THE BONN JOURNAL OF ECONOMICS

Volume V(1) July 2016

DEPARTMENT OF ECONOMICS UNIVERSITY OF BONN

Legal Information

Publisher	The Bonn Journal of Economics Adenauerallee 24-42 53113 Bonn http://bje.uni-bonn.de bje@uni-bonn.de
V. i. S. d. P.	JProf. Tobias Berg University of Bonn
Place of Publication	Bonn, Germany
ISSN	2195-3449

THE BONN JOURNAL OF ECONOMICS

FOUNDED IN 2012

VOLUME 5

ISSUE 1

DEPARTMENT OF ECONOMICS
UNIVERSITY OF BONN

The Bonn Journal of Economics

Contents

Predicting Swiss Unemployment

Christoph Schwerdtfeger[1]

Individuals' job search outcomes[2] are inherently hard to predict from standard demographic data: There are copious amounts of unobserved properties of jobseekers and there is the undeniable influence of luck in the job search process—not to speak of macroeconomic influences like business cyclicality. On the other hand there is practical[3] as well as research[4] interest in *ex ante* prediction of likely outcomes for a newly unemployed person. A precise estimate of job search durations would allow for better treatment of individual jobseekers by the unemployment agency (such as assigning appropriate training). This could improve both the personal welfare of the jobseeker and aggregate welfare (since resources are allocated less inefficiently).
Predictions from a very noisy signal should utilize any information available, unearthing every pattern. The possibility of deriving inferential conclusions from the prediction process must appear remote, thus linearity or interpretability is not a requirement. This thesis applies strong machine learning, specifically support vector machines (**SVM**), to a focused prediction problem that originally stems from prior research but had to be adapted quite significantly.

1 Datasets and prior research

This thesis is based on two datasets. Both of them have been described in great detail by Fechner 2014. They are not publicly available since they contain highly personal information about individual jobseekers.

Swiss unemployment data The first dataset contains the universe of Swiss unemployment and job search data, including detailed demographics, past earnings, job search efforts, sanctions, trainings and other active labor market policy measures. The dataset is organized by unemployment spell.[5] It spans every spell that was registered between 2008 and 2014 in all

[1]Christoph Schwerdtfeger received his degree (B.Sc.) from the University of Bonn in 2015. The present article refers to his bachelor thesis under supervision of Prof. Dr. Hans Martin von Gaudecker, which has been submitted in July 2015.

[2]The primary outcome of interest is the duration of unemployment (in terms of being registered with the unemployment agency). This is the only outcome considered in this thesis.

[3]e.g. Matty 2011 for the UK, Riipinen 2007 for Finland and Rosholm, Svarer, and Hammer 2004 for Denmark.

[4]Hasluck 2008 gives a broad overview

[5]A per person identifier is available

of Switzerland (this amounts to approximately two million spells).

Jobchancen Barometer (JCB) data The second dataset originated from an experimental research design by Patrick Arni that was implemented in the Swiss canton Fribourg in 2012-2014: Caseworkers (**CW**) who administered the job seekers were asked to fill in a questionnaire after the first longer contact with a newly unemployed person. This (computer-based) questionnaire asked the CWs to assess soft facts such as *How realistic is the job seeker's earning expectation?*. It is available for 8606 spells in total.

1.1 JCB experiment

The research idea behind the JCB experiment (which has not yet been published) was to show CWs a computer generated prediction of the length of the unemployment spell of a newly unemployed person. The goal was to prove that a reliable prediction at the CWs disposal could counteract their well known tendency to underestimate the likelihood of long term unemployment. This prediction, called the "Jobchancen-Barometer", was shown to the CW by random assignment. Since prediction from demographics alone is very imprecise, additional information was collected by querying the CWs' impressions of job seekers. This information (and available demographics) was used to generate a prediction for the expected duration of unemployment in days.
The prediction shown during the experiment was not very precise - it was not significantly better than the CW's own predictions (they were asked to report their expectation for the unemployment duration before seeing the system's prediction). This was due to three facts: The prediction was trained in a first stage (the calibration phase) of four months from October 2012 to January 2013 on a sample $\mathbb{J}_C$ of 1200 observations (which were heavily censored), it was a point prediction and the impact of the inherent noisiness of job search.

2 Machine Learning Setup

This thesis contributes a novel approach for the prediction of individual unemployment outcomes. A threefold alteration to the prediction process will be made, trying to enhance the predictive quality as compared to the original research. The calculations will only utilize data that was available at the time the JCB's predictor was trained (1200 observations $\mathbb{J}_C$ plus historical demographic data $\mathbb{H}$), thus simulating the original predictor's training conditions.

2.1 Support Vector Machines

Support vector machines will be used. They, like similar machine learning algorithms, can lever nonlinear structure in a given dataset to a high degree and might be capable of extracting more information as compared to the original predictor that employed nearest neighbor propensity matching and multinomial logit models.

Support vector machines[6] are one of the go-to prediction algorithms in machine learning. They have been hugely successful in the 1990s and early 2000s for problems in speech and image recognition as well as for recommender systems and finance and are still one of the mainstays in supervised machine learning (i.e. learning from labeled training data). Only the recent development of deep learning with neural nets has led to a class of problems with millions of available data points (such as recognizing cat pictures on Facebook) that is tractable with these newer algorithms and not with SVMs (since their running time is quadratic in the sample size which renders learning from millions of examples infeasible). The constraints of this publication do not allow for a full discussion of SVM. See Schwerdtfeger 2015, Hastie et al. 2009, Neumann 1944 and Vladimir Naumovich Vapnik 1998 for details. What follows is a short and high level overview.
The guiding idea behind SVM is to search for an optimal hyperplane that separates input data $\mathbb{X}$ according to a property or labeling $\mathbb{Y}$: All points on one side of the plane have $\mathbb{Y}$, the points on the other side do not. $\mathbb{X}$ might be a stack of pictures, some of them show dogs, some of them don't (that is $\mathbb{Y}$). SVM searches for the best clean separation hyperplane B of the feature space (the pixels in each picture) to distinguish pictures with dogs from those without any. Once (and if) such a separation B is found a new picture x is analyzed purely in terms of where x is located relative to B?
SVM, like any supervised learning algorithm, consists of two parts: A learner that chooses the best classifier and that classifier itself. For SVM the classifier is B. To find B, a not-too-small set of examples called the training set $\mathbb{T}$ is fed into the learner and the learner returns B. The training is very computationally expensive, the prediction on the other hand is almost free.
Separabiltity by a hyperplane is a strong requirement and usually only trivial data will be linearly separable (For example: Is a Corgi more similar to a St. Bernhard or to a cat? How about dogs shown from the side?). Still SVM is one of the standard machine learning algorithms.
The reason lies in the specific formalization of "goodness" of hyperplanes that leads to an optimization problem that requires little knowledge of $\mathbb{X}$—all that is necessary to find the optimal hyperplane are inner products of vectors in the training data $\mathbb{T}$, in other words: their relative geometry. This opens the door for a powerful extension of the method, the kernel trick. This tool allows SVM to borrow the geometric intricacies of a highly complex space $\mathscr{H}$ to get a better description of the training data without paying the computational price of explicitly visiting $\mathscr{H}$. The resulting decision boundary is both simple and complex: it is a hyperplane in $\mathscr{H}$, but its projection into the feature space $\mathbb{X}$ is nonlinear.

2.2 Binary outcome

Point predictions are, intuitively speaking, much harder than binary or categorical predictions—at least as long as there is no assumed underlying functional form. Additionally, many machine learning algorithms are better suited for the classification of binary or categorical outcomes; this is

[6] As introduced in Boser, Guyon, and Vladimir N Vapnik 1992 and Cortes and V. Vapnik 1995

especially true for SVM.
For the JCB experiment there is a natural reduction to a binary outcome which is aligned with the original research interest: The CWs tend to underestimate the likelihood of extreme unemployment outcomes, both for very short and very long durations. This is consequential as they might choose a different course of action (e.g. training measures, consultation density or required job applications per month) for jobseekers at risk of long term unemployment. The binary[7] variable of choice: long term unemployment, encoded by $\mathbb{Y} = \{-1,1\}$ where $\mathbb{Y} = 1$ corresponds to unemployment for more than 360 calendar days. An enhanced experimental setup could warn the CWs only if they do not predict long term unemployment themselves while the system does, thus rarifying the signal and presumably increasing its importance to the CWs.
The reduction to binary outcomes sacrifices direct comparability to the JCB's original predictor, but an adapted and enhanced version with a binary target is available.

2.3 Censoring

The reduction to a binary outcome does not solve the more pressing matter with the data: At the time of the original calibration most observations in $\mathbb{J}_C$ were censored. The job seekers were still unemployed and, thus, the real duration of unemployment was not known. In fact, at the time of original training, more than 90% of spells were censored. The original idea for this thesis was to develop an additional mechanism to simulate and impute outcomes, generating a two stage process: One, take historical spells with demographic information, find similar spells therein and train an arsenal of SVMs to impute the outcomes for the samples in $\mathbb{J}_C$. And two, use $\mathbb{J}_C$ with augmented outcomes to train a second stage SVM with demographic and CW generated data to predict outcomes for spells from the second phase of the JCB—this would have simulated the original experiment's conditions and, thus, any gain in prediction quality would have been available for similar future experiments.
The realization of the extent of outcome censoring in $\mathbb{J}_C$ must lead to the conclusion that this endeavor (and thus the original research proposal) is basically futile: 90% of outcomes in $\mathbb{J}_C$ did not have a realized outcome at the time the original experiment trained the predictor for the experiment's second phase. Imputing 90% of spells (plus the fact that all realized outcomes are mechanically "short term") implies that the second stage cannot outperform the first stage. Even if the JCB questionnaire generated very rich additional information the second stage could, essentially, only reproduce the first stage—if its generalization performed flawlessly. Then it has to be asked: Why not use the first stage on all participants of the JCB? Doing so eliminates the need to fill in the questionnaire without sacrificing any predictive quality.
Furthermore building a SVM with perfect outcome information (utilizing the power of hindsight)

[7]It would be possible to predict interval membership such as $\{(0,90],(90,180],(180,360](360,\infty)\}$ (in days). This would reduce accuracy levels and increase computational cost to train a classifier quadratically in the number of intervals. Accuracy is paramount and, from a welfare perspective, long term unemployment is more pressing. Interval setups are methodically identical, the `R` code as developed is fully prepared for such a design

on $\mathbb{J}_C$ (thus simulating a perfect first stage prediction) shows that the sample size is too small: Using this SVM to predict outcomes for the second phase data of the JCB yields $\kappa \approx 0.21$.[8] A machine $\mathscr{G}$ trained on all new spells registered in Switzerland during the calibration phase without any adaptations returns a prediction with $\kappa \approx 0.25$. The size of $\mathbb{J}_C$ is decidedly too small to train a useful SVM on it.

This realization has to lead to the **abandonment** of the original *two-stage approach*.

Still, the programming code for the second stage has been fully developed: One research idea had been to test what percentage of realized outcomes and sample size would be necessary to gain an advantage from the additional "soft" information by artificially extending the calibration phase into the original experiment's second phase. It turns out: Even the maximized JCB sample (the complete first and second stage) does not gain much when compared to demographics alone—this is in line with research from Fechner 2014.

The first stage remains a well posed (and inherently interesting) problem. $\mathbb{J}_C$ can be viewed as an average regionally and temporally restricted subsample of unemployment spells in Switzerland (under the assumption that the addition of the JCB survey to the unemployment registration process did not fundamentally alter job search outcomes). A conventional econometric JCB outcome imputation for binary outcomes that can serve as a benchmark has been calculated. It utilizes logistic regressions and propensity score matching and is generated from the same data sets as the first stage. This prediction will be used to (broadly) compare the quality of prediction that is achievable with classical econometric methods to that of non-linear machine learning techniques.

3 A composite SVM classifier

Sticking with $\mathbb{J}_C$ as the reference and target sample a set of SVMs will be constructed from historical spells with high similarity to subsets of $\mathbb{J}_C$.

3.1 Theoretical design

The spells in $\mathbb{J}_C$ are grouped into `biweekly` intervals for the time from the date of registration to the end of the calibration phase. Therefore the information that a given spell in the JCB sample already has a positive duration associated with it at the time of censoring is not lost (it is much more likely for x to become long-term unemployed given x is unemployed for more than four months already: roughly 40% compared to 25% at the spell's beginning). For every group a SVM predictor with similar historical spells is computed and used to predict the censored outcomes.

In more precise words: Take historical spells with demographic information (including past spells and earnings) $\mathbb{H}_x$ whose outcome is known (they are realized) and which are similar to a given $x \in \mathbb{J}_C$ and train a classification algorithm f on them and use the resulting classifier $\tilde{f}_{\mathbb{H}_x}$ to predict the outcome for x, yielding $y(x) = \tilde{f}_{\mathbb{H}_x}(x) \in \{-1,1\}$.

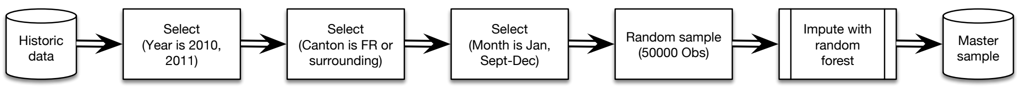
[8] A prediction quality measure that will be introduced in the next paragraph, higher values are better.

Similarity between $x \in \mathbb{J}_C$ and a subset $\mathbb{H}_x$ of the historical sample $\mathbb{H}$ is defined as those spells $h \in \mathbb{H}$ satisfying (1) h begins before February 2012, (2) the realized duration of h measured in `biweekly` intervals is at least as long as the (censored) duration for x and (3) h is in Fribourg or one of the surrounding cantons.

The sampling choice and similarity definition is made pragmatically and heuristically: `biweekly` similarity intervals are chosen because there are 10 of them in $\mathbb{J}_C$ and none of them is too sparsely populated. The end of the historic inflow window is chosen to ensure that every spell in it is realized (in terms of the binary outcome $\mathbb{Y}$)—there is no censoring. The restriction to Fribourg and surroundings assures some shielding from regional effects.

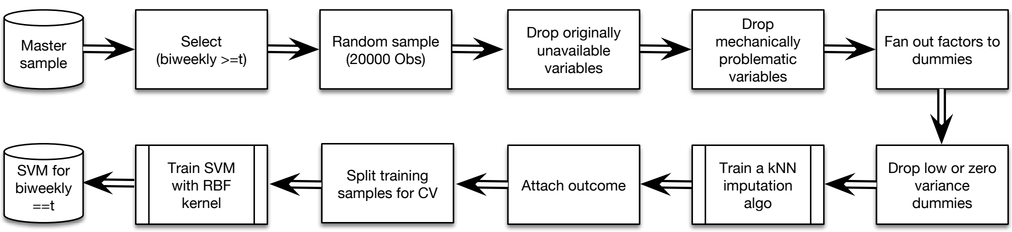
The historical datasets were prepared as follows: Begin with all spells in the relevant cantons. Choose those registered in the two years before February 2012. Take the subset of spells registered from September to December plus January (to match the second stage training sample) and draw 50000 spells randomly. This is the master dataset $\mathscr{D}$. $\mathscr{D}$ has missing values in some features, especially past earnings are missing in about 20% of the cases and had to be imputed.

Figure 1.1: Preparation of the master sample for the composite SVMs



For every interval a random sample of 20000 spells was drawn from $\mathscr{D}$ that matches the minimum `biweekly` duration condition.[9]

Figure 1.2: Training one of the 10 composite SVMs for time $t \in \{1,...,10\}$



3.2 Practical design

The practical design of the composite SVM setup has been quite demanding in terms of both theoretical construction and computational resources. Only the most important decisions are mentioned.

History similarity choice A number of history sample specifications are *ex ante* plausible, e.g. *Is a more recent sample more valuable than a seasonally matched sample?* The most efficient

[9]The size of 50000 was chosen to ensure that there are 20000 available for every `biweekly` interval. 20000 was chosen since this appears to be the threshold for reasonable feasibility in terms of computation time.

one had to be chosen via trial-and-error.

Imputation $\mathscr{D}$ has missing values in a number of entries, most prominently past earnings: 24% of spells do not have an entry. These had to be imputed since missing information is not independent of other properties of the jobseekers: A sample drawn only from complete spells is not perfectly random for the whole population. A random forest (cf. Murphy 2012) based method was used—the resulting SVMs showed surprising gains compared to nearest neighbor based imputation methods. This is not a trivial gain: The imputation process alone took about one 20h on a 64 core server.

Feature selection Feature design and selection is considered to be the hardest part of practical machine learning tasks since it requires both domain expertise and methodological awareness: Which variables are to be kept? How to structure the data for maximum accuracy without risking overfitting and without breaking the algorithms?

1. Most variables in demographic data are categorical. They were fanned out to one dummy per level as a first preparation stage. There is a tradeoff for ordinal variables, e.g. education: treating them as factors ignores the ordering information but enhances their comparability to other categories.
2. Some variables are problematic for domain specific reasons: The calendar year of the historic sample is misaligned with the JCB sample and has to be dropped. But: Seasonal information is valuable. Thus a feature of day of the year of registration is kept.
3. More problematic: The ID of the CW. This ID does carry tremendous information in the training sample as there is endogenous selection—specific CWs have specific cases assigned to them, they are specialized (and they also might differ in terms of talent and enthusiasm). What does it imply out-of-sample if a CW is not among the training CWs? This is unclear, as is the SVMs predictive reaction to such a case—the variable has to be dropped altogether.

3.3 Prediction quality measurement

Swiss job market data is not a balanced sample in terms of long-term unemployment outcomes: The majority of job seekers exits the system after less than a year (in a ratio of approximately 3:1). Hence predicting "not long-term" for every test sample will yield a naive accuracy rating of 75%.[10] Any good measure of predictive performance will have to take this imbalance into account. The most common measure to judge the quality of binary predictions in machine learning is ROC[11], an error measure that takes into account the tradeoff between errors of first and second type, thus

[10] A prediction is counted as "accurate" if it is right and "not accurate" else.

[11] cf. Nevin 1969. ROC is used synonymously with area under the ROC curve, AUROC.

implicitly compensating for unbalanced training sets. For a training set as severely unbalanced as Swiss job search data Cohen's κ (cf. Cohen 1968) is better suited. $\kappa := \frac{P[A]-P[E]}{1-P[E]}$ where $P[A]$ is the proportion of observed agreement and $P[E]$ is the proportion of expected agreement due to class size. κ is widely used in pharmaceutical research and explicitly accounts for unbalanced classes by calculating the relative accuracy of predictions on a sample corrected for class size. According to Landis and Koch 1977 a value of $\kappa \in [0,0.2]$ is considered slightly better than always choosing the modal outcome, $\kappa \in (0.2,0.4]$ is considered fair but not convincing. It is not a perfect measure as shown by Brennan and Prediger 1981 and its values for different underlying populations are not fully comparable, but it is well suited for hyperparameter choice within a class of predictors. Its utility when comparing the prediction quality on the same underlying dataset is limited: A higher κ is always better, but very different classifiers can have similar values of κ. In all optimizations (finding optimal hyperplane parameter values) κ was used as a target function.
A survey of corresponding unemployment duration prediction literature shows a surprising lack of awareness for the necessity of imbalance compensation: As an example, Rosholm, Svarer, and Hammer 2004 rely on naive accuracy to assess model performance even though the dataset is 60/40 unbalanced and compare their model's performance to models done in other countries without considering the possibility of differing underlying population. From the data given in Rosholm, Svarer, and Hammer 2004 it is possible to calculate a $\kappa \approx 0.26$ for their prediction.

4 Results

All results presented in this section are out-of-sample predictions generated on 10% of the training samples that were kept aside and did not figure in the training.
In terms of κ the 10 SVMs all perform roughly the same: they yield $\kappa \in [0.33,0.35]$. This value is considered "fair": It is better than blind guessing of the modal outcome class—but not by much. In terms of accuracy the optimal machine for the `biweekly==1` sample (i.e. there are no time-in-unemployment restrictions) shows an accuracy of 0.86 and a ROC value of 0.83. This is about five ROC percentage points better than previous results from England (see Matty 2011), even though their data contained survey results assessing personality traits of the jobseekers as well as demographic and other information. The English and Swiss job markets are certainly not comparable, but this indicates that the individual predictors work and that the machine learning approach might be capable of uncovering additional structure in the data.
The number of support vectors is nearly constant at 90 for all machines, almost independent of training sample size choices. This is a good indicator that the prediction is not in danger of overfitting.
All 10 classifiers severely underpredict long term outcomes, a pattern that is true for most well documented unemployment duration predictions in the literature as surveyed for this thesis. On average about 67% of those predicted to become long term unemployed will in fact be long term

unemployed while only 30% of real long-term unemployed are predicted to be. The second value, called *specificity*, could be considered the most important value in predicting unemployment outcomes: These are the people at risk a classifier rightly identifies as such (maximizing in terms of specificity is nonsensical, a predictor would always predict "long-term" given such a target function).

4.1 Over- and Undersampling

Low specificity of the "naive" SVMs stems from outcome class imbalance. Since the "short-term" class is significantly larger, a soft-margin SVM will tend to sacrifice precision on the smaller "long-term" class for gains on the larger.[12]
This behavior can be treated with oversampling: Adding synthetic samples interpolated within the underrepresented outcome class to the training set can be used to artificially level outcome class sizes.[13] Alternatively, dropping random observations from the larger class (undersampling) may lead to even better results if enough data is available as is the case with $\mathbb{H}$. This treatment skews the prediction algorithm towards a higher likelihood of prediction of the underrepresented outcome, thus trading errors of type I for errors of type II, false positives for false negatives.
As an example of oversampling, consider a naive SVM and an oversampled SVM (with a 50/50 split of long/short-term outcomes) trained on a 10000 observation history sample. The two SVMs out of sample (within the history data set with `biweekly==10`) predictions: Choosing an undersampling

(a) naive

		Real outcome	
		Short	Long
Prediction	Short	62%	18%
	Long	4%	16%

out-of-sample $\kappa = 0.38$

(b) undersampled

		Real outcome	
		Short	Long
Prediction	Short	48%	11%
	Long	18%	23%

out-of-sample $\kappa = 0.35$

ratio between the real distribution of short and long-term outcomes and a 50/50 split would allow adjusting the ratio of error types. It is remarkable that the (out-of-sample) κ does not change by a lot: The undersampled machine does not perform significantly worse in this metric—this might allow the tentatively drawn conclusion that the tradeoff is almost linear, only weighted by class size.
The oversampling approach leads to a tertiary (and almost political) optimization problem: how many false positives are acceptable to eliminate a false negative?

[12] Returning to hard margin SVMs is not advisable, they would be very unstable under outliers.
[13] This extension is not mentioned in textbooks used for this thesis. The technique is in similar use in gene expression and disease research.

