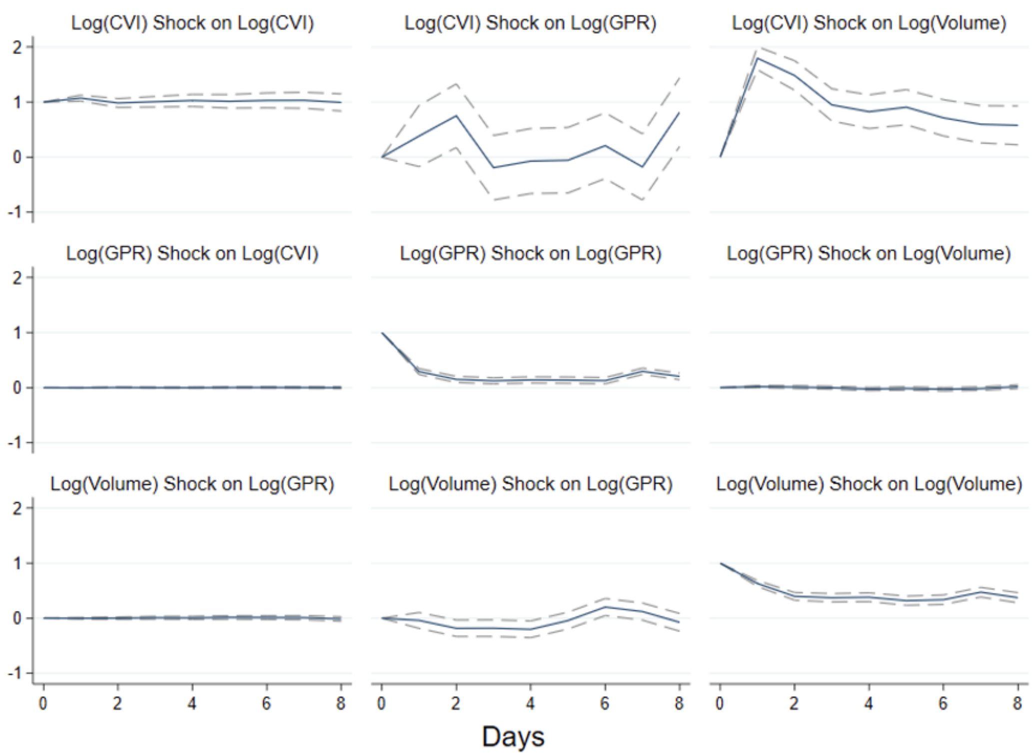


Fig. 13: Impulse Response Functions
Note. This figure shows the impulse responses to a one-unit shock in the impulse variable. The blue line denotes the impulse response with the 95% confidence interval captured by dashed lines. The vertical axes display the unit change response, corresponding to a multiplicative change in the untransformed response variable. A shock in market volatility (CVI) led to a marked increase in Tether volume the following day, with trading volume remaining elevated the following week. In contrast, a shock in geopolitical risk has a negligible effect on Tether volume. A shock to market volatility positively affects market volatility in the subsequent week, suggesting that volatility is persistent over time.

endogenous variables. I used a lag length of 12 days (Figure 25). The war index and exchange rate Granger cause ruble-denominated Tether volume, and the war index does not have predictive power for estimating the exchange rate at the 10% level (Figure 26). While the estimated effect of the shock is not statistically different from zero at the 5% level, a shock to the USD-RUB exchange rate (a depreciation of the ruble relative to the dollar) results in a moderate increase in ruble-denominated Tether volume in the following days (Figure 9), which supports the initial hypothesis that individuals may turn to stablecoins as stores of value to protect against depreciation. A one-unit shock to the logarithm of the *RussiaWarIndex* results in a substantial increase in ruble-denominated Tether volume in the following weeks, with statistically significant increases in Tether transacted five days after the shock. Volume remains elevated two weeks later, suggesting that individuals in Russia may view stablecoins as a safe haven asset during times of geopolitical uncertainty. The increase in stablecoin volume could be driven by individuals seeking to protect their wealth from potential economic disruptions and capital controls that may be imposed during times of conflict. Following the invasion, leaders in the United States, the United Kingdom, Japan, and other countries endorsed disconnecting Russian financial firms from SWIFT, and broad economic sanctions were placed on Russia's central bank (Greene, 2022). It is plausible that these actions, which limit the ability of Russian individuals and businesses to access the global financial system and conduct cross-border transactions, resulted in increased demand for stablecoins to circumvent these restrictions. While this theory cannot be directly confirmed from the available data, future research should examine stablecoin transactions involving Russian wallets to more conclusively determine whether stablecoins were used to bypass economic sanctions.

VI. Results

In analyzing the relationship between the variables described and changes in presidential voting share, I ran four corresponding regressions on the complete dataset of county data available. For the data analysis, I used analytic weights by population size to neutralize the influence of disproportionately populated counties. As this paper concerns county-level effects, regressions were run with fixed effects to absorb the influence of broader state correlations. Further, robust regressions were used instead of Ordinary Least Squares (OLS) regressions. The OLS method is traditionally very sensitive to data outliers, and in the context of the often-volatile indicators in my county-level data robust regressions was the better technical choice. Finally, as the regressions evaluate twenty independent variables, I have additionally included adjusted R-Squared values. Adjusted R-Squared values decrease if independent variables do not possess explanatory power, so they prove useful to determining whether the variables examined are statistically significant.

A similar VAR model was estimated for Ukraine using a lag length of 10 days (Figure 27). However, unlike in Russia, the USD-UAH exchange rate did not remain floating after the onset of the war. In Ukraine, "the dollar-hryvnia exchange rate is a measure of inflation, purchasing power, and the overall health of the economy." (Gorodnichenko, 2022). To prevent widespread panic, the National Bank of Ukraine (NBU) fixed the exchange on the first day of full-scale war. While in the short-run, the public regained confidence in the Ukrainian economy, in July 2022, the NBU had to devalue the hryvnia by roughly 25% to relieve pressure on foreign exchange reserves, improve the competitiveness of Ukrainian exports, raise more fiscal revenue, and align the official and cash-based exchange rates (Gorodnichenko, 2022). As a result of this policy, the exchange rate remains fixed apart from slight fluctuations throughout the sample, with a significant increase during July, months after the onset of the war. This pattern in the data allows us to more precisely isolate increased Tether purchases resulting from concerns over currency devaluation from Tether purchases stemming from the war and its effects more broadly. The logarithm of the *RussiaWarIndex* and the log change in the exchange rate Granger cause hryvnia-denominated Tether volume at the 10% level (Figure 28). A shock to the log change of the USD-UAH exchange rate increased hryvnia-denominated Tether volume in the following days. In contrast, the increase following a shock in the war index was not statistically different from zero (Figure 10). This observation suggests that currency devaluation is a significant driver of Tether purchases in Ukraine and that individuals may turn to stablecoins as stores of value. Since Ukraine did not face the same economic sanctions as Russia, it is unlikely that individuals were using Tether to evade financial restrictions, which could explain the discrepancy in impulse responses between the two countries.

VII. Discussion

The results above contradict my initial hypothesis that cryptocurrencies with positive geopolitical betas have low future returns relative to cryptocurrencies with negative geopolitical betas, which challenges the theory outlined by Long et al. (2022). Their study used less recent cryptocurrency data and did not explicitly include defunct cryptocurrencies to prevent survivorship bias, which could explain the discrepancy in results. Lack of replicability and inconsistency of results across studies is a common challenge in the extant cryptocurrency literature, and future research should continue to examine this phenomenon. The rolling regressions resulted in geopolitical betas that evolve over time and point to an alternative explanation for the geopolitical risk premium in the cross-section of cryptocurrency returns. Cryptocurrencies with volatile geopolitical betas had

$$\text{Vol}_t = \alpha_1 + \sum_{i=1}^{12} \beta_{1i}\text{Vol}_{t-i} + \sum_{i=1}^{12} \gamma_{1i}\ln(\text{USD-RUB}_{t-i}) + \sum_{i=1}^{12} \theta_{1i}\ln(\text{War}_{t-i}) + \epsilon_1$$

$$\ln(\text{USD-RUB}_t) = \alpha_2 + \sum_{i=1}^{12} \beta_{2i}\text{Vol}_{t-i} + \sum_{i=1}^{12} \gamma_{2i}\ln(\text{USD-RUB}_{t-i}) + \sum_{i=1}^{12} \theta_{2i}\ln(\text{War}_{t-i}) + \epsilon_2$$

$$\ln(\text{War}_t) = \alpha_3 + \sum_{i=1}^{12} \beta_{3i}\text{Vol}_{t-i} + \sum_{i=1}^{12} \gamma_{3i}\ln(\text{USD-RUB}_{t-i}) + \sum_{i=1}^{12} \theta_{3i}\ln(\text{War}_{t-i}) + \epsilon_3$$

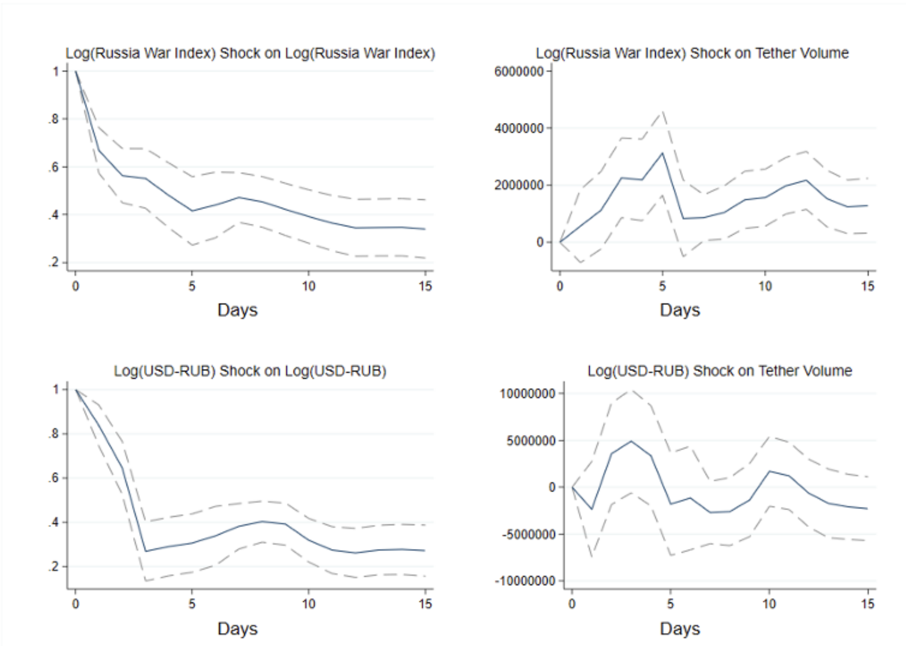


Fig. 14: Impulse Response Functions
Note. This figure shows the impulse responses to a one-unit shock in the impulse variable. The blue line denotes the impulse response with the 95% confidence interval captured by dashed lines. Vertical axes display unit changes in the response variable and are of different scales to retain the interpretability of the results. The estimated effect of a shock in the USD-RUB exchange rate on ruble-denominated Tether volume is not statistically different from zero. In contrast, there is a statistically significant positive effect on ruble-denominated Tether volume in response to a shock to the Russia War Index, with volume remaining elevated two weeks later.

higher returns than those with more stable betas, even when accounting for known cross-sectional factors in the cryptocurrency market. This finding suggests that investors pay a premium for assets that move predictably in response to geopolitical shocks, which aligns with risk diversification theory. Cryptocurrencies with volatile geopolitical betas have risks that cannot be diversified away, which may explain the premium observed in returns. This insight has broader implications for the cryptocurrency market and sheds light on investor behavior in response to geopolitical shocks.

When looking at the global cryptocurrency market, Tether volume is more responsive to changes in market volatility than changes in geopolitical risk. While this time series analysis does not establish causality, it suggests that investors use stablecoins to balance portfolio risk during market uncertainty. It was found that geopolitical threats have more predictive power than geopolitical acts in estimating future Tether volume, indicating that speculation about geopolitical events is a more significant driver of stablecoin volume in the global cryptocurrency market than the realization of these events. One potential limitation of this approach is that it considered cumulative Tether volume across

$$\mathrm{Vol}_t = \alpha_1 + \sum_{i=1}^{10} \beta_{1i} \mathrm{Vol}_{t-i} + \sum_{i=1}^{10} \gamma_{1i} (\ln(\mathrm{USD\text{-}UAH}_{t-i}) - \ln(\mathrm{USD\text{-}UAH}_{t-i-1})) + \sum_{i=1}^{10} \theta_{1i} \ln(\mathrm{War}_{t-i}) + \epsilon_1$$

$$\ln(\mathrm{USD\text{-}UAH}_t) - \ln(\mathrm{USD\text{-}UAH}_{t-1}) = \alpha_2 + \sum_{i=1}^{10} \beta_{2i} \mathrm{Vol}_{t-i} + \sum_{i=1}^{10} \gamma_{2i} (\ln(\mathrm{USD\text{-}UAH}_{t-i}) - \ln(\mathrm{USD\text{-}UAH}_{t-i-1})) + \sum_{i=1}^{10} \theta_{2i} \ln(\mathrm{War}_{t-i}) + \epsilon_2$$

$$\ln(\mathrm{War}_t) = \alpha_3 + \sum_{i=1}^{10} \beta_{3i} \mathrm{Vol}_{t-i} + \sum_{i=1}^{10} \gamma_{3i} (\ln(\mathrm{USD\text{-}UAH}_{t-i}) - \ln(\mathrm{USD\text{-}UAH}_{t-i-1})) + \sum_{i=1}^{10} \theta_{3i} \ln(\mathrm{War}_{t-i}) + \epsilon_3$$

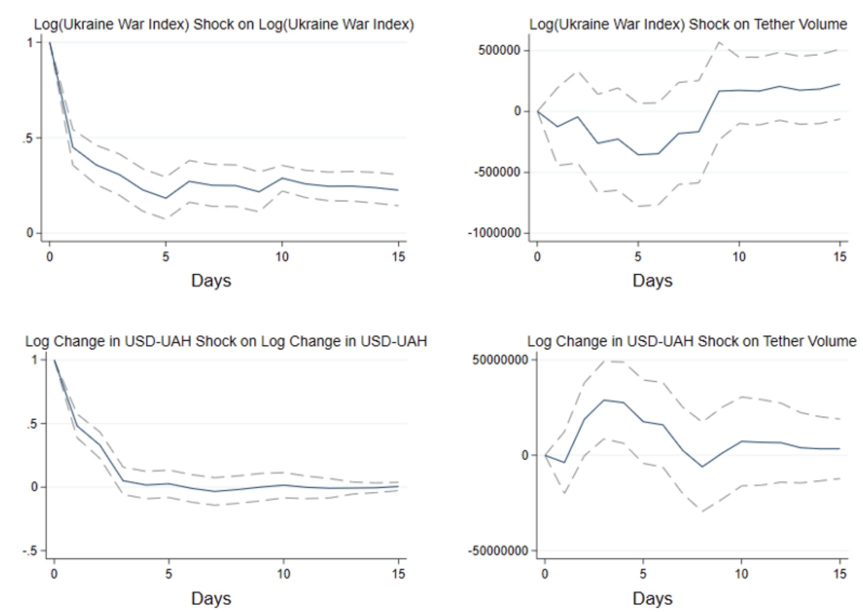


Fig. 15: Impulse Response Functions
Note. This figure shows the impulse responses to a one-unit shock in the impulse variable. The blue line denotes the impulse response with the 95% confidence interval captured by dashed lines. Vertical axes display unit changes in the response variable and are of different scales to retain the interpretability of the results. A shock in the log change of the USD-UAH exchange rate has a statistically significant positive effect on hryvnia-denominated Tether volume three days later. In contrast, the estimated effect of a shock in the *UkraineWarIndex* on hryvnia-denominated Tether volume is not statistically different from zero.

exchanges and did not differentiate between individuals purchasing Tether with traditional money versus other cryptocurrencies. Future research should explore these differences to gain a more nuanced understanding of the factors driving stablecoin volume and their use as stores of value and mediums of exchange.

When examining the Russo-Ukrainian War directly, currency-denominated stablecoin volume in Russia and Ukraine responded differently to changes in the exchange rate and interest in the war. Changes in the Google Trends war index in Russia resulted in a more substantial increase in future ruble-denominated Tether volume than shocks to the exchange rate. This finding could be explained by the economic sanctions levied on Russia due to the war and increased stablecoin use to circumvent these restrictions. While this theory cannot be tested empirically with the available data, if stablecoins were used to bypass sanctions, this would have significant implications for the global financial system and call into question the effectiveness of economic regulation. Future studies should examine stablecoin transactions involving Russian wallets to gain further insight into whether these cryptocurrencies were used to circumvent sanctions, and policymakers need to consider regulation to maintain the integrity of the global financial system.

Conversely, in Ukraine, stablecoins were used primarily as stores of value rather than means of exchange, with shocks to the exchange rate having a statistically significant effect on hryvnia-denominated Tether volume. This discovery suggests individuals in Ukraine used Tether to protect against the devaluation of their local currency, which further bolsters my hypothesis that stablecoins are used in place of traditional safe-haven assets when access is restricted, such as in the case of war. However, it is essential to note that these findings are specific to Tether and may not necessarily apply to other stablecoins. Another limitation of this study is that Google Trends data was used as a proxy for interest in the war and may not necessarily reflect actual sentiment. Additional measures should be considered in future studies, such as mentions of the war in international news.

VIII. Conclusion

Motivated by the increased prevalence of cryptocurrencies and their effect on global financial markets, this paper furthers the literature on the relationship between cryptocurrencies and geopolitical risk by accounting for changing sensitivities over time and examining stablecoin adoption following geopolitical events to shed light on the evolving role of cryptocurrencies in the global financial system. While earlier studies have attempted to identify a geopolitical risk premium, no study has examined the evolution of these geopolitical risk sensitivities over time and how this results in abnormal returns. Moreover, this study identifies the role of stablecoins as potential stores of value and mediums of exchange during geopolitical turmoil, which has meaningful implications for future economic policy and raises concerns about the lack of regulatory oversight in the cryptocurrency space. Stablecoins, in particular, have come under increased scrutiny from regulators due to their potential for illicit activities such as money laundering and terrorism financing (Wilson, 2021). While stablecoins act as substitutes for traditional safe-haven assets, directly benefiting those affected by conflict, regulators still need to monitor cryptocurrencies closely when making policy decisions given their potential to help bypass economic sanctions.

References

[1] All Cryptocurrencies. (2022). Retrieved November 9, 2022, from https://coinmarketcap.com/all/views/all/

[2] Aysan, A. F., Demir, E., Gozgor, G., & Lau, C. K. M. (2019). Effects of the geopolitical risks on Bitcoin returns and volatility. Research in International Business and Finance, 47, 511–518. https://doi.org/10.1016/j.ribaf.2018.09.011

[3] Banz, R. W. (1981). The relationship between return and market value of common stocks. Journal of Financial Economics, 9(1), 3–18. https://doi.org/10.1016/0304-405X(81)90018-0

[4] Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. The Journal of Finance, 32(3), 663–682. https://doi.org/10.1111/j.1540-6261.1977.tb01979.x

[5] Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. Journal of Banking & Finance, 34(8), 1886–1898. https://doi.org/10.1016/j.jbankfin.2009.12.008

[6] Binance Data Collection. (2022, December 8). https://data.binance.vision/

[7] Board of Governors of the Federal Reserve System (US). (2001, August 3). 4-Week Treasury Bill Secondary Market Rate, Discount Basis. FRED, Federal Reserve Bank of St. Louis; FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/WTB4WK

[8] Brauneis, A., & Mestel, R. (2018). Price discovery of cryptocurrencies: Bitcoin and beyond. Economics Letters, 165, 58–61. https://doi.org/10.1016/j.econlet.2018.02.001

[9] Caldara, D., & Iacoviello, M. (2018). Measuring Geopolitical Risk. International Finance Discussion Paper, 2018(1222), 1–66. https://doi.org/10.17016/IFDP.2018.1222

[10] Caldara, D., & Iacoviello, M. (2022). Measuring Geopolitical Risk. American Economic Review, 112(4), 1194–1225. https://doi.org/10.1257/aer.20191823

[11] Campbell, J. Y., & Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. The Review of Financial Studies, 1(3), 195–228.

[12] Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. The Journal of Finance, 52(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

[13] Choi, S., & Shin, J. (2022). Bitcoin: An inflation hedge but not a safe haven. Finance Research Letters, 46, 102379. https://doi.org/10.1016/j.frl.2021.102379

[14] Chow, G. C., & Lin, A. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. The Review of Economics and Statistics, 53(4), 372–375. https://doi.org/10.2307/1928739

[15] Country-Specific Geopolitical Risk Index. (n.d.). Retrieved November 6, 2022, from https://www.matteoiacoviello.com/gpr_country.htm

[16] Crypto Volatility Index Chart (CVI). (2022, December 8). Investing.Com. https://www.investing.com/indices/crypto-volatility-index-chart

[17] Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics, 182(1), 119–134. https://doi.org/10.1016/j.jeconom.2014.04.012

[18] Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. 22.

[19] Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

[20] Fama, M., Fumagalli, A., & Lucarelli, S. (2019). Cryptocurrencies, Monetary Policy, and New Forms of Monetary Sovereignty. International Journal of Political Economy, 48(2), 174–194. https://doi.org/10.1080/08911916.2019.1624318

[21] Fratrič, P., Sileno, G., Klous, S., & van Engers, T. (2022). Manipulation of the Bitcoin market: An agent-based study. Financial Innovation, 8(1), 60. https://doi.org/10.1186/s40854-022-00364-3

[22] Godbole, O. (2022, February 28). Ruble-Denominated Bitcoin Volume Surges to 9-Month High. https://www.coindesk.com/markets/2022/02/28/ruble-denominated-bitcoin-volumes-surges-to-9-month-highs/

[23] Google Trends. (n.d.). Google Trends. Retrieved April 19, 2023, from https://trends.google.com/trends/explore

[24] Gorodnichenko, Y. (2022, November 2). No fear of floating for the hryvnia. https://voxukraine.org/en/no-fear-of-floating-for-the-hryvnia
[25] Greene, R. (2022, March 4). How Sanctions on Russia Will Alter Global Payments Flows. Carnegie Endowment for International Peace. https://carnegieendowment.org/2022/03/04/how-sanctions-on-russia-will-alter-global-payments-flows-pub-86575
[26] Han, W., Newton, D., Platanakis, E., Sutcliffe, C., & Ye, X. (2021). Cryptocurrency Factor Portfolios: Performance, Decomposition and Pricing Models.
[27] Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance, 48(1), 65–91. https://doi.org/10.2307/2328882
[28] Kharif, O. (2022, March 11). Stablecoin Tether Disregards Ukraine's Plea to Halt Russian Use. Bloomberg.Com. https://www.bloomberg.com/news/articles/2022-03-11/stablecoin-tether-disregards-ukraine-s-plea-to-halt-russian-use
[29] Kim, A., Trimborn, S., & Härdle, W. K. (2021). VCRIX — A volatility index for crypto-currencies. International Review of Financial Analysis, 78, 101915. https://doi.org/10.1016/j.irfa.2021.101915
[30] Leonhard, R. (2019). Decentralized Finance on the Ethereum Blockchain (SSRN Scholarly Paper No. 3359732). https://doi.org/10.2139/ssrn.3359732
[31] Liebi, L. J. (2022). Is there a value premium in cryptoasset markets? Economic Modelling, 109, 105777. https://doi.org/10.1016/j.econmod.2022.105777
[32] Liu, W., Liang, X., & Cui, G. (2020). Common risk factors in the returns on cryptocurrencies. Economic Modelling, 86, 299–305. https://doi.org/10.1016/j.econmod.2019.09.035
[33] Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common Risk Factors in Cryptocurrency. The Journal of Finance, 77(2), 1133–1177. https://doi.org/10.1111/jofi.13119
[34] Long, H., Demir, E., Będowska-Sójka, B., Zaremba, A., & Shahzad, S. J. H. (2022). Is geopolitical risk priced in the cross-section of cryptocurrency returns? Finance Research Letters, 49, 103131. https://doi.org/10.1016/j.frl.2022.103131
[35] Ranaldo, A., & Söderlind, P. (2010). Safe Haven Currencies*. Review of Finance, 14(3), 385–407. https://doi.org/10.1093/rof/rfq007
[36] Reuters. (2022, July 21). Ukraine devalues hryvnia currency by 25% against U.S. dollar. Reuters. https://www.reuters.com/markets/rates-bonds/ukraines-central-bank-devalues-hryvnia-by-25-against-us-dollar-2022-07-21/
[37] Schwartz, L. (2022, September 14). Global crypto adoption dominated by emerging markets, Chainalysis finds. Fortune. https://fortune.com/crypto/2022/09/14/global-crypto-adoption-dominated-emerging-markets/
[38] Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. The Journal of Finance, 19(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x
[39] Sigaki, H. Y. D., Perc, M., & Ribeiro, H. V. (2019). Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market. Scientific Reports, 9(1), Article 1. https://doi.org/10.1038/s41598-018-37773-3
[40] Sun, J., Zhou, Y., & Lin, J. (2019). Using machine learning for cryptocurrency trading. 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS), 647–652. https://doi.org/10.1109/ICPHYS.2019.8780358
[41] There's a crypto safe haven amid the Russian invasion of Ukraine. (2022). Fortune. Retrieved December 4, 2022, from https://fortune.com/2022/02/24/stable-coin-crypto-holding-value-amid-putin-invasion-ukraine/
[42] USD/RUB. (2023). Yahoo! Finance. Retrieved April 1, 2023, from https://finance.yahoo.com/quote/RUB=X/
[43] USD/UAH. (2023). Yahoo! Finance. Retrieved April 1, 2023, from https://finance.yahoo.com/quote/UAH=X/
[44] Wilson, T. (2021, October 28). Mass use of stablecoins could boost illicit finance, money-laundering watchdog says. Reuters. https://www.reuters.com/article/fintech-crypto-idUSL4N2RN2QI
[45] Youssef, M., Mokni, K., & Ajmi, A. N. (2021). Dynamic connectedness between stock markets in the presence of the COVID-19 pandemic: Does economic policy uncertainty matter? Financial Innovation, 7(1), 13. https://doi.org/10.1186/s40854-021-00227-3

Appendix

Descriptive Statistics

Variables	Obs.	Mean	Std. Dev.	Min	Max	1st Percentile	99th Percentile
4-Week Treasury Bill Rate	253	1.099	.99	0	3.58	.01	3.25

Fig. 16: Summary of One-Month Treasury Bill Rate (Collected Weekly)
Note. Data was aggregated from FRED (2022).

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
GPR Index	1765	100.379	60.44	3.57	539.58
GPR Acts	1765	69.746	54.256	0	393.73
GPR Threats	1765	126.545	84.9	0	811.53

Fig. 17: Summary of GPR Index
Note. Data was aggregated from Caldara & Iacoviello (2022).

Descriptive Statistics

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
USDT-RUB Volume in Number of Tether Transacted	*666*	*4,538,882.30*	*4,164,478.40*	*359,605*	*37,313,669*
USDT-UAH Volume in Number of Tether Transacted	*666*	*2,904,145.40*	*2,310,790*	*297,989.50*	*14,673,679*

Fig. 18: USDT-RUB Volume Over Time
Note. Data was aggregated from Binance (2023)

Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
War in Russia	19.364	13.798	8	100
Ukraine in Russia	13.227	16.077	3	100
War in Ukraine	22.061	18.725	3	100
Russia in Ukraine	22.621	17.884	8	100
Russia War Index	16.295	14.743	5.5	100
Ukraine War Index	22.341	17.607	6	100

Fig. 19: Google Trends Data
Note. Data was aggregated from Google Trends (n.d.).

Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
USD-RUB	70.89	14.39	52.47	138.97
USD-UAH	31.15	4.17	25.79	36.86

Fig. 20: Exchange Rate Data
Note. Exchange rate data from 10/03/2021 to 01/01/2023 (Yahoo! Finance, 2023).

VARIABLES	$ln(Mean\ Weekly\ Market\ Cap)$
$ln(\sigma(B^{GPR}))$	-0.408***
	(0.00997)
Constant	15.73***
	(0.0335)
Observations	83,776
R-squared	0.023

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Fig. 21: Correlation Between Standard Deviation of Rolling Geopolitical Betas Over Past Six Months and Market Cap

VARIABLES	ln(Volume)	ln(GPR)	ln(CVI)
L.ln(Volume)	0.631***	-0.0397	-0.00220
	(0.0277)	(0.0734)	(0.00719)
L2.ln(Volume)	0.00338	-0.145*	0.00416
	(0.0325)	(0.0863)	(0.00846)
L3.ln(Volume)	0.118***	-0.0168	0.00606
	(0.0325)	(0.0862)	(0.00846)
L4.ln(Volume)	0.0577*	-0.0561	-0.00739
	(0.0327)	(0.0866)	(0.00850)
L5.ln(Volume)	-0.00339	0.150*	0.0110
	(0.0327)	(0.0868)	(0.00851)
L6.ln(Volume)	0.0466	0.257***	-0.0107
	(0.0327)	(0.0867)	(0.00851)
L7.ln(Volume)	0.157***	-0.0398	0.00115
	(0.0327)	(0.0867)	(0.00850)
L8.ln(Volume)	-0.109***	-0.117	-0.0243***
	(0.0328)	(0.0871)	(0.00854)
L9.ln(Volume)	0.00279	-0.0317	0.0155*
	(0.0330)	(0.0874)	(0.00858)
L10.ln(Volume)	-0.00472	-0.00934	0.00477
	(0.0330)	(0.0874)	(0.00858)
L11.ln(Volume)	0.00411	-0.157*	0.000184
	(0.0330)	(0.0875)	(0.00858)
L12.ln(Volume)	0.00688	0.0144	0.0131
	(0.0329)	(0.0872)	(0.00856)
L13.ln(Volume)	0.0243	0.296***	-0.00159
	(0.0327)	(0.0867)	(0.00851)
L14.ln(Volume)	0.123***	-0.0867	-0.0138*
	(0.0311)	(0.0824)	(0.00808)
L15.ln(Volume)	-0.0745***	-0.00639	0.00854
	(0.0259)	(0.0687)	(0.00674)
L.ln(GPR)	0.0193*	0.293***	-0.00136
	(0.0105)	(0.0278)	(0.00272)
L2.ln(GPR)	-0.00401	0.0656**	0.00550*
	(0.0108)	(0.0286)	(0.00281)

L3.ln(GPR)	-0.0183*	0.0635**	-0.00490*
	(0.0108)	(0.0285)	(0.00280)
L4.ln(GPR)	-0.0185*	0.0749***	0.000878
	(0.0108)	(0.0286)	(0.00281)
L5.ln(GPR)	0.00575	0.0595**	0.00298
	(0.0108)	(0.0287)	(0.00281)
L6.ln(GPR)	-0.0248**	0.0329	-0.000468
	(0.0108)	(0.0287)	(0.00282)
L7.ln(GPR)	0.0109	0.201***	-0.00102
	(0.0108)	(0.0287)	(0.00281)
L8.ln(GPR)	0.0305***	0.00988	-0.00224
	(0.0109)	(0.0290)	(0.00284)
L9.ln(GPR)	-0.0156	-0.0762***	-1.89e-05
	(0.0107)	(0.0285)	(0.00279)
L10.ln(GPR)	0.00374	-0.0104	-0.00555**
	(0.0107)	(0.0285)	(0.00279)
L11.ln(GPR)	0.00272	0.00787	0.000691
	(0.0107)	(0.0284)	(0.00278)
L12.ln(GPR)	-0.0165	-0.0486*	0.00680**
	(0.0107)	(0.0283)	(0.00278)
L13.ln(GPR)	-0.0127	0.0748***	-0.00317
	(0.0107)	(0.0283)	(0.00278)
L14.ln(GPR)	0.0175	0.156***	-0.00103
	(0.0107)	(0.0283)	(0.00278)
L15.ln(GPR)	0.0205**	-0.0243	-0.00246
	(0.0104)	(0.0275)	(0.00269)
L.ln(CVI)	1.796***	0.384	1.072***
	(0.107)	(0.283)	(0.0278)
L2.ln(CVI)	-1.582***	0.299	-0.162***
	(0.163)	(0.433)	(0.0425)
L3.ln(CVI)	-0.0735	-0.814*	0.121***
	(0.170)	(0.450)	(0.0441)
L4.ln(CVI)	-0.149	0.388	-0.0375
	(0.170)	(0.450)	(0.0442)
L5.ln(CVI)	0.102	-0.114	0.00130
	(0.170)	(0.451)	(0.0442)
L6.ln(CVI)	-0.123	0.0109	0.00526
	(0.170)	(0.450)	(0.0441)

L7.ln(CVI)	-0.0793	-0.881**	-0.000533
	(0.169)	(0.449)	(0.0440)
L8.ln(CVI)	-0.178	1.280***	-0.0347
	(0.169)	(0.448)	(0.0440)
L9.ln(CVI)	0.520***	-0.448	0.0485
	(0.170)	(0.450)	(0.0441)
L10.ln(CVI)	-0.395**	-0.0459	-0.0780*
	(0.170)	(0.451)	(0.0442)
L11.ln(CVI)	0.132	0.0312	0.00553
	(0.170)	(0.452)	(0.0443)
L12.ln(CVI)	-0.0318	-0.143	0.00442
	(0.170)	(0.451)	(0.0442)
L13.ln(CVI)	-0.0393	0.0548	0.0104
	(0.169)	(0.449)	(0.0440)
L14.ln(CVI)	-0.105	-0.560	-0.0151
	(0.168)	(0.446)	(0.0437)
L15.ln(CVI)	0.192*	0.508*	0.0274
	(0.115)	(0.305)	(0.0299)
Constant	0.492**	0.473	0.0561
	(0.198)	(0.525)	(0.0515)

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fig. 22: Model Coefficients for VAR Model with Endogenous Variables $ln(TetherVolume)$, $ln(GPR)$, and $ln(CVI)$ Using a Lag Length of 15 Days

VARIABLES	ln(Volume)	ln(Acts)	ln(Threats)	ln(CVI)
L.ln(Volume)	0.646***	0.192	-0.157*	-0.00307
	(0.0323)	(0.117)	(0.0931)	(0.00807)
L2.ln(Volume)	-0.0289	-0.244*	-0.124	0.00782
	(0.0377)	(0.136)	(0.109)	(0.00944)
L3.ln(Volume)	0.157***	0.174	-0.0134	0.00949
	(0.0378)	(0.137)	(0.109)	(0.00946)
L4.ln(Volume)	0.0551	-0.197	0.0252	0.00192
	(0.0382)	(0.138)	(0.110)	(0.00956)
L5.ln(Volume)	-0.00807	0.213	0.0471	0.01000
	(0.0382)	(0.138)	(0.110)	(0.00956)
L6.ln(Volume)	0.0903**	0.242*	0.451***	-0.00934
	(0.0381)	(0.138)	(0.110)	(0.00954)
L7.ln(Volume)	0.192***	-0.0851	-0.0196	0.000248
	(0.0369)	(0.133)	(0.107)	(0.00924)
L8.ln(Volume)	-0.117***	-0.266**	-0.212**	-0.0129*
	(0.0307)	(0.111)	(0.0885)	(0.00767)
L.ln(Acts)	0.0177*	0.183***	0.0518*	-0.000939
	(0.00942)	(0.0340)	(0.0272)	(0.00236)
L2.ln(Acts)	-0.00672	0.0946***	0.0217	-0.00110
	(0.00955)	(0.0345)	(0.0275)	(0.00239)
L3.ln(Acts)	-0.00607	0.0984***	0.0561**	-0.00300
	(0.00965)	(0.0349)	(0.0278)	(0.00241)
L4.ln(Acts)	-0.0169*	0.0214	-0.00229	0.00209
	(0.00965)	(0.0349)	(0.0278)	(0.00241)
L5.ln(Acts)	0.000169	0.110***	0.00683	-0.00107
	(0.00950)	(0.0343)	(0.0274)	(0.00238)

L6.ln(Acts)	-0.0161*	0.0717**	-0.0140	0.00119
	(0.00953)	(0.0344)	(0.0275)	(0.00238)
L7.ln(Acts)	0.0116	0.0713**	-0.00207	0.00259
	(0.00945)	(0.0341)	(0.0273)	(0.00236)
L8.ln(Acts)	0.0145	0.0335	-0.00197	0.000114
	(0.00933)	(0.0337)	(0.0269)	(0.00233)
L.ln(Threats)	0.0171	0.0515	0.294***	-0.000696
	(0.0116)	(0.0420)	(0.0335)	(0.00291)
L2.ln(Threats)	-0.0213*	0.00262	0.0507	0.00581*
	(0.0119)	(0.0429)	(0.0342)	(0.00297)
L3.ln(Threats)	-0.00357	0.0182	0.0524	-0.00585**
	(0.0119)	(0.0430)	(0.0344)	(0.00298)
L4.ln(Threats)	-0.0174	-0.0412	0.0383	0.00102
	(0.0119)	(0.0429)	(0.0342)	(0.00297)
L5.ln(Threats)	0.00287	0.0452	0.0315	0.00398
	(0.0119)	(0.0429)	(0.0343)	(0.00297)
L6.ln(Threats)	-0.0308***	0.0432	0.109***	0.000350
	(0.0119)	(0.0430)	(0.0343)	(0.00297)
L7.ln(Threats)	0.0127	0.150***	0.214***	-0.00310
	(0.0120)	(0.0432)	(0.0345)	(0.00299)
L8.ln(Threats)	0.0374***	-0.0587	-0.0305	-0.00342
	(0.0116)	(0.0418)	(0.0333)	(0.00289)
L.ln(CVI)	1.664***	0.990**	0.814**	1.079***
	(0.131)	(0.473)	(0.378)	(0.0327)
L2.ln(CVI)	-1.472***	-0.835	-0.263	-0.189***
	(0.200)	(0.724)	(0.578)	(0.0501)
L3.ln(CVI)	0.0155	-0.922	-0.418	0.162***
	(0.206)	(0.744)	(0.594)	(0.0515)
L4.ln(CVI)	-0.350*	-0.199	-0.00754	-0.0952*
	(0.208)	(0.750)	(0.599)	(0.0519)
L5.ln(CVI)	0.0595	1.438*	-0.588	0.00663
	(0.208)	(0.750)	(0.599)	(0.0519)
L6.ln(CVI)	0.0191	-0.119	0.951	0.0118
	(0.209)	(0.755)	(0.603)	(0.0523)
L7.ln(CVI)	-0.328	-1.236*	-1.455**	0.0364
	(0.207)	(0.750)	(0.599)	(0.0519)
L8.ln(CVI)	0.369***	0.680	0.961**	-0.0449
	(0.140)	(0.505)	(0.403)	(0.0349)
Constant	0.455**	0.521	0.739	0.0550
	(0.228)	(0.823)	(0.657)	(0.0570)

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fig. 23: Model Coefficients for VAR Model with Endogenous Variables $ln(Volume)$, $ln(GPRActs)$, $ln(GPRThreats)$, and $ln(CVI)$ Using a Lag Length of 8 Days

Equation	Excluded	chi2	df	Prob>Chi2
ln(Volume)	ln(Acts)	13.093	8	0.109
ln(Volume)	ln(Threats)	28.091	8	0.000
ln(Volume)	ln(CVI)	186.750	8	0.000
ln(Volume)	ALL	254.100	24	0.000
ln(Acts)	ln(Volume)	19.119	8	0.014
ln(Acts)	ln(Threats)	25.742	8	0.001
ln(Acts)	ln(CVI)	19.173	8	0.014
ln(Acts)	ALL	76.563	24	0.000
ln(Threats)	ln(Volume)	41.144	8	0.000
ln(Threats)	ln(Acts)	12.939	8	0.114
ln(Threats)	ln(CVI)	15.103	8	0.057
ln(Threats)	ALL	65.503	24	0.000
ln(CVI)	ln(Volume)	12.340	8	0.137
ln(CVI)	ln(Acts)	4.186	8	0.840
ln(CVI)	ln(Threats)	10.918	8	0.206
ln(CVI)	ALL	32.535	24	0.114

Fig. 24: Granger Causality Wald Tests
Note. $ln(GPRThreats)$ Granger causes $ln(TetherVolume)$ at the 1% level, but we cannot reject the null hypothesis that $ln(GPRActs)$ does not Granger cause $ln(TetherVolume)$.

VARIABLES	ln(USD-RUB)	Russia Volume	ln(War Index)
L.ln(USD-RUB)	0.834***	-5.038e+06**	-0.000207
	(0.0482)	(2.533e+06)	(0.191)
L2.ln(USD-RUB)	-0.0328	6.992e+06**	-0.209
	(0.0632)	(3.325e+06)	(0.251)
L3.ln(USD-RUB)	-0.248***	1.880e+06	0.178
	(0.0633)	(3.325e+06)	(0.251)
L4.ln(USD-RUB)	0.293***	-5.528e+06*	-0.120
	(0.0638)	(3.353e+06)	(0.253)
L5.ln(USD-RUB)	-0.0184	212,537	0.158
	(0.0651)	(3.423e+06)	(0.259)
L6.ln(USD-RUB)	-0.0453	3.111e+06	0.0192
	(0.0644)	(3.386e+06)	(0.256)
L7.ln(USD-RUB)	0.180***	-7.398e+06**	0.126

	(0.0642)	(3.376e+06)	(0.255)
L8.ln(USD-RUB)	-0.0611	5.618e+06	-0.315
	(0.0651)	(3.423e+06)	(0.259)
L9.ln(USD-RUB)	-0.00262	-1.738e+06	0.477*
	(0.0630)	(3.314e+06)	(0.250)
L10.ln(USD-RUB)	0.0639	-3.855e+06	-0.206
	(0.0616)	(3.239e+06)	(0.245)
L11.ln(USD-RUB)	0.0193	1.039e+07***	-0.180
	(0.0620)	(3.258e+06)	(0.246)
L12.ln(USD-RUB)	-0.0642	-4.582e+06*	0.172
	(0.0470)	(2.470e+06)	(0.187)
L.Russia Volume	1.73e-09*	0.400***	-5.45e-09
	(9.01e-10)	(0.0474)	(3.58e-09)
L2.Russia Volume	7.71e-10	0.0695	5.80e-09
	(9.62e-10)	(0.0506)	(3.82e-09)
L3.Russia Volume	2.22e-09**	-0.00456	6.43e-09*
	(9.57e-10)	(0.0503)	(3.80e-09)
L4.Russia Volume	-9.21e-10	0.223***	-3.83e-09
	(9.59e-10)	(0.0504)	(3.81e-09)
L5.Russia Volume	3.66e-10	-0.0491	-3.43e-09
	(9.80e-10)	(0.0515)	(3.89e-09)
L6.Russia Volume	1.15e-09	0.0699	-1.97e-09
	(9.28e-10)	(0.0488)	(3.69e-09)
L7.Russia Volume	-1.13e-09	0.360***	-9.11e-09**
	(9.27e-10)	(0.0487)	(3.68e-09)
L8.Russia Volume	-1.76e-09*	-0.0725	2.97e-09
	(9.91e-10)	(0.0521)	(3.93e-09)
L9.Russia Volume	1.13e-09	-0.0870*	5.17e-09
	(9.56e-10)	(0.0503)	(3.80e-09)
L10.Russia Volume	-8.35e-10	0.0266	-2.78e-09
	(9.62e-10)	(0.0506)	(3.82e-09)
L11.Russia Volume	7.40e-10	0.00769	7.09e-09*

	(9.56e-10)	(0.0503)	(3.80e-09)
L12.Russia Volume	1.23e-09	-0.133***	-3.69e-09
	(8.92e-10)	(0.0469)	(3.54e-09)
L.Russia War Index	0.0106	279,218	0.708***
	(0.0123)	(649,221)	(0.0490)
L2.Russia War Index	-0.00364	347,179	0.0758
	(0.0145)	(764,797)	(0.0578)
L3.Russia War Index	-0.0204	1.138e+06	0.130**
	(0.0144)	(758,651)	(0.0573)
L4.Russia War Index	0.0187	9,449	-0.0163
	(0.0145)	(760,938)	(0.0575)
L5.Russia War Index	-0.00244	1.141e+06	-0.0304
	(0.0145)	(762,098)	(0.0576)
L6.Russia War Index	0.00414	-1.890e+06**	0.0719
	(0.0146)	(766,255)	(0.0579)
L7.Russia War Index	0.0212	490,449	-0.0456
	(0.0147)	(774,302)	(0.0585)
L8.Russia War Index	0.0120	196,899	0.152***
	(0.0148)	(776,442)	(0.0586)
L9.Russia War Index	-0.0438***	2.273e+06***	-0.0575
	(0.0148)	(779,853)	(0.0589)
L10.Russia War Index	0.00285	-1.821e+06**	0.130**
	(0.0150)	(789,793)	(0.0597)
L11.Russia War Index	0.00594	-347,607	-0.216***
	(0.0145)	(762,783)	(0.0576)
L12.Russia War Index	-0.000446	-1.332e+06**	0.0447
	(0.0114)	(600,426)	(0.0454)
Constant	0.319***	-654,334	-0.280
	(0.0810)	(4.260e+06)	(0.322)

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fig. 25: Model Coefficients for VAR Model with Endogenous Variables Ruble-Denominated Tether Volume, $ln(USD-RUB)$, and $ln(RussiaWarIndex)$ Using a Lag Length of 12 Days

Equation	Excluded	chi2	df	Prob>Chi2
ln(USD-RUB)	Russia Volume	36.757	12	0.000
ln(USD-RUB)	ln(War Index)	17.257	12	0.140
ln(USD-RUB)	ALL	62.846	24	0.000
Russia Volume	ln(USD-RUB)	21.817	12	0.040
Russia Volume	ln(War Index)	45.760	12	0.000
Russia Volume	ALL	70.858	24	0.000
ln(War Index)	ln(USD-RUB)	9.090	12	0.695
ln(War Index)	Russia Volume	24.279	12	0.019
ln(War Index)	ALL	34.710	24	0.073

Fig. 26: Granger Causality Wald Tests
Note. The Russia War Index and USD-RUB exchange rate Granger cause ruble-denominated Tether volume at the 5% level of significance. The war index and exchange rate Granger cause ruble-denominated Tether volume and the war index does not have predictive power for estimating the exchange rate at the 10% level.

