ISSN 2328-7144
DOI:10.17265/2328-7144

From Knowledge to Wisdom

Economics World

Volume 6, Number 6, Nov.-Dec. 2018

David Publishing Company
www.davidpublisher.com

Economics World

Volume 6, Number 6, Nov.-Dec. 2018 (Serial Number 31)

David Publishing Company
www.davidpublisher.com

Publication Information:
Economics World is published monthly in hard copy (ISSN 2328-7144) and online by David Publishing Company located at 616 Corporate Way, Suite 2-4876, Valley Cottage, NY 10989, USA.

Aims and Scope:
Economics World, a monthly professional academic journal, covers all sorts of researches on Economics Research, Macroeconomics, Microeconomics, International Economics, Labor Economics, Econometrics, Development Economics, Industry Economic, Regional Economic, Social Economic, Political Economy, Environmental Economics, Tourism Economics, Finance, Banking, Insurance, and other latest findings and achievements from experts and scholars all over the world.

Editorial Board Members:
Michael Ba Banutu-Gomez (USA), Iordanis Petsas (USA), Eduard Babulak (USA), Yutaka Kurihara (Japan), Tomoko Kishi (Japan), Ignacio Falgueras-Sorauren (Spain), Ralph Sonntag (Germany), Marcelo Miranda de Melo (Brazil), Graziella Benedetto (Italy), Ubaldo Comite (Italy), Celso José Costa Junior (Brasil), Krzysztof Piasecki (Poland), Domenico Marino (Italy), Karima Aly Korayem (Egypt), Juliana Leite (Brazil), Bonny TU (China)

Manuscripts and correspondence are invited for publication. You can submit your papers via Web Submission, or E-mail to economicsworld@davidpublishing.org, economics.world@hotmail.com. Submission guidelines and Web Submission system are available at http://www.davidpublisher.com.

Editorial Office:
616 Corporate Way, Suite 2-4876, Valley Cottage, NY 10989, USA
E-mail: economicsworld@davidpublishing.org; economics.world@hotmail.com

Copyright©2018 by David Publishing Company and individual contributors. All rights reserved. David Publishing Company holds the exclusive copyright of all the contents of this journal. In accordance with the international convention, no part of this journal may be reproduced or transmitted by any media or publishing organs (including various websites) without the written permission of the copyright holder. Otherwise, any conduct would be considered as the violation of the copyright. The contents of this journal are available for any citation, however, all the citations should be clearly indicated with the title of this journal, serial number and the name of the author.

Abstracted / Indexed in:
EBSCO, USA
Ulrich, USA
Cabell, USA
Summon Serials Solutions, USA
ProQuest
Google Scholar
Index Copernicus, Poland
Qualis/Capes Index, Brazil
GetCITED, Canada
CiteFactor, USA
Polish Scholarly Bibliography (PBN), Poland
Open Academic Journals Index (OAJI), Russia
NewJour, USA
InnoSpace, USA
Publicon Science Index, USA
Scholarsteer, USA
Turkish Education Index, Turkey
Universal Impact Factor, USA
WorldCat, USA

Subscription Information:
Price (per year): Print $480
David Publishing Company, 616 Corporate Way, Suite 2-4876, Valley Cottage, NY 10989, USA
Tel: +1-323-984-7526, 323-410-1082 Fax: +1-323-984-7374, 323-908-0457
E-mail: order@davidpublishing.com
Digital Cooperative Company: www.bookan.com.cn

David Publishing Company
www.davidpublisher.com

Economics World

Volume 6, Number 6, Nov.-Dec. 2018 (Serial Number 31)

Contents

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 425-432
doi: 10.17265/2328-7144/2018.06.001

Is the Grass Greener on the Other Side? Work-Related Migration From Western-Hungary to Austria

Előd Kovács
Széchenyi István University, Győr, Hungary

After the 2004 EU accession of Hungary, the evolvement of the cross-border regions has started to emerge there as well, followed by the decrease—and in case of the Schengen Area the complete break off—of the separation role of the borders as a result of which the citizens of the member states have free movement in the greater part of the European Union. The basic goal of the EU enlargement was the establishment of the free movement of capital, goods, information, and workers. A very strong workforce migration has immediately started and is still ongoing among the EU member states, directing in first row from the less developed economies to the better developed ones. This of course is the same in case of Hungary, where the active population in a rapidly growing number makes the decision to take a job in the well-developed countries of the EU for a shorter or longer period. According to Rechnitzer, people living in the Austrian-Hungarian border region (among whom many were Hungarian citizens) were capable of novation and were active performers in general already before the transition period. The study focuses on the key motivation factors and the possible grouping of the people living in Western-Hungary having decided to work in Austria.

Keywords: Austria, labour market, migrants, key motivational factors

The EU Faces New Challenges

In the great majority of the European nation-states, the aging of societies is typical; the ratio of elder generations within the total population is getting larger and larger. In the next couple of decades, this trend is expected to keep on going, which challenges the nation-states. Current population forecasts count with advanced labour force, population aging, and a drop in the number of those belonging to the active population (Van der Gaag & Van der Erf, 2008; EU, 2012). The forecasted aging of the population has a serious effect on not only the outlook of the stability of the pension system, but also endangers the health and social care systems. It also influences the labor market and the living conditions of the people (Kotowska, 2006; Prskawetz et al., 2008). In the last few years, the EU had to face new challenges unexpectedly. The first one is the mass-migration of refugees heading towards the EU from outside its borders trying to escape from war; but also the number of those economic migrants targeting the well-developed EU member states coming from the poor countries (mainly in Africa) has been multiplied. Politicians of the individual EU countries see these issues quite differently. The anti-migration politics builds on the justified and causeless fears as well and supports the

Előd Kovács, Ph.D. candidate, Doctoral School of Regional and Economic Sciences, Széchenyi István University, Győr, Hungary.
Correspondence concerning this article should be addressed to Előd Kovács, EQQ Equities Ltd., H-9789 Sé, Mátyás király u. 4, Hungary.

reclusion. In the politics of many member states, the “swing to the rigt” can be observed. While at the same time, the common report of the OECD & EU (2014) attracts the attention to the utilization of the potential hiding in the workforce migration. Based on the findings of the analysis, being on the way out of the hard times caused by the 2008 world economic crisis, the EU could stand on much more stabile feet if it would exploit the skills and knowledge of the migrant workforce.

The Situation of Hungary as COO

The type of system settled in Hungary in the times of socialism has left a very diverse labour market. The transition period has brought mass unemployment which the state tried to ease with the implementation of the unemployment benefit and the restructuring of the retirement system. With the slow stabilization of the economy sometimes stopped by the smaller or bigger economic crisis, followed by the chance of getting employed abroad arising with the EU membership and later with the expansion of public work program, the Hungarian labour market has undergone drastical changes. Hungary is very much a sender country in terms of work related migration. It still has a dual economy, and the public work program indeed has mainly a social goal and focus, debiting the budget of the state instead of being profitable (Barta & Kovács, 2017). The aging of the population in Hungary is a critical issue. In the last couple of years, a new situation has emerged on the Hungarian labor market: the torturing shortage of labour.

The wages in the public sector kept on a relatively low level for over many years; the spread of minimum wage workplaces and the starvation wages of the public work program all effect on the wage levels of the private sector. The Hungarian wage levels do not only lag behind the EU average, but also tail off that of the neighbouring countries. Due to the fact that the wages are much lower than in the neighbouring countries, Hungarians decide to (and actually do) work abroad in massive quantities. The processes of the Hungarian labour market were of course deeply influenced by the complete opening of the Austrian labor market in 2011, which has resulted in the work related migration of many active members of the younger generation to Austria (Barta & Kovács, 2017); in 2012, the most people migrating there come already from Hungary (ÖIF, 2014). Out of the estimated 500,000 Hungarians working abroad (World Bank, 2011), 94,550 Hungarians are legally working in Austria (Hauptverband der Österreichischen Sozialverischerungsträger, 2018), which—together with the growing mobility and willingness of Hungarians to work or even to migrate abroad in general—has resulted in the lack of workforce in Hungary. The lack of workforce can be felt the most in Western-Hungary, which is caused by the fact that this region is the border region of Austria.

The Situation of Austria as Target Country

Austria is a successful member of the European Union, with dynamic development figures. The strength and importance of the country are proven on different fields of economy. Its population is growing year by year and due to the continuously rising number of migrants—who actually are willing to take lower valued jobs as well—coming to the country, Austria is less endangered by the lack of workforce in some specific endangered types of professions, and also by the aging itself. Based on the latest statistics of the Austrian Integration Fund (ÖIF, 2018), Austria is one of the key migration-aimed target countries within the European Union: Last year, 154,700 people immigrated into Austria and only 110,119 people emigrated. The forecasts show that the due to the increasing migration—counted with an estimated yearly net increase with about 32,000 persons—and due to the increase in life expectancy, the current population of Austria is supposed to grow from 8.6 to 9.4 million

people by 2050 (Dialog, 2011). But the actual statistics show that the yearly net increase is actually much higher: In 2017, for example, it was 64,676 (ÖIF, 2017). Austria has its very special pull effect concerning work-related migration in Central- and Eastern-Europe with its low unemployment rate, financial stability, and social welfare system. According to the newest Austrian official statistics in 2017, more than one fifth of the total population (1.97 million people) had migration background (ÖIF, 2018).

Theoretical Background

Many Hungarian and Austrian researchers have already investigated and analyzed the relations between Hungary and Austria, the advantages of the border region, and the different types of co-operation and their potential outcome (Csapó, 1999; Rechnitzer, 2005; Hárs, 2002; 2009; Hardi, 2009). The main focus earlier laid on the assumption of the number of emigrants (Kapitány & Rohr, 2013) and the characteristics of the migration activities (Blaskó, 2014); and based on the available data, the analyzes of those being affected by migration for not longer than a year (Blaskó, Ligeti, & Sik, 2014; Hárs, 2012; Bodnár & Szabó, 2014). The researches partly touched the field of grouping the migrant workers based on their actual intentions and the reasons of migration, but were not directly focusing on this topic. The findings of Hardi (2005) showed as key reason for work related migration from Hungary to Austria the much higher wages, while the chance to achieve the better social benefits was almost not even present. Another finding carried out by Austrian researchers has also stated that for Hungarian employees working in Burgenland province of Austria, the much higher earning is the explicit motivational factor (Wasserbacher & Michenthaler, 2007).

Upon analyzing the complete period of the 1990s, Cseres-Gergely (2003; 2005) had found that unemployment and the differences in wages play a role in the migration decision, but only with less importance. According to Kertesi and Köllő (1997), the key reason for the migration is the pull effect of the well-prospering economic region. According to Juhász (2006), besides the push factors of the country of origin the political and the family reasons also appear. The study of Kaczmarczyk and Okólski (2005) declared that in Central- and Eastern- Europe, the current migration behavior is caused by the interaction of three different imbalances: demographical, economical, and political.

The Empirical Research

In order to get an up-to-date picture about the current situation concerning the work related migration from Western-Hungary to the bordering Austria, the author carried out a quantitative research using an online questionnaire with the help of Google Forms. With the help of a Facebook group (called "Ausztriai Álláslehetőségek-Friss") carefully built up as part of a business service, the Facebook group was followed by 105,000 people, most of whom are Hungarian citizens, either working or living in Austria already, or being already on the search for a job there. All followers of the Facebook group living in Western-Hungary were asked to participate in the research by filling out the questionnaire between the 18th and the 25th of January 2018 and giving the adequate answers on the questions. After the given time was up, the online access to the questionnaire has been blocked. The questionnaires have been processed by the author, the data-coding, their cleaning, and the follow-up work has been done with the help of the SPSS 20.0 statistical programme.

The final number of sample elements has been 611, as four elements have not been adequate to the first and most important criteria of the research; the geographical impoundment, as the respondents were not living in one of the three counties of the Western-Hungarian regions. Before the initiation of the research, a few key

questions were formulated in advance with the intention to see whether there are changes in the motivating factors for the work related migration from Hungary to Austria, than set in the researches and literature introduced earlier in this study:

Q1: Is salary still the most important for Hungarians upon making the decision to work in Austria?

Q2: Is the social benefit related factor also strong?

Q3: Can other important factors be determined?

Q4: Is it possible to group people based on the factors?

In the following chapters, the author intends to introduce the answers on the questions above presenting the results of the questionnaire.

Description of the Population of the Responders

The gender of the responders was almost in an equivalent relation, as out of the complete sample (N = 611), we had 310 women and 301 men. The narrow difference in the gender of the responders rises our attention to the fact that nowadays it is already not only the privilege of men to seek work abroad (as it was the case earlier in CEE), but women also appear as independent and dominant persons on the labour market. The respondents (N = 611) are mainly aged between 25 and 39 years old; the members of which age group are unfortunately the most active, innovational, and hard-working ones. The great majority (60%) of the respondents (N = 611) do already have a family with one or two children, which sooner or later leads to the sad fact that the grade of bonding of the people to Hungary gets looser, the children will start to go to school in Austria already, their socializing takes place already abroad, and the chance of complete immigration grows. Ten percent of the respondents already have a temporary residency in Austria and half of them (N = 611) are already working there; 80% of the respondents (N = 611) does speak German at some level.

Results Related to Weighing of the Individual Motivational Factors

The author has analyzed the factors influencing the respondents (N = 611) in their choice of working abroad in Austria with the help of eight attitude statements defined by the author at the beginning.

The average values 1 to 5 on the Likert scale below the statements express to what extend the respondents (N = 611) agree with them. Based on the results exposited in Figure 1, the most important motivational factor for work-related migration is the fact that people can earn much more money in Austria than in Western-Hungary. Therefore, the answer on Q1 is definitely “yes”.

The second most important factor for them is the chance to improve their German language knowledge, which together with the possibility of higher earning got a result over 4 on the Likert scale. For the responders the third, but still important factor is the chance to improve their professional knowledge. This possible (wished) development can be considered as a brain gain process, which at the end again can lead to another step forward, followed by a potential increase in their salaries. Professional improvement only got a 3.43 on the Likert scale, but can still be regarded as very important, due to the fact that the upcoming factors do not even reach 3 on the scale. As a result of the values given by the respondents on the importance of the improvement needs (both language- and professional knowledge), Q3 gets a clear “yes” answer. Upon introducing the situation of Hungary as a country of origin in this study, the author has stated that there is already a lack of workforce in Western-Hungary on some specific fields. Based on this, it is quite normal that the factor saying that there are no available jobs in Hungary only got 2.79 on the Likert scale. The extent of the social benefits is not a critical

issue for the respondents; this factor is only the fifth most important out of the eight attitude statements with a value of 2.36 on the Likert scale. This actually answers the Q2 with saying a clear "no": The social benefit related factor is not that important for the respondents intending to work in Austria. At the back, we find the political factors and the temporary solution seekers, both slightly above 2.1 on the Likert scale. And at the very end (already below the value 2) we can find those, who have the intention of getting higher education in Austria.

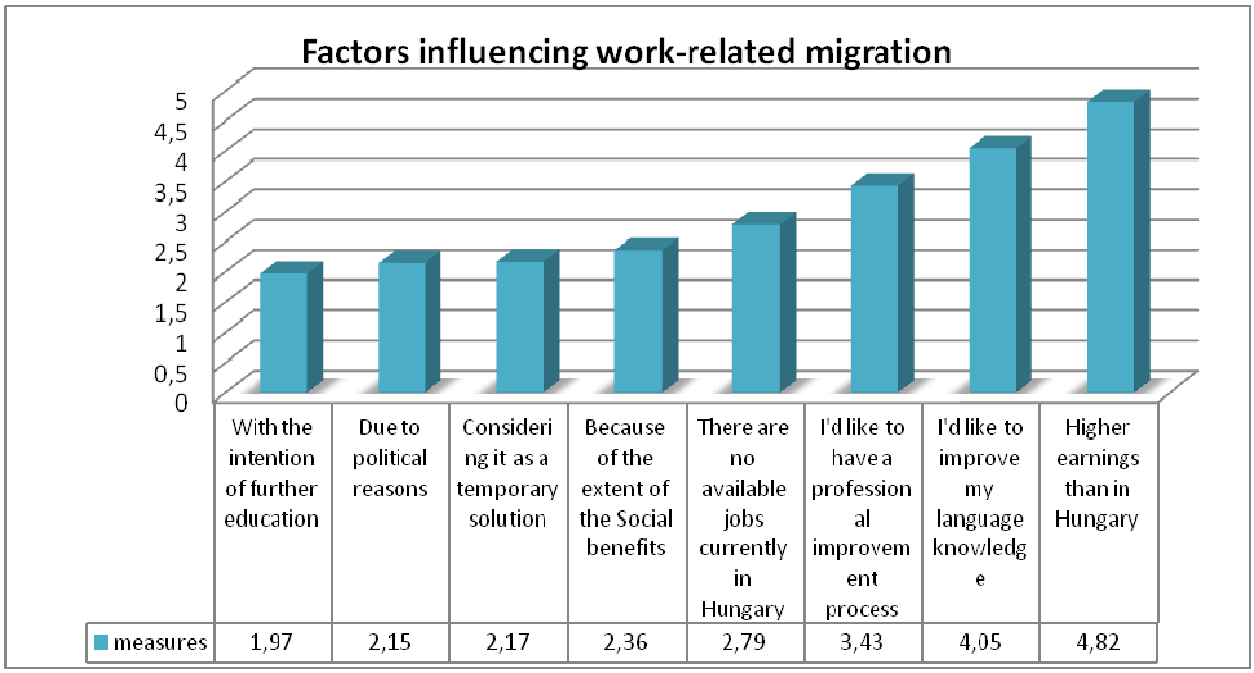

Figure 1. Factors influencing work-related migration. Source: Own research (Kovács, 2018).

Results of the Cluster Analysis

Q4 is looking for an answer on whether it possible to group people based on the factors. In order to be able to give an accurate answer on this one, the author has made a cluster analysis related to the eight attitude statements, which does group the similar elements. Upon deciding on the number of clusters, it is of key relevance that the groups shall be as heterogeneous as possible, but at the same time, it is important that the numerical differences are not supposed to be too big. The 2 and 3 cluster solutions based on the Ward-method both show too big numerical differences, but based on the frequency tables, the 4 cluster solution is the most optimal, because in this case, the cluster members do stand the closest to each other. The best tool for defining the group of clusters is the discriminant analysis. Based on the mean and deviation of the individual variables of groups and total, we can define the point of view of the members of the 4 clusters related to the eight attitude statements.

The attitude statements and their values in each cluster are summarized in Table 1. As stated earlier, not only the high values but also the low once can be of specific importance. As the outraging values have been highlighted, we can see that according to the members of Cluster 1, the higher earnings have key priority, followed by the need for the improvement of their language skills, and quite interesting: The political reason as factor is also very important for them.

For the members of Cluster 2, the higher earnings are of highest importance, and just like for the members of Cluster 1, the improvement of their foreign language skills is on second place. On Rank 3, they also want to improve their professional knowledge and a very low value (1.21) is given by them to the political reasons, which means this is indeed very uninteresting for them.

Cluster 3 very much differs from the previous two, although the higher salary is for them also a very crucial issue. Concerning the remaining seven attitudes, they gave quite low values: Four attitudes are below 2 and three of them actually only slightly above 1. We can clearly see that the members of this cluster actually really are only keen on earning "big" money in Austria.

Cluster 4 again puts the higher earnings on first place, just like the members of all other clusters, but for them the higher extent of the social benefits gets on Rank 2. Political reasons, the intention to participate in further education and considering working in Austria only as a temporary solution, are undervalued by the members of this cluster, therefore they do not regard these factors as important.

The four clusters have some relevant points in common: For all of them, the first and most critical motivational factor for working abroad in Austria is the possibility to legally earn much more money, than they can earn in their homeland. The second most important factor is the possibility to improve their language knowledge abroad, basically in a mother-tongue environment. Most irrelevant is the work-related migrational activity due to political reasons and the high importance of the different extent of the social benefits can also not be determined: It did not even get an average ranking based on the answers.

As we can see, indeed people can be grouped based on factors into clusters. Q4 is this way answered with a clear "yes".

Table 1

Summary of the Four Clusters Related to the Attitude Statements

Attitude statements	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Value	Members	Value	Members	Value	Members	Value	Members
There are no available jobs currently in Hungary.	3.03	181	2.79	248	2.36	72	2.64	106
Higher earnings than in Hungary.	4.78	181	4.83	248	4.76	72	4.91	106
I would like to have a professional improvement process.	3.61	181	3.92	248	1.58	72	3.14	106
I'd like to improve my language knowledge.	4.34	181	4.26	248	3.11	72	3.67	106
Considering it as a temporary solution.	2.53	181	2.01	248	2.79	72	1.53	106
Due to political reasons.	3.99	181	1.21	248	1.21	72	1.73	106
With the intention of further education.	2.52	181	2.03	248	1.07	72	1.47	106
Because of the extent of the social benefits.	2.7	181	1.85	248	1.08	72	3.85	106

Source: Own research (Kovács, 2018).

Conclusion

The analyzes of the factors with the help of attitude statements influencing the decision of taking a job in Austria, has clearly proven that for the people living in Western-Hungary the most dominant argument is the much higher salary they can get on a monthly basis. The higher social benefits are not that important as thought prior to the research, and except for the need of the Western-Hungarians to improve their German language knowledge and their professional skills, no other dominant factor can be determined. The current political situation in either Hungary or Austria has raised very low attention, and almost no one chooses Austria instead

of Hungary as a temporary solution. Also, concerning the further education, the author has got a relatively negative feedback, meaning that it cannot be detected as a key motivational factor for the people of Western-Hungary. Just as described at the beginning of this study upon introducing Hungary as a COO country in this case, the fact that we face shortages in the number of potential employees was returned in the answers, as members of the four clusters did not rank very high the statement that there are no jobs available currently in Hungary.

With the help of cluster analyzes, the author has proven that people can actually be grouped based on their motivation and intention why to work in Austria; the author was able to set up and deeper analyze four clusters. In general, the paper has again verified many of the previous literature (also stated in the study), searching for the key motivational factors upon taking the decisions to work in a foreign country abroad.

References

Barta, G., & Kovács, E. (2017). Older persons on the Austrian and Hungarian labor markets. In Karlovitz (Ed.), *Essays in economics and business studies*. Komarno: International Research Institute.

Blaskó, Z. (2014). Surveying the absentees: Surveying the emigrants. *Seemig Working Papers Series 4*, Hungarian Demographic Research Institute, Budapest.

Blaskó, Z., Ligeti, A. S., & Sik, E. (2014). Magyarok külföldön: Mennyien? Kik? Hol? *Társadalmi Riport*, 351-372. Retrieved from http://old.tarki.hu/adatbank-h/kutjel/pdf/b337.pdf

Bodnár, K., & Szabó, L. T. (2014). A kivándorlás hatása a hazai munkaerőpiacra. *MNB-tanulmányok 114*, Magyar Nemzeti Bank, Budapest, 2014. Retrieved from https://www.mnb.hu/letoltes/a-kivandorlas-hatasa-a-hazai-munkaeropiacra.pdf

Csapó, T. (1999). Határ menti együttműködések a munkaerőpiac területén, különös tekintettel Vas és Zala megyére. In M. Nárai and J. Rechnitzer (Eds.), *Elválaszt és összeköt: A határ* (pp. 269-295). Pécs: MTA RKK.

Cseres-Gergely, Z. (2003). The effects economic incentives on residential mobility and migration in Hungary, 1990-1999: Society and Economy. *Journal of the Budapest University of Economic Sciences and Public Administration, 3*, 351-382.

Cseres-Gergely, Z. (2005). County to county migration and labour market conditions in Hungary between 1994 and 2002. *Zeitschrift für ArbeitsmarktForschung (ZAF), 37*(4), 425-436.

Dialog, B. I. (2011). *Auswirkungen der demographischen Entwicklung auf Arbeitsmarkt und soziale Systeme: Positionen der österreichischen Sozialpartner*. Retrieved from http://www.sozialpartner.at/wp-content/uploads/2015/08/Bad-Ischl-2011-Kurzversion.pdf

Hardi, T. (2009). Határon átnyúló ingázás, munkavállalás egy magyar-osztrák és szlovák-magyar példán keresztül. In T. Hardi, Z. Hajdú,and I. Mezei (Eds.), *Határok és városok a Kárpát-medencében*. Győr-Pécs: MTA RKK.

Hárs, Á. (2002). A munkaerő migrációja és az uniós csatlakozás (Labour migration and the enlargement of the EU) *KOPINT-DATORG Working Papers 34*, p. 41, Budapest.

Hárs, Á. (2012). *A munkaerő migrációja Magyarországon a kilencvenes és a kétezres években: Bevándorlás és elvándorlás a munkaerő-felmérés adatai alapján*. Budapest: MTA KRTK KTI.

Hárs, Á. (kutatásvezető). (2009). Magyarok az osztrák munkaerőpiacon: Ingázók, bevándorlók, munkaerő-migránsok? *Szintetizáló Tanulmány, Kutatási Zárójelentés,* p. 76, Budapest.

Hauptverband der Österreichischen Sozialverischerungsträger. (2018). *Statistische Daten aus der Sozialversicherung: Beschäftigte in Österreich*. Retrieved from http://www.sozialversicherung.at/cdscontent/load?contentid=10008.600539& version=1539602425

Juhász, J. (2006). Migráció és feketemunka Európában: Migration and Irregular Work in Europe (MIGIWE). *Záró Tanulmány*, MTA Földrajztudományi Kutatóintézet, Panta Rhei Társadalomkutató Bt. Budapest.

Kaczmarczyk, P., & Okólski, M. (2005). *International migration in Central and Eastern Europe*: *Current and future trends*. New York: UN Population Division, Department of Economic and Social Affairs.

Kapitány, B., & Rohr, A. (2013). A Magyarországon állandó lakcímmel rendelkező 18-49 éves magyar állampolgárok. *Korfa, 13*(3). Budapest: KSH-NKI.

Kertesi, G., & Köllő, J. (1997). Reálbérek és kereseti egyenlőtlenségek, 1986-1996: A bérszerkezet átalakulása Magyarországon, I. *Közgazdasági Szemle, 7-8*, 612-634.

Kotowska, I. E. (2006). Older workers in the labour market and retirement policies. In R. Palomba, and I. E. Kotowska (Eds.), *The economically active population in Europe: Population Studies, No. 40* (pp. 55-89). Strasbourg: Council of Europe Publishing.