(c) naive

		Real outcome	
		Short	Long
Prediction	Short	71%	19%
	Long	3%	7%

(d) undersampled

		Real outcome	
		Short	Long
Prediction	Short	57%	9%
	Long	16%	18%

(e) classical

		Real outcome	
		Short	Long
Prediction	Short	55%	16%
	Long	17%	12%

4.2 Putting it all together

Here are the results of an 50/50 undersampled history sample with a complete 10 SVM composite `biweekly` matched approach (sample size is 10000 per SVM) on $\mathbb{J}_C$ compared to the naive composite SVM predictor (with 20000 sample size) and to the benchmark result with classical econometric techniques:[14]

The naive SVM shows a better overall accuracy than the classical prediction, but it heavily underestimates the likelihood of long term unemployment, leading to low specificity. The undersampled SVM *clearly outperforms* the classical techniques by a good margin: Specificity is enhanced significantly while no penalty in terms of type II error has to be paid.
The composite 10 machine prediction beats the individual machines when predicting $\mathbb{J}_C$ in terms of κ with a result of $\kappa \approx 0.39$. This implies that the 10 stage concept of utilizing the additional information of time already spent in unemployment is useful and contributes to the prediction gains.

5 Discussion

The gain in predictive power appears impressive at face value. But it only applies to a very artificial sample, a snapshot taken at one point in time of all unemployed people registered in the past few months in a specific Swiss canton. While it might make sense to build a large battery of regionally and temporally restricted predictors with specifically adjusted oversampling, this is, after all, nothing more than an expansion of training sample size that remains computationally feasible by clustering the sample space. Much of the prediction gain seems to be generated by the composite algorithm's knowledge that some spells already lasted for some time. It is possible that some of the gain is generated by putting a focus on the time interval of October to January, which has odd and seasonally specific properties and is, therefore, informative.

[14]The classical method used a slightly tuned subsample of $\mathbb{J}_C$ and a marginally different definition of long term unemployment. This leads to the divergence of relative outcome class size, but does not drive the result. The performance of the classical model appears to be broadly on par with other predictions found in the literature, but a direct comparison is not possible since the underlying populations differ.

It is not yet clear if, how and to what extend the result is going to be generalizable: The sample is small, the design of the composite predictor is quite specific and multiple random draws were used to generate subsamples.[15] Still, the gain achieved might warrant further research to examine generalizability and to test robustness for different sampling on every stage, different target populations and different outcome specifications. Additional enquiry into the properties of the proposed techniques and a systematic analysis of oversampling behavior could be worthwhile.
Furthermore, a verification on a broader range of data, both within Switzerland and internationally, would be needed to establish direct comparability with existing methods and to examine the universal applicability of the oversampled and composited machine learning approach.
And, finally, computational resources have to be taken into account: Generating the two composite classifiers for a sample of 1207 spells (including the imputation of missing values) took approximately 60h on a high performance 64 core server at IZA Bonn (and this was after all the kinks have been worked out). Efficiency needs to be increased for larger scale application.

It stands to reason that even if half the magnitude of improvement carries over to more general applications this could be considered a significant gain in predictive power for long-term unemployment outcomes. And still, even the improved prediction detects only two thirds of real long term unemployed while erroneously assigning "long" about 50% of the time. Is this good enough to guide action and advise people in the real world? Is this relevant information?

References

Boser, Bernhard E, Isabelle M Guyon, and Vladimir N Vapnik (1992). "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, pp. 144–152.

Brennan, Robert L and Dale J Prediger (1981). "Coefficient kappa: Some uses, misuses, and alternatives". In: *Educational and psychological measurement* 41.3, pp. 687–699.

Cohen, Jacob (1968). "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit." In: *Psychological bulletin* 70.4, p. 213.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20.3, pp. 273–297.

Fechner, Patrick (2014). "Predicting Unemployment Durations in Switzerland". MA thesis. Universität Bonn.

Hasluck, Chris (2008). "The use of statistical profiling for targeting employment services: The international experience". In: *New European Approaches to Long-Term Unemployment: What role for public employment services and what market for private stakeholders*.

Hastie, Trevor et al. (2009). *The elements of statistical learning*. Vol. 2. 1. Springer.

[15] Time restrictions did not permit repetition with independent draws, but the sample sizes of 10000-20000 per SVM make representative sampling likely.

Landis, J Richard and Gary G Koch (1977). "The measurement of observer agreement for categorical data". In: *biometrics*, pp. 159–174.

Matty, Simon (2011). "Predicting likelihood of long-term unemployment: the development of a UK jobseekers' classification". https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210303/WP116.pdf.

Murphy, Kevin P (2012). *Machine learning: a probabilistic perspective*. MIT press.

Neumann, John von (1944). "Theory of Games and Economic Behavior (1944)". In: *Nature* 246, pp. 15–18.

Nevin, John A (1969). "SIGNAL DETECTION THEORY AND OPERANT BEHAVIOR: A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. 1". In: *Journal of the Experimental Analysis of Behavior* 12.3, pp. 475–480.

Priest, George L. and Benjamin Klein (1984). "The Selection of Disputes for Litigation". In: *The Journal of Legal Studies* 13.1, pp. 1–55.

Riipinen, Toni (2007). "Risk profiling of long-term unemployment in Finland". http://ec.europa.eu/social/BlobServlet?docId=7583&langId=en.

Rosholm, Michael, Michael Svarer, and Bo Hammer (2004). "A Danish profiling system". In:

Schwerdtfeger, Christoph (2015). *Predicting swiss unemployment with machine learning techniques*. Bachelor thesis.

Vapnik, Vladimir Naumovich (1998). *Statistical learning theory*. Vol. 1. Wiley New York.

Advances in Mechanism Design: Information Management and Information Design

Anne-Katrin Roesler[1]

1 Introduction

The rapid advance of information technologies has changed our access to information and the nature of decision making. Nowadays, a large amount of information is fully and freely available. The gathering of information has become easier, whereas the processing of information has become increasingly challenging. The changes in our information environment are also reflected in some recent developments in economic theory. Questions such as the following currently receive increased attention. When facing a decision, should one use all of the available information? Or could there be benefits from committing to ignore information? Is ignorance bliss? If so, which information should one focus on?

The recent literature on information management and information design addresses these kind of questions, and most of my research falls into this area. In this article, I wish to provide a brief overview on what information design is, and illustrate some questions and findings.

Let me start by explaining how the area developed, what *information design* and *information management* are, and why these topics can be considered as advances or a subfield of mechanism design.

Microeconomic theory is the study of individual behavior, incentives and the allocation of limited resources. A central element in the toolbox of a micro-theorist is game theory. It provides a basis to build mathematical models of economic situations, which are then used to make predictions about outcomes, based on the assumption that the players who interact with each other behave strategically in order to achieve their own goals. Game theoretic models have helped us to better understand the incentives in markets, organizations, and political campaigns, just to name a few.

[1]Anne-Katrin Roesler is assitant Professor in Economics at the University of Michigan in Ann Arbor, USA. akroe@umich.edu. The topics discussed are questions that caught my interest during my doctoral studies at the University of Bonn and at Yale University. The present material is based on insights from my dissertation, Roesler 2015, and Roesler and Szentes 2016. I am grateful to my advisors Benny Moldovanu, Dirk Bergemann, and Daniel Krähmer for their support, advice and numerous insightful discussions throughout my studies and beyond.

Mechanism design takes this idea one step further. Game theory helps us to better understand incentives, whereas mechanism design shifts the focus to influencing incentives. Mechanism design studies how to design a game (or institution) in order to best achieve a desired objective, taking into account the constraints imposed by the private information of agents.[2] Depending on the application, the designer can be a player in the game or a third party. For example, the designer can be a seller aiming to design a mechanism that maximize his revenue from selling an object to privately informed buyers, or a benevolent planner who wishes to maximize welfare and implement the efficient outcome in a market.

Traditionally, in game theory and mechanism design the information structure is given exogenously: There is common knowledge about the structure of the game, a common prior belief about the distribution of players' types, and players may hold private information about their types. A natural step is to endogenize players' private information and include the information generating process into the mechanism design problem. In this case, the designer has to take into account that now the mechanism has a dual purpose. It influences the incentives to acquire or disclose private information, and is then also used to elicit this private information.

The aforementioned changes in our information environment – advancing from a time in which information gathering was costly to today's information age in which enormous amounts of information are freely available – is reflected in the development of the literature. The literature on mechanism design with endogenous information started off by looking at models in which information is costly.[3] Typically information is provided through a specific information technology, and players are able to increase its precision through costly investments. The literature evolved to studying flexible information acquisition and disclosure, and to also consider costless information.[4]

In this article, two concepts that go hand-in-hand and that I consider to bring new, intriguing ideas to mechanism design are discussed: Information management and information design. *Information management* is based on the observation that in various situations of economic interest a designer may not have the authority to change the rules of the game, for example the market structure, or the organizational structure of a company. Still, he may be able to manage the information available to players in the game, that is, he can influence the information environment and hence use information management to influence the incentives for players. *Information design* builds on the same idea, but provides the designer with even more power to design information. First, it aims to understand how in a given game the information environment affects incentives, and to identify the set of outcomes that can be reached by (re-)designing the information environment. The size of the set of achievable outcomes captures how effective information design can be in a specific setting. Second, it analyzes how to design the information environment to influence the

[2]Information asymmetries create incentives for players to strategically misrepresent their types, which imposes restrictions on the outcomes that can be implemented. A mechanism first has to elicit the private information of players, to then use it to determine the outcome.

[3]See for example Persico 2000; Ganuza and Penalva 2010; Shi 2012.

[4]Examples include Kessler 1998 and Bergemann and Pesendorfer 2007.

incentives of self-interested players and to best achieve a given objective. Information design is an addition to the toolbox of a mechanism designer.

2 Illustrations

To illustrate some basic ideas and prospects of information management and information design, I discuss two examples of problems that I have worked on: The role of the information environment – especially the buyer's private information – in an optimal pricing problem, and information management in a promotion contest.

2.1 Buyer's Information and Monopoly Pricing

Consider a standard, simple buyer-seller model, in which a seller (she) wants to sell one object to a buyer (he). Both players are risk-neutral. The buyer's valuation for the object, v is randomly drawn from the uniform distribution on $[0,1]$. The seller's marginal cost is zero. The seller's cost and the distribution of the buyer's valuation are common knowledge.

The seller makes a take-it-or-leave-it offer to the buyer. If a buyer with valuation v and the seller trade the object at price p, then the seller's payoff (*revenue*) is $r = p$, and the buyer's net payoff (*surplus*) is $u = v - p$.

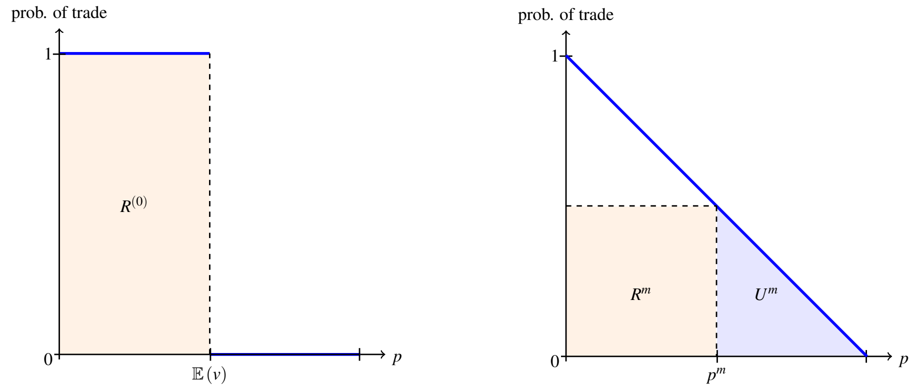
The buyer's private information. In most environments, a buyer has some private information about his valuation for an object. To get an idea of how the private information of the buyer influences the outcome, consider first the two extremes: the case in which the buyer has no additional information about his valuation, and the case in which he is privately informed about his valuation.

In the first case, information is symmetric: both the buyer and the seller only know the distribution of the buyer's valuation. For any price p announced by the seller, the buyer can base his decision whether to buy or not only on the distribution of his valuation. For a risk-neutral buyer it is optimal to trade if and only if his expected value, here $\mathbb{E}(v) = \frac{1}{2}$, is at least as high as the price. It is optimal for the seller to charge a price equal to this expected value, $p = \frac{1}{2}$, and for the buyer to always buy the good. Trade takes place with probability one, and the potential gains from trade are fully realized. The seller extracts all surplus from trade. Her expected revenue is $R^{(0)} = \frac{1}{2}$, the buyer obtains zero surplus.

The second case, in which the buyer is privately informed and knows his true valuation, is the standard monopoly pricing problem. In equilibrium, the seller will charge the revenue-maximizing monopoly price

$$p^m = \arg\max\{p \cdot (1-p)\} = \frac{1}{2},$$

Figure 2.1: Demand function, seller's revenue and buyer's surplus for (a) an uninformed buyer, and (b) a fully informed buyer.



(a) Uninformed buyer.

(b) Fully informed buyer.

and the buyer only buys the good if his true valuation is greater or equal to the price.[5] The seller has to leave information rents to the buyer. Trade is not efficient – there is some deadweight loss.[6] The expected realized gains from trade are $T^m = \frac{3}{8}$, which is split between the buyer and the seller. The seller's expected revenue is $R^m = \frac{1}{4}$, and the buyer's expected surplus is $U^m = \frac{1}{8}$. These two cases with an uninformed and a fully informed buyer, respectively, are illustrated in Figure 2.1.

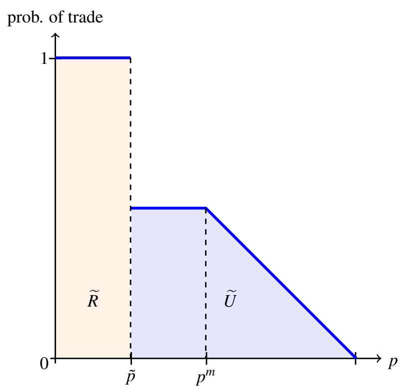
This example already illustrates that the outcome in a pricing model depends on the information held by the players. Here, in the case with symmetric information, trade occurs with probability one, all gains from trade are realized and the seller extracts the full surplus. By contrast, if the buyer is privately informed about his true valuation, he extracts some information rents. Moreover, under the revenue-maximizing price trade is not efficient – not all gains from trade are realized.

Based on the preceding discussion, it may at first seem like the relation between a buyer's private information and his expected payoff is monotone. If the buyer has more information, he can extract more information rents and hence is better off. However, this conjecture turns out to be wrong. The buyer only benefits from information if he can extract information rents. Hence, he does not benefit from information that separates values below the price that the seller charges – which is $\frac{1}{2}$ in the above example. If instead of becoming fully informed, the buyer only obtains a perfectly informative signal for valuations above $\frac{1}{2}$, and otherwise just learns that his valuation is lower, this does not affect his payoff as long as the price remains the same. Upon observing the low signal, the buyer's expected valuation is $\frac{1}{4}$ and he does not buy at price $\frac{1}{2}$; for higher valuations he buys the good. Notice moreover, that the seller now faces an effective demand function with a mass point at $\frac{1}{4}$ and a linear demand for prices of $\frac{1}{2}$ and above. This is illustrated in Figure 2.2. The seller is indifferent between charging a price of $\frac{1}{2}$ and trade half of the time, and charging a

[5] It is irrelevant whether the buyer buys the good or not if he is indifferent, since the event that the buyer's valuation is equal to the price is a zero-probability event.

[6] Not all gains from trade are realized since buyers with valuation below $\frac{1}{2}$ do not participate in trade.

Figure 2.2: Demand function, seller's revenue and buyer's surplus for a partially informed buyer.

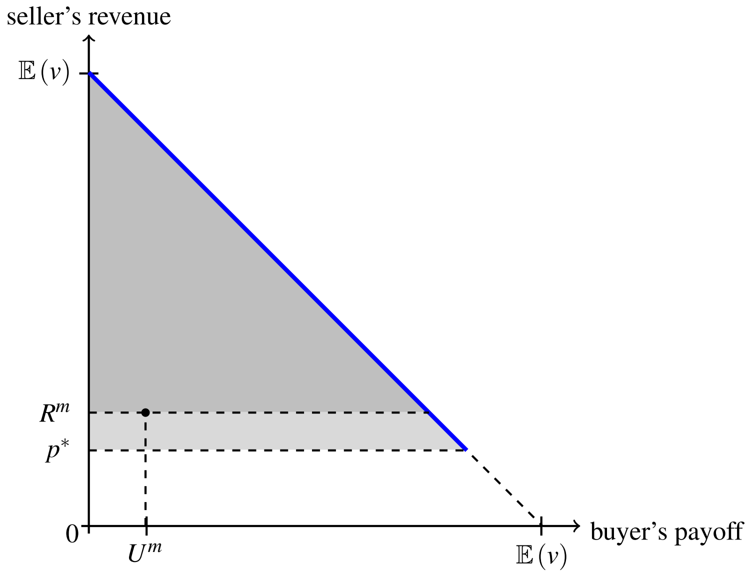


price of $\frac{1}{4}$ and sell for sure. In both cases, her revenue is $\widetilde{R} = \frac{1}{4}$. Hence, there exists an equilibrium in which trade occurs with probability one, all gains from trade are realized, and seller and buyer both obtain a surplus of $\frac{1}{4}$. Under this information structure, the buyer's expected surplus doubles compared to the case in which he knows his true valuation. Here, the buyer is better off by knowing less.

What is the economic intuition behind this result? If the buyer is oblivious to information that would separate low values, he effectively commits to sometimes buy at a price that is higher than his true valuation. As a result additional gains from trade are realized. This induces the seller to offer better terms of trade, that is, a lower price, in return for the increase probability of trade. The positive effect of the additionally realized gains from trade reverberates back to the buyer.

The discussion illustrates how information design may be used to achieve a desired outcome. The outcome of the pricing model depends on the information environment, here, the buyer's private information. The example shows that the buyer may be better off by knowing less. But is this the end of the story? Which outcomes can be achieved for different information structures of the buyer? And what is the optimal information environment for the buyer? These questions are studied in Roesler and Szentes 2016. It is shown that the buyer can do even better. Roesler and Szentes 2016 identify the posterior distribution and the price induced by a buyer-optimal signal. The buyer-optimal signal induces a unit-elastic demand function for the seller, who charges a price equal to the lower bound of the support of the posterior distribution. Trade occurs with probability one. Given the distribution of her value estimate, the buyer always buys the good at the equilibrium price, even though the price may exceed her true valuation. This offers the seller a higher probability of trade at intermediate prices. The intuition from the above example still applies: Under the buyer-optimal signal, the seller offers better terms of trade in return for the increased probability of trade. The buyer's optimal learning is driven by the goal of generating a demand function which induces the lowest possible price p^* subject to all gains from trade being realized. To understand the unit-elastic demand property, notice that if all gains from trade are realized, then the buyer's surplus just depends on how the total surplus is shared. The seller has to choose a price.

Figure 2.3: Surplus triangle. The darker shaded triangle is the set of outcomes attainable by changing the seller's information. The larger, lighter-shaded triangle is the set of outcomes attainable for all possible information environments in the pricing problem.



The unit-elastic demand property of the buyer-optimal signal, pins the seller to a revenue level. It leaves her with just enough surplus to guarantee that she does not want to deviate to a higher price.

Naturally, one can ask the complementary question, how the seller's information affects the outcome in a pricing problem. This problem is studied in Bergemann, Brooks, and Morris 2015. They consider a privately informed buyer and analyze how the information of the seller influences the outcome. More information on the side of the seller means that he can price discriminate. Bergemann, Brooks, and Morris 2015 identify the set of outcomes that is achievable by changing the information held by the seller. It is the set of the surplus triangle for which the seller's revenue is bounded below by the monopoly revenue, the buyer's surplus is non-negative, and total surplus is weakly less than the maximally realizable gains of trade. This is the smaller, darker-shaded triangle in Figure 2.3.

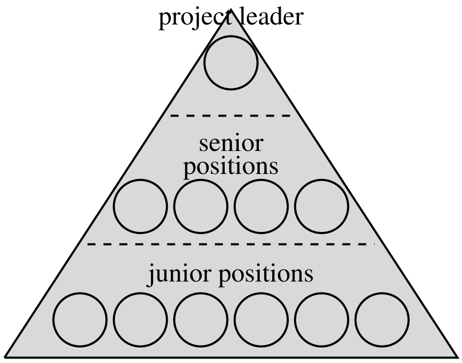
Combining the insights about how the buyer's and seller's information influence the outcome of the pricing problem, yields the set outcomes that are attainable for all possible information environments, for a given underlying prior distribution of the buyer's valuations. This result is formally established in Roesler and Szentes 2016, and illustrated as the larger, lighter-shaded triangle in Figure 2.3.

2.2 Information Management and Feedback in Promotion Tournaments

Information management plays an important role in situation in which the designer cannot change certain features of the environment, that is, if from a theoretical perspective the game is fixed.

In order to illustrate this consider the following example. A designer wishes to create an incentive scheme to maximize total efforts within a specific division of a company. He faces the problem that the organizational structure of the division, as well as the compensation-scheme for the positions is fixed (wages, benefits, etc.) and cannot be changed. For concreteness, suppose that the division has 11 position, a project leader, four senior consultants, and six junior consultants. This organizational structure is illustrated in Figure 2.4.

Figure 2.4: Organizational structure.



Suppose, that the company employs an up-or-out policy: every five years, employees at the same level are ranked according to their efforts. Based on this ranking it is decided whether an employee is promoted to a higher position, otherwise he has to leave the company.[7] Employees find themselves in a promotion contest in which they exert costly effort in order to be promoted to a higher position.

Let us set up a simple, stylized model for this situation. It is common to model contests as an all-pay auction with multiple prizes.[8] Bids correspond to exerted effort, and winning a prize in the contest corresponds to being promoted. Consider the situation for the junior consultants. There are six contestants $i \in \{1,\dots,6\}$, described by their *types* x_i, which are drawn independently from the uniform distribution on $[0,1]$. Players' types reflect their skill levels or abilities. The positions for which they compete are modeled as *prizes* of value $(y_1,\dots,y_6) = (4,3,2,1,0,0)$, where a prize of value 0 represents no promotion.[9] A player with type x_i obtains payoff $x_i y_j$ from winning prize y_j. This payoff structure implies that employees with higher types can generate higher payoff from

[7]For simplicity, we abstract away from the situation of the project leader, and simply assume that this position becomes available after five years. The project leader leaves the division and moves on e.g. to a different company, or to another position within the company.

[8]See for example Moldovanu and Sela 2001 and Olszewski and Siegel 2016b.