VARIABLES	Ukraine Volume	ln(War Index)	Log Change in USD-UAH
L.Ukraine Volume	0.626***	-5.98e-09	-1.70e-10
	(0.0474)	(1.38e-08)	(2.73e-10)
L2.Ukraine Volume	0.0336	7.98e-09	9.98e-11
	(0.0552)	(1.61e-08)	(3.19e-10)
L3.Ukraine Volume	0.0857	1.28e-08	2.75e-10
	(0.0550)	(1.61e-08)	(3.17e-10)
L4.Ukraine Volume	-0.0259	-8.86e-09	-9.85e-11
	(0.0546)	(1.60e-08)	(3.15e-10)
L5.Ukraine Volume	0.00928	3.11e-09	-3.42e-10
	(0.0543)	(1.59e-08)	(3.13e-10)
L6.Ukraine Volume	0.0985*	-5.93e-10	4.36e-10
	(0.0542)	(1.58e-08)	(3.12e-10)
L7.Ukraine Volume	0.126**	4.28e-10	-3.39e-10
	(0.0535)	(1.56e-08)	(3.08e-10)
L8.Ukraine Volume	-0.0372	-1.34e-08	-2.48e-10
	(0.0537)	(1.57e-08)	(3.10e-10)
L9.Ukraine Volume	-0.0522	1.33e-08	2.68e-10
	(0.0535)	(1.56e-08)	(3.08e-10)
L10.Ukraine Volume	0.0126	-1.09e-08	-0
	(0.0428)	(1.25e-08)	(2.47e-10)
L.ln(War Index)	-125,331	0.449***	0.000106
	(162,612)	(0.0475)	(0.000937)
L2.ln(War Index)	89,605	0.154***	-0.000411
	(178,304)	(0.0521)	(0.00103)
L3.ln(War Index)	-228,059	0.0786	-0.000638
	(179,877)	(0.0526)	(0.00104)
L4.ln(War Index)	60,471	0.00108	-0.00138
	(180,531)	(0.0528)	(0.00104)

L5.ln(War Index)	-134,742	0.0110	0.000156
	(179,561)	(0.0525)	(0.00104)
L6.ln(War Index)	69,439	0.125**	0.000393
	(179,693)	(0.0525)	(0.00104)
L7.ln(War Index)	193,770	0.0206	-0.000636
	(180,485)	(0.0527)	(0.00104)
L8.ln(War Index)	3,301	0.0279	0.00152
	(180,267)	(0.0527)	(0.00104)
L9.ln(War Index)	359,643**	-0.0160	0.000757
	(177,513)	(0.0519)	(0.00102)
L10.ln(War Index)	-40,468	0.106**	0.000137
	(164,430)	(0.0481)	(0.000948)
L.Log Change in USD-UAH	-3.764e+06	3.767	0.481***
	(8.271e+06)	(2.417)	(0.0477)
L2.Log Change in USD-UAH	2.347e+07**	-3.801	0.0958*
	(9.167e+06)	(2.679)	(0.0529)
L10.Log Change in USD-UAH	1.124e+06	1.766	0.0237
	(8.348e+06)	(2.440)	(0.0481)
Constant	-396,918*	0.127*	0.000750
	(233,845)	(0.0683)	(0.00135)

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fig. 27: Model Coefficients for VAR Model with Endogenous Variables Hryvnia-Denominated Tether Volume, $ln(USD-UAH)$ in First Differences, and $ln(RussiaWarIndex)$ Using a Lag Length of 10 Days

Equation	Excluded	chi2	Prob>Chi2
Log Change in USD-UAH	Ukraine Volume	0.098	0.754
Log Change in USD-UAH	ln(War Index)	8.316	0.598
Log Change in USD-UAH	ALL	8.417	0.676
Ukraine Volume	Log Change in USD-UAH	16.194	0.094
Ukraine Volume	ln(War Index)	17.966	0.056
Ukraine Volume	ALL	33.250	0.032
ln(War Index)	Log Change in USD-UAH	12.963	0.226
ln(War Index)	Ukraine Volume	3.278	0.974
ln(War Index)	ALL	17.741	0.604

Fig. 28: Granger Causality Wald Tests

Note. The logarithm of the $RussiaWarIndex$ and the log change in the exchange rate Granger cause hryvnia-denominated Tether volume at the 10% level.

The Impact of the Russian Food Embargo on International Agri-Food Trade

Brendon Carter

University of Illinois Urbana-Champaign

Abstract— **In 2014, after the Russian annexation of Crimea and the resulting Western sanctions on Russia, Russia responded by imposing a ban on most food imports from the sanctioning countries, which included the European Union, the United States, Canada, Norway, and Australia. This study uses a structural gravity model to empirically analyze the impacts of the Russian food embargo on bilateral trade flows, welfare, and prices, paying particular attention to the spillover impacts of this trade restriction on third countries, or countries outside of the ban. The key finding from this analysis is that many third countries underwent price impacts due to the embargo that were as large in magnitude as those for the directly-embargoed countries. The results of this analysis show that sanctions can have strong multilateral impacts that go well beyond their intended scope.**

I. Introduction

After Russia forcefully annexed Crimea from Ukraine in February 2014, several Western nations enacted sanctions on Russia in response. The United States, European Union, and several other nations imposed targeted sanctions on firms and individuals linked to the invasion and a trade ban on military-related goods, among other measures. Russia responded with counter-sanctions in the form of an embargo on most agriculture and food imports from the United States, the European Union, Canada, Norway, and Australia. The ban, which was approved on August 7th, 2014, covered meat, fish, milk and dairy products, fruits and nuts, vegetables, and several derivative products. The embargo targeted a significant volume of trade worth over $8.5 billion in 2013, which made up from 30 to 45% of total Russian imports in the affected industries.

This paper uses a structural gravity model to quantify the effects that the food embargo had on consumers and producers of all countries through the use of multilateral resistance terms, as described by Anderson and van Wincoop (2003). This paper specifically focuses on quantifying the impact on countries not a part of the ban in 2014, or "third countries." Using a counterfactual simulation, it estimates the change in trade flows between countries as a result of the embargo and uses these changes to predict the impacts on consumer prices and producer incidence of trade costs. Given the magnitude of the trade disruption caused by the Russian embargo, other studies have attempted to quantify its impacts on trade, prices, and welfare. Crozet and Hinz (2020) estimate the trade losses caused by sanctions related to the 2014 annexation of Crimea for both sides and for both the embargoed products and total trade. They conclude that most of the economic losses to the Western countries were due to the Western sanctions on Russia, as opposed to the countersanctions (i.e., the Russian food embargo). Cheptea and Gaigné (2020) use monthly trade data in combination with Russian customs data to estimate the effects of the ban with a gravity model. They find that the embargo led to EU exports being diverted to other markets at lower prices, and that these diverted exports were more often to other EU countries. In quantifying the amount of trade diversion that occurred because of the embargo, they find that Russia increased its monthly imports of banned goods from non-embargoed countries by 182.5 million Euros, while European Union countries increased their monthly exports to other markets by 230.5 million Euros.

A number of other works have analyzed the Russian food embargo based on computable general equilibrium (CGE) approaches. Boulanger et al. (2016) compute the estimated changes in production, trade, and prices for the banned products with a CGE model. They report the major third country destinations for diverted EU exports and sources for new Russian imports. For Russia, these new sources include Iran, Turkey, Ecuador, China, Belarus, and New Zealand. They report North Africa as the largest destination for new fruit, vegetable, and dairy imports from the EU countries. Kutlina-Dimitrova (2015) uses a similar CGE method to analyze impacts on major exporters and importers. She estimates that the rest of the world (i.e. the third countries as a whole) overall experiences an increase in welfare, defined by equivalent variation, due to the changes in trade flows.

Of the several studies on the Russian food ban, there has been little examination of the effect of the embargo on third countries, and in particular, the implications of the embargo for their consumer prices. There are several reasons for which such impacts are important to consider. First, the countries indirectly affected by trade sanctions (i.e., third countries) such as the Russian food ban are often least developed countries (LDCs) at greater risk for food insecurity. Even small price ripple effects can have significant impacts on countries for which a large share of household income is spent on food. There is a dichotomous effect of price increases on poverty: They can help households who are involved in the production of food, but also hurt consumers within the same country, though existing evidence suggests that the latter effect dominates the former. Along these lines,

Acknowledgment: I would like to thank Dr. William Ridley for his mentorship, support, and guidance throughout this project.

Ivanic and Martin (2008) find that price increases for staple foods in low-income countries increased poverty much more often than they reduced it and that the price-driven increases in poverty tended to be larger than income-driven poverty reductions. Changes in food prices have historically had profound political implications that go beyond their direct economic impacts. Eating is, of course, a necessity, and high food prices often elicit a public response, more so than other prices consumers face. Along with underlying political and social factors, food price inflation is often cited as having contributed to the Arab Spring uprisings (Johnstone and Mazo, 2011). Anticipating sudden price increases in necessities like food and energy could help predict social turmoil in politically sensitive places.

Furthermore, the Russian food embargo warrants study because it is a prominent and recent example of a sudden trade disruption, with such disruptions becoming more frequent in the last decade largely due to geopolitical factors. The use of economic policy is becoming an increasingly common way nations exert their influence on the global stage due to increased globalization. The United States has used sanctions as a bargaining chip with countries like Iran, and the 2022 Russian invasion of Ukraine brought on a flurry of Western sanctions in unprecedented scope. The latter event, along with the Black Sea grain blockade imposed by Russia, is associated with profound changes in food and energy prices all across the globe. These spillover effects are extremely significant in policy discussions of inflation and Western resilience to maintain sanctions for the long term. The Russian food embargo provides a recent historical example of such events and is also comparable to non-agricultural sanctions and embargoes, like that of Western countries against Russian oil.

This paper follows the methodology described by Anderson and van Wincoop (2003). The inward multilateral resistance (IMR) term measures the price index of a product group for consumers of an importing country. Put differently, IMRs capture the average incidence of trade costs that consumers from a country face from all sources, including their own country. Therefore, changes in IMRs can be used to capture changes in theoretical price indices through trade costs. The gravity model methodology allows for the recovery of this term via the use of theoretically consistent fixed effects, which I then use to analyze the impacts the ban had on consumers around the world. The outward multilateral resistance (OMR) term similarly measures the ease for producers in a country to sell their product in all countries, including their own, by capturing their average barriers to trade. Changes in the OMR can therefore be analyzed to see how producers were affected by a policy change.

The use of multilateral resistance terms is an ideal way to measure multilateral trade, welfare, and price impacts. For example, Anderson and Yotov (2010) use these terms to measure buyers' and sellers' incidences of trade costs in Canadian provinces. Along similar lines, Anderson et al. (2018) used changes in IMR and OMR terms to analyze the effects of hypothetically removing all international borders. The advantage of this method is that it allows us to analyze the welfare impacts on countries that are not directly involved with the policy event being studied since the multilateral resistance terms capture the event's effects on trade costs between all countries and thus the multilateral impacts arising from changes in bilateral trade policy.

The results of this study indicate that some third countries were as impacted by the ban as the embargoed countries themselves in terms of changes in prices for consumers and costs for producers. More third countries experienced increases in IMR (relative to a reference country) than decreases, and the increases in some cases were substantial, particularly in the dairy industry[1] . The vegetable industry saw the most widespread price index changes as a result of the ban, with over 40 third countries experiencing a greater than 1% increase in their IMR for fresh vegetables. Third country price increases due to the embargo are estimated to be the largest in many western and central Asian countries, while African countries tend to undergo the largest third-country price decreases. Countries that are estimated to experience the largest decrease in their IMR terms because of the embargo typically also experience a large increase in their OMR terms, or incidences of trade costs for producers, and vice versa. The OMR saw changes greater in magnitude than the changes in IMR. The geographic dispersion of the multilateral resistance changes correlates with the trade diversion as a result of the ban. Both the banned European countries and Russia traded more with other parts of the world to make up for the lost trade. Some countries, like Belarus, essentially blunted the effects of the ban by absorbing imports from Europe and increasing exports to Russia.

The implications of these results are that sanctions and trade policies can yield indirect effects on third countries which are comparable to those on the countries directly affected by the restrictions. These impacts can manifest in either higher or lower prices for consumers in third countries, with the overall direction dependent on the magnitude of trade diversion that occurs in response to the policy change. Policymakers, therefore, should consider that sanctions could impact the stability of the region of the targeted country in their decision to enact such measures. The impacts of sanctions can also be lessened if there are third countries that can easily act as an intermediary for banned goods. If policymakers seek to ensure that the impacts of sanctions remain localized to the targeted country, and have a more severe effect on its economy, measures should be designed in such a way that they cannot be easily circumvented through trade diversion or trade via intermediary countries.

The remainder of the paper is structured as follows. Section 2 introduces the empirical methodology based on the structural gravity framework. Section 3 describes the data on trade flows by food industry. Section 4 describes the

[1] In the structural gravity model used in this paper, the IMR and OMR are in terms of the IMR of a reference country, as will be described in Section 5. South Africa was chosen because of its relatively low and stable trade levels with the embargoed countries and Russia.

econometric results of the gravity equation. Section 5 details the methodology for computing the counterfactual simulation analysis. Section 6 presents the estimates on the effects of the embargo on trade volumes and prices, with the latter captured by counterfactual adjustments in the multilateral resistance terms. Finally, Section 7 concludes with a discussion of the results and their implications.

II. Methodology

To estimate the effects of the Russian food embargo on trade and prices, I use the gravity model as specified by Anderson and van Wincoop (2003):

$$x_{ijt}^{k} = \frac{y_{it}^{k} y_{jt}^{k}}{y_{t}^{Wk}} \left(\frac{y_{ijt}^{k}}{\Pi_{it}^{k} p_{jt}^{k}} \right)^{1-\sigma^{k}} \tag{1}$$

where x_{ijt}^{k} is the monetary value of trade from exporter i to importer j in industry k during year t; $y_{it}^{k} = \Sigma_j x_{ijt}^{k}$ is the total value of production of goods in industry k exported from country i to all countries j, including country i itself, in year t; $y_{jt}^{k} = \Sigma_i x_{ijt}^{k}$ is the total value of consumption of goods from k imported by country j from all countries i, including j itself; y_{t}^{Wk} is the total value of world production in industry k and year t; ${}_{ijt}^{k}$ are the bilateral trade costs in industry k between exporter i and importer j in year t; $\sigma^{k} > 1$ is the Armington-CES elasticity of substitution between goods by country in industry k; Π_{it}^{k} is the outward multilateral resistance term that represents the average (multilateral) impediments to trade faced by country i in exporting goods to every other country including itself, and P_{jt}^{k} is the inward multilateral resistance for country j that captures the average (multilateral) incidence of trade cost on j's consumers and which is also equal to the CES consumer price index.

The outward and inward multilateral resistance terms are defined respectively as:

$$\Pi_{it}^{1-\sigma^{k}} = \Sigma_j \left(\frac{Y_{ijt}^{k}}{P_{jt}^{k}} \right)^{1-\sigma^{k}} \left(\frac{y_{jt}^{k}}{y_{t}^{Wk}} \right)$$
$$\text{and} \quad P_{jt}^{1-\sigma^{k}} = \left(\frac{y_{ijt}^{k}}{\Pi_{it}^{k}} \right)^{1-\sigma^{k}} \left(\frac{y_{it}^{k}}{y_{t}^{Wk}} \right) \tag{2}$$

The OMR term is constructed as a weighted average of bilateral and multilateral trade costs between exporter i and its trade partners j, with weights defined as j's share of total world consumption. Similarly, the IMR term for importer j is a production-share weighted average of bilateral and multilateral trade costs across j's consumption sources. These terms reflect that changes in trade costs and volumes between two countries impact other countries outside of the pair.

The Poisson pseudo maximum likelihood (PPML) method is well suited to estimate the gravity model because it allows for the inclusion of zero trade flows and is consistent and unbiased in the presence of heteroskedasticity, which is common amongst trade data (Santos Silva and Tenreyro, 2006). I use the PPML method estimate the following multiplicative form of the gravity equation:

$$x_{ijt}^{k} = \exp\{\beta_0^{k} + \beta_1^{k} \mathit{EMBARGO}_{ijt} + \alpha_{it}^{k} + \delta_{jt}^{k} + \mu_{ij}^{k}\} + \epsilon_{ijt}^{k} \tag{3}$$

where $\alpha_{it}^{k} = -(1-\sigma^{k})\log\Pi_{it}^{k} + \log y_{it}^{k}$ are exporter-time fixed effects, $\delta_{jt}^{k} = -(1-\sigma^{k})\log P_{jt}^{k} + \log y_{jt}^{k}$ are importer-time fixed effects, μ_{ij}^{k} are time-invariant country-pair fixed effects. The exporter-time and importer-time fixed effects control for changes that affect trade volume specific to individual exporters or importers, such as changes in domestic policies or production and consumption levels. The country-pair fixed effects control for factors between countries that affect trade volume but do not change over time, such as distance. Egger and Nigai (2015) showed that using country-pair fixed effects in this way controls for time-invariant factors better than including variables on factors like distance directly. Therefore, because of the fixed effects, no other direct controls are necessary. EMBARGO is a variable to measure if an exporter-importer pair was affected by the embargo in year t. This variable is equal to 1 for every year the embargo was fully in place, but to account for the ban being enacted in the middle of 2014, it is set at 0.4 for that year to represent the fraction of the year it was in force[2]. Finally, ϵ_{ijt}^{k} is a mean-zero error term.

III. Data

To estimate the gravity model given in Equation 3, I use data on international trade flows in selected industries from the International Trade and Production Database for Estimation (ITPD-E) (Borchert, Larch, Shikher, Yotov, 2022). The ITPD-E covers trade flows for 256 countries in 170 industries, and importantly for the structural gravity model, contains information on domestic (intra-national, i.e., $x_{ijt}^{k}$for$i = j$) trade flows as well as international (Yotov, 2022)[3] . I look at five specific industries from 2000 to 2019 that were directly covered by the Russian food embargo: fresh fruit, fresh vegetables, processing/preserving of meat, processing/preserving of fish, and dairy products, which together encompass the specific commodities targeted by Russia's import ban. The Russian food embargo encompasses a list of agricultural products classified by Harmonized System (HS) commodity codes, whereas the ITPD-E classifies several industries based on several different industry and commodity nomenclatures. I match the banned HS commodity codes to their corresponding ITPD-E industries for this

[2]While trade is likely to not be uniformly distributed throughout the year, setting it as a fraction of time is the best approximation absent monthly trade data.

[3]Yotov (2022) lists several reasons why domestic trade flows are important to include in the gravity model. Especially pertinent to this study, they enable the identification of trade diversion as a result of bilateral trade policies like the embargo.

analysis[4].

The goal of this paper is to evaluate the impacts the ban had on food categories as a whole rather than the exact products specified in the ban, so it is not essential for the included industries to completely align with the banned products. Nonetheless, given the wide coverage of the ban, the five chosen industries correspond reasonably exactly to the products covered by the ban. This is evident by the dramatic drop in exports in these products from the embargoed countries to Russia by the year 2015, as shown in Figure 1. The non-banned industries (all other industries besides the five indicated in Figure 1) only had a slight decrease in 2014, as shown in spill over, and fell more in 2015 but by a much smaller amount than the banned industries. This demonstrates that the effect of other political outcomes correlated with the sanctions did not lead to a significant decrease in trade in the year of analysis, 2014. Moreover, the figure makes clear that long-term trade in the non-banned industries persisted while trade in the banned industries remained close to zero.

IV. Econometric Results

I present the econometric results from equation (3) in Table 1. The EMBARGO variable coefficient, β_1^k, is estimated to be negative and statistically significant at the $p < 0.01$ level for every industry. The effect of the Russian food ban on trade from an embargoed exporter to Russia in industry k is estimated to be (exp$\{\beta_1^k\} - 1$)*100 percent. Based on the estimates for β_1^k for each industry, this corresponds to average reductions in bilateral exports of –92.4% for fresh fruit, –91.3% for fresh vegetables, –97.1% for dairy products, –95.1% for meat, and –96.3% for fish.

V. Counterfactual Simulations

To quantify the effects of the Russian food embargo, I define a counterfactual scenario in which the embargo was never enacted and simulate the predicted values of trade in this setting based on the methodology developed by Anderson et al. (2018). The effect of the food ban on trade is captured in the EMBARGO variable coefficient, so the simulation consists of switching this variable to zero for all observations and then re-estimating the OMR and IMR terms and trade between countries. The new IMR and OMR values therefore represent the trade costs faced by consumers and producers, respectively, throughout the entirety of 2014 if the embargo never occurred. The simulation holds all factors besides the embargo's estimated effect on trade flows constant. Since these other factors, like domestic policies, are unchanged, we can then assume the differences in outcomes in the simulation are caused solely by the embargo. For trade, the differences represent the change in the value of trade between two countries because of the embargo. The differences in the IMR terms represent how the price index for consumers changed, aggregated through the entirety of 2014, and the differences in OMR terms similarly represent how multilateral trade costs for producers changed.

The multilateral resistance terms are interpreted relative to a reference country (chosen for this purpose as South Africa) whose importer fixed effects are excluded from the estimation to avoid perfect collinearity between the fixed effects and the intercept. I chose South Africa because its trade with Russia and the embargoed country was stable throughout the initial ban period and because it has a satisfactory amount of trade across the industries, given that its foreign trade levels are above the average of the dataset for 2014. Each IMR change can therefore be interpreted by how much more or less consumers must pay for the different food categories as compared to South African consumers over the course of 2014, the year of analysis.

The simulations require a value for σ^k , which cannot be directly identified from the estimation of equation (3). σ^k is used to compute the estimates of the IMR and OMR terms, as shown in equation (2), and also is used in the full general equilibrium steps to compute changes in factory-gate prices. To assign a value for σ^k, I use the trade elasticities estimated by Fontagné et al. (2022) for this value. I take the average elasticities from the banned meat groups (4-digit HS headings 0201–0203, 0207, and 0210) to obtain an elasticity of substitution value of 6.5, from the banned fish groups available in the data from Fontagné et al. (4-digit HS headings 0301–0307) to get an elasticity of 4.7, from the banned vegetable groups (4-digit HS headings 0701–0714) to get an elasticity of 5.6, from the banned fruit groups (0803-0811, 0813) to get an elasticity of 8.0, and from the banned dairy groups (0401–0406) to get an elasticity of 5.4.

I exclude banned HS codes 1901 (which includes food preparations from flour, starch, etc., as well as from dairy products) and 2106 (food preparations, not elsewhere specified) to avoid mixing products from the different industries. I also exclude the nut products (0801–0802) from the fruit groups because nuts are recorded under a separate industry in the ITPD-E data and were not covered by the Russian embargo.

Using the parameter estimates yielded by estimation of equation (3), I compute baseline bilateral trade flows for each industry in the year 2014 under the baseline scenario where the food embargo was in place:

$$x_{ij14}^{k,B} = \exp\{\hat{\beta}_0^k + \hat{\beta}_1^k \mathit{EMBARGO}_{ij14} + \hat{\alpha}_{i14}^{k,B} + \hat{\delta}_{j14}^{k,B} + \hat{\mu}_{ij}^k\} \quad (4)$$

[4]Fresh fruit, fresh vegetables, and dairy products are classified into the agriculture sector, which use FAOSTAT Classification List (FCL) codes. To prevent double counting, ITPD-E counts most meat products in the manufacturing sector, which is based on the International Standard of Industrial Classification (ISIC) system. The HS codes for all meat products stipulated by the ban (0201–0203, 0207, 0210, and 1601) correspond to ISIC4 classification 1010, which make up the ITPD-E's processing/preserving of meat industry. The HS codes for the banned fish and seafood products (0301-0308), correspond mostly to ISIC4 codes 03 and 1040, which themselves correspond to the "Fishing" and "Processing/Preserving of Fish" ITPD-E industries, respectively. The trade value of the latter, however, is 10 times larger than the former, so I chose "Processing/Preserving of Fish" to analyze the effect of this part of the ban.

Figure 1: Exports from Embargoed Countries to Russia in Selected ITPD-E Industries

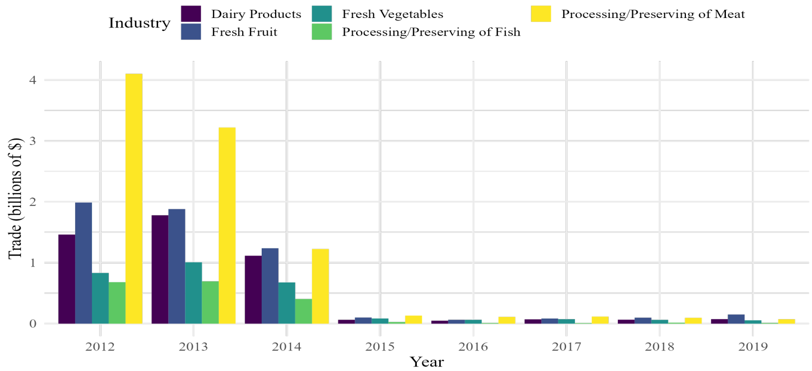
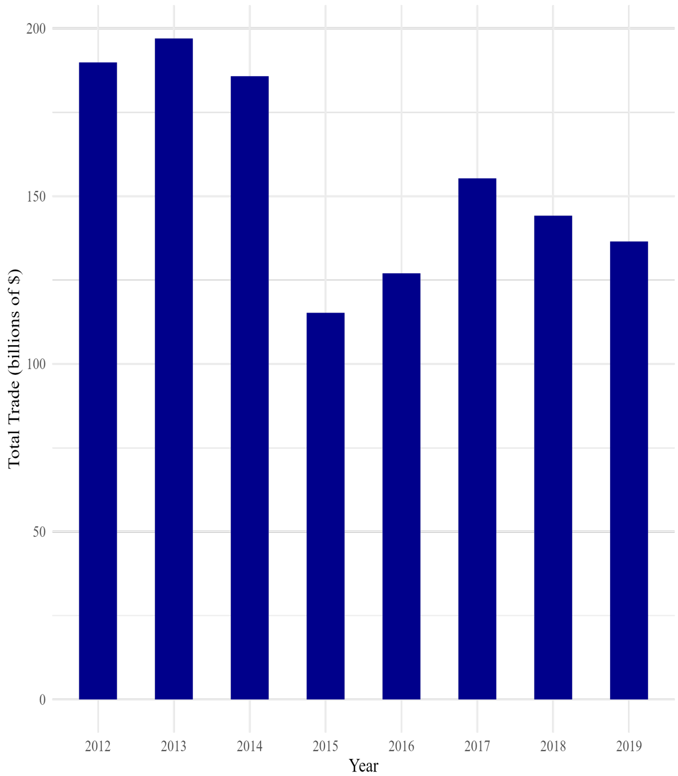


Figure 2: Exports from Embargoed Countries to Russia in Non-Banned Industrie



Here, "$\wedge$" signifies the parameter estimates estimated by equation (3). Equation (4) generates the baseline trade flows, $x_{ij14}^{(}k, B)$, which are later compared to the trade flows from the counterfactual. To get the baseline IMR and OMR terms, I use the importer and exporter fixed effects to calculate equation (5) as per Fally (2015):

$$\Pi_{it}^{\widetilde{B^{1-\sigma^k}}} = \hat{y}_{R14}^{k,B}\hat{y}_{it}^{k,B}\exp\{-\hat{\alpha}_{i14}^{k,B}\text{and}P_{jt}^{\widetilde{B^{1-\sigma^k}}} = \frac{\hat{y}_{it}^{k,B}}{\hat{y}_{R14}^{k,B}}\exp\{-\hat{\delta}_{j14}^{k,B}\} \quad (5)$$

Then I create the counterfactual simulation to evaluate the effect of the food embargo by first re-estimating α_{i14}^{k}, δ_{i14}^{k} with the observed trade flows from the ITPD-E and the counterfactual $EMBARGO$ variable ($EMBARGO_{ij14}^{C}$), which is set to zero for all countries to represent the counterfactual scenario where Russia did not impose the food embargo:

Table 1: Estimates of Gravity Equation for Each Industry

	EMBARGO Variable			
Industry	Estimate	Std. err.	Obs.	Pseudo R^2
Fresh Fruit	–2.580***	(0.219)	250,512	0.995
Fresh Vegetables	–2.447***	(0.193)	209,108	0.998
Dairy Products	–3.547***	(0.201)	297,569	0.994
Meat	–3.025***	(0.215)	316,602	0.994
Fish	–3.292***	(0.490)	367,382	0.972

Notes: Dependent variable is the value of bilateral trade for the indicated industry. Estimation method is PPML. Rows correspond to estimates on the EMBARGO variable obtained from separate regressions of equation (1) by industry. Robust standard errors are reported in parentheses.*** represents $p < 0.01$.

$$x^k_{ij14} = \exp\{\hat{\beta}^k_0 + \hat{\beta}^k_1 \mathit{EMBARGO}^C_{ij14} + \alpha^k_{i14} + \delta^k_{i14} + \hat{\mu}^k_{ij}\} + \epsilon^k_{ij14} \tag{6}$$

This step computes the conditional general equilibrium model, where total output and expenditures are unchanged, generating new conditional fixed effects $\hat{\alpha}^C_{i14}$ and $\hat{\delta}^C_{j14}$. From these fixed effects, I get the conditional changes in IMR and OMR in the same manner as the baseline step. Using the new estimated values, I compute counterfactual bilateral trade flows:

$$x^{k,C}_{ij14} = \exp\hat{\beta}^k_0 + \hat{\beta}^k_1 \mathit{EMBARGO}^C_{ij14} + \hat{\alpha}^{k,C}_{i14} + \hat{\delta}^{k,C}_{j14} + \hat{\mu}^k_{ij} \tag{7}$$

This step generates the conditional counterfactual changes in trade. Next, to get estimates for the full general equilibrium effects, I begin an iterative procedure to account for changes that the conditional equilibrium effects would have on factory-gate prices, output, and expenditure as described by Anderson et al. (2018). First, I allow factory-gate prices, denoted by $p^{k,CFL}_{it}$, to adjust to the changes in the multilateral resistances:

$$\Delta p^{k,CFL}_{i14} = \left(\frac{\exp(\hat{\alpha}^{k,C}_{i14})/y^k_{R14}}{\exp(\hat{\alpha}^{k,B}_{i14})/y^k_{R14}}\right)^{\frac{1}{1-\sigma^k}} \tag{8}$$

where $\hat{\alpha}^{k,C}_{i14}$ for the first loop of the iterative procedure is the importer fixed effect from the conditional general equilibrium step and y^k_{R14} is the expenditures of the reference country, South Africa, in year 2014. With these changes in factory-gate prices, I then compute new values for income, y^{CFL}_{i14}, and expenditures, y^{CFL}_{j14}, for each country:

$$\begin{aligned} y^{k,CFL}_{i14} &= \left(\frac{p^{k,CFL}_{i14}}{p^{k,B}_{i14}}\right) y^{k,B}_{i14} \\ \text{and} \quad y^{k,CFL}_{j14} &= \left(\frac{p^{k,CFL}_{j14}}{p^{k,B}_{j14}}\right) y^{k,B}_{j14} \end{aligned} \tag{9}$$

These new values are used to recalculate the trade flows, which I then use to re-estimate the structural gravity model in the same manner as the conditional general equilibrium step. This produces new values for the fixed effects, which are then used for the new $\hat{\alpha}^{k,CFL}_{i14}$ to repeat the iterative procedure again until the change in factory-gate prices is sufficiently close to zero. Once this final step is completed, the generated fixed effect and trade flows are used to create the full general equilibrium metrics, which are the full estimates for the counterfactual scenario in which Russia never enacted the food ban.

VI. Simulation Results

A. *Trade Diversion*

By comparing the baseline trade flows from equation (4) with the trade flows predicted in the full general equilibrium model, we can estimate the embargo's overall impact on trade between countries. These differences for the year 2014 are reported in Table 2.

The largest estimated trade losses are found in the meat industry, while the vegetable industry undergoes the smallest decrease; however, the estimated impacts are comparable

Table 2: Change in Exports from Embargoed Countries to Russia due to Embargo by Industry

	Vegetables	Fruit	Dairy Products	Meat	Fish
Total Change (million USD)	–485.5	–705.6	–810.0	–1,490.6	–820.3
Percentage Change	–46.2%	–46.5%	–48.1%	–53.3%	–63.8%

in magnitude across all the industries. Overall, the trade loss totaled to \$4.3 billion in 2014, a 51.7% decrease. The considerable lost trade led to both parties, Russia and the embargoed countries, diverting trade to other markets. For embargoed countries, the trade diversion can be to third countries, or to other embargoed countries. Table 3 shows the changes in value and percentage for trade in the countries affected by Russia's import ban for each industry. The difference between the baseline model and the full model is given by the rows labeled "Full Change" and is negative for vegetables, dairy products, and meat. The negative values do not indicate that these countries were less inclined to trade with one another due to the embargo, but rather that the effect the embargo had on decreasing the prices for these foods was dominant over the change in the overall quantity of trade. The contrasting signs on several of the results between the conditional and full general equilibrium models demonstrate this, because the conditional model does not allow prices and production to adjust.

Trade diversion was considerable for many third countries. Table 4 reports the countries that received the most diverted trade throughout the five studied industries. Geographically, it can be seen that the embargoed countries traded the most with large economies in East Asia, followed by Ukraine and Belarus in Eastern Europe, and then with several countries near the Middle East and North Africa region. Given that many of the targeted agricultural goods are highly perishable (e.g., meat, dairy), the EU's immediate neighbors in Eastern Europe and the MENA region are the most logical choice for new trade. China is the obvious exception, but most of the diverted exports to China were in the meat and fish industries, accounting for \$198 million of the \$255 million increase. Many of Russia's new trade partners, on the other hand, were located far away in South America. This reflects the fact that while food consumption is more evenly spread throughout the world, food production is highly regional.

Also notable is the scale of Russian trade diversion compared to that of the embargoed countries. In total, Russia is estimated to have imported around \$3.7 billion more in food products from the third countries as a result of the embargo. The embargoed countries only exported \$1.7 billion more to third countries in comparison. This result is partially explained by the price effect; the EU countries had to sell their exports at a cheaper price while Russia had to pay more for its imports. However, the disparity is not entirely due to the price effect since the analogous changes estimated under the conditional step (in which prices are held constant) amount to around \$2.1 billion and \$3.4 billion for the embargoed countries and Russia, respectively. The remaining difference is mostly due to the embargoed countries trading amongst themselves.

Most third countries experienced an increase in foreign trade as a result of the embargo. Arkolakis et al. (2012) showed that a decrease in the share of expenditures devoted to domestic goods implies a welfare gain, so most third countries benefited from the embargo on net. However, as will be demonstrated in the next section, consumers generally were worse off in most third countries while producers gained from the food embargo. Some third countries, however, did experience decreases in trade. These countries were often net importers who bought from the exporters that Russia pivoted towards because of the embargo.

B. Effects on Consumer Prices

The effect on the price index for consumers for each of the product groups k is estimated with the change in the IMR term for a country j in year t: P_{jt}^{k}. I will use the terms IMR and consumer price index interchangeably. Estimates of this term for each country are interpreted as changes relative to the designated reference country, South Africa.

Unsurprisingly, Russian consumers faced the largest increase in prices in each of the industries. In contrast, the largest decreases in consumer prices are typically estimated to occur for the embargoed European countries, and generally those countries more proximate to Russia. Notably, the increase in prices faced by Russia was always larger than the largest price decrease of any embargoed country.

Certain third countries, however, faced price changes that were larger in magnitude than many of the embargoed countries. This was particularly true for countries facing price increases. Table 5 reports the third countries that experienced the largest consumer price increases, showing that the largest impacts were in the dairy industry. Within this industry, the largest IMR decrease for the embargoed countries was Lithuania, which is estimated to have undergone a –3.2% change in its consumer price index for dairy compared to the reference country of South Africa. Kyrgyzstan, Kazakhstan, Belarus, Ukraine, Tajikistan, Moldova, Mongolia, Turkmenistan, Armenia, and Azerbaijan all undergo estimated IMR increases greater in magnitude than that, demonstrating the significance of the indirect effects of the food embargo. Other studies on trade disruptions typically neglect these indirect effects, but this analysis shows that they can

Table 3: Change in Trade Among Embargoed Countries due to Embargo by Industry

	Vegetables	Fruit	Dairy Products	Meat	Fish
Conditional Change (million USD)	164.1	336.6	52.6	171.5	310.3
Conditional Change (%)	0.8%	1.5%	0.1%	0.3%	1.4%
Full Change (million USD)	–13.9	136.4	–196.0	–206.8	167.2
Full Change (%)	–0.1%	0.6%	–0.5%	–0.3%	0.8%

Table 4: Countries with largest trade diversion effects

Country	Change in Imports from Embargoed Countries (million USD)	Country	Change in Exports to Russia (million USD)
China	254.5	Brazil	731.2
Japan	133.1	Belarus	369.5
Ukraine	128.1	China	323.5
Belarus	104.6	Turkey	236.2
Hong Kong	96.0	Argentina	233.1
Kazakhstan	93.5	Ecuador	206.6
Saudi Arabia	55.0	Ukraine	178.1
Egypt	50.4	Chile	154.3
United Arab Emirates	48.9	Uruguay	132.5
Algeria	47.6	Paraguay	127.8

be very large. Kyrgyzstan's increase was only 2.8 percentage points less than Russia's IMR increase, which was 12.4%. Countries that increased their exports to Russia often had price increases resulting from the increased Russian import demand. However, many countries that did not undergo significant shifts in their trade with Russia also experienced price increases if they were geographically close to Russia, such as Tajikistan, because they import from the same dairy producers as Russia and were therefore indirectly affected by the increase in Russia's import demand resulting from the embargo.

The dairy industry had the most profound IMR changes for third countries out of the five industries, but also notable was the number of countries impacted by the ban on vegetables. Over 40 nations besides Russia experienced at least a 1% increase in the price index for fresh vegetables as a result of the Russian ban, the highest being Azerbaijan, Armenia, Tajikistan, Iran, and Turkey. Map 1 depicts these countries, showing that the price effects rippled through Asia, Eastern Europe, and the Middle East. The third country increases in the fish industry were quite small, with no country estimated to undergo an IMR increase greater than 1%.

While price increases were predominant, there were also some third countries that experienced price decreases because of the ban. What is significant about these third country examples is that the effects were sometimes larger than many of the embargoed countries themselves. Gambia, for example, experienced a larger IMR decrease than 27 out of the 32 embargoed countries for fresh vegetables. However, price decreases were generally small in magnitude and less prevalent than price increases. The loss of trade between the embargoed countries and Russia represents a much larger share of total trade for Russia, since Europe provided much of Russia's food imports, but Russia was merely one of many

Table 5: Largest Third-Country Consumer Price (IMR) Increases by Industry

Fresh Vegetables		Fresh Fruit		Dairy Products		Meat		Fish	
Country	% Change	Country	% Change	Country	% Change	Country	% Change	Country	% Change
Azerbaijan	4.55	Azerbaijan	2.19	Kyrgyzstan	9.60	Kazakhstan	2.61	Tajikistan	0.52
Armenia	3.29	Armenia	1.25	Kazakhstan	9.04	Ukraine	1.88	Burkina Faso	0.23
Tajikistan	3.28	Serbia	1.24	Belarus	8.30	Moldova	1.31	Cuba	0.20
Iran	3.01	Moldova	1.01	Ukraine	6.87	Azerbaijan	1.20	Ivory Coast	0.20
Turkey	2.99	Uzbekistan	1.00	Tajikistan	6.83	Mongolia	1.15	China	0.19

export destinations for the EU. This fact explains why the consumer price changes in the embargoed countries were much smaller than those for Russia, and thus why more third countries experienced spillover price increases.

C. Effects on Producer Incidence of Trade Costs

Lastly, I analyze the effect of the embargo on producers with the change in the OMR term for a country i in 2014: Π^k_{i14}. This approach follows Anderson and Yotov (2010), who use the OMR as the sellers' incidence of the bilateral trade costs for a study of Canadian trade. The estimated changes in the OMR terms can therefore be used to measure the effect of the trade embargo on the multilateral trade costs faced by producers. As in the preceding analysis, the OMR is also interpreted relatively to the designated reference country's (South Africa) IMR.

Table 6 shows the countries with the five largest increases and decreases in their OMR term for third countries. In contrast with the patterns observed in the results on the IMR term, there were more countries with estimated decreases than increases in OMR and the decreases tended to be larger. Notably, many of the countries that undergo the largest increases in their IMR term also experience the largest decline in their OMR term. For example, in the dairy industry, Belarus is estimated to undergo the third highest increase in the consumer price index (+8.3%) and also experiences the largest overall decrease in the OMR term (–16.0%).

The OMR changes tended to be larger than the IMR changes, implying that producers are estimated to have borne a greater share of the incidence of changes in multilateral trade costs. In line with the estimates on IMR changes, dairy also had the highest OMR changes among the five industries. The OMR effects were most geographically widespread in the vegetable industry, with over 50 third countries having greater than a 1% change as a result of the ban.

Several third countries also experienced significant OMR increases because of the embargo. Fresh fruit and fresh vegetables were the only industries for which any countries experienced a greater than 1% increase in OMR, which include the Central African Republic and the Democratic Republic of the Congo for vegetables and The Gambia for fruit. While the countries with the greatest decrease in OMR tended to be in Central Asia, the top 5 countries by increase in OMR in the fruit and vegetable industries were all in Africa. Many of these countries are lower income countries, which illustrates that the embargo had large impacts on people at risk of food insecurity.

VII. CONCLUSION

Sanctions and trade embargoes have become an increasingly prevalent form of trade disruption in the modern global economy. The disruptions caused by such bans on trade have significant impacts on prices, welfare, and trade flows for the directly-affected nations as well as third countries. While there have been studies on the impacts on directly-affected nations for such policies, there has been less attention paid to the third-country effects, effects which I find in my analysis to be large and widespread. Given that countries' food economies are inextricably linked via the global trading system, any disruptions to particular trading linkages will necessarily have spillover effects on third countries not directly involved in the disruption, effects which, as this study shows, can be considerable. This study uses a structural gravity framework to quantify the effects of the Russian food embargo, making use of the gravity model's multilateral resistance terms to evaluate the effects on trade volumes and prices, and in particular, the multilateral impacts caused by this significant trade disruption.