OECD and EU. (2014). *Matching economic migration with labour market needs.* Paris: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/social-issues-migration-health/matching-economic-migration-with-labour-market-needs_9789264216501-en

Österreichischer Integrationsfond (ÖIF). (2014). Migration & integration. *Zahlen.daten.indikatoren*. Retrieved from https://www.bmeia.gv.at/fileadmin/user_upload/Zentrale/Integration/Integrationsbericht_2014/migration_integration2014-web.pdf

ÖIF. (2017). Migration & integration. *Zahlen.daten.indikatoren* Retrieved from https://www.integrationsfonds.at/fileadmin/content/AT/Downloads/Publikationen/Statistisches_Jahrbuch_migration_integration_2017.pdf

ÖIF. (2018). Migration & integration. *Zahlen.daten.indikatoren*. Retrieved from https://www.integrationsfonds.at/fileadmin/content/AT/Downloads/Publikationen/Statistisches_Jahrbuch_2018.pdf

Rechnitzer, J. (2005). Az osztrák-magyar határ menti együttműködés múltja, jelene. *Tér és Társadalom, 2*, 7-29.

Van der Gaag, N., & Van der Erf, R. (2008). EUROPOP 2008 compared with EUROPOP 2004. *European Observatory on Demography and the Social Situation: Demography Network, European Commission*, Brussels.

Wasserbacher, H., & Michenthaler, G. (2007). *Entwicklung des Arbeitsmarktes im Grenzraum Burgenland/Ungarn während der Übergangsperiode*. Retrieved from http://expak.at/expak/img/uploads/expak636.pdf

World Bank. (2011). *Migration and remittances Factbook.* Retrieved from https://siteresources.worldbank.org/INTLAC/Resources/Factbook2011-Ebook.pdf

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 433-443
doi: 10.17265/2328-7144/2018.06.002

Reform of the Indonesian Civil Service: Looking for Quality

Prijono Tjiptoherijanto
University of Indonesia, Depok, Jakarta, Indonesia

Important aspects toward restoring trust in government among others are the role of leadership in building the credibility of reforms implemented after 1988 in Indonesia. Since the Reformation Era, the objectives are improving democratic and local governance, moving governance closer to the people, and strengthening transparency and accountability of the government, and the role of bureaucracy. With respect to civil services, it is very important to develop a modern and efficient government system.

Keywords: decentralization, civil service commission, pay system

Introduction

Indonesia had a large member of civil servants, approximately 4.35 million or 1.3% of the 2017 population. In adapting to the globalization era, the Indonesian government has to improve the structure of its bureaucracy, both in terms of enhancing the quality of government employees, and developing a modern and efficient government system.

Since decentralization has been implemented in 1999, Indonesian local governments are demanded to be more creative in managing their human resources due to their responsibility increasing competitiveness and developing local autonomy. Local government institutions should consist of high competent bureaucrats in order to provide better public services and develop its local government capacity.

Size of Civil Service

Size of civil servants in Indonesia was growing-up since the end of colonialization period. At the end of the Dutch colonial rule, around 50,000 persons were employed in the civil service, with only a small portion of Indonesians occupying senior positions (Bintoro, 1991, p. 75). The number of civil servants increased drastically after Indonesia gained her independence. In 1950, the civil service employed around 303,500 persons and in 1960 around 393,000. The number of civil servants per 1,000 inhabitants increased from an estimated 1.1 in 1940 to 3.7 in 1950 and 4.1 in 1960. The second large increase in the number of civil servants came in the 1970s, when increased revenue from oil allowed the government to expand its scope of activities. From around 525,000 in 1970, the number of civil servants increased to more than two million in 1980. Calculated per 1,000 inhabitants, the number increased from 4.4 to 13.9 during the same period (all figures from Evers & Schiel, 1988). From 1986 to 1992, the civil service grew by 25 percent, with a significant lower growth rate in the latter years of this period. The number of civil service reached its peak in 1993 with slightly more than four million positions, or 21.8 civil servants per 1,000 inhabitants. This number remains same with

Prijono Tjiptoherijanto, Ph.D., prof., Faculty of Economics and Business, University of Indonesia, Depok, Jakarta, Indonesia.
Correspondence concerning this article should be addressed to Prijono Tjiptoherijanto, Faculty of Economics and Business University of Indonesia, Depok, West Java16424, Indonesia.

slightly increase to became 4.35 million in 2017.

In terms of percentages, it is around 1.9% of the total population. These percentages are similar to those other countries in the region, such as India (1.2%), Pakistan (1.5%), the Philippines (2.1%), and Vietnam (3.2%) (Schiavo-Campo, 1998). The growth of civil servants in Indonesia can be seen from Table 1.

Table 1

Growth of the Civil Servants (1950-2017)

Year	Number of civil servants	Civil servants per 1,000 inhabitants
1950	303,500	3.7
1960	393,000	4.1
1970	515,000	4.4
1980	2,047,000	13.9
1993	4,009,000	21.8
2012	4,646,357	19.7
2017	4,348,698	18.5

Source: Own calculation.

Educational Background and Rank

Based on data from The National Civil Service Agency (www.bkn.go.id, 2017), the total number of the-public servants in Indonesia is 4,348,698 in June 2017. However, the largest number in Indonesian public servants is graduated from senior high school occupied 23.5 percent. Meanwhile the percentage of civil servants who achieved master degrees and doctoral degrees is only 6.91 percent, almost similar to the total number of those who finished elementary school and of the junior high school (2.1 percent). Table 2 shows the educational background of public servants in Indonesia.

Table 2

Educational Background of Indonesian Public Servants per June 2017

No	Educational background	Total	Percentage (%)
1.	Elementary-junior high school	91,323	2.1
2.	Senior high school	1,021,945	23.5
3.	Diploma 1 s.d 3	795,811	18.2
4.	Diploma 4 and undergraduate degree (S1)	2,139,559	49.2
5.	Master degree (S2) and doctoral degree (S3)	300,060	6.91
Total		4,348,698	100.00

Source: The National Civil Service Website (www.bkn.go.id, 2017).

In terms of *golongan* (rank), civil servants in Indonesia are divided into four ranks, from I (the lowest) to IV (the highest). Rank I through III are divided into four grades (a, b, c, and d), and Rank IV has five grades (a, b, c, d, and e), making a total of 17 grades from I (a) to IV (e). Individual civil servants' ranks are based on their educational qualifications and seniority. Ranks III and IV require and university degree, while Rank I only need an elementary and secondary education and Rank II mostly occupied by who has finished a senior high school.

As see in Table 3, the majority of civil servants occupy Rank III, about 47.43 percent. It is followed by Rank II (30. 81 percent) and Rank IV (19.11 percent). The smallest percentage is Rank I, about 2.65 percent.

Table 3

The Proportion of Public Servant's Rank per June 2017

No.	Rank	Total	Percentage (%)
1.	Rank I	125,445	2.65
2.	Rank II	1,458,055	30.81
3.	Rank III	2,244,785	47.43
4.	Rank IV	904,187	19.11
Total		4,732,472	100.00

Source: The National Civil Service Website (www.bkn.go.id, 2017).

Salary System

Even though the number of civil servants in Indonesia is equivalent to only about 1.9 percent of the total population, the quality of government employees is low. This is partly an outcome of the unattractive salary system. To attract effective, efficient, and uncorrupt government employees, they need to be provided with appropriate salaries and benefits. Appropriate compensation will not only have an impact on staff turnover and on employees' productivity and quality of work, but also will reduce tendencies for civil servants to engage in corrupt practices.

Salaries for Indonesian civil servants are determined by the level of responsibility, the type of job, and the cost of living. The salary system for government employees in Indonesia is classified in a combination scale system since it combines the single scale system and the double scale system. Under a single scale system, employees at the same rank receive the same salary regardless of the type of job and the level of responsibility. Under a double scale system, salaries are determined based on employees' level of responsibility and type of job. Job performance is not generally taken into account. Under the combination scale system, some civil servants might have significantly higher salary than their colleagues at the same rank.

The basic salary for a civil servant at Rank I (a) (primary and junior high school graduates), regardless of the job held and the level of responsibility is around US$114 per month, or a little over US$4 per day.

The salary for an employee at Rank IV (e) with 32 years of service is the less than a chief executive officer of an Indonesian state owned enterprise. In fact, income disparities between the private and the public sector are widening. The income earned by civil servants in Indonesia is just one-quarter, or at best one third, of what employees of private companies receive (Tjiptoherijanto, 1995).

Table 4 presents a comparison of government and private pay by education in 1998. On average, government earnings at 414,000 rps./month exceeds the national, non-government average of 274,000 rps./month. This is not surprising since government is more education-intensive than the private wage sector.

In the report from Central Bureau of Statistic in 1998, around 49 percent of workers that are engaged in the private wage sector have a primary education or less as compared to only 5 percent for workers employed by government. When disaggregated by education level, a government pay premium remains at lower education levels; close to pay parity is achieved for graduates of senior high school; and a private sector premium emerges for those with some tertiary education ("Diploma I/II" or "Akademi./Diploma III") or a university degree (Universitas/Diploma IV).

The pattern of government pay exceeding private compensation for less educated workers and private pay exceeding government compensation for more educated workers—the problem of government salary compression—is a pattern common to other civil services (Nunberg, 1994). Indonesia's situation does not

appear unique.

Table 4

Monthly Earnings by Education Level, 1998 (Percent of Wage Earners in Category)

Education level	Earnings 1,000 rupiahs/month (percent)		
	Private sector	Government	Private to government pay ratio
Primary or lower	192 (42.2)	290 (0.7)	0.7:1
Junior high school	239 (13.7)	379 (1.2)	0.6:1
Senior high school	337 (23.5)	392 (8.2)	0.9:1
Some tertiary	530 (3.2)	458 (2.0)	1.2:1
University or higher	771 (3.3)	520 (2.1)	1.5:1
All levels	274 (85.8)	414 (14.2)	0.7:1

Source: Authors' analysis of Sakernas (National Manpower Survey): Central Bureau of Statistic (1998).

In recent years, the government has become aware of the need to link civil servants' salaries to those paid in the private sector if they are to attract and retain the talent necessary to improve and sustain public sector performance. When income inequality among staff is deliberately increased, senior management positions become more attractive than was previously the case. In theory, an egalitarian pay structure is more attractive to those in the lower ranks of the civil service, whereas a pay structure that more clearly differentiates between staffs at different levels is conducive to recruiting and retaining talent that might move to the private sector (United Nations, 2005). However, Indonesia's salary structure is moving towards an egalitarian system, resulting in most of its best graduates from well-known and highly-qualified universities uninterested in becoming government employees. Moreover, the low salary tends to encourage wrongdoings and illegal activities, such as accepting bribes and asking for compensation for services provided.

In Indonesia, as in many developing countries, allowances and in-kind benefits play a substantial role in remunerating public sector employees, which is why determining the right balance between pay and benefits, and allowance is very important. In Zambia, for instance, permanent secretaries earn 50 times as much as the lowest paid civil servants when in-kind benefits (housing, cars, telephones, and so on) are taken into account, but if such benefits are excluded, the difference is only fivefold (Kenleers, 2004). Moreover, where "moonlighting" and corruption prevail, senior civil servants will earn more than junior ones, as they are likely to have more opportunity to engage in such activities. The income of civil servant in Indonesia consists basically of three elements:

(1) The basic salary which based on the rank and grade of the civil servant.

(2) Various standardized allowances, like rice and family allowances, structural allowances (for holders of structural positions), functional allowances (for holders of functional allowances), and special allowances for civil servants working in remote areas (like Papua in the past).

(3) Other salary supplements in cash or kind, like *Idul Fitri* (the Moslems holiday) bonuses, provisions of transport to and from the office, housing, daily subsistence allowances for official travel, and medical care.

As an example of the allowances given to the Indonesia's civil servants, who are holding structural

positions, Table 5 given information on allowance for structural positions.

Table 5

Allowance for Structural Position for Civil Servants 2017

Echelon/rank	Allowance (IDR)
I A (4e)	5,500,000
I B (4d-4e)	4,375,000
II A (4c-4d)	3,250,000
II B (4b-4c)	2,025,000
III A (4a-4b)	1,260,000
III B (3d-4a)	980,000
IV A (3c-3d)	540,000
IV B (3b-3c)	490,000
V A (3a-3b)	360,000

Source: Government Regulation No. 26 Year 2017 dated December, 2017.

In 2017, as a result of the bureaucracy reform initiative implemented in the Ministry of Finance, an additional allowance to increase staff take-home pay were substantial as seen in Table 6 below.

Table 6

Monthly Special Allowances Provided Under the Additional Grading Scheme in Presidential Decree, 2015.

No.	Position	Grade	Special allowance (in Rupiah)
1	Director general	27	46,950,000
2	Secretary, directorate general	23	24,100,000
3	Head of division	19	13,670,000
4	Head of sub division	15	7,474,000

Source: Presidential Decree number 156/2014.

Benefits from Decentralization

Although in principle, the conditions of service regarding salaries, allowances, working hours, etc. are uniform throughout the Indonesian civil service, the regional government since implementation of the decentralization policy practiced attractive compensation to the servants in respected regions.

Proponents of decentralization see it as a process that enables more efficient allocation of resources, reduces information asymmetries, increases transparency, promotes citizen participation, and enhances accountability, thereby improving governance. Local governments are often more aware of and attuned to the needs of local populations than the central government, which means that local government may have a clearer sense of which projects and polices people living in their jurisdictions would favor. This will have an impact on the duties of civil servants in different regions.

In Indonesia, the decentralization process coincided with civil service reform, and particularly changes to pay scales for civil servants. Prior to the implementation of decentralization measures beginning in early 2001, one issue of concern had been the extremely low salaries for government employees, which were considered to give an insufficient incentive for work in public service while minimizing the temptation of corruption. As a result of decentralization policy, the increased ability to legislate at the local level has now given local governments the ability to increase salaries employ new incentive structures to reward performance, and increase the differential between the highest- and lowest-paid civil servants. For example, as of December 2006,

a decree by the governor of Riau Province in West Sumatra gave the lowest-ranking civil servants (I a) an additional Rp$1.6 million (approximately US$160) per month, while the highest ranking (IV e) received a pay increase of Rp$4.5 million (approximately US$450) per month.

In addition to regional civil servants being paid more in line with their rank, functional professions are also recognized by means of additional functional allowances. Example, in East Kutai Regency in East Kalimantan Province, since 2006, elementary and high school teachers have been paid an additional allowance of Rp$ 1.2 million (approximately US$120) per month. Consequently, teachers at the rank of II or III now receive monthly incomes of approximately US$250-290, which is significantly higher than the province's minimum wage, which was set by government decree at roughly US$150 per month in that year.

However, the changes in salary levels at the regional and provincial level, which relate to the resources and wealth of the respective area, and particularly, the discrepancy between the lowest- and highest-paid civil servants have created some problems. Whereas civil servants at the regional level may receive additional bonuses and allowances, civil servants within central government must be satisfied with the salaries given according to the law. Thus, the most senior civil servants in Riau are paid more than twice the basic salary received by central government civil servants of the same rank, which is only around US$207 per month. With a salary of US$657 (based pay of US$207 plus US$450) per month, civil servants in Riau earn almost as much as middle managers in the business sector in Jakarta, the capital.

Nevertheless, despite some improvements following the implementation of decentralization policy or regional autonomy, Indonesia's public sector still needs to undergo substantial change especially with regard to human resources development programs. Which are relevant to the civil service system, in order to have a higher posotion?

Human Development Programs

In countries, such as Indonesia where civil servants, like politicians, are key government decision-makers; government employees are sometimes viewed as community leaders. In this sense, civil servants may be expected to perform many duties in the community where they live, following practices established during the Dutch colonial era. Such role calls for adherences to norms and morality meaning the civil servants must avoid irregularities and always obey the rules when conducting their activities (Magnis, 1996; Natakusumah, 1990). Therefore, civil servants should not engage in illegal activities, such as bribery, corruption, and other misconducts.

Friederich (1940) noted growing importance of internal values and moral and professional standards among bureaucrats. In their absence, abuse of power can easily arise in the government sector.

A study by Meir and O'Toole (2006) showed that bureaucratic values are far more important in explaining bureaucratic output and outcomes than political factors. This should be taken to mean that external political control is unimportant, but it does show that paying serious attention to the values of civil servants is important.

Ensuring that civil servants give high priority to honesty, responsibility and integrity with regard to their routine duties can be accomplished through well-planned human resource development. Human resource development for civil servants starts with their recruitment and continues until they leave government service. Recruits should undertake job and requirement analyses before undertaking recruitment activities. Furthermore, to allow the civil service to select the best candidates, the recruitment process should be fair and open.

The next step in human resource development for civil servants is education and training. This should be

provided regularly for those at every level, as is already done in the armed forces. Considering the importance of trainings, in Indonesia, training and education plays a major role in the effort to increase the quality of civil service. However, training and education (pendidikan dan pelatihan-diklat) aims not only at the improvement of job- and work-related skills and knowledge. Forming the attitudinal and behavioral characteristics of civil servants, and ensuring their political allegiance with the program of the government has always been an integral part of the human resources development programs of the government.

In order to have maximum result training should be linked to career development and personnel planning. The training process should cover training needs identification, implementation, and evaluation of training. Training in the Indonesian civil service can be analyzed under different dimensions as shown in Figure 1. The dimensions which are based on status, on career, and on tasks, consist of:

(a) Structural training is career development training for those civil servants holding structural positions. In order to be eligible for promotion to the Eselon III, Eselon II, and Eselon I levels, civil servants have to attend the respective training courses (Diklatpim III for Eselon III, Diklatpim II for Eselon II, and Diklatpim I for Eselon I). Attendance of these courses is by selection and assignment. Since structural positions are general management positions, the structural training courses cover a broad range of subjects including functional management, general aspects of management and leadership, management techniques, interpersonal and communication skills as well as political and economic issues. The main objective of the training is to increase the management capabilities of the participants, and to build up a cadre for the administrative (and political) leadership of the bureaucracy.

(b) Functional training is training for the holders of functional positions. Although it has an influence on the career development of the respective civil servants, its main purpose is to increase the knowledge in the respective professional field. Functional training is differentiated depending on whether it aims at improving the general competence and knowledge of the participants (diklat fungsional keahlian), or whether it aims at the improvement of specific skills (diklat fungsional ketrampilan). Functional training can also be attended by holders of structural positions, if the functions of the structural position require that specific knowledge.

(c) Technical training is related to the immediate tasks of a civil service position, irrespective whether it is a structural or functional position. While technical training linked to the direct job assignment institution (diklat teknis substantif) is implemented by the individual government institution, general technical training (diklat teknis umum) (e.g., in project management, job analysis) is implemented centrally by the National Institute for Administration (LAN).

(d) General Administration training is preparatory administrative and management training for those civil servants who are being promoted to the structural positions of Eselon V and IV levels. The main purpose is to provide the participants with basic managerial capabilities (like integrated work planning) and administrative skills.

(e) The National Leadership Training or Diklat PIMTI is a new training course for holders of Eselon I positions, its main purpose is to broaden the political and administrative perception of these most senior civil servants, to increase their understanding of government policies and of the factors influencing the policy making of the government. Recently, the Indonesian government has introduced another kind of leadership training so-called the Reforrn Leaders Academy (RLA), in which the main objectives is equipping Eselon I and Eselon II with the bureaucratic reform’s skills and expertises which become a top priority of the government’s programs.

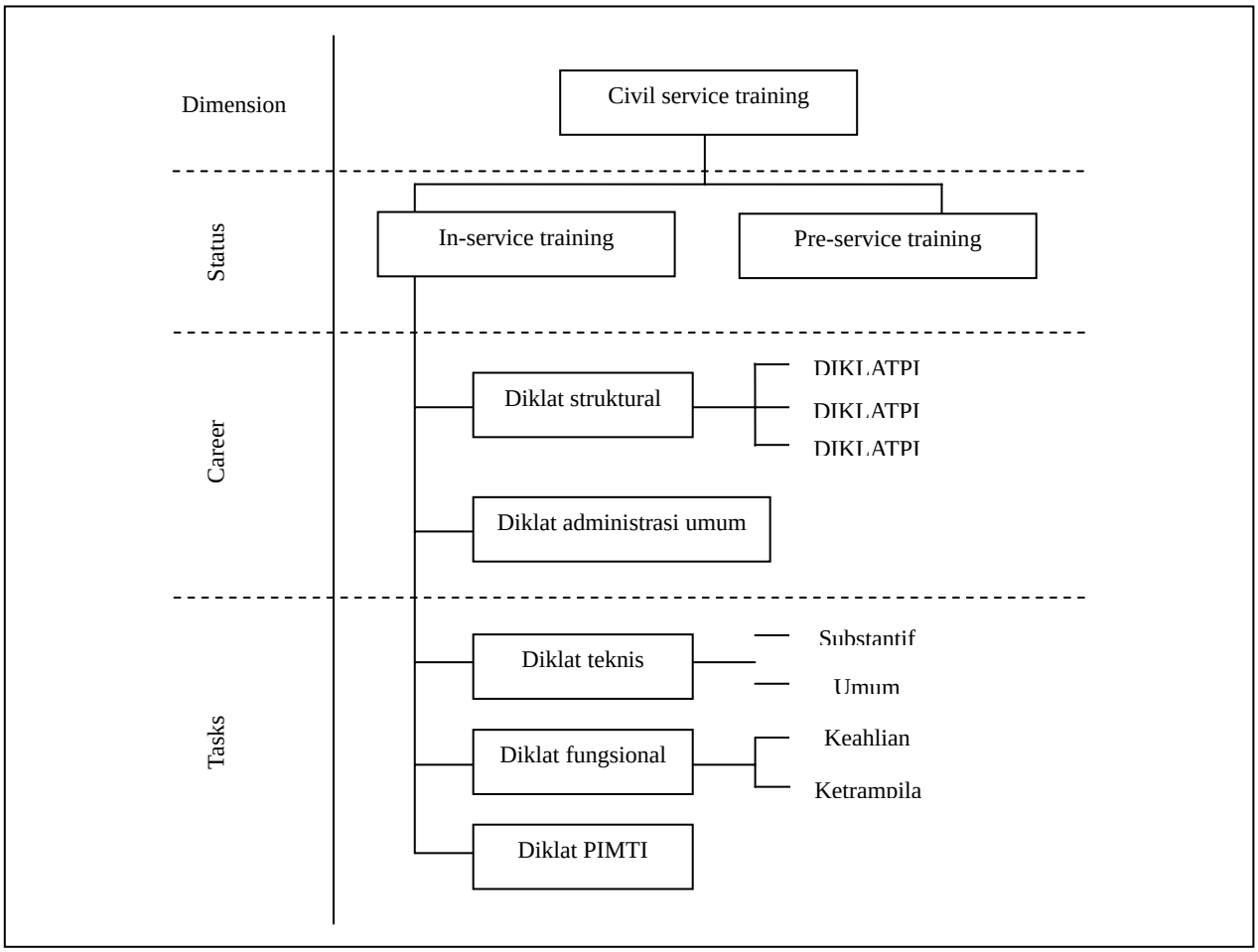

Figure 1. Dimensions of civil service training. Source: Author. *Notes.* *) Reform Leader Academy (proposed recently); Each of these training categories has its own characteristics in terms of content, target group objectives and its link with the career development; shown by rank or structural position (eselon) of the participants.

Reform Initiatives

Currently, a number of new initiatives are planned to accomplish further changes to the incentive system, the size of the civil service, recruitment, performance management, remuneration, and probity as part of the coverall civil service strategy, in line with recommendations by a World Bank report for the Indonesian government (World Bank, 2001)

A remaining issue involves the establishment of a civil service commission (CSC). Despite the improvement that followed implementation of the Regional Autonomy Law—Law No. 22 of 1999, revised by Law No. 32 of 2004, Indonesia's public sector still needs to undergo substantial change, especially towards improving governance and enabling the country to compete in the global arena. To have an effective and efficient public service, most governments have set up a civil or public service commission as a special institution responsible for human resource management. For example, the Republic of Korea established a CSC in 1999, which has been leading the country's major civil service reform initiatives. In 2004, those personnel management functions that still remained under the purview of the Ministry of Government Administration and Home Affairs were transferred to the CSC, thereby resulting in a single, central personnel authority for the

government (Kong, 2006). In New Zealand, in 1999, the state service commissioner asked to be given responsibility for developing a solution to the lack of corporate capacity in the public service. Since that time, New Zealand's public service has increasingly moved to address a wide range of service and human resource management issues from a corporate perspective (United Nations, 2005).

In the case of Indonesia, early in 2014 with the issuance of law No. 5 of 2014, currently have a CSC. Due to this, the division of responsibilities in relation to human resources among line ministries and other public sector entities is as shown in Table 7. This table also illustrates that human resource management within the civil service is being carried out not by an independent body that reports directly to the president, but by institutions that are part of the government bureaucracy. Therefore, more innovative actions are still needed in relation to the governmental institutions setting in Indonesia.

Once a civil service commission has been established, questions often arise pertaining to the commission's relationship with line ministries and agencies. Therefore, once a government decides to establish a CSC, it must clearly delineate the division of responsibilities in relation to resource management among central government departments and agencies. In many countries, responsibilities for human resource management in the civil service are along the lines shown in Table 8. The structure outlined in Table 8 resembles the model prevalent in the Commonwealth of Nations, especially with respect to the role of the CSC, but countries, such as the Republic of Korea and Thailand have similar arrangements in place.

Table 7

Institutions Responsible for Human Resource Management in Indonesia

Agency	Function
Office of the president (State Secretariat and Cabinet Secretariat)	Overall government policies
Ministry of Finance	Civil service pay and pensions (state-owned enterprises are responsible for their own pay and pensions under the supervision of the State Ministry for State-Owned Companies).
State Ministry of Administrative Reforms (MenPAN)	Supervision, coordination, monitoring, and evaluation of all civil services matters, including supervision and coordination of the National Civil Service Agency, and the National Institute of Public Administration.
National Civil Service Agency (BKN)	Appointment, promotions (except at the highest levels, which are managed by a team chosen by the president), transfer and civil service data base.
National Institute of Public Administration (LAN)	Education, training, organizational design, and research for bureaucracy matters.
Civil Service Commission (CSC)	Supervision and Control of an implementation of the merit system and code of conduct for civil servant

Source: Author.