[9]Notice that this payoff structure implies that positions on the senior level are ranked, i.e., there are more and less attractive positions on that level.

being promoted than lower types.[10] Employees compete for positions by exerting costly effort. If a player of type x_i exerts effort $b \in [0,1]$ and wins prize y_j, his payoff is $x_i y_j - b$. Employees are promoted according to their effort, the employee with the highest effort wins y_1, the second highest y_2, and so on. The described setting is an all-pay auction, a standard model in auction theory. For the case in which players' types are private information to the players, following Vickrey 1961, Clarke 1971, and Groves 1973 there exists an equilibrium in which players are matched to prizes assortatively according to their types, and effort levels are given by VCG-payments: The expected payment of each player is equal to the externality that he imposes on the other players. In a contest with n players, for the player with the i^{th}-highest type – which we denote by $x_{i:n}$ – his expected payment is

$$t_{(i)} = \sum_{j=i}^{n} \mathbb{E}\left(x_{i+1:n}\right) \cdot \left(y_j - y_{j+1}\right), \tag{1}$$

where $\mathbb{E}\left(x_{i+1:n}\right)$ is the expected value of the $(i+1)^{th}$ highest type among n players.[11] The designer wishes to maximizes total efforts which is given by the sum of individual efforts

$$\begin{aligned} T &= \sum_{i=1}^{n} t_{(i)} = \sum_{i=1}^{n} \sum_{j=i}^{n} \mathbb{E}\left(x_{i+1:n}\right) \cdot \left(y_j - y_{j+1}\right) \\ &= \sum_{i=1}^{n} i \cdot \mathbb{E}\left(x_{i+1:n}\right) \cdot \left(y_i - y_{i+1}\right). \end{aligned} \tag{2}$$

This problem has been analyzed as a mechanism design problem. As shown in Moldovanu and Sela 2001 the optimal contest which maximizes total effort would only provide one prize and have contestants compete for it.[12] However, in the current example we assume that the organizational structure of the division is fixed and cannot be altered by the designer. This is where information management comes into play. A way to influence the incentives of employees to exert effort is to design the feedback system of the division. Feedback is given to employees to provide them with information about their "type" in the company, which allows them to form a better estimate of how they rank compared to their peers and hence their prospects to be promoted. This may encourage or discourage employees to exert more effort.

Consider the following simple information technology to model feedback. Each employee obtains a private signal $s \in [0,1]$, which with probability $\alpha \in [0,1]$ is his true type, and with probability $1-\alpha$ is pure noise. The employee cannot identify whether he observed his true type or noise. The employee updates his belief based on his private signal. Upon observing s, his posterior type is

$$\mathbb{E}\left(x|s\right) = \alpha s + (1-\alpha)\mathbb{E}(x). \tag{3}$$

[10] This is true for many position, think for example about bonus payments, or the option to engage in side-projects.

[11] Technically, $x_{i:n}$ denotes the i^{th} order statistics, that is, is distributed according to the i^{th}-highest among n random draws from the distribution F – here the uniform distribution on $[0,1]$.

[12] Olszewski and Siegel 2016a confirm and generalize this result. They show that a contest with a single prizes is optimal whenever prize valuations are linear or convex and effort costs are linear, or when prize valuations are linear and effort costs are linear or concave.

Here, the parameter α captures the precision of the information technology; for higher α, the signal is more precise.

Consider the following extension of the contest model. A priori contestants have no private information about their type but learn about it through the feedback provided to them before entering the contest.[13] Formally, they receive a private (partially) informative signal s_i through an information technology of precision α. With this private information players then enter the promotion contest. In the contest, the same considerations as presented above apply. Equilibrium efforts still satisfy (1) and (2), but now the (distribution of) true types $x_1,\ldots,x_6$ has to be replaced by the player's (distribution of) posterior types, $\mathbb{E}\left(x|s_1\right),\ldots,\mathbb{E}\left(x|s_6\right)$.

Suppose now that the designer can choose the precision of the information technology. Notice that for a more precise information technology the support of posterior types gets larger. To see this consider the two extremes: no feedback and perfect feedback. For no feedback – corresponding to an information technology that is pure noise, $\alpha = 0$ – the posterior type $\mathbb{E}\left(x|s\right)$ given by (3) always equals the prior mean $\mathbb{E}(x)$. The support of posterior types is a singleton. A signal contains no information for a player and hence, in the contest all players act as if their type were $\mathbb{E}(x)$. The resulting allocation is random and players' total exerted efforts given by (2) are $T^{(0)} = 5$. For perfect feedback, corresponding to an information technology with $\alpha = 1$, the signal realizations correspond to the players' true types and so do the posterior types, $\mathbb{E}\left(x|s_i\right) = s_i$ for all $s_i \in [0,1]$. The posterior types are distributed uniformly on $[0,1]$, and players total efforts are $T(1) = 4\frac{2}{7}$. In this case, the optimal, effort-maximizing feedback policy is to provide no feedback to contestants.

Consider the contest that the senior consultants face. At this level, four contestants compete for one prize, say of value $y = 1$. In this case, with feedback of precision α, total effort of contestants as given by (2) is $\frac{5+\alpha}{10}$. It is easy to see that this is increasing in α and hence providing perfect feedback is optimal at this stage.

What is the economic intuition behind this observation? Increasing the precision of feedback and hence the private information held by contestants has two important effects: an *allocation effect* and a *competition effect*. On the one hand, if contestants receive more precise feedback, this increases the probability that the contestant with the highest type receives the best feedback, hence exerts the highest effort and is promoted to the best position. More precise information allows for a better allocation. On the other hand, the information precision affects the competition among employees. More precise information will increase the effort levels of high (posterior) types since they now know that they are competing against their peers – other high types. However, feedback will discourage low types to exert effort. For more precise information the posterior types are more in line (correlated) with the underlying true types, which results in a better allocation of prizes to contestants. Hence, for a low-type contestant the probability that he is lucky and receives a prize decreases, which results in reduced effort.

This intuition suggests that the optimal feedback policy depends on the ratio of contestants to

[13] For example, this could be a performance report given to employees after being hired or promoted.

prizes. If there are only a few prizes then only the contestants with a high posterior type will be matched. Hence, for a more precise signal, the increased competition among high-type contestants is the driving force and will result in higher total efforts. By contrast, if a high proportion of contestants expects to receive a prize, the effort-reducing effect of more precise information for low-type contestants becomes more relevant. Total efforts may decrease in the precision of feedback.

In Roesler 2015 it is shown that the insights from this example generalize.[14] The optimal precision of information or feedback in a contest depends on the ratio of prizes to contestants. For any given set of prizes, there exists a number $\hat{n}$ such that the effort-maximizing feedback policy is to provide perfect information if the number of competitors is sufficiently large (i.e. $n \geq \hat{n}$) and no information otherwise.

Notice that we only consider a very simple, stylized example here. We do not consider coarse, more flexible, or dynamic feedback or information technologies, and do not take into account that implementing more precise feedback may be costly. It is possible to incorporate these and other features in the model which is the subject of ongoing research. However, the simple example presented here already shows that information management and design are important additions to the toolbox of a (mechanism) designer.

The observations presented here and the results in Roesler 2015 suggest that we should observe different feedback systems in companies, depending on their organizational structure. In organizations with steeper hierarchies in which employees face fierce competition for job promotions, well-established, precise feedback systems are optimal. By contrast, for organizations with flat hierarchies or promotion by seniority practices, less sophisticated feedback structures are optimal. These predictions seem to be in line with common practices. For example, highly competitive environments like large consulting firms are known to have very rigorous feedback systems.

3 Conclusion

This article provided a brief introduction to information design and presented two examples to illustrate some of the questions that are studied on this topic. There is plenty of exciting research in this area that could not be covered in this short article. Bayesian persuasion is one of these topics. A seminal paper is Kamenica and Gentzkow 2011, in which the authors analyze how a sender can design the information environment of the receiver in order to persuade the receiver to take the sender's preferred action. For the interested reader the note by Bergemann and Morris 2016 provides and excellent introduction and overview on related topics.

One of the intriguing aspects of information design for me is that similar questions and topics

[14]To more general information technologies and prior distributions of types.

have been around and studied by micro-theorists for a while. However, the developments in our information environment have shifted the focus of research in information economics and brought forward new questions. Being involved in research in an area that is motivated by and has partially evolved from changes in our environment is very exciting, and allows to be inspired by experiences and occurrences in everyday situations. Without a doubt there are plenty of open questions related to the topic of information design. I am curious to see how this research area will develop and grow.

References

Bergemann, Dirk, Benjamin A. Brooks, and Stephen Morris (2015). “The Limits of Price Discrimination”. In: *American Economic Review* 105, pp. 921–957.

Bergemann, Dirk and Stephen Morris (2016). “Information Design, Bayesian Persuasion and Bayes Correlated Equilibrium”. In: *American Economic Review, P&P* forthcoming.

Bergemann, Dirk and Martin Pesendorfer (2007). “Information Structures in Optimal Auctions”. In: *Journal of Economic Theory* 137, pp. 580–609.

Clarke, Edward H. (1971). “Multipart Pricing of Public Goods”. In: *Public Choice* 11 (1), pp. 17–33.

Ganuza, Juan-Jose and Jose S. Penalva (2010). “Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions”. In: *Econometrica* 78.3, pp. 1007–1030.

Groves, Theodore (1973). “Incentives in Teams”. In: *Econometrica* 41.4, pp. 617–631.

Kamenica, Emir and Matthew Gentzkow (2011). “Bayesian Persuasion”. In: *American Economic Review* 101, pp. 2590–2615.

Kessler, Anke S. (1998). “The Value of Ignorance”. In: *The RAND Journal of Economics* 29, pp. 339–354.

Moldovanu, Benny and Aner Sela (2001). “The Optimal Allocation of Prizes in Contests”. In: *American Economic Review* 91.3, pp. 542–558.

Olszewski, Wojciech and Ron Siegel (2016a). “Effort-Maximizing Contests”. working paper, Northwestern University.

— (2016b). “Large Contests”. In: *Econometrica* 84. working paper, Northwestern University, pp. 835–854.

Persico, Nicola (2000). “Information Acquisition in Auctions”. In: *Econometrica* 68.1, pp. 135–148.

Roesler, Anne-Katrin (2015). “Information Disclosure in Markets: Auctions, Contests, and Matching Markets”. Working Paper.

Roesler, Anne-Katrin and Balazs Szentes (2016). “Buyer-Optimal Learning and Monopoly Pricing”. Working Paper.

Shi, Xianwen (2012). “Optimal Auctions with Information Acquisition”. In: *Games and Economic Behavior* 74.2, pp. 666–686.

Vickrey, William (1961). “Counterspeculation, Auctions, and Competitive Sealed Tenders”. In: *The Journal of Finance* 16, pp. 8–37.

Demographic Change – The Relation between Fertility, Female Labor Force Participation, and Economic Growth

Lilian Kreitz[1]

1 Introduction

In the light of the ongoing trends of decreasing fertility and mortality rates, coupled with an increase in life expectancy over the past decades, more and more countries are facing aging societies. Demographic change towards an older population tends to induce both lower labor force participation rates and lower savings. Thus, while societies are aging, concerns about the sustainability of social security systems, as well as the growth potential of economies are rising (Bloom et al., 2011). It is therefore fundamental to observe how changes in demographics affect a country's economy and growth potential.
This work focuses on the changes in fertility, its effect on per capita income in the medium run, and how women adjust their labor supply as a response to lower birth rates.

For my analysis, I investigate the effects of a declining fertility on labor, capital, and income over three different channels. Firstly, I will examine the impact of lower birth rates on the economic performance within the framework of the neoclassical Solow-Swan model. Secondly, I will expand this growth model by including the population's age structure. A central concept for this approach are age dependency ratios. Thirdly, I will further include one behavioral effect, the changing labor force participation of women.

The growth effects associated with a decline in fertility are studied empirically using data from Japan. There are two reasons, why I am using Japan as a country for my simulation. First of all, it is one of the first countries to experience the demographic upheaval. Since Japan's trajectory is not unique, yet still of an unprecedented nature globally, other countries could learn from how its economy is affected by such demographic change, and, more importantly, how the country is dealing with the consequences. Secondly, it represents the prime example for a country with high potential of working age women, but low female labor force participation (FLP) rates (Süssmuth-Dyckerhoff, Wang and Chen, 2012).

[1]Lilian Kreitz reveived her degree (B.Sc.) from the University of Bonn in 2014. The presented article refers to her bachelor thesis under supervision of JProf. Michael Boehm, which has been submitted in September 2014.

2 Effects of Demographic Change on Economic Growth

2.1 Key Concepts

Two key concepts in this paper are fertility rates and age dependency ratios. A population's total fertility rate is a theoretical measure for the number of children that would be born per 1,000 women in reproductive age between 15-49, if they were bearing children according to a current schedule of age-specific fertility rates (The World Bank, 2013). Generally, the fertility rates are lowest for the oldest group of women in childbearing age.

The latter key concept illustrates an age-population ratio of the dependent young and elderly part, which are typically not in the labor force, and the productive part, which typically participates in the labor force. More precisely, the old-age dependency ratio is the ratio of people older than 64 years to the working-age population, between 15-64, while the youth dependency ratio is the ratio of people younger than 15 years to people in the age between 15-64 years. The sum of both is the age dependency ratio (The World Bank, 2013). As Bloom et al. (2011) outline: "[...] a country with large cohorts of youth and elderly is likely to experience slower growth than one with a high proportion of working-age people".

2.2 Solow Model, Age Structure Effects and Behavioral Responses

In the basic Solow-Swan growth model, full employment is assumed. Thus, the amount of employed labor equals the population size, or alternatively, the labor force grows at the same rate as the population (Prettner and Prskawetz, 2010). Population growth fosters capital dilution and slows per capita output growth, but does not affect economic growth in the long-run (Weber, 2010). By neglecting changes in age structures of the population, the neoclassical growth theory, however, creates substantial errors when demographic change is taking place (Tyers and Shi 2007). Therefore, more recent attempts adopt life cycle perspectives and more realistic age structures to their growth models. The key premise of these new approaches is that labor supply, productivity, as well as savings vary over the life cycle. Generally, labor supply and savings are high among working-age adults, while the consumption to production ratio is high for young and old cohorts (Bloom et al., 2011). Thus youth and old age dependency ratios are vital for economic growth. On the one hand, demographic change comes along with an increasing old age dependency ratio (OADR), caused by a decline in recent mortality rates, and increased life-expectancy. The elderly cohort, and therefore their dependency on the working age cohort, is further enlarged through past age dynamics. Baby boomers of the years 1943-1960, the group with the highest participation rate in the labor market in the last decades, are steadily moving onward to age categories with much lower participation rate.

On the other hand, the demographic transition is caused by declines in fertility rates, inducing lower youth dependency ratios (YDR) (Bloom, Canning and Sevilla, 2003). A smaller share of young people coexists with an increased output per capita, if everything else stays constant (Brander and Dowrick 1994; Kelley and Schmidt 1995).

During demographic transition not only the population's structural features change, but private and public sectors also adjust their behavior (Bloom et al., 2009). Changes in behavior should therefore be included in the growth analysis.

For this work, the labor market decisions of women as response to the fertility changes are central. Generally, a decline in fertility and thus smaller family sizes are expected to lead to an increased labor force participation of women (Bloom et al., 2011). I chose the FLP as behavioral factor, because it is gaining importance for the economic performance: While education and productivity of women have increased steadily over the last decades, labor force participation is still significantly lower for women than for men. Thus, higher FLP rates could potentially increase the labor supply despite declining and aging populations. There are three other behavioral responses, disregarded in the model, but, as I think, crucial for the discussion of demographic change: the increased investment in human capital and thus higher productivity of the young generation, later retirement and more working years, as well as increased savings as a response to rising life expectancy.

3 Model Applied to Japan

3.1 Features of Japan's Demographics and Labor Market

With 23.1 percent, Japan has the highest percentage of people aged 65 years and above worldwide (Statistics Bureau, 2011), while its fertility rate is below the replacement level. After the country experienced a large demographic dividend from the 1960's to the 1980's, Japan is now growing older faster than any other country in the world (Steinberg and Nakane, 2012). Life expectancy is also amongst the highest on the globe (UN, 2012). Thus, Japan is one of the first countires to experience the consequences of aging.

Common key responses to aging and a decreasing labor force are increased immigration, increased retirement age, and increased participation of women on the labor market.

As Steinberg and Nakane (2012) pose, both, immigration and FLP rates, are well below OECD level. And, similar to other countries, the retirement age in Japan did not align to the rising life expectancy.

Migration is a controversial topic in Japan, and immigration laws are not likely to respond considerably to the demographic transition. Furthermore, since most companies in Japan feature seniority-based payment systems, the older working cohort poses a growing burden for them, and, thus, they struggle with the implementation of rising retirement ages (Haghirian, 2010). Hence, a larger FLP represents a likely optimal response to Japan's demographic change.

Amongst OECD countries, young Japanese women are the second most educated, with an average of 14.3 years of schooling in their late 20's (Steinberg and Nakane, 2012). At the same time, the female labor force is one of the smallest compared to other OECD nations. Currently, 48.5 percent of women participate in the labor force, while the percentage of men is 71.6 percent (Süssmuth-Dyckerhoff, Wang and Chen, 2012).

The age group of women between 25 and 29 years has the highest participation with 77.1 percent, which drops off drastically around the age of 30 (Ministry of Internal Affairs and Communications, 2011). Six out of ten women quit working after giving birth to their first child and a large number of them never starts working again. Re-entering the labor market, furthermore, takes much longer than in other countries. Researchers call the picture, illustrating the female labor force participation in dependency of the age "*Japan's M-curve*" (Steinberg and Nakane, 2012) (Figure 1).

According to the World Economic Forum's "Global Gender Gap Report of 2010", Japan is ranked 100 out of 135 countries with a gap in economic participation and opportunities for women. Moreover, the wage for women is on average only 60 percent of what male workers earn in the same position. One reason for these circumstances is the two-track employment system in Japan[2]. In 2010, 53.8 percent of female employees were non-regular workers, while only 18.9 percent of the male labor force fell into this category (Ministry of Internal Affairs and Communications, 2011). In comparison to regular workers, the working hours are the same, whereas wages and premiums are much lower. Moreover, while permanent employees in Japan enjoy almost irredeemable employment, non-regular workers are hired under temporary contracts (Haghirian, 2010) and are likely to be laid off first during an economic turndown.

3.2 Data

All the data I am using, is given in five-year groups, covering the period from 1960 until 2010, plus a forecast until 2050. The gender-specific population data are grouped in five-year age increments, i.e. from 0-4 until 95-99 and 100+, and are provided by the United Nation's "World Population Prospects: The 2012 Revision". In addition to historical data recorded until 2010, they prepared different scenarios to forecast future population movements. I assume that the decline in birth rates persits and therefore use data from the low fertility scenario until 2050. Fertility as well as mortality rates come from the same source, the United Nations. The historical and forecasted fertility cover the fertile age groups from 15 until 49 years, whereas the mortality rates concern all ages.

Labor market participation data for men and women are taken from the International Labor

[2]Right after graduation, Japanese students have to make the most important individual labor market decision: They either take the career track and generally stay with the chosen corporation for their entire working life, or they decide to take the non-career track and thereafter often become non-regular workers (Yamaguchi, 2008).

Organization's "Database on Labor Statistics", and include the working-age population between 15 to 64 years.

3.3 Simulation for Japan

The empirical analysis of this work is based on the paper "Fertility, Female Labor Force Participation, and the Demographic Dividend" (2007) by Bloom, Canning, Fink, and Finlay. I will closely follow their simulation on long-run income dynamics given demographic changes, which they performed with data from South Korea.

In order to include behavioral changes of the labor market decision of women in Japan to my simulation, I will moreover use some of their regression results, which are summarized in Table 1.

To illustrate the growth effects caused by changes in the demographics, I will use a simple Cobb-Douglas production function, where the total output Y is given for every 5-year period t by

$$Y_t = AK_t^{\alpha} L_t^{1-\alpha} \tag{1}$$

In my simulation the focus is set on the question, how a decreasing fertility rate affects labor L, the physical capital stock K, and the income per capita. Since we are not interested in technological progress A during this simulation, I normalize this variable to 1. Also the capital share in income α is assumed to be one third, and accordingly the labor share $(1-\alpha)$ is set to two third.

The physical capital stock is determined by

$$K_t = sY_{t-1} + (1-\delta)K_{t-1} \tag{2}$$

for each period t. The variable s is the aggregate savings rate and δ is the depreciation rate. Bloom et al. assume a savings rate s of 24 percent and the annual capital stock depreciation rate δ to be 8 percent for South Korea. In favor of an easier comparison to their results, I chose the same values for both variables.

The total labor force for each period t is given by

$$L_t = \sum_{i=0}^{100} \sum_{g=m,f} P_{it}^{g} \cdot \rho_{it}^{g} \tag{3}$$

where P_{it}^{g} denotes the population and ρ_{it}^{g} the participation rate of gender g and age-group i.

To calculate each gender- and age-specific population group separately, I am using two different equations. The male and female population of the youngest birth cohort with the age 0-4 years in time t is given by

$$P_{0t}^{g} = \lambda_g \sum_{i=16}^{45} f_{it} \cdot P_{it}^{f} \tag{4}$$

where λ_g is the sex ratio at birth, which remained constant in the case of Japan with 51 percent of male births and 49 percent of female births (UN, 2012). The age and time specific fertility rate is presented by f_{it}, and only women from 16 till 45 years are considered to be fertile.

Each other gender-specific age cohort [3] at time t is given by

$$P_{it}^{g} = P_{i-1,t-1}^{g} \cdot (1 - \sigma_{i-1}^{g}) \quad \text{for } \ i \geq 1 \tag{5}$$

where σ_i^g captures the age- and sex-specific mortality rates, which are fixed at their 1960's level during the entire simulation, since I only want to observe the impact of fertility decline on population, and not the effect of increasing life-expectancy [4].

The fertility rate and the female labor participation rate are the only exogenous variables I will change during the simulation.

To demonstrate how a declining fertility rate affects the variables of the production function, I will firstly set up a baseline scenario, in which fertility rates remain constant at their 1960's level until 2050. Thereafter, I will change the fertility to their actual rates for the period of 1960 until 2010, and to the forecasted rates from the UN survey (2012), and estimate population, labor, capital and output again. To observe the impact of the fertility change in detail, I will divide the effects into Solow, age-structure, and behavioral effect.

The simulation is initialized with Japan's actual population structure in the period before 1960.

First of all, I calculate the female population for age-groups older than the "newborns" from 0-4 years with formula (5). Hereafter, I can estimate the population for the youngest age cohort for both women and men by using equation (4) since they depend on the fertile women at time t. Finally, I calculate the male population again with (5). From these estimates for the population of each male and female age group from 1960 until 2050, and the on 1960's levels fixed labor participation rate, the labor supply for each period can be calculated with formula (3).

Assuming that the economy starts in a steady state in 1960, the steady state capital-output ratio in 1960 equals savings rate divided by depreciation rate, thus yields three [5]. Now the physical capital stock as well as the economys total output can be calculated for all periods by using the equations (2) and (1). The results for the baseline scenario, where fertility, mortality, and LFPR for men and women remain constant at 1960's levels in every period until 2050, are summarized in Table 2.

The second step is to consider the scenario in which the fertility rate changes with respect to the actual and predicted rates. In similar manner as for the baseline scenario, the population is estimated via equations (4) and (5). The impact of the fertility change on labor, and subsequently also on capital and output, will differ depending on which effect is regarded (Figure 2).

---

[3] This means age group 5-9 until age group 100+ so that $i \geq 1$

[4] This approach is consistent with Bloom et al.'s method

[5] See the calculation in the appendix A.3. With (**??**) K_t can be calculated.