My simulation estimates depict that Russia diverted more of its trade to third countries than did the embargoed countries, and this trade diversion was thus associated with much larger price increases for consumers in third countries than price decreases. The price increases were often geographically clustered around Russia in Eastern Europe and Central Asia, while price decreases were most commonly found for

Map 1: Countries with Greater than a 1% Increase in Consumer Prices for Vegetables

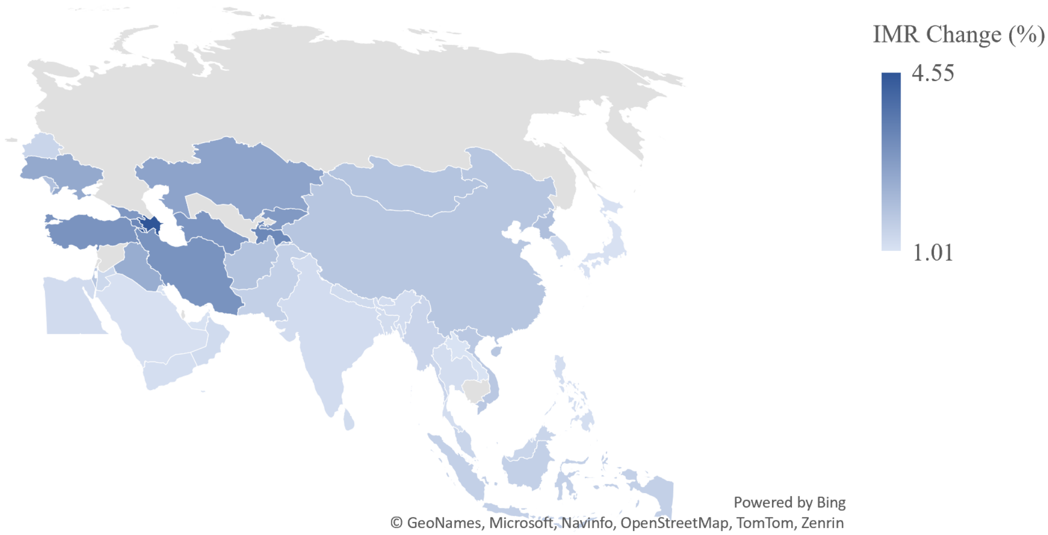


Table 6: Largest Third-Country Producer Incidence Decreases by Industry

Fresh Vegetables		Fresh Fruit		Dairy Products		Meat		Fish	
Country	% Change	Country	% Change	Country	% Change	Country	% Change	Country	% Change
Azerbaijan	–5.95	Azerbaijan	–2.72	Belarus	–16.03	Belarus	–7.49	Belarus	–5.68
Uzbekistan	–4.23	Moldova	–1.83	Kyrgyzstan	–12.93	Palau	–4.4	Armenia	–3.12
Armenia	–4.22	Serbia	–1.79	Kazakhstan	–12.33	Kazakhstan	–3.03	Kazakhstan	–0.99
Tajikistan	–4.15	Armenia	–1.55	Ukraine	–9.65	Paraguay	–2.95	Kyrgyzstan	–0.75
Iran	–3.79	Uzbekistan	–1.45	Tajikistan	–9.04	Mongolia	–2.66	Faroe Islands	–0.64

African countries. The embargoed countries predictably had the largest price decreases for consumers, but the third-country price increases were often greater in magnitude, which highlights the significant multilateral effects of the embargo. Notably, the embargoed countries were estimated to undergo an expansion in the quantity of trade with each other, which helps to explain in part why these countries undertook relatively less third-country trade. However, the value of this intra-embargoed trade was actually lower with the embargo because of the embargo's downward effect on prices in these countries. Producers similarly faced a lower incidence of trade costs in many third countries as consumers faced higher prices, usually in the same countries that are closer in proximity to Russia.

Ultimately, the results of this analysis underscore the consequences of bilateral trade disruptions for the global economy, and critically, the multilateral/spillover effects that such disruptions engender. By illustrating how price and welfare effects ripple across borders via the reallocation of trade, the findings of this study can help predict how countries will be affected by trade disruptions in the future. This analysis is additionally applicable to trade issues beyond agri-food industries. With the recent introduction of a number of large-scale trade sanctions or embargoes, for instance, Western bans on Russian oil as part of the sanctions over the 2022 Russian invasion of Ukraine the analysis of events such as the Russian food embargo are key to developing policy recommendations towards mitigating the harmful effects of such disruptions.

REFERENCES

[1] Anderson, J. E., Larch, M., and Yotov, Y. V. (2018). GEPPML: General equilibrium analysis with PPML. The World Economy, 41(10), 2750–2782.

[2] Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. American Economic Review, 93(1), 170–192.

[3] Anderson, J. E. and Yotov, Y. V. (2010). The changing incidence of geography. American Economic Review, 100(5), 2157–2186.

[4] Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? American Economic Review, 102(1), 94-130.

[5] Borchert, I., Larch, M., Shikher, S., and Yotov, Y. V. (2022). "The International Trade and Production Database for Estimation - Release 2 (ITPD-E-R02)," USITC Working Paper 2022–07–A.

[6] Borchert, I., Larch, M., Shikher, S., and Yotov, Y. V. (2021). The International Trade and Production Database for Estimation (ITPD-E). International Economics, 166, 140–166.

[7] Boulanger, P., Dudu, H., Ferrari, E., and Philippidis, G. (2016). Russian roulette at the trade table: A specific factors CGE analysis of an agri-food import ban. Journal of Agricultural Economics, 67(2), 272–291.

[8] Cheptea, A. and Gaigné, C. (2020). Russian food embargo and the lost trade. European Review of Agricultural Economics, 47(2), 684–718.

[9] Crozet, M. and Hinz, J. (2020). Friendly fire: The trade impact of the Russia sanctions and counter-sanctions. Economic Policy, 35(101), 97–146.

[10] Egger, P. H., and Nigai, S. (2015). Structural gravity with dummies only: Constrained ANOVA-type estimation of gravity models. Journal of International Economics, 97(1), 86-99.

[11] Fally, T. (2015). Structural gravity and fixed effects. Journal of International Economics, 97(1), 76–85.

[12] Fontagné, L. Guimbard, H., and Orefice, G. (2022). Tariff-based product-level trade elasticities. Journal of International Economics, 137, 103593.

[13] Ivanic, M. and Martin, W. (2008). Implications of higher global food prices for poverty in low-income countries. Agricultural Economics, 39, 405-416.

[14] Johnstone, S. and Mazo, J. (2011). Global warming and the Arab Spring. Survival, 53(2), 11–17.

[15] Kutlina-Dimitrova, Z. (2017). The economic impact of the Russian import ban: A CGE analysis. International Economics and Economic Policy, 14, 537–552.

[16] Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. Review of Economics and Statistics, 88(4), 641–658.

[17] Yotov, Yoto V. (2022). On the role of domestic trade flows for estimating the gravity model of trade. Contemporary Economic Policy 40(3), 526–540.

Inference with Machine Learning: Commercial Banking Background and Corporate Innovation

Yongzhe Wang

University of North Carolina, Chapel Hill

Abstract— **This study reveals that commercial banking backgrounds of corporate executives negatively impact corporate innovation by reducing patent applications, researcher employment, and R&D expenses. The effects intensify as the number of executives possessing such backgrounds increases. A novel machine learning framework, ML-DID, is outlined to mitigate endogeneity concerns, including omitted variable bias and selection bias in the setting of "Panel Data with Multiple Treatment Timing." Its robustness is tested against various specifications using the random forest algorithm. Additional analysis demonstrates that the impact of a banking background is strongest among firms with financial constraints and in less competitive industries.**

I. Introduction

"Commercial Banking Backgrounds" refer to corporate executives who have previously served in the banking industry for commercial banks. Such a background holds immense value for corporations as it often provides opportunities for obtaining larger loans and receiving preferential treatment from banks, endowing them with distinctive advantages over their counterparts. Executives with banking experience can leverage their financial expertise to benefit their corporations. However, it is worth noting that those in these positions may display conservative tendencies. Bamber et al. (2010) suggest that executive managers with financial backgrounds often make conservative revenue forecasts that underestimate future earnings. Considering these contrasting factors, the impact of banking backgrounds on corporations, particularly their capability to innovate—a crucial element for long-term growth—remains uncertain.

Corporate innovation catalyzes competitive advantage, enabling companies to differentiate themselves, foster growth, and expand into new markets. It also plays a crucial role in addressing societal challenges such as environmental sustainability, income inequality, and healthcare (Robinson, 1999; Muthuri et al., 2012). Hence, this study provides valuable insights for public policymakers and corporate managers. They can understand the impact of executives with banking backgrounds and formulate effective strategies to utilize these bank-firm connections better. So, how will executives possessing banking backgrounds affect innovation? The presence of banking backgrounds could conceivably bolster a firm's innovative abilities by endowing it with sufficient financial resources necessary for research projects. On the other hand, these backgrounds, which contribute to the firm's financial stability, along with the risk-averse nature commonly observed in these executives, might impede long-term endeavors such as innovation (Bamber et al., 2010). Pressure from investors to prioritize short-term profits can further discourage the pursuit of innovation as a growth strategy (Stein, 1989; Holmstrom, 1989; Acharya & Lambrecht, 2015).

Based on an empirical analysis using a comprehensive dataset on Chinese publicly listed firms from the China Stock Market & Accounting Research (CSMAR) database, the findings demonstrate that corporations where at least one executive possesses a commercial banking background tend to exhibit a lower level of innovation. This is reflected in the reduced number of patent applications, expenditure on research and development (R&D), and fewer R&D personnel compared to firms without such a banking relationship. The relationship is both economically and statistically significant. Notably, this effect is most pronounced in industries with lower competition levels and higher financial constraints. Corporations often prioritize financial stability over innovation in these sectors, contributing to the observed lower innovation rates. Conversely, in highly competitive industries and firms with ample funding, I observed a weak influence of commercial banking backgrounds on corporate innovation. This is attributed to the intense market competition, which compels firms to prioritize ongoing development efforts to stay competitive, and the availability of sufficient financial resources to support innovation.

Endogeneity concerns can be a severe problem in this study. Reverse causality and self-selection bias will either increase or decrease the point estimate. Instead of executive banking backgrounds causing lower innovation, there are possibilities that firms with lower innovation might attract executives with banking backgrounds due to reasons such as risk aversion or a preference for conservative management styles. Additionally, executives with banking backgrounds may choose to work in industries or firms that align with their skills and preferences, which could result in less innovation. This bias could arise from the executives' own preferences or the hiring practices of firms.

To this end, I propose a new machine learning framework that tests and mitigates the abovementioned biases. Resembling the difference-in-differences approach, ML-DID (Machine Learning-Driven Difference-in-Differences) leverages the characteristics of each firm to train a machine learning model that predicts the firm's inherent potential for innovation in the present context. In other words, the model plays a role in estimating the counterfactual innovation capacity of firms in the absence of a bank-firm relationship during the years when such a relationship was present. This method offers several advantages, particularly because it circumvents the need for a parallel trend assumption. The estimator is trained on the same set of firms as the sample used to estimate the effect of banking backgrounds. An inherent strength of ML-DID lies in its ability to accommodate potential unobserved shocks that might impact firm innovation under the data structure of "Panel Data with Multiple Treatment Timing"—that is, the binary treatment indicator has a unique timing of assignment for each entity. By utilizing data from each year in the training process under the "multiple treatment timing" setting, the model becomes empowered to account for broad-level fluctuations spanning the entire sample period and capture patterns that linear models cannot capture (Mullainathan & Spiess, 2017). An additional advantage is that researchers are free to add additional features (variables) to train the model, as machine learning algorithms will automatically exclude unimportant features. ML-DID effectively neutralizes any confounding effects, similar to the standard difference-in-differences method.

To further ensure the robustness of the results and the validity of this new methodology, I have also emphasized traditional econometric methods, controlling for several variables related to a firm's motivations to innovate. These controls include industry- and year-fixed effects. Moreover, I have incorporated additional proxies for innovation, such as patent shock, patent citation, and R&D per employee. Additionally, I have employed propensity scores matching with various specifications. The main results remain unaffected despite these changes in specifications. Remarkably, the negative relationship between banking background and corporate innovation persists across these different approaches. This persistence strongly suggests a causal effect of banking background on corporate innovation and further supports the validity of the methodology in this context.

This paper makes two marginal contributions. First, it fills a literature gap by combining research on commercial banking and corporate innovation from a unique perspective. Investigating the importance of a banking background as an alternative means of establishing close connections between firms and banks sheds light on a previously unexplored area. This empirical finding has far-reaching implications for policy-making and corporate management. Policymakers aiming to foster innovation and economic growth can utilize this knowledge to design targeted initiatives and frameworks that reduce excessive bank-firm connections and create an environment conducive to innovation. Corporate managers, too, will gain insights into the potential long-term consequences of reducing innovation efforts once they recognize the effect of a banking background. Second, the paper introduces a machine learning approach that is easy to implement and allows for various specifications. This approach serves the essential purpose of mitigating self-selection bias and omitted variable bias, which are common challenges in empirical research. The method's versatility is particularly valuable in overcoming limitations in traditional econometric methods, especially when the difference-in-differences approach becomes invalid due to unparallel pre-trends. Though the current method may still be in its early stages of development, its potential to mitigate biases in ways that traditional econometric methods cannot makes it a promising avenue for future research. By providing a robust tool for addressing bias concerns, the proposed machine learning approach represents a valuable contribution to empirical research on banking and corporate innovation.

The paper is organized as follows. In Section 2, a review of existing literature is undertaken. In Section 3, I introduce the data set employed in this study while expanding on the methodological approach utilized. Subsequently, in Section 4, an exposition of the empirical model is provided. Section 5 presents the results derived from a series of sensitivity tests to augment the robustness of the findings. In Section 6, additional analyses are undertaken, exploring various parameters to unravel the nuances of the relationship under investigation. Section 7 concludes.

II. Literature Review

The exploration of banking backgrounds within the realm of financial economics assumes importance, as commercial banks play a pivotal role in shaping the financial landscape of firms. Executive managers possessing such backgrounds foster a unique connection with banks, which provides advantages including mitigating information asymmetry, bolstering the monitoring capabilities of banks, and fortifying the resource allocation and corporate management of firms (Fama, 1985; James, 1987; Besanko & Kanatas, 1993; Gorton & Schmid, 2000; Levine, 2002; Yildirim, 2020). Consequently, firms exhibiting a predilection for banking institutions reap the benefits of lower loan rates and more relaxed collateral requirements (Berlin & Mester, 1999). Furthermore, bank loans are vital in ameliorating a firm's financial constraints (Barth et al., 2006; Pan & Tian, 2015). Aslan (2016) and Dwenger et al. (2020)'s studies also show that the strength of this relationship holds substantial influence over investment structures, and a reduction in lending activities decreases firm employment. Additionally, broader evidence from Acharya and Xu (2017)'s study on US firms suggests that publicly firms engage in external finance such as borrowing from banks appear to have greater capabilities for innovation than their private counterparts.

Nevertheless, another cluster of empirical evidence, put forth by Gao (2008) and Lin et al. (2009), indicates that firms with robust bank-firm connections often exhibit poorer performance than their counterparts. This can be due to insufficient investment as a result of bank shareholding. Dass and Massa (2011) argue that a strong bank-firm connection reduces the firm's stock liquidity by introducing information asymmetry, which leads to adverse selection for other market participants. Given the mixed results of the bank-firm relationship on various aspects of corporate finance, it is important to examine the effects carefully. This present study makes a marginal contribution to the existing body of literature by illustrating the ramifications of a specific form of bank-firm relationship—the banking background of corporate executives—on corporate innovation. The exploration of corporate innovation assumes great importance, as it occupies a central position in a firm's long-term advancement and bolsters the firm's competitive position within the industry (Solow, 1957). Benfratello et al. (2008) discovered that the establishment of banks augments the likelihood of process innovation, particularly in high-tech sectors and small-scale enterprises. Cornaggia (2015) revealed that banking competition curtails innovation at the state level but fosters innovation within private firms, especially those reliant on external financing. Lastly, Francis et al. (2012) proposes an inverse relationship, contending that patenting activities furnish banks with privileged information, thereby cultivating stronger bank-firm connections and facilitating superior loan arrangements.

It is crucial to acknowledge Holmstrom's (1989) observation that corporate innovation is an endeavor fraught with a considerable likelihood of failure owing to its reliance on a multitude of unpredictable factors. Given compelling evidence from literature, it becomes clear that financially constrained firms tend to exhibit a subdued proclivity for innovation, driven by their desire to circumvent the risks associated with innovation failures (Savignac, 2008; Zhao & Zhang, 2023). Following this logic, the presence of executives with a banking background within corporations could potentially serve as a conduit for alleviating these financial constraints, fostering innovative activities. However, my study has provided different evidence on this subject—bank backgrounds of executives lead to a decrease in corporate innovation activities. This can be attributed to the fact that bank-firm connections may lead to a reliance on established financial relationships, discouraging organizations from risk-taking. The emphasis on stability and risk mitigation can impede the willingness to take calculated risks and explore unconventional avenues for innovation.

In the end, although these previous studies have delved deep into the market structure and corporate governance in relation to bank competition and innovation, only a few studies have examined how private or personal connections of firm executives and banks could affect corporate innovation. For example, Liu et al. (2020) presents a very similar study to this paper, examining the effect of a financial background on corporate innovation with data on Chinese firms. However, that paper focuses on a broader range of backgrounds, including investment banking, venture capital, or management consulting. Additionally, problems with endogeneity have yet to be fully resolved despite the various specifications employed by the authors. However, this study provides early evidence of the detrimental effect of executives with a general financial background on corporate innovation. Hence, this paper fills a literature gap, provides a new methodology, and offers valuable insight into another imperative channel through which banks could affect firm innovation and long-term progress.

At last, this study also relates to a body of literature that combines machine learning algorithms with causal inference. While the mainstream literature in this domain centers on the acquisition of previously unavailable economic data or factor analysis (Kang et al., 2013; Hoberg & Phillips, 2016; Donaldson & Storeygard, 2016; Leippold et al., 2022; Obaid & Pukthuanthong, 2022, among others), only a limited subset of studies delve into the application of machine learning methodologies to mitigate the problems posed by endogeneity and confounding variables. Lee et al. (2010) introduces a novel approach employing machine learning algorithms to estimate propensity scores. Athey and Imbens (2016) employ sample partitioning to attain a valid confidence interval for estimating treatment effects. Furthermore, Farrell et al. (2019) investigate the application of deep neural networks to enable semi-parametric inference, encompassing treatment and decomposition effects. Nonetheless, these methods often rely on rigorous assumptions such as unconfoundedness that are either infeasible to fulfill or challenging to apply. Thus, this study aims to present a straightforward framework that enables researchers to address the issue of endogeneity within the context of a binary independent variable across multiple time periods.

III. Data

Given the substantial role personal connections play in bank loaning, China's socio-political context offers an avenue for exploring the relationship between corporate innovation and banking background. To investigate the influence of possessing a banking background on corporate innovation, I obtain data from the China Stock Market & Accounting Research (CSMAR) Database, which provides information regarding publicly listed firms, including asset structure, quarterly performance, and executive backgrounds. It is important to note that the scope of my investigation is limited to private-owned firms, thereby excluding state-owned enterprises from the analysis, as the latter often exhibit different characteristics. Additionally, financial sector firms are excluded following the approach of Hirshleifer et al. (2012), and data is winsorized at the 1st and 99th percentiles to mitigate extreme values. Upon processing the data and excluding observations with missing innovation measures and banking background indicator, the final dataset contains 20,532 observations, encompassing 74 industries and spanning over 2008 to 2020. This newly recorded data allows me to provide valuable current information for policymakers and corporate managers. Definitions of all the variables can be found in Table I.

A. Banking Background Measures

The CSMAR database contains details on corporate executives' financial backgrounds for individual firms, including Regulatory Authorities, Securities Companies, Investment Management Companies, Commercial Banks, and more. Upon processing this data, I construct a binary variable Bank that assumes a value of 0 when none of the executives possess any experience in the realm of commercial banking. Conversely, it assumes a value of 1 when at least one executive boasts such a background. Additionally, to study the nuance of this relationship, I also construct the variable BankNum that represents the number of corporate executives having commercial banking backgrounds. These two variables are sufficient to capture the variations in banking backgrounds and allow me to conduct further investigations.

B. Corporate Innovation Measures

I implement several variables that measure various dimensions of a firm's capability for innovation, drawing upon established literature. Firstly, to assess a firm's commitment to innovative projects, I utilize the metric of corporate expenditure dedicated to research and development (RD). Secondly, the quantity of patents granted ($Patent$) is a direct and tangible proxy for outcomes resulting from innovative activities (Hirshleifer et al., 2012). However, it is important to note that while the number of patents approved offers valuable insights, it fails to capture the precise timing of the innovation process accurately. Hence, I employ the number of patent applications instead, as it better aligns with the actual timing of the development of the patent (Li, 2011). Thirdly, the number of employees ($Personnel$) engaged in research and development activities is a significant indicator of a firm's potential for innovation (Ouimet & Zarutskie, 2014; Chang et al., 2015; Hong et al., 2016).

In addition to these three main proxies, I incorporate the variable $Citations$, which capture the number of times a firm's patents are referenced on an annual basis (Trajtenberg, 1990; Albert et al., 1991; Hall, Jaffe and Trajtenberg, 2001, 2005). Self-citations are excluded from this analysis to ensure robustness. Chang et al. (2015) points out that this proxy may be prone to truncation bias, as patents developed long ago are afforded more time to accumulate citations. In order to mitigate this bias, I created a two-year gap by excluding data from 2021 and 2022 to allow citations to accumulate and adopt a fixed-effect approach, which involved scaling the citations within a given year by the mean citation counts across all industries during that same year. Furthermore, I construct the variable $PatShock$, denoting patent shock, which expresses patents in relation to the firm's total assets, measured in millions of RMB, following the methodology proposed by Fang et al. (2017). $PatShock$ offers insights into the magnitude of a firm's patent activity relative to its overall financial standing, thereby capturing the relative importance of innovation within the company's framework. At last, I formulate the variable $RDPerson$, which signifies the R&D expenditure per employee involved in research projects. This measure indicates the resource allocation and investment intensity dedicated to each individual engaged in corporate innovation.

C. Control Variables and Features

To ensure the robustness of my analysis and isolate the exogenous variation of banking background, I included an array of control variables including firm characteristics and macro fluctuations from the CSMAR database. Following Hall and Ziedonis (2001)'s argument that capital-intensive firms generate more patents, I control for total assets (TA) and fixed assets (FA) (Luo et al., 2011; Pan & Tian, 2015). I also include return on assets (ROA) to capture the profitability of operation. Operating revenue growth (ORG) and market-to-book ratio (MBR) are also included to capture growth opportunities of a firm. To account for the life cycle of firms, I also control for enterprise age (Age). Macro level factors consumer price index (CPI) and gross domestic product (GDP) are included to control for fluctuations across years. Tobin-Q accounts for investment opportunities and Herfindahl-Hirschman Index (HHI) accounts for market concentration (Atanassov, 2013; Chemmanur & Tian, 2011). At last, I include the total number of executives for each firm ($ExeNum$), which accounts for corporate governance.

The above variables are also utilized to train the machine learning models that are used to predict the counterfactual innovation outcome if the firm does not possess executives with banking backgrounds. However, to demonstrate that these features are sufficient to capture the variation in innovation, I also employ additional variables on firm characteristics to ensure that the firm's current situation is captured. These 50+ variables are documented in the Online Appendix.

D. Measurements of Market Competition and Financial Constraints

To study the effect of market competition on the relationship between banking background and corporate innovation. I rely on two indices of market concentration. The concentration ratio index is calculated as follows,

$$CR = \left(\sum^{n} TOR_i\right) / TOR \tag{3.1}$$

where TOR_i is the Total Operating Revenue of top n firms, while TOR presents the entire industry. Following this formula, $CR4$ represents the top four firms' concentration ratio of the industry, and $CR8$ represents the top eight firms' concentration ratio of the industry.

Additionally, two measures of financial constraints are adopted. Following Kaplan and Zingales (1997) and Wei et al. (2014), the construction of the KZ index for each firm is through estimating the regression (3.2):

$$\begin{aligned} KZ_t =& \beta_1 \cdot \frac{CF_t}{TA_{t-1}} + \beta_2 \cdot \text{Lev}_t + \beta_3 \cdot \frac{D_t}{TA_{t-1}} \\ &+ \beta_4 \cdot \frac{CASH_t}{TA_{t-1}} + \beta_5 \cdot \text{TobinQ}_t \end{aligned} \tag{3.2}$$

where CF_t is cashflow at year $t; TA_{t-1}$ is total assets in the previous year; DIV_t is the cash dividend of year t, Lev_t is the ratio of long-term liabilities to assets, and CASH_t is the cash holding of year t.

Referencing Whited and Wu (2006), the WW Index is constructed as follows:

$$\begin{aligned} WW_{i,t} = & -0.091 \cdot CF_{i,t} - 0.062 \cdot DivD_{i,t} + \\ & 0.021 \cdot Lev_{i,t} - 0.044 \cdot TA_{i,t} + 0.102 \cdot ORG_{j,t} - \\ & 0.035 \cdot ORG_{i,t} \end{aligned} \tag{3.3}$$

where $CF_{i.t}$ is the cashflow of firm i in year t; Div $D_{i,t}$ is the dummy variable of cash dividend payment; $Lev_{i,t}$ is the ratio of long-term liabilities to assets; $ORG_{j,t}$ is the industry operation revenue growth and $ORG_{i,t}$ is the firm operation revenue growth.

E. Descriptive Statistics

Table II displays the descriptive statistics for the variables discussed earlier in the preceding sections. This table provides several valuable insights. Firstly, it is revealed by the $Bank$ variable that approximately 23 percent of the firm-year observations have at least one commercial banking background. Secondly, the maximum number of connections between banks and firms in this particular sample is 6, as indicated by the variable $BankNum$. Additionally, the three primary independent variables each exhibit a range from 0 to 19.508, 0 to 4.942, and 0 to 7.898, respectively. Comparatively, these ranges appear skewed when juxtaposed with their respective means. Consequently, log-transformations are applied to these variables to address this asymmetry.

Figure 1 illustrates the distribution of commercial banking backgrounds in various industries. The size of the circle indicates the size of the industry. It is evident that there are indications of a selection bias across different industries, although the bias is not significant as the ratios generally hover around the average of 0.238.

Figure 1 also displays the distribution of commercial banking backgrounds across industries. The x-axis presents the two-digit level industries, while the y-axis presents the ratio of corporations with banking backgrounds over the total number of corporations. The radius of each circle indicates the size of the industries in the sample.

Table III unveils the correlations among key variables. A noteworthy pattern emerges as the correlation between $Bank$ and all three independent variables manifests as negative. This preliminary analysis sets the stage for further exploration in my study. Furthermore, it is crucial to emphasize that $Bank$ exhibits relatively weak correlations with the control variables, relieving concerns of multicollinearity.

IV. Empirical Model

A. Baseline Model

I first present the baseline fixed-effect model for estimating the effect of banking background on innovation,

$$\begin{aligned}\text{Innov}_{i,j,l,t} = \beta_0 + \beta_1 \text{Bank}_{i,j,l,t} + \beta_n X_{i,j,l,t} \\ + \alpha_t + \alpha_j + \varepsilon_{i,k,t}\end{aligned} \tag{4.1}$$

where $\text{Innov}_{i,j,l,t}$ denotes the firm-level innovation indicators including RD, $Patent$, and $Personnel$. $Bank_{i,j,l,t}$ denotes the binary variable of bank-firm connection of firm i in industry j and city l at year t. $X_{i,j,l,t}$ denotes an array of control variables at the firm level and macro level. Details about these variables can be found in Section 3.3.

In order to capture the variations across different industries, industry-level fixed effects (α_j) are included in the analysis. These fixed effects serve to control unobservable factors that are specific to each firm and may have an impact on both corporate innovation and the employment of executives with a commercial banking background. Moreover, to account for the economic conditions and potential macroeconomic shocks, time-fixed effects (α_t) are introduced. These fixed effects help to control changes in corporate innovation over time, ensuring that temporal fluctuations do not influence the estimated effects of interest. An alternative combination of time-fixed effects and city-fixed effects is also tested to validate the consistency of the estimate. Additionally, industry clusters are employed at the single-digit level, allowing for interactions between firms within the same industry cluster. This approach mitigates problems of autocorrelation within industries.

B. ML-DID Framework

The issue of endogeneity poses a significant challenge to identification in the context of the current study. One possible approach to address this problem is to utilize instrumental variables in a two-stage least regression. In their research on the relationship between financial background and firm innovation, Liu et al. (2020) suggest two instruments: the average number of executives with financial backgrounds within a province and the industry's average ratio of executives with financial backgrounds. However, it is important to note that even with these instruments, the problem of endogeneity persists. This is because the provincial mean number of executives may reflect the financial development of the province, and provinces with varying levels of financial development might exhibit different levels of innovation. Similarly, the average ratio of executives with financial backgrounds across industries can influence innovation through channels other than just financial backgrounds. Furthermore, when I tested these instruments with the data from the present study, the results were negative but statistically insignificant. As a result, alternative strategies or additional instruments may be required to improve the identification of causal effects and draw more robust conclusions.

Therefore, I suggest utilizing a machine learning framework that can be applied to diverse scenarios of "Panel Data with Multiple Treatment Timing" to address the problem of endogeneity and yield a more accurate estimation of the impact of commercial banking background. This approach combines machine learning techniques with the conventional difference-in-differences method.

Figure 2 provides a clear visual representation of the proposed methodology. Panel A demonstrates how the intervention or treatment effect (δ_{treat}) is represented and calculated using the conventional differencing approach. This method relies on the assumption that the control group follows a similar trend as the treatment group, thereby canceling out any confounding effects. However, in practical applications, it is often challenging to satisfy this "Parallel Trend Assumption." To address this limitation, the ML-DID method, illustrated in Panel B, utilizes a machine learning model trained solely on pre-treatment data from the treatment group. The model predicts the outcome of post-treatment data based on the pre-treatment information. The difference between the predicted and actual outcomes represents the treatment effect (δ_{treat}). In simple terms, the machine learning model does not know the treatment factor (corporate background) in this context, but it considers all other firm characteristics. Consequently, the prediction also reflects the trend of innovation in the absence of treatment for the treated group. This allows for a comparison of expected outcome values and facilitates the calculation of the treatment effect.

Notes: The figure displays the canonical and ML-DID approach for a standard difference-indifference design. Panel A presents the canonical approach: a treatment group (red line) and a control group (green line), while Panel B presents the ML-DID approach: a treatment group (red line), and a predicted trend (grey line). In both panels, τ denotes the time when intervention is implemented and β_{treat} denotes the treatment effect.

It is crucial to recognize that machine learning models are not infallible in predicting outcomes. The error term $(\varepsilon_{\mathrm{i}})$ that arises from these predictions captures the influence of unobserved factors that the model lacks information about. Nonetheless, if the machine learning model possesses an equivalent predictive capability for both the treatment group and the control group, any confounding effects can be mitigated. Based on this observation, I make the following assumption:

Assumption 1. Equal Predictive Power Assumption

The prediction errors from the machine learning algorithm are constant across the treatment and control groups.

$$\mathrm{E}\left[\varepsilon_{\mathrm{i}1}(0)\right]=\mathrm{E}\left[\varepsilon_{\mathrm{i}1}(1)\right]$$

To examine the validity of this assumption, a straightforward approach is to independently apply the same machine learning model specification to both the treatment group and the control group, utilizing the same dataset. This enables a comparison of the average errors between the two groups using a two-sample t-test. By assessing whether there is a statistically significant difference in the mean errors, researchers can gain insights into the comparability of predictive power between the treatment and control groups.

Before elaborating on the identification strategy, we first construct our predictor with a random forest:

$$T_i(X, Y)=\sum_{j=1}^{M_i} \operatorname{Leaf}\left(j, X, Y, \theta_{i, j}\right) \tag{4.2}$$

where M_i is the number of leaves (terminal nodes) in the i-th decision tree. $\operatorname{Leaf}\left(j, X, \theta_{i, j}\right)$ is a function that evaluates to a value based on the features X, outcome Y, and the parameters $\theta_{i,j}$ associated with leaf j in tree i. Then, we have our random forest estimator $\Psi(X, Y)$ as,

$$\Psi(X, Y)=\frac{1}{N} \sum_{i=1}^N T_i(X, Y) \tag{4.2}$$

Assume $\Psi(T, X)$ is the machine learning algorithm employed, where T is the treatment indicator of whether training data is the control group (0) or treatment group (1), and X is a set of features that are used to predict an outcome. $\mathrm{E}\left[\mathrm{Y}_{\mathrm{i}1}(0) \mid \mathrm{D}=1\right]$ is the actual outcome. Then, the prediction is composed of two parts:

$$\mathrm{E}\left[\Psi_{\mathrm{i}}\left(\mathrm{Y}_{\mathrm{i}0}(0), \mathrm{X}_{\mathrm{i}0}\right) \mid \mathrm{D}=1\right]=\mathrm{E}\left[\mathrm{Y}_{\mathrm{i}1}(0) \mid \mathrm{D}=1\right]+\mathrm{E}\left[\varepsilon_{\mathrm{i}1}(0)\right] \tag{4.3}$$

$$\mathrm{E}\left[\Psi_{\mathrm{i}}\left(\mathrm{Y}_{\mathrm{i}0}(1), \mathrm{X}_{\mathrm{i}0}\right) \mid \mathrm{D}=1\right]=\mathrm{E}\left[\mathrm{Y}_{\mathrm{i}1}(1) \mid \mathrm{D}=1\right]+\mathrm{E}\left[\varepsilon_{\mathrm{i}1}(1)\right] \tag{4.4}$$

This allows us to establish the following equation for ATT:

$$\mathrm{E}\left[\Psi_{\mathrm{i}}\left(\mathrm{Y}_{\mathrm{i}1}(1), \mathrm{X}_{\mathrm{i}1}\right)-\Psi_{\mathrm{i}}\left(\mathrm{X}_{\mathrm{i}0}, \mathrm{Y}_{\mathrm{i}0}(0)\right) \mid \mathrm{D}=1\right] \tag{4.5}$$

Proof:

$$\begin{aligned}
ATT &= E\left[Y_{i1}(1)-Y_{i1}(0) \mid D=1\right] \\
&= E\left[Y_{i1}(1) \mid T=1\right]-E\left[Y_{i1}(0) \mid D=1\right] \\
&= E\left[\Psi_i\left(Y_{i1}(1), X_{i1}\right) \mid D=1\right]-E\left[E_{i1}(1)\right] \\
&- E\left[\Psi_i\left(X_{i0}, Y_{i0}(0)\right) \mid D=1\right]+E\left[E_{i1}(0)\right] \\
&= E\left[\Psi_i\left(Y_{i1}(1), X_{i1}\right) \mid D=1\right] \\
&- E\left[\Psi_i\left(X_{i0}, Y_{i0}(0)\right) \mid D=1\right] \\
&= E[\Psi i(Yi1(1), Xi1)-\Psi i(Xi0, Yi0(0)) \mid D=1]
\end{aligned}$$

Some might argue that in machine learning, using the same data for training and testing can lead to inaccurate error estimates because of the risk of overfitting. However, in this specific scenario, this approach is still valid. The models are designed to generalize and avoid overfitting, and in the case of panel data, when predicting the treatment outcome using a model trained on the control group, the predictions are still applicable to the same firms. The purpose of this test is to show that the model produces similar errors for both the treatment and control groups, allowing for the cancellation of the error. It's important to note that this assumption is different from the parallel trend assumption commonly used in difference-in-differences analysis. This approach acknowledges the presence of selection bias and focuses solely on evaluating the predictive power of the machine learning algorithm. Its goal is to determine whether the features from the treatment and control groups contain similar information for predicting the outcome.

For detailed steps on testing this Equal Error Assumption, please refer to Table IV, which provides researchers with a reliable framework for assessing these considerations. To clarify, steps 5 and 6 are the first and second stages of the ML-DID (Machine Learning Difference-in-Differences) method. In the first stage, the method predicts the outcome for the treatment group using data from the control group. Then, in the second stage, a regression analysis is performed using the predicted outcomes and the actual outcomes of the treatment group.

This study, which focuses on commercial banking and corporate innovation, aims to determine the impact of having a corporate executive with a banking background on a firm's innovation capabilities. The paper introduces two definitions to facilitate the analysis: Bank Years and Non-Bank Years. Bank Years refer to firm-year observations where the variable $Bank$ is equal to one, indicating the presence of a bank-firm connection. In other words, these are the years when the firm has a connection with a bank. On the other hand, Non-Bank Years encompass firm-year observations where the variable $Bank$ is equal to zero. However, to ensure the integrity of our machine learning predictor, all firms without any bank-firm connection throughout the entire sample period are excluded from this category. This approach guarantees that irrelevant firms do not influence the machine learning model in the analysis.

In this study, I use a generic predictor and features representing firm characteristics and macro conditions. The predictor is trained using data from Bank Years, taking into account the fluctuations that occur when firms establish bank-firm relationships. This trained predictor is then capable of predicting firm innovations, such as patents and R&D expenses, without specific knowledge of bank-firm relationships. Following the approach of Hirshleifer, Low, and Teoh (2012), I exclude all firms that did not file any patents between 2008 and 2020. To ensure the robustness of the ML-DID approach, I conduct tests with various parameters and features. The appendix table provides detailed information on these features. Furthermore, I define a reduced version of the estimator, trained on firm characteristics and macro indicators $(X_{i,j,l,t})$, which are used in equation (4.1) for the OLS estimator. Additional features, such as year, industry, and city dummies, are also included in this reduced version of the estimator.

To explore the consistency of ML-DID, I tested various parameters and features. I defined the reduced version of the estimator as trained on the firm characteristics and macro indicators used in equation (4.1) for the OLS estimator. The full version of the estimator includes 50+ features, encompassing various characteristics of the firm. Details of these features can be found in the appendix table. The main challenge lies in selecting the appropriate machine learning models. Two popular models are tested: deep neural networks and random forest. However, upon pre-analysis, the results from deep neural networks did not pass the Equal Error Assumption test. Therefore, random forest (RF) is utilized for estimation. Additionally, the issue of extreme values can severely impact the validity of the estimates. To address this concern, Huber weights (Huber, 1964) are used in the second-stage estimate. This helps mitigate the influence of outliers and ensures more robust results.

C. OLS Results

The results from equation (4.1) are presented in Table V. This section focuses on the main proxies for corporate innovation: R&D expenses, patent applications, and research personnel. In columns (1) through (3), the estimates for R&D expenses are reported. The coefficients are all negative and statistically significant. This suggests that having a corporate executive with a banking background has a detrimental effect on a firm's expenses in research and development. Moving on to columns (4) through (6), the estimates for patent applications are presented. Similar to R&D expenses, the coefficients are negative and statistically significant at the 1 percent level. This indicates that a banking background reduces the number of patent applications, which is a measure of innovation output. Similarly, columns (7) through (9) show the estimates for employees engaged in research and development. The negative and significant coefficients demonstrate that a bank-firm connection can also impact the employment levels within a firm's research departments.

The finding of this study aligns with past literature. Yang et al. (2021) found that financial expert CEOs have a significant negative impact on firm innovation, and Liu et al. (2020) found that the financial background of executive officers has detrimental effects on innovation. It is also worth noting that the inclusion of fixed effects has resulted in significant changes in the estimates. This suggests the possibility of endogeneity, indicating the need for further investigation and consideration of potential confounding factors. Further analysis and robustness checks are necessary to establish a causal relationship between a corporate executive's banking background and a firm's innovation outcomes.

Notes: Columns (1) through (3) reports estimate with research and development expense as the dependent variable; Columns (4) through (6) reports estimate with annual patent applied as the dependent variable; Columns (7) through (9) reports estimate with annual number of employees in research field as the dependent variable. Columns (2), (5), and (8) include controls, year, and industry fixed effects, while Columns (3), (6), and (9) include controls, year, and city fixed effects. Single-letter industry cluster standard errors are applied to all columns except (1), (4), and (7). Columns (1), (4), and (7) employ heteroskedasticity-robust standard errors. All continuous variables are winsorized at the first and 99th percentiles. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed.

D. ML-DID Results

Table VI presents the estimates from ML-DID for patent applications, using different architectures represented as Architecture 1 through 10, which correspond to various parameter combinations. Two essential parameters of random forest are the estimators and max depth. The former represents the number of trees in the forest, while the latter represents the max depth each branch of the trees can expand to. By adjusting these two parameters, I can prevent the random forest algorithm from overfitting or underfitting.

Panel A focuses on the estimates obtained using reduced features to train the random forest model. Remarkably, there is little variation across all ten architectures, indicating that the results from ML-DID are robust and not heavily influenced by different parameter settings. The t-statistics for the first stage are all insignificant, suggesting that the errors are similar between the treatment and control groups. Moving on to Panel B, the estimates are derived using full features to train the model. Although the estimates are slightly higher compared to those from the reduced version, the 95 percent confidence intervals indicate that these estimates are still within an acceptable range. Furthermore, when comparing the estimates from ML-DID to those in Table V (which include industry and city fixed effects) that are approximately -0.0985 and -0.1245, it can be observed that the estimates from ML-DID, around -0.27 and -0.29, are reasonably higher. This demonstrates that endogeneity indeed introduces bias into the OLS results of the relationship between commercial banking background and corporate innovation.

To visualize the predictions, I present Figure 2 which showcases the predicted outcome and actual outcome for both the reduced and full versions. In Panel A and Panel C, a notable issue arises with extreme values, where the actual values are significantly larger compared to the predicted values at the extreme end. Huber regression is employed to address this issue, which tends to be more efficient and valid than ordinary least squares regression in such contexts.

Moreover, the estimated values derived from the predictions generally show higher figures overall compared to the actual values. This implies that companies led by executives with a background in commercial banking tend to have fewer patent applications compared to firms whose executives lack such a background. On the other hand, Panel B and Panel D present a clear depiction of the estimated values for the initial 100 observations. The predicted and actual values display similar patterns, although the predicted values remain slightly higher than the actual ones.

Table VII displays the estimates for R&D and Personnel. All of these estimates are negative and statistically significant, meeting the requirement of the Equal Error Assumption. The coefficients for R&D are quite similar to those obtained from the fixed effect regression in Table IV. It is important to note that ML-DID excludes all the observations from the control group (firms with no bank-firm connections throughout the entire sample period), and still yields consistent estimates as ones from OLS, suggesting that the estimates are reliable. Both the reduced and full models produce estimates of similar magnitude, as indicated by the confidence intervals.

Finally, an important advantage of machine learning methods is the estimation of individual treatment effects—that is, researchers could identify how the treatment effect is for each individual firm-year observation. Figure 4 demonstrates such individual treatment effects for Patent. Due to the large number of observations, only the first 200 observations are shown in the graph. Blue dots represent the actual values for Patent for a firm that possesses executives with commercial banking backgrounds, while orange dots represent the predicted values. Two dots with a grey line connecting them are from the same firm-year observation. As shown by the box plot, the median value of the predicted is higher than the median value of the actual values, indicating that banking background has a negative effect on patent applications. A similar trend is observed for R&D expenses and research personnel, which can be found in the Online Appendix Figure 2.

V. Robustness Tests

I perform a series of rigorous tests to assess the robustness of the baseline results. The objective is to minimize the impact of endogeneity by employing traditional econometric methods such as substituting dependent variables and propensity score matching. The results reveal that companies led by executives with a banking background continue to exhibit a lower rate of innovation. These findings strengthen the validity of the initial results and provide additional evidence for the negative relationship between a corporate executive's banking background and firm innovation outcomes.

A. Citation, R&D per Person, and Patent Shocks

To further validate the baseline findings, I incorporate additional dependent variables that serve as indicators for corporate innovation. By testing the results using different specifications, as presented in Table VIII, I observe that all coefficients are statistically significant at a 1 percent level. This indicates that our model and results remain robust across various model specifications.