Table 8

Responsibility for Human Resource Management in Central Government Agencies: A General Model

Agency	Function
Office of the Prime Minister	Overall government policy
Ministry of Finance	Pay and pensions
Ministry of Public Service	Deployment and conditions of service for civil servants
CSC	Appointment, promotion, transfers and discipline
National Administration Staff College	Staff training and development

Source: Adapted from United Nations (2005, Table 6).

A Civil Service Commission is the best central management structure if the main issue is to protect the civil service from political interference and other undue influence. If the issue is to reform the civil service, the establishment of CSC within a central government entity is more effective.

Conclusion

Since the 1980s, many countries, including Asian countries, have engaged in major efforts to promote administrative reform, focusing on openness, transparency, and accountability of government administration. Each country, regardless of their economic circumstances or development stage, requires good governance. For some Asian countries this becomes particularly important after the 1997 Asian financial and economic crisis.

In Indonesia, following the fall of the New Order Government in 1998, a political movement emerged which pursued reforms in relation to politics, the economy, the judicial system, and public administration. Law No. 22/1999 on Decentralization and Law No. 43/1999 and Law No. 5/2014 on Civil Service Matters opened new possibilities for public service reform in Indonesia, but the country still has a long way to go in achieving a high-quality civil service. As with any reform, strong and determined leadership is crucial. While good governance is central for anticipating the challenges of global competition, Indonesia must also undertake civil service reforms to achieve a cleaner and more efficient bureaucracy.

Civil service reform is needed to sustain the important institutional reform result achieved over the last 18 years in various sectors and policy areas and to further consolidate Indonesia as a progressing to become middle income country.

However, in all reforms activities, the role of leadership will be key in building the credibility of new reforms, providing the populace with an overall vision of a future, in which government earns and fully merits citizen confidence, and helping to commit different socio-economic actors and resources to this long-term goal.

References

Bintoro, T. (1991). *Public Administration in Indonesia (mimeo)*. State Administration Agency: Jakarta.

Bisnis Indonesia. (2007). Tunjangan Depkeu naik Rp. 4.3 trilliun (The Aplication of The Ministry of Finance up to 4.3 Trillion)). *Daily Newspaper*, July 6, Jakarta.

Evers, H., & Schiel, T. (1988). *Strategische grappen*: *Vergleichende studien zu staat, burokratie und klassenbildung in der Dritten Welt* (Strategic Groups : Comparative Studies on State, Bureaucracy and Class Formation in the Third World). Berlin Dietrich Reimer Verlag.

Friederich, C. (1940). Public policy and the nature of administration responsibility. In C. J. Friederich and E. S. Mason (Eds.), *Public policy* (pp. 3-24). Cambridge, MA: Harvard University Press.

Kenleers, P. (2004). Key issues for consideration when assisting civil service personnel management reforms in developing countries (Unpublished paper Unites Nations Development Programe, Subregional Resource Facility for Pacific, Northeast, and Southeast Asia, Bangkok).

Kong, D. (2006). Reinventing South Korea's bureaucracy toward open and accountable governance. Paper presented at the *Asian Public Nanning Reform Forum*, China.

Magnis, S. F. (1996). Morality in bureaucracy (in Bahasa Indonesia). Paper presented at the *Meeting on "The Efficiency and Effectiveness of Bureaucratic work Patterns and the Quality of Nine Years of Elementary Education in Relation to the Era of Globalization: Especially in 2003 and Beyond"*, Jakarta.

Meier, K. J., & O'Toole, L. J. Jr. (2006). Political control versus bureaucratic values: Reframing the debate. *Public Administration Review, 66*(2), 177-192.

Natakusumah, P. (1990). Quality improvement of government employee (in Bahasa Indonesia) (Unpublished paper, Lembaga Administrasi Negara [National Institute of Administration], Jakarta).

Nunberg, B. (1994). Experience with civil service pay and employment reform: An overview. In D. Lindauer and B. Nunberg (Eds.), *Rehabilitating government: Pay and employment reform in Africa*. World Bank, Washington D.C.

Schiavo-Campo, S. (1998). Government employment and pay: The global and regional evidence. *Public Administration and Development, 18*, 457-478.

Tjiptoherijanto, P. (1995). Efforts toward clean, respected government. *The Jakarta Post, Daily Newspaper*, 17 March, Jakarta.

United Nations. (2005). *World Public Sector Report 2005: Unlocking the human potential for public sector performance*. New York, N.Y.: Department of Economic and Social Affairs.

World Bank. (2001). Indonesia: The imperative for reform. *Report No. 23093-IND*, The World Bank, Jakarta.

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 444-449
doi: 10.17265/2328-7144/2018.06.003

Property Valuation Challenges for Properties Condemned for Environmental Public Upgrades

Odysseas Kopsidas
Technological Educational Institute of Thessaloniki, Thessaloniki, Greece
Leonidas Fragkos-Livanios
Hellenic Army Academy, Attiki, Greece

No property may be taken for public use without just compensation. The issue is what is "just compensation" and how it is determined in a particular circumstance. Regardless of the type of valuation, the Greek Constitution (Article 17, paragraphs 2-5) clearly forbids estimation based on possible future projections. In this way, future externalities, positive or negative, deriving from the change of use are not incorporated. This contravenes the Pareto optimal resource distribution and even challenges the Caldor-Hicks compensation criterion, especially in cases of partial takings. We suggest valuation to include a best-case and a worst-case scenario for possible future values. A combination of them, e.g., the average, will be used to determine benefits and costs and the final level of compensation. The most important externality under consideration, especially in cases of land properties condemned for public large-scale constructions, such as environmental public utilities, should be the environmental effect. The internalisation of possible negative externalities should be internalised by the construction benefactor.

Keywords: valuation, property taking, Pareto optimum, environmental externalities

Introduction

The procedure for the selection process is called optimization or maximization. In the business world, almost all the decisions may be considered as optimization problems. The optimization can be applied not only to maximize profit and minimize the cost of production, the use of the optimal size, but also in the employment of the highest quantity of a productive factor. Optimization without constraints means that a business seeks to maximize the benefits or minimize costs without putting constraints on resources that will be used. The optimization effort can be not only one but several things simultaneously, so the problem becomes more complicated. There is a problem with optimization constraints when seeking to maximize utility or minimize the cost of an activity with the restriction that is certain pore size that can be placed (Carson & Mitchell, 1993).

The individual seeks to maximize his/her own benefit, cares only for his/her private benefits and costs, and utterly is indifferent to the consequences of acts of other members of society. The relationship between risk and

Odysseas Kopsidas, Dr., Ph.D., Department of Civil Engineering, Technological Educational Institute of Thessaloniki, Thessaloniki, Greece.
Leonidas Fragkos-Livanios. Dr., Ph.D., Division of Natural Sciences and Applications, Hellenic Army Academy, Attiki, Greece.
Correspondence concerning this article should be addressed to Odysseas Kopsidas, Department of Civil Engineering, Technological Educational Institute of Thessaloniki, 17th Km Thessaloniki-Sindos, Thessaloniki 57400, Greece.

expected return is the key to whether or not to an investment. Often, projects or activities have external effects, which involve social costs or benefits, in addition to private. External influences, because they are internalized in market prices and create tension between market prices and social benefits or costs. With the economic analysis or cost-benefit analysis, it identified and quantified the benefits and costs of activity or policy, and considered whether it is appropriate and beneficial implementation of the whole economy and society (Flad, 1997).

Implementation

To properly place a cost-benefit analysis, all sizes should reflect the image of society. The various inputs and outputs that occur on a project from markets can function effectively or not. When markets function efficiently, the use of the purchase price is good approximation of the social cost. When you come from markets that not operate effectively, the price does not reflect the true social costs them. With the "Cost-Benefit Analysis (CBA)", all critical parameters problem attributed to a single base assessment, which facilitates decision-making process. The basis of evaluation is to compare benefits and costs. If the benefits are larger, then the project (or activity) is socially desirable (Hochman & Ofek, 1979).

Otherwise, it is considered socially beneficial. "Weak" Treaty Pareto: A project or a policy measure is socially acceptable when improving the welfare of every member of society. "Strong" condition Pareto: A project or a policy measure is socially acceptable when ensure improved welfare even one person without reducing the welfare of another. The Pareto principle is based on individual conception of welfare, whereby the people regarded as the best exponents of their own prosperity through their options. He has limited use, since there is almost no action to improve. As someone while continuing to deteriorate the position of others. Not discussed concepts, such as social justice or income redistribution (Karadimitriou, 2013).

The search for suitable instruments or for the best possible combination of the use of command, control, and economic instruments nowadays constitutes one of the most complex points of discussion on environmental economics. The environmental reassessment of economic procedures and the change in production and consumption of non-conservation friendly models, which constitute the fixed position of the European Union and the OECD, could be achieved by using suitable economic instruments. We are focalising his present study on environmental taxes, the most well-known and widespread category of economic instruments, by studying the advantages and disadvantages of their enforcement (Karadimitriou, 2013).

Next, we will present the experience at an international level, focusing on their use as well as the consequences on international competitiveness. According to the first theorem of economic prosperity, under certain conditions, a competitive economy guarantees a Pareto-optimal economic outcome. In other words, a competitive market leads to allocations of resources to the property that any position cannot improve not worsen the position of another. This allocation is done automatically through the price mechanism, e.g., where there is a demand that the price goes up and when a bid price it falls. The adjustment of prices solves the problem of distribution of goods. The second fundamental theorem states that the competition of firms has the objective maximization of profits consumers. Therefore, it is necessary to have a central designer will decide who gets what in the economy. In fact, the free market can lead to great disparities that can be removed by state intervention (e.g., taxation). According to the criterion of a Pareto distribution, it is effective when there is no other way allocation to improve one's position without diminishing someone else.

The externalities or the external economies occur when a person acts or a business affect other people or companies when a company imposes a cost on others, but does not compensate, or end, when a company brings

benefits in other businesses, but does not receive remuneration for providing this benefit. We can distinguish two types of externalities, public (e.g., air pollution, the water that affects the welfare of many people) and private (e.g., a person casts trash in the yard of neighbour). This movement affects the welfare of the neighbour and any other. The cases where the activity of an individual or business impose costs others refer to as negative externalities or external costs. When induced positive externality in the production of a commodity, the social costs production is less than the private cost. The optimal quantity of good "Q" optimum is greater than the equilibrium quantity "Q" market.

Notice that in both cases, either external charges or external economies, the price mechanism does not give enough information to the recipient of decisions. In one case, the values do not represent the actual cost and the other is not represent a real benefit. We say market failure. According to Pigou in his "The economics of Welfare", taxation is a effective tool for addressing the external charge. In the case, foreign economy given subsidy represents the real benefits of business. Unlike the Coase in his "The problem of social cost" where there is a way of supporting addressing externalities awarding property rights over natural resources. He argues that if the contaminant obtains a right of victims of pollution, then pollution will pay the first to stop or reduce the polluting activity. Unlike the pollutants to be able to benefit from the natural resource should compensate the victims, which have acquired the right to operate.

When there is a clearly defined system of property rights, the market mechanism will lead to an efficient allocation of resources. In environmental policy, the polluter (whether company, individual, or the State) pays applicable in several countries the world. This is automatically an incentive to reduce pollution at least at the level where the marginal cost of reducing pollution equals the marginal cost of damage causing this pollution. Also, many countries apply the system of subsidies for the pollution control. This suggests that property rights are particularly important in the formation of environmental policy. Residents should have rights property, the polluter or the victim worked (Coase, 1960). In theory, it places great emphasize on the importance of ownership of natural resources and to negotiate between those who pollute and those who suffer from pollution. One of the conclusions of R. Coase was that under certain conditions, the creation of property rights can lead the parties that are on opposite sides to have interest in negotiating among themselves to find an agreed solution on the level pollution would be considered socially acceptable (Lee, McElheny, & Gibbons, 2007).

The adoption sustainable development as a central policy choice, but as a principle of both international and European and domestic legal systems (especially after the revision of the Article 24 of the Greek Constitution) creates new standards for the role, nature, and function of environmental policy tools. In particular, the passage of regulatory approach to environmental protection, which was based mainly on the use of direct intervention on the strategic and integrated approach, requires an overall strategy for sustainable development. Main aim towards sustainable development is the environmental redefinition of economic processes and a fixed position of the European Union and Organisation for Economic Co-operation and Development (OECD) agreed at the World Johannesbur (McWilliams, 2011).

Environmental redefinition of economic processes and changing unsustainable patterns of production and consumption agreed at Johannesburg, cannot achieve with tools to intervene directly, but rather the use of economic tools. The same should be accepted and to solve the environmental problems the second generation, such as climate change, biodiversity loss, and soil erosion, as taking effective measures in this direction requires the use of other tools except those of direct intervention. The key feature of economic instruments is that the type of conduct which guides the operators of production processes associated with a particular economic advantage.

The logic function consists, in particular to internalize partially or completely, of "externalities", i.e., the impact on the environment, which is secondary effects of production processes and consumption and which are not calculated as a cost to those who cause it. This is also an established position in economic theory. It should also be noted that all financial instruments not show the same degree of compatibility with the market mechanisms (e.g., permits emissions have the greatest degree of compatibility with respect to subsidies, which a minimum).

These tools provide economic incentives for environmental change behaviour either through direct changes in the levels of prices and costs through fees products, duties on carbon or on raw materials, or through indirect changes in prices or the cost through financial and fiscal instruments, such as direct subsidies, loans, or end through creating new markets for environmental good, such as tradable licenses, etc. The production and consumption of goods and services have resulted create adverse impacts on the environment. Starting thus with the principle "I live, so befoul" and realizing that one cannot speak for the elimination of pollution, the problem lies in "how much pollution." In other words, in what will be the "optimum" level of environmental pollution or environmental protection from pollution, based on various economic, technological, social, psychological, and other parameters that apply to a society in a given period (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006).

According to literature (McWilliams, 2011), environmental degradation may be defined in economic terms, as external costs. The internalisation of these costs occurs when polluters pay a tax or a fee. A tax or fee is defined as the payment for each unit pollution deposited and is attracted to the environment or any environmental unit destruction/degradation. The main economic reason for using taxes in environmental policy is the integration the costs of pollution and any other use of the environment on commodity prices and services produced by economic activities. Such costs are called "externalities" because they are side effects of economic activity and not part the prices paid by producers or consumers who are directly involved. When the externalities are not included in prices cause major market distortions encouraging activities that are costly to society even if the private benefits are important, the calculation of the economic value of externalities is not easy. For example, recent calculations show that the external environmental costs of road transport, such as increased costs resulting from air pollution, climate, and disturbance from noise, is quite large and growing. These 40 externalities cost the EU an average of about 5.5% of Gross National Product (GDP). If you include the costs associated with accidents, costs are as high as 7.8% of GDP (EEA, 2000, Environmental Taxes: Recent Developments in Tools Integration). An environmental tax tries to incorporate these external costs on prices (internalising externalities), so that both social and private costs to come closer. The better prices allow the markets to work more efficiently, leading to a reallocation of resources under "fair and efficient" prices through the redistribution of costs. Environmental taxes also help in implementing the principle "the polluter pays", after facing those who cause pollution to the full costs of polluting their activity. In practice, there is little or no agreed data on the economic costs of externalities or their distribution and therefore the people making policies determine the price environmental taxes on those levels that they believe will achieve their goals their policy (McWilliams, 2011).

The impact of environmental taxes on competitiveness is an issue which is constantly on the agenda of discussions. The potential adverse effects on international competitiveness concern to designers of environmental policy when considering their application in energy and other products related to environmental problems. Some factors affecting the efficiency of environmental policy and effects on international trade are: the size of its economy and its influence on pricing internationally, the homogeneity or diversity of products produced in areas affected from taxation, the size of import substitution and innovation, mobility of production factors, the trade agreements between countries and as they affect the possibility of tax adjustment and other measures to reduce

adjustment costs, and the introduction of similar measures in other. Initially, the solution of environmental problems was used in system direct control (strategy command and control). The inability of the system for control and imposition of rules by the State in conjunction with the economic inefficiency characterizing the system cost of pollution control, led to the shift environmental policy, adopting economic tools to solve them. According to experience so far, financial tools, can under certain circumstances act as flexible, efficient, and effective mechanisms to achieve environmental objectives (Loures, 2015).

Concluding Remarks

In this analysis, it is considered the natural environment as a public good and environmental pollution as an external economy fails the price mechanism to internalize. In all three cases, the approach of foreign trade was with the "willingness to pay (WTP)" method and calculated the external costs generated by the degradation of the environment from the responses of respondents in monetary units. Respondents answered without knowing it was the environment to its original condition and not expect it to return to its original form. The quality of the clean environment and therefore the valuation of foreign economic burden are caused by the contamination with personal criteria and with personal endorsement of the value of that public good. In the case of archaeological monuments, residents have built their buildings. In case of lakes, respondents have developed an urban way of life around the lakes. In the case of industrial units, residents have supported throughout the local economy on them. The initial state of the environment is unknown and undefined. Also, the natural environment that is altering the original state cannot be determined. Human works and buildings create new values in the region and therefore the external costs can be measured only with the expected quality of the environment and is not lost. Allowances, taxation, and value of land use are calculated solely on the expected image of the landscape.

Therefore, the Pareto optimal socio-economic lines status is defined according to the new form of environment created after the regeneration of areas and not according to the initial state of the environment. Also in Kaldor, compensation should be determined based on the economic valuation of public goods by their own people, who judge based on expectations rather than on the past. The expected form natural environment varies from respondent to respondent and its approach to social welfare units can only be done through alternative scenarios best and worst scenario. In any case, the society wants to reach the minimum point of the charge received from the pollution and what can be achieved by the "invisible hand", but the regulation and government intervention. History has shown that the charge received by the society because of pollution varies with the socio-economic status of citizens. The more low-income residents who are more elastic is the loss of the natural environment. The elasticity of citizens deprived of the physical environment or not is measurable size.

References

Carson, R. T., & Mitchell, R. C. (1993). The value of clean water: The public's willingness to pay for boatable, fishable, and swimmable quality water. *Water Resources Research, 29*(7), 2445-2454.

Flad, H. K. (1997). Country clutter: Visual pollution and the rural roadscape. *Annals of the American Academy of Political and Social Science, 533*, 123-125.

Hochman, O., & Ofek, H. (1979). A theory of the behavior of municipal governments: The case of internalizing pollution externalities. *Journal of Urban Economics, 6*(4), 416-431.

Karadimitriou, N. (2013). Planning policy, sustainability and housebuilder practices: the move into (and out of?) the redevelopment of previously developed land. *Progress in Planning, 82*, 1-41.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car Naturalistic Driving Study data. *Report DOT HS 810 594*, National Highway Traffic Safety Administration, Washington DC, USA.

Lee, S. E., McElheny, M. J., & Gibbons, R. (2007). Driver performance and digital billboards: Final report prepared for the Foundation for Outdoor Advertising Research and Education. *Virginia Tech Transportation Institute*, Center for Automotive Safety Research, Blacksburg, VA, USA.

Loures, L. (2015). Post-industrial landscapes as drivers for urban redevelopment: Public versus expert perspectives towards the benefits and barriers of the reuse of post-industrial sites in urban areas. *Habitat International.* Special Issue: *Exploratory Spatial Analysis of Urban Habitats, 45(*2), 72-81.

McWilliams, A. (2011). Creating and capturing value: Strategic corporate social responsibility, resource-based theory, and sustainable competitive advantage. *Journal of Management, 37*(5), 1480-1495.

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 450-456
doi: 10.17265/2328-7144/2018.06.004

Econophysical Applications for Economic Progress: Monopoly and Competition Cases

Shahlar Gachay Askerov
Baku State University, Baku, Azerbaijan

This study attempts to investigate the relationship between monopoly and competition and the philosophy of progress, using the methods, models, and terms from physics. The term "progress" is a newly adopted term and is defined as the value of an increase in the production rate per unit of time. It is shown that to achieve progress, it is necessary to increase the production on a non-linear basis over time. Therefore, it is enough to have many firms that interact with each other under the influence of "market forces". It is important to have a high level of university education, a legal environment for competition and indestructible antitrust laws. Even without strong science, you can make progress by acquiring technology and creating conditions for competition in the market. A factor that strongly influences progress, of course, is technology. For the development of technology, there is a great need for science. Science is a very powerful factor that affects the non-linear change in the economic development. It is shown that during the transition period from an absolute monopoly to an imperfect monopoly the system becomes more complex, and its output characteristics become non-linear as a function of time. It was found that the relationship between monopoly and competition is very simple, and there is no contradiction between them. Initially, the market is born as a monopoly, and then with the creation of similar firms, competition, as the natural market process, begins between them.

Keywords: econophysics, progress, competition, monopoly, duopoly, production rate, Schottky diode, revers İ-V characteristic Schottky diodes

Introduction

In the modern world, competition is the driving force of the economy and an important factor in the development of the economy. Therefore, the interaction between monopoly and competition is widely studied. This problem is considered in the light of the econophysics in this article.

Econophysics is an interdisciplinary and standalone research field which has been set to form since the 1990s of the last century. The term "econophysics" was coined by Eugene Stanley (Mantenga & Stanley, 2000). Nowadays, econophysics is developing rapidly. To solve some economic problems, it is sometimes useful to apply models, methods, and laws of the science of physics. One of the brightest examples for the application of physics to solve economic problems is the gravity model of international trade. This model in 1954 was first introduced in economics world by Walter Isard (1954).

Shahlar Gachay Askerov, doctor of Science, professor, Faculty of Physics, Baku State University, Baku, Azerbaijan.
Correspondence concerning this article should be addressed to Shahlar Gachay Askerov, 23 Zahid Khalilov str, Baku State University, Baku AZ-1148, Azerbaijan.

In our opinion, some of the results we obtain in different fields of physics can be applied in the market economy. Such phenomenas include low-energy cathode sputtering (Askerov & Sena, 1969; Askerov, 1970) and the reverse branch of the current-voltage characteristic of the Schottky Diodes (Askerov, Gasanov, & Abdullayeva, 2018).

Research Methods

Similarity between the mechanisms of low-energy sputtering and the market economy was show in the report "The Study of the Market Economy Problem by the Method of Econophysics" (Sh.G. Askerov & A.Sh. Askerov, 2017). To achieve progress in the area of the market economy, one can borrow scientific results obtained in the field of cathode sputtering at low ion energies. Both cathode sputtering and the market economy have a clear relationship between the structure and the property. Thus, it is possible to alter the properties of the object by changing the structure and to convert its linear output characteristics to non-linear characteristic which is very impotent in this case.

Another physical phenomenon that is useful for the market economy is a reverse branch of the current-voltage (I-V) characteristic of the Schottky diodes (SD). Schottky diode is called the diode which is formed between the semiconductor and metal. In this case, semiconductor must definitely have a monocrystalline structure. But the metal may have polycrystalline structure. This type of diode was researched and implemented over the last hundred years. The reverse branches of the current-voltage characteristic is that this diodes vary for their shapes, which are presented in Figure 1. As it is shown in Figure 1, the shapes of the İ-V characteristic have "hard" (a) and "soft" (d) breakdowns, and sometimes is observed I-V characteristic with breaks (see Figure 1, b and c).

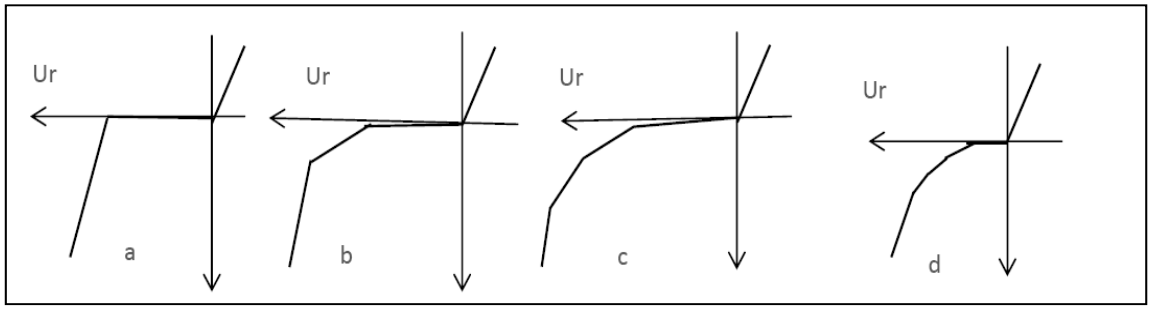

Figure 1. Reverse branches of the I-V characteristic of the metal-semiconductor contact. *Note.* Ur and Ir show the reverse voltage and the current flowing through the diodes.

Figure 1(a) shows that the SD has an ideal (hard breakdown) I-V characteristic and reverses current breaks only at one point. With an increase in the reverse voltage (Ur), the reverse current grows linearly. This means the interface of investigated diode is uniform and not only semiconductor, but also metal has a monocrystalline structure. And the diode consists only from one sub-diode. Figure 1(b) shows that the reverse I-V characteristic is broken at two points. This means that the investigated diode consists of a parallel combination of two sub-diodes which have different parameters, such as: (i) the breakdown voltage; (ii) diode area; and (iii) barrier heights. After breakdown of the first diode, the reverse current rises according to the avalanche breakdown mechanism. When the voltage reaches to the breakdown voltage of the second diode, the second diode breaks, and the reverse current breaks one more time. In the case, when the number of sub-diodes is three, the third bending appears in the reverse current-voltage characteristic (see Figure 1, c). In the same way, it is possible to

explain the I-V characteristic of Schottky diodes when the number of sub-diodes are more than 4. Obviously, if the number of sub-diodes is large, then the inverse I-V characteristic of the common diode will be non-linear (see Figure 1, d).

Thus, in the opposite direction, the linearity or non-linearity of the current-voltage characteristic of a SD depends on the degree of homogeneity of the interface. By changing the degree of inhomogeneity with the technology, it is possible to change the number of contact-forming sub-diodes. The number of sub-diodes can be varied by the shape of the I-V characteristic. At $n = 1$, the Schottky diode has an ideal "sharp", and at $n > 4$, a "soft" current-voltage characteristic. When $n = 2$ or 3, there are İ-V characteristics with 2 or 3 bends. As a result of geometrical addition of linear reverse, currents I-V characteristics, the common current-voltage characteristic becomes soft, i.e., non-linear.

It can be concluded from the aforementioned that, if the numbers of sub-diodes increases, the angular coefficient of the I-V characteristic increases continuously. As a result, the straight-forward character of the sub-diodes is transformed to become non-linear.