As mentioned above, I will first observe how the fertility affects the per capita income when only the Solow effect is considered. Therefore, the labor-to-population ratio is assumed to be constant at its 1960's level. Expectedly, a decline in fertility implies a lower population growth and thus declining labor supply. Hence, they lead to a higher capital-labor ratio and higher per capita income. In our case, the output, and especially the output per capita increases indeed, but the capital-labor ratio declines on average relative to the baseline (Table 3).

Overall, the Solow effect, which allows fertility to decline, but keeps the labor-to-population ratio constant, raises the per capita output by 12 percent relative to the baseline.

As mentioned, the demographic transition also translates into changes in the age-structure. By keeping the participation rates constant at 1960's level, but allowing for labor to react to the declining population, the Solow-plus-age-structure effect can be portrayed (Table 4). Theoretically, a fertility decline leads to a shrinking youth cohort, thus a declining youth dependency ratio, and is considered to have a positive effect on the labor supply and the income per capita. In my case, however, the impact of the age-structure itself on the per capita output is on average 13 percent lower than in the baseline scenario. Combining Solow and age-structure effect, the per capita output is still 2 percent smaller relative to the case where the fertility does not change. Why does the shift in age-structure have such a negative impact on economic growth in Japan? The answer is quite simple: While the youth-dependency ratio indeed decreases, the old-age-dependency ratio increases more strongly (Figure 3 and 4).

Given the fertility decline Japan saw, and is predicted to see over the given period, the YDR is smaller in every period, while the OADR is larger in every period compared to the baseline. Starting in 2010 the OADR begins to appear noticeably higher than in the baseline scenario. Overall, the per capita output showed positive growth from 1960 until 2005 when considering the age structure effect, due to a predominant decline in YDR, but for the period from 2010 to 2050 the simulation predicts rapid decreases, due to a dominant OADR. One reason for the downward trend in this period is, that Japan's baby boomers, who were born between 1947 and 1949, are gradually retiring between 2007-2014. [6] Combining Solow and age structure effect, I find higher growth from 1960 until 2010, and lower growth from 2015 until 2050 than in the baseline, and altogether, the lower growth rate outweighs the higher growth of income per capita relative to the baseline.

Furthermore, the changes in fertility are likely to induce behavioral changes as well. As Bloom et al. showed in their paper "Fertility, Female Labor Force Participation, and the Demographic Dividend" (2007)[7], a decline in fertility has a strong positive impact on the female labor participation. To include the effect of this behavioral response of women, I allow the female labor force participation rate to adjust to the lower fertility in line with their regression result for β_i reported in

[6] Many in the baby boom generation left the labor market at the age of 60, which used to be the legal retirement age until 2001. Since then, the government gradually raised the statutory retirement age to 65 years.

[7] See appendix A.1 and A.2 for a detailed explanation of Bloom et al.'s approach

Table 1.

In the medium run, this response leads to an increase of the steady state income per capita of 4 percent additional to the Solow-plus-age-structure effect (Table 5). Combining all three, the Solow, the age structure and the behavioral effect, implies 1.3 percent increase of the per capita output relative to the baseline (Table 6). [8]

4 Discussion

First of all, it is striking that the estimated population decreases that drastically in all scenarios. When reality checking the numbers with Japan's historical population data until 2010, there exists a large discrepancy. It needs to be further tested whether this is due to the fact that mortality rates are fixed on relatively high levels of 1960, or due to any other errors.
Overall, the model I used, which aims to observe the effects of fertility decline on the economic performance, might not have been the eligible approach in the case of Japan. Even though the country is a prime example for an aging society, with low fertility and mortality rates, and high life expectancy, it did not see drastic declines in the birth rates, the crucial factor for the model. The Solow effect did not affect Japan's economy over an increased capital-to-labor ratio for all periods, as it was expected because de facto, the labor did not decline for the second half of the given period. The behavioral effect of FLP indeed showed a positive growth rate of income per capita compared to the baseline, but did not occur as powerful as I hoped. This is due to the fact that I used Bloom et al's estimations for the marginal effect of fertility on FLP, which they found to be negative for all age groups between 20 and 64 years. Since the absolute change in fertility compared to the baseline does not show a downwards trend in every period, Japan's FLP decreases in those periods. Moreover, Bloom et al. stated in their paper that there exists a positive correlation between female labor supply and fertility for OECD countries, where fertility rates are already low. Therefore, Bloom et al.'s regression results on the marginal effect of fertility on FLP, were, supposedly, not the right choice to explain how Japanese women's behavior towards labor supply changes, given the demographic change. In Japan the FLP is low despite a relatively low fertility rate. Maybe, it would have been more beneficial to include another instrument than fertility, to display the behavioral change of women more precisely.

Nevertheless, the age structure effect was helpful to understand Japan's demographics and the challenges the country is facing now and in future. The large negative effect the baby boomers pose currently, pursuant to my simulation results, would be even more grave if declined mortality and increased life expectancy are considered. At the same time, one should bear in mind that while the elderly are growing in absolute numbers, they have also become substantially healthier (Bloom et al., 2011). This phenomenon, referred to as the *compression of morbidity*, whose empirical

[8] The effects are roughly multiplicative, even though the labor supply tends to reduce the capital-to-labor ratio somewhat.

occurrence is supported by most studies, suggests that the years people spend ill are compressed over the last decades of the life (Fries, 1989; Crimmins, 2004; Costa, 2002). This means, that the burden of demographic aging is smaller than anticipated (Fries, 1980).

For Japan, it is notable that actually 42.2 percent of the population over 65 years are still participating in the labor force. This percentage is almost twice as high as the OECD level (OECD, 2013). When taking into consideration the increased longevity, the demographic transition towards an aging society is likely to result in increasing labor participation of the old-aged cohort, especially when taking into account the good state of health of the Japanese population (The Commonwealth Fund, 2013), and the compression of morbidity.

Thus, the age-structure effect presumably has a more positive impact on the per capita income.

5 Conclusion

At first glance the demographic transition towards a smaller and older society appears to be ominous. Developed countries are forecast to be severely affected, especially since they already "exhausted" the demographic dividend brought about by the post-war baby boomers, and are now suffering from rising old-age dependency ratios. However, many research papers indicate that the impact of changes in demographics will not be as grave as predicted and might even affect economic growth in a positive manner.

Supportively, my simulation shows, that, even for matured economies as Japan, a fertility decline comes along with increased income per capita, despite a "disadvantageous" age structure and high old-age dependency ratio.

The simulation further alludes to the significance of observing the different channels through which changes in demographic factors, influence economic growth. The classical growth theory approach by Solow neglects shifts in age structure, by holding the ratio of labor participation rate to the population constant. Thus, the Solow-effect underestimates the per capita income, if the decline in youth dependency ratio dominates the increases in old-age dependency ratio. Conversely it overestimates the growth if the relative share of old-aged, compared to the number of people in working-age, overweighs the declined relative share of young people in relation to the working-age cohort.

Moreover, as my simulation shows, behavioral effects should also be added to describe the relationship between fertility and per capita output more accurately. While demographic indicators alter, individual behavior adapts in order to maximize the utility, given the changing environment. Central for this paper was the increased female labor supply as response to declining fertility. As estimated by Bloom et al. (2007) the negative impact of birth rates on FLP rates is large and shows high persistence up to women in infertile age.

Like Japan, many other developed countries have birth rates below replacement level. Thus, it is fundamental for governments to implement policies that help increase FLP rates, even without a further decline in fertility.

For Japan, approaches which support the combination of family and work are desirable. Furthermore, policies should be designed with regard to the currently existing labor market and social structures. These create various barriers for women who want to participate in the workforce. In light of the inexorable pressure from the rising old-age dependency ratio, the female labor force presents an auspicious chance for the Japanese economy.

Bibliography

[1] Bloom, D. E., D. Canning and B. Graham (2003). Longevity and Life-Cycle Savings. Scandinavian Journal of Economics 105(3).

[2] Bloom, D. E., D. Canning and G. Fink (2011). Implications of Population Aging for Economic Growth. Oxford Review of Economic Policy, Volume 26(4). *Referred to as Bloom et al., (2011).*

[3] Bloom, D. E., D. Canning and J. Sevilla (2003). The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change. Population Matters Monograph MR-1274, RAND, Santa Monica.

[4] Bloom, D. E., D. Canning, G. Fink and J. E. Finlay (2007). Demographic Change, Institutional Settings, and Labor Supply. Mimeo.

[5] Bloom, D. E., D. Canning, R. K. Mansfield and M. Moore (2007). Demographic Change, Social Security Systems and Savings. Journal of Monetary Economics 54(1).

[6] Bloom, D.E., and D. Canning (2000). The health and wealth of nations. Science, 12079.

[7] Bloom, D.E., D. Canning, G. Fink, and J. Finlay (2007). Fertility, Female Labor Force Participation, and the Demographic Dividend. NBER Working Paper 13583. Referred to as Bloom et al. (2007).

[8] Bloom, D.E., D. Canning, and G. Fink (2009). The Graying of Global Population and Its Macroeconomic Consequences. Program on the Global Demography of Aging, Working Paper 47.

[9] Bongaarts J. and G. Feeney (1998). On the quantum and tempo of fertility. Population and Development Review 24(2).

[10] Bongaarts, J. (2004). Population Aging and the Rising Cost of Public Pensions. Population Council, Policy Research Division Working Paper.

[11] Brander, J. A. and S. Dowrick (1994). The Role of Fertility and Population in Economic Growth. Journal of Population Economics 7(1).

[12] Costa, D.L. (2002). Changing chronic disease rates and long-term declines in functional limitation among older men. Demography 39(1).

[13] Crimmins, E. (2004). Trends in the health of the elderly. Annual Review of Public Health 25. Crimmins, E., Y. Saito, D. and Ingegneri (1997). Trends in disability-free life expectancy in the U.S.: 1970-1990. Population and Development Review, 23(3)

[14] Duncan, R. and C. Wilson (2004). Global Population Projections: is the UN getting it wrong? Rural Industries Research and Development Corporation41(4), RIRDC, Canberra.

[15] Dychtwald, K. (1999). Ken Dychtwald on the Future. San Francisco Chronicle, San Francisco.

[16] European Commission (2010). Green paper: Towards adequate, sustainable and safe European pension systems. Green Paper.

[17] European Union (2010). Working Group on Ageing Populations and Sustainability. Economic Policy Committee.

[18] Fries, J (1989). The compression of morbidity: near or far? Milbank Quarterly 67.

[19] Fries, J. (1980). Aging, Natural Death and the Compression of Morbidity. New England Journal of Medicine 303.

[20] Goldin, C. (2005). From the Valley to the Summit: A Brief History of the Quiet Revolution That Transformed Women's Work. Regional Review 14.

[21] Haghirian, P. (2010). Understanding Japanese Management Practices. Business Expert Publishing.

[22] ILO Bureau of Statistics (2009). ILO Database on Labour Statistics. International Labour Organization.

[23] Kelley, A. C. and R. M. Schmidt (1995). Aggregate Population and Economic Growth Correlations: The Role of the Components of Demographic Change. Demography 32(4).

[24] Kulish M., Smith K. and Kent, C. (2006) Ageing, retirement and savings: a general equilibrium analysis. Reserve Bank of Australia Research Discussion Paper (2006-06).

[25] Lee, R. and A. Mason (2010). Fertility, Human Capital, and Economic Growth over the Demographic Transition. European Journal of Population 26.

[26] Lee, R. D. (2003). The demographic transition: three centuries of fundamental change. Journal of Economic Perspectives 17(4).

[27] Lutz, W., A. Goujon, S. KC, and W. Sanderson (2007). Reconstruction of population by age, sex and level of educational attainment of 120 countries for 1970-2000. Vienna Yearbook of Population Research.

[28] Mammen, K. and C. Paxson (2000). Women's Work and Economic Development. The Journal of Economic Perspectives 14(4).

[29] Ministry of Internal Affairs and Communications (2011). The Statistical Handbook of Japan 2011.

[30] OECD (2013). Pensions at a glance 2013: Country Profiles Japan.

[31] Prettner, K. and A. Prskawetz (2010). Demographic Change in Models of Endogenous Economic Growth. A Survey. Working Paper, Vienna Institute of Demography.

[32] Sanderson, W.C., and S. Scherbov (2010). Remeasuring Aging. Science 239.

[33] Smith, J. P. and M. P. Ward (1985). Time-Series Growth in the Female Labor Force. Journal of Labor Economics 3(1).

[34] Statistics Bureau (2011). Preliminary Sample Tabulation of the 2010 Population Census of Japan. Director-General for Policy Planning and Statistical Research and Training Institute, Ministry of Internal Affairs and Communications.

[35] Steinberg, C. and M. Nakane (2012). Can Woman Save Japan? IMF Working Papers, Asia and Pacific Department.

[36] Süssmuth-Dyckerhoff, C., J. Wang, and J. Chen (2012). Women Matter: An Asian Perspective. McKinsey & Company.

[37] The Commonwealth Fund (2013). International Profiles of Health Care Systems, 2013.

[38] The World Bank (2013). Gender Statistics. World DataBank.

[39] Toossi, M. (2011). A behavioral model for projecting the labor force participation rate. Monthly Labor Review.

[40] Tyers, R. and Q. Shi (2007). Global Demographic Change, Policy Responses and Their Economic Implications. 43(3).

[41] United Nations (2012). World Population Prospects: The 2012 Revision.

[42] Vitali, A., A. Aassve, F. Furstenberg (2013). Wealth Inequalities across Generations. Centre for Population Change, Working Paper 46.

[43] Weber, L. (2010). Demographic Change and Economic Growth Simulation on Growth Models. Springer Verlag, Berlin, Heidelberg.

[44] World Economic Forum(2010) The Corporate Gender Gap Report of 2010.

[45] Yamaguchi, K. (2008). The Way to Resolve the Wage Gap between Male and Female Workers: Theoretical and Empirical Evidence of Irrationality of Statistical Discrimination. The Japanese Journal of Labor Studies 574.

The Selection Effect by Priest and Klein (1984): Robustness towards the relaxation of initial assumptions

Silvan Häs[1]

1 Introduction

In their 1984 article Priest and Klein discuss which factors affect whether a dispute gets settled or litigated in court. Their model, refered to as Priest-Klein-model, provides a framework under which a *selection effect* can be derived from certain assumptions. This means, that not an arbitrary collection of disputes gets settled or litigated, but that the selection very well depends on e.g. the cases' value as well as litigation costs, settlement costs and so on.

The following paper will derive under which assumptions such a selection effect can be observed and how much these assumptions can be relaxed.

In the second section I will provide a brief introduction on the ideas of the Priest-Klein model, section three will focus on the mathematical properties of the original model and how they can be relaxed according to Lee and Klerman (2014). Section four discusses a discrete model by Schweizer (2014). In section five argues whether or not the selection effect can still be observed when some assumptions have been relaxed.

2 The Selection Effect by Priest and Klein (1984)

2.1 The Priest-Klein-model

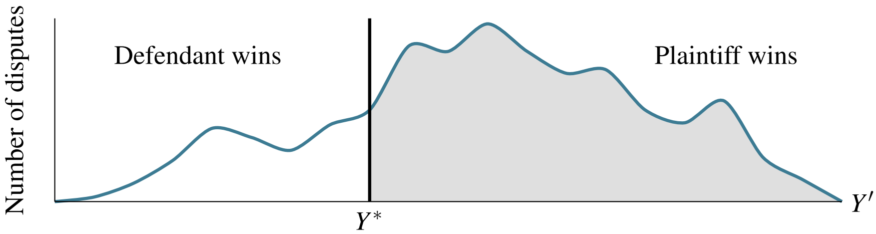
Priest and Klein derive in their article "The Selection of Disputes for Litigation" from the year 1984 the existence of a selection effect in disputes before civil courts. Intuitively one might argue, that "very close" cases are more likely to be fought about in court than other, less close, disputes.

Below I will give a brief introduction to the Priest-Klein-model.

First of all we assume, that all cases of interest either get settled out of court or end with litigation. That means we exclude the possiblity of the plaintiff dropping the case for some reason.

[1]Silvan Häs received his degree (B.Sc.) from the University of Bonn in 2015. The present article refers to his bachelor thesis under supervision of Prof. Dr. Urs Schweizer, which has been submitted in February 2015.

Figure 4.1: Example for arbitrary distribution of disputes - adapted from Priest and Klein (1984)



Let J_p denote the cases value to the plaintiff (e.g. the value of the plaintiff's car) and J_d the cases value to the defendant.[2] For now we will focus on the case $J_p = J_d = J$.

Furthermore we assume Y to measure the "‘level of fault"’. You can think of some (for our argument not specified) function which translates some real-life factors (e.g. was the defendant speeding, did he drink and drive, etc.) into some number Y. We denote the true value of some case as Y'. Let Y^* describe the court's decision standard. If the true value of some case is below the decision standard, i.e. $Y' < Y^*$, the defendant wins. If the true value is higher than the decision standard, the plaintiff wins (see figure 4.1).

The true values of cases are distributed by some function $g(Y')$. For reasons of simplicity Priest and Klein expect this density function to be normally distributed.[3]

Before deciding whether they want to settle their dispute or go to court both plaintiff and defendant try to estimate the cases' true value, Y'_p and Y'_d respectively:

$$Y'_p = Y' + \varepsilon_p \tag{1a}$$

and

$$Y'_d = Y' + \varepsilon_d \tag{1b}$$

Priest and Klein assume the error terms ε_p and ε_d to be independently and normally distributed with an expected value of Zero and standard deviation $\sigma_p = \sigma_d = \sigma_\varepsilon$ (i.e. $\varepsilon_p,\varepsilon_d \sim \mathcal{N}(0,\sigma_\varepsilon)$). Here we expect the degree of uncertainty to be equal for both plaintiff P and defendant D.

From the assumption made above we can derive the probability the *plaintiff* wins from the plaintiff's and defendant's perspective, respectively:

$$P_p = P\left(Y' \geq Y^* \mid Y'_p\right)$$

and

$$P_d = P\left(Y' \geq Y^* \mid Y'_d\right)$$

[2] An intuitive example for different values of J_p und J_d provided by Priest and Klein would be that e.g. the defendend is a well known firm that does not only face the monetary threat of losing the case in court, but might also lose reputation leading to smaller profits in the future.

[3] Lee and Klerman relax this assumption. See section three.

In words: The probability that P wins is the probability that the true level of fault is bigger or equal the court's decision standard, given the signal plaintiff or defendant observed, Y_p' or Y_d'. By combining the equations above with (1a) and (1b) we get:

$$P_p = P\left(\varepsilon_p \leq Y_p' - Y^*\right) \tag{2a}$$

and

$$P_d = P\left(\varepsilon_d \leq Y_d' - Y^*\right) \tag{2b}$$

It is convenient to normalize the decision standard at Zero.[4] Combined with our distributional assumption about the error term (ε_p,$\varepsilon_d \sim \mathcal{N}(0, \sigma_\varepsilon)$) we get the following subjective probabilities:

$$P_p = F_p\left(Y_p'\right) \tag{3a}$$

and

$$P_d = F_d\left(Y_d'\right) \tag{3b}$$

For a better understanding of the above equations think of the following example: If the plaintiff estimates the true value of the case to be two, which means he receives the signal $Y_p' = 2$, he wins in court, as long as his error term has a value smaller or equal to two.

Furthermore assume the costs of litigation to be C_p and C_d, respectively. Analogously we can define the settlement-costs by S_p and S_d. It is quite intuitive to assume that litigation costs exceed settlement costs. The plaintiff is willing to accept any settlement offer that exceeds his expected value of litigation

$$Z_P^{\min} = P_p J - C_p + S_p \tag{4}$$

and the defendant is willing to pay any settlement offer which is smaller than her expected loss of litigation:

$$Z_D^{\max} = P_d J + C_d - S_d \tag{5}$$

In order for a dispute to end up in court, the defendant's maximum settlement offer must be smaller than the plaintiff's minimum demand, i.e. $Z_P^{\min} > Z_D^{\max}$. After inserting the above expressions in this inequality and rearranging we get:

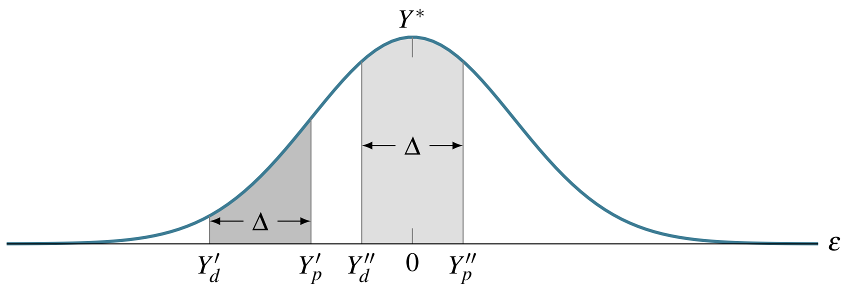
$$P_p - P_d > \frac{C - S}{J} \tag{6}$$

Where $C = C_p + C_d$ and $S = S_p + S_d$. Note that, if $C > S$, we get $0 < \frac{C-S}{J} < 1$.

At this point Priest and Klein present a simple graphical intuition for the existence of a selection effect on which disputes get litigated in court and which cases are settled (see figure 4.2). Note

[4]Note that this holds without loss of generality.

Figure 4.2: Different estimates of Y' - adapted from Priest and Klein (1984)



that equation (6) and (3) combined, for the case displayed in figure 4.2, are:

$$F_p\left(Y'_p\right) - F_d\left(Y'_d\right) > \frac{C-S}{J}$$

and

$$F_p\left(Y''_p\right) - F_d\left(Y''_d\right) > \frac{C-S}{J}$$

Y'_p and Y'_d represent one possible set of signals far away from the decision standard. Y''_p and Y''_d are signals very close to the decision standard. For the first equation the shaded area and for the second equation the light-toned area need to be bigger than $\frac{C-S}{J}$ in order for a dispute to get litigated in court. Since the location of the signals relative to the decision standard seems to have a great impact on the probability for litigation, figure 4.2 provides a first graphical intuition for the existence of some selection effect. Suppose the true level of fault in the first and second case is right in the middle of Y'_p, Y'_d and Y''_p, Y''_d, respectively. Since the error term is independent of the location of decision standard and level of fault, the observations Y'_p and Y'_d, given Y', as well as Y''_p and Y''_d, given Y'', are equally likely. Still it might be the case that the first dispute gets settled and the second one gets litigated, just because of the different location of the level of fault.

One step further Priest and Klein set up the 50%-hypothesis. This hypothesis states that the smaller the disputants errors about the true value of the case, the fewer cases are litigated and those which are tend to a verdict-rate of 50% (For graphical intuition, see figure 4.3).[5]

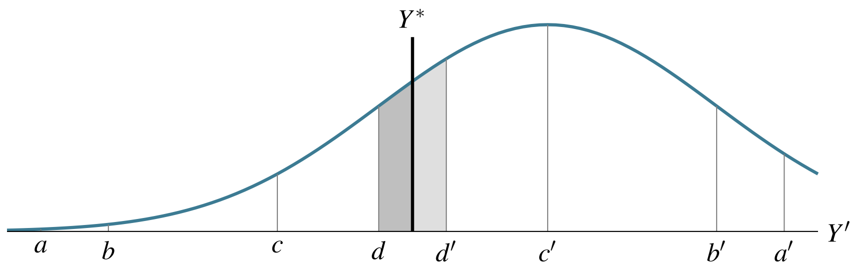
2.2 Summary

Priest and Klein conclude that under certain assumptions like $J_p = J_d = J$ and normally distributed error terms with equal standard deviation for both plaintiff and defendant a selection effect exists. For standard deviation σ_ε tending to Zero Priest and Klein even derive the 50%-hypothesis, which holds regardless where decision standard and true fault level are.