It is also important to highlight that, like how a banking background can decrease patent applications, it can also lead to a reduction in citations. This suggests that a corporate executive's banking background may have a broader impact on various aspects of firm innovation, influencing not only the quantity of innovation output but also its recognition and acknowledgment in the academic and research community. These additional findings further support the negative relationship between a banking background and corporate innovation.

B. Propensity Score Matching

To address the potential impact of self-selection bias on the regression results, I employ propensity score matching using two specifications with nearest neighbor matching. The propensity scores are calculated by estimating first-stage probit regressions that include all relevant controls and fixed effects.

In columns (1), (3), and (5) of the analysis, I present the results with three matches per observation, while columns (2), (4), and (6) display estimates with six matches per observation. Despite these variations, the results remain statistically significant at a 1 percent level. This suggests that the relationship between a banking background and corporate innovation is not solely driven by endogenous factors. The propensity score matching technique helps mitigate potential self-selection bias and provides additional support to the robustness of the findings regarding the negative association between a corporate executive's banking background and firm innovation outcomes.

VI. Additional Analysis

To further study the underlying factors that contribute to the relationship between commercial banking background and corporate innovations, I partition my sample in two different ways: by industrial competition and by financial constraints.

A. Industrial Competition on Banking Background and Innovation's Relationship

The relationship between corporate executives' banking background and corporate innovation may be influenced by the level of industrial competition. In industries with higher competition, there is often a greater emphasis on innovation to gain a competitive advantage. Aghion (2005) argues that competition and innovation have an inverse-U relationship, where lower levels of competition are associated with lower levels of innovation. Based on this, I propose the hypothesis that the negative impact of corporate executives' banking background on innovation will be more pronounced in industries with lower levels of competition.

Table X displays the results of the role of industrial competition on the relationship between banking background and corporate innovation. Significantly larger coefficients are found among those less competitive industries with low concentration ratios.

B. Financial Constraint on Banking Background and Innovation's Relationship

Financial constraints can also have a strong impact on the relationship between innovation and banking background. Firms with severe financial constraints might experience a greater reduction in innovation compared to those without such constraints. Table XI presents the results. When comparing the lower 25 percent of firms with less financial constraints to those with strong financial constraints, the latter group shows a statistically significant negative coefficient, indicating a more pronounced detrimental effect of a banking background on innovation for financially constrained firms. This finding suggests that financial constraints play a crucial role in influencing the relationship between a corporate executive's banking background and firm innovation outcomes.

C. Number of Banking Backgrounds on Corporate Innovation

Table XII presents regressions with $BankNum_{i,j,l,t}$ as the explanatory variable, which indicates the number of executives with a banking background possessed by a firm. All coefficients are negative and significant, indicating that as the number of executives with a banking background increase, there is a greater reduction in corporate innovation. This finding suggests that having a higher proportion of executives with banking backgrounds in a firm is associated with a more pronounced negative impact on innovation outcomes. Thus, the presence of multiple executives with banking backgrounds may exacerbate the detrimental effect on corporate innovation.

VII. Conclusion

This paper delves into the relationship between corporate executives' commercial banking background and corporate innovation and uncovers a statistically significant negative relationship, indicating that firms led by executives with a banking background exhibit reduced innovation outcomes.

One of the critical contributions of this paper lies in its focus on the nuance of the banking background on corporate innovation, specifically emphasizing commercial banking backgrounds. By delving into this aspect, the study sheds light on how low industrial competition and high financial constraints can strengthen the observed relationship, providing valuable insights primarily overlooked in the existing literature on executive backgrounds. Moreover, the paper proposes and develops a novel machine learning framework akin to the difference-in-differences approach to tackle endogeneity issues and omitted variable bias, enhancing the credibility of the findings. Despite its contributions, the study acknowledges certain limitations. One notable limitation is the need for a comprehensive theoretical foundation for the ML-DID method used in the analysis. This points to an avenue for further research and refinement of the proposed framework.

The findings have implications for policymakers and corporate governance practices. Understanding the effects of corporate executives' banking background on innovation can offer valuable guidance to policymakers seeking to foster and promote innovation within firms. The insights from this study can inform policy decisions aimed at encouraging a diverse set of executive backgrounds, which may positively impact innovation and overall firm performance.

References

[1] Acharya, V., Lambrecht, B.M., 2015. A Theory of Income Smoothing When Insiders Know More than outsiders. *Review of Financial Studies* 28, 2534–2574

[2] Acharya, Viral, and Zhaoxia Xu. "Financial Dependence and Innovation: The Case of Public Versus Private Firms." *Journal of Financial Economics* 124.2 (2017): 223-243.

[3] Aghion, Philippe, et al. "Competition and innovation: An inverted-U relationship." *The Quarterly Journal of Economics* 120.2 (2005): 701-728.

[4] Albert, M., Avery, D., Narin, F., McAllister, P., 1991. Direct Validation of Citation Counts as Indicators of Industrially Important Patents. *Research Policy* 20 (3), 251–259. http://dx.doi.org/10.1016/

[5] Aslan, Hadiye. "Do Lending Relationships Affect Corporate Financial Policies?" *Financial Management*, vol. 45, no. 1, 2016, pp. 141–73.

[6] Athey, Susan, and Guido Imbens. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113.27 (2016): 7353-7360.

[7] Bamber, Linda Smith and Jiang, John (Xuefeng) and Wang, Isabel, What's My Style? The Influence of Top Managers on Voluntary Corporate Financial Disclosure (June 1, 2008). *Accounting Review*, Vol. 85, No. 4, pp. 1131-1162, July 2010, Available at SSRN: https://ssrn.com/abstract=1312904

[8] Barth, James R., Gerard Caprio, and Ross Levine. *Rethinking Bank Regulation: Till Angels Govern*. Cambridge University Press, 2008.

[9] Benfratello, Luigi, Fabio Schiantarelli, and Alessandro Sembenelli. "Banks and Innovation: Microeconometric Evidence on Italian Firms." *Journal of Financial Economics* 90.2 (2008): 197-217.

[10] Berlin, M., Mester, L.J., 1999. Deposits and Relationship Lending. *The Review of Financial Studies* 12, 579–607.

[11] Besanko, D., and G. Kanatas. 1993. Credit Market Equilibrium with Bank Monitoring and Moral Hazard. *Review of Financial Studies* 6(1):213-32.

[12] Chang, Xin, Kangkang Fu, Angie Low, Wenrui Zhang, Non-executive Employee Stock Options and Corporate Innovation, *Journal of Financial Economics*, Volume 115, Issue 1, 2015, Pages 168-188.

[13] Cornaggia, Jess, et al. "Does Banking Competition Affect Innovation?." *Journal of Financial Economics* 115.1 (2015): 189-209.

[14] Dass, Nishant, and Massimo Massa. "The Impact of a Strong Bank-Firm Relationship on the Borrowing Firm." *The Review of Financial Studies*, vol. 24, no. 4, 2011, pp. 1204–60.

[15] Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction." *American Economic Review*, 109 (12): 4178-4219.

[16] Donaldson, Dave, and Adam Storeygard. "The View From Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30.4 (2016): 171-198.

[17] Dwenger, Nadja, Frank M. Fossen, and Martin Simmler. "Firms' Financial and Real Responses to Credit Supply Shocks: Evidence From Firm-Bank Relationships in Germany." *Journal of Financial Intermediation* 41 (2020): 100773.

[18] Fama, E.F., 1985. What's different About Banks? *Journal of Monetary Economics* 15, 29–39.

[19] Farrell, Max H., Tengyuan Liang, and Sanjog Misra. "Deep Neural Networks for Estimation and Inference." *Econometrica* 89.1 (2021): 181-213.

[20] Francis, Bill, et al. "Do Banks Value Innovation? Evidence from US Firms." *Financial Management* 41.1 (2012): 159-185.

[21] Gao, Wenlian. "Banks as Lenders and Shareholders: Evidence from Japan." *Pacific-Basin Finance Journal* 16.4 (2008): 389-410.

[22] Gorton, G., Schmid, F., 2000. Universal Banking and the Performance of German Firms. *Journal of Financial Economics* 58, 29–80.

[23] Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The NBER Patent and Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper no. 8498

[24] Hoberg, Gerard, and Gordon Phillips. "Text-based Network Industries and Endogenous Product Differentiation." *Journal of Political Economy* 124.5 (2016): 1423-1465.

[25] Holmstrom, B., 1989. Agency Costs and Innovation. *Journal of Economic Behavior and Organization* 12, 305–327.

[26] Hong, Jin, et al. "Do Government Grants Promote Innovation Efficiency in China's High-Tech Industries?." *Technovation* 57 (2016): 4-13.

[27] Huber, P. J. 1964. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 35: 73–101. https://doi.org/10.1214/aoms/1177703732.

[28] James, C. 1987. Some Evidence on the Uniqueness of Bank Loans. *Journal of Financial Economics* 19(2):217-36,

[29] Kang, Jun Seok, Polina Kuznetsova, Michael Luca, and Yejin Choi. 2013. "Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews." EMNLP 2013: 2013 *Conference on Empirical Methods in Natural Language*

[30] Kaplan, Steven N., and Luigi Zingales. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?." The Quarterly Journal of Economics 112.1 (1997): 169-215.

[31] Lee, Brian K., Justin Lessler, and Elizabeth A. Stuart. "Improving Propensity Score Weighting Using Machine Learning." *Statistics in Medicine* 29.3 (2010): 337-346.

[32] Leippold, Markus, Qian Wang, and Wenyu Zhou. "Machine Learning in the Chinese Stock Market." *Journal of Financial Economics* 145.2 (2022): 64-82.

[33] Levine, R. 2002. Bank-Based or Market-Based Financial Systems: Which Is Better? *Journal of Financial Intermediation* 11(4):398-428.

[34] Li, X., 2011. Sources of External Technology, Absorptive Capacity, and Innovation Capability in Chinese State-Owned High-Tech Enterprises. *World Dev.* 39 (7), 1240–1248.

[35] Lin, Xiaochi, Yi Zhang, and Ning Zhu. "Does Bank Ownership Increase Firm Value? Evidence from China." *Journal of International Money and Finance* 28.4 (2009): 720-737.

[36] Liu, Baohua, et al. "Corporate Executives with Financial Backgrounds: The Crowding-out Effect on Innovation Investment and Outcomes." *Journal of Business Research* 109 (2020): 161-173.

[37] Luo, Wei, Yi Zhang, and Ning Zhu. "Bank Ownership and Executive Perquisites: New Evidence From an Emerging Market." *Journal of Corporate Finance* 17.2 (2011): 352-370.

[38] Mullainathan, Sendhil, and Jann Spiess. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31.2 (2017): 87-106

[39] Muthuri, Judy N., Jeremy Moon, and Uwafiokun Idemudia. "Corporate Innovation and Sustainable Community Development in Developing Countries." *Business & Society* 51.3 (2012): 355-381.

[40] Obaid, Khaled, and Kuntara Pukthuanthong. "A Picture is Worth a Thousand Words: Measuring Investor Sentiment by Combining Machine Learning and Photos from News." *Journal of Financial Economics* 144.1 (2022): 273-297.

[41] Ouimet, Paige, and Rebecca Zarutskie. "Who Works for Startups? The Relation Between Firm Age, Employee Age, and Growth." *Journal of financial Economics* 112.3 (2014): 386-407.

[42] Pan, Xiaofei, and Gary Gang Tian. "Political Connections and Corporate Investments: Evidence from the Recent Anti-Corruption Campaign in China." *Journal of Banking & Finance* 119 (2020): 105108.

[43] Robinson, James C. *The Corporate Practice of Medicine: Competition and Innovation in Health Care*. Vol. 1. *Univ of California Press*, 1999.

[44] Savignac, Frédérique. "Impact of Financial Constraints on Innovation: What Can be Learned from a Direct Measure?" *Econ. Innov. New Techn.* 17.6 (2008): 553-569.

[45] Sevilir, M., Tian, X., 2013. Acquiring Innovation. Unpublished working Paper. *Social Science Research Network*

[46] Solow, R., 1957. Technological Change and the Aggregate Production Function. *Review of Economics and Statistics* 39, 312–320.

[47] Tan, Y., Tian, X., Zhang, X., Zhao, H., 2015. The Real Effects of Privatizations: Evidence from China's Split Share Structure Reform. *University of Otago Working Paper*.

[48] Trajtenberg, M., 1990. Economic Analysis of Product Innovation: The Case of CT Scanners vol. 160. Harvard University Press, MA.

[49] Wei, Zhihua, et al. "Minimum Wage Effects on Firms' R&D investment: Evidence from China." *International Review of Economics & Finance* 87 (2023): 287-305.

[50] Whited, Toni M., and Guojun Wu. "Financial Constraints Risk." *The Review of Financial Studies* 19.2 (2006): 531-559.

[51] Yang, Chenlin, et al. "CEO Financial Career and Corporate Innovation: Evidence from China." *International Review of Economics & Finance* 74 (2021): 81-102.

[52] Yildirim, Alev. "The Effect of Relationship Banking on Firm Efficiency and Default Risk." *Journal of Corporate Finance* 65 (2020): 101500.

[53] Zhao, Meiyu, and Bo Zhang. "Credit Corruption, Financial Constraint and Corporate Innovation: Evidence on China." *Finance Research Letters* 53 (2023): 103658.

VIII. Appendix

Table 1- Variable Definitions

Dependent Variables	
RD	Annual research and development expenses in Yuan [Log transformed]
$Patent$	Annual patent applied [Log transformed]
$PatShock$	Annual patent applied divided by total assets in millions of Yuan [Fraction]
$Citations$	Annual citations divided by mean citation counts across all industries [Fraction]
$Personnel$	Total employees engage in research and development [Log transformed]
$RDPerson$	Annual research and development per employee in Yuan [Fraction]
Independent Variable	
$Bank$	1 if any corporate executive has banking background; 0 if corporate executive has no banking background [Binary]
$BankNum$	Number of corporate executives have banking backgrounds [Log transformed]
Control Variables	
ROA	Annual rate of return on assets [Fraction]
ORG	Operating revenue growth from $t-1$ year to t year
TA	Firm total assets [Log transformed]
CPI	Annual consumer price index [Log transformed]
GDP	Annual growth domestic product [Log transformed]
$Tobin-Q$	Market Value divided by Total Assets [Fraction]
FA	Firm fixed assets [Log transformed]
MBR	Market to book ratio [Fraction]
Age	Years since the date of establishment
HHI	Herfindahl-Hirschman Index that measure of market concentration
CF	Annual cash flow
TFA	Annual total financial assets held [Log transformed]
Lev	Financial Leverage Ratio
$ExeNum$	Total number of executives of a firm.
Other Variables	
$CR4$	Top 4 firms' concentration within the affiliated industry [Ratio]
$CR8$	Top 8 firms' concentration within the affiliated industry [Ratio]
KZ	Measurement of financial constraint. (Kaplan and Zingales, 1997)
WW	Measurement of financial constraint. (Whited and Wu, 2006)

TABLE I: Variable Definitions

Table 2- Descriptive Statistics

Variables	Obs.	Mean	SD	Min	Max	1%	99%
Dependent Variables							
Bank	20,532	0.238	0.426	0	1	0	1
BankNum	20,532	0.289	0.576	0	6	0	2
Independent Variables							
RD	20,532	4.085	6.545	0	19.477	0	19.31
Patent	20,532	7.275	17.424	0	134	0	104
Personnel	20,532	2.884	2.784	0	7.876	0	7.876
PatShock	20,532	0.273	0.635	0	13.624	0	3.068
CitationScaled	20,532	1.004	5.709	0	384.355	0	14.631
RDPerson	20,532	2.632	5.218	0	21.278	0	17.077
Control Variables							
ROA	20,532	0.045	0.057	-0.196	0.188	-0.196	0.188
ORG	20,532	0.217	0.58	-0.596	4.1	-0.596	4.1
TA	20,532	21.655	1.126	19.114	26.668	19.114	24.939
CPI	20,532	101.143	8.021	84.276	111.386	84.276	111.386
GDP	20,532	13.472	0.323	12.674	13.829	12.674	13.829
$Tobin-Q$	20,532	2.211	1.391	0.885	8.534	0.926	8.534
FA	20,532	19.544	1.575	14.926	24.592	14.926	23.192
MBR	20,532	4.582	4.003	0.632	28.433	0.79	28.433
Age	20,532	16.197	5.863	1	62	4	30
HHI	20,532	0.198	0.121	0.081	0.846	0.081	0.759
CF	20,532	0.009	0.83	-24.993	8.294	-2.208	0.634
TFA	20,532	12.29	8.422	0	25.338	0	22.113
Lev	20,532	0.456	0.206	0.071	1.211	0.071	1.211
Other Variables							
CR4	20,532	0.534	0.191	0	0.986	0	0.986
CR8	20,532	0.632	0.214	0	1	0	1
KZ	20,532	0.967	2.701	-15.297	17.195	-7.456	7.644
WW	20,532	-1.057	3.698	-522.008	-0.108	-1.181	-0.818

TABLE II: Descriptive Statistics

Table 3- Correlations of Selected Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) *Bank*	1.000												
(2) *RD*	-0.025	1.000											
(3) *Patent*	-0.023	0.127	1.000										
(4) *Personnel*	-0.078	0.229	0.245	1.000									
(5) *ROA*	-0.032	-0.043	0.093	0.036	1.000								
(6) *ORG*	0.018	-0.008	-0.007	-0.021	0.168	1.000							
(7) *TA*	0.037	0.130	0.232	0.277	-0.052	0.040	1.000						
(8) *GDP*	-0.043	0.186	0.125	0.683	-0.090	-0.064	0.278	1.000					
(9) $Tobin-Q$	0.020	0.015	-0.012	-0.020	0.113	0.038	-0.353	-0.041	1.000				
(10) *MBR*	0.034	-0.008	-0.008	-0.031	0.116	0.147	-0.348	-0.046	0.681	1.000			
(11) *Age*	-0.003	0.042	-0.007	0.216	-0.129	-0.028	0.197	0.417	0.047	0.011	1.000		
(12) *HHI*	0.035	-0.066	-0.081	-0.072	-0.048	-0.000	0.044	0.002	0.006	0.032	0.017	1.000	
(13) *CF*	-0.060	-0.021	-0.062	-0.020	0.022	-0.153	-0.345	-0.028	0.087	0.047	-0.043	-0.020	1.000

TABLE III: Correlations of Selected Variables

Step 1.	*Define the machine learning specifications*
Step 2.	*Use the same specifications and train the data on both the control group and treatment group separately.*
Step 3.	*Predict with the model separately using the training set data for control and treatment groups.*
Step 4.	*Conduct 2 sample t-tests to estimate if the expected values of the errors from the predictions are the same.*
Step 5.	*[First stage ML-DID] Predict the treatment outcome with the model trained with control group data.*
Step 6.	*[Second stage ML-DID] Run regressions comparing the predicted outcome and true outcome for the treatment group data.*

TABLE IV: Test of Equal Error Assumption and ML-DID Procedure

	$RD_{i,j,l,t}$			$Patent_{i,j,l,t}$			$Personnel_{i,j,l,t}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Bank_{i,j,l,t}$	-0.4302	-0.4704	-0.3397	-0.1447	-0.0985	-0.1245	-0.2771	-0.1038	-0.1146
	(0.1052)	(0.0898)	(0.0850)	(0.0193)	(0.0122)	(0.0132)	(0.0322)	(0.0219)	(0.0254)
Observations	20,532	20,532	20,532	20,532	20,532	20,532	20,532	20,532	20,532
Cities	338	338	338	338	338	338	338	338	338
Industries	74	74	74	74	74	74	74	74	74
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	N	Y	Y	N	Y	Y	N	Y	Y
Industry FE	N	Y	N	N	Y	N	N	Y	N
City FE	N	N	Y	N	N	Y	N	N	Y
Industry Clusters	N	Y	Y	N	Y	Y	N	Y	Y
R^2	0.0621	0.1868	0.2092	0.1373	0.2550	0.2489	0.5142	0.7243	0.7045

TABLE V: ML-DID Estimates with Random Forest for Patent

$Patent_{i,j,l,t}$						
Panel A. Reduced Features						
Architecture	method	n estimators	max depth	Estimate	Assumption t-statistic	CI
1	RF	500	None	-0.2665 (0.0147)	0.5317	[-0.2953, -0.2376]
2	RF	300	40	-0.2660 (0.0147)	0.6194	[-0.2949, -0.2371]
3	RF	300	50	-0.2669 (0.0148)	0.7030	[-0.2960, -0.2379]
4	RF	300	60	-0.2668 (0.0148)	0.7518	[-0.2958, -0.2377]
5	RF	500	40	-0.2656 (0.0147)	0.4568	[-0.2943, -0.2368]
6	RF	500	50	-0.2646 (0.0147)	0.4665	[-0.2935, -0.2358]
7	RF	500	60	-0.2655 (0.0147)	0.4789	[-0.2943, -0.2366]
8	RF	700	40	-0.2668 (0.0146)	0.3178	[-0.2955, -0.2382]
9	RF	700	50	-0.2660 (0.0147)	0.3405	[-0.2948, -0.2373]
10	RF	700	60	-0.2659 (0.0147)	0.3355	[-0.2947, -0.2371]
Panel B. Full Features						
Architecture	method	n estimators	max depth	Estimate	Assumption t-statistic	CI
1	RF	500	None	-0.2933 (0.0139)	0.0625	[-0.3206, -0.2660]
2	RF	300	40	-0.2846 (0.0140)	0.0119	[-0.3121, -0.2571]
3	RF	300	50	-0.2891 (0.0140)	0.1369	[-0.3166, -0.2615]
4	RF	300	60	-0.2888 (0.0140)	0.1217	[-0.3163, -0.2613]
5	RF	500	40	-0.2905 (0.0139)	-0.0310	[-0.3178, -0.2633]
6	RF	500	50	-0.2933 (0.0139)	0.0754	[-0.3207, -0.2660]
7	RF	500	60	-0.2933 (0.0139)	0.0611	[-0.3207, -0.2660]
8	RF	700	40	-0.2917 (0.0139)	-0.0699	[-0.3189, -0.2645]
9	RF	700	50	-0.2929 (0.0139)	0.0390	[-0.3202, -0.2657]
10	RF	700	60	-0.2925 (0.0139)	0.0266	[-0.3198, -0.2653]

Notes: Panel A reports estimates with a reduced list of features as training data, while Panel B reports estimates with full list of features. Equal Error Assumption t-statistic and 95 percent confidence interval are reported. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed. Industry clustered standard errors in brackets. Estimates are from Huber Regressions.

TABLE VI: Effect of Bank-Firm Connection on Corporate Innovation

	$RD_{i,j,l,t}$			$Personnel_{i,j,l,t}$		

Panel A. Reduced Features

Architecture	Estimate	Assumption t-statistic	CI	Estimate	Assumption t-statistic	CI
1	-0.3313	-0.3491	[-0.4259,	-0.1129	0.0465	[-0.4259,
	(0.0482)		-0.2368]	(0.0164)		-0.2368]
2	-0.3329	-0.4477	[-0.4260,	-0.1135	0.0968	[-0.4260,
	(0.0475)		-0.2398]	(0.0164)		-0.2398]
3	-0.3343	-0.4363	[-0.4275,	-0.1134	0.1168	[-0.4275,
	(0.0475)		-0.2412]	(0.0163)		-0.2412]
4	-0.3353	-0.4151	[-0.4284,	-0.1133	0.1405	[-0.4284,
	(0.0475)		-0.2422]	(0.0164)		-0.2422]
5	-0.3440	-0.3025	[-0.4380,	-0.1142	0.0982	[-0.4380,
	(0.0479)		-0.2500]	(0.0164)		-0.2500]
6	-0.3394	-0.2941	[-0.4337,	-0.1133	0.1043	[-0.4337,
	(0.0481)		-0.2451]	(0.0164)		-0.2451]
7	-0.3401	-0.2849	[-0.4344,	-0.1130	0.1320	[-0.4344,
	(0.0481)		-0.2458]	(0.0164)		-0.2458]
8	-0.3377	-0.4422	[-0.4326,	-0.1145	0.1259	[-0.4326,
	(0.0484)		-0.2428]	(0.0164)		-0.2428]
9	-0.3361	-0.4354	[-0.4309,	-0.1143	0.1337	[-0.4309,
	(0.0484)		-0.2413]	(0.0164)		-0.2413]
10	-0.3358	-0.4266	[-0.4306,	-0.1140	0.1498	[-0.4306,
	(0.0484)		-0.2409]	(0.0164)		-0.2409]

Panel A. Full Features

Architecture	Estimate	Assumption t-statistic	CI	Estimate	Assumption t-statistic	CI
1	-0.3304	-0.0092	[-0.4239,	-0.1319	-0.4477	[-0.1628,
	(0.0477)		-0.2369]	(0.0158)		-0.1009]
2	-0.3405	0.0312	[-0.4345,	-0.1301	-0.4234	[-0.1611,
	(0.0479)		-0.2465]	(0.0158)		-0.0990]
3	-0.3393	0.0281	[-0.4333,	-0.1304	-0.4397	[-0.1614,
	(0.0480)		-0.2452]	(0.0158)		-0.0993]
4	-0.3393	0.0302	[-0.4334,	-0.1304	-0.4419	[-0.1614,
	(0.0480)		-0.2453]	(0.0158)		-0.0994]
5	-0.3327	-0.0036	[-0.4262,	-0.1315	-0.4362	[-0.1624,
	(0.0477)		-0.2392]	(0.0158)		-0.1005]
6	-0.3303	-0.0087	[-0.4238,	-0.1319	-0.4445	[-0.1628,
	(0.0477)		-0.2368]	(0.0158)		-0.1009]
7	-0.3304	-0.0092	[-0.4239,	-0.1319	-0.4477	[-0.1628,
	(0.0477)		-0.2369]	(0.0158)		-0.1009]
8	-0.3402	-0.0600	[-0.4335,	-0.1313	-0.4466	[-0.1623,
	(0.0476)		-0.2468]	(0.0158)		-0.1004]
9	-0.3403	-0.0518	[-0.4337,	-0.1319	-0.4609	[-0.1628,
	(0.0476)		-0.2469]	(0.0158)		-0.1009]
10	-0.3407	-0.0466	[-0.4341,	-0.1318	-0.4620	[-0.1627,
	(0.0476)		-0.2474]	(0.0158)		-0.1008]

Notes: Panel A reports estimates with a reduced list of features as training data, while Panel B reports estimates with the full list of features. The Equal Error Assumption t-statistic and 95 percent confidence interval are reported. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed. Industry clustered standard errors are shown in brackets. The estimates are obtained from Huber Regressions.

TABLE VII: ML-DID Estimates with Random Forest for R&D and Personnel

	$Citations$		$PatShock$		$RDPerson$	
	(1)	(2)	(3)	(4)	(5)	(6)
$Bank_{i,j,l,t}$	-0.2480	-0.2183	-0.0335	-0.0248	-0.2517	-0.3334
	(0.0929)	(0.0484)	(0.0102)	(0.0039)	(0.0842)	(0.0839)
Observations	20,532	20,532	20,532	20,532	20,532	20,532
Controls	Y	Y	Y	Y	Y	Y
Industry Clusters	N	Y	N	Y	N	Y
Year FE	N	Y	N	Y	N	Y
Industry FE	N	Y	N	Y	N	Y
R^2	0.0298	0.0464	0.0625	0.1077	0.0459	0.1683

Notes: The table reports estimate with additional dependent variables that measure corporate innovation. Columns (1), (3), and (5) reports estimate without fixed effects and industry cluster, while other columns contain both. Clustered standard errors are in brackets. All continuous variables are winsorized at the first and 99th percentiles. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed.

TABLE VIII: Additional Dependent Variables Robustness Tests

	$RD_{i,j,l,t}$		$Patent_{i,j,l,t}$		$Personnel_{i,j,l,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$Bank_{i,j,l,t}$	-0.5491	-0.5104	-0.0844	-0.0882	-0.1415	-0.1474
	(0.1156)	(0.1070)	(0.0208)	(0.0197)	(0.0352)	(0.0320)
Observations	20,532	20,532	20,532	20,532	20,532	20,532
N-Neighbor	3	3	3	6	6	6
Controls	Y	Y	Y	Y	Y	Y
Industry Clusters	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y

Notes: The table reports estimate from propensity score matching. The first stage is estimate via probit regression. Columns (1) through (3) employ nearest neighbor of 3, while columns (4) through (6) employ nearest neighbor of 6. Year and firm fixed effects are included for all columns. Clustered standard errors are in brackets. All continuous variables are winsorized at the first and 99th percentiles. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed.

TABLE IX: Propensity Score Matching

	$Patent_{i,j,l,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$Bank_{i,j,l,t}$	CR4<50%	CR4<25%	CR8<50%	CR4>50%	CR4>75%	CR8>50%
	-0.1449	-0.1269	-0.1189	-0.0150	-0.0356	-0.0411
	(0.0154)	(0.0569)	(0.0208)	(0.0277)	(0.0462)	(0.0246)
Observations	11,640	5,743	11,713	11,847	5,854	11,774
Controls	Y	Y	Y	Y	Y	Y
Industry Clusters	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
R^2	0.2171	0.1624	0.2181	0.2638	0.2468	0.2610

Notes: The table reports estimate from partitioning the sample in terms of market competition. Column (1) displays results with firms that are affiliated with industries that are below 50 percentiles competitive according to the concentration ratio of top 4 corporates. Column (2) includes only below 25 percentiles and column (3) uses concentration ratio of top 10 corporates. Columns (4) through (6) resembles columns (1) through (3) but takes the top half of the sample in terms of competition. Industry clusters are included for all columns. Year and firm fixed effects are included for all columns. Clustered standard errors are in brackets. All continuous variables are winsorized at the first and 99th percentiles. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed.

TABLE X: Role of Industrial Competition on Relationship between Banking Background and Innovation

	$Patent_{i,j,l,t}$			
	(1)	(2)	(3)	(4)
$Bank_{i,j,l,t}$	KZ<25%	KZ>75%	WW<25%	WW>75%
	-0.0134	-0.1203	-0.0343	-0.0369
	(0.0132)	(0.0288)	(0.0373)	(0.0205)
Observations	5,871	5,871	4,671	5,871
Controls	Y	Y	Y	Y
Industry Clusters	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
R^2	0.2377	0.2934	0.3440	0.2220

Notes: The table reports estimate from partitioning the sample in terms of financial constraint. Column (1) and (2) displays results with firms that are affiliated with industries that are top and bottom 25 percentiles competitive according to the KZ index. Columns (3) through (4) displays results with firms that are affiliated with industries that are top and bottom 25 percentiles competitive according to the WW index. Year and firm fixed effects are included for all columns. Clustered standard errors are in brackets. All continuous variables are winsorized at the first and 99th percentiles. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed.

TABLE XI: Role of Industrial Competition on Relationship between Banking Background and Innovation

	$RD_{i,j,l,t}$	$Patent_{i,j,l,t}$	$Personnel_{i,j,l,t}$
	(1)	(2)	(3)
$BankNum_{i,j,l,t}$	-0.0523	-0.3424	-0.0733
	(0.0100)	(0.0831)	(0.0193)
Observations	20,532	20,532	20,532
Controls	Y	Y	Y
Industry Clusters	Y	Y	Y
Year FE	Y	Y	Y
Industry FE	Y	Y	Y
R^2	0.2525	0.1865	0.7239

Notes: The table reports estimate the number of executives with commercial banking backgrounds as independent variable. All columns contain fixed effects and industry clusters. Clustered standard errors are in brackets. All continuous variables are winsorized at the first and 99th percentiles. The sample period covers 2008 to 2020. Estimates for the fixed effects and controls are suppressed.

TABLE XII: Number of Executives with Commercial Banking Backgrounds

Year [Dummy]	Year [Dummy]
Industry Code [Dummy]	Industry Code [Dummy]
City Code [Dummy]	City Code [Dummy]
Return on Assets	Return on Assets
Operating Revenue Growth	Operating Revenue Growth
Total Assets	Total Assets
Inflation CPI	Inflation CPI
GDP	GDP
Tobin-Q	Tobin-Q
Fixed Assets	Fixed Assets
Market-to-Book Ratio	Market-to-Book Ratio
Financial Leverage Ratio	Financial Leverage Ratio
Enterprise Age	Enterprise Age
Cash Flow	Cash Flow
Herfindahl-Hirschman Index	Herfindahl-Hirschman Index
KZ Index	KZ Index
	Return on Equity
	Selling Expense Ratio
	Administrative Expense Ratio
	Financial Expense Ratio
	Return on Investment
	Cash Ratio
	Working Capital
	Interest Coverage Ratio
	Net Cash Flow from Operating Activities / Current Liabilities
	Liabilities to Assets Ratio
	Ratio of Net Profit to Total Profit
	Total Liabilities
	Debt to Assets Ratio
	Management Expense Ratio
	Income Tax Rate
	Earnings Volatility
	Cash Flow Volatility
	Growth Rate of Fixed Assets
	Growth Rate of Total Assets
	Growth Rate of Net Profit
	Growth Rate of Operating Profit
	Earnings Volatility
	Risk Coefficient
	Ratio of Current Assets
	Ratio of Cash Assets
	Ratio of Intangible assets
	Ratio of Shareholders' Equity to Fixed Assets
	Ratio of Current Liabilities
	Ratio of Operating liabilities
	Ratio of Current Assets
	Ratio of Current Assets

TABLE XIII: Reduced vs. Full Features

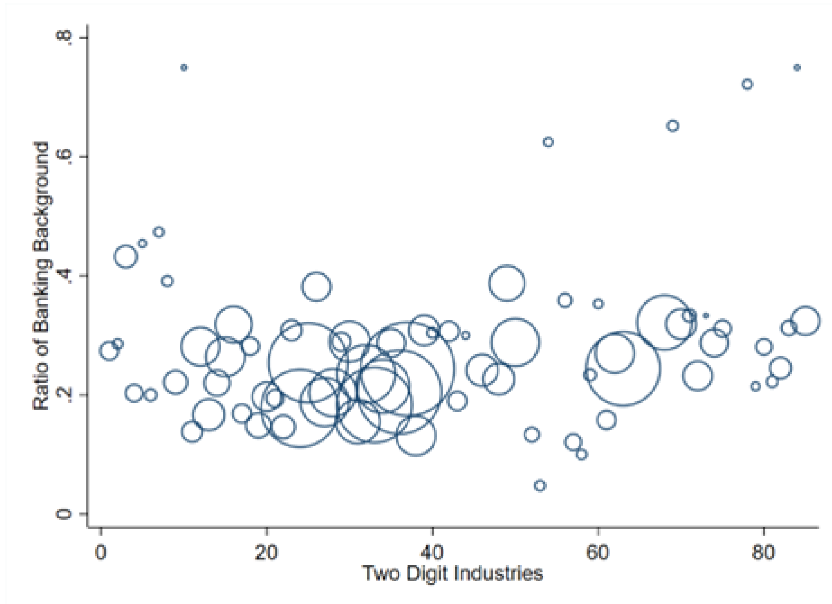


Fig. 1: Distribution of Commerical Banking Background

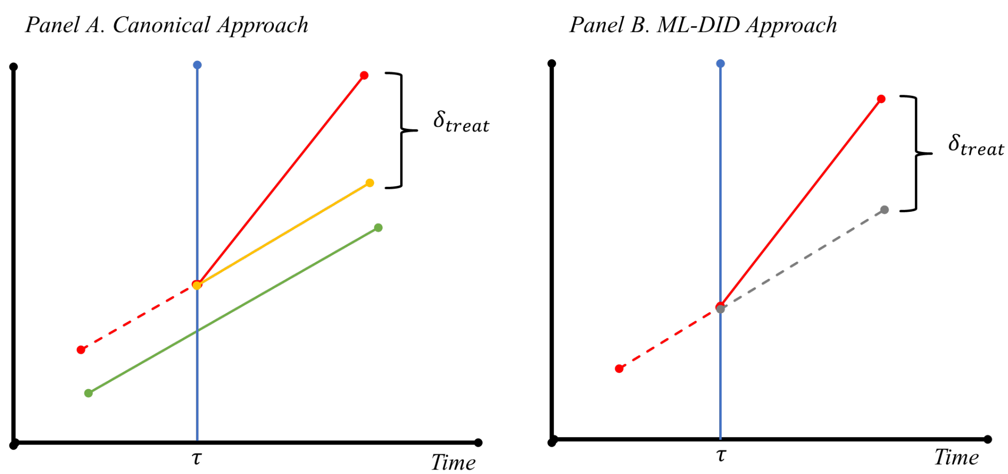


Fig. 2: Difference-in-Difference Canonical and Machine-Learning Approach

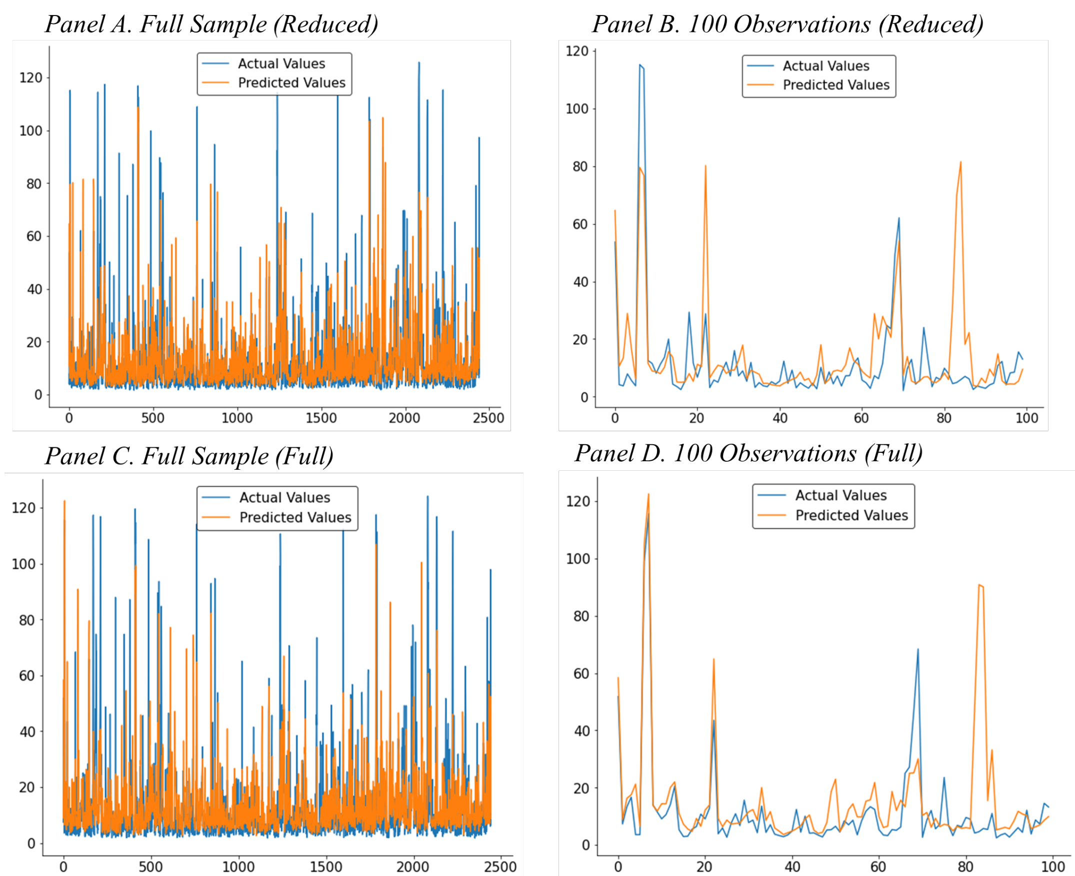


Fig. 3: Predicted Patent and Actual Patent Distributions

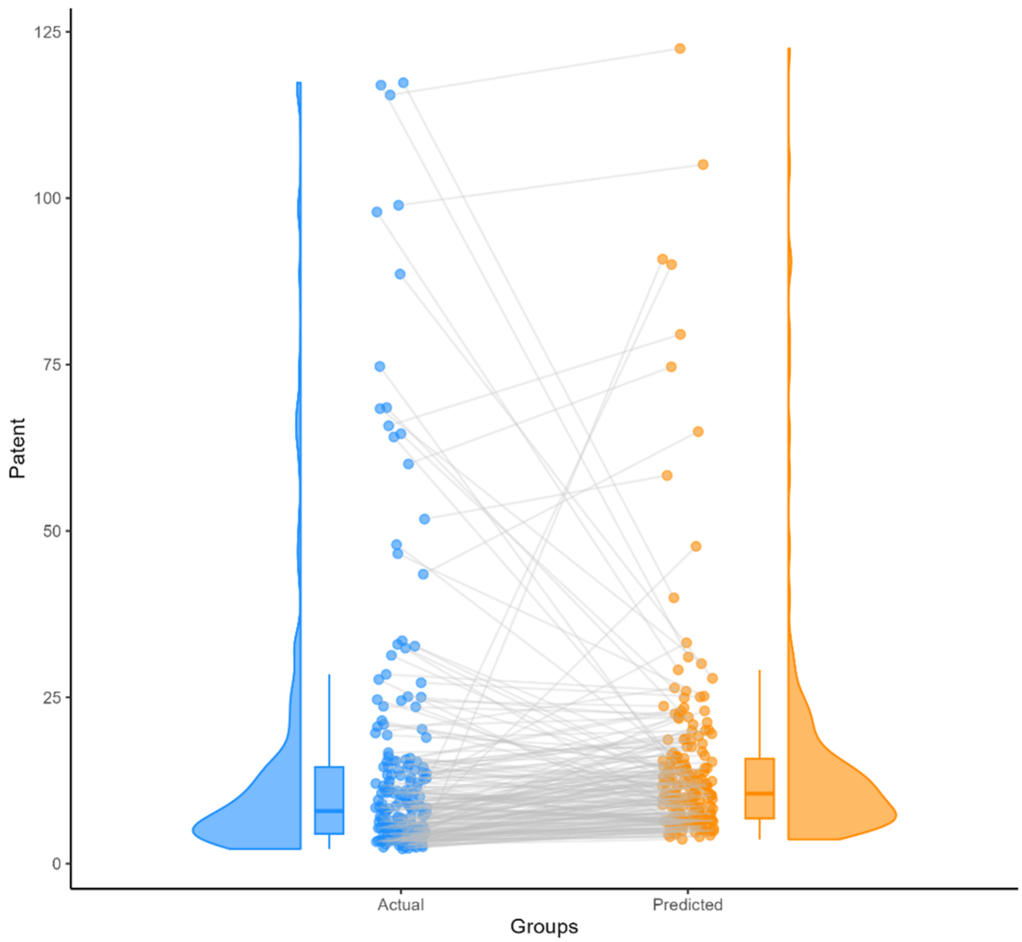


Fig. 4: Distribution of Individual Treatment Effect

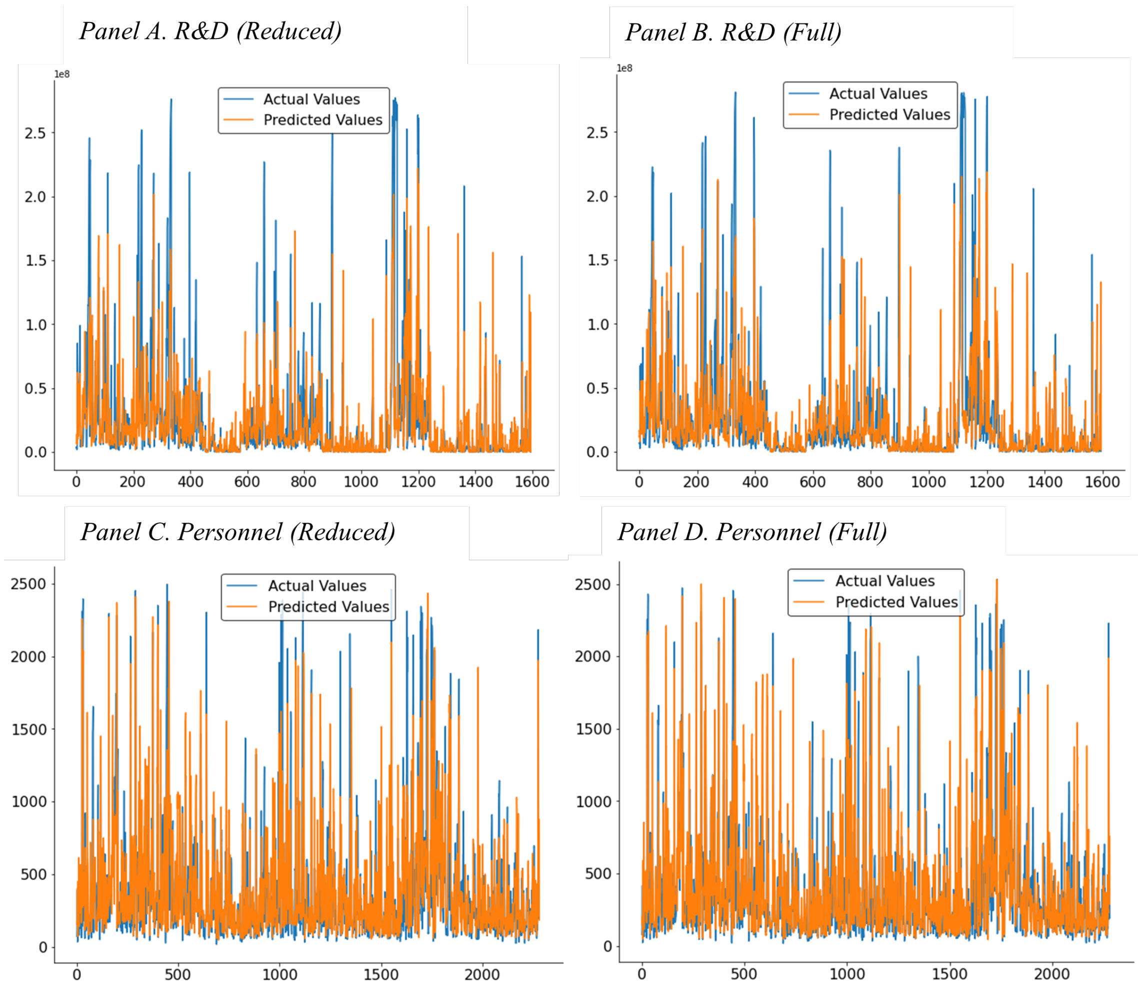
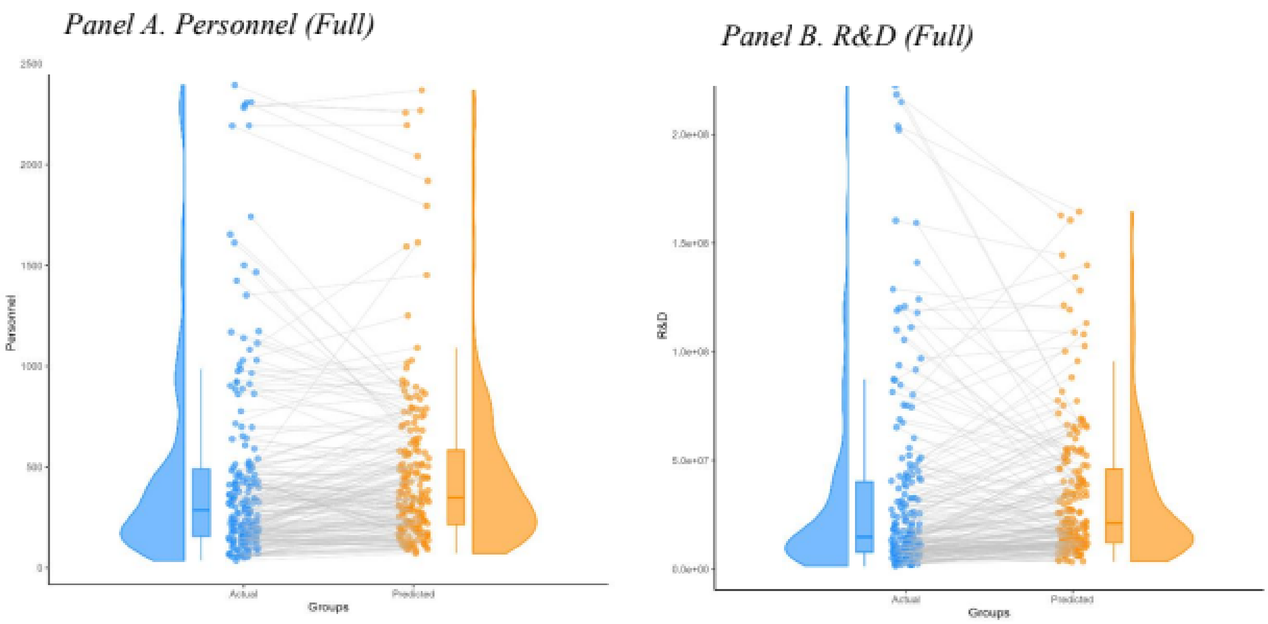


Fig. 5: Additional Predicted Patent and Actual Patent Distributions



Notes: Figure 3 presents individual treatment effects for patent applications for the first 200 observations. Patent is represented on the y-axis, while the two groups are shown on the x-axis.