According to the author, the above mentioned approach can be applied in the market economy as well.

Results and Discussion

Let us suppose that there is a linear relationship between the volume of production ($\dot{I}$) and the time (t):

$$\dot{I} = b\ t \tag{1}$$

Here, b is the proportionality coefficient, which shows the volume of production produced per unit of time. İt can also be called the production rate. At $b = 0$, there is no production, i.e., commodity is not produced.

We can assume that in the case of a pure monopoly, b remains constant ($b = b_m$ = const) and does not change over time. In this case, we have the production, but we don't have the progress.

To achieve progress, it is necessary that the volume of products produced per unit of time (the speed of production) increases over time. In other words, to achieve progress the linear $\dot{I}$(t) dependence should become non-linear, or change non-linearly. It means that the in the progress coefficient of proportionality, b (in formula 1) should gradually increase. According to the author, this is a necessary condition for progress.

Figure 2 shows the dependence of the production volume of goods (or services) on time under different market structures. The production activity of the pure monopolist firm is represented by a straight line mm, while the activity of the second similar firm is shown in the straight line dd. As it is easy to see, the presence of the second independent producer of identical goods (dd) in the market creates competition and the dependence $\dot{I}$(t) is transformed from a straight line into a broken line 1-2. If another new independent producer of goods (oo lines) appears on the market, the competition will increase even more, the number of broken lines will increase to three (1-2-3).

In the economic theory, the case of 1-2 is called duopoly and the case 1-2-3 is called oligopoly (Econophysics, 2007). The angular coefficient b will respectively increase as the number of firms increases.

The parallelism of the lines mm, dd and oo in Figure 2 means that the technologies used in the production of goods by competing firms are at the same level. In addition, as can be seen from the graph, if the number of competitors grows, the output volume per unit of time will also increase: $b_o > b_d > b_m$.

Thus, in the presence of competition, the transition process from an absolute monopoly to an imperfect monopoly (oligopoly) makes market more complex, and its output characteristics differ significantly from the output characteristics of individual independent producers on the market.

The increase in the quantity of production per unit of time b in case of duopoly can be represented by the following formula:

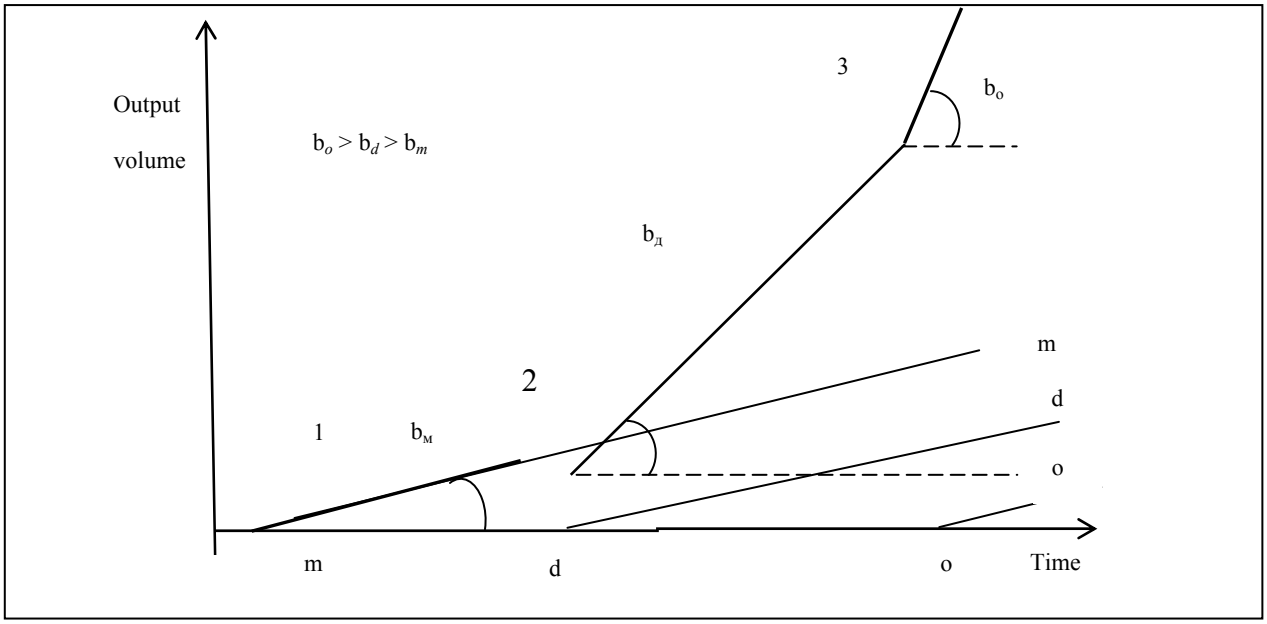

Figure 2. Dependence of the production volume of goods (or service) on time in case of pure monopoly (mm), duopoly (dd) and oligopoly (oo) market structures.

$$b_d = b_m + \alpha t \tag{2}$$

Here, b_d and b_m are production rates of a duopoly and pure monopolistic firm, respectively; α is a constant characterizing progress. According to the formula (2), progress can be defined as a value equal to the change in b per unit time, and its unit of measurement is the production volume (goods or services)/t^2.

It is known from the school physics that, the speed of the alternating motion at any time is defined by the following formula:

$$v_t = v_0 + at \tag{3}$$

where v_t is the velocity at time t, and v_0 is the velocity at the time of the reference, and a is the acceleration. Acceleration in physics shows a change in speed over time and measured by the unit: distance/t^2 . Acceleration in the economy characterizes the progress and shows the change of production rate over time and measured by the unit: production volume/t^2.

If we take formula (2) into account in formula (1), we can get that $\dot{I}$ is non-linearly time-dependent. Even with perfect competition (many firms) and an identical technology, the amount of output per unit time proportionally increases over the square of time, which leads to the emergence of progress:

$$\dot{I} = b_м\ t + \alpha\ t^2 \tag{4}$$

In such conditions, there is no need for the development of science. İt is sufficient to have a high level of university education, and a legal environment for competition. Some countries which belong to so-called emerging economies often choose this way of development.

Another factor that strongly influences the progress is, of course, technology. It is well known that technology depends on the level of science. The need for technological development leads to a greater need for science. For this reason, in developed countries the society allocates big financial resources for science.

The intellectual way to the progress goes through these stages: science technolog progress.

The relationship between the change in the level of technology (ΔT) and science (S) in a certain time interval ΔT can be represented in the following form:

$$\Delta T = \kappa_1 S \, \Delta t \tag{5}$$

where k_1 is the coefficient of proportionality.

Figure 3 shows the time dependence of the growth of production of goods by competing firms using different technologies. The graph shows the case where each newly created company uses more advanced technology ($b_o > b_д > b_м$) for economic growth.

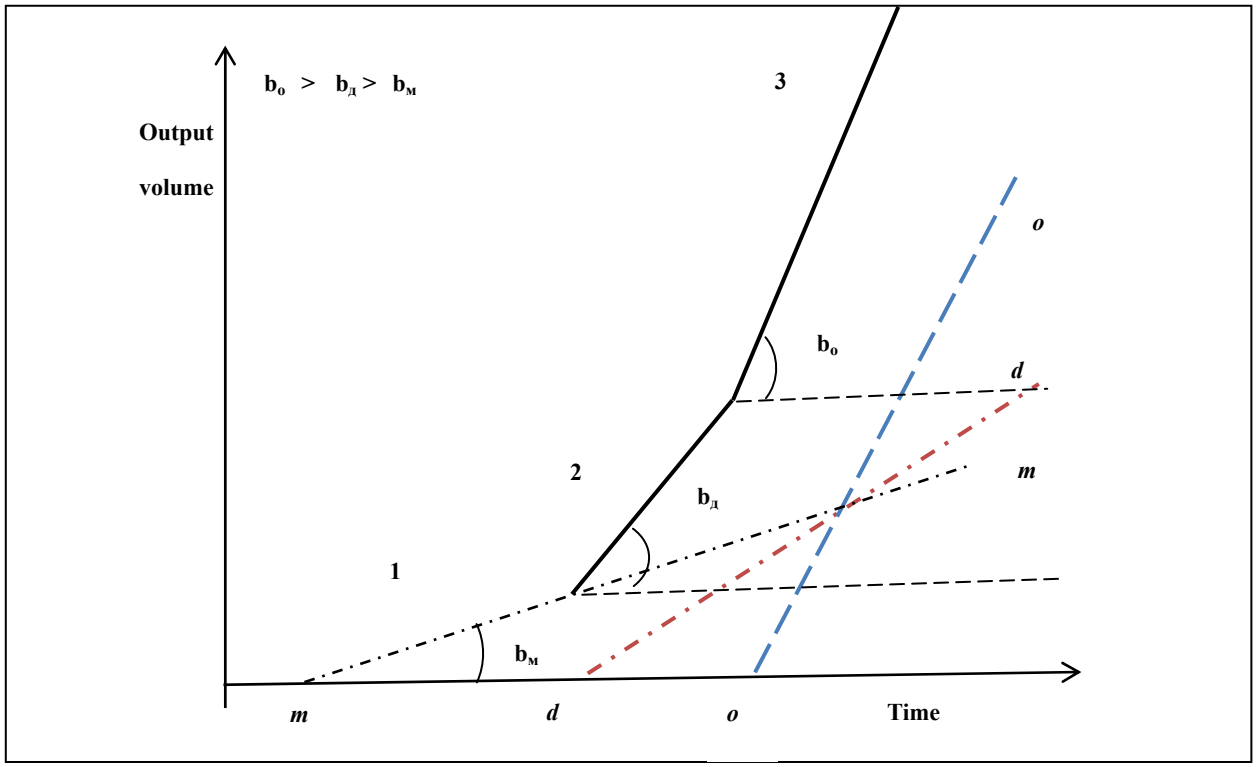

Figure 3. Time dependence of output characteristics of production with application of various technologies: mm-old technology, dd-standard technology, oo-high technology.

As it is seen from Figure 3, technology is a powerful factor to increase productivity over unit a time. For this reason, developed countries are pioneers in education, science, and technology.

Thus, we can conclude that with the increase of the number of competing firms, the types of market structures change, and their sequence from pure monopoly to perfect competition can be represented by the following chain:

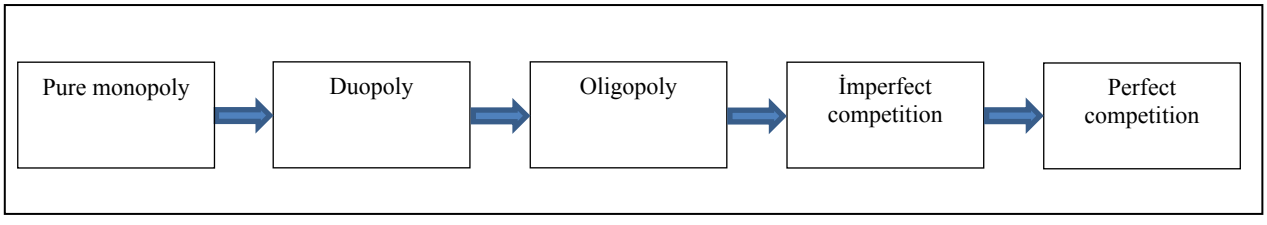

Figure 4. Dynamics of transition from pure monopoly to pure competition.

As can be seen, the relationship between monopoly and competition is very simple and there is no contradiction between them. Initially, the market is born as a monopoly, and then analogical firms were created, competition between firms starts.

From the foregoing, it can be concluded that in order to achieve progress over time, the production of non-linear products is important. To do this, it is sufficient to have a number of firms interacting with each other under the influence of competition. It is noted that if the number of firms grows, the market is subject to the law of complex systems. By acquiring technology and creating conditions for competition in the market, one can achieve progress even without having a strong science. In other words, free competition and strong antitrust laws are very important for the economic growth of countries.

Technology is a factor that strongly influences the progress. There is a great need for science in order to develop the technology. Science is a very strong factor that influences the non-linearly changing growth of the economy. Scientists are subjects that develop the science. Scientists are generators of ideas; they stand at the center of world development.

Again, as the value of the reverse voltage applied to the Shottky diode increases, the number of sub-diodes involved in the breakdown process increases. As a result, the I-V characteristics of diodes with different shapes are expressed in different linear and non-linear laws. There is also a similar event in the market economy. If the number of firms producing the same product grows, the structure of the market varies, competition takes place, and the quantity of products produced at the unit time increases.

It is shown that with the growth of the number of competing firms, the types of market structures change, and their sequence from pure monopoly to perfect competition can be represented by the following chain: pure monopoly-duopoly-oligopoly-pure competition and again a monopoly arising from technological advancement.

Conclusions

To summarize, as there is no conflict between monopoly and competition, or there is no “magic hand”, there is a natural market interaction. Like an electric charge which creates around itself an electric field, each company creates its own “market field” and interacts with each other through this field. The following results were obtained from this study:

(1) In order to achieve progress, it is necessary to bring together several firms that produce the same product (service). This makes for a complex system and creates a significantly competitive environment. Unlike the straight line characteristics of monopoly firms, the output characteristic of the complex system is non-linearity. The non-linearity of output characteristics is important for progress. Another important prerequisite for progress is the protection of entrepreneurs from monopoly, i.e., and the need for tough anti-monopoly laws.

(2) As a result of the research, the term “progress” has a new meaning, indicating an increase in production rate over a unit time, and measured by the production volume/square of the unit time.

(3) Technology is a key to rapid economic development. It is possible, to acquire technology, to create a competitive environment, and to achieve progress. There is no need for serious science here.

(4) A high technology is essential for speedy development. Technology develops by science, but the science improves by scientists. Hence, scientists are at the center of development. This fact is well-understood by developed countries. Of course, this result is not new, since the values of scientists have been understood by the society since many centuries.

References

Askerov, Sh. G. (1970). On the character of the dependence of the coefficient, sputtering on the ion energy in the threshold energy region. *Soviet Physics, Solid State, 12*, 660-661. (English Translation, Fiz Tverd Tela Trans.)

Askerov, Sh. G. (2018). The study of the market economy problem by the method of econophysics. In *4th International Conference on Global Issues in Multidisciplinary Academic Research*, February 01-02, Tokyo, Japan.

Askerov, Sh. G., & Askerov, A. Sh. (2017). Econophysical analysis of the relationship between monopoly and competition. *Journal of Economic Sciences: Theory & Practice, 74*(1), 4-6. Retrieved from http://ecosciences.edu.az/view.php?lang=en&menu=cjournal&id=344

Askerov, Sh. G., & Sena, L. A. (1969). Cathode sputtering of metals by slow mercury ions. *Soviet Physics, Solid State, 11*, 1288-1293. (English Translation, Fiz Tverd Tela Trans.)

Askerov, Sh. G., Gasanov, M. G, & Abdullayeva, L. K. (2018). The influence of the metal Microstructure on the breakdown mechanism of schottky diodes. *Materials Physics and Chemistry, 1*(2), 43-56. doi:10.63019/mpc.v1i2.124

Kharitonov, V. V., & Ezhova, A. L. (2007). *Modern physics in search of economic theory* (p. 624). M.: MİPhİ.

Kuznetsov, B. G. (2010). *Physics and economics: At the roots of econophysics* (p. 344). M.: URSS.

Leonidov, A., & Roumyantsev, E. (2013). Network topology based estimate of systemic risks in the Russian interbank market. *Journal of New Economic Association, 3*(19), 65-80. (in Russian)

Loskutov, A. Yu., & Mikhailov, A. S. (2007). *Fundamentals of the theory of complex systems*. Moscow: Izhevsk.

Mantenga, R. N., & Stanley, H. E. (2000). *An introduction to Econophysics: Correlations and complexity in finance*. Cambridge University Press.

Romanovsky, M. Yu., & Romanovsky, Yu. M. (2012). *Introduction to econophysics: Statistical and dynamic models*. Moscow: Izhevsk.

Vlasov, M. P., & Shimko, P. D. (2005). *Modeling of economic processes*. Rostov-on-Don: Feniks Publication.

Walter, I. (1954). Location theory and trade theory: Short-run analysis. *Quarterly Journal of Economics, 68*(2), 305.

Zaitsev, S., Zaitsev A., Leonidov, A., & Trainin, V. (2009). Market mill dependence pattern in the stock market: Multiscale conditional dynamics. *Physica A: Statistical Mechanics and its Applications, Elsevier, 388*(21), 4624-4634.

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 457-471
doi: 10.17265/2328-7144/2018.06.005

Cross-Sectional Variations of Illiquidity on Stock Returns, Idiosyncratic Volatility Biases in the Shanghai A' Share Stock Market

Chen Yang
Xi'an International Studies University, Xi'an, China

Under the NTS Reform (Non-Tradable Share Reform), this paper explores the cross-sectional relations between illiquidity and stock returns by considering the idiosyncratic volatility biases in the Shanghai A' Share stock market. Differing from prior studies, stock returns are decreasing in a stock's illiquidity both before and after the NTS Reform. Regarding the negative relation between illiquidity and stock returns, we find that stock returns show no clear relation with illiquidity after controlling for idiosyncratic volatility biases. Furthermore, we use residual approach to eliminate the effect of idiosyncratic volatility, and find there exists a positive relation between illiquidity and stock returns after the NTS Reform.

Keywords: NTS reform, illiquidity, idiosyncratic volatility, Shanghai A' Share stock market

Introduction

Market microstructure models have shown that liquidity is one of the most important market fractions that influence asset prices (e.g., Easley & O'Hara, 1987). In a seminal paper, Amihud and Mendelson (1986) show a positive relation between illiquidity and stock returns by using the bid-ask spread for illiquidity proxy. Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998), Datar, Naik, and Radcliffe (1998), and Amihud (2002) all show that stocks with low liquidity acquire higher expected returns as compared to stocks with high liquidity. Pastor and Stambaugh (2003), Acharya and Pederson (2005), Chordia, Sarkar, and Subrahmanyam (2005), Liu (2006) have shown that illiquidity is important for explaining asset returns and should be priced by the market.

However, fewer studies attempt to simultaneously consider the effect of liquidity and idiosyncratic volatility on expected stock returns into account. O'Hara (2003) develops an asymmetric information asset pricing model that incorporates the transactions costs of liquidity as well as the risks of price discovery. In particular, Stoll (1978) and Ho and Stoll (1980; 1981) propose an inventory model that dealer provides

Acknowledgements: This work is financially supported by Ishii Memorial Securities Research Promotion Foundation and the Scientific Research Project of Xi'an International Studies University (No. 16XWC02). Also this work is the Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 18JK0641). The author is greatly thankful to the Nippon Finance Association and Eleventh Annual Conference of the Asia-Pacific Economic Association for their helpful comments and suggestions.

Chen Yang, doctor of Economics, School of Japanese Studies, Xi'an International Studies University, Xi'an, China.
Correspondence concerning this article should be addressed to Chen Yang, No. 1, South Wenyuan Road, Guodu Educational Sci-tech Industrial Zone, Xi'an 710128, China.

immediacy by investor trading to market who is faced with risks on his inventory, which is partly due to the de-diversification caused by dealing in few stocks (idiosyncratic volatility risk). Dealers are unwillingness to provide immediacy (liquidity) when they are faced with higher return uncertainty on their inventory. With information asymmetry, an increase in volatility of underlying security returns implies higher uncertainty about the expected value of the security, which leading a higher adverse selection cost faced by liquidity provider, due to the increased possibility of trading with informed traders. This in turn leads to lower liquidity due to higher transaction costs and higher volatility because of higher bid-ask bounce (see for Easley, Kiefer, & O'Hara, 1996; O'Hara, 2003). Other studies like Easley and O'Hara (2010) develop a model in which illiquidity arises from uncertainty during financial crisis; Brunnermeier and Pedersen (2008) link the asset's market liquidity and traders' funding liquidity, and imply that market liquidity is related to volatility.[1]

Among empirical studies, Spiegel and Wang (2005) investigate the interaction relation between the two factors—idiosyncratic volatility and liquidity and find that the explanatory power of idiosyncratic volatility is strong and even could eliminates the liquidity effect on US stock market. Bao, Pang, and Wang (2011), Kalimipalli and Nayak (2012), and Kalimipalli, Nayak, and Perez (2013) examine the jointly effect of the two factors on bond market and find that both idiosyncratic volatility and liquidity remain their significant influence on bond yields. Han and Lesmond (2011) show that controlling for liquidity bias in the estimated idiosyncratic volatility eliminates the pricing ability of idiosyncratic volatility (e.g., Ang, Hodrick, Xing, & Zhang, 2006; 2009; Bali & Cakici, 2008; Bali, Cakici, & Whitelaw, 2011; Fu, 2009; Huang, Liu, Rhee, & Zhang, 2009; Khovansky & Zhylyevskyy, 2013; Hou & Loh, 2016)[2]. Motivated by these studies, we examine the cross-sectional correlations between illiquidity and stock returns by considering the idiosyncratic volatility bias in the Chinese stock market. As for our knowledge, there has no article to study the effect of liquidity on stock returns with considering idiosyncratic volatility.

Chinese stock market has developed such remarkably that catches a great deal of attention from the world in recent years. Different from developed markets like US, Europe countries, Japan, we conjecture that Chinese stock market may show some typical features related to Chinese investment and Chinese behavior due to the numerous constraints only in Chinese market[3]. The most represent feature is the Non-Tradable Share Reform (April 2005, hence for NTS Reform), which aimed at overcoming split share structures by converting non-tradable shares into tradable shares. This reform directly improves the market liquidity and influences volatility. Thus it is very necessary as well as important to study whether NTS Reform affects illiquidity, idiosyncratic volatility, or both. Therefore the purpose of this paper is to investigate correlations between illiquidity and stock returns by considering the idiosyncratic volatility bias under the NTS Reform in the Chinese stock market.

The results are as follows. Firstly, differing from prior studies, stock returns are decreasing in a stock's illiquidity (illiq_zero) both before the NTS Reform and after the NTS Reform. This finding means that investors require higher compensation on liquid stocks than illiquid stocks, which is inconsistent with illiquidity premium (e.g., Amihud & Mendelson, 1986; Amihud, 2002; Pastor & Stambaugh, 2003). Thus we call there exists illiquidity puzzle in the Chinese stock market.

---

[1] Vayanos and Wang (2011) show that the positive relationship between expected returns and idiosyncratic volatility might be partly due to illiquidity.

[2] Ang et al. (2006; 2009) examine the pricing ability of idiosyncratic volatility and find that stocks with high idiosyncratic volatility have earn lower average returns.

[3] Other features such as short-sales constraints are remained now, which also arises the transaction costs and information asymmetry, leading low liquidity and high volatility.

To seek for the negative correlation between illiquidity and stock returns, we consider the impact of idiosyncratic volatility on illiquidity as well as stock returns by using bivariate sorts portfolio analysis. To this end, we firstly sort stocks into quintiles by idiosyncratic volatility, and then within each idiosyncratic volatility quintile portfolios, we further sort stocks into quintiles based on illiq_zero, and this produces rebalance every month. After considering the idiosyncratic volatility basis, stock returns show no clearly relations to illiquidity, especially after the NTS Reform. This result is similar with Spiegel and Wang (2005), who also show that idiosyncratic volatility reduces the pricing ability of liquidity.

Following Fama and French (2008), we use residual illiq_zero by orthogonalizing the idiosyncratic volatility from the illiq_zero measure. And then we sort stocks into quintiles by the residual illiq_zero and find that positive difference for arbitrage portfolio (lowest illiq_zero quintile—highest illiq_zero quintile) but insignificant before the NTS Reform, while there exists a positive relation between illiq_zero and stock returns after the NTS Reform.

Prior studies on Chinese stock market are all about the influence of liquidity on stock returns or idiosyncratic volatility on stock returns[4]. For example, Su and Mai (2004), Wu and Song (2007), Zhang, Tian, and Wirjanto (2009), Nayan and Zheng (2010), i.e., examine the relation between liquidity and stock returns by using turnover ratio as liquidity measure, and find a positive relation between illiquidity and stock returns. With the illiquidity measure of Amihud (2002), Li and Wu (2003) provide evidence that supports a negative relations between illiquidity and stock returns; whereas Nayan and Zheng (2010) show aggregate illiquidity is a priced risk factor, which is positively related to stock returns (see also for Chen, Tu, & Lin, 2007). Other related studies like Jin and Yang (2002) explore the effects of stock price, trading volume and volatility on market liquidity, and find that factors such as trading volume, stock price and volatility of return can give significant explanation to different liquidity level.

Comparing with prior studies on Chinese stock market, this paper makes three efforts. It is the first attempt to study the jointly effect of liquidity and idiosyncratic volatility on stock returns with using a new illiquidity measure to capture the liquidity of Chinese stock market. Second, we consider the influences of the NTS Reform to examine whether the NTS Reform affects the relations between liquidity and idiosyncratic volatility or liquidity and returns. Third, we propose a residual approach to eliminate the impact of idiosyncratic volatility on liquidity.

The rest of this paper is organized as follows. Section 2 provides the data description, liquidity and idiosyncratic volatility measures used in this study. Section 3 presents the portfolio analyses for univariate sorts analysis and bivariate sorts analysis with considering the impact of idiosyncratic volatility. Section 4 presents illiq_zero residual approach to further explore the correlation between illiq_zero and stock returns. Section 5 provides concluding comments.

Data and Descriptive Statistics

We obtain daily and monthly stock returns, market returns and trading volumes from CSMAR (China Stock Market & Accounting Research) database over the period December 2000 through January 2012, which includes all common stocks traded on the Shanghai A' Share Stock Exchange. As discussed earlier, we focus

[4] Yang and Han (2009) document a negative relation between idiosyncratic volatility on stock returns, Chen et al. (2009) also obtain the similar results with Yang and Han (2009), while Deng and Zheng (2011) find that idiosyncratic volatility is positively related to stock returns.

on the A' Share market since the size and the trading value of the A' Share market account for 80 percent of the whole stock market. We also use annual accounting data for calculating the book-to-market ratio (B/M) and market value (MV) from the balance sheet of each firm. In addition, monthly risk-free rate is converted from the annual risk free rate based on compound interest calculation from CSMAR database. Here we exclude stocks whose trading days in a month is less than 10 days.