The attentive reader might have found some weak spots in the Priest-Klein-model. First of all any possible bargaining about the value of a possible settlement payment is excluded. It is simply

[5] I will not go into further detail in this paper. For more information on the 50%-hyopthesis read Priest and Klein (1984) and for a formal proof Lee and Klerman (2014).

Figure 4.3: 50%-hypothesis with normally distributed Y' - adapted Priest and Klein (1984)



assumed that, as soon as $Z_P^{\min} < Z_D^{\max}$ the case is settled even though $Z_P^{\min}$ as well as $Z_D^{\max}$ are, at least at the beginning, private information.

Furthermore, if the result of the bargain only depends on the signals plaintiff and defendant received, information must be exchanged between the disputants at least to a certain degree. This additional information is not used by the disputants in order to re-evaluate their estimate of P_p and P_d respectively.

Regardless of possible objections on the Priest-Klein model, I will point out the idea of the model in order to clarify it for further discussion. Two disputants, plaintiff and defendant, have a dispute about some damage J. If the defendants fault level were bigger than some decision standard, plaintiff would win in court and the defendant is obliged to pay plaintiff some compensation J. If the fault level is sufficiently low, i.e. the court's decision standard is not exceeded by the fault level, plaintiff gets nothing. Both plaintiff and defendant try to estimate defendant's fault level. From this estimation and some assumptions about the estimation's error term, plaintiff and defendant derive their own estimates about the probability that plaintiff wins. This probability leads to an expected compensation payment (defendant) and compensation received (plaintiff). The dispute stays out of court as long as the defendant's willingness to pay exceeds the plaintiffs expected value of going to court.

3 Formal analysis of the selection effect

Lee and Klerman (2014) provides a mathematically formal analysis of the selection effect introduced above. So far our derivation of the selection effect has been a graphical and intuitive one. In this section the emphasis will be on deriving a formal representation of the selection effect dependant on assumptions as week as possible.

3.1 Derivation of the selection effect according to Lee and Klerman (2014)

The notation introduced in section 2.1 will be used and extended in this section.

In order to make a formal argument the proportion of all cases decided in court in the plaintiff's favour is of interest. Since all cases with a true fault-level Y' bigger or equal to the court's decision standard Y^* lead to a verdict in plaintiff's favour, and the true value of all cases is distributed with some density function $g(Y')$, plaintiff's winning rate would be:

$$\int_{Y^*}^{\infty} g(Y')\mathrm{d}Y' \tag{7}$$

Note that $g(Y')$ states the relative frequency certain cases occur. Lee and Klerman relax the assumptions on the distribution of the error terms ε_p and ε_d. For our purposes it is sufficient to let ε_p and ε_d be distributed according to the two-dimensional normal distribution $f_{\sigma_p,\sigma_d}(\varepsilon_p, \varepsilon_d)$ with mean 0 and standard error σ_p und σ_d. Note that Priest and Klein assumed $\sigma_p = \sigma_d = \sigma_\varepsilon$. Also we expect ε_p and ε_d to be uncorrelated. In the next step we standardize the expected winning probabilities of Plaintiff over the different σ-values:

$$P_p = F_p\left(\frac{Y' + \varepsilon_p - Y^*}{\sigma_p}\right) \text{ sowie } P_d = F_d\left(\frac{Y' + \varepsilon_d - Y^*}{\sigma_d}\right)$$

Note that now both F_p and F_d are the distribution function belonging to the standard normal distribution. For the condition for litigation derived in (6) we now get:

$$F_p\left(\frac{Y' + \varepsilon_p - Y^*}{\sigma_p}\right) - F_d\left(\frac{Y' + \varepsilon_d - Y^*}{\sigma_d}\right) > \frac{C - S}{J}$$

It is important to understand that here $F_p(\cdots)$ and $F_d(\cdots)$ are univariate standard-normal distributed even though they are part of the two dimensional density-function $f_{\sigma_p,\sigma_d}(\varepsilon_p, \varepsilon_d)$.

Lee and Klerman define a function R indicating which tuples $(\varepsilon_p, \varepsilon_d)$ fulfil the condition for litigation:

$$R_{\sigma_p,\sigma_d}\left(Y', Y^*\right) = \left\{ (\varepsilon_p, \varepsilon_d) \in \mathbb{R}^2 \;\middle|\; F_p\left(\frac{Y' + \varepsilon_p - Y^*}{\sigma_p}\right) - F_d\left(\frac{Y' + \varepsilon_d - Y^*}{\sigma_d}\right) > \frac{C - S}{J} \right\}$$

Note R gives us all values $(\varepsilon_p, \varepsilon_d)$ for which, given decision standard and true value of the case, the case gets litigated in court.

Now we want to compute the probability that a case with given values for Y' and Y^* gets litigated in court in stead of settled. Let Π denote this probability function. By using R and the error-distribution-function's properties described above, we get:

$$\Pi_{\sigma_p,\sigma_d}\left(Y', Y^*\right) = \iint\limits_{R_{\sigma_p,\sigma_d}(Y',Y^*)} f_{\sigma_p,\sigma_d}(\varepsilon_p, \varepsilon_d)\,\mathrm{d}\varepsilon_p \mathrm{d}\varepsilon_d \tag{8}$$

Intuitively the double integral sums up all probabilities of all touples leading to litigation. Note that R_{σ_p,σ_d} depends on the case's true value and the court's decision standard. Since Π_{σ_p,σ_d} depends on R_{σ_p,σ_d}, the probability function Π_{σ_p,σ_d} is also dependant on decision standard and fault level.

The final step of the formal derivation of the selection effect is to find an argument that gives us the fraction of cases won by the plaintiff *in court*. If this fraction is different from the plaintiff's hypothetical winning rate if all cases were litigated, then clearly there is some non-arbitrary selection of cases which are litigated in court.

The Function Π indicates with which probability a case of value Y' gets litigated in court with decision standard Y^*. $g(Y')$ gives us the probability that a certain case of value Y' occurs. Hence the product of those two functions gives us the probability that a random case out of the set of all cases gets litigated. Since the set of all cases is continuous and a case gets won by the plaintiff if $Y' > Y^*$, we get the following fraction of cases litigated in court and won by plaintiff:

$$W_{\sigma_p,\sigma_d}(Y^*) = \frac{\int_{Y^*}^{\infty} \Pi_{\sigma_p,\sigma_d}(Y',Y^*)g(Y')\mathrm{d}Y'}{\int_{-\infty}^{\infty} \Pi_{\sigma_p,\sigma_d}(Y',Y^*)g(Y')\mathrm{d}Y'} \tag{9}$$

Recall that above we derived the fraction of cases won by plaintiff if all cases were litigated as $\int_{Y^*}^{\infty} g(Y')\mathrm{d}Y'$ (see (7)), which is clearly different from our argument of $W_{\sigma_p,\sigma_d}(Y^*)$. Hence Lee and Klerman formally proved the existence of a selection effect introduced in the Priest-Klein-model.

3.2 Conclusion

Lee and Klerman (2014) not only proved the existence of the selection effect, but also managed to relax the assumptions necessary to do so. Unlike the assumptions made above, which base upon those imposed by Priest and Klein (1984), Lee and Klerman allow the error-terms to be correlated and be distributed according to other functions than the normal distribution. Only for the prove of the 50%-hypothesis it is necessary to assume that the distribution function is log-concave.

Hence for our purposes of proving the existence of the selection effect, no restrictions on the behaviour of the error-terms need to be made.

Also Priest and Klein expect the true value of all cases Y' to be normally distributed. Here Lee and Klerman don't need to make any assumptions about $g(Y')$ other than that is is a density function which is bound above, strictly larger than 0 and continuous at $Y*$.

4 Discrete Example of the Selection Effect

4.1 Introduction

In this subsection I will introduce a discrete model by Schweizer (2014). In contrast to the work of Lee and Klerman, the article by Schweizer (2014) builds up a new and simplified model based on the ideas of Priest and Klein (1984).

Note that, in order to keep the notation consistent through this article, it differs substantially from the original notation used by Schweizer (2014).

4.2 Formal structure

In contrast to the original Priest-Klein model Schweizer assumes that cost of litigation gets split up according to the "english rule" in stead of - as in the Priest-Klein model - every party pays its own fees. Note that this difference does not to different outcomes for our view on the existence of the selection effect.

As a result, we get the following willingness to accept and willingness to pay for plaintiff and defendant, respectively:

$$\begin{aligned} Z_P^{\min} &= P_p \cdot J - (1-P_p) \cdot C \\ Z_D^{\max} &= P_d \cdot (J+C) \end{aligned}$$

A case ends up in court iff $Z_P^{\min} > Z_D^{\max}$. Therefore after some rearranging we get the following inequality similar to (6):

$$P_p - P_d > \frac{C}{C+J} \tag{10}$$

Like in the Priest-Klein model Schweizer assumes here, that every case has a true value Y'. This value is not expected to be distributed according to a continuous function $g(Y')$ (e.g. the normal distribution), but discrete with

$$Y_0 < Y_1 < \cdots < Y_n < Y_{n+1} \tag{11}$$

and related probabilities $g_i(Y' = Y_i)$, with property $\sum_{i=0}^{n+1} g_i(Y_i) = 1$. Plaintiff and Defendant, as before, make estimates Y'_p and Y'_d about the case's true value. But now those estimates are assumed to be at most one step away from the true value. Therefore the estimates are made like in (1a) and (1b) with the difference that the estimation error $\varepsilon_{p,d}$ is distributed as

$$\varepsilon_p = \begin{cases} 1 & \text{with probability } \xi_p \\ 0 & \text{with probability } 1-2\xi_p \\ -1 & \text{with probability } \xi_p \end{cases} \tag{12a}$$

and

$$\varepsilon_d = \begin{cases} 1 & \text{with probability } \xi_d \\ 0 & \text{with probability } 1-2\xi_d \\ -1 & \text{with probability } \xi_d \end{cases} \tag{12b}$$

respectively, where the following holds:

$$\xi_{p,d} = \text{prob}\underbrace{\left(Y'_{p,d} = Y_{i-1} \mid Y' = Y_i\right)}_{Y' \text{ undererstimated}} = \text{prob}\underbrace{\left(Y'_{p,d} = Y_{i+1} \mid Y' = Y_i\right)}_{Y' \text{ overerstimated}}$$

and for the right estimation of Y':

$$1 - 2\xi_{p,d} = \text{prob}\left(Y'_{p,d} = Y_i \mid Y' = Y_i\right)$$

Since in a discrete model with finitely many possible true values the behaviour of those probabilities at the corners are different and estimation therefore gets biased, the above statements only hold in the interval $1 \leq i \leq n$. At the corners Schweizer assumes the estimations to be correct:[6]

$$\text{prob}\left(Y'_{p,d} = Y_0 \mid Y' = Y_0\right) = \text{prob}\left(Y'_{p,d} = Y_{n+1} \mid Y' = Y_{n+1}\right) = 1 \tag{13}$$

Similar to (2a) and (2b) we now derive the probability that plaintiff wins out of both parties perspective:

$$\begin{aligned} P_p^i &= \text{prob}\left(Y' \geq Y^* = Y_s \mid Y'_p = Y_i\right) \\ P_d^i &= \text{prob}\left(Y' \geq Y^* = Y_s \mid Y'_d = Y_i\right) \end{aligned}$$

Here Schweizer makes another technical assumption in order to avoid problematic corner solutions. He assumes $Y^* = Y_s$ to be in the interval $2 \leq s \leq n-1$.[7]

In the Schweizer model if a player receives a signal more than one step below the decision standard, he knows for sure that plaintiff will lose in court. Also, if he receives a signal higher that the decision standard, he knows that plaintiff will win. Formally: $P^i_{p,d} = 0$ for $i < s-1$ and $P^i_{p,d} = 1$ for $i > s$. In Case of $i = s-1$, plaintiff only wins if the signal $Y'_{p,d}$ is one step below the case's true value, i.e. $\varepsilon_{p,d} = -1$. This corresponds to the following probability of plaintiff winning:

$$P^{s-1}_{p,d} = \text{prob}\left(Y' = Y_s \mid Y'_{p,d} = Y_{s-1}\right) = \xi_{p,d}$$

For $i = s$, plaintiff wins, if $Y'_{p,d}$ is either correct or underestimates the case:

$$P^s_{p,d} = \text{prob}\overbrace{\left(Y' = Y_s \mid Y'_{p,d} = Y_s\right)}^{\text{estimate correct}} + \text{prob}\overbrace{\left(Y' = Y_{s+1} \mid Y'_{p,d} = Y_s\right)}^{\text{underestimated}} = 1 - \xi_{p,d}$$

Since litigation cost C as well as the value J can be expected bigger than 0, inequality (10) can only hold if $P_p > P_d$.

At this point Schweizer makes some crucial assumptions rather different from the idea of the Priest-Klein model. First he claims any case in which plaintiff receives a signal of $Y'_p = Y^* = Y_s$

[6] Note that this assumption violates the claim in the Priest-Klein model, that the distribution of error terms must be independent of Y'. But this appears to be a minor problem, since for sufficiently large n the probability of observing a true value at one of the corners tends to zero.

[7] If e.g. $Y^* = Y_s = Y_1$, the agent getting a signal of $Y'_{p,d} = Y_1$ would know for sure that plaintiff wins.

and defendant receives $Y_d' = Y_{s-1}$ will get litigated in court:

$$\underbrace{1-\xi_p}_{=P_p^s} - \underbrace{\xi_d}_{=P_d^{s-1}} > \frac{C}{C+J} \tag{14}$$

Also the following assumption is made:[8]

$$\xi_p < \frac{C}{C+J} \quad \text{and} \quad \xi_d < \frac{C}{C+J} \tag{15}$$

These two assumptions combined prevent any other cases than the following to be litigated in court:

$$\left(Y_p', Y_d'\right) = \left\{ \left(Y_s, Y_{s-2}\right), \left(Y_s, Y_{s-1}\right), \left(Y_{s+1}, Y_{s-1}\right) \right\}$$

For each of those three cases in which litigation occurs we can attach some possible true value. Let $\omega = \left(Y', Y_p', Y_d'\right)$ denote such a combination of the true value and plaintiff's and defendant's signal. Note that plaintiff always wins, if the true value of the case is bigger or equal than some decision standard. Therefore:

$$\omega \in \Omega_s^p \text{ if } \omega = \left(Y_s, Y_{s+1}, Y_{s-1}\right) \text{ or } \left(Y_s, Y_s, Y_{s-1}\right)$$

Defendant wins in court if $Y' < Y^*$:

$$\omega \in \Omega_s^d \text{ if } \omega = \left(Y_{s-1}, Y_s, Y_{s-1}\right) \text{ or } \left(Y_{s-1}, Y_s, Y_{s-2}\right)$$

Where Ω_s^p (Ω_s^d) denotes the set of combinations for which plaintiff (defendant) wins the case. Here the existence of the selection effect in the Schweizer model gets easy to show. If all cases were litigated in court, similar to the derivation in the previous section, the proportion of all cases won by plaintiff would be $\sum_{i=s}^{n+1} g_i(Y_i)$.

Since only those cases mentioned above are litigated in court, the proportion of all cases won by plaintiff in court can be denoted as $\frac{\pi(\Omega_s^p)}{\pi(\Omega_s)}$ where $\Omega_s = \Omega_s^p \cup \Omega_s^d$ and $\pi(\cdot)$ denotes a probability. Those probabilities can be computed as stated below:

$$\begin{aligned}
\pi\left(\Omega_s^p\right) &= \underbrace{\pi\left[\omega = \left(Y_s, Y_{s+1}, Y_{s-1}\right)\right]} + \underbrace{\pi\left[\omega = \left(Y_s, Y_s, Y_{s-1}\right)\right]} \\
&= g\left(Y_s\right) \cdot \xi_p \cdot \xi_d + g\left(Y_s\right) \cdot \left(1-2\xi_p\right) \cdot \xi_d \\
\pi\left(\Omega_s^p\right) &= g\left(Y_s\right) \cdot \left(1-\xi_p\right) \cdot \xi_d
\end{aligned}$$

[8]The actual assumption made in Schweizer (2014) is $\left|P_p^s - P_d^s\right| = \left|P_p^{s-1} - P_d^{s-1}\right| = \left|\xi_p - \xi_d\right| < \frac{C}{C+J}$. This assumption is not sufficient to exclude some cases Schweizer does not discuss in the rest of his article. Therefore I changed the assumption in order to make it sufficiently strong. The change in assumption does not affect any of the results made in the article.

For the defendant:

$$\begin{aligned}
\pi\left(\Omega_s^d\right) &= \pi\left[\omega=\left(Y_{s-1},Y_s,Y_{s-1}\right)\right] &+& \quad \pi\left[\omega=\left(Y_{s-1},Y_s,Y_{s-2}\right)\right] \\
&= g\left(Y_{s-1}\right)\cdot\xi_p\cdot\left(1-2\xi_d\right) &+& \quad g\left(Y_{s-1}\right)\cdot\xi_p\cdot\xi_d \\
\pi\left(\Omega_s^d\right) &= g\left(Y_{s-1}\right)\cdot\left(1-\xi_d\right)\cdot\xi_p
\end{aligned}$$

Plugging in and rearranging leads to the following statement:

$$\frac{\pi\left(\Omega_s^p\right)}{\pi\left(\Omega_s\right)}=\frac{g\left(Y_s\right)}{g\left(Y_s\right)+g\left(Y_{s-1}\right)\cdot\frac{\xi_p\cdot\left(1-\xi_d\right)}{\left(1-\xi_p\right)\cdot\xi_d}} \tag{16}$$

This statement can be simplified if we plug in the assumption made by Priest and Klein, that the error terms for both plaintiff and defendant are identical, i.e. $\xi_p=\xi_d$:

$$\frac{\pi\left(\Omega_s^p\right)}{\pi\left(\Omega_s\right)}=\frac{g\left(Y_s\right)}{g\left(Y_s\right)+g\left(Y_{s-1}\right)} \tag{17}$$

Since both (16) and (17) are different from the proportion of cases won by plaintiff, if all cases were litigated, $\sum_{i=s}^{n+1} g_i(Y_i)$, a selection effect exists also in the discrete adaption of the Priest-Klein model by Schweizer (2014).[9]

5 Sumary of results

The aim Priest and Klein had in setting up their model was to find a model which gives reason to the observation, that cases litigated in court might not be a random sample taken from all disputes. In their model they derive the existence of a selection effect, causing that cases are more likely to be litigated if their true value lies close to some decision standard.

In section 3 this article discussed a formal proof of this selection effect by Priest and Klein. This proof manages to relax many assumptions Priest and Klein made for reasons of simplicity. Therefore the cases are allowed to be distributed with some function $g(\cdot)$, which needs only to be continuous at Y^*, bounded above and strictly bigger than zero. Furthermore the error terms ε_p and ε_d were only required to be distributed so some joint-probability function, even allowing for some correlation between the error terms, as well as different standard errors.[10]

Section 4 introduced an alternate model by Schweizer (2014). In contrast to the original Priest-Klein model, the Schweizer model is discrete, i.e. there is some finite number of values the case's true value, decision standard and plaintiff and defendant's observations (and therefore error

[9]The attentive reader might have noticed that in this adaption the earlier mentioned 50% plaintiff winning rate only holds, if $g(\cdot)$ is uniformly distributed, which is a very strong assumption and a big difference compared to the results of the Priest-Klein model.

[10]In order to prove the 50%-hypothesis the distribution function needs to be log-concave.

terms) can have. Also Schweizer restricts the range the received signals are allowed to be off the case's true value, which is also a difference to the original model.

Robustness in the sense of model's results means, that it is possible to make different assumptions that still lead to the same or very similar basic result. Since the selection effect introduced by Priest and Klein could be proven to hold under relaxed assumptions (Lee and Klerman 2014) as well as under an entirely different discrete setting (Schweizer 2014) we can conclude that the selection effect is very robust.[11]

References

Lee, Yoon-Ho Alex and Daniel Klerman (2014). "The Priest-Klein Hypothesis: Proofs, Generality and Extensions". USC CLASS Research Paper No. 14-34; USC Law Legal Studies Paper No. 14-43.

Priest, George L. and Benjamin Klein (1984). "The Selection of Disputes for Litigation". In: *The Journal of Legal Studies* 13.1, pp. 1–55.

Schweizer, Urs (2014). "Litigated Cases: the Selection Effect Revisited". Mimeo.

[11] As an outlook note that the 50%-hypothesis appears to be way less robust. In the Schweizer model the hypothesis can only be derived under the assumption that the true value is uniformly distributed and also does not depend on the limit of uncertainty tending to 0.

High-Frequency Trading and Fundamental Price Efficiency

Simon Niklas Maria Schmickler[1]

1 Introduction

In a regression of medium-term price efficiency on high-frequency trading (HFT) presence, controlling for market level variables, stock exchange fixed effects, year fixed effects and clustering standard errors, the coefficient of HFT presence is -0.06 and statistically significant with a p-value of 0.017. This means price efficiency halves for the average S&P 500 firm upon the start of HFT. But market prices determine the allocation of real resources in the economy. Thus, capital flows less efficiently when HFT is present. My empirical findings are consequently consistent with the notion that HFT decreases welfare. I explain my findings with HFT front running. HF traders detect large orders of institutional investors and execute the same trade faster, free-riding on investors' private information. This discourages institutional investors from becoming informed. As a consequence, they bring less information into market prices.

For price efficiency, I use a measure which captures how well market valuations forecast future profits. I construct it similarly to Bai, Philippon, and Savov 2016 for annual forecasting horizons from one to five years. The recently estimated HFT starting dates from Aitken, Cumming, and Zhan 2015 capture HFT presence. I then identify the impact of HFT on price efficiency in a multi-event difference in differences (DiD) analysis. My data cover 13 stock exchanges. Broadly, HFT started in the United States at the turn of the century, then spilled over to Europe and other developed countries until it finally reached the emerging market India in 2009. China and South Korea, where HFT is not present yet, serve as counterfactuals. This staggered start of HFT makes my identification strategy appealing because it makes it unlikely that a simultaneous, unrelated event drives my results.

There is no consensus about the overall welfare impact of HFT. Stiglitz 2014 claims HFT is a negative sum game costing real resources while contributing no social value. The Nobel price laureate calls HFT front running and suggests the introduction of a tax. The theoretical literature largely supports this notion (Baldauf and Mollner 2015, Dugast and Foucault 2016, Jarrow and

[1]Simon Niklas Maria Schmickler received his degree (B.Sc) from the University of Bonn in 2015. The present article refers to his bachelor thesis under supervision of Prof. Dr. Hendrik Hakenes, which has been submitted in July 2015.