Fig. 6: Distribution of Actual and Predicted Values

Fortune Favors the Bold: Impact of International Knowledge Transfer on Economic Growth

Michael Chen

Northwestern University

***Abstract*— Why do some poor countries exhibit rapid economic growth while others do not? Traditional growth models, which rely on factors of production, theorize that there will be economic convergence since there are diminishing returns to input growth, yet empirical evidence exists both for and against this theory. In this paper, I research the extent to which knowledge growth drives economic growth. In particular, since knowledge can be transferred between countries cheaper than traditional inputs like raw goods or labor, embracing change can be instrumental towards growth. I develop a model of economic growth as a function of domestic knowledge growth and international knowledge sharing. I then show that under this model, there is economic convergence. I test my theory with worldwide developmental data. Finally, I conclude that absolute and conditional economic convergence has been observed in the past 30 years.**

I. Introduction

The various efforts to reduce global income inequality have led to many multilateral initiatives to aid development and fight poverty. Of course, there have been many approaches to addressing this disparity, and growth economics offers an unusual perspective: Do nothing and inequality will fix itself. Many popular development models, such as the Solow growth model, predict that lower income countries will experience faster growth than their wealthier counterparts. This phenomenon is known as economic convergence.

Despite the prominence of economic convergence within popular growth models, the observed growth patterns of countries do not necessarily indicate the same conclusions. While there are cases of convergence, such as the rapid growth of South Korea in the later 20th Century (Figure 1), in other cases, countries fail to experience fast growth, and sometimes even grow at a slower rate than their wealthier counterparts, such as Zimbabwe during the same time period (Figure 2).

I would like to first thank my advisor, Mark Witte, for providing me with inspiration and feedback throughout the development of this thesis. His guidance and support have been invaluable in helping me explore this field and motivating me throughout the process. I am grateful for his patience and willingness to share his knowledge and expertise with me. I would also like to thank my friends Drew and Will for helping me brainstorm ideas and providing me with valuable insights. Their feedback and suggestions have been instrumental in helping me develop my research. Finally, I would like to thank my family for their unwavering support in my academic pursuits. Their love and encouragement have been a constant source of strength throughout this journey. I am grateful for their belief in me and for always being there to provide me with the support I needed.

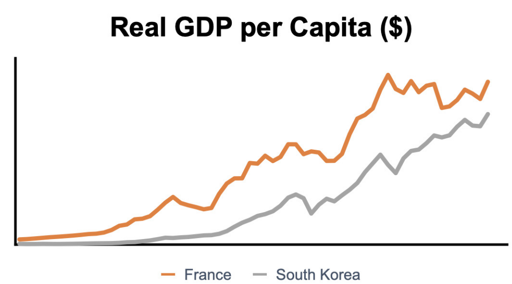


Fig. 1: Real GDP per capita in modern dollars between France and South Korea, 1960–2021

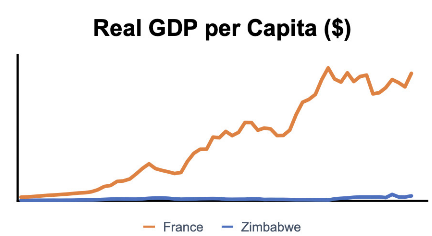


Fig. 2: Real GDP per capita in modern dollars between France and Zimbabwe, 1960–2021

This paper will look at why some countries follow convergence trends while others do not. Particularly, it will investigate the role knowledge growth within a country and knowledge transfer between countries has on economic growth.

II. Literature Review

Neo-classical growth theory has largely supported economic convergence. A growth model introduced by Solow (1956), and extended upon by Cass (1965), proposed that income convergence occurs because there are decreasing returns to growth in the factors of production. In this model, there is an optimal capital to labor ratio that leads to a steady state growth. Countries further from that ratio, i.e. poorer countries, grow faster and growth slows down the closer countries are to this optimal ratio. On the other hand, endogenous growth theories proposed that there will actually be a divergence in income (Romer 1994). Wealthier nations are able to invest more in human capital, leading to higher levels of growth in human capital and subsequently the national income.

Empirical research has also been split on the issue of economic convergence. Earlier studies have shown a lack of convergence and, in some cases, a divergence in the national income. Barro (1991) studied countries from 1960-1985 and showed that there was weak divergence, and Pritchett (1997) showed that there was strong divergence during the period of 1870-1990, where the ratio of per capita income between the richest and poorest countries increased by a factor of five. Conversely, in one study using data since the 1990s, Kremer, Willis, and You (2021) showed a trend towards convergence since 1990 as well as convergence since 2000, with both a slowdown in growth at the frontier as well as faster catch-up growth.

Because of the lack of clear evidence on convergence, empirical research has focused on the factors of steady state income, the level of income associated with a country effectively optimizing its resources like labor or capital. Barro and Sala-i-Martin (1992) found that economies further away from the steady state income level grew at a faster rate. Moreover, their research also found support for conditional convergence, which is a weaker form of convergence. When looking for conditional convergence, the researchers hold constant a set of variables that proxy for differences in characteristics that determine steady-state income. Some factors that researchers would look to account for differences in would be the difference in culture or level of technology. In another study, Evans and Karras (1996) also found that there was conditional convergence in the 48 contiguous U.S. states as well as in a group of 54 countries.

Culture is a crucial determinant in a steady state equilibrium, and as such, many researchers have tried to model its role in development. The role of culture has been modeled by Boyd and Richerson (1985) and Rogers (1988). In these models, there are infinite periods in a game where there is a certain probability that the state of the world changes from period to period. The payoff of an action depends on the state of the environment. Individuals can choose to follow the tradition of the previous generation or incur a learning cost to discover the true state of the world. If the action matches the correct state of the world, the agent receives a payoff, β. Because learning is costly, the model shows that there are certain benefits of having culture over pure rationality. Giuliano and Nunn (2021) derive a general equilibrium where there is a certain proportion of traditionalists depending on the cost of learning. They do this by deriving payoffs for being fully non-traditional and having some level of traditionalism. Formally, by treating the environment as an infinite period game, the payoffs can be written as

$$\pi_T = \frac{\beta(1-x)(1-\Delta)}{1-x(1-\Delta)}; \pi_{NT} = \beta - K.$$

In this scenario, the optimal choice is to forego some level of learning and stick to historical knowledge to save on costs. Nunn (2021) extends upon this model to include a country's level of technology within its profit function. Particularly, Nunn formalizes the arguments of Mokyr (2018) on the relationship between the level of technology, , level of traditionalism, x, and probability of the state changing, Δ, by adding the following requirements to the model:

$$(i)\Delta(x), \Delta'(x) < 0;$$
$$(ii)\frac{\dot{\mu}}{\mu}(x), \frac{\dot{\mu}'}{\mu}(x) < 0;$$
$$(iii)\Delta\left(\frac{\dot{\mu}}{\mu}(x)\right), \Delta\left(\frac{\dot{\mu}'}{\mu}(x)\right) < 0.$$

From the inclusion of technology, Nunn concluded the following payoff function for being traditional, π_T, and non-traditional, π_{NT}, are

$$\pi_T = \mu + \frac{\beta(1-x)(1-\Delta)}{1-x(1-\Delta)}; \pi_{NT} = \mu + \beta - K.$$

Nunn also gave a representation of Δ, where Δ can be defined as $1 - \Delta(x) = x^{\theta}$ for $\theta \in 0, 1$, which satisfies the aforementioned requirements. By plugging in this definition for the level of technology and solving for the optimal level of traditionalism, Nunn concluded two equilibria within the model: all traditionalists or no traditionalists.

Another factor of economic growth that has been studied is the development of technology. Jones (2021) explores the context of combinatorial growth with the draw of new ideas coming from a Weibull distribution. Using a thin-tailed distribution, Jones encapsulates the idea that it is easy to come up with a bad idea but hard to come up with good ones. In this environment, for each period, t, a country has a set of ingredients, N, that it can use to create new recipes. The number of ingredients that can be evaluated at each period is given by

$$\dot{N}_t = \alpha R_t^{\lambda} N_t^{\phi}, \phi < 1$$

where R_t is the number of researchers such that the number of researchers plus the number of laborers equals to the total population. Furthermore, Jones derived a constant asymptotic growth rate for N such that

$$g_N \equiv \frac{\dot{N}_t}{N_t} = \frac{\alpha R_t^{\lambda}}{N_t^{1-\phi}} = \frac{\lambda g_L}{1-\phi}$$

where g_L is the rate of population growth.

Now let Z_K be the maximum value from K number of draws from a thin-tailed distribution, which represents the best recipe available. As the population grows, there are more researchers so there are more chances to create new recipes, and K increases. Particularly, Jones derived the relationship between the growth of Z_K, g_Z; the growth of K, g_K; and g_N as

$$g_Z \equiv \frac{\dot{Z}_{Kt}}{Z_{Kt}} = \frac{g_{\log K}}{\beta} = \frac{g_N}{\beta} \tag{1}$$

where β is a parameter that governs the thinness of the tail of the distribution. Subsequently, the growth in output per capita, g_y, is defined as

$$g_y = g_Z = \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}.$$

Previous literature has largely revolved around the movement of factors of production (neoclassical growth), the movement of human capital (endogenous growth), or the growth in technology (combinatorial growth). In the remainder of this paper, I will look at the role of the flow of knowledge in economic growth, and in particular, how the transfer of knowledge between countries affects each country's economic growth. I will first outline my general model. I will then analyze the model in instances of one country, two countries, and a continuum of countries. Afterwards I will explain the data and empirical framework I will use to explore my model using historical data. Finally, I will discuss the results I concluded in my empirical research and relate it back to my theoretic model.

III. Model

We will now combine Nunn's and Jones' frameworks to create a new model that defines a country's payoff for a given period as a function of its level of traditionalism, level of technology, and the level of technology of other countries. Using these inputs, this model will encapsulate the idea that there is both internal knowledge growth and a flow of knowledge from one country to another. The model also captures that the level of knowledge within a country is a key factor in the utility or payoff of a country, similar to how human capital plays a key role in endogenous growth theories. First consider Nunn's framework of the role of culture. For each period, t, each country, i, has a certain proportion of traditionalists, $x_{it} \in [0,1]$. Now consider the combinatorial growth environment in Jones (2021). For each period, there are N_{it} number of ingredients available for each country to build with. The level of traditionalism determines how "risky" the researchers in Jones' environment are willing to be in utilizing novel technologies at period t. Therefore, the number of ingredients evaluated at each period t is

$$\dot{N}_{it} = \alpha R_{it}^{\lambda(1-x_{it})}.$$

Following Jones' derivation,

$$\frac{\dot{N}_{it}}{N_{it}} = \frac{\alpha R_{it}}{N_{it}^{1-\phi}}$$

and then

$$g_N \equiv \frac{\dot{N}_{it}}{N_{it}} = \frac{\lambda(1-x_{it})g_L}{1-\phi}$$

Plugging this definition of g_N into Jones' equilibrium (eq. 1), there is the result that the output per capita growth rate in this environment is

$$g_{y_{it}} = \frac{1}{\beta}\frac{\lambda(1-x_{it})g_L}{1-\phi}.$$

Furthermore, let there be multiple countries where each one is trying to maximize its per period payoff. Each country can choose to be fully non-traditional and choose to pay some sum k to re-learn everything and get a guaranteed payoff

$$\pi_{it}^{NT} = \mu_{it} + y_{it} - k_{it}$$

for some technology level, μ, that satisfies the restrictions outlined in Nunn's environment, output per capita, y, and non-trivial cost, k.

Each country can also choose a level of traditionalism, $x_i t$. In this case, the partial payoff at each period t related to country i's chosen level of traditionalism would be

$$\pi_{it}^{T} = \mu_{it} + \frac{y_{it}(1-x_{it})(1-\Delta_t)}{1-x_{it}(1-\Delta_t)}$$

for some technology level, μ, that satisfies the restrictions outlined in Nunn's environment, output per capita, y, and probability of the state of the world changing, $\Delta \in (0,1)$, that also satisfies the restrictions outlined in Nunn's environment.

Now consider that when a country chooses some level of traditionalism, it can not only rely on its own knowledge, but also knowledge from other countries. The total payoff is then a function of its technology level, output per capita, and relative technology factor

$$\pi_{it}^{T} = \mu_{it} + \frac{y_{it}(1-x_{it})(1-\Delta_t)}{1-x_{it}(1-\Delta_t)} + (1-x_{it})\sum_{j\neq i}(\mu_{jt}-\mu_{it})^2.$$

The second-degree difference between technologies represents the idea that countries at the extremes benefit the most from exchanging information. Countries at the lower extreme are able to play catch up with more modern technologies, while the countries at the higher extreme benefit from being able to create the next invention that is most compatible with their current cookbook. In contrast, countries at the middle of the technology spectrum cannot learn from those who are less advanced and might have different, but comparable in technological level, ingredients to their more advanced counterparts so they cannot build the new technologies without having to first replace their current ingredients.

One representation of μ is Z_K, the best ingredient available to a country, found in Jones' environment, where

$$\mu_{it} \triangleq Z_{K_{it}}$$

satisfies the restrictions on μ and Δ in Nunn's environment. Moreover, using Nunn's representation of Δ, $1-\Delta(x) = x^\theta$ for $\theta \in 0, 1$, the payoff for traditionalism can then be written as

$$\pi_{it}^T = Z_{K_{it}} + \frac{y_{it}(1-x_{it})x_{it}^\theta}{1-x_{it}^{\theta+1}} + (1-x_{it})\sum_{j\neq i}(Z_{K_{jt}} - Z_{K_{it}})^2.$$

A. One Country

First look at the scenario where there is only one country. In this environment, $\{Z_{K_j} : j \neq i\} = \{\phi\}$ since there is only one country. Therefore, the relative technology factor can be rewritten as

$$\sum_{j\neq i}(Z_{K_{jt}} - Z_{K_{it}})^2 = 0$$

and

$$\pi_{it}^T = Z_{K_{it}} + \frac{y_{it}(1-x_{it})x_{it}^\theta}{1-x_{it}^{\theta+1}}.$$

Now solve for the optimal level of traditionalism in this environment. The derivative of the payoff function with respect to traditionalism is lower bounded by

$$\frac{\partial}{\partial x_{it}}\pi_{it}^T > 0 \tag{2}$$

for all $x_{it} \in (0, 1)$ and $\theta \in (0, 1)$.

Thus, the optimal level of traditionalism to optimize payoff is x_{it} as close as possible to 1. Assuming k_{it} is in a range where neither π_{it}^T or π_{it}^{NT} is dominant for all values of x_{it}. Then, the optimal decision for maximizing payoff would be to either be non-traditional or choose to be fully traditional.

Now look at the growth of the payoff. Following the definitions of growth in Jones' paper, the growth rate of the payoff is

$$g_{\pi_{it}} = \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(1-x_{it}) + \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\frac{(1-x_{it})^2 x_{it}^\theta}{x_{it}^{\theta+1}} \tag{3}$$

where for $x_{it} \in [0, 1], g_{\pi_{it}}$ is maximized for some $x_{it} \in (0, 1)$.

Thus, by optimizing for the optimal return and choosing a level of traditionalism at 0 or 1, countries are growing at a sub-optimal rate.

B. Two Countries

In the context of two countries, there is a singular relative technology factor term so

$$\pi_{it}^T = Z_{K_{it}} + \frac{y_{it}(1-x_{it})x_{it}^\theta}{1-x_{it}^{\theta+1}} + (1-x_{it})(Z_{K_{jt}} - Z_{K_{it}})^2.$$

In this environment, solving for the optimal level of traditionalism gives the following derivative

$$\begin{aligned}\frac{\partial}{\partial x_{it}}\pi_{it}^T =& \frac{1}{(1-x^{\theta+1})^2}\Big[y_{it}(1-x^{\theta+1})(\theta x^{\theta-1} - (\theta+1)x^\theta) \\ &+ y_{it}(1-x)x^\theta(\theta+1)x^\theta\Big] - (Z_{K_{jt}} - Z_{K_{it}})^2.\end{aligned}$$

There are two scenarios depending on the relative technology factor. If the relative technology factor is small, then the optimal decision is to be fully traditional or non-traditional, similar to the one country case, which leads to sub-optimal growth. On the other hand, if the relative technology factor is sufficiently large, then the optimal decision is to choose a level of traditionalism $x \in (0, 1)$.

Formally, we can describe these scenarios as for

$$\begin{aligned}&(Z_{K_{jt}} - Z_{K_{it}})^2 \\ &< \lim_{a\to 1}\frac{1}{(1-a^{\theta+1})^2}\Big[y_{it}(1-a^{\theta+1})(\theta a^{\theta-1} - (\theta+1)a^\theta) \\ &\quad + y_{it}(1-a)a^\theta(\theta+1)a^\theta\Big]\end{aligned}$$

The same results hold for the two-country scenario as in the one-country scenario, particularly

$$\frac{\partial}{\partial x_{it}}\pi_{it}^T > 0 \text{ for all } x_{it} \in (0, 1), \theta \in (0, 1). \tag{4}$$

On the other hand, for

$$\begin{aligned}&(Z_{K_{jt}} - Z_{K_{it}})^2 \\ &> \lim_{a\to 1}\frac{1}{(1-a^{\theta+1})^2}\Big[y_{it}(1-a^{\theta+1})(\theta a^{\theta-1} - (\theta+1)a^\theta) \\ &\quad + y_{it}(1-a)a^\theta(\theta+1)a^\theta\Big]\end{aligned}$$

there is the result that

$$\text{there exists } a \in (0, 1) \text{ such that } \frac{\partial}{\partial x_{it}}\pi_{it}^T(a) = 0 \tag{5}$$

so there is a maximum payoff achieved for some mixed level of traditionalism.

Now, looking at the growth of the payoff. We have the following growth term

$$\begin{aligned}g_{\pi_{it}} =& \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(1-x_{it}) + \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\frac{(1-x_{it})^2 x_{it}^\theta}{1-x_{it}^{\theta+1}} \\ &+ 2(1-x_{it})(Z_{K_{it}} - Z_{K_{jt}})\frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(x_{it} - x_{jt}).\end{aligned}$$

When adding a country, the growth rate of a particular country is also dependent on the attributes of the other country. Under certain conditions for the relationship between one country's level of technology and traditionalism and their peer's levels, there is a guarantee of positive growth. Formally, there is positive growth when

$$(Z_{K_{it}} - Z_{K_{jt}})(x_{it} - x_{jt}) > -1/2. \tag{6}$$

Generally, countries that lack significantly in their level of technology to their peers would need to be more traditional than their peers in order to maintain positive growth. Also under this condition, the growth rate behaves similarly to its counterpart in the one-country scenario. Particularly, the growth rate is maximized for $x \in (0, 1)$.

When the condition is not satisfied,

$$g_{\pi_{it}}(0) < 0. \tag{7}$$

This does not preclude the growth rate from becoming positive for some value of traditionalism. As shown in the proof for (7), the change in growth rate at $x_{it} = 0$ approaches infinity regardless of the condition. So, the growth rate does increase for initial increments of the level of traditionalism, but it might not increase enough for the growth rate to ever become positive.

C. Multiple Countries

To generalize the model from two countries to multiple countries, assume that there is a continuous distribution of technology levels such that $Z_K \in (0, 1)$. Let P represent the distribution of Z_K and $p(Z_K)$ be the density function of P. Then, the payoff for country i during period t for being traditional is

$$\begin{aligned}\pi_{it}^T =& Z_{K_{it}} + \frac{y_{it}(1-x_{it})x_{it}^{\theta}}{1-x_{it}^{\theta+1}} \\ &+ (1-x_{it})\int_0^1 (Z_{K_{jt}} - Z_{K_{it}})^2 dP(Z_{K_{jk}}).\end{aligned}$$

Like the case with two countries only, the payoff function in the multiple-country scenario depends on the relative technology factor. When the relative technology factor is small

$$\begin{aligned}&\int_0^1 (Z_{K_{jt}} - Z_{K_{it}})^2 dP(Z_{K_{jk}}) \\ &< \lim_{a\to 1} \frac{1}{(1-a^{\theta+1})^2}\Big[y_{it}(1-a^{\theta+1})(\theta a^{\theta-1} - (\theta+1)a^{\theta}) \\ &+ y_{it}(1-a)a^{\theta}(\theta+1)a^{\theta}\Big]\end{aligned}$$

and the one-country conclusions still hold, particularly

$$\frac{\partial}{\partial x_{it}}\pi_{it}^T > 0 \text{ for all } x_{it} \in (0,1), \theta \in (0,1).$$

Moreover, following the same derivation as the case for the two-country scenario. When

$$\begin{aligned}&\int_0^1 (Z_{K_{jt}} - Z_{K_{it}})^2 dP(Z_{K_{jk}}) \\ &> \lim_{a\to 1} \frac{1}{(1-a^{\theta+1})^2}\Big[y_{it}(1-a^{\theta+1})(\theta a^{\theta-1} - (\theta+1)a^{\theta}) \\ &+ y_{it}(1-a)a^{\theta}(\theta+1)a^{\theta}\Big]\end{aligned}$$

there is also a similar result in that

$$\text{there exists } a \in (0,1) \text{ such that } \frac{\partial}{\partial x_{it}}\pi_{it}^T(a) = 0.$$

Now looking at the growth rate in the multiple-country scenario,

$$\begin{aligned}g_{\pi_{it}} =& \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(1-x_{it}) + \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\frac{(1-x_{it})^2 x_{it}^{\theta}}{1-x_{it}^{\theta+1}} \\ &+ 2(1-x_{it})\frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(x_{it} - x_{jt}) \\ &\cdot \int_0^1 (Z_{K_{jt}} - Z_{K_{it}})(x_{it} - x_{jt}) dP(Z_{K_{jk}}).\end{aligned}$$

The results for the growth rate in the multiple-country scenario are also similar to the two-country scenario, but because of the presence of multiple countries, it differs in that now there is a change in distribution of $Z_{K_{jt}}$, whereby denote the change in the density function as p_t' . Similar to the derivation in the two-country scenario, there is still a necessary condition to guarantee positive growth where

$$\int_0^1 (Z_{K_{jt}} - Z_{K_{it}})(x_{it} - x_{jt}) dP(Z_{K_{jk}}) > -1/2.$$

When the condition is not satisfied, there is the same result that the growth rate is not guaranteed to be positive since

$$g_{\pi_{it}}(0) < 0$$

Likewise, this scenario does not mean a failure of the condition leads to a negative growth rate for all values of traditionalism. The growth rate still increases for initial increments of the level of traditionalism, but as before, it might not increase enough for the growth rate to obtain a positive value.

IV. Empirical Setup

A. Empirical Model

There are three relationships between growth, traditionalism, denoted by T, and relative technology factor, denoted by RT, that are tested for in the historical data. According to the model:

1) Countries at the extremes of the level of technology should have a lower level of traditionalism;
2) Countries at the extremes of the level of traditionalism should grow at a slower rate;

3) Countries with a lower relative technology factor should grow at a faster rate.

Since there will be time series data across multiple countries, the empirical model will also be using country and year fixed effects, denoted by FE, to compare the effects of the exogenous variables across time periods and countries. For (1), the empirical model regressed relative technology on traditionalism. For practicality purposes, the relationship is generalized as a quadratic function, which captures the idea that the level of traditionalism is the highest in the middle values and decreases with movement towards either direction. So, Regression 1 is as follows:

$$T_{it} = \beta_0 + \beta_1 \cdot RT_{it} + \beta_2 \cdot RT_{it}^2 + FE.$$

For (2) and (3), the empirical model regressed the growth in GDP on traditionalism and relative economy. Following the model, the model used a linear relationship between the relative technology factor and growth. For traditionalism, the model also uses a quadratic function to generalize the relationship. Formally, Regression 2 is as follows

$$Growth_{it} = \beta_0 + \beta_1 \cdot T_{it} + \beta_2 \cdot T_{it}^2 + \beta_3 \cdot RT_{it} + FE.$$

B. Data

I used data from the World Bank on literacy rates and GDP per capita. Using literacy rates, I proxied the level of traditionalism a country has as a proportion of male literacy to female literacy, where a more traditional country would have a higher ratio. I chose to use the ratio of literacy rates because more traditional countries are less motivated to address the literacy gap between their male and female populations. Therefore, the ratio of literacy rates is a good representation of the level of traditionalism for a country. Traditionalism in this context has been seldom studied. This particular measure of traditionalism can also be correlated to poverty as wealthier countries are more able to educate their female population. Thus, I will be using the percent of population under the poverty headcount of $2.15 a day as a robustness check for this result. I also took the difference between each country's GDP per capita to the mean GDP capita at that time period as a measure of the relative technological level of each country. I chose to use GDP per capita as a measure for technology because countries with higher levels of technology have more efficient tools, effective processes, and streamlined organizational structures that should allow them to produce more with the same labor force. I also used GDP per capita at a specific time period as the total payoff a country receives per time period. I chose this measure because of the many factors considered by policy makers, not only is it the most accessible, but it is also one of the most important factors considered.

For my data collection process, I chose three time periods to measure the growth between, 1990, 2000, and 2010. While there is extensive GDP per capita data for every country in these time periods, the literacy data is less complete. Thus, because of insufficient literacy data, I averaged literacy data from 1986-1995 for 1990, 1996-2005 for 2000, and 2006-2015 for 2010. I did this because even if there were changes to the literacy rates, the changes would need time to take effect so the difference in literacy rate year over year should be pretty stable. Thus, averaging between a small window should not have a major effect.

V. Results

In this section, the results of Regression 1 will be discussed first, followed by the results of Regression 2. Finally, there will be a discussion on the limitations in the empirical strategy that might impact the results and provide alternatives for future works.

A. Regression 1

In Figure 3, I provide the results for Regression 1. In both a linear and quadratic setting, there is no statistical significance in the results. Thus, while I cannot say there is no relationship between the level of traditionalism and the relative technology factor, there is no evidence for the presence of a relationship.

Moreover, there is an argument that the relationship between the relative technology factor and level of traditionalism in fact does not exist. In the model, there is the assumption that for each period, every country is maximizing its current payoff. Under this assumption, there is the result that countries with a high or low relative technology factor will satisfy the condition where the maximum payoff is for a mixed level of traditionalism. For countries that have a middling relative technology factor, the maximum payoff is to be fully traditional. There are two potential issues that could make this not a plausible assumption. First, countries are not required to take a near-sighted view of the current payoff. In fact, there are many examples of countries taking a forward-looking approach towards policy making, such as the various climate pledges made by different countries to be carbon neutral by certain future dates. Second, the model only accounts for two main aspects: the level of technology and a measure of expected payoff per capita. While these are two important components of the policy making process, there are a plethora of other variables that are not accounted for that might have some influence on policy makers. Thus, there could be parts of the story that are not told by this model that would lead to differing results when looking at historical data.

	(1)	(2)
Constant	1.0455***	1.0053***
	(0.132)	(0.147)
RT	0.0428	0.1206
	(0.08)	(0.145)
RT^2		−0.0159
		(0.025)

* - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$

Fig. 3: Regression results of Regression 1

B. Regression 2

The results for Regression 2 are in Figure 4. As expected in the model, there is a negative linear relationship between the relative technology factor and growth. Moreover, this result is statistically significant with and without the traditionalism term. Thus, countries that are more advanced generally grow at a slower rate than their peers. Importantly, by using real GDP per capita in the calculation for the relative technology factor, this result shows the existence of convergence. Countries with a higher relative technology factor are represented in the data by countries that have a significantly higher real GDP per capita than the mean during the specific time period. Whereas countries with a low relative technology score are those with a significantly lower real GDP per capita than the mean. The negative linear relationship implies that countries that have weaker economies experience faster growth than those with stronger economies, and so leading to the desired convergence result. Furthermore, since the coefficient for the relative technology factor is negative and significant with and without the traditionalism variable, this regression indicates the presence of both conditional and absolute convergence.

	(1)	(2)	(3)
Constant	0.7934	1.4071	−2.7689
	(0.798)	(1.115)	(0.217)
T		−0.587	4.5176*
		(0.743)	(2.478)
T^2			−1.0018**
			(0.465)
RT	−1.6869***	−1.6618***	−1.6949***
	(0.485)	(0.487)	(0.474)

* - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$

Fig. 4: Regression results of Regression 2

Moreover, the coefficients for traditionalism also characterize the model's prediction, which is that countries at the extreme levels of traditionalism should grow at a slower rate. In the case of a linear relationship between traditionalism and growth, there is no statistical significance. Instead, when I regressed using a quadratic relationship between traditionalism and growth, both the first-degree and second-degree polynomial terms are statistically significant. This result highlights the curved nature of this relationship. In the linear case, there was a model misspecification leading to no significance in the result. In the quadratic case, the model was close enough to the specification of the real relationship to pick up on a significant result. Thus, even if the relationship isn't exactly a quadratic, as we used a quadratic in this test to simplify the regression, it follows some kind of shape similar to a quadratic relationship. Moreover, the coefficient on the two traditionalism terms also characterizes the modeled shape for the quadratic relationship. The first-degree polynomial is positive, whereas the second-degree polynomial is negative. Together, these two terms form a concave function, which matches the model prediction of a higher growth rate for the middling levels of traditionalism.

C. Limitations

There are some limitations to the empirical setup and data collection methodology that could have affected the results. In the empirical setup, any convex relationship in the model was generalized as a second-degree polynomial. While the second-degree polynomial possesses the desired properties, the model itself is not defined by a second-degree polynomial. This would reduce the fit of the model onto the data as the model being fitted is not the right shape, albeit being similar. On the other hand, it might also be easier to fit a second-degree polynomial to the data instead of a more complex relationship because of the limited sample data used. There might not be enough data points to draw statistically significant results using a more complicated model.

Another potential issue is the choice of proxies. There is no defined empirical measure of traditionalism. For the time period of this empirical analysis, more traditional or conservative cultures were less motivated to teach literacy to their female population. However, if this were not the case, then no conclusion can be drawn for the relationship between traditionalism and growth since the proxy, ratio between male and female literacy rates, would then not be correlated with traditionalism. Similarly, there is no measurable statistic for a country's level of technology. The intuition behind using real GDP per capita is that a more technologically advanced country would be able to produce more with the same level of population. An issue with this is that some productions are more labor intensive, so if a country specializes in something that requires a lot of labor for less production, it might reflect negatively on its relative technological score. In such a case, the empirical model would believe that the country specializing in the more labor intensive product as having lower technology because its production per capita would be lower. Then, the conclusion would instead be that countries specializing in more labor intensive products grow at a faster rate. In the future, more research could be done on using a more technology-based rating.

Finally, there could be an issue with GDP per capita appearing in the calculation of both the growth rate and the relative economy. This is the most concerning of the limitations because there would be a negative bias on the results for relative economy since if all else were equal, a lower GDP per capita would lead to a lower relative economy and higher growth. A more technology-based proxy for the relative level of technology would help amend this issue as well.

VI. Conclusion

In this paper, I analyzed the economic convergence debate through the lens of knowledge growth and flow of knowledge between countries. Building upon previous literature in growth economics, I formulated a growth model that relates the economic growth of a country to its relative level of technology and level of traditionalism. Using this model, I drew three major conclusions, which I then tested using historical data from 1986-2015. This model provides a framework for addressing income inequality. Particularly, I argue that economic convergence exists, but sometimes a country's willingness to accept novel ideas inhibits their ability to leverage the vast amount of knowledge already learned by other countries to grow at a faster rate. Here, technology doesn't necessarily need to strictly refer to what we commonly consider as high-tech, such as faster computers or more accurate models. Technology can also refer to how we do things, like various organizational systems to manage a company or an assembly line or different institutions that define how society operates. Take the case of South Korea and Zimbabwe. One could argue that South Korea took extensive lengths to modernize, and adapt the institutions and technologies of modern countries, particularly democratic ones, whereas Zimbabwe hasn't experienced the same modernization. From the model in this paper, I would argue that Zimbabwe's failure to adapt to modern institutions and technologies inhibited their convergence.

Another piece of the puzzle that could be interesting to look at would be some factor of willingness to exchange information. In the model described in this paper, countries are allowed to learn from each other as long as they are willing. However, sometimes countries are reluctant to help each other, such as when countries in an alliance are not as willing to share information to those not in the alliance. In my model, this reluctance to share can be accounted for by giving a country a higher traditionalism level if other countries are not willing to work with a specific country, but there could be value in modeling this interaction explicitly too.

Moreover, this research can be used in other fields beyond just economic growth. For example, if each country is seen instead as a company, then a parallel model can be made for company growth as opposed to country growth. In this scenario, the level of technology can be intellectual property of the company, both in terms of how much IP a company has and how advanced the IP is. In this case, traditionalism would be a measure of how willing a company would be to invest in new technologies or processes. While there would need to be adjustments made to the model, since there are many times more companies than countries and many of them fail, there could be interesting results from analyzing the life cycle of a company using a parallel to traditionalism and relative technology factor. One such adjustment would be that companies cannot directly gain knowledge from other companies. However, they are able to observe both the choices of other companies and the corresponding results. One example would be the development of smart phones. Rival companies frequently "steal" features based on how popular a new release is.

REFERENCES

[1] Barro, R. J. (1991). Economic growth in a cross section of countries. The Quarterly Journal of Economics, 106(2), 407–443. https://doi.org/10.2307/2937943

[2] Barro, R. J., & Xavier Sala-i-Martin (1992). Convergence. Journal of Political Economy, 100(2), 223–251. http://www.jstor.org/stable/2138606

[3] Boyd, Robert, and Peter J. Richerson (1985). Culture and the evolutionary process. University of Chicago Press.

[4] Cass, David (1965). Optimum growth in an aggregative model of capital accumulation. The Review of Economic Studies, 32(3), 233-240. https://doi.org/10.2307/2295827

[5] Evans, Paul and Georgios Karras (1996). Convergence revisited. Journal of Monetary Economics, 37(2), 249-265. https://doi.org/10.1016/S0304-3932(96)90036-7

[6] Giuliano, Paula and Nathan Nunn (2017). Understanding cultural persistence and change. National Bureau of Economic Research. http://doi.org/10.3386/w23617

[7] Jones, Charles I. (2021). Recipes and economic growth: a combinatorial march down an exponential tail. National Bureau of Economic Research. http://doi.org/10.3386/w28340

[8] Kremer, Michael, Jack Willis, and Yang You (2021). Converging to convergence. National Bureau of Economic Research. http://doi.org/10.3386/w29484

[9] Mokyr, Joel (2018). The past and the future of innovation: some lessons from economic history. Explorations in Economic History, 69, 13-26. https://doi.org/10.1016/j.eeh.2018.03.003.

[10] Nunn, Nathan (2021). History as evolution. In Alberto Bisin and Giovanni Federico, Handbook of Historical Economics (41-91). Academic Press.

[11] Pritchett, Lant (1997). Divergence, big time. The Journal of Economic Perspectives, 11(3), 3-17. http://www.jstor.org/stable/2138181.

[12] Rogers, Alan R. (1988). Does biology constrain culture? American Anthropologists, 90(4), 819-831. https://doi.org/10.1525/aa.1988.90.4.02a00030.

[13] Romer, Paul M. (1994). The origins of endogenous growth. The Journal of Economic Perspectives, 8(1), 3-22. http://www.jstor.org/stable/2138148.

[14] Solow, Robert M. (1956). A contribution to the theory of economic growth. The Quarterly Journal of Economics, 70(1), 65-94. https://doi.org/10.2307/1884513.

APPENDIX

PROOF OF (2)

To show this result, first note that

$$y > 0 \text{ and } (1 - x_{it}^{\theta+1})^2 \text{ for all } x_{it} \in (0, 1), \theta \in (0, 1)$$

and

$$(1 - x^{\theta+1})(\theta x^{\theta-1} - (\theta + 1)x^{\theta}) + (1 - x)x^{\theta}(\theta + 1)x^{\theta} = \theta x^{\theta-1} - x^{\theta}\theta + x^{2\theta}.$$

Note that $x \in (0, 1)$ so it suffice to factor out an x^{θ} and find a lower bound on

$$f(x) = \theta x^{-1} - \theta - 1 + x^{\theta}.$$

There are the following properties of $f(x)$: it is continuous for $x \in (0, 1), \lim_{a\to 0} f(a) = 1$ for $x = 0, f(x) = 0$ for $x = 1$, and $f(x) \neq 0$ for all $x \in (0, 1)$. Under the intermediate value theorem, if $f(x) < 0$ for some $x \in (0, 1)$ then for some y between 0 and 1, $f(y) = 0$, which contradicts one of the properties of $f(x)$. Therefore, $f(x) > 0$ for all $x \in (0, 1)$.