Following Fama and French (1992; 1993), we form size portfolios and book-to-market portfolios to calculate SMB and HML factor. In the end of each year *t* from 2001 to 2012, all common stocks are ranked based on market value and then split them into two groups, small and big (S and B). We also break all these common stocks into three book-to-market equity groups based on the breakpoints for the bottom 30% (Low), middle 40% (Medium), and top 30% (High) of the ranked values of B/M.

Then we construct six portfolios (S/L, S/M, S/H, B/L, B/M, B/H) from the intersections of the two MV and three B/M groups. For example, the S/L portfolio contains the stocks in the small MV group that are also in the low B/M group, and the B/H portfolio contains the big MV group that also have high B/M. Monthly value-weighted returns on the six portfolios are calculated each year. SMB is the difference between the returns on the small- (S/L, S/M, S/H) and big- (B/L, B/M, B/H) stock portfolios with about the same weighted average book-to-market equity. As the same way, HML means the difference between the returns on the low- (S/L, B/L) and high- (S/H, B/H) stock portfolios with about the same weighted average MV.

Estimation of Illiquidity and Idiosyncratic Volatility

Consider that there has a large percentage of non-trading days due to high trading costs in Chinese stock market, we propose a new liquidity measure, Illiq_zero—the revised version of Amihud (2002) for our illiquidity measure, which can be calculated as follows:

$$\text{Illiq\_}Zero_{i,t} = \left[\ln\left(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}\left|R_{i,d}\right|/VOLD_{i,d}\right)\right] + NT\%_{i,t} \tag{1}$$

where $N_{i,t}$ is the number of days on which stock *i* is traded in month *t*, $\left|R_{i,d}\right|$ is the absolute value of returns on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month. Therefore, if the stock return on day d is not zero, the new illiquidity measure is the logarithm of the Amihud illiquidity measure, whereas if the stock return on day d is zero, the new illiquidity measure will be the same as that of Lesmond, Ogden, and Trzcinka (1999)[5]. This new illiquidity measure captures the price reaction to trading volume as well as the trading cost. Therefore, higher the new illiquidity measure, lower the stock liquidity.

Following Ang et al. (2006, 2009), we define idiosyncratic volatility as the standard deviation of the residuals from the Fama and French (1993) model. In each month, daily excess returns of individual stocks are regressed on the daily Fama-French three factors: the excess return on a broad market portfolio ($R_m - r_f$), SMB and HML factors,

$$R_{id} - r_{f\mathrm{d}} = \alpha_{it} + \beta_{it}\left(R_{md} - r_{fd}\right) + s_{it}SMB_d + h_{it}HML_d + \varepsilon_{id} \tag{2}$$

where d is the subscript for the day and t is the subscript for the month, $d \in t$, and β_i, s_i, h_i are factor sensitivities or loadings. We run a time-series regression for each stock in each month. The idiosyncratic

[5] Lesmond et al. (1999) propose an illiquidity measure, by using the numbers of zero return days to numbers of trading days over some intervals, to capture the trading costs dimension of liquidity.

volatility of a stock is computed as the standard deviation of the regression residuals.

Figure 1 plots the time-varying illiq_zero measure and idiosyncratic volatility during the period January 2001 to December 2012. Overall, it seems to have a negative relation between illiq_zero measure and idiosyncratic volatility since when the illiq_zero is high, idiosyncratic volatility of the whole market comes to be low. On one hand, illiq_zero measure has a high value before the year of 2006, and sharply decreases during the period 2006 through 2012 due to NTSR. On the other hand, idiosyncratic volatility varies dramatically during the increasing period of Chinese stock market (2001-2003) and financial crisis (2007-2008) while displays flat in the rest period.

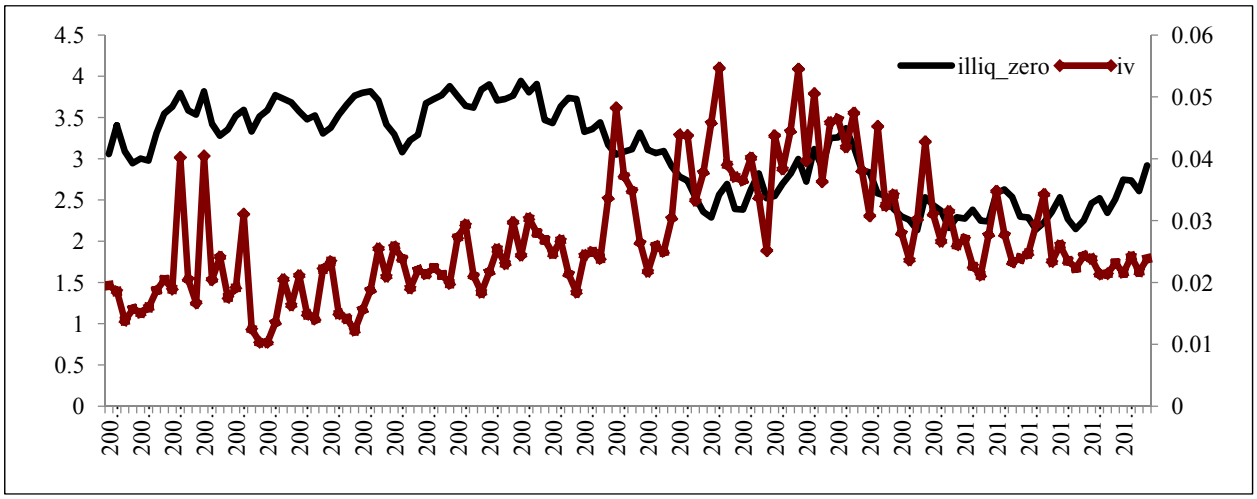

Figure 1. Illiq_zero measure and idiosyncratic volatility. *Notes.* The sample includes stocks from Shanghai A' Share market through 2001 to 2012. The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,m}$ is the number of trading volume days of stock *i* in month *t*, $|R_{i,d}|$ is the absolute return on stock *i* on day *d*, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock *i* on day *d*. $NT\%_{i,t}$ is the percentage of zero-return days within a month. Idiosyncratic volatility is defined as the standard deviation of the residuals from the Fama-French (1993) model.

Summary Statistics

Table 1 shows summary statistics for main variables examined here. Fama-French three factors, for example, SMB have a mean of -0.0073 while the mean value of HML is -0.0106. As also shown in Figure 1, illiq_zero has a mean of 2.0197, varying from -3.7793 to 12.1131. Idiosyncratic volatility varies from 0.0001 to 5.6571 and receives a mean of 0.0294.

Correlations for these variables are provided in Table 2. Firstly, the relation between individual stock returns and SMB factor is positively 0.3073, suggesting that small stocks have larger returns than big stocks, while the correlation for HML factor and stock returns shows negative. The correlation between illiq_zero and stock returns is -0.1888, and for illiq_zero and market returns is -0.1900, which is contrast with prior studies. Studies such as Amihud and Mendelson (1986), Amihud (2002) document that stocks with lower liquidity should be compensated by higher returns. Moreover, illiq_zero and idiosyncratic volatility are negatively correlated with a value of -0.0795.

To precisely observe these relations, we divide our sample period into two intervals—before NTSR and after NTSR—since NTSR may have influence on our variables. Figure 2 plots the cross-sectional relationship

between illiq_zero and stock returns before the NTS Reform and after the NTS Reform. Obviously, there has a negative relationship between illiq_zero and stock returns before the NTS Reform, meaning that higher the liquidity, larger the stock returns. One possible interpretation is that individual investors dominate tradable share market and they are willing to hold more liquid stocks for seeking short-period profits. After the NTS Reform, we could not obtain any clearly relationship between illiq_zero and stock returns. Figure 3 shows the cross-sectional relations between illiq_zero and idiosyncratic volatility before the NTS Reform and after the NTS Reform. There exists a clearly positive relationship between illiq_zero and idiosyncratic volatility after the NTS Reform, while it is negative before the NTS Reform.

Table 1

Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
R	78,162	0.0130	0.1430	-0.7703	1.8945
Size	78,527	14.4319	1.3289	11.4431	21.3179
B/M	78,497	7.4670	0.9731	3.1713	12.7957
Rm	132	0.0070	0.0896	-0.2693	0.2925
SMB	132	-0.0073	0.0479	-0.1504	0.0984
HML	132	-0.0106	0.0302	-0.1024	0.0637
Illiq_zero	78162	2.0197	1.5436	-3.7793	12.1131
Iv	78162	0.0294	0.0438	0.0001	5.6571

Notes. The sample includes stocks from Shanghai A' Share market between Jan 2001 to Dec 2012. Size is the market capitalization of all common stock, B/M is the observation of book-to-market ratio of all common stocks. Rm is the return of market portfolios, SMB and HML are the returns of mimicking portfolios related to market value and book-to-market ratio, respectively. Illiq_zero is the illiquidity of all common stocks, which is calculated as follows. $\text{Illiq\_}Zero_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}| / VOLDi,d) + NT\%i,t\right.$, where Ni,t is the number of trading volume days of stock i in month t, Ri,d is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month. And Iv is the idiosyncratic volatility of all common stocks, which is defined as the standard deviation of the residuals from the Fama-French (1993) model.

Table 2

Cross-sectional Correlations

	R	Rm	SMB	HML	Illiq_zero	Iv
R	1.0000					
Rm	0.6740	1.0000				
SMB	0.3073	0.1453	1.0000			
HML	-0.1104	0.0512	-0.4239	1.0000		
Illiq_zero	-0.1888	-0.1900	-0.1221	0.1209	1.0000	
Iv	0.0277	0.0097	0.0360	-0.0470	-0.0795	1.0000

Notes. The sample includes stocks from Shanghai A' Share market between Jan 2001 to Dec 2012. Size is the market capitalization of all common stock, B/M is the observation of book-to-market ratio of all common stocks. Rm is the return of market portfolios, SMB and HML are the returns of mimicking portfolios related to market value and book-to-market ratio, respectively. Illiq_zero is the illiquidity of all common stocks, which is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}| / VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,t}$ is the number of trading volume days of stock i in month t, $|R_{i,d}|$ is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month. Iv is the idiosyncratic volatility of all common stocks, which is defined as the standard deviation of the residuals from the Fama-French (1993) model.

We sort stocks into quintiles by illiq_zero, form equally weighted illiq_zero quintile portfolios. The summary statistics of these illiq_zero quintile portfolios are shown in Table 3. The Amihud measure, which is defined by the price impact to trading volume, increased monotonically across all illiq_zero quintiles, indicating our new measure, illiq_zero is highly correlated with Amihud (2002). The idiosyncratic volatility are also shown to decrease (except for illiq_zero quintile 1), which is consistent with Table 2, though it is not obvious. Also with the increasing of illiq_zero, size and book-to-market value for illiq_zero quintiles reduce monotonically.

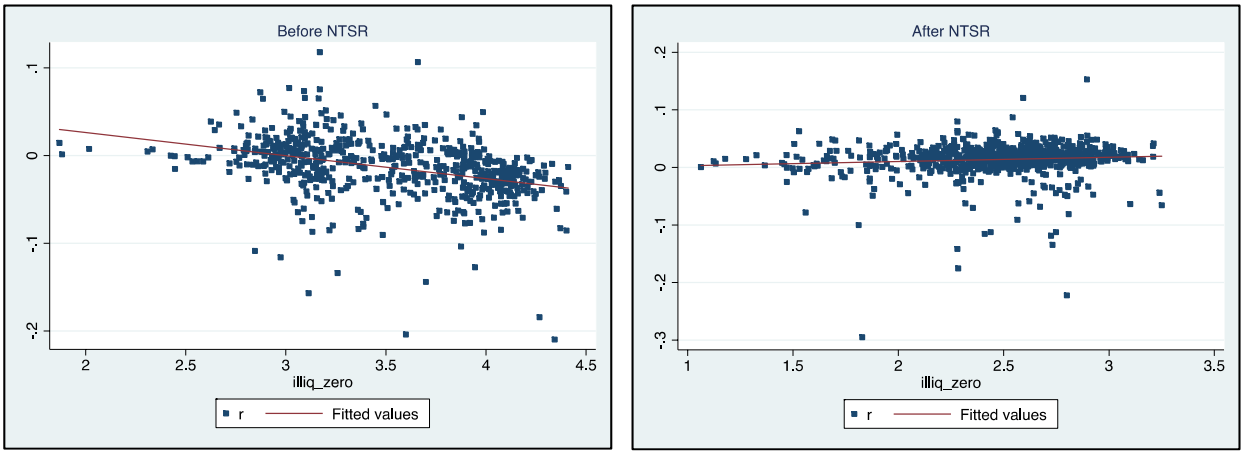

Figure 2. Cross-sectional relations between illiq_zero and return. *Notes.* The sample includes stocks from Shanghai A' Share market within two periods—before NTS Reform (Jan 2001 to Apr 2005) and after NTS Reform (Oct 2008 to Dec 2012). r represents the returns of each firm. The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,t}$ is the number of trading volume days of stock i in month t, $|R_{i,d}|$ is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month.

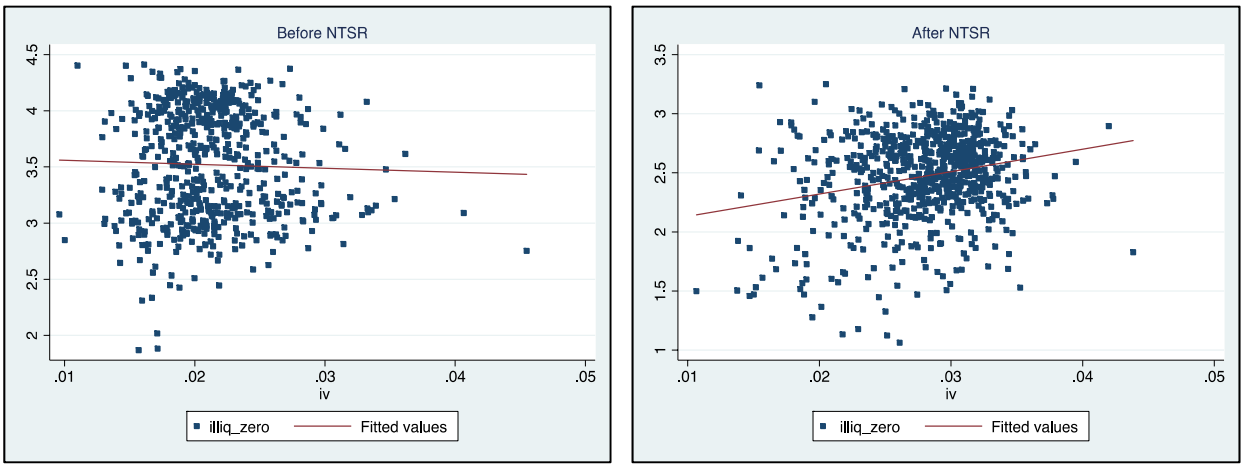

Figure 3. Cross-sectional relations between illiq_zero and idiosyncratic volatility. *Notes.* The sample includes stocks from Shanghai A' Share market within two periods—before the NTS Reform (Jan 2001 to Apr 2005) and after the NTS Reform (Oct 2008 to Dec 2012). The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLDi,d)+NT\%i,t\right.$, where Ni,t is the number of trading volume days of stock i in month t, Ri,d is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month. Idiosyncratic volatility is defined as the standard deviation of the residuals from the Fama-French (1993) model.

Table 3

Correlations of Various Variables on Portfolios Sorted by Illiq_zero

	Illiq_zero	Amihud	Iv	Lnsize	lnbk
1 (Low)	2.355	0.044	0.0280	15.591	16.902
	(0.58)	(0.05)	(0.01)	(1.45)	(1.96)
2	2.821	0.112	0.0282	14.656	15.674
	(0.54)	(0.12)	(0.01)	(1.09)	(1.15)
3	3.057	0.194	0.0280	14.289	15.331
	(0.56)	(0.22)	(0.01)	(1.03)	(1.02)
4	3.268	0.305	0.0275	13.991	15.095
	(0.57)	(0.36)	(0.01)	(1.01)	(0.99)
5 (High)	3.580	0.541	0.0272	13.674	14.835
	(0.62)	(0.79)	(0.01)	(0.98)	(0.98)

Notes. The sample includes stocks from Shanghai A' Share market within two periods—before the NTS Reform (Jan 2001 to Apr 2005) and after the NTS Reform (Oct 2008 to Dec 2012). This table presents the means and standard deviation (in parentheses) for quintile sorts by illiq_zero. The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,t}$ is the number of trading volume days of stock i in month t, $|R_{i,d}|$ is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month. Amihud is the illiquidity measure of Amihud (2002), defined as the price impact caused by trading volume. Idiosyncratic volatility is defined as the standard deviation of the residuals from the Fama-French (1993) model. Lnsize and lnbk are the logarithm of size and book value.

Univariate and Bivariate Portfolio Analysis

In this section, we form portfolios based on one- and two-way sorts to explore the cross-sectional relations between illiquidity and stock returns both before and after the NTS Reform. As numerous studies have document that illiquidity is highly related to idiosyncratic volatility (e.g. Ho & Stoll, 1978; Brunnermeier & Pederson, 2005), we consider the impact of idiosyncratic volatility in two-way sorts portfolio analysis.

Univariate Sorted Portfolio Returns

We sort stocks into quintiles based on their monthly estimates of illiq_zero, form equally weighted quintile-sorted portfolios, and then difference the lowest and highest illiq_zero quintiles. Specifically, each month, stocks are sorted into five quintiles based on the illiq_zero and then rebalanced. We then regress the quintile portfolio returns against either the CAPM model or the three-factor (Fama-French) model to estimate the CAPM or the Fama-French alpha, respectively. We finally compare the performance between the portfolio with the lowest illiq_zero (1 Low) and the portfolio with highest illiq_zero (5 High). The difference is the abnormal return one would earn on a zero-cost (arbitrage) portfolio formed by taking a long position in the lowest-quintile portfolio and taking a short position in the highest-quintile portfolio (P1-P5).

The results are presented in Table 4, with Panel A focusing on the results before the NTS Reform and Panel B focusing on results after the NTS Reform. As shown in Panel A of Table 4, the raw return on quintile 1 (lowest illiq_zero) is 0.692% on quantile 5 (highest illiq_zero) is -2.669% per month. This results a difference of 3.361% in the arbitrage portfolio, with a *t*-statistic of 19.00. And the decreasing trend performs monotonic, which is supportive of Figure 2, meaning higher the liquidity, larger the stock returns. After controlling for CAPM or the Fama-French model, the adjusted returns remain decreasing with illiq_zero. For example, the Fama-French alpha of the arbitrage portfolio (P1 = P5) has earned a significantly positive return of 1.708% (*t* =

2.93). These results suggest that including the market return, size and value factor do not sufficiently control for the effect of illiq_zero on stock returns.

Table 4

Returns on Portfolios Sorted by Illiq_zero

Panel A: Before NTS Reform						
	1 (Low)	2	3	4	5 (High)	P1-P5
r	0.692	-0.887	-1.762	-2.205	-2.669	3.361
	5.26	-6.75	-13.83	-18.51	-22.54	19.00
CAPM alpha	1.881	0.506	-0.301	-0.719	-1.253	3.134
	16.70	4.85	-3.15	-8.76	-14.71	12.60
FF-3 alpha	1.378	0.507	0.252	0.047	-0.33	1.708
	10.61	4.18	2.31	0.51	-3.53	2.93
Panel B : After NTS Reform						
	1 (Low)	2	3	4	5 (High)	P1-P5
r	2.135	2.089	2.045	1.564	1.270	0.865
	10.49	10.44	10.32	8.09	6.85	3.14
CAPM alpha	1.232	1.221	1.185	0.682	0.394	0.838
	8.11	7.87	7.76	4.74	2.98	2.68
FF-3 alpha	0.799	0.471	0.409	0.228	-0.11	0.909
	4.77	2.89	2.61	1.55	-0.83	0.71

Notes. The sample includes stocks from Shanghai A' Share market within two periods—before NTS Reform (Jan 2001 to Apr 2005) and after NTS Reform (Oct 2008 to Dec 2012). Stocks are ranked into quintiles on the basis of their illiq_zero every month. Panel A and B present the monthly raw returns, adjusted returns (alpha) relative to CAPM, as well as adjusted returns (FF-3 alpha) relative to Fama-French 3 factor models before and after NTSR, respectively. The column (P1-P5) reports the return differences between low illiq_zero portfolio and high illiq_zero portfolio. The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,t}$ is the number of trading volume days of stock i in month t, $|R_{i,d}|$ is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month.

In Panel B of Table 4 (After the NTS Reform), which the raw return as well as the CAPM and Fama-French alpha show similar trend with Panel A. The Fama-French alpha declines from 2.135% to 1.270% with the increasing of illiq_zero, which yields a positive alpha of 0.909% per month.

Notable in the results is the monotonically decreasing trend of stock returns from the lowest-ranked illiq_zero quintile to the highest-ranked illiq_zero quintile. This means investors would earn a positive return on arbitrage portfolio formed by taking long positions in lowest illiq_zero quintile and taking short positions in the highest illiq_zero quintile portfolio. Therefore we call there exists an illiquidity puzzle in Chinese stock market.

Bivariate Sorted Portfolio Returns

We then form two-way sorted portfolios to simultaneously control for the impact of idiosyncratic volatility while examining the cross-sectional effect of illiquidity on stock returns. Each month, we sort stocks into quintiles by idiosyncratic volatility firstly, and then we form portfolios by illiq_zero into quintile for each idiosyncratic volatility quintile. This formation would yield 25 idiosyncratic volatility-illiq_zero portfolios. Table 5 reports the Fama-French alpha both before (Panel A) and after (Panel B) the NTS Reform. The column (P1-P5) shows the difference on portfolio formed by taking long positions in lowest illiq_zero quintile and

taking short positions in highest illiq_zero quintile. The last row—"Control for Iv" shows the Fama-French alpha across illiq_zero quintiles for a portfolio that is equally weighted across idiosyncratic volatility quintiles.

Table 5

Portfolios Sorted by Illiq_zero After Controlling for Idiosyncratic Volatility, FF 3 Factor Alphas

Panel A : Before NTS Reform						
	Illiq_zero 1 (Low)	2	3	4	5 (High)	P1-P5
Iv 1 (Low)	0.604	0.223	-0.126	-0.260	-0.940	1.544
	3.27	1.16	-0.72	-1.58	-6.25	5.65
2	0.720	0.542	-0.238	-0.387	-0.457	1.177
	3.36	2.53	-1.30	-2.33	-3.15	2.54
3	0.781	0.221	-0.175	-0.491	-0.391	1.172
	3.04	0.96	-0.91	-2.96	-2.38	1.85
4	0.569	0.177	0.461	-0.267	-0.190	0.759
	2.02	0.69	2.01	-1.29	-0.96	0.80
Iv 5 (High)	3.415	2.036	1.248	1.303	0.934	2.481
	8.38	5.54	3.53	4.08	2.92	2.59
Control for Iv	1.219	0.642	0.238	-0.018	-0.209	1.428
	9.47	5.36	2.18	-0.18	-2.19	1.87
Panel B : After NTS Reform						
	Illiq_zero 1 (Low)	2	3	4	5 (High)	P1-P5
Iv 1(Low)	-0.968	-1.495	-1.514	-1.588	-1.478	0.510
	-4.72	-6.53	-6.45	-6.91	-7.77	7.53
2	-1.382	-1.312	-0.892	-1.726	-1.932	0.550
	-4.92	-4.86	-3.58	-7.77	-9.18	7.88
3	-0.592	-1.258	-1.075	-1.433	-1.597	1.005
	-1.87	-4.17	-3.72	-5.43	-7.09	5.88
4	1.108	0.248	-0.076	-0.067	0.176	0.932
	3.19	0.73	-0.23	-0.22	0.62	0.56
Iv 5 (High)	6.523	5.469	5.421	5.731	4.660	1.863
	13.19	11.15	12.16	12.11	10.87	10.18
Control for Iv	0.941	0.335	0.376	0.188	-0.027	0.968
	5.63	2.05	2.46	1.24	-0.20	0.18

Notes. The sample includes stocks from Shanghai A' Share market within two periods—before NTSR (Jan 2001 to Apr 2005) and after NTSR (Oct 2008 to Dec 2012). In each month, stocks are ranked into quintiles on the basis of their idiosyncratic volatility firstly, and then sorted (into quintiles) by their illiq_zero in each idiosyncratic volatility quintile. Panel A and B present the monthly adjusted returns (alpha) relative to Fama-French 3 factor models before and after NTSR, respectively. The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,t}$ is the number of trading volume days of stock i in month t, $|R_{i,d}|$ is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month. And idiosyncratic volatility is defined as the standard deviation of the residuals from the Fama-French (1993) model.

Before the NTS Reform, the decreasing trend is also matched in each idiosyncratic volatility quintile (Iv 1~Iv 5). That is lowest illiq_zero quintile 1 which has earned largest returns while highest illiq_zero quintile 5 has earned lowest returns. The Fama-French alphas on arbitrage portfolios (P1-P5) are all positive, but not all of these results are significant. For instance, holding for idiosyncratic volatility quintile 4, the Fama-French alphas are decreasing from 0.569% to -0.190%, which yields a positive value of 0.759%, while the *t*-statistic is

0.80. Furthermore, in the "Control for Iv" row, the Fama-French alpha shows a monotonically decreasing trend and the result on (P1-P5) is positively 1.428%, but insignificant. Thus idiosyncratic volatility seems to have some impact on the relation between illiq_zero and stock returns.

We now turn to the results after the NTS Reform. According to each idiosyncratic volatility quintile, the Fama-French alphas in arbitrage portfolios (P1-P5) are all positive, but the decreasing trend is not all monotonically across each idiosyncratic volatility quintiles. For idiosyncratic volatility quintile 2, the Fama-French alpha in illiq_zero quintile 3 is largest, -0.892%, while the result in illiq_zero quintile 1 is -1.382%. After controlling for idiosyncratic volatility (the row "Control for Iv"), illiq_zero quintile 1 has earned largest value of 0.941%, then illiq_zero quintile 3 is 0.376%, illiq_zero quintile 2 is 0.335%, illiq_zero quintile 4 is 0.188%, illiq_zero quintile 5 has earned smallest value of -0.027%. However, only the result of illiq_zero quintile 1 is significant under 1% statistic level. Finally, the *t*-statistic on the arbitrage portfolio (P1-P5) is only 0.18. Considering the impact of idiosyncratic volatility, the negative relation between illiq_zero and stock returns is eliminated.

