Journal of Applied Economics. Vol XX, No. 1 (May 2017), 1-28

MAN VS. MACHINE: AN INVESTIGATION OF SPEEDING TICKET DISPARITIES BASED ON GENDER AND RACE

Sarah Marx Quintanar[*]
University of Arkansas at Little Rock

Submitted June 2015; accepted March 2016

This paper analyzes the extent to which police behavior in giving speeding tickets differs from the ticketing pattern of automated cameras, which provide an estimate of the population of speeders. The novel data are obtained from Lafayette, Louisiana court records, and provide specific details about the ticketed driver as well as a wide range of violation characteristics. In contrast to the automated cameras, the probability of a ticketed driver being female is consistently and significantly higher when the ticket was given by a police officer. For African-American drivers this effect is less robust, though in general still positive and significant. This implies that police use gender and race as a determining factor in issuing a speeding ticket. Potential behavioral reasons for this outcome are discussed. The validity of using automated cameras as a population measure for police-issued tickets is thoroughly investigated and supportive evidence is provided.

JEL classification codes: J71, K42
Key words: gender bias, racial bias, police, ticketing, automated traffic cameras

I. Introduction

Since the seminal work of Becker (1957), which created the theoretical foundation of economics of discrimination, researchers have empirically investigated the existence of discrimination in a variety of settings ranging from wages to murder

[*] 2801 S. University Drive, Little Rock, AR; email: smquintanar@ualr.edu. I would like to thank Kaj Gittings, Robert Newman, Sudipta Sarangi, and Naci Mocan for guidance in many matters related to this research. Also thanks to Carter Hill for insight regarding econometric concerns and to Lafayette City Court for their cooperation and willingness to help in the data collection process and related data questions. Christian Raschke was extremely helpful in constructing the ticket maps and has my gratitude. Thank you as well to seminar participants in the LSU Economics Department, University of Arkansas seminar series, Southern Economic Association Conferences and CeMent Participants. All provided invaluable comments and suggestions.

trials.[1] Of particular relevance to the present work is the investigation of racial and gender bias of police in motor vehicle searches and ticketing for driving violations, which is costly to innocent individuals of a targeted race or gender (Durlauf 2006). As a result of little to no external validity of studies across different police departments, findings are mixed: some researchers find evidence of racial and/or gender discrimination (Antonovics and Knight 2009, Blalock et al. 2007, Makowsky and Stratmann 2009, Anwar and Fang 2006), while others report evidence of no discriminatory behavior by law enforcement officers (Knowles et al. 2001, Persico and Todd 2007, Grogger and Ridgeway 2006).

This paper exploits detailed, unique data from automated speed detection to measure differences in the proportion of speeding tickets issued to gender and racial groups in Lafayette, Louisiana. By comparing the proportion of women and African-Americans who receive tickets from police officers to those who receive tickets from an automated source, it is possible to determine if police use gender or race as a determinant in issuing speeding tickets. I find strong, statistically significant evidence that police consider gender when deciding to ticket speeders, and some evidence that race is also a factor even when accounting for potential endogeneity of the location of officers and automated sources.

In the context of this analysis, it is impossible to distinguish between tastes versus statistical motives for differential ticketing; however, the first-order issue is whether or not these types of behaviors exist at all. Preference-based discrimination means police derive an additional non-monetary benefit by ticketing these individuals. Differential treatment based on gender (or race) is considered statistical discrimination if police officers use gender (or race) as a proxy for a relevant characteristic which is difficult to observe. For example, perhaps police frequently ticket women because, on average, they are considered more dangerous, more likely to change their future behavior as a result of the stop, or even more likely to pay a speeding ticket fine instead of going to court to contest it (Blalock et al. 2007). Though taste for discrimination cannot be ruled out, later I present evidence that police behave rationally in that they issue tickets more frequently to those who speed 16 miles an hour or more over the limit (rather than those who were only traveling 5-15 miles an hour above the speed limit), which is associated with higher fines.

[1] For example, Munnell et al. (1996) control for credit worthiness, labor characteristics, race, gender, age, job history, and neighborhood characteristics in identifying the impact of race on mortgage rejection rates. Argys and Mocan (2004) investigate the impact of race and gender on death row commutation by controlling for characteristics of the criminal and crime, as well as the governor's party affiliation, race, and gender.

Researchers approach the question of police discriminatory behavior by utilizing variations of two general methods: the outcome test approach and the benchmarking test approach (Becker 1957, Anwar and Fang 2006). Mainly utilized in stop and search related questions, the outcome test approach is best explained by considering the work of Knowles, Persico, and Todd (2001). Utilizing highway data from Maryland, they illustrate equal success rates for searches of motor vehicles driven by African-Americans and whites despite the fact that more searched cars are driven by African-Americans. This implies police are engaging in statistical discrimination: once a car has been stopped, police are more likely to search if the driver is African-American because, on average, it is more likely that they will find drugs or contraband. Race is statistically related to likelihood to carry drugs. However, Antonovics and Knight (2009) use the same dataset and find that if the officer's race is different from that of the offender, the vehicle is more likely to be searched, implying a motive of preference-based discrimination. The outcome test approach requires data which provides two stages: for example, which driver and cars are stopped and then, of those, which cars are subsequently searched.

The benchmarking test compares a population to the sample of interest to determine if the sample ticketed by police is different by race or gender, all else equal. One drawback of this approach is the inability to distinguish between preference- based and statistical discrimination, however, the current paper provides a distinct advantage over previous literature due to the richness and reliability of the population data. In previous studies, the benchmarking test approach has not been reliable due to the denominator problem, which occurs if the "benchmark" used is not the same population observed by police. For example, comparing the population of all drivers in a city to those ticketed by police may not be reliable if the population of drivers committing an offense differs from the city driver population by race and/or gender. Then, the underlying comparison and any conclusions drawn would be invalid.

For the present work, the automated tickets provide an entirely objective measure of the speeding population in a given location where every speeding car that passes receives a ticket, and from which police ticket a subsample of drivers. Due to the uniqueness of this dataset, the denominator problem is eliminated. This assumption is reinforced in Section VI where I employ a similar methodology to Grogger and Ridgeway (2006), who utilize differences in daylight savings time to compare tickets issued by police during daylight to those issued during darkness, when race is likely unobservable. They find no significant evidence of racial profiling in Oakland, California, whereas I find evidence consistent with my main results, reinforcing the validity of using automated cameras as the population measure for police-issued tickets.