Protter 2012, Gao, Ivković, and S. Z. Li 2014, Yang and Zhu 2016). Yet, according to the empirical literature, HFT tends to have a positive impact on welfare by improving liquidity and lowering trading costs. Conrad, Wahal, and Xiang 2015 find an association between higher quoting activity and a lower effective spread. Further, Brogaard, Hagströmer, and Riordan 2015 find that increased speed available to market makers improves liquidity. Still, Tong 2015 finds that HFT increases execution costs for institutional investors.

In particular, the empirical literature links HFT to improved price efficiency. Using NASDAQ transaction level data, Brogaard, Hendershott, and Riordan 2014 show HF trades are correlated with positive returns over the next one and two seconds. During this time horizon HFT trades in the direction of permanent price changes while trading against pricing errors. Next, Aitken, Cumming, and Zhan 2015 use data from surveillance authorities to capture manipulation cases and find HFT reduces end-of-day market price manipulation. Finally, Boehmer, Fong, and Wu 2014) study the effect of algorithmic trading on price efficiency in an international setting. They document a positive effect of HFT on price efficiency, defined as the absence of short-term return predictability. In general, HFT has a positive effect on short-run measures of liquidity and price efficiency. But this does not necessarily imply that HFT has a positive effect on the informativeness of prices in a fundamental sense. In fact, an improvement of short-term measures and a decrease in medium- and long-term price efficiency can co-exist.

However, Stiglitz 2014 argues extreme short-term price efficiency is irrelevant to the real economy. It is thus essential to determine the causal effect of HFT on price efficiency at a longer horizon. My paper aims at closing this gap in the empirical literature. I explicitly investigate the causal effect of HFT on medium- and long-term price efficiency. I find evidence consistent with the claims of Stiglitz 2014 and go beyond. Concordantly with decreasing medium- and long-run price efficiency, HFT tends to decrease real price efficiency. In addition, my results associate HFT with a temporary increase in real information production by financial analysts. Analysts' earnings forecasts are less erroneous and coincide more.

HFT is typically characterized by four traits Jones 2013. These are first, an extreme trading speed; second, using stock exchanges' co-location services to maximize speed; Third, very short holding periods and lastly, submission and immediate cancellation of a large number of orders. According to Aitken, Cumming, and Zhan 2015, the start of HFT is not directly observable and not even precisely known to directors of stock exchanges. Aitken, Cumming, and Zhan 2015 estimate HFT starting dates using co-location, average trade size and number of canceled orders. The first only occurs years after HFT actually starts. The second does not directly come from the definition of HFT. Consequently, I use the third in my main analyses. I also show the robustness of my results to the definition of HFT.

The space constraint forces this summary to omit theoretical justifications and substantial empirical results, in particular further regression tables. Please email me at `snms@princeton.edu` if you are interested in a full version of this article.

2 Hypotheses

I aim to determine the effect of HFT on the quality of market prices. The key to this relationship is rational, generally institutional investors. Extending the argument and to be able to evaluate additional empirical findings, I introduce financial analysts as their suppliers of information. Institutional investors buy information from analysts and bring them into market prices through trades. HFT impacts on institutional investors. A change in their behavior then impacts on both, price efficiency and financial analysts.

There are two competing lines of argument. According to Stiglitz 2014, HFT reduces the rents to information via front running. This makes information acquisition less attractive and I infer that demand for information consequently drops. First, this decreases price efficiency. Second, in the short run, this implies an increase in competition between the suppliers of information, financial analysts. According to Sun 2011 this increases the quality of analyst reports. However, this relationship changes when investment banks adjust the number of analysts to the new, lower level of demand for information. In the long run, closer to the new equilibrium, analysts' forecast quality partially reverses as competition eases. This implies that from the short to the long run analyst measures deteriorate.

The contrasting line of argument is that HFT increases liquidity and decreases trading costs which increases the gains from information acquisition. Thus, demand for information increases and the effect runs exactly in the opposite direction. First, Fundamental price efficiency increases. Second, Analyst report quality decreases in the short run because competition eases and partially reverses as the system reaches a new equilibrium.

3 Research design and data

3.1 Data and variable construction

I use annual firm level data from 1990 to 2014. Accounting data are from Compustat, market data from CRSP. Analyst based variables stem from IBES databases. I greatly thank Dr. Jasmin Gider and Dr. Christian Westheide for giving me access to the data. The HFT start dates are from Aitken, Cumming, and Zhan 2015. I use the GDP deflator from the World Bank to turn nominal into real values. I also convert all values to USD using exchange rates from the Federal Reserve System. As in Bai, Philippon, and Savov 2016, I measure stock market value at the next end of march after the close of the firm's fiscal year and drop financial firms from my sample. Further, I drop all observations which do not have an HFT start date. My sample then comprises 28.270 firms and around 365.000 observations.

Table 5.1 gives an overview of my data. Comparing them to Bai, Philippon, and Savov 2016, the ratios are in line but my accounting means are substantially larger for two reasons: first, my time frame begins 30 years later. Second, while Bai, Philippon, and Savov 2016 use equity market values, I look at total market values. Total market value is the sum of equity market value and market value of debt. I use the book value of debt as proxy for its market value. Bai, Philippon, and Savov 2016 only use this specification in a robustness check. I use total market values instead of equity market values because this reduces noise in the price efficiency variable. For example, even if future EBIT was certain, a firm with great prospects but high debt may have less market value of equity per total assets than a firm with poor prospects and no debt. Here, using only market value of equity would diagnose poor price efficiency, while using total market value would truthfully reveal high price informativeness.

I use two types of dependent variables. First, I construct price efficiency similar to Bai, Philippon, and Savov 2016. This variable measures how well market prices predict future profits. It describes forecasting price efficiency. I estimate:

$$\frac{E_{i,t+k}}{A_{i,t}} = a_t \ln\left(\frac{M_{i,t}}{A_{i,t}}\right) \times 1_t + b_t \ln\left(\frac{E_{i,t}}{A_{i,t}}\right) \times 1_t + c_{s(i,t),t}(1_{SIC1}) \times 1_t + \varepsilon_{i,t} \tag{1}$$

where E is EBIT, A is total assets, M is market value of equity, $SIC1$ is the first digit of the SIC code and $k = 1,...,5$. All ratios entering equation 1 are winsorized at the 1% level. Bai, Philippon, and Savov 2016 use predicted variation$_t = a_t \times \sigma_t[ln(M_{i,t}/A_{i,t})]$ as main dependent variable. However, for my application, this variable does not have a unique interpretation. Whether more predicted variation means higher or lower price informativeness depends on total variation $= \sigma_t[E_{i,t+k}/A_{i,t}]$. Only if the share of predicted variation of total variation increases, price informativeness increases. I thus choose share of predicted variation, price efficieny k = Eff^k = predicted variation/total variation, as central regressand for my analysis. Ideally, this should yield 312 yearly stock exchange level observations from 13 exchanges over 24 years for $k = 1$. However, in 38 cases regression 1 fails for lack of observations. This leaves 274 data points.

In general, the fewer observations I construct the measure from, the more often the resulting data point is an outlier. On average, I compute it from a regression with 783 data points. But for example for the national stock exchange in Mumbai in 1993, I generate the measure from only seven observations. It takes the value minus one, even though price efficiency should intuitively be and largely is between zero and one. As I do not want outliers computed from few observations to confound my results, I winsorize price informativeness at the 2,5% level.

Similar to price efficiency, Bai, Philippon, and Savov 2016 also develop real price efficiency variables which measure how well market valuations predict future investment in R&D and capital expenditure. I estimate:

$$\begin{aligned}\frac{x_{i,t+k}}{A_{i,t}} = a_t \ln\left(\frac{M_{i,t}}{A_{i,t}}\right) \times 1_t + b_t \ln\left(\frac{x_{i,t}}{A_{i,t}}\right) \times 1_t \\ + c_t \ln\left(\frac{E_{i,t}}{A_{i,t}}\right) \times 1_t + d_{s(i,t),t}(1_{SIC1}) \times 1_t + \varepsilon_{i,t}\end{aligned} \tag{2}$$

where x is investment either measured as expenses for research and development (R&D) or capital expenditure (CAPX). I change R&D and CAPX to zero when they are missing in my dataset. Before the change, no firm had zero investment in R&D in my dataset. But in reality many, especially small firms, do not invest in R&D. I reason this does not appear in my dataset because Compustat reports zero R&D as missing. The same is valid for CAPX. As before, all ratios are winsorized at the 1% level and $k = 1,...5$. They then generate predicted variation as above. Again, I use total market value instead of market value of equity and choose real price efficiency RPE-(x)^k = predicted variation/total variation for my analysis.

Second, I use the analyst based measures from Kang and Liu 2008.

$$\text{Error} = \frac{\text{mean} - \text{actual}}{\text{S}} \tag{3}$$

$$\text{Disp} = \frac{\sigma_{i,t}(\text{forecasts})}{\text{S}} \times 100 \tag{4}$$

where mean is analysts' mean earnings forecast, actual is actual earnings, S is the stock price and $\sigma_{i,t}(\text{forecasts})$ is the standard deviation of analysts' earnings forecasts. Error and Disp measure how well financial analysts predict future profits. I winsorize the two variables at the 1% level to avoid obtaining results which are driven by extreme outliers.

Furthermore, I measure inter-day stock volatility as the monthly standard deviation of logarithmic daily return factors. My measure for liquidity is the Amihud illiquidity measure from Amihud 2002. Moreover, I construct the variable Placebo +(-)ty, with $t = 1,2,...,5$ like the main variable HFT. The only difference being that I shift the HFT starting date back or forth by t years. Lastly, I divide some measures by a constant to simplify the interpretation of their coefficients. Market value (MV), total assets, EBIT, CAPX and R&D are in million USD. Market size and average capitalization are in trillion USD. The five year rolling volatility of EBIT, average market Amihud illiquidity are in millions and average volatility is in thousands.

3.2 Empirical model and identification strategy

Similar to Christensen, Hail, and Leuz 2015 and using annual stock exchange level observations, I draw inference from the multi-event DiD model:

$$\text{Eff}^k_{e,t} = \beta_0 + \beta_1 \text{HFT}_{e,t} + \delta' X_{e,t} + \eta_t + \eta_e + \varepsilon_{e,t} \tag{5}$$

e indicates the stock exchange and t the year. Eff^k represents the market-level price efficiency measure I constructed similarly to Bai, Philippon, and Savov 2016 for the time horizons $k = 1,...,5$. HFT is zero prior to the HFT starting date and one for all following years. It is the fraction of time in which HFT was present for the year in which it started. X is a vector of control variables which consists of total market size, average market capitalization and electronic, a dummy variable capturing the effect of the transition from floor to electronic trading. η_t are year fixed effects. η_e

are stock exchange fixed effects and $\varepsilon_{e,t}$ is the error term.

I also use Analyst based measures to test my hypotheses. For this, I estimate a similar model using annual firm level observations:

$$\mathrm{IBES}_{i,t} = \beta_0 + \beta_1 \mathrm{HFT}_{i,t} + \delta' X_{i,t} + \eta_t + \eta_i + \varepsilon_{i,t} \tag{6}$$

i indicates the firm. IBES stands for the analyst based measures analyst dispersion and analyst error. Here, logarithmic market value, Tobin's Q, electronic and the rolling five year volatility of EBIT constitute the vector of control variables. η_i are firm fixed effects. Both models feature year and panel variable fixed effects. The former flexibly eliminates common trends. The latter eliminates the impact of unobservable firm or stock exchange specific characteristics. The results are thus driven by within firm and within stock exchange variation.

The key of my identification strategy is the staggered start of HFT across markets which I illustrate in figure 5.1. HFT started in North America, spilled over first to Europe, other developed countries and finally also reached emerging markets but not China and South Korea. Both countries do not have a co-location event in Aitken, Cumming, and Zhan 2015. In China it is even illegal to sell a share on the day of its acquisition Guo, Z. Li, and Tu 2012. This staggered introduction of HFT remedies the concern my results could be driven by unrelated macroeconomic shocks. Because of the use of fixed effects, they can only drive my results if they occur simultaneously in the same order going from west to east, while leaving China and South Korea unaffected. This is very unlikely. As in Christensen, Hail, and Leuz 2015, this and the concrete specification of my empirical model also imply it is not crucial to establish parallel trends.

I now tackle the threats of reverse causality and endogeneity. HFT is not a direct consequence of lower price efficiency or higher analyst report quality. Still, which mechanisms drive the dispersion in the main independent variable HFT? HF traders (HFTs) may self-select into markets where they anticipate lower price efficiency or better analyst reports. It seems possible that HFTs make larger profits in inefficient markets or with better reports. This would bias my results. However, this is not the case for three reasons. First, I do not expect HFTs to buy and study analyst reports. The typical purchaser of an analyst report is an institutional investor. Algorithms can search a report for key words which trigger trades but they cannot interpret complicated qualitative analyses. Thus, self-selection into markets with better analyst reports is not a concern. Second, HFT started in the USA in stocks of large firms Aitken, Cumming, and Zhan 2015; Arnuk and Saluzzi 2012. Price efficiency is three times as high for these stocks (S&P 500) than the global average. HFT did not start in the USA because price efficiency was low but because the technological innovation was developed there. It then spilled over first to other developed countries, especially in Europe, because these markets have similar human capital and infrastructure. HFT reached the emerging markets last because it was more difficult to implement there. This explanation for the dispersion in HFT, i.e. its staggered start across markets, is more plausible than HFTs correctly anticipating lower price efficiency in the same order.

Yet, this leads to another endogeneity concern. I argued that the dispersion in the start of HFT is driven by human and physical capital. If human and physical capital were correlated with lower price efficiency, this could introduce an endogeneity bias. However, it is not. Stock exchanges in countries with higher human and physical capital generally show higher price informativeness. For example, Eff^k is on average 30% higher at the NYSE than at the Bombay stock exchange. Additionally, the inclusion of stock exchange fixed effects mitigates this issue.

Third, HFTs are especially active in large, liquid stocks Brogaard, Hendershott, and Riordan 2015; Aldridge 2013; Tong 2015. Price informativeness is more than twice as high for the upper than for the lower half of firms, with respect to market value. A HF trader with a preference for stocks with low price efficiency would trade in small, illiquid stocks. As HFTs do not, self-selection into inefficient markets does not endanger the validity of my results. Altogether, HFT does not fall from the blue sky, but the mechanisms driving it do not threaten with an overestimation but an underestimation of its impact.

Beside HFT, there are other factors which cause variation in the dependent variable price efficiency. I control for these factors to eliminate concerns. First, there may be several time and stock exchange specific characteristics influencing the informativeness of prices. Thus, I include year and exchange fixed effects. This already captures many factors driving price informativeness. However, it does not capture the impact on price efficiency of temporary characteristics of stock exchanges or characteristics of years common to a subgroup of exchanges. Hence, I control for further characteristics. These are average market capitalization, total market value and the transition from open outcry to electronic trading. In a robustness check I additionally control for Amihud illiquidity and average volatility.

There are also other factors causing dispersion in the explained variables analyst error and dispersion. In particular, I control for firm size, the five year rolling volatility of earnings and Tobin's Q. In a robustness check, I include other measures indicating the uncertainty of profits. These are R&D expenditure, tangibility and volatility. I also include EBIT and a binary variable for the global analyst research settlement in the robustness check. Finally, for the same reasons as in the paragraph before, I control for the transition to electronic exchanges, year and panel variable fixed effects and include volatility and Amihud illiquidity in a robustness check.

Using normal ordinary least squares standard errors may lead to an overestimation of the t-statistic because of correlation between residuals and robust standard errors need not be consistent in fixed effects models either. In this subsection, I consequently use cluster-robust standard errors as described in each table. Hence, heteroscedasticity and correlation between residuals in the specific clusters do not introduce a bias Cameron and Miller 2015. The robustness checks then rule out remaining concerns about error correlation.

4 Empirical results

4.1 Identified effects

Table 5.2 shows my central result. I estimate my main empirical model with price efficiency as dependent variable. Consistent with the first line of argument, my findings suggest a negative association between HFT and price informativeness. The coefficients of HFT are significant at the 5% level for the time horizons one to four and at the 10% level for time horizon five. Taking into account the small sample size of only 271 annual stock exchange level observations, this effect is highly statistically significant. It is also economically significant. HFT halves price efficiency for the average S&P 500 firm for all five time horizons. The adjusted R^2 is high. For example, it is 0.47 for time horizon one.

I now estimate the main model, but use real price efficiency as explained variable. HFT tends to decrease real price informativeness even though the impact is less statistically significant. The effect is significant for the time horizons four and five when analyzing CAPX and the time horizons five and six for R&D expenditures. The results associate HFT with both, real price efficiency with repect to R&D and CAPX by 30% compared to their respective means. The coefficients are significant at the 5 or 10% level.

I now test the implications with respect to financial analysts. I estimate the main empirical model of this paper with analyst based measures as regressands. Table 5.3 associate HFT with better analysts' earnings forecasts. Forecast error and forecast dispersion decrease by almost half their mean value when HFT starts. This effect is highly statistically significant. With HFT, analysts predict earnings better.

The first line of argument also predicts that error and dispersion increase from the short to the longer run. To test this, I drop the observations where HFT is absent. Using my main model, I then test whether error and dispersion actually increase two, three and four years after the start of HFT. This is indeed the case. The coefficient of the dummy indicating whether HFT is present for at least four years is positive and significant for analyst dispersion and error. When taking three instead of four years it is still positive and significant for analyst dispersion. When using only two years, it is not significant. All in all, the results coincide with the first line of argument.

4.2 Robustness

To show the robustness of my findings, I first demonstrate the robustness to the definition of HFT. Aitken, Cumming, and Zhan 2015 estimate HFT starting dates as defined by number of cancellation orders per 1000 traded shares, changes in average trade size and co-location. My results are robust to these definitions. To demonstrate this, I estimate my main empirical models 5 and 6 with the explanatory variable *HFT* defined by changes in average trade size and co-location instead of, as before, by cancellation orders.

Defining HFT by changes in average trade size leaves the results almost unchanged. Only in the regression with price efficiency with horizon five years the HFT coefficient becomes just insignificant. Defining HFT by co-location weakens the statistical significance of the coefficients. This is not surprising. HF traders are active for years before the co-location event. This means when I define HFT by co-location, I assign the value zero to observations where there is in fact active HFT. Still, the coefficients of HFT remain negative and for the horizons one, two and five, as for the analyst based variables, statistically significant. Also, for both alternative definitions, the economic significance of the effect decreases only slightly.

Another concern is that my main results presented in figure 5.2 may be significant, because in this particular specification the coefficient of electronic is not significant and they closely coincide. However, first, there is a substantial gap between the transition to electronic markets and the start of HFT. While the former happened mostly during the 1990s, the latter mainly occured during the last decade. Second, when I estimate the main model excluding the HFT variable the coefficients of electronic are still insignificant. Hence, my results are not driven by this concern.

Continuing, Bertrand, Duflo, and Mullainathan 2004 highlight serial correlation in panel datasets can cause biased standard errors in DiD analyses. This leads to overrejection of the null hypothesis that a coefficient is equal to 0. I use cluster-robust standard errors in the subsection above and continue now, following Petersen 2009, to eliminate further possible issues related to correlation between errors.

When analyzing Eff^k, residuals of a stock exchange may be correlated across years because of unobservable stock exchange characteristics and residuals of a year may be correlated across stock exchanges because of unobservable year characteristics. The rows of table 5.5 show the coefficients of HFT from estimating four different versions of the main model. By using fixed effects and clustering on two dimensions, the first two rows show that neither unobservable year or stock exchange characteristics, nor temporary unobservables introduce a bias. I use the stata-command *cluster2* written by Mitchell A. Petersen which is based on Thompson 2011 and Cameron, Gelbach, and Miller 2011. The third row gives the results from a further specification adressing the same challenges. The last row uses the wild bootstrap t-technique technique from Cameron, Gelbach, and Miller 2008 to mitigate the concern that there are only few stock exchange clusters. I use the stata-command *cgmwildboot* written by Judson Caskey to implement the approach. Every estimate remains negative and statistically significant. Also, the analyst report quality results pass the equivalent robustness check. Thus, Correlation between residuals does not confound my results.

Additionally, I run a placebo test. When Bertrand, Duflo, and Mullainathan 2004 generated placebo treatments in serially correlated data, they found that 45% of the coefficients of placebo treatment variables were significant when estimating simple DiD models. Similarly, I generate random HFT starting dates and use them to construct a placebo for HFT presence. I then use this placebo as independent variable in my main model. I do this 1000 times for each of my main regressions 5 and 6 with price efficiency with time horizons one to five and analyst error and dispersion as dependent variables, respectively. I count how many coefficients of the randomly generated placebo are statistically significant at the one, five and ten percent level and calculate the resulting rejection rates. For each of the seven regressions, all three rejection rates are lower than the respective significance level. For example, in the regression with analyst error as regressand, the coefficient of the placebo is only in 0,4% of all cases significant at the 1% level. This test shows that my identification strategy is very conservative.

In a further robustness check, I estimate the main empirical models 5 and 6 with additional control variables. I excluded them previously because they are either potential channels for the effect or because I consider them redundant. For the regressions with price efficiency as regressand I add average Amihud illiquidity and average volatility. For the regressions with the analyst based measures as dependent variables I additionally add EBIT, volatility, tangibility, R&D expenditures and dummy variable for the global analyst research settlement. For both models, all coefficients of HFT remain negative and statistically significant. Also, economic significance decreases only slightly.

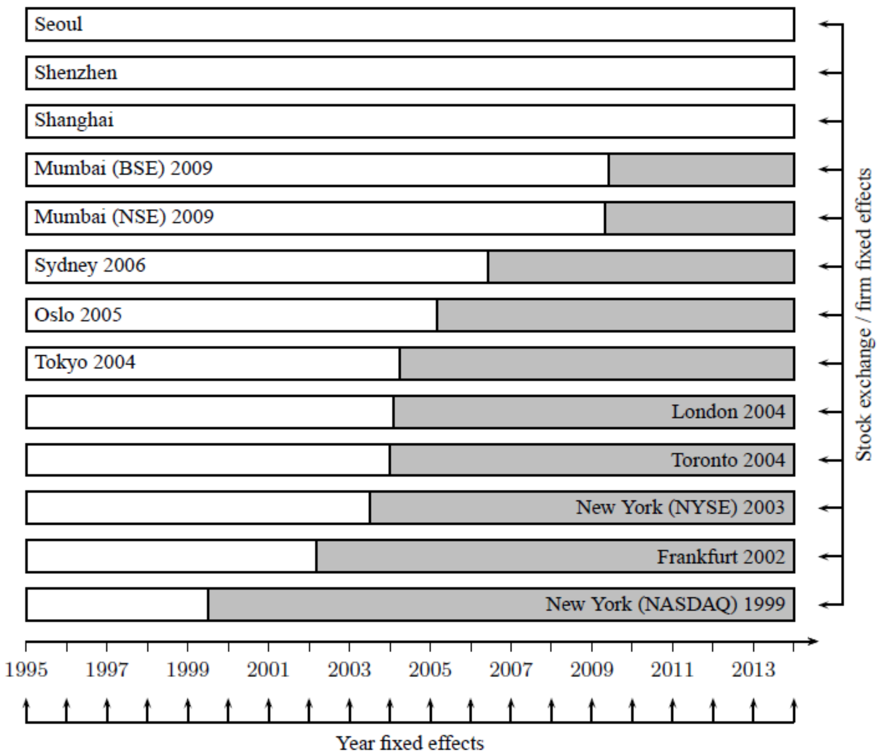
5 Conclusion

The empirical evidence I present suggests that HFT reduces price efficiency. With HFT, market valuations predict future earnings less precisely. HFT also tends to decrease real price efficiency. Further, HFT is associated with a temporary increase in the quality of analysts' forecasts until report quality deteriorates again three to four years later. This is because HFT induces investors to buy less information from financial analysts which temporarily raises competition inducing them to perform better. This effect subsides when the market for information adapts to the external shock from HFT. My empirical results also hold for alternative definitions of HFT, alternative methods to compute standard errors and pass further robustness checks. They are consistent among one another and with theoretical arguments.