PROOF OF (3)

To achieve this result, note that $g_{\pi_{it}}$ is continuous for $x \in [0, 1]$ where $g_{\pi}(0) \geq 0$ and $\lim_{a\to 1} g_{\pi}(a) = 0$. Now, to show that the growth rate is maximized for $x \in (0, 1)$, there exists $a \in (0, 1)$ such that $g_{\pi}(a) > g_{\pi}(0)$ since there is already $g_{\pi}(0) \geq \lim_{a\to 1} g_{\pi}(a)$. To show this, take derivative of the growth function as

$$\frac{\partial}{\partial x_{it}} g_{\pi_{it}} = -\frac{1}{\beta}\frac{\lambda g_L}{1-\phi} + \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\Big[\frac{(1 - x_{it}^{\theta+1})(\theta x_{it}^{\theta-1}(1 - x_{it})^2 - 2(1 - x_{it})x_{it}^{\theta})}{(1 - x_{it}^{\theta+1})^2} + \frac{()1 - x_{it})^2 x_{it}^{\theta}(\theta + 1)x_{it}^{\theta}}{(1 - x_{it}^{\theta+1})^2}\Big]$$

Focusing on

$$\begin{aligned}
&\frac{(1 - x_{it}^{\theta+1})(\theta x_{it}^{\theta-1}(1 - x_{it})^2 - 2(1 - x_{it})x_{it}^{\theta})}{(1 - x_{it}^{\theta+1})^2} + \frac{(1 - x_{it})^2 x_{it}^{\theta}(\theta + 1)x_{it}^{\theta}}{(1 - x_{it}^{\theta+1})^2} \\
=&\frac{(1 - x_{it}^{\theta+1})(\theta(1 - x_{it})^2 - 2(1 - x_{it})x_{it}^{\theta}x_{it}^{\theta-1})}{(1 - x_{it}^{\theta+1})^2 x_{it}^{\theta-1}} + \frac{(1 - x_{it})^2 x_{it}^{\theta}(\theta + 1)x_{it}^{\theta}}{(1 - x_{it}^{\theta+1})^2} \\
=&\frac{(1 - x_{it}^{\theta+1})(1 - x_{it})(\theta(1 - x_{it}) - 2x_{it}) + (1 - x_{it})^2 x_{it}^{\theta+1}(\theta + 1)}{(1 - x_{it}^{\theta+1})^2 x_{it}^{\theta-1}}
\end{aligned}$$

where

$$(1 - x_{it}^{\theta+1})(1 - x_{it})(\theta(1 - x_{it}) - 2x_{it}) + (1 - x_{it})^2 x_{it}^{\theta+1} > 0$$

for $x_{it} \in [0, \theta/(\theta + 2))$ and

$$(1 - x_{it}^{\theta+1})^2 x_{it}^{1-\theta} > 0$$

for $x_{it} \in (0, 1)$. Therefore,

$$\lim_{a\to 0^+} \frac{\partial}{\partial x_{it}} g_{\pi_{it}}(a) = \infty.$$

So for a small increase ϵ in x,

$$g_{\pi}(\epsilon) > g_{\pi}(0).$$

PROOF OF (4)

This result can be shown by first showing an upper bound on the second derivative. The second derivative is

$$\begin{aligned}
\frac{\delta^2}{\delta x_{it}^2}\pi_{it}^T =& \frac{y_{it}(1 - x_{it}^{\theta+1})^2(\theta(\theta - 1)x_{it}^{\theta-2} - \theta^2 x_{it}^{\theta-1} - \theta x_{it}^{\theta-1} + 2\theta x_{it}^{2\theta} - 1)}{(1 - x_{it}^{\theta+1})^4} \\
&- \frac{y_{it}(\theta x^{\theta-1} - \theta x^{\theta} - x^{\theta} + x^{2\theta})2(10x_{it}^{\theta+1}(\theta + 1)x_{it}^{\theta})}{(1 - x_{it}^{\theta+1})^4}
\end{aligned}$$

where

$$y > 0$$

and

$$(1 - x^{\theta+1})^2$$

for all $x_{it} \in (0,1), \theta \in (0,1)$ so we need to find an upper bound on

$$(1-x^{\theta+1})^2(\theta(\theta-1)x^{\theta-2} - \theta^2 x^{\theta-1} - \theta x^{\theta-1} + 2\theta x^{2\theta-1}) - (\theta x^{\theta-1} - \theta x^{\theta} - x^{\theta} + x^{2\theta})2(1-x^{\theta+1})(\theta+1)x^{\theta}$$

where $(1-x^{\theta+1})^2 > 0$ and

$$(1-x^{\theta+1})(\theta+1)x^{\theta} \text{ for all } x \in (0,1), \theta \in (0,1)$$

which leaves finding an upper bound on

$$h(x) = (\theta(\theta-1)x^{\theta-2} - \theta^2 x^{\theta-1} - \theta x^{\theta-1} + 2\theta x^{2\theta-1}) - (\theta x^{\theta-1} - \theta x^{\theta} - x^{\theta} + x^{2\theta})$$

To find the upper bound, start with

$$p(x) = \left(\theta(\theta-1)x^{\theta-2} - \theta^2 x^{\theta-1} - \theta x^{\theta-1} + 2\theta x^{2\theta-1}\right).$$

There's the following properties of $p(x)$: it is continuous for $x \in (0,1), \lim_{a\to 0^+} p(a) = -\infty$ for $x = 0, p(x) = 0$ for $x = 1$, and $p(x) \neq 0$ for all $x \in (0,1)$. Similar to before, under the intermediate value theorem,

$$p(x) < 0.$$

Now look at the second expression

$$q(x) = \theta x^{\theta-1} - \theta x^{\theta} - x^{\theta} + x^{2\theta}$$

Similarly, there's the following properties of $q(x)$: it is continuous for $x \in (0,1), \lim_{a\to 0^+} q(a) = -\infty$ for $x = 0, q(x) = 0$ for $x = 1$, and $q(x) \neq 0$ for all $x \in (0,1)$. So,

$$q(x) > 0.$$

With these two results,

$$h(x) < 0$$

so

$$\frac{\partial^2}{\partial}\pi_{it}^T < 0 \text{ for all } x_{it} \in (0,1), \theta \in (0,1).$$

Since the second derivative is negative, that means that the minimum value for the first derivative positive as $x \to 1$. So, it is sufficient to say that if the first derivative is always positive without $(\mu_{jt} - \mu_{it})^2$, as discussed in the one country section, and $(\mu_{jt} - \mu_{it})^2$ is less than the minimum value of

$$\frac{y_{it}(1-x_{it}^{\theta+1})(\theta x_{it}^{\theta-1} - (\theta+1)x_{it}^{\theta}) + y_{it}(1-x_{it})x_{it}^{\theta}(\theta+1)x_{it}^{\theta}}{(1-x_{it}^{\theta+1})^2},$$

then the first derivative will always be positive.

Proof of (5)

Show this by first noting that as shown before, the minimum value of the first derivative is when $x \to 1$. Under this condition,

$$\lim_{a\to 1} \frac{\delta}{\delta x_{it}} \pi_{it}^T(a) < 0$$

thus there is somewhere the first derivative is negative. There is also somewhere where the first derivative is positive because

$$\lim_{a\to 0^+} \frac{\delta}{\delta x_{it}} \pi_{it}^T(a) \to \infty$$

and since the derivative function is continuous for $x \in (0,1)$ under the intermediate value theorem, there has to be some value of traditionalism where the derivative is zero.

PROOF OF (6)

To show this, first note that

$$\frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\frac{(1-x_{it})^2 x_{it}^{\theta}}{1-x_{it}^{\theta+1}} > 0 \text{ for all } x_{it} \in (0,1),\ \theta \in (0,1)$$

To guarantee positive growth,

$$\frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(1-x_{it}) > 2(1-x_it)(Z_{K_{jt}} - Z_{K_{it}})\frac{1}{\beta}\frac{\lambda g_L}{1-\phi}(x_{it} - x_{jt})$$

which after simplifying, gives us the aforementioned result.

PROOF OF (7)

Following similar logic as the one country scenario, under this condition: $g_{\pi_{it}}$ is continuous for $x(0,1)$ where $g_\pi(0) \geq 0$ and $g(a) = 0$. The derivative of this growth function is

$$\frac{\delta}{\delta x_{it}} g_{\pi_{it}} = -\frac{1}{\beta}\frac{\lambda g_L}{1-\phi} + \frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\left[\frac{(1-x_{it})^{\theta+1}\left(\theta x_{it}^{\theta-1}(1-x_{it})^2 - 2(1-x_{it})x_{it}^{\theta}\right)^2}{(1-x_{it})^{\theta+1}} + \frac{(1-x_{it})^2 x_{it}^{\theta}(\theta+1)x_{it}^{\theta}{}^2}{(1-x_{it})^{\theta+1}} + 2\left(Z_{K_{jt}} - Z_{K_{it}}\right)\frac{1}{\beta}\frac{\lambda g_L}{1-\phi}\left[-(x_{it}-x_{jt}) + (1-x_{it})\right]\right]$$

where there is still

$$\lim_{a \to 0^+} \frac{\delta}{\delta x_{it}} g_{\pi_{it}}(a) \to \infty$$

and for a small increase $\epsilon \in x$,

$$g_\pi(\epsilon) > g_\pi(0).$$

So, the maximum growth rate is some value of mixed traditionalism, $x_{it} \in (0,1)$, under the condition that

$$(Z_{k_{jt}} - Z_{K_{it}})(x_{it} - x_{jt}) > -\frac{1}{2}.$$

PROOF OF (8)

To show this, note that the second derivative of the payoff with respect to traditionalism is the same between the two-country scenario and multiple country scenario. Thus, the derivation for the bound on the second derivative still stands in this scenario. Similarly, we can conclude that the first derivative is always positive because the relative technology factor is not a big enough negative number to make a difference.

ROBUSTNESS CHECK FOR REGRESSION 2

Figure 5 shows the results of Regression 2 with the addition of a poverty control variable. In this result, the coefficient for relative technology is still negative and statistically significant. Thus, we still have evidence for economic convergence. However, none of the coefficients are statistically significant for traditionalism or poverty proxies in both the linear and quadratic regressions. There are three interpretations for this result. One could be that given the limited sample size, there is too much noise to reliably find the true relationship. Another interpretation could be that there is no relationship that exists. Finally, there could also be the case that the introduction of the poverty variable creates a model misspecification that leads to poor estimates for the true relationship between the traditionalism proxy and economic growth.

	(1)	(2)	(3)	(4)
Constant	0.7934	1.4071	-2.7689	0.6505
	(0.798)	(1.115)	(0.217)	(3.319)
T		-0.587	4.5176^{*}	0.3069
		(0.743)	(2.478)	(4.127)
$\mathbf{T}^2$			-1.0018^{**}	-0.0432
			(0.465)	(0.977)
RT	-1.6869^{***}	-1.6618^{***}	-1.6949^{***}	-1.4498^{**}
	(0.485)	(0.487)	(0.474)	(0.676)
Poverty				-0.0182
				(0.015)

* - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$

Fig. 5: Robustness check regression results of Regression 2

Rationalizability, Iterated Dominance, and the Theorems of Radon and Carathéodory

Roy Long

University of Chicago

***Abstract*— The game theoretic concepts of rationalizability and iterated dominance are closely related and provide characterizations of each other. Indeed, the equivalence between them implies that in a two player finite game, the remaining set of actions available to players after iterated elimination of strictly dominated strategies coincides with rationalizable actions. I prove a dimensionality result following from these ideas. I show that for two player games, the number of actions available to the opposing player provides a (tight) upper bound on how a player's pure strategies may be strictly dominated by mixed strategies. I provide two different frameworks and interpretations of dominance to prove this result, and in doing so, I relate it to Radon's Theorem and Carathéodory's Theorem from convex geometry. These approaches may be seen as following from point-line duality. A new proof of the classical equivalence between these solution concepts is also given.**

I. Introduction

Strategic dominance is a standard concept in game theory. It encapsulates the intuition that in a strategic environment, one should never take some action a if in all possible scenarios (formally, all profiles of the other players' actions), another action a' does strictly better.

In this section, we review the concepts of dominance and rationalizability and summarize the content of the paper.

We consider finite, two player strategic games. For $i = 1, 2$, player i's (pure) strategies is the set of actions A_i.

Definition: A pure strategy for player i, $a_i \in A_i i$ is strictly dominated by another pure strategy a_i' if $u_i(a_i, a_{-i}) < u_i(a_i', a_{-i})$ for any a_{-i}.

Acknowledgment: I would like to thank Scott Gehlbach at the University of Chicago Political Science department (and Harris' Political Economy program) for teaching me game theory during my first year SSI Formal Theory class. Prof. Gehlbach introduced the concepts of iterated dominance and rationalizability to me, and this paper grew out of some observations I made from examples in his class. I would also like to thank various UChicago economics and computer science professors for kindly taking a look at the paper and giving feedback during Winter/Spring 2022. Roger Myerson also provided some important suggestions about improving the details of the paper, and the last proposition showing the bound is tight is partly motivated by his observation.

Dominance can be generalized to mixed strategies, where players randomize over actions. In this case, the expected payoff is compared.

Definition: A (pure) strategy $a_i \in A_i$ is strictly dominated by a mixed strategy s_i over the set $\{a_j\}$ if $u_i(a_i, a_{-i}) < U_i(s, a_{-i}) = \sum_j p_j u_i(a_j, a_{-i})$ for all a_{-i} where (p_j) is a probability distribution over $\{a_j\}$ specified by s.

If the game is finite, the iterated elimination of strictly dominated strategies may be performed. The simple algorithm is to repeatedly remove strategies (for either player) which are strictly dominated by some mixed strategy from the game, until no such strategies are left. It is easy to show that the set of remaining strategies are those that were not strictly dominated in the original game, and any strictly dominated strategy is removed by this process.

More recently, the solution concept of *rationalizability* was developed. For two player games the definition is as follows:

Definition: An action $a_i \in A_i$ is *not rationalizable* if for any belief $q = (q_1, \ldots, q_m)$ player i has about player $(-i)$'s strategy s_{-i}, there exists some other action $a_i' \in A$ (which may depend on q) such that $U_i(a_i, s_{-i}) < U_i(a_i', s_{-i})$

One can also give a recursive definition of k-level rationalizability ($k \in \mathbb{N}$) and repeatedly remove strategies that are not rationalizable until all remaining actions are rationalizable.

It is easy to show that any strictly dominated strategy is not rationalizable, as one intuitively expects. It turns out that the converse is true as well, which was first shown in 1984 independently by both Bernheim and Pearce.

Theorem I.1. *(Pearce, Bernheim). Given a finite, two player strategic game G, the set of (pure) strategies that remain after iterated elimination of strictly dominated strategies are precisely the strategies that are rationalizable in G.*

Bernheim's proof represents strategies as payoff vectors in $\mathbb{R}^d$ and uses the separating hyperplane theorem. Pearce's proof uses the minimax theorem for zero sum games. See Fudenberg or Yildiz for a presentation of the former and Obara for the latter.

The central result of this paper is **Theorem 2.1**. We prove a tight bound on the minimum number of strategies needed to form a mixed strategy that strictly dominates another strategy. There is a rich connection to convex geometry in our analysis, and there are several representations of dominance and rationalizability that lead to classical results in this area of mathematics.

The methods in section 2 also yield a new proof of **Theorem 1.1**, mirroring the separating hyperplane proof of the equivalence. We defer our proof to the appendix.

In section 2, we provide a proof of **Theorem 2.1** using Radon's Theorem motivated by a natural view of the equivalence between rationalizability and dominance. We also show that the bound still holds when a player can have infinitely many actions.

In section 3, we represent dominance using vectors in $\mathbb{R}^d$. We give another proof of **Theorem 2.1**, and discuss the mathematical connections between the two proofs.

We may also the two approaches as resulting from a kind of point-line duality. In section 3, we view strategies as payoff vectors (or points), while in section 2, strategies are represented by hyperplanes. Omitted proofs can be found in the Appendix.

II. Connections to Convex Geometry

We study the following question: Given a two player finite strategic game G with action sets A, B, suppose some player's (without loss of generality player 1) pure strategy $a_i \in A$ is strictly dominated by a mixed strategy on a set of actions $\{a_j\} = A' \subseteq A$. Can we bound the size of the support of this mixed strategy, $|A'|$?

An obvious bound is $|A'| \leq |A|$, and furthermore $|A'| \leq |A| - 1$, as if $a_i \in A'$ then a_i is still dominated by some mix on $A' \setminus \{a_i\}$.

Interestingly, it turns out that we actually can bound $|A'|$ by the number of actions available to player 2, i.e. $|B|$. This is not at all obvious.

More precisely, we have the following result:

Theorem II.1. *Let G be a finite two player game, where player 1's set of actions is A and player 2's set of actions is B. If $a_i \in A$ is strictly dominated by a mix of strategies over $\{a_j\} \subseteq A$, then in fact a_i is strictly dominated by a mixed strategy $\{a'_j\}$ consisting of at most $\min(|A|, |B|)$ actions. An analogous bound holds for player 2.*

Remark II.2. The bound for player 1 can be improved to $\min(|A| - 1, |B|)$ based on the simple observation above. In fact, we will show that this bound is tight.

Now the result for player 1 is trivial if $|A| \leq |B|$, but it is not immediately clear why such a result should be true otherwise. Why does the number of actions the opponent has affect (and moreover provides a bound on) how a player's pure strategies are dominated by mixes of their own strategies?

For example, if player 2 has $|B| = 4$ available actions and player 1 has $|A| = 10$ available actions, then this tell us that if some action $a_i \in A$ of player 1 is strictly dominated, it's in fact dominated by some set of just 4 of player 1's available actions.

When $|B| = m = 1$, it is trivial, as player 2 only has one action that she must play. When $m = 2$, this tells us that if a pure strategy is strictly dominated, either it is dominated by another pure strategy or by a mix of two strategies.

Henceforth, we assume von Neumann Morgenstern preferences, i.e. preferences over strategies are compared based on their expected payoffs.

To motivate this theorem, we start by first examining the case for $m = 2$. Consider the simple 3×2 game, with the payoffs as listed, where rows are player $1's$ actions and columns are player $2's$ actions.

1/2	L	R
U	(6,1)	(0,3)
M	(2,1)	(5,0)
D	(3,2)	(3,1)

Fig. 1: payoff matrix

Then if player 1 has belief $(q, 1-q)$ over $\{L, R\}$ about player 2's strategy, then player 1's expected payoffs by playing U, M, D are, respectively:

$$E_U = 6q$$
$$E_M = 2q + 5(1-q) = 5 - 3q$$
$$E_D = 3$$

By plotting these as graphs on the domain $q \in [0, 1]$, we can see that for every $q \in [0, 1]$, the $E_D(q)$ is less than at least one of $E_U(q)$ and $E_M(q)$. Rationalizability then tells us that D is strictly dominated by some mixed strategy of U and M.

Let $E_U = f_1, E_M = f_2, E_D = f_3$ for ease of notation. Since these functions are all linear and therefore convex, their maximum is also convex. So by taking the upper convex hull of the lines, ie. $f_{MAX} = \max_{1 \leq i \leq 3}\{f_i\}$, the graph of f_3 lies entirely below the upper convex hull f_{MAX} and is strictly

dominated by it.
Now let's look at a different game, where player 2 again has two actions $B = \{b_1, b_2\}$, and suppose player 1 has five actions $\{a_1, \dots, a_5\}$ corresponding to the following expected payoffs for player 1, if she has belief $(q, 1-q)$ over $\{b_1, b_2\}$:

$$E_1(q) = 1.2q + 0.4$$
$$E_2(q) = -1.3q + 1.3$$
$$E_3(q) = 0.5q + 0.8$$
$$E_4(q) = -0.8q + 1$$
$$E_5(q) = 0.8.$$

The graph below shows the expected payoffs of these strategies as a function of the belief $q \in [0, 1]$, where the black curve is E_2, the green curve is E_4, the horizontal purple curve is E_5, the positive sloping purple curve is E_1, and the blue curve is E_3.

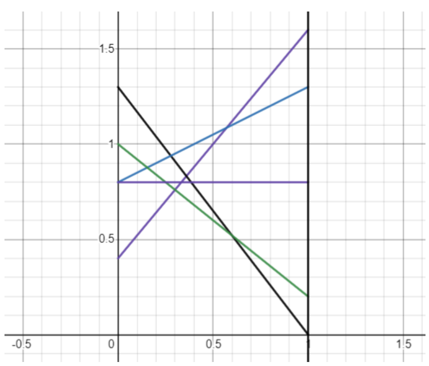


Fig. 2: expected payoffs given beliefs

We can see that a_4 and a_5 are strictly dominated by a_1, a_2, a_3 as

$$E_4(q), E_5(q) < \max\{E_1(q), E_2(q), E_3(q)\}$$

for all $0 \le q \le 1$. For example, a_4 is strictly dominated by a mix over a_1, a_2, a_3 with mixed strategy $q = (0.2, 0.3, 0.5)$, which results in an expected payoff of $0.2(1.2q+0.4)+0.3(-1.3q+1.3)+0.5(0.5q+0.8) = 0.1x + 0.87 > 0.8 = E_5(q)$ for all $q \in [0.1]$, and a_5 is strictly dominated by a mix over a_1, a_2, a_3 with mixed strategy $q = (0.2, 0.7, 0.1)$ which results in an expected payoff of $0.2(1.2q+0.4)+0.7(-1.3q+1.3)+0.1(0.5q+0.8) = 1.07 - 0.62q > -0.8q + 1$ for all $q \in [0, 1]$.

Notice that in both cases, we only need to choose two out of the three to guarantee a mixed strategy that strictly dominates them. For example, if we consider a mix over with a_1, a_2 with probability distribution $(0.3, 0.7)$, then the expected payoff is $0.3(1.2q + 0.4) + 0.7(-1.3q + 1.3) = 1.03 - 0.55q > -0.8q + 1 = E_4(q)$ for all $q \in [0, 1]$. Similarly, a_5 can be dominated by mixes over just $\{a_1, a_2\}$, as well as a mix of $\{a_2, a_3\}$. Graphically, we see that for a strictly dominated action, we can choose just two of the graphs of other functions such that the maximum of them lies strictly above the dominated strategies' graph.

Unfortunately, there are some edge cases where strict dominance never happens and the graph of every f_i is part of the upper convex hull $\max f_i$ at some point.

Thus, the equivalence of rationalizability and elimination of dominated strategies gives us the following result:

Proposition II.3. ***(Theorem 2.1 for** $d = 2$**)*** *Let $f_1, f_2, \dots, f_k : [0, 1] \to \mathbb{R}$ be linear, i.e. $f_i(x) = a_i x + b_i$ for all i, and $q \in [0, 1]$. Suppose $g : [0, 1] \to \mathbb{R}$ is also linear and satisfies: for all $x \in [0, 1]$, the $\max_i(f_i(x)) > g(x)$. Then there exists weights $0 \le r_1, \dots, r_k \le 1$ such that $r_1 + \dots + r_k = 1$ and $\sum r_i f_i(x) > g(x)$ for all $x \in [0, 1]$. In fact, it is possible to choose the r_i such that all but at most two of them are 0.*

In other words, if the graph of an action's expected payoff lies strictly below the upper convex hull (i.e., it's strictly dominated), it is in fact strictly dominated by a mix over a set of at most two actions.

Proof: The first part of the statement follows from the equivalence of rationalizability and iterated dominance, so we show the second part. First, note that we may assume that $g(x) = 0$ for $x \in [0, 1]$, as we consider $f_i'(x) = f_i(x) - g(x)$ for all i, which are still linear functions and the condition becomes $\max_i f_i' > 0$ with $g' = 0$.

First, remove all functions f_i where $f_i(x) \le g(x)$ for all $x \in [0, 1]$, as they will never be part of the upper convex hull. Now, for the graphs of the remaining functions, either they intersect the graph of g or they don't. If they don't, then they lie entirely above g, i.e. $f_k(x) > g(x)$ for all $x \in [0, 1]$, and we can just let $r_k = 1$ and all other $r_i = 0$.

Otherwise, they have a unique intersection point. For simplicity, as we noted above, we may assume g is zero. Then, we can look at where each f_k intersects $[0, 1]$ on the x-axis. Let L be the leftmost x-intercept in $[0, 1]$ such that the corresponding f has positive slope, and R be the rightmost x-intercept in $[0, 1]$ such that the corresponding f has negative slope. Then, it's easy to see that $L < R$ because otherwise $\max f$ is not greater than zero at points between L and R, so it reduces to only having two functions f_1, f_2, where $f_1 = a_1(x + b_1)$ and $f_2 = a_2(x + b_2)$ where $a_1 > 0, a_2 < 0$ and $0 \le b_1 < b_2 \le 1$.

Then, we have $r_1 f_1 + r_2 f_2 = r_1 a_1(x+b_1) + r_2 a_2(x+b_2)$, and by setting $r_1 a_1 = -r_2 a_2$, we just get a constant function which is greater than zero ($r_1 f_1 + r_2 f_2$ goes through the intersection point of f_1, f_2 which lies above the x-axis.)

Notice that in the last step of the previous proof, we didn't need to explicitly construct the function $r_1 f_1 + r_2 f_2$. In fact, the equivalence of rationalizability and iterated strict dominance tells us if the graph of $g(x)$ is strictly below the upper convex hull formed by some set of f_i, then g is strictly dominated by those f_i, so in particular, we're guaranteed that such r_i exists for the corresponding f_i that make the weighted sum greater than g. The main result of the previous proof was that we only needed to choose two of the f_i and it would be sufficient that the $\max f_i$ over those two functions would be greater than zero at all points.

Now, we generalize the preceding intuition to higher dimensions, i.e. when player 2 has more than two actions. In essence, what we did in the previous proof was this: after subtracting g from all the f_i, we were left with graphs of linear functions whose overall maximum is strictly greater than 0, i.e. their upper convex hull lies strictly the x-axis. Then, we looked at all linear functions that intersected $[0, 1]$. We can assume this because if the linear function had an x-intercept outside of $[0, 1]$, then its restriction to $[0, 1]$ will clearly be positive. Thus, we only needed to consider the cases of positive and negative slope that intersected $[0, 1]$. Then, if a positive sloped line intersected $[0, 1]$ at x_0, the $\max_i f_i(x) > 0$ for all $x > x_0$, and if a negative sloped line intersected $[0, 1]$ at x_0, we know $\max_i f_i(x) > 0$ for all $x_0 < 0$.

Thus, the problem is reduced to: Given a set of 1-dimensional rays or half lines that cover $[0, 1]$, we want to show that there exists a subset of at most two of them that also covers $[0, 1]$.

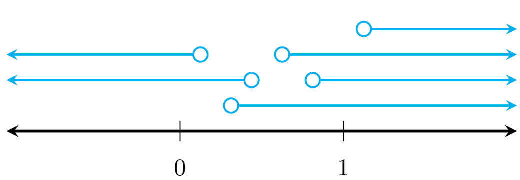


Fig. 3: Covering of $[0, 1]$ by half-lines

In general, a belief player 1 can have about player 2 where player 2 can choose from m possible actions $b_1, \dots, b_m$ is an m-tuple $(q_1, \dots, q_m) \in [0, 1]^m$ where $\sum_{i=1}^m q_i = 1$.
Letting $q_m = 1 - q_1 - \cdots - q_{m-1}$, the set of possible beliefs is represented by the $m-1$-dimensional simplex S^{m-1} (including its interior) of points $x = (x_1, \dots, x_{m-1}) \in [0, 1]^{m-1}$ where $\sum_{i=1}^{m-1} x_i \leq 1$. For example, when $m = 2$, this is just the line segment $[0, 1]$. When $m = 3$, this is the closed triangle region bounded by the x-axis, the y-axis, and the line $x + y = 1$. When $m = 4$, it is a tetrahedron with vertices at $(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)$, and so forth.

Let us examine this for the $d = 3$ case. We can plot the mixed strategies (beliefs) of player 2 in q_1-q_2 space, which is S^2, the right isosceles triangle in $\mathbb{R}^2$ bounded by $q_1, q_2 \geq 0, q_1 + q_2 \leq 1$. Then, the expected payoff for player 1 by choosing action a_i, with belief $q = (q_1, q_2, q_3 = 1 - q_1 - q_2)$, is a function $f_i : S^1 \subset [0, 1]^2 \to \mathbb{R}$ defined by $f(q_1, q_2) = a_{i,1} q_1 + a_{i,2} q_2 + a_{i,3}(1 - q_1 - q_2)$. Thus, the graph of f_i is the restriction of a plane to S^1.

Finally, we have the graph of g, which is also a plane. Again, we can consider $f' = f - g$, and the resulting functions are still planes, so we have $\max\{f'(x)\} > 0$ for all $x \in S^1$. Now, each of the planes f'_i will intersect the plane $\mathbb{R}^2 \times \{0\}$ at some line l in the q_1-q_2 plane. First, if any f'_i is parallel to $\mathbb{R}^2$, then clearly it is constant and strictly greater than zero, and so that single function $f_i > g$ for all $x \in S^2$. This means that the pure strategy a_i strictly dominates g so we may ignore this case. The cross sections will be lines that intersect the closed simplex S^2 (i.e., the closed triangle with vertices $(0, 0), (0, 1), (1, 0)$) and we will have open half planes $h_1, h_2, \dots, h_k$ that cover S^2.

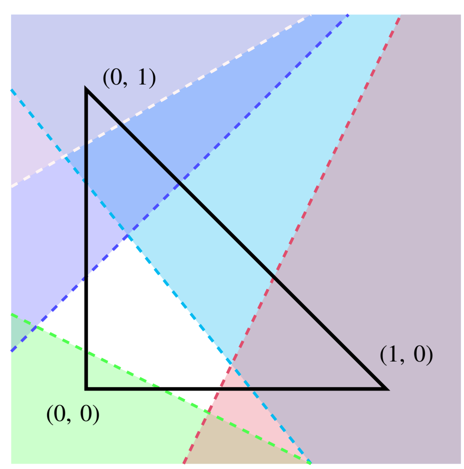


Fig. 4: A partial covering of S^2 with half planes

So in the general case, each action a_i has an expected utility $f_i : S^{m-1} \to \mathbb{R}$ whose graph is a $m-1$ dimensional hyperplane in $\mathbb{R}^n$, restricted to the domain S^{m-1}. After taking a transformation $f_i' = f_i - g$, the condition turns into $\max\{f_i'\} > 0$. Then, each hyperplane f_i' intersects $\mathbb{R}^{m-1} \times 0$, which cuts $\mathbb{R}^{m-1}$ into two half spaces, where one of the (open) half spaces represents the set of points where $f_i' > 0$, and the other (closed) half space represents the set of points where $f_i' \leq 0$. Of course, we only care about the intersection of the half spaces with the domain $S^{m-1} \subset \mathbb{R}^{m-1}$.

We are thus left with the following problem:

Proposition II.4. ***(Theorem 2.1, geometric form)*** *Let $Q \subset \mathbb{R}^d$ be a d-dimensional convex polytope (including its boundary and interior). Suppose a set of open half spaces $h_1, h_2, \ldots, h_n$ covers Q. Then, there exists a subset of at most $d+1$ of these h_i that also cover Q.*

The key observation is that we want to remove redundant half spaces. If a half space covers a region that is also covered by another set of half planes, then we can remove the first half plane and the covered region will be the same. A half space being covered by another half space corresponds to dominance by a pure strategy. Being covered by a set of half spaces corresponds to being dominated by a mixed strategy. We define a **minimal configuration** to be one where no half plane can be removed and the remaining half planes still cover the polytope, i.e. each half plane contains a point that is uniquely covered by it.

First, we gain some intuition for this result by showing it for $d = 3$, i.e. a covering of S^2 with half planes.

Proposition II.5. ***(Theorem 2.1 for d=3)*** *Let $Q \subset \mathbb{R}^2$ be a closed convex polygon including its interior. Suppose n open half planes covers Q. Then, there exists a subset of at most 3 half planes that cover Q.*

Proof: Suppose for contradiction that there exists a covering of the polygon Q with half planes $\{h_i\}$ such that any subset of $\{h_i\}$ that also coves Q contains at least four planes. Consider any minimal configuration $\{h_j'\} \subset \{h_i\}$ that covers Q. By assumption, $|\{h_j\}| \geq 4$.

Since $\{h_j'\}$ is minimal, we cannot remove any $h \in \{h_j'\}$ to get a smaller subset of half planes that also covers Q. Thus, for every h_j', there exists some point $p_j \in Q$ that is only covered by h_j'.

Thus, we have four points $p_1, p_2, p_3, p_4 \in Q$ such that p_j is uniquely covered by h_j' for all $1 \leq j \leq 4$.

There are a few possible configurations. We check the cases based on the convex hull H of p_1, p_2, p_3, p_4.

Suppose H is a line segment, i.e. $p_1p_2p_3p_4$ are collinear, in that order. Then, we've just reduced it to the $d = 2$ one dimensional case, as each plane bisects the line at some intersection point and covers everything to one side. So we see that there must be some plane that covers at least two points as there are more than two planes.

Now suppose H is a triangle, say $H = \triangle p_2p_3p_4$. Then p_1 lies in $\triangle p_2p_3p_4$. If p_1 lies strictly in the interior of $\triangle p_2p_3p_4$, then clearly any half planes that cover p_1 also intersect $\triangle p_2p_3p_4$, and we see that one of those vertices are necessarily covered as well. (If the plane doesn't cover any of the three vertices, then the entire triangle lies on the other side of the plane and is also not covered so p_1 is not covered, which can't happen.)

If p_2 lies on the boundary (one of the sides) of $\triangle p_2p_3p_4$, then we need to be a bit careful but the same result holds as in the previous subcase.

Finally, suppose H contains all four points p_1, p_2, p_3, p_4, so the points form a non-degenerate convex quadrilateral, without loss of generality $p_1p_2p_3p_4$ in clockwise order. Then, note that p_1p_3 intersects p_2p_4 at some point x inside H, and again we see that any point that covers x also covers at least two of the vertices of H, so this case is covered.

The result follows, i.e. any covering of S^2 with n half planes contains a subcover using at most three half planes.

For the general case in $\mathbb{R}^d$, we utilize **Radon's Theorem**, which states:

Theorem II.6. ***(Radon)*** *Let P be a set of $n \geq d+2$ distinct points in $\mathbb{R}^d$. Then, P can be partitioned into two nonempty subsets V_1 and V_2 such that the convex hulls of V_1 and V_2 intersect.*

For the general case in $\mathbb{R}^d$, let $Q \subset \mathbb{R}^d$ be a convex polytope (i.e., including the boundary and interior). Suppose there are $n \geq d+2$ half spaces $h_1 \ldots h_n$ which cover Q, such that each half space contains a point $p_i \in Q$ that is only covered by h_i.

By Radon's Theorem (as $n \geq d+2$), we can partition $P = \{p_1, \ldots, p_n\}$ into two sets P_1 and P_2 such that their convex hulls $H(P_1), H(P_2)$ intersect. Consider a point $x \subset H_1(P_1) \cap H(P_2)$ in this intersection, which lies in the original polytope

Q by convexity: as $P_1, P_2 \subset P \subset Q$, we have $H(P_1), H(P_2) \subset H(P) \subset H(Q) = Q$.

However, x cannot be covered by any h_i with corresponding p_i in P_1, as x is in H_2, and so any such h_i would have to intersect H_2 and contain some other point in P_2. Analogously x cannot be covered by any h_i with p_i in P_2. So $x \in Q$ is not covered by $\bigcup h_i$, which is a contradiction. This uses the fact that for an open half space, if a set of points all lie on the other side, then their convex hull also lies entirely on the other side.

So letting $Q = S^{n-1}$, the $n-1$-dimensional simplex, we have proven **Theorem 2.1**. ■

An immediate result of **Theorem 2.1** is that it gives us a characterization of which actions can be removed by iterated strict dominance, i.e. which ones are rationalizable. Indeed, as we've shown, if f_i lies under the upper convex hull of $f_1, \ldots, f_m$, then f_i is strictly dominated. In practice, we should be able to solve this with linear programming. If f_i is part of the upper convex hull at any point $q \in S^{m-1}$, then $f_i(q) \geq f_k(q)$ for all $1 \leq k \leq m$, so no mixed strategy beats f_i with belief q. Finally, we point out the case of weak dominance. When $m = 2$, it's easy to see that either the graph of the function lies strictly below the upper convex hull, coincides with the upper convex hull on some set that contains an interval, or intersects the upper convex hull exactly at a single point. The last case is when the action is weakly dominated. For $m = 3$, we see that weak dominance can result in a weak set (where the graph of the function coincides with the upper convex hull) of a point or a line, and so forth.

We also note that **Theorem 2.1** still holds even if one player has infinitely many actions.

Corollary II.7. *Let G be a two player game where player 2 has a finite number of actions $B = \{b_1, \ldots, b_m\}$, and player 1 has infinitely many actions $A = \{a_i\}_{i \in I}$ for some index set I. Then, if some action a_i is strictly dominated by a mixed strategy over a set of actions $\{a_j\} = A' \subset A$, in fact a_i is strictly dominated by a mixed strategy over a finite set of at most m actions.*

Proof: We simply note that we are covering $Q = S^{n-1}$ with open half spaces h_{a_j}, and since Q is compact (closed and bounded), $\bigcup_{a_j \in A'} h_{a_j}$ forms an open cover of Q, so there exists a finite subcollection $a'_1, a'_2, \ldots, a'_k \in A'$ such that $\bigcup_{a'_j \in A'} h_{a'_j}$ also covers Q. Then by the previous result, we again know that there exists a subset of at most m of these $h_{a'_j}$ that covers Q, and we are done.

III. A Linear Algebra Perspective

Now, let us turn to a completely different representation of dominance. This approach yields a more natural, geometric, interpretation of **Theorem 2.1**, which can be seen fundamentally as a dimensionality result.

Again suppose player 2 has m actions $b_1, \ldots, b_m$. We view the payoffs for player 1 as a matrix $A = (a_{ij})$ (linear transformation), and each action a_i as a row vector $v_i = (a_{i,1}, \ldots, a_{i,m})$.

Assuming von-Neumann Morgenstern Preferences, by taking a positive affine transformation, we may first transform the matrix to $A' = A + cJ$, where J is the $n \times m$ matrix with all 1's and c is a positive constant, so that all coordinates are strictly positive.

We view each payoff action as a vector in $\mathbb{R}^m$. Then, if we play a mixed strategy over some actions $\{a_j\}$ with probabilities $\{p_j\}$, then in vector form this represents $\sum p_j v_j$, where $\sum p_j = 1$.

In other words, if we take the convex hull of all the v_j, along with the origin, the possible mixed strategies over these vectors is represented by the outer boundary B of the convex hull H.

Similarly, we see that the convex polytope formed by taking the convex hull of the vectors is exactly the region

$$\left\{ \sum_{i=1}^{m} \lambda_i v_i \mid 0 \leq \lambda_1, \ldots, \lambda_m \leq 1 \text{ and } \sum_{i=1}^{m} \lambda_i \leq 1 \right\},$$

i.e. the region contained by the affine hull of the vectors $v_1, \ldots, v_m$.

The natural connection here is that the characterization of probability distributions over m elements using the simplex and the linear algebra formulation of convex hull of points using linear combinations are very similar, which leads to a re-interpretation of the probabilities/weights as coefficients of vectors.

With this formulation in mind, we now prove several preliminary results about strict dominance.

Proposition III.1. *A pure strategy a_i is strictly dominated by another pure strategy a_j if and only if the kth component of v_i is strictly less than the kth component of v_i for all $1 \leq k \leq m$.*

Proof: First, suppose that for some strategies a_i, a_j, that $a_{i,k} < a_{j,k}$ for all components $1 \leq k \leq m$. Clearly for any belief $q = (q_1, \dots, q_m) \in [0,1]^m, \sum q_k = 1$ we have $\sum_{k=1}^m q_k(a_{i,k} - a_{j,k}) < 0$ because each term is less than or equal to zero and not all of them are zero, so $\sum_{k=1}^m q_k a_{i,k} < \sum_{k=1}^m q_k a_{j,k}$, i.e. $E(a_i) < E(a_j)$, and a_i is strictly dominated by a_j.

Now, suppose $v_i = (a_{i,1}, \dots, a_{i,m})$ is strictly dominated by $v_j = (a_{j,1}, \dots, a_{j,m})$. Then, for all possible beliefs $q = (q_1, \dots, q_m)$, we must have

$$\sum_{k=1}^m q_k a_{i,k} < \sum_{k=1}^m q_k a_{j,k},$$

so

$$\sum_{k=1}^m q_k(a_{i,k} - a_{j,k}) < 0.$$

By letting

$$q = (1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1),$$

we see that $a_{i,k} < a_{j,k}$ for all $1 \leq k \leq m$.

It turns out that more generally, if we consider strict dominance by a mixed strategy, we can treat the mixed strategy as a pure strategy vector and compare them as if we were comparing two pure strategies' vectors. This leads us to the following proposition.

Proposition III.2. *Consider a mixed strategy over actions $\{a_j\}$ with probability distribution $\{\lambda_j\}$, which we represent as $u = \sum \lambda_j v_j$. Then, a pure strategy a_i is strictly dominated by the mixed strategy over $\{a_j\}$ if and only if the kth component of v_i is strictly less than the kth component of u, i.e. v_i is strictly dominated by the pure strategy vector representation of $\{a_j\}$ with lottery $\{\lambda_j\}$.*

Proof: Let $q \in [0,1]^m$ be any belief player 1 can have about player 2's possible strategy. Then, the expected payoff with the pure strategy a_i is simply the dot product $E(a_i) = \langle v_i, q \rangle$, and the expected payoff by playing the mixed strategy is

$$E(\{a_j\}) = \sum \lambda_j \langle v_j, q \rangle = \sum \langle \lambda_j v_j, q \rangle$$
$$= \left\langle \sum \lambda_j v_j, q \right\rangle = \langle u, q \rangle,$$

as with probability λ_j, player 1 plays action a_j, in which case the expected payoff is $\langle v_j, q \rangle$. Thus, we see that $E(a_i) < E(\{a_j\})$ if and only if $\langle v_i, q \rangle < \langle u, q \rangle$ for all possible beliefs q, which from the previous claim is equivalent to the components of v_i being strictly less than the components of u.

Now, we can start classifying which types of pure strategies are dominated by a mix over other strategies. Let's look at a set of pure strategies $\{a_j\}$ with corresponding vectors $\{v_j\}$. For example, consider the four vectors in $\mathbb{R}^2$ as below. Then, a good geometric intuition for whether a vector v is strictly dominated by some mix over $\{v_j\}$ is if the vector is contained in the convex hull formed.

For example, consider the following payoffs for player 1 in a game where player 2 has two actions and player 1 has three actions.

$$A = \begin{pmatrix} 1 & 5 \\ 5 & 1 \\ 2 & 2 \end{pmatrix}$$

We see that $a_3 = (2,2)$ is strictly dominated by a mix of over a_1, a_2 with equal probability for both actions, i.e. the mixed vector is $v' = \frac{1}{2}(1,5) + \frac{1}{2}(5,1) = (3,3)$ which strictly dominates $(2,2)$.

If we graph them, we see that $(2,2)$ lies between (inside) the vectors $(1,5)$ and $(5,1)$, i.e. the point $(2,2)$ lies inside the convex hull or triangle with vertices $(0,0), (1,5), (5,1)$. So a good intuitive guess is that points that lie within the convex hull formed by the vectors $(1,5), (5,1)$ will be dominated by some mix of the two.

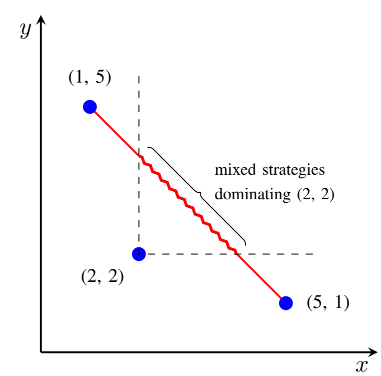


Fig. 5: vector representation of dominance

And indeed, this easily follows from the previous claims.

Proposition III.3. *If a strategy vector v_i lies inside the convex hull (affine hull) of the vectors $H\left(\bigcup v_j\right)$ (but not on the outer faces/boundary), then the corresponding strategy a_i is strictly dominated by a mix over $\{a_j\}$.*

Proof: By the linear combination definition of convex hull, this means that we can write $v_i = \sum \lambda_j v_j$, where $0 \leq \lambda_j \leq 1$. Remove all vectors that have trivial contribution so we may assume that all $\lambda_j > 0$. Note

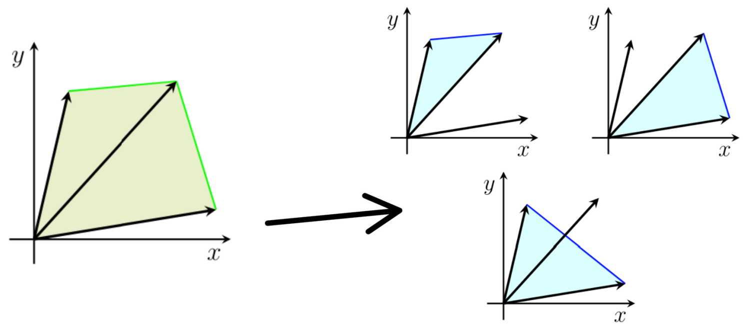


Fig. 6: covering of green convex hull by blue triangles

that $\sum \lambda_j = s < 1$ as the the vector v_i does not lie on the outer boundary of the convex hull, so we may extend v_i to a vector $u = \lambda v_i$ with $\lambda > 1$ that intersects the outer boundary of the convex hull.

In particular, let $u = \sum \lambda'_j v_j = \sum \frac{\lambda_j}{s} v_j$ which represents a mixed strategy of $\{a_j\}$ with probability distribution $\{\lambda_j/s\}$ as $\sum \frac{\lambda_j}{s} = 1$. Note that $\lambda_j < \frac{\lambda_j}{s}$ for all terms in the sum. Then have

$$\begin{aligned} E(\{a_j\}) &= \langle u, q\rangle \\ &= \left\langle \sum \frac{\lambda_j}{s} v_j, q \right\rangle \\ &> \langle \sum \lambda_j v_j, q\rangle \\ &= \langle v, q\rangle \\ &= E(a_i), \end{aligned}$$

and a_i is dominated by this mix over $\{a_j\}$ as claimed.