The method of data collection itself can be a problem for either approach: if police do not report all incidents of tickets or stops, analyses will be biased. This issue is referred to as nonreporting, and is a consistent problem within the literature (Grogger and Ridgeway 2006, Knowles et al. 2001, Persico and Todd 2007). Mainly arising in conjunction with post-lawsuit data (Grogger and Ridgeway 2006, Blalock et al. 2007, Knowles et al. 2001, Persico and Todd 2007, and Makowsky and Stratmann 2009), nonreporting occurs because police officers are legally required to record all stops, not only the ones which result in a ticket. However, in an effort to avoid repercussions from the lawsuit and suspected racial discrimination, police officers may change their behavior as a result of the knowledge of data collection.[2] For example, Makowsky and Stratmann (2009) find that Hispanics are more likely to be fined, but there is no difference in fines issued to African-American drivers, which they state may be a product of widespread knowledge of the study and data collection by the police department.[3] Audit studies have found a large discrepancy between actual stops and reported stops, especially in initial data collection, where up to 70% of stops were not recorded (Grogger and Ridgeway 2006).[4] The dataset used in this paper has a distinct advantage because the data were collected after the speeding tickets were given, include every ticket issued by police officers during the sample period with no prior knowledge by police officers, and thus avoids the problem of post-lawsuit data.[5]

II. Data source and descriptive statistics

Lafayette, Louisiana is a city in southern Louisiana with a population of 120,623, about 60 miles west of Baton Rouge (Census 2010). About 64% of Lafayette residents

---

[2] Data collection typically begins after the lawsuit is filed.

[3] Due to this potential change in police behavior, studies which employ post-lawsuit data provide a lower-bound estimate of the extent of racial/gender profiling. That is, if police officers change their behavior in order to avoid punishment or stigma, the results obtained from the analysis of post-behavioral change data will reflect a lower amount of racial or gender bias than truly exists.

[4] A related problem is the omitted variables problem, where police may observe something which the researcher cannot: behavior of the driver, for example. This problem exists for both the outcome approach and the benchmarking approach insofar as this omitted variable is related to race or gender.

[5] Even without a formal lawsuit it is important that police had no suspicion that this type of study might occur, so that they would not intentionally alter their behavior. There is no history of legal action taken against the police department in Lafayette, however, the issue of racial profiling within Louisiana has been of interest to the media after the period of data used in this analysis. For instance, a 2009 report by the American Civil Liberties Union claims there is widespread racial profiling in Louisiana, and House Representative Rickey Hardy of Lafayette (unsuccessfully) pushed a bill requiring police to track the race of individuals stopped for traffic violations in 2010 (Pierce 2010). No substantial policy or news changes came as a result of either of these publications.

are white and about 31% African-American. Lafayette encompasses five zip codes, 70501, 70503, 70506, 70507, and 70508. Each of these areas has quite different characteristics. Specifically, 71.8% of 70501 residents are African-American, as opposed to 70503 and 70508, where less than 11% of residents are African-American (Census 2010). The gender composition throughout the city does not vary significantly between zip codes, ranging from 47.7% male to 48.9% male (Census 2010). However, income disparity seems to follow a similar pattern as the city's racial composition. Per capita income in the northern area of the city, where there are many more African-American residents, is the lowest, at $15,491, while in the wealthiest areas (70508 and 70503) it is nearly $38,000 (Census 2007-2011).

Lafayette began implementing automated speed cameras in October 2007, with the help of Redflex, the company who created and helps to run these programs across the U.S. and Australia. The dataset is compiled of speeding tickets given by the automated cameras and all speeding tickets given by the Lafayette Police Department. Specific details of the data and how they were collected are in the following subsections.

A. Lafayette City police issued tickets

The Lafayette City Court database contains every misdemeanor ticket given by an officer in the Lafayette police department within the city limits.[6] The database includes information on the ticketed individual (date of birth, race, gender), the badge and name of the police officer who wrote the ticket, time, place, legal speed limit, and speed traveled at the time of the violation.[7] Police officers use discretion in issuing speeding tickets, but do not have any influence over the fine charged for a specific violation. Instead, Lafayette City Court sets fines. This is vital in understanding potential discrimination in this setting, especially in reference to existing research where police motives in issuing tickets may also affect the fine amounts (Makowsky and Stratmann 2009, for example).

Previous studies have utilized officer variation to aid in identifying the type of discrimination occurring, if it is occurring, however, the majority of officers in the

---

[6] In the Lafayette City Court computer database, speeding violations are specifically coded as 86-incident number. When a speeding ticket is reduced to a lesser charge, it is coded as a speeding ticket amended to something else (seatbelt violation for example). Tickets given outside of the city limits or given by State Troopers in the city limits are not in this database.

[7] Information specific to the offender is taken from the driver's license and officer observation. More specifically, name, gender, age, and home address are printed on Louisiana licenses, but race is not. Officers input driver race when they issue a ticket, which is also input into the computer system and provided in the dataset.

Lafayette Police Department are white males (Antonovics and Knight 2009 and Anwar and Fang 2006). Even more strikingly, less than 3% of tickets were given by officers who are not white males. Due to lack of variation of officer characteristics it is not useful to control for the officer's race or gender in the analysis.

B. Automated tickets

Lafayette Consolidated Government, and not the police department, decided to implement the Redflex program[8] and oversee its technology in an attempt to improve traffic safety. The speed cameras are available in two forms: a fixed camera at traffic lights to catch both speeders and vehicles that run red lights, and also in "speed vans" which park at different locations throughout the city to catch speeders. Tickets issued by the fixed cameras will not be included in the present analysis because they likely induce changes in speeding behavior for two main reasons: these locations are publicized and well known so drivers will intentionally drive safer to avoid being fined[9] and they are at intersections where drivers likely are more cautious as opposed to other stretches of the same road.