All in all, my results confirm the arguments of Stiglitz 2014. HFT reduces the gains from information for institutional investors by front running. Hence, institutional investors acquire less information. Consequently, market prices reflect economic value less which distorts the basis for real resource allocation. This is an effect which decreases total welfare. However, to evaluate the overall impact of HFT on welfare, we need to quantify HFT's full range of welfare effects. While a decrease in price efficiency decreases welfare, lower trading costs and higher liquidity create direct welfare gains. Thus, all in all, it is not obvious which effect dominates.

6 Figures

Figure 5.1: Illustration of the identification strategy



This figure illustrates the identification strategy of this paper. The sample includes observations from global stock exchanges. Gray shades indicate HFT presence. The graphic shows the staggered start of HFT starting in North-America, then reaching Europe, other developed countries and finally the emerging market India, but not China and South Korea, which I include as counterfactuals in my analysis. This makes it very unlikely that simultaneous but unrelated events drive my results. It also allows me to include year and panel variable fixed effects. Year fixed effects model a flexible time trend. The panel variable is a stock exchange ID when I use price efficiency and a firm ID when I use analyst error or dispersion as dependent variable. After controlling for the transition from floor to electronic trading, firm or market level variables and clustering standard errors, I robustly identify the impact of HFT.

7 Tables

Table 5.1: Descriptive statistics

	S&P 500			All Firms		
	Mean	Median	sd	Mean	Median	sd
Market Value	27.398	10.501	54.736	4.900	354	41.460
Total Assets	14.816	5.255	32.082	18.824	208	871.960
EBIT	1.462	540	3.445	1.064	9	60.967
R&D	300	2	1.000	76	0	7.845
CAPX	814	235	2.086	666	4	41.942
ln(MV/Total Assets)	0,67	0,58	0,54	0,39	0,26	0,66
EBIT/Total Assets	0,11	0,11	0,10	0,02	0,05	0,21
R&D/Total Assets	0,03	0,00	0,06	0,02	0,00	0,07
CAPX/Total Assets	0,07	0,05	0,06	0,06	0,03	0,08
Price Efficiency 1	0,14	0,16	0,11	0,07	0,07	0,11
Price Efficiency 3	0,12	0,11	0,12	0,05	0,03	0,13
Price Efficiency 5	0,13	0,15	0,12	0,06	0,04	0,12
Analyst Dispersion	0,14	0,05	0,57	8,32	0,26	36,57
Analyst Error	0,00	0,00	0,01	0,23	0,00	1,14
Observations		10.715			364.215	

I use annual observations from 13 stock exchanges between 1990 and 2014. Data are from Compustat, CRSP, IBES, the World Bank and the Federal Reserve System. Accounting measures are real and in million USD. As in Bai, Philippon, and Savov 2016, I measure stock market value at the next end of march after the close of the firm's fiscal year and drop financial firms from my sample. It then comprises 28.270 firms. Ratios and Analyst measures are winsorized at the 1% and price efficiency at the 2,5% level. Section three explains this choice and gives variable definitions.

	(1) Price Efficiency 1	(2) Price Efficiency 2	(3) Price Efficiency 3	(4) Price Efficiency 4	(5) Price Efficiency 5
HFT	-0.059**	-0.060**	-0.065**	-0.064**	-0.064*
	(-2.57)	(-2.17)	(-2.17)	(-2.30)	(-2.09)
Electronic	0.011	0.038	0.0069	-0.016	-0.042
	(0.41)	(1.07)	(0.14)	(-0.49)	(-1.30)
Market Size	0.011***	0.0069	0.013**	0.011**	0.011*
	(3.05)	(1.18)	(2.41)	(2.43)	(1.93)
Average Cap	-8.94**	-3.52	-7.21	-6.81	-8.16
	(-2.20)	(-0.61)	(-1.08)	(-1.25)	(-1.45)
Constant	0.14***	0.037**	0.014	0.11***	0.11***
	(10.08)	(2.19)	(0.60)	(5.49)	(7.83)
Year FE	yes	yes	yes	yes	yes
Exchange FE	yes	yes	yes	yes	yes
Adjusted R2	0.47	0.46	0.39	0.38	0.40
N	274	261	248	236	224

Table 5.2: Price efficiency: main DiD model

Data and variables are as described in section three. This table presents the results of my main multi-event DiD model. I regress price efficiency with horizon k on HFT, a set of control variables and year and stock exchange fixed effects:

$$\text{Eff}^k_{e,t} = \beta_0 + \beta_1 \text{HFT}_{e,t} + \delta' X_{e,t} + \eta_t + \eta_e + \varepsilon_{e,t}$$

I cluster standard errors by year. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 5.3: Analyst variables: main DiD model and market level triple differences cross-sectional analysis.

	(1) Analyst Error	(2) Analyst Dispersion
HFT	-0.13***	-3.18***
	(-5.92)	(-4.42)
ln(MV)	-0.13***	-3.22***
	(-11.99)	(-11.11)
EBIT vola	-0.17	-8.33
	(-1.21)	(-1.13)
Tobin's Q	-0.014***	-0.68***
	(-3.95)	(-5.81)
Electronic	0.21***	6.45***
	(9.31)	(8.37)
Constant	0.78***	22.4***
	(13.62)	(13.11)
Year FE	yes	yes
Firm FE	yes	yes
Adjusted R2	0.55	0.63
N	77,907	61,887

I use annual firm level observations from 13 stock exchanges between 1990 and 2014. All values are real and in USD when applicable. The five year rolling volatility of EBIT is in millions. IBES stands for analyst dispersion and analyst error. The first two columns present the results of my main multi-event DiD model. I regress the analyst based measures on HFT, a set of control variables and year and firm fixed effects: $\text{IBES}_{i,t} = \beta_0 + \beta_1 \text{HFT}_{i,t} + \delta' X_{i,t} + \eta_t + \eta_i + \varepsilon_{i,t}$. Columns three and four present the results of my firm level cross-sectional analysis of the analyst based measures. I regress error and dispersion on HFT, the interaction of HFT with logarithmic market value, ln(MV), a set of control variables and year and firm fixed effects: $\text{IBES}_{i,t} = \beta_0 + \beta_1 \text{HFT}_{i,t} + \beta_2 \text{HFT}_{i,t} \times ln(MV)_{i,t} + \beta_3 ln(MV)_{i,t} + \delta' X_{i,t} + \eta_t + \eta_i + \varepsilon_{i,t}$. All variables are defined as in section three. I cluster standard errors by firm. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Analyst Error	(2) Analyst Dispersion	(3) Analyst Error	(4) Analyst Dispersion	(5) Analyst Error	(6) Analyst Dispersion
Placebo +2y	-0.026	0.66				
	(-1.24)	(1.00)				
Placebo +3y			0.020	1.88^{***}		
			(1.10)	(3.30)		
Placebo +4y					0.031^{*}	1.61^{***}
					(1.88)	(3.11)
EBIT vola	0.028	1.02	0.0077	0.066	0.0041	0.36
	(0.10)	(0.10)	(0.03)	(0.01)	(0.01)	(0.04)
ln(MV)	-0.13^{***}	-2.56^{***}	-0.13^{***}	-2.52^{***}	-0.13^{***}	-2.51^{***}
	(-15.20)	(-9.52)	(-15.07)	(-9.38)	(-15.00)	(-9.34)
Tobin's Q	-0.0011	-0.23^{**}	-0.0014	-0.24^{**}	-0.0013	-0.23^{**}
	(-0.28)	(-1.98)	(-0.37)	(-2.06)	(-0.35)	(-2.00)
Electronic	0.023	-0.16	-0.0012	-1.14	-0.0069	-0.85
	(1.02)	(-0.23)	(-0.05)	(-1.56)	(-0.30)	(-1.22)
Constant	0.98^{***}	22.2^{***}	0.98^{***}	22.2^{***}	0.98^{***}	22.2^{***}
	(18.01)	(12.49)	(18.01)	(12.49)	(18.02)	(12.48)
Year FE	yes	yes	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes	yes	yes
Adjusted R2	0.63	0.68	0.63	0.68	0.63	0.68
N	38308	31229	38308	31229	38308	31229

Table 5.4: Analyst variable DiD: HFT present and shifted

Variables, data, t statistics and significance stars are as in table 5.3. This table presents the results of the main multi-event DiD model when I shift HFT by x years and drop observations with HFT $= 0$. I cluster standard errors by firm.

$$\text{IBES}_{i,t} = \beta_0 + \beta_1(\text{Placebo +xy}_{i,t}) + \delta' X_{i,t} + \eta_t + \eta_i + \varepsilon_{i,t}$$

	(1) Price Efficiency 1	(2) Price Efficiency 2	(3) Price Efficiency 3	(4) Price Efficiency 4	(5) Price Efficiency 5
(1) HFT	-0.059***	-0.060**	-0.065**	-0.064**	-0.064**
	(-2.65)	(-2.18)	(-2.12)	(-2.05)	(-1.99)
(2) HFT	-0.12***	-0.12***	-0.12***	-0.11***	-0.12***
	(-6.07)	(-3.71)	(-4.88)	(-3.84)	(-3.20)
(3) HFT	-0.099***	-0.100***	-0.11***	-0.11***	-0.13***
	(-4.40)	(-3.08)	(-3.75)	(-3.75)	(-3.37)
(4) HFT	-0.099***	-0.100***	-0.11***	-0.11***	-0.13**
	(-3.12)	(-2.77)	(-3.13)	(-3.13)	(-2.54)

Table 5.5: Price efficiency: robustness checks

This table presents the coefficients of HFT when I estimate four variations of the main empirical model as robustness checks. The four models differ with respect to the inclusion of fixed effects and clustering of standard errors. I use annual stock exchange level observations from 13 stock exchanges between 1990 and 2014. All values are real and in USD when applicable. Market size and average capitalization are in trillion USD. I regress price efficiency with horizon k on HFT, a set of control variables and include year and stock exchange fixed effects in some of the models:

$$\mathrm{Eff}_{e,t}^{k} = \beta_0 + \beta_1 \mathrm{HFT}_{e,t} + \delta' X_{e,t} [+\eta_t + \eta_e] + \varepsilon_{e,t}$$

All variables are defined as in section three. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The four models are as follows:

(1) Year FE, firm FE, normal OLS standard errors
(2) No FE, standard errors clustered by year and by stock exchange
(3) Year FE, standard errors clustered by stock exchange
(4) Year FE, standard errors clustered by stock exchange with wild bootstrap t-technique

References

Aitken, Michael, Douglas Cumming, and Feng Zhan (2015). "High frequency trading and end-of-day price dislocation". In: *Journal of Banking & Finance* 59.

Aldridge, Irene (2013). *High-frequency trading: A practical guide to algorithmic strategies and trading systems*. 2nd ed. Wiley trading series. Hoboken, N.J: Wiley. ISBN: 9781118343500. URL: `http://lib.myilibrary.com/detail.asp?id=477746`.

Amihud, Yakov (2002). "Illiquidity and stock returns: Cross-section and time-series effects". In: *Journal of Financial Markets* 5.1, pp. 31–56.

Arnuk, Saleem and Joseph Saluzzi (2012). *Broken markets: How high frequency trading and predatory practices on Wall Street are destroying investor confidence and your portfolio*. Upper Saddle River, N.J: FT Press. ISBN: 9780132875240. URL: `http://proquest.tech.safaribooksonline.de/9780132875288`.

Bai, Jennie, Thomas Philippon, and Alexi Savov (2016). "Have Financial Markets Become More Informative?" In: *Journal of Financial Economics* forthcoming.

Baldauf, Markus and Joshua Mollner (2015). *High-Frequency Trade and Market Performance. Working Paper*.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates". In: *Quarterly Journal of Economics* 119.1, pp. 249–275. ISSN: 00335533. URL: `http://qje.oxfordjournals.org/content/by/year`.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu (2014). *International Evidence on Algorithmic Trading. Working Paper*.

Brogaard, Jonathan, Björn Hagströmer, and Ryan Riordan (2015). "Trading Fast and Slow: Colocation and Liquidity". In: *Review of Financial Studies* 28.12, pp. 3407–3443.

Brogaard, Jonathan, T. Hendershott, and Ryan Riordan (2014). "High frequency trading and price discovery". In: *Review of Financial Studies* 27.8, pp. 2267–2306.

— (2015). "High Frequency Trading and the 2008 Short Sale Ban". In: *Working Paper*.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors: 90 / 3". In: *Review of Economics and Statistics*, pp. 414–427. ISSN: 00346535. URL: `http://www.mitpressjournals.org/loi/rest`.

— (2011). "Robust inference with multiway clustering". In: *Journal of Business & Economic Statistics* 29.2, pp. 238–249.

Cameron, A. Colin and Douglas L. Miller (2015). "A practitioner's guide to cluster-robust inference". In: *Journal of Human Resources* 50.2, pp. 317–372.

Christensen, Hans B., Luzi Hail, and Christian Leuz (2015). "Capital-market effects of securities regulation: The role of implementation and enforcement". In: *Working Paper*.

Conrad, Jennifer, Sunil Wahal, and Jin Xiang (2015). "High-frequency quoting, trading, and the efficiency of prices". In: *Journal of Financial Economics* 116.2, pp. 271–291.

Dugast, Jérôme and Thierry Foucault (2016). "Data Abundance and Asset Price Informativeness". In: *HEC Paris Research Paper No. FIN-2014-1036*.

Gao, Feng, Zoran Ivković, and Sophia Zhengzi Li (2014). "Trading Frequency and Information Efficiency: Theory and Evidence from US and Chinese Markets". In: *Working Paper*.

Guo, Ming, Zhan Li, and Zhiyong Tu (2012). "A unique "T + 1 trading rule" in China: Theory and evidence". In: *Journal of Banking & Finance* 36.2, pp. 575–583.

Jarrow, Robert A. and Philip Protter (2012). "A Dysfunctional Role of High Frequency Trading in Electronic Markets". In: *International Journal of Theoretical and Applied Finance* 15.3, pp. 1–15. ISSN: 02190249. URL: `http://www.worldscientific.com/loi/ijtaf`.

Jones, Charles M. (2013). *What do we know about high-frequency trading?. Working Paper*.

Kang, Qiang and Qiao Liu (2008). "Stock trading, information production, and executive incentives". In: *Journal of Corporate Finance* 14.4, pp. 484–498.

Petersen, Mitchell A. (2009). "Estimating standard errors in finance panel data sets: Comparing approaches". In: *Review of Financial Studies* 22.1, pp. 435–480.

Stiglitz, Joseph E. (2014). *Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy? Speech at the Federal Reserve Bank of Atlanta 2014 Financial Markets Conference*. Atlanta, Georgia.

Sun, Lei (2011). "How Does Competition Affect Opinion Dispersion?" In: *Working Paper*.

Thompson, Samuel B. (2011). "Simple formulas for standard errors that cluster by both firm and time". In: *Journal of Financial Economics* 99.1, pp. 1–10.

Tong, Lin (2015). "A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors". In: *Working Paper*.

Yang, Liyan and Haoxiang Zhu (2016). *Back-Running: Seeking and Hiding Fundamental Information in Order Flows. Working Paper*. Rotman School of Management Working Paper No. 2583915.

Wars and Innovation: Quantifying Long Term Effects

Timon Lucas Dreyer[1]

1 Introduction

"War is the father of all" is the common reduction of a famous quote by Heraclitus. The quote has become popular to stress the crucial influence of wars on every aspect of society. Even more popular is the notion that wars repeatedly yielded groundbreaking progress in technological fields that revolutionised society beyond wartime. Classic examples are the advancement in the chemical industry during World War I or the developments in rocketry during World War II. The latter ultimately laid the foundation for all satellite-related communication. Undoubtedly, as wartime research focuses heavily on weapons development, there have been inventions of dubious value for humanity: the Manhatten Project brought the problematic ability to extinguish complete cities with nuclear warheads.

Whereas the destruction potential of these nuclear, chemical, or biological weapons is beyond question, the view that other war-related inventions have a significant influence on innovation in times of peace remains unproven. Therefore, it seems reasonable to get beyond mere assertions and find evidence - or consider its absence.

Innovation during wars seem to be attributable to three channels. First, inventions have resulted from massively altered resource allocation (i.e. Research and Development, R&D) by warring countries Ruttan 2006. As an example, the direct costs of the Manhattan Project are calculated as 18.5 billion US-Dollar in prices of 2005. Bakker 2013 estimates the opportunity costs as worth about 90% of 2005 United States Gross Domestic Product (GDP). Information as precise as above is unfortunately hardly available for other historical government financed research projects. Second, other inventions were made for reaons of scarcity. The search for substitutes lead to otherwise undiscovered solutions, more generally pointed out by Hayek 1968. Those two channels represent either the concentration or the lack of resources. Third, wars may be of eminent importance for national development, as they cause institutional innovation. Not only pre-modern

[1]Timon Lucas Dreyer received his degree (B.Sc.) from the University of Bonn in 2015. The present article refers to his bachelor thesis under supervision of Prof. Dr. Moritz Schularick, which was submitted in September 2015.

wars Schofield 2000 but also modern wars may play an important role: the discovery of large-scale state intervention in the economy can be named as well as the collapse of tyrannical or oligarchic political systems.

In this study, I examine the hypothesis of war-induced extraordinary technicological developments in a quantitative way. To assess effects on innovation, patent counts are employed as the dependent variable approximating technological development. I construct a dataset and conduct a panel fixed effects analysis. On the right side of the regression, dummy variables are incorporated to indicate times of war and subsequent postwar periods. Their statistical significance or insignificance is expected to shed light on potential war effects. Further on, these dummy variables are replaced in order to measure conflict intensity: every war is weighted with an approximation for its intensity. Country and time fixed effects as well as relevant control variables for education, wealth, and investment are used to mitigate omitted variable bias.

2 Empirical Study: Innovation and Patents

Before any regression can be conducted to shed light on war effects on innovation, *measuring* innovation in the first place must be discussed in brief. The standard approach for innovation is using patent counts (Griliches, 1990 or Moser, 2013) due to its close relation to the inventive sector as well as availability of data, especially for the time since World War II. Even though the rich sources for *current* data provided by the United States Patent and Trademark Office (USPTO), the National Bureau of Economic Research, but also the EU and Japan allow for an extensive study of citation behaviour[2] a broad analysis for several economies considering the *long run* would obviously be an enormous task, even if the necessary data was available (and it is not). Nearly all states assumed to be responsible for substantial innovative progress have (fortunately) fought only few wars abroad and none on their own territory since World War II Wimmer and Min 2009. Therefore, in order to analyse a potential relation between wars and innovation and for reasons of statistical validity, data must be available for several decades before 1950. For those periods, patent application and granting counts are the best available approximation for innovation.

The concern may arise that a significant part of wartime inventions may not be patented due to their military importance in an ongoing conflict. Then, patents would fail to be a viable measure for innovation, especially for the given task. However, I examine effects on the *after-war* society. Inventions and developments based on important wartime discoveries are likely to be documented by the patent system sooner or later. If they are not documented (e.g. due to prolonged nondisclosure), they are of no benefit for society other than military and not of interest for the topic.

[2]Nagaoka, Motohashi, and Goto 2010 use patent meta data such as forward and backward citations, as well as innovation surveys. Lanjouw and Schankerman 2004 construct an index for the innovative value of an invention consisting of the number of patent claims, forward citations, backward citations, and family size. Again, this is only possible with data which is not available for the bulk of the time periods considered in this study.

The two measures available are the number of patents granted and the number of patent applications. Given civil servants are capable of evaluating applications, granted patents are a better measure in terms of quality, but they are also a function of the patent office's funds. As yet another robustness check, both measures are employed. "Patent counts" hereafter subsumes both granted patents as well as filed patent applications. The patent system established itself during the second half of the 19th century Lerner 2005. Thus, 1870 is used as the starting point for granted patents and 1883 for patent applications, as there is no data on patent applications before 1883. Due to a massive increase in patent applications in recent decades Golden 2006, but most importantly because of the Patent Cooperation Treaty (PCT)[3], effective January 1978, the examined time period is terminated in 1977.

2.1 Econometrical Framework

I assess the effects of wars on innovation using a panel fixed effects approach. The underlying methodology is the so-called Difference-in-Differences approach. The best way to study the effect of a policy change or any shock is an experiment. If the effect of a certain education reform on test scores is to be studied, a sufficiently high number of similar school classes has to be assigned randomly to either the control group or the treatment group in order to avoid systematical differences between them Angrist and Pischke 2014. Now, such experiments are impossible in macroeconomics. The idea of Difference-in-Differences is to *imitate* the environment of an experiment. This is also applicable to the school classes: instead of conducting an experiment, one could choose some classes at the two sides of the border between two German states when one of the states decides to conduct an education reform. Note that the effects could only be studied if those classes *would have* developed equally in absence of a policy change. This is the crucial requirement, more generally called similar trends. Clearly, it cannot be observed whether this requirement holds.

Applied in a more general sense to countries, this method allows studying the effects of macroeconomic events. Obviously, it is not sufficient to just compare patent counts of countries at peace with countries at war. Many other factors influence patent counts and distort the effects of war. Regression enables controlling for these other factors and thus approximate *otherwise* (i.e. in absence of war) similar trends. The war effect is measured with a dummy variable that takes on the value of 1 if a war is fought in the particular country. In a perfectly controlled regression, the coefficient of the dummy would give the difference in patent counts between fighting a war and being at peace *ceteris paribus*.[4] "Would give" as there is no perfectly controlled regression. Accordingly, the coefficient of a postwar dummy gives the difference between being in this very

[3] The PCT facilitates the filing of patents in that an application for a patent in one member state automatically implies an application in every other member state. For the mentioned reason of explosive growth in patent applications roughly since 1980, there is no sense in extending the time horizon much longer. Therefore, the end is set before the new rule went effective to avoid possible distorting effects.

[4] For a refreshing read on the history of counterfactual economic analysis see McAfee 1983.

postwar situation and the counterfactual. If the notion of substantial innovative effects of wars is true, then those postwar dummies should (i) have positive coefficients and (ii) be significantly different from zero.