In sum, there exists illiquidity puzzle in Chinese stock market both before and after the NTS Reform. However, the puzzle could be partly eliminated by controlling for the impact of idiosyncratic volatility, especially after the NTS Reform. We provide the following approach to further explore the illiquidity puzzle with the effect of idiosyncratic volatility.

Illiq_Zero Regression Residual Approach

The bivariate sorts portfolio analysis makes it clear that the correlation between illiq_zero and stock returns may not provide a conclusive test without considering the impact of idiosyncratic volatility. As inventory control model implies, sorting on illiq_zero could be in effect sorting on the idiosyncratic volatility. Thus, in effort to purge the idiosyncratic volatility effects from the illiq_zero, we provide a residual illiq_zero as a sorting variable (e.g. Chen et al., 2002; Fama & French, 2008). The residual illiq_zero is estimated by regressing the illiq_zero on idiosyncratic volatility and then obtains the residual of this regression. This approach will orthogonalize the idiosyncratic volatility from the illiq_zero measure by using Fama-MacBeth setting. Therefore we could isolate the impact of idiosyncratic volatility to examine the correlation between illiq_zero and stock returns.

The Relation Between Illiq_Zero and Idiosyncratic Volatility

We present in Table 6 a series of Fama-MacBeth (1973) regressions. We implement the Fama-MacBeth regressions as follows. Each month, we run a separate cross-sectional regression before as well as after the NTS Reform. We repeat such cross-sectional regressions for 52 months before NTSR and 39 months after NTS Reform, this gives us a total of 10,475 regressions before NTS Reform and 26,089 regressions after NTS Reform. Table 6 reports the mean coefficients across these regressions, along with the associated *t*-statistics and the adjusted *R*-squared values.

Before the NTS Reform, the idiosyncratic volatility is negatively related to illiq_zero, which is different with prior studies. In contrast, the correlation between idiosyncratic volatility and illiq_zero turns to positive after the NTS Reform. As we noted above, the composition on tradable share market has changed through the NTS Reform, and a large population of institutional investors or legal persons enters to the market. These changes directly influence the liquidity and volatility in the Chinese stock market and cause market to be more

efficient and close to developed markets (i.e. US stock market). Thus it is reasonable to explain the strong positive relation between illiq_zero and idiosyncratic volatility after the NTS Reform.

Table 6

Illiq_zero and Idiosyncratic Volatility

Dependent variable: Illiq_zero						
	Coef.		Std. Err.	*t*-value	No. of Obs	Adj. R^2
Before NTS Reform						
Iv	-4.886	***	1.68	-2.91		
_cons	3.549	***	0.05	78.18	10,475	1.13%
After NTS Reform						
Iv	5.760	***	0.23	24.98		
_cons	2.312	***	0.01	337.56	26,089	2.21%

Notes. The sample includes stocks from Shanghai A' Share market within two periods—before NTS Reform (Jan 2001 to Apr 2005) and after NTS Reform (Oct 2008 to Dec 2012). The table presents Fama-MacBeth regression results of the illiq_zero, estimated using $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}|/VOLD_{i,d})\right] + NT\%_{i,t}$, where $N_{i,t}$ is the number of trading volume days of stock *i* in month *t*, $|R_{i,d}|$ is the absolute return on stock *i* on day *d*, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock *i* on day *d*. $NT\%_{i,t}$ is the percentage of zero-return days within a month on the influence of idiosyncratic volatility, which is defined as the standard deviation of the residuals from the Fama-French (1993) model. Significance at the 1% level is given by ***.

Residual Illiq_zero Sorted Portfolio Analysis

Each month, we sort stocks into quintiles by the estimated residual illiq_zero, form equally weighted quintile portfolios, and compare the performance of the quintile portfolios and the (P1-P5) arbitrage portfolio. Table 7 reports the cross-sectional raw returns, CAPM-α, Fama-French-α on residual illiq_zero quintile portfolios before the NTS Reform (Panel A) and after the NTS Reform (Panel B).

As shown in Panel A of Table 7 before the NTS Reform, the raw returns of the residual illiq_zero quintile portfolios perform a decreasing trend, which is similar with the results of illiq_zero quintiles. And the arbitrage portfolio (P1-P5) has earned a return of 0.377% per month, with a *t*-statistic of 1.91. However, among the Fama-French alphas of the residual illiq_zero quintiles, only the residual illiq_zero quintile 1 is significant. The lack of significance in abnormal performance extends to the (P1-P5) extreme quintile difference (from the case of 1.888% shown in Panel A of Table 7).

Conversely, after the NTS Reform in Panel B, the results are remarkable. The negative correlations between illiq_zero and stock returns are converted, indicating that stock returns are increasing with illiq_zero. Illiq_zero quintile 1 earns lowest raw return of -0.770%, then illiq_zero quintile 2 is -0.223%, illiq_zero quintile 3 is 0.294, illiq_zero quintile 4 is 2.065% and illiq_zero quintile 5 earns largest raw return of 7.708%, and all are significant. Also the CAPM or Fama-French alpha of residual illiq_zero quintiles shows a monotonically increasing trend with residual illiq_zero. The raw return of the arbitrage portfolio (P1-P5) is significantly -8.478%, suggesting that taking long positions in the highest illiq_zero quintile and taking short positions in the lowest illiq_zero quintile would earn 8.478% per month. After adjusted by CAPM model or Fama-French three-factor model, the alpha is reduced to 8.084% (CAPM alpha) or 6.963% (Fama-French alpha), but still significant at the 1% level.

These results are indicative of the "illiquidity puzzle" after removing the impact of idiosyncratic volatility on illiq_zero. In particular, after the NTS Reform, investors require profits when they taking long positions in

illiquid stocks. The result also implies that the NTS Reform has improved the liquidity and volatility of tradable share market and further promotes the market efficiency.

Table 7

Returns on Portfolios Sorted by Residual Illiq_zero

Panel A: Before NTS Reform						
	1 (Low)	2	3	4	5 (High)	P1-P5
r	-0.694	-1.788	-1.818	-1.478	-1.071	0.377
	-3.94	-13.78	-15.76	-13.98	-12.02	1.91
CAPM alpha	1.148	-0.240	-0.433	-0.221	-0.152	1.300
	7.99	-2.55	-5.26	-2.94	-2.20	1.36
FF-3 alpha	1.780	0.168	-0.013	0.035	-0.108	1.888
	10.97	1.55	-0.14	0.41	-1.34	0.85
Panel B : After NTS Reform						
	1 (Low)	2	3	4	5 (High)	P1-P5
r	-0.770	-0.223	0.294	2.065	7.708	-8.478
	-5.93	-1.39	1.63	10.30	29.60	-29.07
CAPM alpha	-1.445	-1.038	-0.586	1.116	6.639	-8.084
	-15.47	-9.41	-4.69	7.91	33.35	-42.67
FF-3 alpha	-1.410	-1.452	-1.194	0.275	5.553	-6.963
	-14.12	-12.97	-9.44	1.9	26.47	-33.78

Notes. The sample includes stocks from Shanghai A' Share market within two periods—before NTS Reform (Jan 2001 to Apr 2005) and after NTS Reform (Oct 2008 to Dec 2012). Stocks are ranked into quintiles on the basis of their residual illiq_zero. Residual illiq_zero is estimated by regressing the illiq_zero estimate on idiosyncratic volatility. Panel A and B present the monthly raw returns, adjusted returns (alpha) relative to CAPM, as well as adjusted returns (FF-3 alpha) relative to Fama-French 3 factor models before and after NTSR, respectively. The illiq_zero is calculated as follows. $\text{Illiq\_Zero}_{i,t} = \left[\ln(\frac{1}{N_{i,t}}\sum_{t=1}^{N_{i,t}}|R_{i,d}| / VOLDi,d) + NT\%i,t\right.$, where Ni,t is the number of trading volume days of stock i in month t, Ri,d is the absolute return on stock i on day d, and $VOLD_{i,d}$ is the Chinese yuan trading volume of stock i on day d. $NT\%_{i,t}$ is the percentage of zero-return days within a month.

Conclusions

This paper investigates the cross-sectional correlations between illiquidity and stock returns, with considering the NTS Reform in Shanghai A' Share stock market. Our main findings are shown as follows. Firstly, differing from prior studies, stock returns are decreasing in a stock's illiquidity (illiq_zero) both before the NTS Reform and after the NTS Reform. This finding means that investors require higher compensation on liquid stocks than illiquid stocks. Thus we call there exists illiquidity puzzle in Shanghai A' share stock market. To seek for the negative correlation between illiquidity and stock returns, we then consider the impact of idiosyncratic volatility on illiquidity as well as stock returns by using bivariate sorts portfolio analysis. After considering the idiosyncratic volatility basis, stock returns show no clearly relations to illiquidity, especially after the NTS Reform. Lastly, we use residual approach to minimize the effect of idiosyncratic volatility, and find there exists a positive relation between illiq_zero and stock returns.

References

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics, 77*(2), 375-410.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets,* 5(1), 31-56.

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17*(2), 223-249.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance, 61*(1), 259-299.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics, 91*(1), 1-23.

Bali, T. G., & Cakici, N. (2008). Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis, 43*(1), 29-58.

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics, 99*(2), 427-446.

Bao, J., Pan, J., & Wang, J. (2011). The illiquidity of corporate bonds. *The Journal of Finance, 66*(3), 911-946.

Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics, 49*(3), 345-373.

Brennan, M. J., & Subrahmanyam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics, 41*(3), 441-464.

Chen, G., Tu, H., & Lin, H. (2007). Idiosyncratic volatility puzzle and explanations based on heterogeneous beliefs: Evidence from Chinese stock markets. Workingpaper, Xiamen University.

Chen, J., Hong, H., & Stein, C. J. (2002). Breadth of ownership and stock returns. *Journal of Financial Economics, 66*(2-3), 171-205.

Chen, Y. (2015). An empirical study of liquidity and autocorrelations on Chinese stock markets. *Asia-Pacific Financial Markets, 22*(3), 261-282.

Chordia, T., Sarkar, A., & Subrahmanyam, A. (2005). *The joint dynamics of liquidity, returns, and volatility across small and large firms.* Staff Report, Federal Reserve Bank of New York.

Datar, V. T., Naik, N. Y., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets, 1*(2), 203-219.

Deng, X. C., & Zheng, Z. L. (2011). Is there an idiosyncratic volatility puzzle exist in China's Equity Market. *Journal of Business Economics, 1*(1), 60-67.

Easley, D., Kiefer N. M., & O'Hara, M. (1996). Liquidity, information, and infrequently traded stocks. *The Journal of Finances, 51*(4), 1405-1436.

Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics, 19*(1), 69-90.

Easley, D., & O'Hara, M. (2010). Microstructure and ambiguity. *The Journal of Finances,65*(5), 1817-1846.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56.

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance, 63*(4), 1653-1678.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy, 81*(3), 607-636.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics, 91*(1), 24-37.

Han, Y., & Lesmond, D. (2011). Liquidity biases and the pricing of cross-sectional idiosyncratic volatility. *The Review of Financial Studies, 24*(5), 1590-1629.

Ho, T., & Stoll, H. R. (1980). On dealer markets under competition. *The Journal of Finance, 35*(2), 259-267.

Ho, T., & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics, 9*(1), 47-73.

Hou, K., & Loh, R. K. (2016). Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics, 121*(1), 167-194.

Huang, W., Liu, Q., Rhee, S. G., & Zhang, L. (2009). Return reversals, idiosyncratic risk, and expected returns. *The Review of Financial Studies, 23*(1), 147-168.

Jin, Y. H., & Yang, W. (2002). An empirical analysis of stock price, trade volume and volatility on market liquidity. *Journal of Financial Research, 264*(6), 12-21.

Kalimipalli, M., & Nayak, S. (2012). Idiosyncratic volatility vs. liquidity? Evidence from the US corporate bond market. *Journal of Financial Intermediation, 21*(2), 217-242.

Kalimipalli, M., Nayak, S., & Perez, M. F. (2013). Dynamic effects of idiosyncratic volatility and liquidity on corporate bond spreads. *Journal of Banking & Finance, 37*(8), 2969-2990.

Khovansky, S., & Zhylyevskyy, O. (2013). Impact of idiosyncratic volatility on stock returns: A cross-sectional study. *Journal of Banking & Finance, 37*(8), 3064-3075.

Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A new estimate of transaction costs. *The Review of Financial Studies, 12*(5), 1113-1141.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics, 82*(3), 631-671.

Narayan, P. K., & Zheng, X. (2010). Market liquidity risk factor and financial market anomalies: Evidence from the Chinese stock market. *Pacific-Basin Finance Journal, 18*(5), 509-520.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance, 58*(4), 1335-1354.

Pástor, Ľ., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy, 111*(3), 642-685.

Stoll, H. R. (1978). The pricing of security dealer services: An empirical study of NASDAQ stocks. *The Journal of Finance, 33*(4), 1153-1172.

Spiegel, M. I., & Wang, X. (2005). Cross-sectional variation in stock returns: Liquidity and idiosyncratic risk. Workingpaper, Yale ICF Working Paper No. 05-13, EFA Moscow Meetings Paper.

Su, D. W., & Mai, Y. X. (2004). Liquidity and asset pricing: An empirical exploration of turnover and expected returns on Chinese stock markets. *Economic Research Journal, 2*, 95-105.

Vayanos, D., & Wang, J. (2011). Liquidity and asset returns under asymmetric information and imperfect competition. *The Review of Financial Studies, 25*(5), 1339-1365.

Wu, Y. F., & Song, F. M. (2007). Liquidity risk and stock return. *Proceedings of the Operations Research and Management Science, 16*(2), 117-122.

Yang, H. W., & Han, L. Y. (2009). An empirical study of the relationship between the idiosyncratic volatility and cross-sectional return. *Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 1*(22), 6-10.

Zhang, F., Tian, Y., & Wirjanto, T. S. (2009). Liquidity risk and cross-sectional returns: Evidence from the chinese stock markets. *Finance Research Letters, 6*(4), 219-229.

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 472-481
doi: 10.17265/2328-7144/2018.06.006

Strategy for Safer Agricultural Robots

Eleftheria Mitka
Kimmeria Campus, Xanthi, Greece

Given the prediction that in 2050 agriculture will be doubled due to increasing earth population, the noticeable research question that arises is how do we incorporate safety in agricultural robots? Yet, due to the lack of standardization in this domain, the author will review well-established standards such as autonomous farm equipment, cab-less tractors, and forestry vehicles. Since these robots have come to reduce skilled labour, input costs, and environmental impact, this paper strives to emphasize the need of standardization and regulation framework that will step up the successful commercialization of heave farm machines. The upper goal is to establish a safety assessment improving the compliance with standards and taking account of the public trustworthiness in these emerging applications worldwide.

Keywords: precision agriculture, robots, standardization, safety, risk assessment

Introduction

The farming era represents a crucial part of society due to the attractiveness of the sector and the contribution to human survival. Agricultural field of robotics is both challenging and progressive segment. By 2050 the world's population is likely to increase to 9.6 billion, which will consequently increase by 70% the demand for fruit and vegetables, according to current estimations (Folley, 2011). Standardization and legal framework will lead to safe machine creation in protected horticulture, forestry, and parkland. Figure 1 explores the existing classes of robotics domain. Notwithstanding that precision agriculture is considered crucial and it is from the most hazardous industries. Agriculture involves highly automated machines that fulfil boring tasks such as de-weeding, seedbed preparation, scouting, fertilization, harvesting with fitted intelligence and associated perception relieving the operator from hard working conditions and musculoskeletal disorders (Figure 2). Harvester reduces spillage and users stress increasing maximum up-time, accuracy and productivity protecting farmer's investment. When it comes to crops, robot has embedded computers with high algorithms on decisional level in order to move around autonomously at the end of row, operating water treatment and waste management and to be monitored by laptop or portable tablet. Self-driving tractors are usually low cost integrated with navigation across crop row in farms while individually operating ploughing and irrigation, haying and chopping fast and robust. Recent technology requires real-time positioning data, on-board laser, and share wireless communication such as cooperation of robots intended for precise pesticide and herbicide (Emmi, Gonzalez-de Soto, Pajares, & Gonzalez-de Santos, 2014). Otherwise, when cellular communication is unavailable an optical modem can be used for satellite mode capable in centimetre accuracy. 3G Coverage data calculate the complete

Eleftheria Mitka, Ph.D. candidate/Deng, MEng, MBA, Department of Electrical & Computer Engineering, Kimmeria Campus, Xanthi, Greece.
Correspondence concerning this article should be addressed to Eleftheria Mitka, Department of Electrical & Computer Engineering, Kimmeria Campus, Building B, Office 2.13, Xanthi GR-67100, Greece.

field coverage by row-crop planting, harvesting, or seeding and enhance overlap control significantly. That reduces fuel costs and improves precise ground compaction. A display is embedded that alerts the farmer with manage robot health, shows multiple machines simultaneously, tractors real-time logistics information saving time and seed. Furthermore, the display monitors weather data like humidity, wind speed and direction, and temperature in order to advise your crew.

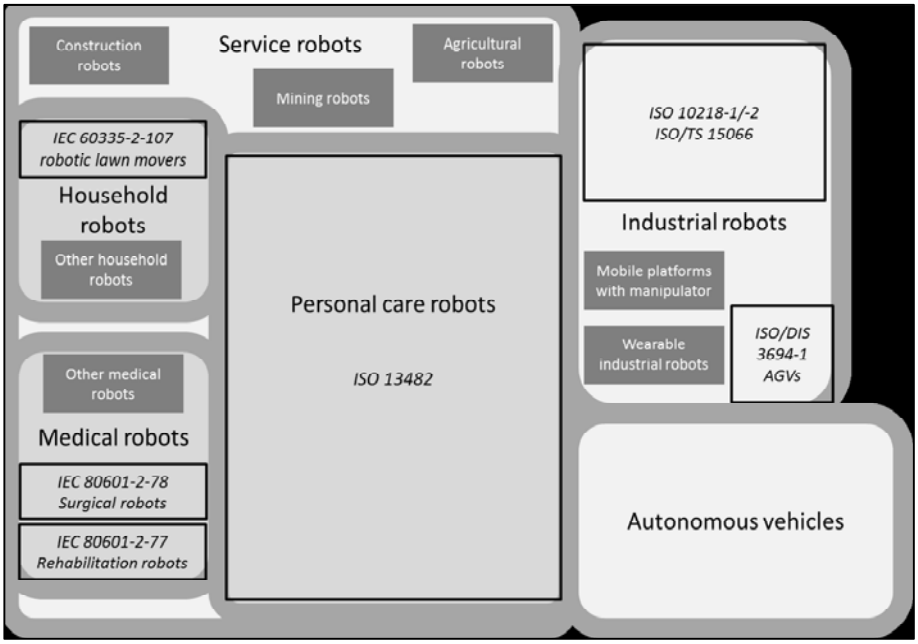

Figure 1. The subclass of agricultural robots as a part of service robotics in the whole community of robots according to ISO/TR 23482-2 draft (ISO/DTR 23482-2, n.d.).

Figure 2. Autonomous vehicle customized for seeding and hydroponic growing system. Source: www.utahbusiness.com.

In the next decades, self-propelled equipment with intelligent farming will cover the rarefaction of skilled labour and add value offering rootworm protection. Most of them incorporate the form of insect and crawl on mechanical legs or mowing wheels driving by two motors or solar panels in order to be eco-friendly. They intended to scan symmetry of leaves size, analyse soil compaction, get ready for cultivation with no or restricted human intervention and auto-steering even in low-visibility conditions. A lot of them are intended for identifying disease, weed, and insects by infrared while ultrasonic sensors arrays keep them on track assisted by telematics, image sensor, and location tracking.

Current Standards

The problem of safety in industrial robotics has been a matter of concern for years, in automotive industry, basically (Mitchell, 2010). However, the answer was to separate the operator for the robot cell with “keep out zones” excluding tasks like maintenance, training and troubleshooting that operator could enter the cell (Taubig, Frese, Hertzberg, Luth, Mohr, Vorobev, & Walter, 2012). In future farming systems, safety standards involve human-robot cooperation and coexistence so it cannot be applied the “principle of separating”. Yet, due to the lack of standardization in this domain, the author will survey the well-established standards such as autonomous vehicle ones. Moreover, ISO 13482 (ISO 13482, 2014) can assist the risk assessment, verification and is significant in overall functional safety. Another standard is ISO 25119-2 (ISO 25119-2, 2010) based on quality assurance of ISO 9001 (ISO 9001, 2015) which defines mitigation measures and requirements for operators and designers for tractors and field machines. Protective measures are described in ISO 61496 (ISO 61496, 2012), electro-sensitive protective equipment, defining how fault shall force the system to enter a safe-mode in vision of industrial settings and how multiple faults shall not influence the aforementioned action. Notwithstanding the fundamental ideas of standardization, the procedures for safety of machinery, as well as the ISO 10218-1 (ISO 10218-1, 2011) providing guidelines for robot-human interactions, should also be applied to agronomy for sustainable development.

All these high-level requirements must be achieved in an open-ended chaotic environment, in farms. IEC 61499 can prove very helpful in formal specification deriving implementation (IEC 61499-1, 2012). ISO/DIS 18497 (ISO/DIS 18497, 2014) describes different level of performance for agricultural equipment establishing the increasing level of hazard counting hardware, mean time to repair, fault diagnosis system and traditional faults. The loose connection in standardization towards hardware specifications can be linked with ISO 13849-1 (ISO 13849-1, 2015) and for ethernet communication architecture for tractors the ISO 11783 (ISO 11783, 2017). ISO 26262 is component-level in address human hazards and risks in electronics (ISO 26262, 2011). Finally, the software architecture for safety of agriculture machines can be assessed based on the IEC 61508 (IEC 61508-1, 2010) providing mathematics, diagrams, computed-oriented tools, and formal specification deriving implementation (Ingibergsson, Schultz, & Kraft, 2015). ISO 14121-1 (ISO 14121-1, 1999) defines the following types of risks: electrical, environmental, hazard by materials and chemical substances, by infrared radiation, by neglecting ergonomic principles in machine, mechanical, vibration-related, noise, radiation.

Risk Assessment

The flow chart represents the risk assessment for operators, maintenance personnel and apprentices that design and fabric robots safe considering their intended purpose by following this procedure (Figure 3). It also provides a safety strategy based on ISO 13849 (ISO 13849-1, 2015) defining the machine limits for all stakeholders and the type of misuse that leads to dangerous situations. This strategy sums up an iterative process that can be used by standardization’s experts and third parties involved, with design measures and, as well as, describing limitations of operators for constant supervision. This strategy involves:

- Risk assessment for each risk according to ISO 14121-1 (ISO 14121-1, 1999) or established standards for industrial robotics, passive safety, safety of electrical equipment, law framework relevant to farms and tractors.
- Determination of machine limits for the intended purpose.
- Identification of the hazards.

- Risk reduction by inherent measures or by guards.
- Risk estimation.

The quantitative risk estimation is validated counting severity of a harmful incident, possibility to avoid dangerous circumstances and probability of harm occurring. The output variable is risk and arranged within three levels: (1) Low, (2) Medium, and (3) High, as shown in Table 1, counts diverse levels of parameters.

Table 1

The Discrete Levels of Input Variables (ISO 25119, 2010)

Severity	Possibility	Probability	Output risk
S0-No harm, bruises, scratch	C0-Controllable/ operator always correct	E0: 0.0001 Improbable	1. Low
S1-Reversible harm	C1-Easily controllable/ some operators cannot correct	E1: 0.001 Rare events 1/year	2. Medium
S3-Irreversible harm	C2-Generally controllable/ average operators correct	E2: 0.01 Sometimes < 1% of operating time	3. High
S4-Fatality/permanent disability	C3-Fatal injuries/operator never able to correct	E3: 0.1 Often < 10% of operating time E4: 1 Frequently 10-100% of operating time	

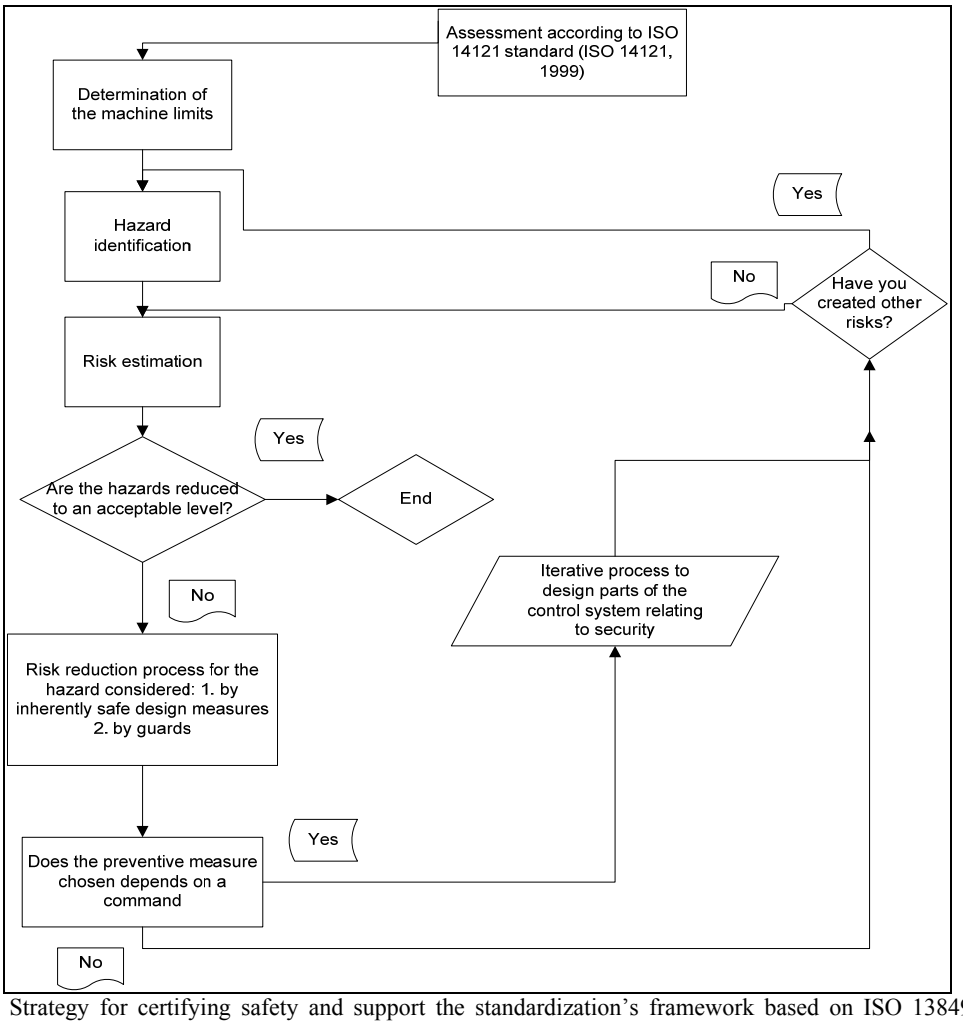

Figure 3. Strategy for certifying safety and support the standardization's framework based on ISO 13849-1 (ISO 13849-1, 2015).