The program began in October 2007 with two speed vans giving citations at about 35 different locations in Lafayette. Though the automated ticketing system still continues today, the sample period used in this paper only extends to February 2008. Over the sample period, October 2007 to February 2008, the speed vans gave citations at 64 different locations. The Department of Traffic and Transportation, a department within Lafayette Consolidated Government, determined acceptable locations from accident statistics and individual requests for vans to be placed in specific areas with a speeding problem. Once the requested locations were verified to be safe for a van location, they were added to the list, and continue to be added and removed over the entire sample. On a particular day and at specific times, the vans are told to locate at randomly selected locations from the overall list, but the public is not informed of these decisions.

Outwardly, speed vans provide a close comparison to police cars (officers). Speed vans move around Lafayette randomly and individuals cannot predict their locations, nor are they significantly easier to identify than a police car. Therefore, drivers should behave in the same manner around police cars and speed vans. In

---

[8] The police department did not take control of the program until months after the sample period considered for this analysis.

[9] As in Bar-Ilan and Sacerdote (2004), where they find individuals do alter behavior in order to avoid an increase in a fine for running a red light. It is not hard to imagine this same behavior in order to avoid a speeding ticket.

opposition to police, cameras on the vans are completely automatic, and take photographs whenever they detect a car that is traveling faster than the speed limit. As soon as the cameras detect a speeder, four photographs are taken: one of the driver, one of the car's license plate, and two of the general area of the car at the time of the violation. Once an individual has been "caught" by the speed cameras, a paper ticket is issued to the car's registered owner (the assumed driver of the car). The automated database contains every ticket given by speed vans. Lafayette Consolidated Government officials estimate that about five to ten percent of the time, the person driving is not the car's registered owner. When someone is issued a ticket, but they were not actually driving, they have two options: pay the ticket anyway, or refute the ticket by naming the actual driver of the car. When a ticket is refuted, it is reissued to the individual who was named as the driver. It is more common for individuals to just pay the ticket instead of arguing, especially instances where a young person was driving a parent's car.[10]

The information available from the automated tickets is: name and home address of the registered owner of the vehicle, location, time and date of the ticket, legal speed limit, and speed traveled. Gender and race can be inferred from the four pictures on each ticket, most importantly, two of the driver.[11] Since automated tickets are easier to give and require less manpower, they are issued much more frequently than police tickets. During the period of October 2007 to February 2008 the average number of automated tickets is 3,100 per month.

C. Data

The sample includes speeding tickets issued between 6:00 A.M. and 6:59 P.M. from October 2007 to February 2008. The police portion of the data includes every ticket issued by a Lafayette city police officer within the city limits. Since the number of automated tickets had to be handled record by record, and each individual's characteristics had to be manually inferred, a 15% random sample was chosen from the population of automated tickets. Because of little or no visibility of individual drivers at night, only daytime tickets are used in the main analysis so that race and

[10] The information in the preceding paragraph was provided through personal communication with Tony Trammel, Director of the Department of Traffic and Transportation. Instances when a ticket was refuted can be observed in the data because a letter is added to the citation number every time the ticket is contested and reassigned. This occurs rarely, in about 7% of the sample.

[11] One is a close up of the driver's seat, the other taken from a further distance which has the entire front of the car in view.

gender can be identified. In a later analysis, a longer time period of police-issued tickets are utilized, to take advantage of differences in visibility in a similar manner to Grogger and Ridgeway (2006).

Table 1 lists descriptive statistics of all ticket data. About 23% of ticketed drivers are African-American and 48% are female. Since the socio-economic characteristics of some of Lafayette's zip codes are drastically different and driving behavior also may differ from area to area, much of the analysis controls for the zip code where the ticket was issued. The majority of tickets are issued in 70506, with nearly one-third issued in the poorest zip code (70501). The average ticketed driver was traveling about 52 miles an hour, with 72% of ticketed drivers speeding between 5 and 15 miles over the legal limit.

To provide a sense of the differences between tickets given by police and the automated system, Table 2 lists descriptive statistics broken down by zip code and source of ticket. The subjective nature of police-issued tickets means that some variables will differ by source: police can only ticket a subsample of the population of speeding drivers, and as such target higher speeders, in lower speed zones (often neighborhoods). For instance, the variables which measure how fast an individual was traveling (*Less than 10 Miles Over, 11-15 Miles Over, 16-20 Miles Over and More Than 20 Miles Over*) illustrate an important difference between the automatically issued tickets and police tickets: the majority of automated tickets are issued at lower severities of speeding.[12] Conversely, most police issued tickets are given in the *16-20 Miles Over* range. Individuals who receive tickets for higher speeds must pay a higher fine,[13] which results in higher revenues for the City of Lafayette, and in turn, likely a higher budget for the police department (Makowsky and Stratmann 2009).

These differences may arise in part because of the different costs faced by automated cameras versus police in issuing tickets. The automated cameras can easily issue tickets to every car that passes, but police must spend time to issue a ticket, and while issuing tickets they must let other speeders pass unpunished. However, it is important to remember that these differences do not violate the underlying assumption: the automated sources capture the population of speeders whereas the police are ticketing a subsample of that population. Further evidence of the validity of this assumption will be provided in Sections III and VI.

---

[12] Though, note that neither police officers nor the automated system issue tickets to speeders traveling 5 miles or less over the speed limit.

[13] Lafayette City Court bases fines on the severity of the speeding violation, however, individuals who have received prior traffic violations or committed the violation in a school or construction zone will have higher fines all else equal.