Proposition III.4. *(Characterization of strategies that are strictly dominated by a mix over a given set)*
Let $\{v_j\}$ be a set of pure strategy vectors corresponding to the actions $\{a_j\}$. Consider the boundary B, or the affine hull of the $\{v_j\}$ defined by

$$B = \left\{ \sum \lambda_j v_j : 0 \le \lambda_j \le 1, \sum \lambda_j = 1 \right\},$$

which represents the set of all possible mixed strategies over $\{a_j\}$. Then the set

$$X_B = \bigcup_{b \in B} \{v : v_{(k)} < b_{(k)} \, \forall \, 1 \le k \le m\}$$

defines the set of pure strategies that are strictly dominated by some mix of $\{a_j\}$. i.e. v is strictly dominated by a mix over $\{v_j\}$ if and only if $v \in X_B$.

Before using this characterization to provide another proof of **Theorem 2.1**, let us return to the case of points that lie strictly within the convex hull of some set of actions' vectors. We will use this to find an elegant and natural interpretation of **Theorem 2.1**. For now, we work with the closed convex hull (i.e., including the boundary B). The result is the same for open hull in the case of points that lie strictly within H.

Let's look at four points v_1, v_2, v_3, v_4. as shown below.
Then, the region of vectors that are strictly dominated by a mix of $\{v_1, v_2, v_3, v_4\}$ is the region on the following page.

We can also look at the pairs of vectors, which bound triangular regions (above right).

These triangular regions represent the regions that are strictly dominated by two vectors. More specifically, the convex hulls are a subset of what points are strictly dominated by that set of vectors.
Now, the motivation for **Theorem 2.1** becomes clear: the polygon region formed by the convex hull of the vectors $\{v_1, v_2, v_3, v_4\}$ is equal to the union of the convex hulls of pairs of the vectors.

Formally, if $H(S)$ denotes the convex hull of a set of points S, then our preceding observation is:

$$\begin{aligned} & H(O, v_1, v_2, v_3, v_4) \\ =& H(O, v_1, v_2) \cup H(O, v_1, v_3) \\ & \cup H(O, v_1, v_4) \cup \ldots \cup H(O, v_3, v_4) \\ =& \bigcup_{1 \le i,j \le 4} H(O, v_i, v_j) \end{aligned}$$

Similarly, if we have have n points in $\mathbb{R}^3$, i.e. $v_1, \ldots, v_n$ in $\mathbb{R}^3$, it's easy to see that taking the union over all triples of points, i.e. $H(O, v_1, \ldots, v_n) = \bigcup_{i,j,k} H(O, v_i, v_j, v_k)$ yields the entire convex hull. This is very intuitively clear. We are essentially considering all tetrahedrons with a fixed vertex O, and by shifting the vertices around, we use these tetrahedra to cover the entire interior of the

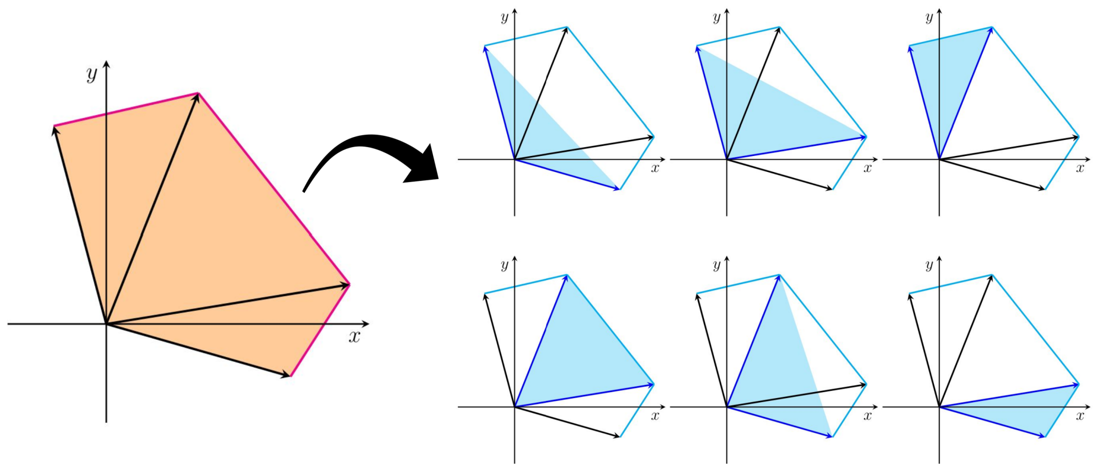


Fig. 7: convex hull is covered by the union of triangles over all pairs of vectors

polyhedra.

More generally, a non-degenerate set of $d + 1$ points in $\mathbb{R}^d$ should determine a d-dimensional figure, and by varying across different points in the set, we should be able to cover all d-dimensional points in the convex hull, which is the following statement: Given n nonzero points $v_1, \ldots, v_n$ in $\mathbb{R}^d$, representing vectors extending from O, then we have:
$H(O, v_1, \ldots, v_n) = \bigcup_{S_d \in \binom{[n]}{d}} H(O, v_{k_1}, \ldots, v_{k_d})$, where S_d varies over all subsets of
$[n] = \{1, \ldots, n\}$ with d elements.

It is now easy to show that the generalization of our intuitive observation follows from **Carathéodory's Theorem**.

Theorem III.5. ***(Carathéodory)*** *If a point $x \in \mathbb{R}^d$ lies in the convex hull of a set P, then x can be written as the convex combination of at most d + 1 points in P. i.e. there is a subset $P' \subset P$ consisting of $d + 1$ or fewer points such that $x \in H(P')$.*

For our proof, we need a variant known as **Conical Carathéodory's Theorem**

Theorem III.6. ***(Carathéodory's Theorem for Conical Hull)*** *Let $n \geq d + 1$ and $v_1, \ldots, v_n$ be n distinct nonzero vectors in $\mathbb{R}^d$. Consider the conical hull formed by this set. Then, if a nonzero point x is in the conical hull, x can be written as the conical combination of at most d of the v_i.*

Corollary III.7. *Furthermore, if x also lies in the convex hull $H(O, v_1, \ldots, v_n)$, i.e. x is a convex combination of the v_i (and $O = 0$, so sum of coefficients of the v_i is less than or equal to one), then x can be written as the convex combination of at most d of the v_i, and O.*

Finally, we can finish the second proof of **Theorem 2.1**. Let $v_1, \ldots, v_n$ be the vector payoffs for player 1. Let $H = H(O, v_1, \ldots, v_n)$ be the convex hull of the set of vectors, and let B be the outer boundary of H.

From the characterization of these points, we know that a vector $v \in \mathbb{R}^n$ is strictly dominated by a mix over a subset of $\{v_j\} = v_1, \ldots, v_n$ if and only if there exists some $b \in B$ such that $v_{(k)} < b_{(k)}$ for all components $1 \leq k \leq m$. But $B \subset H$, and we know from the corollary following Carathéodory's Theorem that all points in H can be represented as the affine sum of at most d distinct vectors. Thus, if v_i is strictly dominated by some such b, we can write b as the affine linear combination of at most d of the $\{v_j\}$, which means that v_i is strictly dominated by a mix over those v_j, and in particular, is dominated by at most d of the other vectors as claimed. ■

We conclude by showing that the bound in **Theorem 2.1** is tight in the following sense:

Proposition III.8. *For every pair $(n, m) \in \mathbb{Z}^2_{\geq 2}$ there exists a game $G_{n,m}$ with $|A| = n$ and $|B| = m$, and a strategy $a_i \in A$ that is strictly dominated by a mixed strategy over a set $\{a_j\} = A'$ of size $|A'| = \min(|A| - 1, |B|)$, and a_i is not strictly dominated by a mixed strategy over any smaller set of actions. An analogous statement holds for player 2.*

Proof: We again represent pure strategies in their vector forms, with a pure strategy a_i is associated to a vector $v_i \in \mathbb{R}^m$, and a mixed strategy over $\{a_j\}$ associated to a convex combination $\sum_j \lambda_j v_j$.

First suppose $|A| - 1 < |B|$. Then, consider

$$\begin{aligned} v_1 &= (n, 0, 0, \dots, 0, 1, \dots, 1) \\ v_2 &= (0, n, 0, \dots, 0, 1, \dots, 1) \\ v_3 &= (0, 0, n, \dots, 0, 1, \dots, 1) \\ &\dots \\ v_{n-1} &= (0, 0, 0, \dots, n, 1, \dots, 1) \\ v_n &= (1, 1, 1, \dots, 1, 0, \dots, 0) \end{aligned}$$

where the k-th coordinate of v_k is n and the j-th coordinate is zero for $1 \leq j \leq n-1$ and is one for $j \geq n$, for each $1 \leq k \leq n-1$, and the first $n-1$ coordinates of v_n are zero. Then, we see that v_n is strictly dominated by the mixed strategy

$$\begin{aligned} &\tfrac{1}{n-1}v_1 + \tfrac{1}{n-1}v_2 + \cdots + \tfrac{1}{n-1}v_{n-1} \\ =&(\tfrac{n}{n-1}, \tfrac{n}{n-1}, \dots, \tfrac{n}{n-1}, 1, \dots, 1) \\ \gg&(1, 1, 1, \dots, 1, 0, \dots, 0), \end{aligned}$$

but v_n is clearly not strictly dominated by a mixed strategy over any smaller set as each v_k for $1 \leq k \leq n-1$ is needed in the mixed strategy.

Now, suppose that $|B| \leq |A| - 1$. Then, consider

$$\begin{aligned} v_1 &= (2m, 0, 0, \dots, 0) \\ v_2 &= (0, 2m, 0, \dots, 0) \\ v_3 &= (0, 0, 2m, \dots, 0) \\ &\dots \\ v_m &= (0, 0, 0, \dots, 2m) \\ v_{m+1} &= (1, 1, 1, \dots, 1) \\ v_k &= (0, 0, 0, \dots, 0) \;\; \forall m+1 < k \leq n \end{aligned}$$

We see that v_{m+1} is strictly dominated by the mixed strategy

$$\begin{aligned} \tfrac{1}{m}v_1 + \tfrac{1}{m}v_2 + \cdots + \tfrac{1}{m}v_m &= (2, 2, 2, \dots, 2) \\ &\gg (1, 1, 1, \dots, 1), \end{aligned}$$

but v_m is clearly not strictly dominated by any mixed strategy over a set of $m-1$ or fewer actions, as v_k for $k \geq m+2$ are all zero vectors, and each v_k for $k \leq m$ is needed as v_k is the only vector that has a positive k-th coordinate.

References

[1] Bernheim, D. (1984) Rationalizable Strategic Behavior. Econometrica 52: 1007-1028.

[2] Pearce, D. (1984) Rationalizable Strategic Behavior and the Problem of Perfection. Econometrica 52: 1029-1050.

[3] Fudenberg, D. and Tirole, J. (1993) Game Theory. Cambridge: MIT Press.

[4] Yildiz, M. Rationalizability lecture notes. https://dspace.mit.edu/handle/1721.1/99213

[5] Obara, I. Rationalizability and Iterated Elimination of Dominated Strategies http://www.econ.ucla.edu/iobara/Rationalizability201B.pdf

APPENDIX

he appendix provides proofs of Caratheodory's and Radon's Theorems, as well as a new proof of **Theorem 1.1** using the representation of dominance in section 2 and half-spaces.

We begin the proof of **Theorem 1.1** with the following covering lemma.

Lemma .1. ***(Rotation covering)*** *Suppose two half spaces $a \cdot x < 0$ (A) and $b \cdot x < 0$ (B) together cover a compact convex set S in $\mathbb{R}^n$. Then there exists $\lambda \in [0,1]$ such that the half space C given by $(\lambda a + (1-\lambda)b) \cdot x < 0$ also covers S.*

Proof: Suppose neither A nor B covers S by itself so $\lambda \in (0,1)$. Also assume that a and b are not parallel (linearly dependent) as that case can be easily handled.

Thus the hyperplanes $a \cdot x = 0$ and $b \cdot x = 0$ intersect at an $n-2$ dimensional subspace P. Note that P lies entirely outside of S.

These two hyperplanes divide $\mathbb{R}^n$ into four quadrant regions based on the signs of $a \cdot x$ and $b \cdot x$. For the region where $a \cdot x < 0$ and $b \cdot x < 0$ clearly $(\lambda a + (1-\lambda)b) \cdot x < 0$. So consider the other two regions. The parts of S in them are $S_A = (S \cap A) \setminus B$ and $S_B = (S \cap B) \setminus A$ which lie in two different quadrants. By assumption that neither A nor B alone cover S, both S_A, S_B are nonempty. Also S_A, S_B are compact and convex.

Consider the plane C_λ: $(\lambda a + (1-\lambda)b) \cdot x = 0$ as λ increases from zero to one. Note that C_0 intersects S_A and C_1 intersects S_B. Suppose for contradiction that C_λ always intersects at least one of S_A and S_B. As S_A, S_B are both closed, this means for some λ^* the plane intersects both S_A and S_B, say at points p and q.

We claim that the line segment $\overline{pq}$ intersects P. Note that $C_{\lambda^*} \subset P \cup Q_1 \cup Q_2$ where $Q_1 = \{x : a \cdot x \geq 0, b \cdot x < 0\}$ and $Q_2 = \{x : b \cdot x \geq 0, a \cdot x < 0\}$ where we used $\lambda \in (0,1)$. Since $p \in Q_1, q \in Q_2$, as as we move from p to q along the segment, the signs of $a \cdot x$ and $b \cdot x$ each change (from nonnegative to negative, or vice versa). But this can only happen at P, which means that $\overline{pq}$ intersects P at some point z.

However as $p, q \in S$, by convexity, this means $z \in S$. This contradicts the fact that P does not intersect S.

Theorem .2. ***(Equivalence of rationalizability and dominance)*** *Given a finite, two player strategic game G, the set of (pure) strategies that remain after iterated elimination of strictly dominated strategies are precisely the strategies that are rationalizable in G.*

Proof: It is easy to show that any strictly dominated strategy is not rationalizable, so we focus on the reverse implication. As before, let player 1 have strategies $\{a_1, \ldots, a_m\}$ and we denote by $S^n = \{x \in \mathbb{R}^n \mid x_1 + \cdots + x_n = 1, x_i \geq 0 \; \forall i\}$ the $n-1$ dimensional simplex of probability vectors where n is the number of actions of player 2. Thus, a belief player 1 has about player 2 is some $q \in S^n$.

As in the framework of section 2, we can represent each pure strategy a_j as a linear function $f_j : S^n \to \mathbb{R}$, whose image is a restricted hyperplane in $\mathbb{R}^{n+1}$. Suppose the pure strategy a_i represented by $f_i = g$ is not rationalizable. Again by taking $f'_j = f_j - g$ for all j we may assume that g is identically 0 on S^n. For a slight abuse of notation, in the following we write a_i to mean the payoff vector $(a_{i,1}, \ldots, a_{i,n})$, so the expected payoff of playing the i-th action under belief q is $f_j(q) = a_i \cdot q$.

For a given f_j the set of beliefs q for which a_j yields strictly higher payoff than a_i are those that satisfy $f_j(q) = a_j \cdot q > 0$, ie. the intersection of an open half-space in $\mathbb{R}^n$ with S^n. Let this half-space be H_j.

If a_i is a never best response, then the union of these H_j over all $j \neq i$ must cover S^n entirely. Similarly, for a mixed strategy $\sigma = r_1 a_1 + \cdots + r_m a_m$, the beliefs for which σ yields higher payoff than a_i are those q satisfying $(r_1 a_1 + \cdots + r_m a_m) \cdot q > 0$, which is also a half-space H_σ. Then, σ strictly dominates a_i iff the half-space H_σ covers S^n.

The proof of the theorem follows from **Lemma 1**, which allows us to repeatedly replace a pair of (pure or mixed) strategies with a single mixed strategy, such that the union of the remaining half-spaces still covers S^n. This reduces the number of strategies while maintaining the covering invariant. Eventually, the process terminates with a single half-space H^* that covers S^n by itself, which corresponds to a mixed strategy σ that dominates a_i.

Indeed, suppose we have $m-1$ half spaces $p_j \cdot x < 0 \;\; (H_j)$ for $j \in [m] \setminus \{i\}$ in $\mathbb{R}^n$ each representing a pure strategy a_j, and their union covers the compact convex set of probability vectors S^n. We can now take two half spaces $p_1 \cdot x < 0$ and $p_2 \cdot x < 0$ and since $H_1 \cup H_2$ covers the convex compact set $S^n \setminus (H_3 \cup \cdots \cup H_m)$, there exists some λ such that the half space $(\lambda p_1 + (1-\lambda)p_2) \cdot x < 0$ covers $S^n \setminus (H_3 \cup \cdots \cup H_m)$, and we can replace H_1, H_2 with this halfspace instead. Thus, we have reduced the total number of half spaces while still covering S^n. Iterating this process, we eventually have a single half space of the form $\lambda_1 a_1 + \cdots + \lambda_m a_m < 0$ which covers S^n. Since we take convex combinations in each step, the weights are such that $\lambda_1 + \cdots + \lambda_m = 1$ with all $\lambda_i \geq 0$, so this defines a mixed strategy σ which strictly dominates a_i.

Theorem .3. ***(Radon)*** *Let P be a set of $n \geq d+2$ distinct points in $\mathbb{R}^d$. Then, P can be partitioned into two nonempty subsets V_1 and V_2 such that the convex hulls of V_1 and V_2 intersect.*

Proof: Consider any set $X = \{x_1, \ldots, x_{d+2}\} \subset \mathbb{R}^d$ of $d+2$ points.
Then there exists multipliers $a_1, \ldots, a_{d+2}$, not all zero, such that:

$$\sum_{i=1}^{d+2} a_i x_i = 0, \qquad \sum_{i=1}^{d+2} a_i = 0.$$

Take some particular nontrivial solution $a_1, \ldots, a_{d+2}$. Let $Y \subseteq X$ be the set of points with positive multipliers and $Z = X \setminus Y$ be the set of points with negative or zero multiplier. Observe that Y and Z forms a partition of X into two subsets with intersecting convex hulls.

Indeed, $H(Y)$ and $H(Z)$ necessarily intersect because they each contain the point

$$p = \sum_{x_i \in Y} \frac{a_i}{A} x_i = \sum_{x_i \in Z} \frac{-a_i}{A} x_i,$$

where $A = \sum_{x_i \in Y} a_i = -\sum_{x_i \in Z} a_i$. So we have expressed p as convex combination of points in Y and also as a convex combination of points in Z, which means that p lies in the intersection of the convex hulls.

Theorem .4. ***(Carathéodory)*** *If a point $x \in \mathbb{R}^d$ lies in the convex hull of a set P, then x can be written as the convex combination of at most d + 1 points in P. i.e. there is a subset $P' \subset P$ consisting of $d+1$ or fewer points such that $x \in H(P')$.*

Proof: Suppose x can be written as a convex combination of $k > d+1$ of the points $x_i \in P$. Without loss of generality, they are $x_1, \ldots, x_k$. Thus, we can write $x = \sum_{j=1}^{k} \lambda_j x_j$, where $0 < \lambda_j \leq 1$ and $\sum_{j=1}^{k} \lambda_j = 1$.

Then, consider $x_2 - x_1, \ldots, x_k - x_1$, which is a linearly dependent set, so there exists μ_j for $2 \leq j \leq k$ not all zero such that $\sum_{j=2}^{k} \mu_j (x_j - x_1) = 0$ and not all are zero. Then, if we set $\mu_1 = -\sum_{j=2}^{k} \mu_j$, this means that $\sum_{j=1}^{k} \mu_j x_j = 0$ as well as $\sum_{j=1}^{k} \mu_j = 0$, and not all of the μ_j are equal to zero, so there exists some $\mu_j > 0$.

Now let

$$\alpha = \min_{1 \leq j \leq k} \left\{ \frac{\lambda_j}{\mu_j} : \mu_j > 0 \right\} = \frac{\lambda_i}{\mu_i}.$$

Note that $\alpha > 0$. Then, we have

$$x = \sum_{j=1}^{k} \lambda_j x_j - \alpha \sum_{j=1}^{k} \mu_j x_j = \sum_{j=1}^{k} (\lambda_j - \alpha \mu_j) x_j.$$

Since $\alpha > 0$, then $\lambda_j - \alpha\mu_j \geq 0$ for all j.
Also,

$$\sum_{j=1}^{k} (\lambda_j - \alpha \mu_j) = \sum_{j=1}^{k} \lambda_j - \alpha \sum_{j=1}^{k} \mu_j = 1 - \alpha \cdot 0 = 1.$$

However, by definition, $\lambda_i - \alpha\mu_i = 0$. Thus, $x = \sum_{1 \leq j \leq k, j \neq i} (\lambda_j - \alpha\mu_j) x_j$, where every $\lambda_j - \alpha\mu_j$ is non-negative and their sum is one.

In other words, x is a convex combination of $k-1$ points in P. Repeating this process, we can write x as a convex combination of at most $d+1$ points in P, as desired.

Theorem .5. ***(Carathéodory's Theorem for Conical Hull)*** *Let $n \geq d+1$ and $v_1, \ldots, v_n$ be n distinct nonzero vectors in $\mathbb{R}^d$. Consider the conical hull formed by this set. Then, if a nonzero point x is in the conical hull, x can be written as the conical combination of at most d of the v_i.*

Proof: Suppose x is expressed as the conical combination of $k > d$ of the v_i, without loss of generality, $v_1, \ldots, v_k$. Then, $x = \sum_{j=1}^{k} \lambda_j v_j$, where $\lambda_j > 0$ for all $1 \leq j \leq k$. Since $k > d$, then $v_1, \ldots, v_k$ is linearly dependent, so there exists μ_j, not all equal to zero, such that $\sum_{j=1}^{k} \mu_j v_j = 0$. By multiplying each μ_j by -1 if necessary, we can ensure that at least one of these μ_j must be positive. Without loss of generality, we may also assume that $\sum_{j=1}^{k} \mu_j \geq 0$ because if the sum were less than zero (then there would exist some negative μ_j), we could multiply each μ_j by -1. Now, note that for any α, we have

$$\sum_{j=1}^{k} (\lambda_j - \alpha\mu_j) v_j = \sum_{j=1}^{k} \lambda_j v_j - \alpha \sum_{j=1}^{k} \mu_j v_j = x.$$

We want to choose some α such that $\lambda_i - \alpha\mu_i = 0$ for some i, and all $\lambda_j - \alpha\mu_j \geq 0$.

Indeed, if we let

$$\alpha = \min_{1 \leq j \leq k} \left\{ \frac{\lambda_j}{\mu_j} : \mu_j > 0 \right\} = \frac{\lambda_i}{\mu_i} > 0,$$

then we do have $\lambda_j - \alpha\mu_j \geq 0$ for all j. If $\mu_j \leq 0$, then since $\lambda_j > 0$, we have $\lambda_j - \alpha\lambda_j \geq 0$. And if $\mu_j > 0$, we have $0 < \alpha = \frac{\lambda_i}{\mu_i} \leq \frac{\lambda_j}{\mu_j}$, and again $\lambda_j - \alpha\lambda_j \geq 0$. Thus, $x = \sum_{j=1}^{k} (\lambda_j - \alpha\mu_j) v_j$ represents a conical combination using at most $k-1$ of the v_i, as all $\lambda_j - \alpha\mu_j \geq 0$, and $\lambda_i - \alpha\mu_i = 0$. We can repeat this process until we are left with at most d of the v_i, and so x is the conical combination of at most d of the v_i as desired.

Corollary .6. *Furthermore, if x also lies in the convex hull $H(O, v_1, \ldots, v_n)$, i.e. x is a convex combination of the v_i (and $O = 0$, so sum of coefficients of the v_i is less than or equal to one), then x can be written as the convex combination of at most d of the v_i, and O.*

Proof: If x lies in the convex hull $H(O = v_0, v_1, \ldots, v_k)$, that means that $x = \sum_{j=0}^{k} \lambda_j v_j$, where $\lambda_j > 0$ and $\sum_{j=0}^{k} \lambda_j = 1$. Since $O = 0$, this is the same as saying that $\sum_{j=1}^{k} \lambda_j \leq 1$. Again, if $k > d$, then we can write $\sum_{j=1}^{k} \mu_j v_j = 0$, where we can assume that $\sum_{j=1}^{k} \mu_j \geq 0$. But then in the preceding proof, we have that

$$\sum_{j=1}^{k} (\lambda_j - \alpha\mu_j) = \sum_{j=1}^{k} \lambda_j - \alpha \sum_{j=1}^{k} \mu_j \leq \sum_{j=1}^{k} \lambda_j \leq 1,$$

as $\alpha > 0$ and $\sum_{j=1}^{k} \mu_j \geq 0$. Again, recall that $\lambda_i - \alpha\mu_i = 0$, and each of the $\lambda_j - \alpha\mu_j \geq 0$.

Thus, x lies in the convex hull of $H(v_0, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$. Like before, by repeating this process, we can show that x lies in the convex hull H^* of at most d of the nonzero v_i, along with v_0, i.e. $x \in H(v_0, v_{j_1}, \ldots, v_{j_d})$.

How Sophisticated Are MLB Players?

Leo Oh

Rice University

Abstract— **Baseball showcases a dynamic between pitchers and batters, where both players must act optimally to maximize their payoff. We frame this interaction as an independent sequence of static simultaneous games, specifically a zero-sum game where the expected run values are used as utilities. This paper aims to determine whether the players learn by updating their belief on the opponent's probability of outcomes (PO) with two different models. The first model contains the non-conditional PO that does not use the first pitch as a prior, while the second model does the opposite and uses the conditional PO. We begin by calculating mixed strategy Nash equilibrium with linear programming but encounter pure strategies, which lead to zero-likelihood problems. Thus, we resort to quantal response equilibrium where human error is incorporated into the model, removing pure strategies. To interpret our results, we use the odds ratio from the two models' likelihood calculated using their equilibrium to compare models empirically. For our chosen group of players, the odds ratios show that most play closer to the sophisticated model's equilibrium, which holds on both aggregate and individual match-up levels. The following result implies that most players learn from the first count and adjust their actions in the second count, especially when experienced. However, limitations to the results include the assumption of homogeneity in players' performance against corresponding strategies regardless of their opponent and the differences in magnitude of the odds ratio, which affects the interpretation of our results.**

I. Introduction

In baseball, countless factors contribute to a team's win. At its core, however, baseball showcases the dynamic between pitchers and batters, where both players must decide on actions that maximize their payoff. Pitchers aim to deliver a pitch that minimizes the likelihood of the batter reaching base. Conversely, batters must decide when to swing based on which pitch offers them the best chance of successfully getting on base. Moreover, baseball is a game where the teams use various applications of statistics to analyze their opponents with play-by-play data. Players can decide their actions by combining these two elements depending on the expected payoff.

This interaction between a batter and a pitcher can be represented in a game theoretic framework. Suppose the players knew their opponents' strategies and payoffs. In that case, they may theoretically reach an equilibrium, where both players in a match-up choose the best response to the strategy of the opposing player (Tadelis, 2013).

For this paper, we assume that the batters are "guess hitters." In other words, the batters cannot react to a pitch once it is thrown. Bahill and LaRitz (1970) proved that batters are unable to keep their eyes on the ball for the entire trajectory unless they anticipate the pitch. We are not doubting batters' ability to spot obvious balls, especially at the MLB level. However, within the *shadow zone* (33% larger than strike zone: See Figure 3), we are suggesting it becomes more difficult for batters to identify the pitch and adjust their action accordingly.

Thus, we frame this interaction as an independent sequence of static simultaneous games, meaning players make their decisions at the same time with imperfect information. More specifically, we structure the match-up as a zero-sum game where the expected run values are used as utilities. We are particularly interested in the change in the probability of outcomes (PO) for the corresponding actions and counts that dictate expected run values.

In Figure 1, the left and right figure's x-axis indicate pitcher and batter's strategies, respectively. For the pitcher, the possible strategy is broken down into four zones and two speed classification which will be discussed in detail in Section III-B. For the given match-up, we see a clear difference in both players' strategy distributions between counts due to their change in utilities. We know that players play a different game on the second pitch, leading to a different equilibrium. However, we do not know how sophisticated these players are. This paper assumes that all players recognize the change in different outcome values between counts. The players should know that a strike at 1-0 is more valuable to the pitcher than at 0-0.

Instead, we investigate whether players adjust their actions according to a less noticeable change in conditional and non-conditional PO. Consider a scenario where the first pitch is low and called a ball. Batters are more likely to think the strike zone is higher on the second pitch. Thus, the batters may swing less at the same speed and location depending on the previous pitch, impacting the PO. Figure 2 displays PO at count 1-0. Both conditional and non-conditional cases show PO when the thrown pitch is in zone 1 and fast, as defined in Section III-B again. However, the conditional case has a second condition, the first pitch being in zone 4 and fast. We see a clear difference between the two distributions, ultimately affecting the expected run values. Thus, we see that the equilibrium may be different and check

I would like to acknowledge Dr. Mahmoud El-Gamal for his creative suggestions and ideas that allowed me to explore interdisciplinary areas in economics and sports. His influence made the experience of writing this paper truly enjoyable. Additionally, I am deeply grateful to my family, friends, and girlfriend. The journey of completing this thesis was demanding, yet their continuous support and encouraging words were a source of inspiration and motivation for me throughout this endeavor.

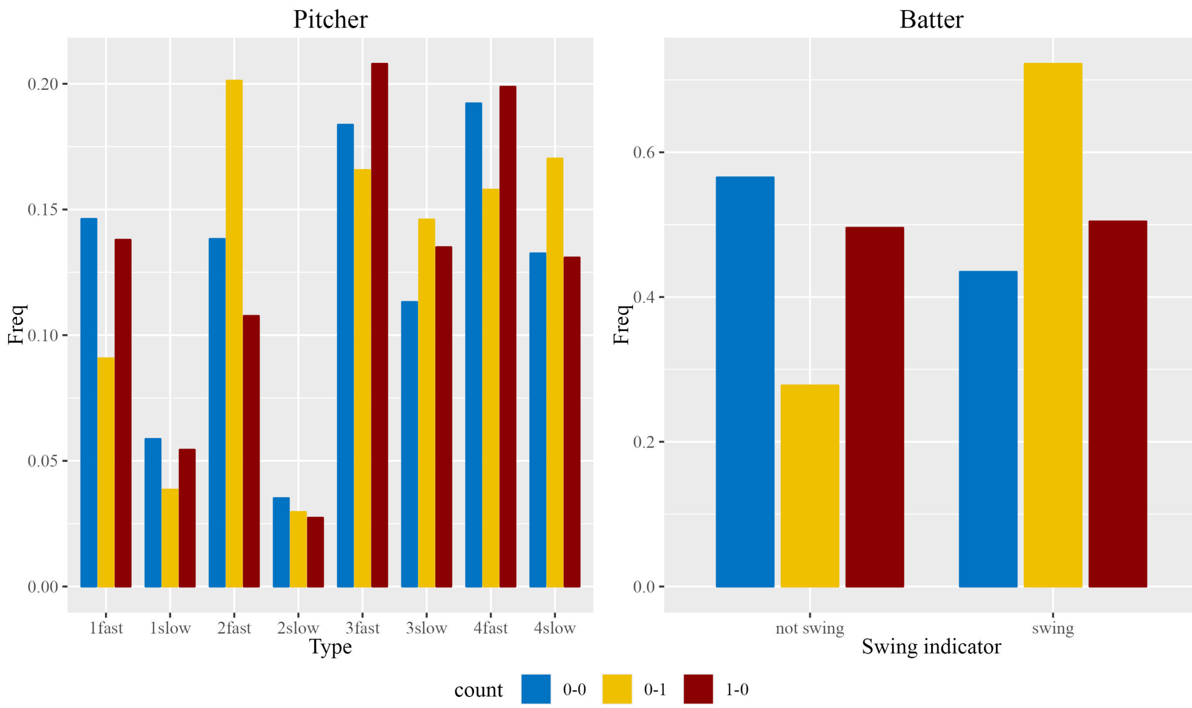


Fig. 1: Pitch and swing distributions between counts: Zach Greinke & Miguel Cabrera

which equilibrium the players adhere to.

The goal is to determine whether the players learn by updating their belief on the opponent's PO with two different models. The first model contains the non-conditional PO that does not use the first pitch as a prior. The second model does the opposite and uses the conditional PO. We first study the change in mixed strategy Nash equilibrium (MSNE) of games but run into the zero-likelihood problem when comparing the models due to pure strategies. Thus, we utilize logit Quantal Response Equilibrium (QRE) to solve for the bounded probability of optimal actions. Ultimately, we aim to determine if players are sophisticated enough to notice the change in PO between pitches by analyzing the odds ratio (*OR*) of the two models' equilibrium.

We also consider heterogeneity among players. While countless variables cause heterogeneity amongst players, we specifically examine the disparity in players' skills; some hitters may be adept at batting against certain types of pitches, whereas some pitchers may be more effective at throwing specific pitches. Since this is reflected in PO, players' ability will also affect the respective payoff matrices. By analyzing the empirical similarities of both aggregate and individual match-ups, we can validate the heterogeneity among players and determine which players exhibit greater sophistication in their skills.

Nevertheless, it is important to emphasize that addressing heterogeneity requires segmenting our dataset, leading to a reduction in sample size. In this study, we opted for a variable closely tied to the sophistication level of MLB players in our discussion. However, this decision assumes that the influence of other variables on the paper's results is minimal, as elaborated in IV-C.

This paper's structure is as follows. Chapter 2 of this paper will discuss previous work that has contributed to this research. Chapter 3 will discuss the methodology that explains the game framework, PO calculation, MSNE, and logit QRE optimization. Chapter 4 will detail the result from our *OR*s and discuss the implication of our empirical analysis. Chapter 5 will be our conclusion to this paper, summarizing the procedure and the results.

II. Literature Review

Douglas et al. (2023) developed a game-theoretic framework where the game was designed as a zero-sum simultaneous game. The batters' actions were binary, choosing to swing or not swing, while the pitchers' actions were complex, which included pitch types and locations. They calculated the game's utility based on on base percentage and used Neural Network to estimate PO, which determined the utility. They aimed to achieve empirical similarity of on base percentage by calculating the equilibrium of their designed game-theoretic frameworks.

Similar to Douglas et al., Melville (2023) had the same action sets for batters and players and utilized Neural Network when estimating PO. However, he diversified his games in three ways: simultaneous, sequential, and decision-point games. Further, he used expected run value, an outcome value that incorporates game context (e.g., count), for his

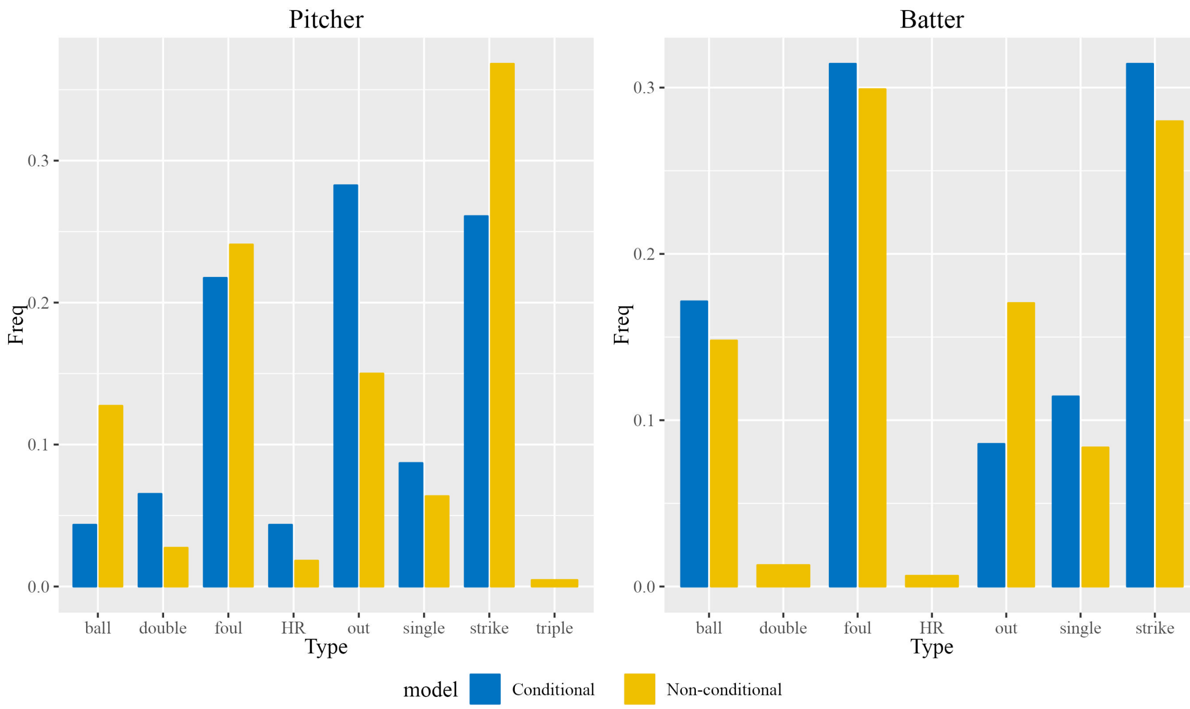


Fig. 2: Conditional and non-conditional PO: John Lackey & Ichiro Suzuki

utility. The main objective was to optimize pitch sequences through his empirical equilibrium.

In addition to these two papers, there are many papers with applications of game theory within the context of baseball (Flanagan, 1998; Kovash and Levitt, 2009; Choe and Kim, 2019). Flanagan showed that MSNE is a good predictor of proportions of right and left-handed batters. Kovash and Levitt showed that pitchers are not playing according to optimal mixed strategies as pitchers are throwing too many fastballs. Choe and Kim also showed that batters and pitchers are not playing optimal strategies and described the effect of players' contracts on players' strategies, where more extended contracts decrease players' incentive to play optimal strategies.

Aside from baseball, some papers explore the dynamic between two players with game theory in different sports. While there are noticeable differences, such as the sets of actions amongst players, the overall process of utilizing game-theoretic models to test a theory is similar.

In soccer, Chiappori et al. (2002) studied whether kickers and goalkeepers would conform to their mixed strategy Nash equilibrium in penalty kicks. However, they faced a challenge with few observations for any given pair, as penalty shootouts are generally rare. Thus, they resorted to aggregating a heterogeneous population, limiting their results' interpretation. Fortunately, we do not face such a problem in baseball, so we utilize aggregate and individual match-up data for our models to get different interpretations.

In tennis, Walker and Wooders (2001) have represented the dynamic between a server and receiver as a static simultaneous game and determined the MSNE in a predetermined framework. They showed that MSNE did a better job explaining the dynamic between experienced players and inexperienced players. However, Anderson et al. (2023) argued that a limited form of super-game exists due to *muscle memory effects* for each score state. Thus, they utilized dynamic programming to solve for optimal strategies dependent on the score state.

Our paper explores a similar concept and looks for a limited notion of super-game effect between a batter and a pitcher based on the count. To capture the full supergame effect, we would need to consider the memory of previous innings and matches at which a given pair had met, but we do not incorporate the effect of these variables in the models.

Regardless of the sport, we see that the research structure remains constant. The dynamic between two players is represented as zero-sum games where the MSNE is calculated using the Minimax theorem and linear programming. The papers above explore implications based on the discrepancy between models and actual strategy distributions.

For this paper, we aim to explore the learning mechanism of players between counts by comparing two game frameworks. We consider a scenario where batters and pitchers adjust their actions based on counts, meaning the utility will differ between pitches. While not resorting to dynamic programming as Anderson et al. did, we created utilities conditional on the previous pitch, capturing the baseball equivalent of *muscle memory effect*.

Melville also covers a similar topic as OptimusPitch, a Neural Network algorithm, incorporates previous pitches to estimate PO. However, due to the time constraints of this thesis, we decided against using machine learning algorithms such as Neural Networks. Instead, we take a historical frequentist approach when calculating PO. Since this method suffers from the exponential increase in dimensionality with conditioning-history length, we restrict our samples to the first two pitches.

The main difference between our research and previous works lies in our objective and the method of calculating the equilibrium for the proposed games. Since MSNE often results in pure strategies, there are game scenarios where batters are expected to swing with certainty, leading to the zero-likelihood problem. Thus, we introduce a concept similar to El-Gamal et al. (1993)'s *tremble*, where players have a chance of deviating away from their optimal equilibrium due to errors. McKelvey and Palfrey (1995)'s QRE models the probability of *trembles* more sophisticatedly with probability of error depending on expected loss in the case of error.

III. The Game and Equilibrium

In baseball, the team on offence can continue accumulating points as long as they have not accumulated three outs. This scoring opportunity arises from the dynamic between the pitcher and the batter. In a face-off between the two, the pitcher aims to get the batter out, while the batter aims to get on base. With each pitch thrown, the possible outcomes are an out, a hit, a strike (foul), or a ball. If the outcome is a strike or a ball, the interaction will continue until a batter gets out or on base. However, we only consider the first two pitches due to the curse of dimensionality, a problem where subsetting dataset leads to exponential loss in sample size. As a result, we narrowed down all possible scenarios into three counts: 0-0, 1-0, and 0-1. If dimensionality was not a problem, we can extend this analysis beyond the first two pitches using the same methodology.

In certain instances, pitchers intentionally walk batters by throwing four clear balls. This strategy is commonly used when a batter has been effective in the game, leading to the pitcher's reluctance to pitch directly to them. Instead, the pitcher opts to let the batter reach first base, avoiding the risk for a worse outcome. The data in subsection III-A does not include these scenarios.

A. Data

For this paper, we have utilized **[?]** to get play-by-play data for each player. The data spans from the 2008 to the 2023 season, meaning data for players who retired before 2008 are unavailable. Ideally, we would look for players who started their careers around 2008 and continued playing for a long time. We needed to find players who faced each other multiple times to separate them into individual match-ups when conducting empirical tests.

Therefore, we selected ten match-ups by the frequency of interactions, measured by the number of plays (or counts), between the two players. Given that we could not filter all match-ups from 2008, our criterion was to identify match-ups with a minimum of 150 counts. The match-ups we have chosen are as follows:

Pitcher	Batter	Number of counts
Trevor Bauer	José Abreu	183
Madison Bumgarner	Yasiel Puig	203
Gerrit Cole	J.D. Martinez	192
Jacob deGrom	Freddie Freeman	299
Zack Greinke	Miguel Cabrera	167
Clayton Kershaw	Paul Goldschimdt	284
John Lackey	Ichiro Suzuki	179
CC Sabathia	David Ortiz	255
Max Scherzer	Andrew McCutchen	187
Justin Verlander	Mike Trout	224

TABLE I: The ten match-ups

The data from Savant contain valuable information relevant to our study. For every count, it provides data on the ball's release speed and xy coordinates when it crosses the plate from the catchers' perspective, allowing us to categorize the type of pitches that pitchers could throw. Further, there is an indicator of whether the batter swung, allowing us to formulate batters' strategies. Lastly, Savant includes the outcomes of every count that is needed to calculate PO.

In total, we have three datasets that contain the elements that we have described above. The initial dataset comprises play-by-play data for ten pitchers' careers dating back to 2008, enabling us to compile their entire match-up history. The second dataset mirrors the first, but it pertains to batters. Lastly, the third dataset contains the precise play-by-play information for the match-ups defined in Table I. We will utilize the first two datasets to generate our payoff matrix, while the third dataset will serve as the test data for our empirical analysis.

B. Framework

We define pitchers' action space as $P = S \times L$, where $S \in \{\text{fast, slow}\}$ and $L \in \{1, 2, 3, 4\}$. We consider a pitch to be *fast* if it is above the pitcher's average release speed. L defines the pitch's location where the numbers represent the quadrants of possible pitch locations as seen in Figure 3. Thus, there are eight discrete choices that a pitcher can make. For batters, the action space is $B = \{\text{swing, not swing}\}$.