Safety Requirements

Taking account of the status-quo of standardization activities towards precision agriculture and the computation of risk, safety restrictions are concentrated so as to apply the outcomes of risk management and ensure safety and reliability of operator and agribot (agricultural robot). The safety precautions for people and all stakeholders are given below for agricultural robots:

Hazard 1: Unexpected start-up

- The manufacturer should take the necessary precautions alerting the operator that regular tests on all protective structures need to be completed. Detailed directions on how to operate these tests should be provided. Unauthorized change of adapters, over speed the engine or rotating driveline is prohibited. The maintenance should be carried out by service experts.
- Resetting only a single button shall be unfeasible. A more complex reset procedure shall be applied to prevent from harmful situations. The target of this requirement is that a simple restart procedure shall not be feasible to solve any arbitrary condition preventing an operator from pushing by mistake a restart button and, as a result, an inappropriate or disoriented start-up most probable leading to an accident.

Hazard 2: Unexpected deceleration/stop

- The restart position shall be tested for its appropriate performance, proper calibration and maintenance.
- An emergency stop shall be embedded and tested that remains functional preventing any hazardous condition (Jiang & Cheng, 1990).
- Make sure stops will not result in a crushing and the embedded pressure safety button will be start if the robot exceeds its limits because of wrong operator command, malfunction in mechanics or bug in software.
- Detailed directions to the farmer shall be given in order to be able to constantly gather data and guidelines on line.

Hazard 3: Unexpected acceleration

- The tractor shall be prevented from running high speed (Bonney & Yong, 1985). High speed can result in loss of control, no ability of braking quickly, damage to equipment and components. Transfer the tractor with good visibility in order to be seen by other drivers, use extremity lights or flashers, pull over for traffic and frequently check for approaching from behind.
- Designer should check tractor's procedures naming manual stop, control of user presence and station, manual controller and drive speed.
- Post-manufacture check of the full scale system, about its maximum, minimum, optimal speed and settings, start/end points, path exceptional accuracy, time, save fuel or efficacy criteria, dead distance.

Hazard 4: Unexpected change of direction

- The emergency stop button shall stop within seconds every harmful movement quickly. The emergency stop should be easily accessible and maybe red color should help (IEC 60204-1, 2005).
- It should be stated clearly in the user's guide that operator will ask for help in case any abnormal or disoriented movement observed. Any possible interference shall not lead to loss of calibration or changing direction with no progress.
- The safety buttons, controllers, and sensors should pass dynamic tests and validated that they are in a good condition.

- The manufacturer should inform the operator that he shall maintain properly in place all guards, shields, buttons, safety features, and sensors. The operator should not turn on the machine if any parts are faulty or do not work.

Hazard 5: Wrong direction at start-up

- The stop and reset switches and process shall be frequently reminded. An audible alert if a shut-off switch stops working and a visual cue or emergency verbal exclamation, should be embedded if needed. The operator needs to be informed about secondary guidelines to halt and how to return to start position.
- The software shall embed a self-awareness algorithm that detects any arbitrary function in the field. If that program does not bear, the operator should not control a shut-off of a suspicious robot when possible.
- Field machine shall incorporate an emergency button to halt the motion on the remote control so as to block its path if it is arbitrary.

Hazard 6: Failure to start

- Control parameter settings should be aligned towards the specifications while the machine stays immobile or out of service and safety precautions ought to be given to the operator (Jiang & Cheng, 1990).
- The user's guide shall inform farmer, that he ought to inspect vehicle for weeds or corns, to clean base with a cloth, to ensure contact among station cables and power supply unit, to make sure that power supply is plugged at main power and check that coiled cord is placed in its place properly, to check the good reliability of the power cable to the connector and to confirm reliable connection in cables.

Hazard 7: Failure to stop

- Children and untrained people must not ride upon, misuse or attack the equipment. Operator shall pay attention to infants since they are careless and in a weak position. Pets shall be prevented from tragic accidents when hunting the robot or disturbing its task.
- Operator ought to possess a checklist of guidelines regarding the equipment's operation in a noticeable position in the farm.
- Ensure that all ranchers are well out of robot's motion path when starting the machine.
- Control a "return-to-home" function following deliberate control-link transmission failure.
- Warning signs of electrical shock shall be placed if needed to warn careless or unauthorized humans that may be careless.

Hazard 8: Failure to react on the command

- Some field robots are able to execute its task with restricted or no supervision. In some cases, limited control may be applied though remote control.
- According to the capabilities of the robot, it generally ought to be as solid as feasible with smallest chassis and heaviness, so as to maneuver effectively in the farm.
- The robot shall be equipped with an audio or visual cue in order to be recognizable in case of changing its path or task.

Hazard 9: Avoid damaging people, infrastructure, and crops

- The machine should be appropriately placed and arm the tools so that operator could halt the motion if a tool is mishandled. Embed a sensor to alert the operator of the mistreated tool and provide an indicator naming a strong noise or visible alert, to warn farmers and ranchers that a tool is misused.
- Do not stay in touch with the machine as it operates. Forbid farmers from expelling fruits or plants on the machine or tools while it executes among the crop rows.

• Communication between robots is necessary when multiple robots are operating on a farm. Multiple machines result in three levels of interaction: motion coordination, frequent communication, and collaboration. It shall be feasible to embed diverse communication protocols, plug and play features, software and hardware minimization that result to modularity and expandability. Vehicles with sprayer, autonomous harvester, truck and heterogeneous tractors can participate in the small team of multi-tractor system.

• All cables and wires must be embedded inside the chassis of the robot to avoid electrical shock.

Hazard 10: Robustness against noise and uncertainties

• The field robot should not be affected by interferences such as high ambient noise, in field radio signal, unshielded computers, infrared remote controllers and magnetic fields in order to ensure reliability and robustness in harsh conditions.

• The operator should ensure that any sensory equipment accompanied with the device is located not close to extreme light, microwave antennas, magnetic waves, vibration, heat, and sound (Jenkins, 1993).

• The robot shall be equipped with a specific audio or visual signal with a frequency that is not within the range of noise frequencies, in case of an audio signal.

Hazard 11: Recognize worn out tools

• Every faulty condition inside the machine should be recognized while the field equipment should carry on working if it does not violate the safety precautions.

• Except for spare of not well-preserved or auxiliary objects, the robot should not demand routine or unexpected treatment more than a time in a year.

Hazard 12: Quick detection and isolation of faults

• The programming architecture that enforces safety constraints on agribot should take into account that programs must not be affected by faults in operation system.

• A specialized diagnostics program ought to be installed so that it will figure out the root cause of a fault condition. The absence of a complete process might result in arbitrary handling rather than really solve the mismanagement.

• Use of well-established programming languages, where possible with a certified compiler. Avoiding error prone code features such as dynamic obstacles, corn, automatic type etc. Use of style guidelines and structured programming methodologies. Architectures where process software and integrity checks run independent from each other.

Hazard 13: Low modelling and low computational requirements

• The software shall divide the program in minor operational loops, the loops shall be divided with sequences and iterations, a restricted number of paths ought to be added in programming, complicated ramifications and leaps shall be prevented, modules shall be connected to input specification settings, and complex calculations should be avoided when making decisions on forks and loops.

• Machine task path simulation in 2-dimension or 3-dimension should be executed from manufacturer to imitate robot motion in an urgent situation. Since divergences among actual and the cybernetic world might take place, simulation information ought to be checked on the real prototypes.

• The manufacturer should check 3D georeferenced position and velocity control, person and object detection, path planning, obstacle avoidance, selective muting of safeguards, application of varying protection fields based on situation, safety-related autonomous decision. The field boundary information is well-known, as well as if it is

planted or no, the wet and dry yield, total yield, average moisture, seeding variety, and the machine location where the tractor can travel with coordinated motion.

- The computers of the main controller shall be connected to the network with point-to-point links considering the amount of sensors, peripherals, operating systems and computing power.

Hazard 14: Manual resetting

- Operator shall be informed about how to select manual resetting, verify that switching the external.
- Control does not lead to dangerous conditions, and how to prevent the event of loss of external control.
- Supportive tutorials and help guides should require practical guidelines, so that operator can have immediate communication in data on how to restart the machine and with a comprehensive knowledge how the agribot can operate its path.
- The values and parameters for the recovery procedure shall be tested, as it is crucial to reset the robot after a malfunction. A standby operation system ought to be triggered by programming in an emergency (Jiang & Cheng, 1990).
- If remote controller is shut off, it should continuous with manual mode (Izard, Ribeiro, Barreiro, & Valero, 2011).
- Expanding path offset and orientation offset, it should be continuous with manual mode.

Hazard 15: A lock to prevent switching on the system by accident

- Hardware control system and/or safety program should be initiated when pressing the machine to halt in an urgent condition, to protect from dangerous events or tearing of tools.
- Embedded diagnostic software should not lead to fail shutting off, in an emergency situation.
- Commands with redundant power for agribot stored in Random Access Memory should be tested so that a power outage would not result in a corruption of program.
- Agribot's parts shall use only backup in Read Only Memory to avoid altering the operational program.

Hazard 16: Leakage of moisture of liquids

- If overheated gasses, liquids, or combustible substances such as pesticide are contained in tanks inside the robot, the designer shall ensure that any increase of temperature will not cause fire. In case of fire, shut-off the equipment instantly, run away smoke, and if it is extended do not try to extinguish it.
- Chemical burn can be harmful in case that battery is mistreated and explosion hazard or leaking of corrosive liquid, if the battery is incorrectly placed. In particular, goggles, gloves, and respirator must be used for protection.
- Moisture content of the soil is another external condition influencing the performance of the robots. Moisture content can be measured by static sensors placed on strategic positions in the field identified by Autotrack guidance. The broad-acre spray for herbicide shall stop under strong wind conditions. The broadcast spray of herbicide should avoid chemical applications on cotton plants or sides without unwanted weeds. The robot distinguishes slight differences among colour and leaves as the machine passes learning as it goes.
- A visual daily check shall be implemented in case of leakage and that every safety features are in place.

Hazard 17: Software or camera failure

- Software crashes: Computer halts via hardware watchdog.
- Problems with camera including: no image, frozen image, wrong field of view, artefacts on display, image

which does not display, scenery (vary light), and delay image (Izard, Ribeiro, Barreiro, & Valero, 2011).

- Software shall embed all safety guidelines preventing from all hazards from bruises to blindness.

Conclusion

In conclusion, researchers rely only on machine learning and artificial intelligence and forget machine ethics, risk assessment and hierarchy among safety goals which is a paradox. The author proposes a strategy of safety assessment with a set of safety rules based on scientific knowledge and well-established standards as hoping to fill the standardization gap of these robots. Safety requirements shall be understandable and easily readable by humans reflecting the trade-off among safety goals. The agricultural robot should be able to react itself on unforeseen events like low sun, rain and snow, sloping terrain or hilly ground combining encoders, gyros, sonar and cameras. Verification and validation using computer vision and neural networks is still an open issue. Not-too-distant future work concerns regulation framework of multi-faceted robotics community using Artificial Neural Networks that they are going to play a significant role beyond "Asimovian" ethics (Asimov, 1950).

References

Asimov, I. (1950). *I, robot*. NY: Gnome Press.

Bonney, M. C., & Yong, D. (1985). *Robot safety*. UK: International Manufacturing Technology. IFS Publications Ltd.

Emmi, L., Gonzalez-de Soto, M., Pajares, G., & Gonzalez-de Santos, P. (2014). New trends in robotics for agriculture: Integration and assessment of a real fleet of robots. *The Scientific World Journal, 2014*, 1-21.

Foley, J. A. (2011). Can we feed the world and sustain the planet? *Scientific American, 305*, 60-65.

IEC 60204-1. (2005). *10.7.1 Devices for emergency stop shall be readily accessible*.

IEC 61508-1. (2010). *Functional safety of electrical/electronic/programmable electronic safety-related systems—Part 1: General requirements.*

IEC 61499-1. (2012). *Function blocks—Part 1: Architecture*.

ISO/DTR 23482-2. (n.d.). *Robotics—Application of ISO 13482—Part 2: Application guide*.

ISO 14121-1. (1999). *Safety of machinery. Risk assessment. Part 1: Principles*.

ISO 10218-1. (2011). *Robots for industrial environments—safety requirements—Part I: Robot*.

ISO 9001. (2015). *Quality management systems—Requirements*.

ISO 13849-1. (2015). *Safety of machinery. Safety-related parts of control systems. Part 1: General principles for design*.

ISO 11783. (2017). *Tractors and machinery for agriculture and forestry—Serial control and communications data network. Part 1: General specifications*.

Ingibergsson, M. T. J., Schultz, P. U., & Kraft, D. (2015). Towards declarative safety rules for perception specification architectures. *International Workshop on Domain-Specific Languages and Models for Robotic Systems*, September 28, Hamburg (Germany).

Izard, G. M., Ribeiro, A., Barreiro P., & Valero, C. (2011). Safety functional requirements for robot fleets for highly effective agriculture and forestry management. *Proceedings of the First International Workshop on Robotics and Associated High Technologies and Equipment for Agriculture*. France.

Jiang, B., & Cheng, O. (1990). A procedure analysis for robot system safety. *International Journal of Industrial Ergonomics, 6*(2), 95-117. Industrial Engineering Department, Auburn University, AL 36849, USA.

Jenkins, F. (1993). Practical requirements for a domestic vacuum-cleaning robot. *Technical Report*.

Mitchell, L. R. (2010). Toyota's lesson: Software can be unsafe at any speed. *Computerworld*.

Taubig, H., Frese, U., Hertzberg, C., Luth, C., Mohr, S., Vorobev, E., & Walter, D. (2012). Guaranteeing functional safety: Design for provability and computer-aided verification. *Autonomous Robots, 32*(3), 303-331.

TC 22. (2011). Road vehicles functional safety. *International Organization for Standardization.* International Standard ISO 26262, 2011.

TC 23. (2010). *Tractors and machinery for agriculture and forestry—Safety-related parts of control systems*. International Standard ISO 25119, 2010.

TC 23. (2014). *Agricultural machinery and tractors—Safety of highly automated machinery*. International Standard ISO/DIS 18497, ISO, 2014.

TC 44. (2012). *Safety of machinery—electro-sensitive protective equipment. International Electronical Commission*. International Standard IEC 61496, 2012.

TC 184. (2014). *Robots and robotic devices—Safety requirements for personal care robots. International Organization for Standardization*. International Standard ISO 13482, 2014.

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 482-486
doi: 10.17265/2328-7144/2018.06.007

Health Care Expenditure Inequality Associated With Pollution in China

Pengfei Sheng
Henan University, Kaifeng, China

This research builds an index to describe the inequality in health care expenditure related to pollution, and the measurement with Chinese provincial data set suggests that health care expenditure inequality related to pollution is obvious negative, which confirms that residents located in high-polluted areas cannot get the better health care resources. Meanwhile, the absolute value of health care expenditure inequality related to pollution is even larger than that related to income, which suggests the inequality related to pollution cannot be ignored.

Keywords: health care expenditure, inequality, pollution

Introduction

China's Gini coefficient is 0.462 in 2015, which indicates that income inequality is of increasing severity, and the most important factor affecting income inequality is the rising inequality in health care expenditure in China. Meanwhile, China's rising health care expenditure inequality is closely related to the pollution, which can do much damage to human's health. Thus, it is valuable to measure the health care expenditure inequality related to pollution.

A number of literatures consider that health inequality would be affected by many factors, such as the difference in social economic status (Lowry & Xie, 2009), the accessibility of medical care service (Evans & Kantrowitz, 2002), the different education level (Glied & Lleras-Muney, 2008), and so on. However, there are few researches discussed the relationship between pollution and health inequality. As suggested by Cropper (1981), pollution would promote the depreciation of health human capital, and many empirical studies argued that pollution can do much damage to residents' health and increase their health care expenditure (Narayan & Narayan, 2008; Coneus & Spiess, 2012; Janke, 2014). Although pollution would cause the same damage on any resident's health, residents would make the unequal decision on health care expenditure. The unequal decision is on account of different resident's constraints, such as income, education, the accessibility of medical care resources, and so on. Thus, the inequality in health care expenditure would be affected by pollution.

The prior researches indicate that health care expenditure inequality is obviously correlated with pollution. Meanwhile, pollution and health care expenditure inequality are both important factors affecting economic

Acknowledgements: This research was supported by the Youth Foundation of Humanities and Social Sciences of the Ministry of Education in China [grant number 15YJCZH139].

Pengfei Sheng, Ph.D., associate professor, Postdoctoral Center for Studies on Advanced Statistics, Henan University, Kaifeng, China.

Correspondence concerning this article should be addressed to Pengfei Sheng, School of Economics, Henan University, the North Part of Jinming Street, Kaifeng 475002, China.

development and social stability in China and other developing countries, and the two problems cannot be solved independently. In this paper, our research tries to provide an index to measure the inequality in health care expenditure related to pollution, and the measurement with China's provincial data set of the urban residents can provide some important implications for the policies to reduce pollution and alleviate the inequality in health.

Method and Data

Method

The concentration coefficient is often used to measure the inequality in health (Wagstaff, Van Doorslaer, & Watanabe, 2003; Islam, Gerdtham, Clarke, & Burstrom, 2010), and this index can take the bivariate distributions into account. Therefore, this research also uses the concentration coefficient to build the index to measure the health care expenditure inequality related to pollution. In the way provided by Kakwani, Wagstaff, and Van Doorslaer (1997), the concentration coefficient can be set as equation (1), and this index takes both the health distribution and the pollution distribution into account.

$$CHP = \frac{2}{N\mu_h}\sum_{i=1}^{N} h_i P_i - 1 \tag{1}$$

where, *CHP* denotes the health care expenditure inequality related to pollution, N is the number of the units in the sample, μ_h is the average level by health care expenditure, h_i is the health value for the i^{th} province. P_i is the rank for the i^{th} unit at the degree of the pollution, and the rank is measured by the percentage and weighted by the province's urban population.

Regarding to the index of *CHP*, its negative value suggests that the health care expenditure inequality may be benefit to the residents located in the less-polluted region, which means the residents in the less-polluted areas can get the better medical care resources and have the better health status. When the value of *CHP* is positive, this indicates that residents located in the more-polluted areas can have the better health status.

Data

In this paper, the research conducts a Chinese provincial data set of urban residents to measure the health care expenditure inequality related to pollution. Since pollution would do the same damage to all the residents located in one region, and then it is feasible choose the urban residents of one province as one group. This choice may underestimate the real health inequality related to pollution in China, because it ignores the inequality among the residents in one province. However, it is also preferable because all cities in the same province should accept the same policies on pollution reduction and health care in China.

The data set in this research covers China's 30 provinces[1] over the period 2005-2014. Health is measured by the health care expenditure per capita, which is derived from the urban household survey carried by China's official statistics bureau. Pollution is described by the emissions of the main pollutants, which comprise the emission of sulfur dioxide in waste gas per capita, the emission of nitrogen oxide in waste gas per capita, and the emission of smoke dust in waste gas per capita. Urban population is calculated by the permanent residents located in the urban area of one province. The data set is collected from *China Statistical Yearbook* and *China Environment Yearbook*, and Table 1 reports the main descriptive statistics.

[1] Due to the data availability, Hongkong, Makou, Taiwan, and Tibet are excluded from the sample in this paper.

Table 1

Descriptive Statistics

Variable	Description	Unit	Mean	Max	Min	Coefficient of variance
Health	Per capita health care expenditure	RMB per capita	886.91	2327.60	326.46	0.37
Pollution	Per capita emission of sulfur dioxide	Ton per 100 thousand persons	127.82	1147.95	11.89	1.49
	Per capita emission of nitrogen oxide	Ton per 100 thousand persons	206.85	2438.87	9.34	1.86
	Per capita emission of smoke dust	Ton per 100 thousand persons	168.79	2246.85	14.04	1.87
Urban population	Permanent residents in the urban area	100 thousand persons	146.19	1469.84	4.34	1.92

Results and Discussion

By using the equation (1), Table 2 reports the health care expenditure inequality related to pollution with China's provincial data set of the urban over the sample period 2005-2014. Meanwhile, we also calculate the health care expenditure inequality related to income, and the comparison with the two indexes can help to examine whether the health care expenditure inequality to pollution is severe and important.

Following the results in Table 2, it is found that all values of health care expenditure inequality related to pollution are negative. The negative health care expenditure inequality related pollution indicates that there is an obvious gap in health care expenditure among the residents located in areas with different degrees of pollution. Meanwhile, the negative value also suggests that the residents located in high-pollution areas cannot get the better medical care resources. Thus these residents would have the worse health in short- and long-run. Over the period 2005-2014, the health care expenditure inequality related to pollution shows an upward trend, which means that residents located in high-polluted areas become worse and worse in health care. Regarding to the health care expenditure inequality related to income, the value is positive and suggests that the wealthier residents can get the better health care resources. However, results also shows that health care expenditure inequality related to pollution is larger than that related to income at the provincial level[2], and this demonstrates that health care expenditure inequality related to pollution cannot be ignored, and policies on health care should pay more attention to the health inequality related to pollution.

As Grossman and Krueger (1994) argue that pollution is highly corrected with income, and then the pollution distribution should conclude some information in income distribution. When the same information is big enough, the measurement of health care expenditure related to pollution would not be efficient. Table 3 reports the correlation coefficient between pollution and per capita disposable income, and results income the two variables are correlated. However, their correlation coefficients range from 0.022 to 0.467, and are not big enough, thus it is concluded that measurement on the health care expenditure inequality should be of efficiency and can describe the actual status in China.

[2] Results cannot provide the evidence that health inequality related to pollution is larger than that related to income in the whole, because our measurement does not take the inequality within the province into account.

Table 2

Health Care Expenditure Inequality Related to Pollution in China

	Health care expenditure inequality related to pollution			Health care expenditure inequality to income
Reference indexes	Per capita emission of sulfur dioxide	Per capita emission of nitrogen oxide	Per capita emission of smoke dust	Per capita disposable income
2005	-0.0944	-0.0982	-0.0405	0.0447
2006	-0.1038	-0.0992	-0.0499	0.0473
2007	-0.1198	-0.1254	-0.0758	0.0744
2008	-0.1231	-0.0717	-0.1001	0.0792
2009	-0.1436	-0.0972	-0.1289	0.0931
2010	-0.1569	-0.1260	-0.1318	0.1057
2011	-0.1451	-0.1374	-0.1185	0.0883
2012	-0.1450	-0.1382	-0.1480	0.1192
2013	-0.1546	-0.1509	-0.1350	0.1123
2014	-0.1400	-0.1551	-0.1137	0.0970

Table 3

Correlation Coefficient Between Pollution and per Capita Disposable Income

Pollution indexes	Per capita emission of sulfur dioxide	Per capita emission of nitrogen oxide	Per capita emission of smoke dust
2005	-0.143	0.022	-0.423
2006	-0.179	0.200	-0.439
2007	-0.195	0.170	-0.427
2008	-0.205	0.130	-0.434
2009	-0.240	0.057	-0.467
2010	-0.277	-0.014	-0.408
2011	-0.344	-0.172	-0.432
2012	-0.346	-0.205	-0.424
2013	-0.321	-0.204	-0.368
2014	-0.334	-0.228	-0.367

Conclusions

Although pollution can make the same damage to residents' health, residents would make different decisions on health care expenditure because they are at different levels of income, education, the accessibility of health care resources, and so on. Thus, the inequality in health care expenditure is closely correlated with pollution. By using the method of concentration coefficient, this paper tries to provide an index to describe the inequality in health care expenditure related to pollution, and conducts a China's provincial data set of the urban residents with the period 2005-2014. The results indicate that the inequality in health care expenditure is negative, which means that residents can get the better health status and use more health care resources when they are located in less-polluted areas. Moreover, the inequality related to pollution is even larger than that related to income by the absolute value, and this confirms that health care expenditure related to pollution cannot be ignored, and policies on health should take the pollution difference into account in China.

References

Coneus, K., & Spiess, C. K. (2012). Pollution exposure and child health: Evidence for infants and toddlers in Germany. *Journal of Health Economics, 31*(1), 180-196.

Cropper, M. L. (1981). Measuring the benefits from reduced morbidity. *The American Economic Review, 71*(2), 235-240.

Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: The potential role of environmental risk exposure. *Annual Review of Public Health, 23*(1), 303-331.

Glied, S., & Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography, 45*(3), 741-761.

Grossman, G. M., & Krueger, A. B. (1994). Economic growth and the environment. *Quarterly Journal of Economics, 110*(2), 353-377.

Islam, M. K., Gerdtham, U., Clarke, P., & Burstrom, K. (2010). Does income-related health inequality change as the population ages? Evidence from Swedish panel data. *Health Economics, 19*(3), 334-349.

Janke, K. (2014). Air pollution, avoidance behaviour and children's respiratory health: Evidence from England. *Journal of Health Economics, 38*, 23-42.

Kakwani, N., Wagstaff, A., & Van Doorslaer, E. (1997). Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics, 77*(1), 87-103.

Lowry, D., & Xie, Y. (2009). Socioeconomic status and health differentials in China: Convergence or divergence at older ages? Research Report. Population Studies Center, University of Michigan.

Narayan, P. K., & Narayan, S. (2008). Does environmental quality influence health expenditures? Empirical evidence from a panel of selected OECD countries. *Ecological Economics, 65*(2), 367-374.

Wagstaff, A., Van Doorslaer, E., & Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics, 112*(1), 207-223.