Table 1. Definitions and descriptive statistics

Variable	Definition	Observations	Mean	Standard deviation
Police	Dummy Variable (=1) if the ticket was given by a police officer, 0 otherwise.	1,914	.49	.50
Automated	Dummy Variable (=1) if ticket was given by an automated camera, 0 otherwise.	1,914	.51	.50
African-American	Dummy Variable (=1) if the ticketed driver was African-American, 0 otherwise.	1,663	.23	.42
Female	Dummy Variable (=1) if the ticketed driver was female, 0 otherwise.	1,681	.48	.50
70501	Dummy Variable (=1) if ticket was given in zip code 70501, 0 otherwise.	1,896	.32	.47
70503	Dummy Variable (=1) if ticket was given in zip code 70503, 0 otherwise.	1,896	.10	.30
70506	Dummy Variable (=1) if ticket was given in zip code 70506, 0 otherwise.	1,896	.39	.49
70508	Dummy Variable (=1) if ticket was given in zip code 70508, 0 otherwise.	1,896	.19	.40
HalfMth 1	Dummy Variable (=1) if violation was given in the first half of the month, 0 otherwise.	1,914	.52	.50
RushHour	Dummy Variable (=1) if violation was given between 7:00 and 8:59 AM or 5:00 and 6:59 PM, 0 otherwise.	1,914	.35	.48
Legal Speed	The speed limit where the ticket was given.	1,892	38.63	10.89
Less than 10 Miles Over	Dummy Variable (=1) if the driver was traveling 10 miles or less over the limit, 0 otherwise.	1,892	.30	.46
11-15 Miles Over	Dummy Variable (=1) if the driver was traveling 11-15 miles over the limit, 0 otherwise.	1,892	.42	.49
16-20 Miles Over	Dummy Variable (=1) if the driver was traveling 16-20 miles over the limit, 0 otherwise.	1,892	.22	.42
More Than 20 Miles Over	Dummy Variable (=1) if the driver was traveling 21 or more miles over the limit, 0 otherwise.	1,892	.05	.22
Weekday	Dummy Variable (=1) if the ticket was issued on a weekday (M,T,W,T,F), 0 otherwise.	1,914	.81	.40
Speed Trav	The speed the driver was traveling when given a ticket.	1,892	51.72	10.24

Table 2. Means and standard deviation, by area and ticket type

	70501		70503		70508		70506	
	Police	Automated	Police	Automated	Police	Automated	Police	Automated
African-American	.35**	.27	.07	.14	.15	.13	.21	.22
	(.47)	(.44)	(.26)	(.35)	(.36)	(.34)	(.41)	(.41)
	[331]	[201]	[15]	[145]	[216]	[112]	[346]	[279]
Female	.50**	.32	.60	.47	.54	.55	.58**	.36
	(.50)	(.47)	(.51)	(.50)	(.50)	(.50)	(.49)	(.48)
	[332]	[205]	[15]	[146]	[213]	[110]	[349]	[293]
Legal Speed Limit	27.52**	48.70	31.54**	40.37	36.27**	38.34	30.81**	49.16
	(5.59)	(4.99)	(5.55)	(8.29	(3.74)	(9.48)	(7.47)	(7.77)
	[328]	[270]	[13]	[174]	[212]	[151]	[343]	[388]
Less than 10 Miles Over	.01**	.29	.23**	.67	.02**	.77	.03**	.62
	(.10)	(.45)	(.44)	(.47)	(.15)	(.42)	(.18)	(.49)
	[328]	[270]	[13]	[174]	[212]	[151]	[343]	[388]
11-15 Miles Over	.38	.63	.54*	.29	.37**	.19	.61**	.32
	(.49)	(.48)	(.52)	(.45)	(.48)	(.40)	(.49)	(.47)
	[328]	[270]	[13]	[174]	[212]	[151]	[343]	[388]
16-20 Miles Over	.51**	.07	.15**	.03	.45**	.02	.31**	.05
	(.50)	(.26)	(.38)	(.18)	(.50)	(.14)	(.46)	(.21)
	[328]	[270]	[13]	[174]	[212]	[151]	[343]	[388]

Table 2. (continued) Means and standard deviation, by area and ticket type

	70501		70503		70508		70506	
	Police	Automated	Police	Automated	Police	Automated	Police	Automated
More than 21 Miles Over	.10**	.01	.08**	.01	.15**	.01	.05**	.01
	(.02)	(.01)	(.28)	(.08)	(.37)	(.11)	(.22)	(.10)
	[328]	[270]	[13]	[174]	[212]	[151]	[343]	[388]
Weekday	.99**	.70	.80	.72	.95**	.80	.99**	.5
	(.08)	(.46)	(.41)	(.45)	(.21)	(.40)	(.12)	(.50)
	[333]	[270]	[15]	[174]	[216]	[151]	[349]	[388]
Speed Trav	43.94**	60.93	45.00*	50.02	53.00**	47.23	45.79**	59.35
	(5.58)	(6.23)	(5.20)	(9.50)	(4.54)	(10.28)	(8.37)	(9.47)
	[328]	[270]	[13]	[174]	[212]	[151]	[343]	[388]
Half Month1	.44	.45	.80*	.57	.47	.54	.57	.55
	(.50)	(.50)	(.41)	(.50)	(.50)	(.50)	(.50)	(.50)
	[333]	[270]	[15]	[174]	[216]	[151]	[349]	[388]
RushHour	.66**	.28	.13	.29	.09**	.32	.38**	.30
	(.48)	(.45)	(.35)	(.45)	(.28)	(.47)	(.49)	(.46)
	[333]	[270]	[15]	[174]	[216]	[151]	[349]	[388]

Note: Standard deviations are in (parentheses). The number of observations is in [parentheses]. * denotes a significant difference between the automated and police means at a 10% level, ** denotes significance at a 5% level.

III. Validity of automated tickets as a measure of the population

In order for the automated issued tickets to provide a valid comparison group to police issued tickets, both ticketing sources must observe the same driving (speeding) population. Police witness the population of speeders in a given location, but are only able to ticket a select number, while the automated cameras ticket the entire population of speeders in that location objectively. If police do not observe the same population, any difference in ticketing may be the result of the different population of speeders and not due to a difference in ticketing behavior. There are some procedural differences that need to be considered, but overall, the populations being measured are shown to be comparable. I first provide convincing descriptive evidence below, and then in Section VI more explicitly account for endogeneity concerns with exploitation of police visibility using daylight savings time.