Within this framework, we calculate the utilities of players where $p \in P$ and $b \in B$:

$$U(p, b, \theta, c) = \mathbb{E}[R(\theta, c)|p, b] = \sum_{i}^{n} R(\theta_i, c) \cdot P(\theta_i|p, b) \tag{1}$$

In Equation 1, θ represents all possible outcomes in a given strategy, while c represents the current count. R is a run value associated with an outcome, which also depends on the count; a strike at 0-2 would be more valuable to the pitcher than at 0-0. In this paper, we utilize Melville's run values derived from run potentials in each scenario, meaning run values would increase if a batter is more likely to score a run after an outcome (see Table XIV in section). Since we are modeling the game as zero-sum, the pitcher's utilities are simply the negative of the batter's

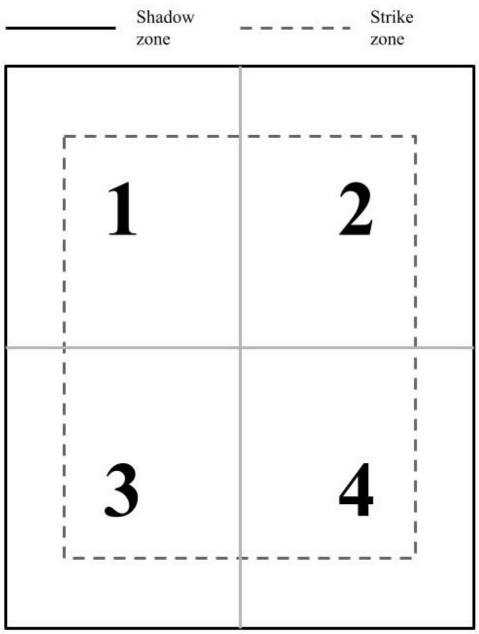


Fig. 3: Different zones and quadrants

For the PO, we take historical distributions of the players' actions by extracting play-by-play data from [**?**]. Under the condition that the batter does not swing, we utilize the pitcher's historical outcome to calculate the utility, as it relies solely on the pitcher to get the pitch within the strike zone. On the other hand, given that the batter does swing, we utilize the batter's historical outcome to calculate the utility and measure the batter's ability to hit different pitches. Since the pitcher's action space is restricted to release speed and location, we disregard other factors, such as the spin of a pitch.

Since we have defined the necessary components for building the two models, we can visualize our game with a payoff matrix as seen in Table II. The values represent the utilities of the pitcher, meaning the batter's would be the negative of the corresponding values. First, we notice that most values are positive, which confirms the common wisdom that the pitcher always holds an advantage regardless of the type of pitch thrown. Batters have multiple tries to get a hit to compensate for the disadvantage. In addition, we see that the values of utilities change between counts due to the change in both the run values and PO.

So far, we have discussed the components of the unsophisticated model. We will now discuss the components of the sophisticated model. For this model, the action space for both players remains the same, but the calculation of utility changes:

		Batter	
		$Swing$	$NotSwing$
Pitcher	$1fast$	0.0462	0.0176
	$1slow$	0.0048	0.0193
	$2fast$	0.0092	0.0160
	$2slow$	0.0058	0.0244
	$3fast$	0.0154	0.0251
	$3slow$	0.0206	0.0236
	$4fast$	0.0123	0.0125
	$4slow$	0.0053	0.0014

(a) Count: 0-0

		Batter	
		$Swing$	$NotSwing$
Pitcher	$1fast$	0.0272	0.0174
	$1slow$	0.0232	0.0209
	$2fast$	0.0152	0.0063
	$2slow$	0.0921	0.0286
	$3fast$	0.0358	0.0312
	$3slow$	0.0493	0.0126
	$4fast$	0.0064	0.0065
	$4slow$	0.0040	-0.0144

(b) Count: 1-0

TABLE II: Unsophisticated model payoff matrix: John Lackey vs Ichiro Suzuki

$$\begin{aligned} U^*(p, b, \theta, c, p_0) &= \mathbb{E}[R(\theta, c)|p, b, p_0] \\ &= \sum_{i}^{n} R(\theta_i, c) \cdot P(\theta_i|p, b, p_0) \end{aligned} \tag{2}$$

In Equation 2, p_0 represents the pitch that the pitcher threw at count 0-0. Thus, there are eight games for 1-0 and 0-1, separated by prior actions. The run values stay the same as they already change based on count. However, the PO would be different. Ultimately, the goal is to see whether the players recognize the change in the PO and play the corresponding equilibrium strategies.

Under the sophisticated model, the utilities at 0-0 remain the same but differ in second pitches as seen in Table III. Even though only one matrix is shown as an example for 1-0, there are eight payoff matrices with eight different equilibria for 1-0 and 0-1.

Lastly, we exclude pitch types from the payoff matrix that do not appear in specific match-ups. Imagine a situation where a batter has faced numerous pitchers throughout his career, contributing to a comprehensive payoff matrix. However, in a particular match-up against a specific pitcher, the pitcher may never use a certain pitch against this batter. Thus, we face a problem where we cannot learn anything about data that has never occurred in the sample due to zero out of zero frequency being undefined. As a result, we remove strategies that have never been played from the payoff matrix.

C. Mixed Strategy Nash Equilibrium

Under our zero-sum game framework, we can utilize the minimax theorem to prove the existence of MSNE. Addition-

		Batter	
		$Swing$	$NotSwing$
Pitcher	$1fast$	0.0462	0.0176
	$1slow$	0.0048	0.0193
	$2fast$	0.0092	0.0160
	$2slow$	0.0058	0.0244
	$3fast$	0.0154	0.0251
	$3slow$	0.0206	0.0236
	$4fast$	0.0123	0.0125
	$4slow$	0.0053	0.0014

(a) Count: 0-0

		Batter	
		$Swing$	$NotSwing$
Pitcher	$1fast$	0.0875	0.0147
	$1slow$	-0.0721	0.0252
	$2fast$	0.0065	0.0218
	$2slow$	0.0572	0.0387
	$3fast$	0.0455	0.0339
	$3slow$	0.0415	0.0058
	$4fast$	0.0663	-0.0022
	$4slow$	0.1085	-0.0002

(b) Count: 1-0 & p_0: $1fast$

TABLE III: Sophisticated model payoff matrix: John Lackey vs Ichiro Suzuki

ally, we can formulate this theorem into two linear programs that are duals to calculate MSNE with ease VonNeumann. Thus, we first define our MSNE as such Tadelis:

$$\mathbb{E}[U(\sigma_i^*, \sigma_{-i}^*)] \geq \mathbb{E}[U(\sigma_i, \sigma_{-i}^*)] \tag{3}$$

In this equation, σ_i and σ_{-i} represent the mixed strategy for the player i and the other, respectively. Since mixed strategies are probabilistic, $\sum_{s_i \in S_i} \sigma_i(s_i) = 1$ and $\sigma_i \geq 0$ where S_i is the strategy set for player i. Having defined our payoff matrix, we can use linear programs' duality to solve for mixed strategies for both players as such Bradley:

Primal:

$$\begin{aligned} \max \quad & W \\ \text{s.t.} \quad & W \leq \sum_i A_{ij} x_i, \quad \forall 1 \leq j \leq 2 \\ & \sum_i x_i \leq 1 \\ & -\sum_i x_i \leq -1 \\ & x_i \geq 0, \qquad \forall 1 \leq i \leq 8 \end{aligned} \tag{4}$$

Dual:

$$\begin{aligned} \min \quad & W \\ \text{s.t.} \quad & W \geq \sum_j A_{ij} y_j, \quad \forall 1 \leq i \leq 8 \\ & \sum_j y_j \leq 1 \\ & -\sum_j y_j \leq -1 \\ & y_j \geq 0, \qquad \forall 1 \leq j \leq 2 \end{aligned} \tag{5}$$

where W is the expected utility, A is the payoff matrix, and x_i & y_j are the mixed strategies for the pitcher and the batter, respectively (see 1 and 2 in Table).

	$W = 0.0232$
$1fast$	0.2525
$1slow$	0
$2fast$	0
$2slow$	0
$3fast$	0.7475
$3slow$	0
$4fast$	0
$4slow$	0
$swing$	0.1966
$not\ swing$	0.8034

(a) Count: 0-0

	$W = 0.03117$
$1fast$	0
$1slow$	0
$2fast$	0
$2slow$	0
$3fast$	1
$3slow$	0
$4fast$	0
$4slow$	0
$swing$	0
$not\ swing$	1

(b) Count: 1-0

TABLE IV: Unsophisticated model MSNE: John Lackey vs Ichiro Suzuki

	$W = 0.0232$
$1fast$	0.2525
$1slow$	0
$2fast$	0
$2slow$	0
$3fast$	0.7475
$3slow$	0
$4fast$	0
$4slow$	0
$swing$	0.1966
$not\ swing$	0.8034

(a) Count: 0-0

	$W = 0.0387$
$1fast$	0
$1slow$	0
$2fast$	0
$2slow$	1
$3fast$	0
$3slow$	0
$4fast$	0
$4slow$	0
$swing$	0
$not\ swing$	1

(b) Count: 1-0 & p_0: $1fast$

TABLE V: Sophisticated model MSNE: John Lackey vs Ichiro Suzuki

Table IV and Table V display the calculated values of x_i^* & y_j^*. We can confirm that the MSNE differs for the sophisticated (conditioning on first pitch) and unsophisticated (unconditional) models. Thus, We can now conduct empirical tests such as comparing the *OR* between models to see which equilibrium best represents the player's dynamic.

However, MSNE often result in pure strategies, meaning $\pi_i(s_i) = 1$. We see a pure strategy in Table IV and Table V, suggesting that pitcher and batter always take the same action. While this approach appears logical in theory, the expectation that players will perform optimally without any margin for error is unrealistic in practical scenarios. Further, it is impossible to measure the *OR* between models due to the zero-likelihood problem. If MSNE estimates an action to be played with a probability of zero but still appears in the sample, the likelihood would always converge to zero. Thus, we resort to QRE in subsection III-D.

D. Quantal Response Equilibrium

At the MLB level, we assume batters and pitchers choose their optimal strategies on average. However, as humans are prone to make mistakes, both players may play a non-optimal strategy, especially in a high-pressure situation. Thus, rather than using a deterministic pure strategy that MSNE produces, we calculate QRE, the noisy version of MSNE. It is crucial that QRE does not abandon the notion of Nash equilibrium and includes a statistical measurement of human errors. For this paper, we use a particular class of quantal response function, the logistic quantal response function, and its equilibrium is defined as such Mckelvey:

$$\pi_i(s_i, ...) = \frac{\exp(\lambda \mathbb{E}[U(s_i, \pi_{-i})])}{\sum_{s_i' \in S_i} \exp(\lambda \mathbb{E}[U(s_i', \pi_{-i})])}, \quad \forall i, s_i \in S_i \tag{6}$$

where λ can interpreted as the statistical measurement of error. $\lambda = 0$ means that players' actions are fully caused by error, while large λ indicates the opposite, where players choose optimal strategies with no error. Thus, if λ is given, we can solve a system of equations to get the QRE.

The problem lies in choosing the optimal λ for our context. Similar to [**?**], we utilize an empirical method, maximum likelihood estimation (MLE), to select λ^*. The log-likelihood functions for our two models are thus:

Unsophisticated:

$$LL_{NS} = \sum_j [\ln \pi_{i,NS}(x_i|\lambda_{NS}) + \ln \pi_{i,NS}(y_i|\lambda_{NS})] \tag{7}$$

$$\lambda^*_{NS} = \underset{\lambda_{NS} \in \Lambda}{arg\ max}\ LL_{NS} \tag{8}$$

Sophisticated:

$$LL_{S} = \sum_j [\ln \pi_{i,S}(x_i|\lambda_{S}, p_0) + \ln \pi_{i,S}(y_i|\lambda_{S}, p_0)] \tag{9}$$

$$\lambda^*_{S} = \underset{\lambda_{S} \in \Lambda}{arg\ max}\ LL_{S} \tag{10}$$

where x and y are the actual strategies that batters and pitchers played, respectively, from the play-by-play data, and j is the length of the data.

Table VI and Table VII show the QRE after λs have been estimated with MLE for the given pair. We see a clear difference from Table IV and Table V in the equilibrium, and the discrepancies indicate that there is an implied loss in expected payoff due to players' erroneous play. If players played without error ($\lambda = \infty$), Table VI would converge to Table IV and Table VII to Table V. However, as players make errors in their decision (λ gets closer to zero), the QRE begin deviating from MSNE as shown in Table VIII.

In sophisticated models with multiple games for each count, we can either estimate a universal λ^* applicable to all games within a specific count or calculate distinct λ_{p_0} values based on the prior pitch. Choosing a common λ implies that players exhibit a consistent error rate irrespective of the prior pitch, whereas opting for individual λ_{p_0} values suggests that players' error rates vary depending on the preceding pitch. Thus, we would utilize the exact maximum likelihood estimation as before but with eight variables now:

$$LL_{S,multi} = \sum_j [\ln \pi_{i,S}(x_i|\lambda_{S,p_0}, p_0) + \ln \pi_{i,S}(y_i|\lambda_{S,p_0}, p_0)] \tag{11}$$

$$\lambda^*_{S,p_0} = \underset{\lambda_{S,p_0} \in \Lambda}{arg\ max}\ LL_{S,multi} \tag{12}$$

[**?**] showed that batters use an estimate of pitch speed to predict pitch height with expectations formed by prior pitch information. The batters would predict how far a ball would drop and form the decision to swing. However, batters

	$\lambda^*_{NS} = 48.3409$
$1fast$	0.2420
$1slow$	0.0955
$2fast$	0.0977
$2slow$	0.1112
$3fast$	0.1417
$3slow$	0.1537
$4fast$	0.0962
$4slow$	0.0619
$swing$	0.4854
$not\ swing$	0.5146

(a) Count: 0-0

	$\lambda^*_{NS} = 3.3438$
$1fast$	0.1247
$1slow$	0.1246
$2fast$	0.1200
$2slow$	0.1412
$3fast$	0.1295
$3slow$	0.1282
$4fast$	0.1183
$4slow$	0.1137
$swing$	0.4842
$not\ swing$	0.5158

(b) Count: 1-0

TABLE VI: Unsophisticated model QRE: John Lackey vs Ichiro Suzuki

	$\lambda^*_S = \lambda^*_{NS} = 48.3409$
$1fast$	0.2420
$1slow$	0.0955
$2fast$	0.0977
$2slow$	0.1112
$3fast$	0.1417
$3slow$	0.1537
$4fast$	0.0962
$4slow$	0.0619
$swing$	0.4854
$not\ swing$	0.5146

(a) Count: 0-0

	$\lambda^*_S = 0.3769$
$1fast$	0.1260
$1slow$	0.1225
$2fast$	0.1243
$2slow$	0.1258
$3fast$	0.1255
$3slow$	0.1247
$4fast$	0.1251
$4slow$	0.1261
$swing$	0.4976
$not\ swing$	0.5024

(b) Count: 1-0 & p_0: $1fast$

TABLE VII: Sophisticated model QRE: John Lackey vs Ichiro Suzuki

count	MSNE payoff	QRE payoff	Implied loss
0-0	0.0232	0.0194	16.3793%
1-0	0.0312	0.0235	24.6795%

TABLE VIII: Unsophisticated MSNE and QRE model implied loss

often are not successful in making accurate predictions of ball trajectory and lead to large errors in perception of ball height when the pitch speed varies between counts Gray. Thus, suppose a pitcher opens up with a *fast* pitch but throws a *slow* pitch in the third or the fourth zone. The batter's optimal strategy may be to swing, but the misjudgement of pitch speed may lead them to think that balls may go past the *shadow zone*, causing them to not swing.

In count 1-0 of Table VI, a single λ value estimates the QRE as an unsophisticated model has one game per count. For Table VII, a single λ value estimates the QRE of eight games (one for each prior pitch), meaning the games would share the λ. Lastly, for Table IX, eight λ_{p_0} values are used to estimate the QRE of the corresponding games.

	$\lambda_S^* = \lambda_{NS}^* = 48.3409$
$1fast$	0.2420
$1slow$	0.0955
$2fast$	0.0977
$2slow$	0.1112
$3fast$	0.1417
$3slow$	0.1537
$4fast$	0.0962
$4slow$	0.0619
$swing$	0.4854
$not\ swing$	0.5146

(a) Count: 0-0

	$\lambda_{S,1fast}^* = 8.9104$
$1fast$	0.1442
$1slow$	0.0829
$2fast$	0.1098
$2slow$	0.1452
$3fast$	0.1355
$3slow$	0.1156
$4fast$	0.1220
$4slow$	0.1447
$swing$	0.4265
$not\ swing$	0.5735

(b) Count: 1-0 & p_0: $1fast$

TABLE IX: Sophisticated model QRE (multi-λ): John Lackey vs Ichiro Suzuki

With the estimated λ^*, we calculate for $\pi_i^*(s_i)$ and see that there are no equilibria where a strategy with probabilities of zero exist in Table VI, Table VII, and Table IX. Thus, we do not run into the zero-likelihood problem anymore and can now calculate the *OR* (see 3, 4, 5 in Table):

$$OR_{NS,S} = \exp(LL_S^* - LL_{NS}^*) \tag{13}$$

$$OR_{NS,S,multi} = \exp(LL_{S,multi}^* - LL_{NS}^*) \tag{14}$$

IV. Results

A. Aggregate

Table X shows the estimated λ^*'s and *OR* for the aggregate unsophisticated and sophisticated model. We see multiple 0.1 as λ values, the minimum threshold set in the optim function. As we mentioned in subsection III-D, λ values closer to 0 indicate that the players' actions are fully caused by errors and disregarding our theoretical optimal strategies completely. As a result, we label the count as *NA* if both λs are 0.1. Interpreting *OR* values do hold much meaning as both models' QREs are far from optimal strategy, leading *OR*s converging to one as a result. count 0-0 should always have *OR* of one as they use the same non-conditional model, but we will still label it *NA* if both λs are 0.1 for the reasons above.

count	λ_{NS}^*	λ_S^*	$OR_{NS,S}$
0-0	0.1	0.1	*NA*
1-0	0.1	3.0355	1.3519
0-1	7.0454	0.1	0.7118

TABLE X: $OR_{NS,S}$ of aggregate data

It also could mean that our way of calculating the payoff aligns differently with how players value specific outcomes. One major element contributing to this misalignment is the batters' swing frequency. In count 0-1 especially, many batters swing more than they should due to the added pressure of a strike. However, batters swinging in these situations lead to more outs and fouls, so the PO still favors them not swinging in these circumstances. Thus, when we aggregate the data, it is likely that the majority of batters fall into this category.

Lastly, players may try to be intentionally "unpredictable" to throw their opponents off their game or prevent them from playing their own side of the MSNE. Regardless, we can see how each batter and pitcher behaves differently in a match-up in subsection IV-B.

Nevertheless, Table X shows cases where λ_{NS}^* and λ_S^* values are different within the same count (excluding 0-0 case as both models use the same payoff matrix). Such outcomes are anticipated since a high λ value suggests that empirical data align more closely with a model's Nash equilibrium. Further, we see different patterns of λ between counts, suggesting that at a 1-0 count, empirical data tend to align better with the Nash equilibrium of the sophisticated model compared to the unsophisticated model, while at 0-1, the situation appears to be reversed. However, it is important to note that higher λ values do not guarantee a higher likelihood for a model, as the QRE is the determining factor for model likelihood.

Thus, we see an interesting change in $OR_{NS,S}$ values between counts. The unsophisticated model seems better at estimating when counts are 0-1, while the sophisticated model performs better when counts are 1-0. This result may be due to the difference in pressure on batters. [**?**] shares an interesting take regarding athletes' self-consciousness in high-pressure situations. The athletes start to question their own performance process instead of the task at hand, resulting in a decline in performance. Thus, batters or pitchers are more prone to make wrong decisions than usual as the situation becomes dire.

At 1-0, both players know that there is a cushion of strikes and balls, so there is no added pressure to force a decision, allowing them to be more "sophisticated." On the other hand, we see that players (most likely the batters) do not consider the first pitch's impact on the second pitch's outcome at 0-1. Thus, we can hypothesize the reason behind the values of $OR_{NS,S}$, but we can also see how the pattern of *OR* changes with different assumptions.

Table XI is the result of using different error estimates,

λ_{S,p_0}, dependent on prior pitch types. The eight λ_{S,p_0} vary significantly with implications of different error rates. Like the lower threshold of 0.1, there is also an upper threshold of 100 to keep denominators and numerators in Equation 6 below the maximum float that R can handle. Under the assumption that players make errors at a different rate, we see a different result in $OR_{NS,S,multi}$.

Table XI show different λ_{S,p_0} values within the same count, suggesting that players' error rates do depend on prior pitches. Further, with multiple λ^*_{S,p_0}, the $OR_{NS,S,multi}$ indicate that players are sophisticated learners in both counts. So far, however, we have discussed the results from aggregating players into a single payoff matrix. Using such a method suffers from generalizing the player population where heterogeneity may exist due to each pitcher and batter's unique skills and behavior. Thus, we resort to separating players into individual match-ups to resolve the heterogeneity of players' characteristics.

B. Individual Match-ups

Table XII separates the aggregated data into ten separate match-ups and shows the QRE fit of using a single λ value. Consistent from before, we see that λ_{NS} and λ_S values are different. However, we can also see that the players have different λ_{NS} and λ_S values. Since each player has different playoff matrix, some play closer to the Nash equilibrium than others. Thus, we can identify which players play closer to our defined model in different counts as λ values differ again between counts in individual match-ups

We also see different patterns of $OR_{NS,S}$ values in the match-ups. Seven out of ten match-ups show values greater than one for 1-0 and less than one for 0-1, consistent with the aggregate data. However, there are three match-ups where $OR_{NS,S}$ are either greater than or less than one in both pitch counts. For Jacob deGrom and Freddie Freeman, $OR_{NS,S}$ is greater than one for both counts, which could mean that these players are more immune to situational pressure and can stay sophisticated in how they play. However, players such as Gerrit Cole and J.D. Martinez may be the opposite as values for $OR_{NS,S}$ are less than one in both counts. The differing $OR_{NS,S}$ confirms heterogeneity in players' ability to process prior pitches' information.

Table XIII show the results when using multiple λ^*_{S,p_0} to estimate models for individual match-ups. Similar to Table XII, there are differing values of λ_{NS} and λ_{S,p_0} between counts, confirming differing error rates between counts. Further, we see that values of λ_{S,p_0} are different between players once again, showing the heterogeneity amongst players in terms of their ability to withstand pressure and play in a sophisticated way. Lastly, we can also see that the values of λ_{S,p_0} within the same count are different, meaning players have different rates of error for each prior pitch, which is consistent with the result of the aggregate model.

However, the major difference between Table XII and Table XIII are in $OR_{NS,S,multi}$. Nine out of ten match-ups show values greater than one for $OR_{NS,S,multi}$ in both counts, meaning players can take prior pitch into their strategy regardless of the count. The match-up between Zack Greinke and Miguel Cabrera is the only case where $OR_{NS,S,multi}$ is not greater than one in both counts. Ultimately, the result indicates that most players are sophisticated enough to learn from the prior pitch to recognize the change in the outcome of the second pitch regardless of the count.

In short, our findings vary based on the assumptions made about players' error rates. Assuming that players maintain a consistent error rate irrespective of the initial pitch, our results suggest that players exhibit less sophisticated gameplay at a 1-0 count. Conversely, if we consider players' error rates as influenced by the first pitch, it appears that players play our defined version of sophisticated way regardless of the count.

C. Limitation of Our Results

While utilizing zero-sum games as our framework provides us with many advantages, the zero-sum assumption may oversimplify the complexity of the game. While individual match-ups may seem zero-sum, baseball is a team sport with cooperative elements. Team strategies, fielding, and base running may introduce elements that go beyond the zero-sum framework. Additionally, since we are treating each sequence of games to be independent, we assume that players are not changing their behaviors over time within the same match-up. However, they may change their learning process throughout their match-up history.

When we break down aggregate data into specific player match-ups, we address the variability among players. However, our models do not account for situational differences that could affect the game, such as rivalries, weather conditions, and other unique aspects of each match. Additionally, we assume that players' performance remains constant regardless of their opponents. Our payoff matrix construction relies on players' PO drawn from their entire career data since 2008. Therefore, we anticipate that players will strategize based on these historical outcomes. This approach implies that we overlook the potential for players to differ in their PO in different match-ups. For example, we assume Ichiro Suzuki's historical performance remains the same when he faces $1fast$ from John Lackey as it does against all other pitchers he has encountered throughout his career.

Further, the result of most players being sophisticated learners in both counts in Table XI and Table XIII may be due to selection bias. As mentioned previously in subsection III-B, our selection criteria for match-ups required a minimum of 150 interactions between the two players. This criterion suggests that both players have had sufficiently long careers to encounter each other numerous times. Consequently, this indicates that the players involved in these match-ups are likely more skilled than the average MLB, given their longevity and frequency of play. Thus, the result can not be generalized to the MLB as a whole.

Finally, the magnitude of *OR*s matter, as they indicate the extent to which the sophisticated model is more or less likely compared to the unsophisticated model. For example, $OR_{NS,S,multi}$ of 3300.8986 for Kershaw and Goldschmidt

count	λ^*_{NS}	$\lambda^*_{S,p_0} \in (\lambda^*_{S,1fast}, \lambda^*_{S,1slow}, ..., \lambda^*_{S,4slow})$	$OR_{NS,S,multi}$
0-0	0.1	0.1	*NA*
1-0	0.1	(24.9427, 39.6840, 0.1, 0.1, 1.9390, 0.1, 0.1, 9.0096)	648.0060
0-1	7.0454	(0.1, 0.1, 0.5045, 4.6753, 89.4736, 0.1, 33.3067, 4.6454)	31654.9559

TABLE XI: $OR_{NS,S,multi}$ of aggregate data

pitcher	*batter*	count	λ^*_{NS}	λ^*_S	$OR_{NS,S}$
Lackey, John	Suzuki, Ichiro	0-0	48.3409	48.3409	1.0
Lackey, John	Suzuki, Ichiro	1-0	3.3438	0.3769	1.4260
Lackey, John	Suzuki, Ichiro	0-1	42.2116	8.7889	0.1989
Greinke, Zack	Cabrera, Miguel	0-0	0.1	0.1	*NA*
Greinke, Zack	Cabrera, Miguel	1-0	0.1	0.1	*NA*
Greinke, Zack	Cabrera, Miguel	0-1	32.9493	1.8941	0.1565
Kershaw, Clayton	Goldschmidt, Paul	0-0	0.1	0.1	*NA*
Kershaw, Clayton	Goldschmidt, Paul	1-0	0.1	1.6326	1.0732
Kershaw, Clayton	Goldschmidt, Paul	0-1	0.1	0.1	*NA*
Scherzer, Max	McCutchen, Andrew	0-0	0.1	0.1	*NA*
Scherzer, Max	McCutchen, Andrew	1-0	0.1	15.9321	2.9053
Scherzer, Max	McCutchen, Andrew	0-1	6.6504	0.1	0.8222
Verlander, Justin	Trout, Mike	0-0	0.1	0.1	*NA*
Verlander, Justin	Trout, Mike	1-0	2.1726	1.8668	1.0297
Verlander, Justin	Trout, Mike	0-1	0.1	0.1	*NA*
Bumgarner, Madison	Puig, Yasiel	0-0	26.4309	26.4309	1.0
Bumgarner, Madison	Puig, Yasiel	1-0	37.3507	4.9403	0.0367
Bumgarner, Madison	Puig, Yasiel	0-1	0.1	0.1	*NA*
deGrom, Jacob	Freeman, Freddie	0-0	0.1	0.1	*NA*
deGrom, Jacob	Freeman, Freddie	1-0	0.1	6.9191	4.9998
deGrom, Jacob	Freeman, Freddie	0-1	9.6379	8.7926	4.2684
Cole, Gerrit	Martinez, J.D.	0-0	0.1	0.1	*NA*
Cole, Gerrit	Martinez, J.D.	1-0	1.3822	0.3069	0.9991
Cole, Gerrit	Martinez, J.D.	0-1	9.4516	1.1047	0.6285
Bauer, Trevor	Abreu, José	0-0	0.1	0.1	*NA*
Bauer, Trevor	Abreu, José	1-0	0.1	0.6586	1.3733
Bauer, Trevor	Abreu, José	0-1	39.1592	5.3637	0.7327
Sabathia, CC	Ortiz, David	0-0	0.1	0.1	*NA*
Sabathia, CC	Ortiz, David	1-0	0.1	0.5573	1.1759
Sabathia, CC	Ortiz, David	0-1	0.1	0.1	*NA*

TABLE XII: $OR_{NS,S}$ of individual match-ups

in Table XIII indicate that the sophisticated model is around 3300 times more likely than the unsophisticated model. On the other hand, $OR_{NS,S,multi}$ of 1.0750 for Sabathia and Ortiz in the same table indicate the sophisticated model is only 1.07 times more likely than the unsophisticated model. Thus, there are varying levels of how much a model outperforms the other.

V. Conclusion

This paper examined whether the players can recognize the change in PO conditioned on the prior pitch. To test this, we created two versions of game-theoretic models. Both models determined player payoffs using the expected run value and PO. However, their approach to calculating PO differed: The simpler model utilized an unconditioned PO, whereas the more complex model utilized a conditioned PO. The goal

pitcher	*batter*	count	λ^*_{NS}	$\lambda^*_{S,p0} \in (\lambda_{S,1fast}, \lambda_{S,1slow}, \ldots, \lambda_{S,4slow})$	$OR_{NS,S,multi}$
Lackey, John	Suzuki, Ichiro	0-0	48.3409	48.3409	1.0
Lackey, John	Suzuki, Ichiro	1-0	3.3438	(8.9104, 3.8895, 0.1, 12.2918, 6.0298, 0.1, 85.8041, 0.1)	9.8404
Lackey, John	Suzuki, Ichiro	0-1	42.2116	(7.5563, 0.1, 25.4955, 10, 16.7723, 0.1, 0.1, 10)	1.2770
Greinke, Zack	Cabrera, Miguel	0-0	0.1	0.1	*NA*
Greinke, Zack	Cabrera, Miguel	1-0	0.1	(0.1, 10, 0.1, 98.6016, 10, 3.2301, 0.1, 0.1)	3.4893
Greinke, Zack	Cabrera, Miguel	0-1	32.9493	(15.1464, 0.4396, 10, 0.1, 0.1, 10, 0.1, 6.8115)	0.3562
Kershaw, Clayton	Goldschmidt, Paul	0-0	0.1	0.1	*NA*
Kershaw, Clayton	Goldschmidt, Paul	1-0	0.1	(16.2664, 10, 100, 0.1, 0.1, 69.3419, 58.5698, 4.5883)	3300.8986
Kershaw, Clayton	Goldschmidt, Paul	0-1	0.1	(0.1, 0.1, 0.1, 24.4230, 0.1, 100, 0.1, 50.8908)	47.7820
Scherzer, Max	McCutchen, Andrew	0-0	0.1	0.1	*NA*
Scherzer, Max	McCutchen, Andrew	1-0	0.1	(32.9205, 21.8861, 33.5700, 10, 10, 10, 13.9356, 0.1)	17.0508
Scherzer, Max	McCutchen, Andrew	0-1	6.6504	(0.1, 12.5991, 0.1, 0.1, 0.1, 1.3765, 6.2228, 14.5466)	2.3210
Verlander, Justin	Trout, Mike	0-0	0.1	0.1	*NA*
Verlander, Justin	Trout, Mike	1-0	2.1726	(20.3959, 10, 0.1, 10, 0.1, 10, 100, 0.1)	6.3449
Verlander, Justin	Trout, Mike	0-1	0.1	(0.1, 10, 0.1, 0.3056, 20.7593, 10, 0.1, 0.1)	1.9130
Bumgarner, Madison	Puig, Yasiel	0-0	26.4309	26.4309	1.0
Bumgarner, Madison	Puig, Yasiel	1-0	37.3507	(5.4137, 0.1, 30.4060, 12.4564, 0.1, 33.8627, 100, 10)	2.7370
Bumgarner, Madison	Puig, Yasiel	0-1	0.1	(1.5867, 10, 0.1, 0.1, 0.1, 17.7683, 0.1, 0.2994)	2.6203
deGrom, Jacob	Freeman, Freddie	0-0	0.1	0.1	*NA*
deGrom, Jacob	Freeman, Freddie	1-0	0.1	(3.7353, 51.2013, 0.1, 24.1301, 0.1, 0.1, 0.1, 29.0615)	186.5501
deGrom, Jacob	Freeman, Freddie	0-1	9.6379	(0.1, 21.6391, 4.0639, 8.2194, 7.4019, 23.1987, 4.3154, 17.9117)	18.4424
Cole, Gerrit	Martinez, J.D.	0-0	0.1	0.1	*NA*
Cole, Gerrit	Martinez, J.D.	1-0	1.3822	(10, 100, 35.2833, 0.1, 10, 10, 2.5785, 0.1)	7.6107
Cole, Gerrit	Martinez, J.D.	0-1	9.4516	(74.8631, 1.2223, 10.6876, 0.1, 0.1, 15.8100, 1.1787, 0.1)	10.1635
Bauer, Trevor	Abreu, José	0-0	0.1	0.1	*NA*
Bauer, Trevor	Abreu, José	1-0	0.1	(1.3724, 30.7924, 100, 0.1, 10, 10, 0.1, 0.5551)	22.0002
Bauer, Trevor	Abreu, José	0-1	39.1592	(5.1780, 10.5742, 0.1, 2.4420, 10, 0.1, 11.3606, 36.0132)	4.1146
Sabathia, CC	Ortiz, David	0-0	0.1	0.1	*NA*
Sabathia, CC	Ortiz, David	1-0	0.1	(100, 0.1, 0.1, 27.1752, 6.6407, 7.3448, 18.4853, 0.1)	197.7592
Sabathia, CC	Ortiz, David	0-1	0.1	(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 4.6510, 0.1)	1.0750

TABLE XIII: $OR_{NS,S,multi}$ of individual match-ups

was to compare the two model's equilibrium with their *OR*.

However, we realized that MSNE resulted in mixed strategies played with zero probability, resulting in a zero-likelihood problem when calculating *OR*. Thus, we utilized QRE, allowing players to make errors and deviate from optimal choice. As a result, we removed the possibility of equilibrium with mixed strategies of zero probability. Calculating QRE involved estimating λ, which measured error. We also considered the possibility of players having different error rates depending on the prior pitch by estimating separate λ for each situation.

When we aggregated our data, we showed that players' sophistication levels depended on the count. However, when we allowed different error rates in the second pitch QRE, the result showed that players were sophisticated regardless of the count. Separating the data into individual match-ups displayed similar results. However, it also gave us an insight into which players were sophisticated under different assumptions and confirmed that there was heterogeneity among players due to the difference in skill level and characteristics.

References

[1] Anderson, A., Rosen, J., Rust, J., and Wong, K.-p. (2023). Disequilibrium play in tennis. http://dx.doi.org/10.2139/ssrn.4383716, last accessed on 11/19/23.

[2] Bahill, A. T. and Karnavas, W. (1993). The perceptual illusion of baseball's rising fastball and breaking curveball. *Journal of Experimental Psychology-human Perception and Performance - J EXP PSYCHOL-HUM PERCEP PERF*, 19:3–14.

[3] Bahill, A. T. and LaRitz, T. (1970). Why can't batters keep their eyes on the ball? *American Scientist*, 72:249–253.

[4] Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of personality and social psychology*, 46(3):610–620.

[5] Bradley, S., Hax, A., and Magnanti, T. (1977). Applied *Mathematical Programming.* Addison-Wesley Publishing Company.

[6] Chiappori, P.-A., Levitt, S., and Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92(4):1138–1151.

[7] Choe, J. and Kim, J. S. (2019). Minimax after money-max: why major league baseball players do not follow optimal strategies. *Applied Economics*, 51(24):2591–2605.

[8] Douglas, C., Witt, E., Bendy, M., and Vorobeychik, Y. (2023). Computing an optimal pitching strategy in a baseball at-bat. *The International FLAIRS Conference Proceedings*, 36(1).

[9] El-Gamal, M. A., McKelvey, R. D., and Palfrey, T. R. (1993). A bayesian sequential experimental study of learning in games. *Journal of the American Statistical Association*, 88(422):428–435.

[10] Flanagan, T. (1998). Game theory and professional baseball: Mixed-strategy models. *Journal of Sport Behavior*, 21(2):121–138. Copyright - Copyright Journal of Sport Behavior, University of South Alabama Jun 1998; Last updated - 2023-09-11

[11] Gray, R. (2002). Behavior of college baseball players in a virtual batting task. *Journal of experimental psychology. Human perception and performance*, 28:1131–48.

[12] Kovash, K. and Levitt, S. D. (2009). Professionals do not play minimax: Evidence from major league baseball and the national football league. Working Paper 15347, National Bureau of Economic Research.

[13] McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

[14] Melville, W. (2023). A game theoretical approach to optimal pitch sequencing. https://scholarsarchive.byu.edu/etd/9910, last accessed on 11/19/23.
[15] Neumann, J. V. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, USA.
[16] Baseball Savant (2023). https://baseballsavant.mlb.com/
[17] Tadelis, S. (2013). Game Theory: An Introduction. *Princeton University Press,* Princeton, NJ
[18] Walker, M. and Wooders, J. (2001). Minimax play at wimbledon. *The American Economic Review*, 91(5):1521–1538.

Appendix

Count	Ball	Called Strike	Swinging Strike	Foul	Hit	Double	Triple	Home Run	Out
0-0	0.034	-0.043	-0.043	-0.043	0.494	0.79	1.068	1.407	-0.289
1-0	0.063	-0.05	-0.05	-0.05	0.46	0.756	1.034	1.373	-0.323
0-1	0.027	-0.062	-0.062	-0.062	0.537	0.832	1.11	1.45	-0.246

TABLE XIV: Expected run values of first two counts

Algorithm 1 MSNE solver: pitcher's side (primal)

Require: *p*, 8×2 payoff matrix
A ← transpose(*p*)
A_equality ← matrix(c(rep(1, ncol(*A*)), rep(-1, ncol(*A*))), nrow = 2, byrow = TRUE) ▷ adding $\sum_i x_i \leq 1$ and $-\sum_i x_i \leq 1$ constraints to A
A ← rbind(*A*, *A_equality*)
A ← cbind(c(-1, -1, 0, 0), *A*) ▷ adding W to A
B ← c(0, 0, 1, -1) ▷ specifying constraint values
C ← c(1, rep(0,ncol(*A*))) ▷ maximizing W, the first column of A
constraints_direction ← c(”¿=”, ”¿=”, ”¡=”, ”¡=”) ▷ primal constraint direction
p_optim ← lp(”max”, *C*, *A*, constraints_direction, *B*) ▷ lp includes $x_i \geq 0$ constraint automatically
Ensure: *p_optim$solution*

Algorithm 2 MSNE solver: batter's side (dual)

Require: *p*, 8×2 payoff matrix
A ← p ▷ the steps are identical to Algorithm 1, but it is a minimization problem with different A shape and constraint directions
C ← c(1, rep(0,ncol(*A*)))
A_equality ← matrix(c(rep(1, ncol(*A*)), rep(-1, ncol(*A*))), nrow = 2, byrow = TRUE)
A ← rbind(*A*, *A_equality*)
A ← cbind(c(rep(-1,nrow(*A*)-2), 0, 0), *A*)
B ← c(rep(0,nrow(*A*)-2), 1, -1)
C ← c(1, rep(0,ncol(*A*)))
constraints_direction ← c(rep(”¡=”,nrow(*A*)))
p_optim ← lp(”min”, *C*, *A*, constraints_direction, *B*)
Ensure: *p_optim$solution*

Algorithm 3 Likelihood generator for non-sophisticated QRE model

Require: (*balls*, *strikes*) = (0,0) ∨ (0,1) ∨ (1,0)
Require: *matchup*(*players*, *balls*, *strikes*), wanted match-up data; contains pitch type & swing indicator
Require: *NC*(*players*, *balls*, *strikes*), wanted payoff matrix
  if unique(*matchup$pitchtype*) $\neq$ unique(*NC$pitchtype*) **then**
    filter out rows from *NC* where the *pitchtype* is not found in *matchup*
  end if
  $FN \leftarrow \frac{\exp(\lambda_{NS}\mathbb{E}[U(s_i,\pi_{-i})])}{\sum_{s'_i \in S_i}\exp(\lambda_{NS}\mathbb{E}[U(s'_i,\pi_{-i})])}$ where $\mathbb{E}[U(s_i,\pi_{-i})] = NC[s_i,] \times \pi_{-i}$ ▷ a system of equations; $\pi = \mathrm{c}(x, y)$
  $LL_{NS} \leftarrow \sum_j [\ln \pi_{i,NS}(x_i|\lambda_{NS}) + \ln \pi_{i,NS}(y_i|\lambda_{NS})]$ ▷ x_i and y_i from *FN* and mapped to *matchup*
  $result \leftarrow$ optim(LL_{NS}) ▷ outputs $\lambda^*_{NS} = \underset{\lambda_{NS}\in\Lambda}{arg\ max}\ LL_{NS}$
  $final \leftarrow LL_{NS}(x^*, y^*, \lambda^*_{NS})$
Ensure: *final*

Algorithm 4 Likelihood generator for sophisticated QRE model

Require: (*balls*, *strikes*) = (0,1) ∨ (1,0)
Require: *matchup*(*players*, *balls*, *strikes*), wanted match-up data; contains pitch type & swing indicator
Require: *C*(*players*, *balls*, *strikes*), a list of containing 8 payoff matrix for each prior type
  FN ← an empty list
  for p_0 in 1:length(*C*) **do**
    if unique(*matchup$pitchtype*) $\neq$ unique(*C*[[p_0]]*$pitchtype*) **then**
      filter out rows from *C*[[p_0]] where the *pitchtype* is not found in *matchup*
    end if
    $FN[[p_0]] \leftarrow \frac{\exp(\lambda_S\mathbb{E}[U(s_i,\pi_{-i})])}{\sum_{s'_i \in S_i}\exp(\lambda_S\mathbb{E}[U(s'_i,\pi_{-i})])}$ where $\mathbb{E}[U(s_i,\pi_{-i})] = C[[p_0]][s_i,] \times \pi_{-i}$ ▷ a list containing systems of equations; $\pi = \mathrm{c}(x, y)$
  end for
  $LL_S \leftarrow \sum_j [\ln \pi_{i,S}(x_i|\lambda_S, p_0) + \ln \pi_{i,S}(y_i|\lambda_S, p_0)]$ ▷ x_i and y_i from *FN* and mapped to *matchup*
  $result \leftarrow$ optim(LL_S) ▷ outputs $\lambda^*_S = \underset{\lambda_S\in\Lambda}{arg\ max}\ LL_S$
  $final \leftarrow LL_S(x^*, y^*, \lambda^*_S)$
Ensure: *final*

Algorithm 5 Likelihood generator for sophisticated QRE model (multi-λ)

Require: (*balls*, *strikes*) = (0,1) ∨ (1,0)
Require: *matchup*(*players*, *balls*, *strikes*), wanted match-up data; contains pitch type & swing indicator
Require: *C*(*players*, *balls*, *strikes*), a list of containing 8 payoff matrix for each prior type
  FN ← an empty list
  for p_0 in 1:length(*C*) **do**
    if unique(*matchup$pitchtype*) $\neq$ unique(*C*[[p_0]]*$pitchtype*) **then**
      filter out rows from *C*[[p_0]] where the *pitchtype* is not found in *matchup*
    end if
    $FN[[p_0]] \leftarrow \frac{\exp(\lambda_{S,p_0}\mathbb{E}[U(s_i,\pi_{-i})])}{\sum_{s'_i \in S_i}\exp(\lambda_{S,p_0}\mathbb{E}[U(s'_i,\pi_{-i})])}$ where $\mathbb{E}[U(s_i,\pi_{-i})]) = C[[p_0]][s_i,] \times \pi_{-i}$ ▷ a list containing systems of equations; $\lambda_S = \mathrm{c}(\lambda_{S,1fast}, ..., \lambda_{S,4slow})$; $\pi = \mathrm{c}(x, y)$
  end for
  $LL_{S,multi} \leftarrow \sum_j [\ln \pi_{i,S}(x_i|\lambda_{S,p_0}, p_0) + \ln \pi_{i,S}(y_i|\lambda_{S,p_0}, p_0)]$ ▷ x_i and y_i from *FN* and mapped to *matchup*
  $result \leftarrow$ optim($LL_{S,multi}$) ▷ outputs $\lambda^*_{S,p_0} = \underset{\lambda_{S,p_0}\in\Lambda}{arg\ max}\ LL_{S,multi}$
  $final \leftarrow LL_{S,multi}(x^*, y^*, \lambda^*_S)$
Ensure: *final*