Economics World, Nov.-Dec. 2018, Vol. 6, No. 6, 487-496
doi: 10.17265/2328-7144/2018.06.008

On a Certain Method of Rejecting Games With Too Much Risk and Not Enough Expected Gain

Ewa Drabik
Warsaw University of Technology, Warsaw, Poland

Many important economic decisions involve an element of risk. Risk aversion is a concept in economics, game theory, finance, and psychology related to the behavior of consumers, players, and investors under uncertainty. Loss aversion is an important component of a phenomenon that has been widely discussed in recent years. It refers to a tendency to feel the pain of a loss more acutely than the pleasure of an equal-sized gain. Many scientists have analyzed the problem of profitability in games. Some authors presented certain features which characterize "safe" games played once. Kahneman and Tversky (1991) showed that the ratio of loss aversion to gain attraction should amount to 1:2. The aim of this paper is to show an asymptotically efficient strategy which enables the risk-averse player to establish boundary variables of loss and gain at each stage of a repeated game.

Keywords: asymptotically efficient adaptive allocation rule, one-armed bandit problem, risk aversion, repeated games

Introduction

In some games the combined value of risk, losses, and threats exceeds any and all benefits the player could derive from the game. In such cases, a rationally behaving player should not participate in the competition. However, in many social games the participation is virtually mandatory, and it is often necessary to cooperate with a group of individuals sharing common goals (e.g. in a workplace environment). The stakes of such a game can be resources, a formal structure, and the individual players' positions in that structure. If one does not participate in the game—consciously or not—they can drop out of this structure. This principle is similar for many other market structures (i.e. financial market) whose existence depends mainly on the behavior of the structure's participants.

It is usually assumed that a rational player has an aversion to risk. Risk aversion is a long known concept in economics, game theory, finance, and psychology. It refers to the behavior of consumers, players, and investors operating under conditions of uncertainty. The degree of risk aversion can be expressed mathematically; the most known measures of risk aversion were introduced by Pratt (1964) and Arrow (1965). One of the most important issues is the problem of participation in undertakings, investments, or games, including those played on the market, which are characterized by high risk. From the point of view of the player (e.g. a market participant) it is extremely important to determine the maximum loss and the minimum

Ewa Drabik, professor, Faculty of Management, Warsaw University of Technology, Warsaw, Poland.
Correspondence concerning this article should be addressed to Ewa Drabik, Faculty of Management, Warsaw University of Technology, Narbutta 85, Warsaw 02-524, Poland.

profit which will make them decide to participate in the game. These problems were discussed by numerous authors (Palacios-Huerta & Serrano, 2006; Rabin, 2000; Tversky & Kahneman, 1991).

The problem of risk aversion was raised and then developed by Kahneman and Tversky (1991) as a part of their study of prospect theory. Their theory states, among other things, that the economy is composed of not only regularities but also people, who do not always behave rationally. They also showed that market participants perceive a loss differently than an equal (when it comes to value) profit. In other words, the pain of a loss can be felt more acutely than the joy of an equal profit. In 2000 Rabin showed that if a player is a risk averse expected utility maximizer and rejects a given gamble of equally likely gain g and loss l, $g > l > 0$, over a given range of wealth level, then they will reject correspondingly larger gambles of gain G and loss L. Palacios-Huerta and Serrano (2006) showed the relationship between risk attitudes for low- and high-stakes gambles.

As far as the rational approach to a game is concerned, it is of major importance to establish l (loss) as well as g (gain) at each stage. Loss aversion was introduced by Kehneman and Tversky (1979) as part of the more general "prospect theory". In 1991 they, as well as others, estimated the loss aversion to gain attraction ratio to be about 2:1. Many authors such as Rabin and Thaler (2001), Segal and Spivak (1990), and Epstein (1992) have discussed the problem of participation in games of different stakes, and examined the participants' attitude to risk in different contexts and with different utility functions.

The problem of rejecting "bad" games was also discussed by Palacios-Huerta and Serrano (2006). They identified some common features of safe games, and at the same time formulated a very important principle which determines the profit and loss thresholds for the game to be considered "safe"—this will be presented in a later section. Aumann and Serrano (2007), basing on the Arrow-Pratt risk aversion measures, defined the so-called riskiness indices whose properties were expressed, among other methods, with axioms.

The aforementioned study considered only single-stage games. The aim of this work is to present a strategy to determine profit and loss thresholds in games played repeatedly. This will be accomplished using an asymptotically efficient strategy which was originally used by Lai and Robbins (1985) to solve the one-armed bandit problem.

This paper is divided into sections in the following way. The second part presents some of the theory related to different attitudes of market participants to risk, as well as the principle of determining profit and loss thresholds at which one should not participate in a game played once (Palacios-Huerta & Serrano, 2006). The third part presents a strategy for calculating profit and loss thresholds at which one should reject a multi-stage game at a specific phase in order for the game to be profitable.

Research Design

Research Subject

The paper aims at discussing an asymptotically efficient allocation rule which enables the risk-aversive player to establish loss/gain thresholds at each stage of a multi-stage game. The specific objective is to present measures of risk aversion that make it possible to assess the situation in the game.

Theory and Hypotheses

The allocation rule based on the one-armed bandit problem solution is asymptotically efficient. Therefore it results in successful rejection of so-called bad games.

Basic Information About Risk Aversion, Neutrality, Risk-Loving Player

Elements of the Expected Utility Hypothesis

Risk aversion is a concept in economics, game theory, finance, and psychology related to the behavior of consumers, players, and investors under uncertainty.

The most famous measures of risk-aversion were introduced by Pratt (1964) and Arrow (1965). For a player (or decision maker) with wealth level *w* and twice differentiable utility function *u*,

Definition 1. The Pratt-Arrow coefficient of absolute risk aversion, at wealth *w*, is:

$$r_A(w,u) = -\frac{u''(w)}{u'(w)} \tag{1}$$

where *u* is a twice differentiable Bernoulli utility function.

The coefficient of relative risk aversion (RRA) is:

$$r_r(w,u) = w \cdot r_A(w,u) \tag{2}$$

Absolute risk aversion is a measure of the player's reaction to uncertainty associated with absolute changes in affluence. Relative risk aversion expresses the player's attitude to the "uncertainty" associated with the relative changes (percentage) in affluence. A player whose risk aversion decreases along with an increase in affluence will spend more and more money on a risky game. There may also be differences in the players' degree of risk aversion.

The expected utility can be expressed as:

$$V = \sum_x p(x)u(x) \tag{3}$$

where

$u : X \to R$ elementary utility function (Bernoulli utility function);

p_i probability of an outcome occurring (representing a probability in simple lottery).

A simple example to illustrate the subsequently discussed concepts is simple lottery. In this case the probability p_i represents the probability of event *i*, while *n* – tuple $L = (p_1,...,p_n), \quad p_i \geq 0$ defines the lottery, while $i = 1, ..., n$ and $\sum_i p_i = 1$. Lottery can also be expressed geometrically as a point in an *n*-dimensional simplex.

$$\Delta = \left\{ p \in [0,1]^n : p_1 + ... + p_n = 1 \right\}.$$

Let $F_z(x) = P\{z \leq x\}$ be a distribution function corresponding to the random variable *z*. The player makes a selection according to the following guideline:

F_z is preferable to F_y, i.e. $F_z \succ F_y$ if and only if

$V(F_z) \geq V(F_y)$ where $V(F_z) = \int u(x)dF_z(x)$ or equivalently $V(F_z) = \sum_x p(x)u(x)$;

(analogically defining $V(F_y)$).

Let us assume that the random variable *z* has two possible values z_1 or z_2. Let *p* denote the probability of

event z_1, and (1-p) the probability of event z_2. u will denote the elemental utility function, whose expected value can be determined as follows:

$$E(u) = pu(z_1) + (1-p)u(z_2).$$

The following denotations will be used.

X—set of outcomes (real line R);

Z—random variable;

M—set of random variables;

$C(Z)$—the income from lottery;

$\Pi(Z) = E(Z)-C(Z)$—risk premium.

It is assumed that the player qualifies as:

- Risk-averse:

A player is "risk-averse" if $C(Z) < E(Z)$ or $\Pi(Z) > 0$ for all $Z \in M$;

- Risk-neutral:

A player is "risk-neutral" if $C(Z) = E(Z)$ or $\Pi(Z) = 0$ for all $Z \in M$;

- Risk-loving:

A player is "risk-loving" if $C(Z) > E(Z)$ or $\Pi(Z) < 0$ for all $Z \in M$.

The following theorem is true.

Theorem 1. Let $u : R \to R$ be an elementary utility function representing preferences $\geq_h$ over M which is monotonically increasing.

Then

(1) u is concave if and only if $\geq_h$ displays risk-aversion;

(2) u is convex if and only if $\geq_h$ displays risk-loving;

(3) u is linear if and only if displays risk-neutral.

Each player with utility u can show:

- Decreasing absolute risk aversion (DARA):

u displays absolute risk aversion if $\Pi(w,u) > \Pi(w+a,u)$ for all $a > 0$.

- Increasing absolute risk aversion (IARA):

u displays increasing absolute risk aversion if $\Pi(w,u) < \Pi(w+a,u)$ for all $a > 0$.

- Constant absolute risk aversion (CARA):

u displays constant absolute risk aversion if $\Pi(w,u) < \Pi(w+a,u)$ for all $a > 0$.

Absolute risk aversion is a measure of the player's reaction to the uncertainty associated with absolute changes in affluence. Relative risk aversion expresses the player's relation to the "uncertainty" associated with the relative (percentage) changes in the level of his wealth. A player with a decreasing risk aversion along with an increase in affluence will spend more and more money on a risky game. Players may also have a varied degree of risk aversion.

The player has an absolute risk aversion if $r_A(w,u)$ is a decreasing function for a given $u(.)$. The selection of $u(.)$ is therefore a very important problem in the discussed theory. Table 1 presents examples of utility functions and their classification in terms of changes in absolute and relative risk aversion.

It should be emphasized that the utility functions presented in Table 1 are characteristic of a risk-averse

investor. Other authors, e.g. Rabin (2000) presented other utility functions, such as

$$u(w) = \frac{w^{1-\gamma}}{1-\gamma} \quad \gamma \geq 0 \tag{4}$$

The $u(.)$ selection problem is widely discussed, because it is assumed that not every player avoids risk to an equal degree. Studies on various issues make use of other utility functions, i.e. with different analytical forms and different properties resulting from the signs of the first and second derivative. It was however found that as a general rule, the basic characteristics of this function, assuming that the players maximize their expected utility, are not very sensitive to the type of individual risk propensity (represented by absolute and relative risk aversion). It should be added that besides the aforementioned Arrow-Pratt measures, other methods of measuring risk aversion also exist (Mass-Colell, Whinston, & Green, 1995).

Table 1

Example Utility Functions and Their Classification in Terms of Changes in Absolute and Relative Risk Aversion

Utility function	Absolute risk aversion	Relative risk aversion
$u(w) = \ln w$	Decrease	Constant
$u(w) = (w+C)^{-\gamma} \quad c > 0, \quad \gamma \in (0,1)$	Decrease	Increase
$u(w) = -e^{-\gamma w} \quad \gamma > 0$	Constant	Increase
$u(w) = e^{-2w^{-\frac{1}{2}}}$	Decrease	Decrease

Source: based on Mass-Colell, Whinston, and Green (1995).

"Safety" Games

As was mentioned in the introduction, in the course of studies on the behavior of players, investors, consumers, and other market participants the problem of participation in high risk ventures developed naturally. Players are not willing to participate in games which bring not enough profit or too much loss. In 1991, Kahneman and Tversky proposed the term "loss aversion" instead of "risk aversion". In 2006, Palacios-Huerta and Serrano (2006) proposed and proved a theorem allowing to calculate the threshold value of loss l to gain g at which one can consider a game to be safe and participate in it.

Theorem 2. Let u satisfy non-increasing absolute risk aversion. Let I be an interval in the positive real line. For every $w \in I$

$$\frac{1}{2}u(w+g) + \frac{1}{2}u(w-l) < u(w) \tag{5}$$

there exists $a^* > 0$ such that the absolute risk aversion coefficient $r_A(w,u)$ is greater than a^* for all $w \in I$.

Moreover, the highest such a^* is the solution to the equation

$$f(a) = e^{al} + e^{-ag} - 2 = 0 \tag{6}$$

where l is loss, g is gain.

Player rejects a game losing l or gaining g.

The importance of theorem 2 is impossible to overstate. It must be noted, however, that it applies only to games played a single time. The next chapter will introduce a method that allows to determine the thresholds U

and μ in multi-stage games (games played repeatedly).

Asymptotically Efficient Adaptive Allocation Rules and One-armed or Multi-armed Bandit

Let us assume that the players participating in a particular game are risk-averse. This implies that the players feel the pain of a loss more acutely than the satisfaction from a gain of equal value. Therefore it is important to set threshold values of gain and loss which would satisfy the participants of games played repeatedly. This problem will be solved using a modified version of the strategy presented in the work of Lai and Robbins (1985) which was originally designed to solve the issue one-armed bandit problem.

The one-armed bandit is a mechanical or electrical machine, most frequently equipped with three drums holding a variety of card faces, used for gambling purposes. This simple machine has been known since 1887, when Charles Fey devised an automatic mechanism to make the trade of offered products more attractive. Since the machines proved to be very successful, in the middle of the 20th century the mechanical constructions were substituted by computerized devices, which were used in casinos. They were used to entertain and divert the ladies whose husbands and fiancés devoted themselves to card games.

First automats worked on a fully random basis. Currently, due to the application of software, the randomness is only partial. The participation in such games and generation of shots attracted the attention of many scholars who predominantly aimed to elaborate on profit-maximizing strategies. These strategies became attributable to many disciplines, such as economics, biology, control theory (compare available literature: Lai & Robbins, 1985; Anantharam, Varaiya, & Warland, 1987).

The three-reel slot machines invented in San Francisco in 1896 by Charles Fey underwent substantial changes over the course of the 20th century. For example, while classical slots were purely mechanical or electro-mechanical, modern ones are electronic and controlled by microprocessors with random number generators. Slot machines are the only casino games where evaluating the house advantage requires information not generally available to the gambling public. The slot machine is essentially a cabinet housing 3 or more narrow cylindrical drums, which are marked with symbols. Vertically placed on a common axis, the reels are caused to revolve freely when a player activates the machine and pulls a lever affixed to the side of the cabinet. Awards, which are paid out automatically, are based on the horizontal alignment of symbols when the spinning reels come to the position of inertial rest (Ethier, 2010).

Unstated but implicit in the definition was the requirement that the players insert one or more coins into a slot (hence the name "slot machine") in order to activate the machine. Nowadays almost all slots machines are electronic and controlled by microprocessors. The microprocessor is programmed with a random number generator that operates continually. The moment the player pulls the handle or presses the button to activate the machine, the most recently generated random numbers almost instantly determine the outcome. Despite all this, the underlying mathematics of the modern slot machine is much different from that of a classical machine.

Slot machines have a lot applications, for instance in statistics. One of them is the asymptotically efficient adaptive allocation rule which was assigned by Lai and Robbins (1985). In the following years the rule was further developed by other scholars and was renamed as the multi-armed bandit problem.

One of the strategies—the asymptotically efficient adaptive allocation rule—which is used in the field of statistics may be described as follows.

Let w_i (i = 1, 2...) denote statistical populations (wealth, treatments, manufacturing processes etc.) specified by a univariate (or another, for instance normal) density function $f(w,\theta i)$ with respective measure ν,

where $f(.,.)$ is known and θi are unknown parameters belonging to a set Θ. Assume that

$$\int_{-\infty}^{\infty} |w| f(w,\theta) dv(w) < \infty \quad \text{for all} \quad \theta \in \Theta \tag{7}$$

We sample w_1, w_2.... sequentially from the k populations in order to achieve the greatest possible expected value $S_n = w_1 + ..., w_2$ as $n \to \infty$. The multi-armed bandit problem is based on an imagined slot machine with $k \geq 2$ arms. When an arm is pulled the player wins a random reward. For each arm j there is an unknown probability distribution of the reward. The player's choices are in some way determined by the rewards of previous trials, so as to maximize the long-run total expected reward.

The method enables the establishment of random variables l and g in a more precise manner. Therefore, it can be counted among the asymptotically efficient adaptive allocation rules, which were applied for the first time by Lai and Robins (1985) to the multi-armed bandit problem with $k \geq 2$ arms. The rule may be explained as follows: in an n-stage game w_1, w_2, …, w_n, on the basis of payouts (historical data) whose distribution is parameterized by an unknown parameter θ belonging to a known parameter space, two auxiliary statistics μ_n and U_n are determined. Statistic μ_n represents the random variable l and generally represents the average payout in the game (mean). U_n may be considered as the upper bound of the confidence interval for a mean (a lower bound). The properties as well as construction of the above statistics will be presented below. In the following section loss l is represented by statistic μ_n, while gain g is represented by statistic U_n.

It is assumed that during a specified number of stages ($T_n \leq n$) the player continues to participate in the game regardless of the payouts received.

$$\mu(\theta) = \int_{-\infty}^{+\infty} w f(w;\theta) d\nu(w) \tag{8}$$

Then

$$ES_n = \sum_{j=0}^{k} \sum_{i=1}^{n} E\left(w_i I_{\{\varphi_i = j\}} \middle| \Im_{i-1}\right) = \sum_{j=0}^{k} \mu(\theta) T_n(j) \tag{9}$$

where

$T_n(j) = \sum_{i=1}^{n} I_{\{\varphi_i = j\}}$ is the number of times that φ samples to stage n. The problem of maximizing ES_n is

therefore equivalent to that of minimizing regret

$$R_n(\theta) = n\mu^* - ES_n = \sum_{j:\mu(\theta_j)<\mu^*} (\mu^* - \mu(\theta_j)) E T_n(j) \tag{10}$$

where

$$\mu^* = \max\{\mu(\theta_0), \mu(\theta_1)\} = \mu(\theta^*) \quad \text{for some} \quad \theta^* \in \{\theta_0, \theta_1\}$$

Let $I(\theta,\lambda)$ denote the Kullback-Leibler number

$$I(\theta,\lambda) = \int_{-\infty}^{\infty} \log \frac{f(w;\theta)}{f(w;\lambda)} \cdot f(w;\theta) dv(w) \tag{11}$$

such that $0 < I(\theta,\lambda) < \infty$ whenever $\mu(\lambda) > \mu(\theta)$.

Generally regret is

$$R_n(\theta) \approx \left\{ \sum_{j:\mu(\theta_j)<\mu^*} (\mu^* - \mu(\theta_j)) / I(\theta_j, \theta^*) \right\} \log n \quad \text{for } n \to \infty \ . \tag{12}$$

The rule φ.

Let $w_1, w_2, \ldots$ be a sequence of random variables whose distribution is parametrized by unknown parameter $\theta \in \Theta$.

Let $g_{nt} : \mathfrak{R}^t \to \mathfrak{R}$ ($n = 1, 2, \ldots; t = 1, \ldots, n$) be a Borel function such that for every $\theta \in \Theta$

(W_1) $P_\theta \{ r \le g_{nt}(w_1, \ldots, w_t) \quad \text{for all} \quad t \le n \} = 1 - o(n^{-1})$

(W_2) $\lim_{\varepsilon \to 0} \left(\limsup_{n \to \infty} \sum_{t=1}^{n} P_\theta \{ g_{nt}(w_1, \ldots, w_t) \ge \mu(\lambda) - \varepsilon \} / \log n \right) \le 1 / I(\theta, \lambda)$

whenever $\mu(\lambda) > \mu(\theta)$.

(W_3) g_{nt} is nondecreasing in $n \ge t$ for every fixed $t = 1, 2, \ldots$

Let $h_t(w_1, \ldots, w_t)$ be a Borel function, $h_t : \mathfrak{R}^t \to \mathfrak{R}$, such that for every $\theta \in \Theta$

(W_4) $h_t \le g_{nt}$ for all $\theta \in \Theta$.

(W_5) $P_\theta \{ \max_t | h_t(w_1, \ldots, w_t) - \mu(\theta) | > \varepsilon \} = o(n^{-1})$ for every $\varepsilon > 0$.

Condition W_5 can be satisfied for the average: $h_t(w_1, \ldots, w_t) = (w_1 + \ldots + w_t)/t$ if $E_\theta w_i^2 < \infty \quad (i = 1, \ldots, t)$
and

$$g_{nt}(w_1, \ldots, w_t) = \overline{w}_t + \sigma (2 a_{nt})^{1/2} \quad \text{for } n \ge t \tag{13}$$

where σ is standard deviation;
a_{nt} ($n = 1, 2, \ldots; t = 1, \ldots, n$) is a positive constant such that for all t, a_{nt} is nondecreasing for all $n \ge t$
and there exists $\varepsilon \to 0$ such that

$$\left| a_{nt} - \frac{\log n}{t} \right| \le \frac{\varepsilon (\log n)^{1/2}}{t^{1/2}} \quad \text{for all } t \le n.$$

Remark. If $f(.,.)$ denotes the density function of the normal distribution i.e.:

$$f(w; \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{(w-\theta)^2}{2\sigma^2} \right\}$$

then

$$I(\theta, \lambda) = \frac{(\theta - \lambda)^2}{2\sigma^2} \ . \tag{14}$$

Let $w_1, \ldots, w_{T_n(j)}$ be the successive rewards obtained from arm j up to stage n. The upper confidence

bound $U_n(j)$ and the point estimate $\mu_n(j)$, the mean reward under arm j, are given by

$$\mu_n(j) = h_{T_n(j)}(w_1,...,w_{T_n(j)})$$

$$U_n(j) = g_{n,T_n(j)}(w_1,...,w_{T_n(j)})$$

Define the "leader" at stage n as the population with the largest estimated mean among all population. The following rule φ is applied to decide which arms to play.

In the first stage we observe w_1, w_2,..., w_n and compute $U_n(j)$ and $\mu_n(j)$. Next, the player can take one of the following decisions.

In stage (n+1) they make one of the following decisions:

(1) provided that in stage n the player has not participated in the game and $\mu_n \leq 0$ they do not enter the game in stage (n+1);

(2) provided that in stage n the player has not participated in the game and $U_n < \mu_n$, they do not enter the game in stage (n+1);

(3) provided that in stage n the player has not participated in the game and $U_n \geq \mu_n$, they enter the game in stage (n+1);

This is to prove that this type of strategy is asymptotically efficient in the sense of maximizing joint payouts (or minimizing the regret) in the game. What also seems interesting is the manner in which the above strategy influences risk-averse indices, which was presented in chapter 2.

Theorem. The rule φ is asymptotically efficient.

Strategy φ is asymptotically efficient, which means that the cost of such a game at $n \to \infty$ converges to an asymptote. This is particularly important for games where participation is mandatory and one can only opt out of participation in individual phases of the game.

The strategy presented in this work is intended for repeated games other than the one-armed bandit or stock market.

Conclusions

Some affluent people are seemingly fascinated by extreme risk and very often engage in activities which are not entirely legal, counting on a large profit. This raises a question related to psychology: why do they risk so much to have even more? There is no rational explanation. Keynes has stated that "when the development of a country becomes a side effect of casino operations, one should not expect positive results."[1] Initially it was thought that financial markets are able to correct prices even if they are rising irrationally. However, an opinion of increasing popularity is that investors behave irrationally in moments of panic, often conducting speculations which are destructive, particularly to their competitors. While it is true that some mathematical models were developed on the assumption that investors rationally calculate risk and gain (e.g. Capital Asset Pricing Model—CAPM), predicting how to construct a wallet and set the price of financial instruments, they proved to be purely theoretical. The assumption of an efficient market caused the emergence of numerous "bubbles", and thus economics as a science was faced with some trouble. Financial markets are far from perfect. However, it is worth trying to counteract such peculiarities, relying among others on the help of mathematicians. The strategy

[1] https://en.wikiquote.org/wiki/John_Maynard_Keynes.

presented in this work can contribute to this goal.

Games that initially bring high profits but are burdened with high risk are also dangerous from the point of view of potential investors (e.g. financial pyramids). Therefore, in the long term perspective, taking too much risk is inadvisable even if the game initially brings profits. It is better to withdraw from the game at a calculated stage, as discussed in the presented strategy ϕ. If the game consistently brings losses, the decision seems obvious—it is best not to play.

References

Anantharam, V., Varaya, P., & Warland, J. (1987). Asymptotically efficient allocation rules for multiarmed bandit problem with multiple plays. *IEEE Transactions on Automatic Control, 32*(11), 968-982.

Arrow, K. (1965). *Aspects of the theory of risk-learning*. Helsinki: Yrjö Johnssonin Sätiö.

Aumann, R. J., & Serrano, R. (2007, July). An economic index of riskiness. *3th Spain Italy Netherlands Meeting on Game Theory (SING 3)*, Amsterdam.

Drabik, E. (2010). Asymptotically efficient adaptive allocation rule as a method of rejecting low and excessively high risk. *Statistical Review, 1*(57), 53-65.

Epstein, L. G. (1992). Behavior under risk: Recent developments. In. J. L. Laffont (Ed.), *Theory and applications in advances in economic theory* (Vol. II, pp. 1-63). Cambridge University Press.

Ethier, S. N. (2010). *The doctrine of chances: Probabilistic aspects of gambling*. Berlin-Heidelberg: Springer Verlag.

Kimball, M. S. (1991). Standard risk aversion. *NBER Technical Working Paper*, 99.

Mass-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory.* New York: Oxford University Press.

Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advanced in Applied Mathematics, 6*, 4-22.

Palacios-Huerta, I., & Serrano, R. (2006). Rejecting small gambles under expected utility. *Economics Letters, 91*, 250-259.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica, 1*(32), 122-136.

Rabin, M. (2000). Risk aversion and expected utility: A calibration theorem. *Econometrica, 68*, 1287-1292.

Rabin, M., & Thaler, R. H. (2001). Anomalies risk aversion. *Journal of Economic Properties, 15*(1), 219-232.

Segal, U., & Spivak, A. (1990). First order versus second order risk aversion. *Journal of Economic Theory, 51*, 111-125.

Tversky, A., & Kahneman, R. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics, 106*, 204-217.

Economics World
Volume 6, Number 6, Nov.-Dec. 2018
David Publishing Company
616 Corporate Way, Suite 2-4876, Valley Cottage, NY 10989, USA
Tel: 1-323-984-7526, 323-410-1082; Fax: 1-323-984-7374, 323-908-0457
http://www.davidpublisher.com, www.davidpublisher.org
economics.world@davidpublishing.org, economics.world@hotmail.com

ISSN 2328-7144
12
9 772328 714181