The first step to show the equivalence of the police-observed population and the automated-observed population is to understand the locating procedures of both ticketing sources. If police have the freedom to patrol where they please, they may choose to target areas where certain groups travel. For example, if police have a preference for ticketing African-Americans, and locate where more African-Americans travel, more African-Americans will receive tickets. If the automated tickets are not given in those specific areas, the number of tickets issued to African-Americans by police would be higher in comparison to automated tickets in other areas, but this would reflect the differential exposure rates, not police discrimination.[14]

In the case of tickets issued by police, the data only specify the location of the violation, but not how or why the officer was located there. Importantly, officers are told where to locate according to precincts in Lafayette, generally: north, south, east or west. More specifically, when complaints have been filed about speeders in specific neighborhoods or areas within this distinction, traffic officers are told to

---

[14] Another scenario may initially seem plausible as well, motivated by the difference in means of speed limit by ticketing type, as seen in Table 2. Since automated cameras ticket on streets with a higher average speed limit than police, perhaps these automated cameras are being placed on busier roads used for commuting, while police are locating in neighborhoods and school areas, where there are other safety concerns besides speeding. If this is the case, and women and African-Americans are more likely to travel in neighborhoods, while men and whites are more likely to travel on the busy commuting routes, then the results herein are being driven by this fact and not police discrimination. This scenario cannot be the driving force of these results however, because the neighborhoods and school zones where police are locating are public schools with a majority of white students, and white neighborhoods. Therefore, if different ticketing populations were the true source of the differential ticketing, whites would receive more tickets from police than automated sources, the opposite of the present findings. Though there is not as simple of an explanation regarding gender, it is unlikely that this type of selection could be driving the entire result.

focus on these areas for the duration (or the majority) of their shift.[15] There is always an officer in each area of the city.[16] Therefore, how police are located to give tickets should not be influenced by preferences to ticket a specific type of individual, because they are told in which areas to locate for each shift.

Although the mobile automated cameras are randomly assigned to a location during the day, the locations themselves are not completely random. First of all, Redflex states that its mobile cameras can be used, "on suburban streets, as well as on higher-speed thoroughfares, either by parking in a safe position on the roadway or nearby for added safety" (Redflex 2010). Since "safe" locations include different types of roads, this should not cause any problem in comparing to police issued tickets since it is feasible that police will also search for speeders in a "safe" spot, despite the fact that this is not explicitly stated in police procedure.

The other source of non-randomness in speed van locations is that the initial acceptable list comprised areas known to have speeding problems; and as such, tended to be busier streets instead of neighborhood roads. Similarly, because the goal of this program was to reduce speeding, the areas that would have the most impact on speeders tended to be busier city streets, as compared to neighborhood roads. This can be seen in Table 2, where the majority of tickets issued by automated sources are issued on streets with relatively high speed limits. Over time, because individuals could request a van be placed in their neighborhood, these neighborhood locations were added to the list, but the number of tickets issued on busier streets is much larger than the number of tickets issued on streets with lower legal speed limits.

Police also locate on busy streets, but they tend to focus more on ticketing speeders in neighborhoods, and specifically near schools. In school zones, the legal speed a car can travel is much lower than larger city streets. This is one reason why the average speed limit for police issued tickets is less than the mean speed for automated issued tickets. Police locate in neighborhoods, but generally on streets with high traffic volume; streets with low speed limits that are used by a large number of travelers. This does not affect the validity of the comparison, because vans locate nearby these same areas.[17] The next section empirically investigates this claim.

---

[15] Though the data do not specify the difference between the two existing types of officers (traffic and patrol), it is obvious that the officer on duty was sent specifically to target speeders when he/she gives numerous tickets in the same location in a short period of time. Traffic officers issue the most speeding tickets, on occasion a patrol officer will observe someone speeding in their area, and give a ticket.

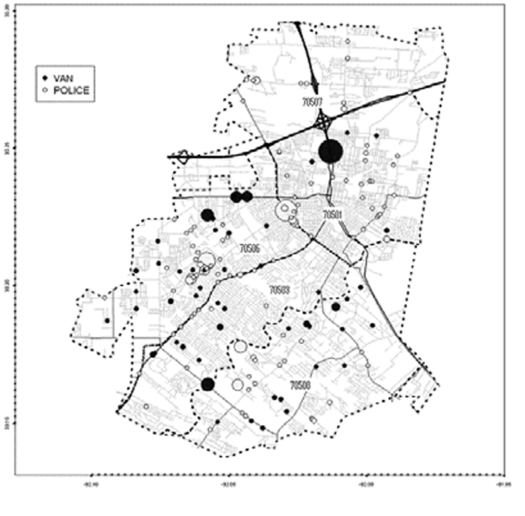
[16] The Lafayette Police Department provided the information in the preceding paragraph through personal communication.

[17] When school zones are excluded from the analysis, the police coefficient is actually larger than before.

The ideal measure of the speeding population that police observe would be to consider drivers at the exact locations where police issue tickets, but this is not feasible for multiple reasons. The most obvious of these reasons is that if automated sources and police chose to locate at the exact same locations, they would not be maximizing speed-deterrence. If a police officer is traveling to a designated spot to target speeders, and upon arriving sees a mobile van, he/she will most likely travel to a nearby street, or nearby block. In the sample, as can be seen from Figures 1 and 2, there are some instances where an automated camera and police officer ticketed a speeder in the exact same location, however, it is more common for tickets to be issued nearby, generally within a block or two. This does not create a bias, because individuals who drive in neighborhoods also must drive on the busier city streets where vans are located nearby.

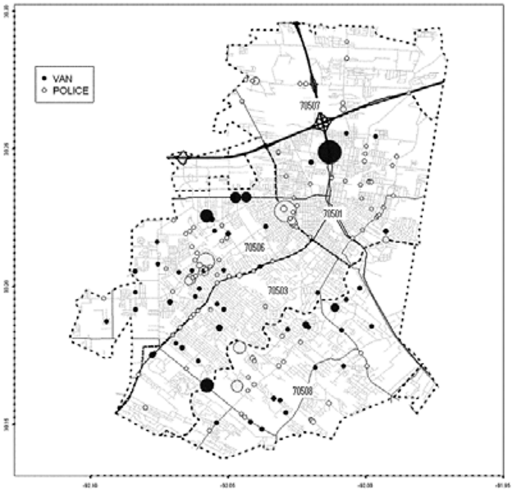
Figures 1 and 2 show the city of Lafayette, with dots representing the frequency of tickets issued by each ticketing source, at specific locations. The dots are sized proportionately to the frequency of tickets that were issued at that location.[18] For example, in many instances only one ticket is issued in a location and these dots are the smallest on Figure 1.