There are control variables (like GDP) directly related to the dependent variable patent counts. To cover effects differing among states and time not identified by these control variables, a set of country dummies and a set of year dummies is added Baltagi 2008. This *does not* mean that there is one specific dummy for country i in year t (which would result in overfitting), but one dummy for each country i and one dummy for each year t. This setting leads to the following generic equation:

$$Y_{i,t} = \beta_j \text{ conflict}_{i,j,t} + \sum_{k=1}^{20} \gamma_{k,j} \text{ year } k \text{ after conflict}_{i,j,t} + \delta_m X_{i,t} + FE + e_{i,t}$$

where
i relates to the country, j relates to the type of conflict, depending on classification (e.g. outcome, type), t relates to the year,
$Y_{i,t}$ is either patent applications or granted patents for country i in year t,
β_j is the coefficient of war type j,
$\text{conflict}_{i,j,t}$ is a conflict of type j (e.g. won war) ongoing in country i in year t,
$\gamma_{k,j}$ is the coefficient of the k^{th} year after a war of type j, i.e. the k^{th} postwar year,
$X_{i,t}$ is a vector of m control variables related to patent counts, and
FE relates to the country and time fixed effects.

The differentiation among conflicts works as follows. The conflict and postwar variables are dummies that take on the value 1 if the specified event occurs and 0 otherwise. Two examples:
1. No differentiation between conflicts: There is only one dummy variable representing wars. In this case, there is also only one dummy per postwar year. For example, the dummy for the 10th postwar year takes on the value 1 for Germany in 1956.
2. Differentiation between won and lost conflicts: There are two dummy variables for ongoing wars: one for wars that were won and one for wars that were lost. There are also two dummies for every postwar year. The dummy for the first year after a *lost* conflict takes on the value 1 for Austria in 1919, as this is the first year after the end of World War 1. The dummy for an ongoing war that was victorious for the country will take on the value 1 for France in 1917, as France was among the victors of World War 1.
Note that due to estimating a panel fixed effects model there is only one variable for the same type of conflict for all countries, i. e. the goal is to find an overall occurring pattern. As can be seen from the sum, I allowed for an effect until twenty years after the end of a war. In case of another

war during these twenty postwar years, the country is not considered undergoing postwar time anymore: In 1949, the dummy for the fourth postwar of a won conflict is 1 for the United States. However, in 1950, the dummy for the fifth postwar will not take on the value of 1, as the United States waged war in Korea from this year on.

2.2 Data: Wars, and Control Variables

A detailed account of the used datasets and my treatment of data gaps can be found in the appendix. The objection may arise about the definition of wars. The two World Wars stand out as they affected every single economy on the planet. This is only the tip of the iceberg, though. Categorisations mostly fall short of creating comparable types of wars as conflicts differ across time and participating countries in multiple dimensions such as strategies, employed weapons, internationality, intensity, and most of all how the civil society is affected. This hardly measurable diversity must be kept in mind for any inference. In order to reduce this issue at least partially, I use different datasets on wars to differentiate by outcome, by type (mainly international vs. intranational), and/or by intensity. The first dataset for armed conflicts, provided by Wimmer and Min 2009, is used to differentiate between types of wars and their locations. It considers six forms of conflict, two interstate war categories - namely wars of conquest and balance of power wars - and four categories of intranational conflicts (e.g. civil wars or ethnic uprisings). Importantly, this dataset regards wars always from the perspective of the countries that they were fought in.[5] This is helpful as it enables reducing the examination of war effects to the countries that provided the battle ground. If there is an effect on scientists' efforts if the conflict is close by, i.e. if a war poses an existential threat, this approach should be useful to investigate it. The dataset was exploited in two different ways: First, conflicts are only considered for the countries they were fought in. Second, conflicts are considered for every participant, i.e. including all participants even though the war did not take place on every participant's ground. Note that it is possible that an external power takes part in an intrastate conflict, e.g. Russia in Hungary in 1956.

The second dataset employed is the so-called "Correlates of War" dataset from Sarkees and Wayman 2010. It reports the victors and the "battle-related combatant fatalities suffered by the state". This information was used to differentiate between effects of lost versus won wars and for a measure of conflict intensity.

Control variables are a common means to maintain unbiasedness of regressions. Certainly, there is no such thing as a perfectly controlled regression. Some control variables that appear to make sense, as for example corporate R&D, or a property protection index, cannot be employed as available time series do not reach back long enough. Restricted by this logistical constraint, I took the following controls for the econometric model into account. In order to control for the

[5] An example: World War II is at first glance surprisingly defined as a balance of power war for Germany. That is because it had in fact become a balance of power war when combat reached Germany in 1942. For the Netherlands, World War II is listed as a war of conquest, as from the dutch perspective it was nothing else when the Wehrmacht attacked in 1940.

standard of education, log school and high school years are incorporated. The size of a country is an important factor of the amount of filed patents, which is why I employ log population counts and the lagged log GDP. The latter comprises the mean log GDP of years three to six prior to year t, i.e. $\frac{1}{4}\cdot(\text{GDP}_{t-6}+\text{GDP}_{t-5}+\text{GDP}_{t-4}+\text{GDP}_{t-3})$. This regressor not only serves to control for size, but it also covers changes in capital and thus its destruction as GDP is considered to be (amongst others) a function of capital Solow 1957. Growth in the five periods prior to the current is incorporated in order to control for economic shocks. It is assumed that investment in research takes approximately three to five years to materialise. The ratio of country i's per capita income to the per capita income of the world's wealthiest nation (in terms of per capita income) serves as a control for the relative sophistication of the economy Lerner 2002; it is hereafter called *relative GDP*. Finally, lagged log investment *level* is incorporated, comprising the mean real investment in the four years prior to period t (in some estimations, each of these years will be considered separately instead of a mean). Investment may be the single most important control variable as it is an approximation for R&D which is in turn a good measure for the (financial) input into research Hall, Griliches, and Hausman 1986. However, there are several issues with investment as a control variable. First, there is consequently probably measurement error in this variable as I want to exploit it as an indicator for investment *in research*. Obviously, investment in research makes up only a fraction of overall investment. Second, data on investment is not as complete as for the other variables. Many countries have data gaps for investment during and after World War II. Third, it could be objected that investment is a bad control as it is dependent on expected earnings which are in turn dependent on valuable inventions. Then, investment would be a function of the outcome variable. Fourth, one may argue that governmental military research expenditures are not covered by investment. In this case, *if* there is a *positive* effect on patent counts during postwar periods, this may be taken into account as an explanation. I acknowledge that investment is a difficult control and therefore estimate and report any model that incorporates investment also once without it.

2.3 Results

I considered two approximate measures of innovation with patent counts, namely log patents and patents per inhabitant, each time once inserting granted patents and once patent applications.

The dataset by Sarkees and Wayman 2010 enables the differentiation between won and lost conflicts. There is a negative relation between victorious wars and *patent applications* for both measures (namely log patent applicatoins and patent applications per inhabitant) Table 6.2 reports the results for several configurations of postwar effects on patent applications and my "standard" model with patents per inhabitants as the dependent variable. Note that the postwar dummies for single years refer to the time after *won* wars only. Dummies for the era after *lost* wars were also estimated in each regression but not reported as they are insignificant (likely due to an insufficient number of observations). A complete assessment for different sets of control variables is reported in the appendix. Table 6.2 shows that countries fighting a war with positive outcome (i.e. they were victorious) remain on a *lower* level of patent applications *ceteris paribus*. Regression (1) shows that the dummies have similar coefficients in a range from -0.07 to -0.20. However, many of them are yet insignificant. In Regression (2), all twenty postwar years are incorporated in one dummy, i.e. the new dummy is just the sum of all dummies for single years and thus gives the mean of the twenty dummies. The coefficient of the dummy is negative and statistically significant at the 5% level. Thus, the level of patent applications is about 16 per cent lower for countries experiencing a postwar period. Also, regression (3) shows that there are no significant deviations from this mean for the complete postwar era. The fourth regression introduces the lagged log investment level as an additional control variable. This reduces the sample to ten countries. Also, data on investment right after World War II misses for most countries and will be lost in those regressions. It is important to note that the postwar variables for the years ten to twenty are not controlled for any reduced investment during the war: If a country is in postwar year 15 but investment is lagged only for the last six years, then any effect that is induced by lower investment during the war will not be captured anymore. This is why the coefficients of the first postwar years are especially important for this question. The importance of the findings related to investment is discussed below. Regression (5) confronts the objection that taking the log of patent counts (given low absolute numbers in the frist years) may have a distorting effect. Patent applications per inhabitants are employed as the dependent variable. The above discussed negative postwar effects persist, they are even a little bit stronger and exhibit higher significance.

Apart from low significance, there are several issues with these results. If granted patents instead of patent applications are used as the dependent variable, there are *no* significant results at all. There might be some agency-theoretical approach explaining this phenomenon: Patent offices receiving less applications might still deliver a comparable output of granted patents either because their capacities were already too low to process all applications in the first place or out of fear to face a budget cutback if less patents are granted. Alternatively, we might observe here an attempt to absorb war-related declines in patent applications, maybe even politically intended.

Another concern could arise in regard of World War II. Given that all countries in the sample were affected by World War II, positive effects could be absorbed by the year fixed effects. However, those year dummies for the war and postwar time are insignificant. Thus, it can be rejected that positive effects occuring all over for all countries are absorbed by the year dummies. Even more, regressing for World War II *only* in a time window from 1920 to 1970 shows significant negative postwar effects on patents per inhabitants. As this time frame is shorter, more countries exhibit complete time series. Therefore, I also conducted the same estimations with a larger sample. Then, negative postwar coefficients have lower standard errors and the null hypothesis can be rejected for more than half of them. Unfortunately, it is not possible to check the influence of investment with this regression, as data on investment during and right after World War II misses for most countries.[6] However, even without controlling for investment, these results are remarkable, as

[6] It is, however, easily arguable that during World War II military research was the important player, hardly covered by

World War II is the single most important basis for narratives of outstanding inventions. However, we see that both conditions for this notion to be confirmed (positive and significant coefficients) are violated. There is also no evidence that World War II is only a negative outlier: if it is excluded as wartime, the postwar dummies' coefficients remain negative but mostly become statistically insignificant.[7]

As mentioned above, *lost* wars and their aftermath do not exhibit significant negative effects. This is possibly due to too few observations. There are only five observations for lost wars in the underlying sample.[8] The low number of lost wars in the sample is related to the constraint of complete time series, there are far more in the larger dataset. The completeness of data surely is correlated with good statistical agencies that are in turn found in wealthy countries. By chance, all countries in the constrained sample happen to be OECD members today. It is remarkable that those mostly rich countries seem to have endured few defeats. I can only hypothesise that there may be some reverse causality or endogeneity in play here, with countries being richer today if they fought victorious wars, or a higher success rate in waging wars.

Weighting by Conflict Intensity

As discussed above, diversity of wars is a crucial problem in assessing patterns across time and country. Sarkees and Wayman 2010 give the "battle-related combatant fatalities suffered by the state". Estimating these figures is complicated, even more for civilians than for combatants. Thus any measure for conflict intensity will be at least vague, if not biased. However, the number of battle deaths allows for an approximation that should not be left unconsidered.

Information on battle deaths is taken into account in two ways. Instead of employing dummies indicating only whether a country is in, say, the fifth postwar year or not, the information is weighted with the battle deaths of the war in millions. That is, for the United States, the variable for the fifth postwar year takes on the value of 0.405 in 1950, as about 405,000 American soldiers died in World War II. Table 6.3 shows that the negative postwar effects are stronger in terms of significance when conflicts are weighted by the number of dead soldiers. In the second configuration, the weight is not million battle deaths, but battle deaths per inhabitants. Estimation results are similar.

I pointed out above that a new conflict eradicates the postwar era of the last conflict in the data. This results in situations like the post World War I time for Germany, which is promptly overlaid by Germany's minor commitment in the Latvian Liberation. Consequently, the 1920s are counted as a period after a war with 750 casualties instead of the 17.7 million lost in World War I. I confront this issue by estimating a model only considering conflicts with more than 5,000

investment.

[7] Excluding World War II as a war means that the dummies indicating an ongoing conflict remain zero from 1939-1945 and that there is no postwar time indicated after World War II. However, I do not allow the war to fall completely out of the sample: Any country undergoing a postwar phase in the 1930s is not regarded as doing so anymore from 1939 on.

[8] There are 27 *won* wars in the underlying sample with World War II and 21 without. This should be a sufficient amount of observations.

casualties. Regressions (4) and (5) in table 6.3 show the results. A remarkable effect of these intense conflicts on postwar patent application levels is observable.

In Regressions (3) and (5), investment controls are incorporated. Most of the coefficients of the first years after war do not seem to be significantly affected anymore, indicating that a big part of the negative war effect is probably related to a drop in investment.[9] These results suggest that wars drag resources out of the research sector leading to a drop in innovative activity and slower patent growth. The negative postwar dummies purport that there is no such thing as technological catch-up growth afterwards, let alone a positive effect.

However, these results are not robust for patents per inhabitants as the dependent variable. They are neither robust for granted patents (instead of the above utilised patent applications) as the dependent variable in which configuration whatsoever. This could be attributable to the mentioned agency problems. Also, I already pointed out that the employed measure for conflict intensity is somewhat noisy. Thus, attenuation bias likely draws the coefficients towards zero. With coefficients nearer by zero, null hypothesis are less likely to be rejected which results in seemingly less significant coefficients.

The results of using the dataset by Wimmer and Min 2009 shall only be mentioned briefly. Differentiating between inter- and intranational conflicts does not yield much, the remaining countries in the sample fought nearly no internal conflicts. Restricting wars to geographically affected countries also drastically reduces the amount of observations, harming any regression's consistency. Consequently, estimations are uninformative; they also yield only insignificant coefficients. However, as it cannot be determined whether insignificance can be attributed solely to statistical problems with insufficient observation counts, it weakens any interpretation of the above presented results.

3 Conclusion

In this thesis I tried to analyse long term effects of wars on innovation. My results suggest that wars cost innovative output and that this lost output is not compensated. The lower innovative output seems to be partly related to lower investment. Other factors could be drafted research personnel or destroyed infrastructure, both physical and intellectual. Also, the efficiency of concentrating resources into military research is questionable. Theoretically speaking: the marginal benefit of a Dollar on top of a billion Dollars already invested into a military project may be lower than an additional Dollar on top of a million Dollars invested in a medical research project, for example. This holds only if science is evolving linearly: If it does, the atomic bomb would have been developed sooner or later. The contrary would be the claim that only the concentrated efforts of the large group of important physicists brought together under the roof of the Manhattan Project

[9] More regressions paying attention to this question while employing different measures of investment can be found in the original thesis.

enabled building the bomb.

This conclusion comes with the following concerns and question marks:
Quantifying innovation appears to be the biggest challenge. The use of patent counts certainly introduces measurement error raising standard errors. In consequence, hypothesis tests are less likely to be rejected and coefficients are more likely to be assessed insignificant. Thus, if it was certain that patent counts are a sensible approximation for innovation, this would turn out in favour of the coefficients' significance. But are we? There are some patents as valuable in terms of inventiveness as thousands of other patents. This may be smoothed by this valuable patent entailing a wide range of patents building on it. That is, we cannot measure the quality of a single patent with patent counts, but the positive effect of this valuable patent on future patent applications is again reflected by the patent count measure. However, it is remarkable that the somewhat robust results for patent applications are not replicated with granted patents. This could be due to patent offices trying to avoid budget cutbacks by granting a comparable amount of patents, capacity constraints so that the output is smoothed, red tape, or the politically intended attempt to absorb any war-related decline in patent applications. To resolve the issue of potentially mismeasured innovation, an application of a method similar to the above mentioned of Lanjouw and Schankerman 2004, that takes into account patent meta data, would deliver much more resilient results. However, the implementation of such a technique with patent documents from the beginning of the century would be very resource-intensive, that is, meta data would need to be read directly from original patents.

There is also the distortion of national patent counts and technological development induced by confiscated inventions or emigration of scientists. German scientists that were commandeered mostly by the Soviet Union and the United States after World War II took valuable rocketry along, crucial for the Space Race. It is also arguable that the Cold War following World War II is comparable to taking part in a conflict abroad, especially with respect to military expenses poured into research.

Another concern is whether the heterogeneity of wars can be adequately considered. Results are robust if conflicts are distinguished by the number of battle deaths per participant. I see an important benefit in this differentiation between minor conflicts abroad and severe wars. A more precise distinction always comes at the cost of lowering observations for each type of war resulting in uselessness for a *quantitative* approach.

A more general question is the identification of the best set of control variables. Clarke 2005 argues that it is *per se* impossible to find out whether an additional relevant control variable reduces or increases the bias. Additionally, this thesis is challenged with a reduced pool of applicable control variables due to the long time horizon. Data on government expenses for military research projects in the beginning of the century would be of exceptional value.

All these challenges in mind, there is *no* quantitative evidence for a *positive* effect of wars on technological progress.

4 Data Description and Tables

Table 6.1: Data Sources

Data	Source	Notes
Granted Patents I	Federico 1964	
Granted Patents II	WIPO 2011	World Intellectual Property Organization
Patent Applications	WIPO 2011	
GDP & Population	Maddison 2010	
Real Investment	Schularick and Taylor 2012	Missing for Belgium and Portugal
Education	Morrisson and Murtin 2009	Mean School and High School Years; Identical Values for Scandinavian Countries and for Benelux + Switzerland
Wars I	Sarkees and Wayman 2010	War Outcome and Fatalities
Wars II	Wimmer and Min 2009	War Type and Location

4.1 Data Gaps

I did enquire into problems with missing data. There are no alarming patterns among missing data points; especially, there are no problems in that data would lack more often for postwar periods. However, for a more detailed account see my complete thesis. Still, I have to point out that some countries exhibit a substantial lack of data. There seem to be three possibilities in coping with missing data: impute, interpolate, or ignore. Every single one has its particular advantages and disadvantages. Imputation seems to be the most sophisticated way of dealing with lacking data Honaker and King 2010, but it takes strong assumptions and, importantly, any interpretation of results in the final analysis ultimately relies on the applied imputation method in the first place Albouy 2012. Interpolation, nothing else than a simple form of imputation, may be helpful at least for the two education variables that both exhibit low variation. Average school and high school years are reported per decade, these figures change rather slowly and will most certainly not endure drastic shocks. Recall that the numbers refer to the whole population, not only the current youth. However, for the crucial patent counts, I decided not to impute but to apply listwise deletion, i.e. only regress on complete observations. This leaves subsistent gaps for several countries. As a compromise between retaining enough countries and minimising data gaps, all countries that lack more than seven observations in patent counts are left out. Further lowering the threshold would leave too few countries. The remaining dataset contains twelve countries.

Table 6.2 is discussed in length in the text. It covers the main regressions for distinguishing conflicts by outcome. The same applies to Table 6.3, where wars are weighted with the number of battle deaths. The underlying definition of war is the one from Sarkees and Wayman 2010 who differentiate between won and lost wars. This is also the definition of war focused on in the main section.

Table 6.2: Effects of Victorious Wars on Postwar Patent Applications. This table is not reproduced here. It can be found online at `bje.uni-bonn.de`

Table 6.3: Conflict Intensity, Dependent Variable: Patent Applications. The table is not reproduced here. It is available online at `bje.uni-bonn.de`

References

Albouy, David Y (2012). "The Colonial Origins of Comparative development: an empirical investigation: comment". In: *The American Economic Review*, pp. 3059–3076.

Angrist, Joshua D and Jörn-Steffen Pischke (2014). *Mastering'metrics: The Path from Cause to Effect*. Princeton University Press.

Bakker, Gerben (2013). "Money for nothing: How firms have financed R&D-projects since the Industrial Revolution". In: *Research policy* 42.10, pp. 1793–1814.

Baltagi, Badi (2008). *Econometric analysis of panel data*. Vol. 1. John Wiley & Sons.

Clarke, Kevin A (2005). "The phantom menace: Omitted variable bias in econometric research". In: *Conflict Management and Peace Science* 22.4, pp. 341–352.

Federico, PJ (1964). "Historical patent statistics". In: *Journal of the Patent Office Society* 76, pp. 89–171.

Golden, John M (2006). "Patent trolls and patent remedies". In: *Tex. L. Rev.* 85, p. 2111.

Griliches, Zvi (1990). "Patent Statistics as Economic Indicators: A Survey". In: *Journal of Economic Literature*, pp. 1661–1707.

Hall, Bronwyn H, Zvi Griliches, and Jerry A Hausman (1986). "Patents and R&D: Is there a lag?" In: *International Economic Review* 27.2, pp. 265–283.

Hayek, Friedrich August von (1968). *Der Wettbewerb als Entdeckungsverfahren*. Kieler Vorträge, Neue Folge 56, Hrsg. E. Schneider, Kiel.

Honaker, James and Gary King (2010). "What to do about missing values in time-series cross-section data". In: *American Journal of Political Science* 54.2, pp. 561–581.

Lanjouw, Jean O and Mark Schankerman (2004). "Patent quality and research productivity: Measuring innovation with multiple indicators*". In: *The Economic Journal* 114.495, pp. 441–465.

Lerner, Josh (2002). *Patent protection and innovation over 150 years*. Tech. rep. National Bureau of Economic Research.

— (2005). "150 years of patent office practice". In: *American Law and Economics Review* 7.1, pp. 112–143.

Maddison, Angus (2010). "Statistics on world population, GDP and per capita GDP, 1–2008 AD". In: *Historical Statistics*.

McAfee, R Preston (1983). “American economic growth and the voyage of Columbus”. In: *The American Economic Review* 73.4, pp. 735–740.

Morrisson, Christian and Fabrice Murtin (2009). “The century of education”. In: *Journal of Human Capital* 3.1, pp. 1–42.

Moser, Petra (2013). “Patents and innovation: evidence from economic history”. In: *The Journal of Economic Perspectives*, pp. 23–44.

Nagaoka, Sadao, Kazuyuki Motohashi, and Akira Goto (2010). “Patent statistics as an innovation indicator”. In: *Handbook of the Economics of Innovation* 2, pp. 1083–1127.

Ruttan, Vernon W (2006). “Is war necessary for economic growth?” In: *Historically speaking* 7.6, pp. 17–19.

Sarkees, Meredith Reid and Frank Wayman (2010). *Resort to War: 1816–2007.*

Schofield, Norman (2000). “Institutional innovation, contingency and war: A review”. In: *Social Choice and Welfare* 17.3, pp. 463–479.

Schularick, Moritz and Alan M Taylor (2012). “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008”. In: *The American Economic Review* 102.2, p. 1029.

Solow, Robert M (1957). “Technical change and the aggregate production function”. In: *The review of Economics and Statistics*, pp. 312–320.

Wimmer, Andreas and Brian Min (2009). “The location and purpose of wars around the world: a new global dataset, 1816–2001”. In: *International Interactions* 35.4, pp. 390–417.

WIPO (2011). *World Intellectual Property Organization; Patent applications and grants by patent office, broken down by resident and non-resident (1883-2010).*