Figure 1. Tickets used in the race estimation sample



[18] Size of the bubbles was determined based on the equation: Size = (Frequency of Tickets Issued / Maximum Frequency of Tickets Issued at One Location).

Figure 2. Tickets used in the gender estimation sample



The western portion of the map, which includes zip codes 70506 and 70503, illustrates a fairly equal coverage of mobile vans and police officers. Since there are automated vans and police officers in near proximity to one another (and, generally on similar road types), it is feasible to assume that both ticketing sources are observing the same population of speeders, when controlling for time of day, and other incident characteristics. While there is a greater discrepancy between police and automated ticket locations of the remaining zip codes, 70501 and 70508, tickets are still issued within blocks of each other. The vans and police officers issue tickets in the same neighborhoods, or a police officer may issue tickets within a neighborhood while a van issues tickets on a nearby street where those residents must travel to get home. Additional robustness checks in Section VI provide supportive evidence that the automated tickets remain a valid measure of the speeding population in these areas as well.

The area encompassing zip code 70507, north of Interstate 10, is excluded from the analysis because no speed van tickets were issued in this area during the sample period. Without an accurate measure of the speeding population, a reliable analysis cannot be conducted. This area of the city does not include the main commercial areas and only 77 police-issued tickets are dropped as a result of this exclusion.

One potential data issue that is not present in other literature arises because Lafayette is a relatively small city, where the majority of officers are white males. If police officers happen to stop individuals they know personally (e.g. another white male), and let them go without a ticket, the results may create an impression of race or gender bias when it is actually a result of corruption, based on personal relationships. Even if this was the case, the effect should be minor since the city is large enough that police officers do not know everyone. Also, the magnitude of the results here are substantial enough that it is unlikely that they are driven by this type of behavior.

In a novel paper, Eeckhout, Persico, and Todd (2010) explore the practice of random crackdowns by police and find that these crackdowns are efficient for crime reduction in the context of automated traffic monitoring. One important distinction from the current work is that individuals in Lafayette are not notified of which roads the automated vans will be located: the "crackdowns" are unknown. Though the use of automated cameras likely reduces speeding on the whole, as found in Eeckhout, Persico, and Todd (2010), it should not impact the validity of comparing police to automated tickets since both observe this reduction simultaneously.

IV. Methods

The ideal way to investigate if police give speeding tickets differentially based on gender or race is to have information on the entire population of speeders, then to compare the population of speeders with those who are ticketed. If the racial and gender composition of speeders who are ticketed by police is different than the racial and gender composition of the entire population of speeders, police are treating individuals differently based on gender and/or race. However, observing the entire population of speeders is costly, and nearly impossible when looking at the speeding behavior of a whole city. The automated ticket system provides a measure of the speeding population in a given location, since the cameras ticket all speeding cars in that area. This also provides an advantage over previous literature, where population measures are not completely objective.[19] If police do not consider race or gender when they issue tickets, then the proportion of tickets issued to certain sub-groups of the population (such as females or African-Americans) should not differ between tickets issued by police and tickets issued by vans.

I will use individual level tickets to investigate police behavior in issuing speeding tickets. Thus, the estimation will pose the question: Given that the driver was caught speeding and issued a ticket, is the probability of being black (or female) the same regardless of the ticketing source, that is:

[19] For example, Grogger and Ridgeway (2006) use tickets issued at night as a population measure, but police can likely still observe car type, which may be correlated with race. Therefore, this may not be a completely objective measure of the population.

PR(*Black*|*Ticket*,*Police*) = PR(*Black*|*Ticket*,*Automated*)?

The analysis will be performed at the individual level, with the dependent variable a dummy equal to 1 if the ticketed individual is African-American and 0 otherwise (or female/male). The advantage of the individual-level analysis is that the richness of the data will allow for control of most factors that police may use to decide whether or not to ticket an individual, such as severity of the speed violation, the speed limit where the ticket was given, as well as other determinants of ticketing, which include the day of the week, and the location of the infraction. The specification is depicted by Equation (1)

$$B_i = \alpha + X_i'\beta + \gamma Pol_i + \varepsilon_i\,, \tag{1}$$

where B_i is equal to 1 if the recipient is black, and zero otherwise (or equal to 1 if the recipient is female and 0 otherwise), X_i includes specific characteristics of the violation, and Pol_i is a dummy variable equal to 1 if the ticket was given by a police officer and 0 otherwise (if the ticket was given by an automated source). In this specification, if the coefficient of the dummy variable for a police-given ticket (γ) is positive and statistically significant, this implies that race (or gender) does play a role in a police officer's decision to pull over and ticket a speeder.

V. Results

Table 3 provides regression results for the whole sample of tickets as well as for subsamples based on street-type where the ticket was issued. The entries are marginal effects; and robust standard errors, clustered by zip code, are reported in parentheses. All columns control for zip code fixed-effects, whether the ticket was given in the first half of the month, whether the ticket was issued during rush hour in the morning or evening, the legal speed limit where the ticket was issued, severity of the speeding violation (*11-15 Miles Over, 16-20 Miles Over, and More than 20 Miles Over*), whether the violation occurred on a weekday, and fixed effects for time blocks during the day.[20]

---

[20] Estimates for each zip code individually can be provided upon request. There is no significant effect of police for African-American tickets, but it is more likely that a ticketed individual will be female if the ticket was issued by a police officer for one zip code. Similarly, if no area controls are included the police coefficient is not statistically significant for African-Americans, but remains significant when considering gender.

Table 3. Probit marginal effects by street type

	Dependent Variable: African-American			Dependent Variable:Female		
	Entire Sample	Main City Streets	Neighborhood Streets	Entire Sample	Main City Streets	Neighborhood Streets
Police	.077**	-.014	.040	.135**	.195**	.338**
	(.025)	(.076)	(.091)	(.056)	(.096)	(.121)
70503	-.143**	-.080	-.279***	.106**	.080	.209
	(.007)	(.047)	(.033)	(.025)	(.074)	(.128)
70506	-.091**	-.007	-.256**	.054**	.001	.118
	(.007)	(.042)	(.103)	(.009)	(.053)	(.133)
70508	-.144**	-.024	-.142	.119**	.052	-.055
	(.009)	(.055)	(.065)	(.021)	(.073)	(.155)
HalfMonth 1	-.010	.024	.005	.025	.012	-.009
	(.023)	(.031)	(.048)	(.022)	(.040)	(.059)
Rush Hour	.009	.035	-.017	-.061	-.125	.026
	(.069)	(.064)	(.050)	(.071)	(.077)	(.063)
LegalSpeed	.002**	.001	.001	-.004**	-.009	.000
	(.001)	(.004)	(.005)	(.001)	(.006)	(.008)
11-15 Miles Over	.004	.078**	-.088	-.054	-.060	-.267**
	(.023)	(.038)	(.099)	(.051)	(.046)	(.128)
16-20 Miles Over	.011	.004	-.091	-.079	-.038	-.217
	(.028)	(.063)	(.090)	(.052)	(.076)	(.138)
More than 20 Miles Over	.001	.095	-.168*	-.120	-.129	-.060
	(.034)	(.109)	(.060)	(.079)	(.101)	(.195)

Table 3. (continued) Probit marginal effects by street type

	Dependent Variable: African-American			Dependent Variable:Female		
	Entire Sample	Main City Streets	Neighborhood Streets	Entire Sample	Main City Streets	Neighborhood Streets
Weekday	-.064	-.057	-.057	.007	-.042	.070
	(.042)	(.037)	(.136)	(.025)	(.045)	(.163)
9:00-11:59 AM	-.018	-.029		-.108	-.154	
	(.098)	(.069)		(.126)	(.091)	
12:00-2:59 PM	-.034	.001		-.085	-.150	
	(.097)	(.072)		(.102)	(.091)	
3:00-5:59 PM	-.037	-.030		-.070	-.090	
	(.069)	(.060)		(.074)	(.082)	
6:00-6:59 PM	.236**	.258**		-.088	-.031	
	(.106)	(.119)		(.206)	(.114)	
N	1628	697	340	1646	711	343
ln L	-836.11	-342.12	-164.58	-1105.99	-460.77	-229.24
BIC	1694.4	788.98	399.10	2234.20	1026.60	528.54

Note: The reported values are the marginal effects, estimated using individual-level data. Robust standard errors, clustered by zip code, are in parentheses. * denotes significance at a 10% level, and ** denotes significance at a 5% level. Time block was excluded from neighborhood regressions because if included 6:00 to 6:59 perfectly predicted driver race and gender. Even when included results are qualitatively the same.

Columns I and IV include all zip codes except for 70507, where no automated van tickets are issued. However, the remaining columns restrict the sample based on the type of street where the ticket was issued to better control for potential differences in exposure to police and differences in driving populations. *MainCityStreets* are comprised of tickets issued on major two to four lane roads within Lafayette: Johnston Street, Ambassador Caffery Parkway, Kaliste Saloom Road, and Pinhook Road. The regression results using the sample including clusters of police and automated tickets issued on *NeighborhoodStreets* are presented in Columns III and VI.

The police coefficient is positive and significant in the first column, where *African-American* is the dependent variable, implying that the probability of being African-American is higher if the ticket was given by a police officer than if it was given by an automated source. However, when controlling more specifically for road type the police coefficient is no longer significant.

When *Female* is the dependent variable, the police coefficient is positive and significant for all regressions, though this impact is larger for tickets issued on neighborhood roads (.34 as opposed to .195 for city streets). The police coefficient of the estimates utilizing the entire sample indicates that conditional on being issued a ticket, the probability of a speeding ticket being received by a female is nearly 14 percentage points higher when the ticket was issued by a police officer.[21] Anecdotally, individuals tend to believe that men receive more tickets than women,[22] but the data illustrate that this is not the case. Blalock et al. (2007) also found that women were more likely to be ticketed than men in three of five study locations.

In order to identify differences in ticketing by source, I control for a number of variables important for driving patterns and population as well as violation severity. Zip code fixed effects are important if the racial or gender makeup of those areas are vastly different, based on living areas or travel patterns. These are most important for race, as seen in the negative coefficients, since compared to 70501 relatively fewer African-Americans are receiving tickets. *HalfMonth 1* is added to test conventional wisdom that police ticket differentially depending on the time of month, but is not significant in any specification. In a similar vein, *RushHour* and

[21] One concern is that *70506* and *70503* may be driving these results. However, even when these zip codes are excluded, the coefficient on the police dummy is smaller, but still significant (.044 at a 5% level). These zip codes include commercial as well as residential areas, similar to the other zip codes in this analysis, so it is unclear why there would be a difference in ticketing based on gender in the area.

[22] For example, in a college-student survey, Blalock et al. (2007) found that the majority of respondents believed a man would more likely be ticketed than a woman if both were stopped for speeding 12 miles over the limit.

Weekday should capture any differences in driving population differences by time of day or work patterns, but both are insignificant in all specifications.

Legal speed where the ticket is issued acts as a proxy for road type, which better explains why it is significant only when estimated using the entire ticket sample. The subsequent columns explicitly control for road type, rendering this control insignificant. The controls for severity of the violation are a range of dummy variables (*11-15 Miles Over, 16-20 Miles Over, More than 20 Miles Over*) which are equal to one if the violation was within the range and 0 otherwise. These controls are not consistently significant in any specification.[23]

Since driving patterns may differ by race or gender based on the time of day the ticket was issued (Grogger and Ridgeway 2006, Blalock et al. 2007), specific time of day variables are also included in the main specification. The additional hour dummy variables are: *6:00 to 8:59 AM, 9:00 to 11:59 AM, 12:00 to 2:59 PM, 3:00 to 5:59 PM, and 6:00 to 6:59 PM,* but are not consistently significant across specifications.

The same analysis was performed utilizing police precincts instead of zip codes as the area identifier, where the police coefficient is still positive, significant, and nearly the same magnitude as earlier estimations. Controlling for police precincts should reduce any impact of differential exposure to police based on area of patrol (since police patrols are assigned according to precincts). This is especially important given the importance of location in an officer's decision to stop a driver (Sanga 2014). Results are provided in Appendix Table A2.[24]

These results provide more evidence that police consider gender when issuing speeding tickets, but introduce doubt into the likelihood of police consideration of

---

[23] A question may arise as to how the severity of the violation impacts these results. If the ticketed drivers included in the regression sample are restricted to those traveling 11 miles or more over the limit, the majority of the results remain consistent to estimations including the entire sample of ticketed drivers. However, the police coefficient is no longer significant when the dependent variable is *African-American*. It is not necessarily the case that there will be greater evidence of racial and/or gender bias when considering these more dangerous speeders. In fact, as cited in other existing literature (Blalock et al. 2007), there is some evidence supporting the idea that police discrimination decreases as the offense severity increases; in this case, for example, for more severe crimes police are less willing to focus on gender and instead the desire to punish the bad behavior regardless of the offenders' gender strengthens.

[24] All specifications were also run including interaction terms between police and zip code dummies. The overall results are consistent, though the police marginal effect is stronger in 70501 and 70506 for both gender and race estimates. Similarly, including both automated sources does not change the overall finding that the probability of a ticketed speeder being a woman or African-American is higher for tickets issued by police officers, in any specification. Lastly, specifications were run excluding tickets issued in December, since this month may be different due to holidays, etc. The results are unchanged. Results for any of these robustness checks can be provided by request.

race in the manner discussed thus far. Due to this contrary finding, the next section further investigates endogeneity by abandoning the assumption that the automated cameras are a valid comparison, and instead, following a procedure similar to Grogger and Ridgeway (2006) where the population depends on daylight, but only utilizes one ticketing source at a time.

VI. Investigating endogeneity by utilizing daylight savings time

Next, I explore a slightly different approach to support the previous analysis. Grogger and Ridgeway (2006) estimate the population at risk of being stopped by police by comparing the racial distribution of drivers stopped at night to the distribution stopped during the day. This analysis relies on a concept they call the “veil of darkness.” During the daytime, when race is visible, it is possible that police use the race of the driver as a determinant of whether or not to stop a car. At night it is unlikely that police can distinguish between different races, and therefore presumably make traffic stops based on actual offenses without regard to the driver’s race. A direct comparison of the two distributions assumes that driving patterns, driving behavior, and police exposure are the same during the day and night, but by exploiting information from daylight savings time, they are able to control for driving patterns while still comparing tickets issued in the dark to those issued when the sun is up. Individuals’ work schedules as well as police patrol schedules, differ by time of day and not by darkness. They find no evidence of racial profiling.

Following the logic of Grogger and Ridgeway (2006), I restrict the estimation sample to police-issued tickets between 6:00 AM and 7:59 AM, and between 5:00 PM and 6:59 PM. I supplement my dataset with sunrise and sunset data taken from the U.S. Naval Base. As a result of daylight savings time, some tickets are issued in the dark while some are issued in daylight, even though the clock time of the issued ticket is the same. In other words, in November the sun sets around 5:30 PM, but in October the sun sets around 6:30 PM. This means that someone who received a ticket in November at 6:00 PM received a ticket when it was dark outside, and the police officer likely could not see inside the vehicle (and thus, could not determine race or gender of the driver). However, if another driver was ticketed at 6:00 PM in October, when it was light outside, police officers could see inside the vehicle.

Assuming that police officers have no driver visibility and cannot observe race or gender when it is dark outside, any difference in issuance to African-Americans or women when it is light as compared to when it is dark implies that police officers do consider race or gender in issuing tickets. Utilizing daylight savings time allows for keeping time of day constant, while providing the ability to compare tickets issued in light to those issued in the dark. All other controls are the same as in Table 3.

The coefficient of interest is *Daylight Visibility*, which equals 1 if it is light outside (if the ticket was issued on that day after the sun rose and before it set), and 0 if it is dark outside (if the ticket was issued on that day before the sun rose or after it set).

Table 4 provides means and standard deviations of this new control variable in terms of gender and race, independently as issued by police and automated sources. Since automated cameras are assumed to measure the population of speeders at a given location, regardless of whether it is light or dark outside, we can compare the proportion of these tickets to those issued by police officers, to determine if there is a difference in issuing based on visibility.

Initially, if we look only at tickets issued during daylight hours, when drivers are visible to police, it is obvious that ticketing behavior is different between police and automated sources. Police issue a greater proportion of tickets to African-Americans as well as women, though this raw difference is only significant for gender. These rough results coincide with the earlier findings of this paper. Conversely, during dark hours when there is no visibility, the proportion of tickets issued to women and African-Americans by police and automated sources are very similar. Since this difference only arises when there is visibility of drivers, this implies that police are using some subjective criteria once observing the speeding driver to determine whether or not to issue a ticket.[25] Recall that only tickets issued between

Table 4. Daylight visibility: means and standard deviation of daylight controls

	=1, visibility		=0, no visibility	
	Police	Automated	Police	Automated
African- American	.285	.274	.267	.263
	(.452)	(.448)	(.458)	(.452)
	[263]	[106]	[15]	[19]
Female	.551*	.385*	.333	.474
	(.498)	(.489)	(.488)	(.513)
	[265]	[109]	[15]	[19]

Note: Recall that only a subset of police issued tickets are being used: those issued between 6:00 AM and 7:59 AM and those issued between 5:00 PM and 6:59 PM. Standard deviations are in (parentheses). The number of observations is in [parentheses].
* denotes a significant difference between tickets issued by police and those issued by automated sources, at a 5% level.

[25] Police may still infer gender or race based on the car model, type, or even color. Therefore, police may still be able to consider these factors, though to a lesser extent.

6:00 AM and 7:59 AM and 5:00 PM and 6:59 PM are included in these estimates, and so it is unlikely that these results are driven by differences in driving patterns. Though these statistics are extremely useful for analyzing trends in the raw data, a more thorough approach needs to be used to provide more reliable results.

The regression results including daylight controls are presented in Table 5, which support the previous results and imply that African-Americans and females are more likely to receive a ticket from a police officer only when race or gender is visible. If the same exercise is performed using only automated issued tickets, the coefficient on *Daylight Visibility* is not significant, as can be seen in Table 6. Since automated

Table 5. Probit marginal effects: investigating the effect of daylight on police-issued tickets

Variable	African-American	Female
Daylight	.110*	.297**
	(.057)	(.124)
70506	-.180**	.134**
	(.040)	(.038)
70508	-.161**	.049
	(.040)	(.177)
HalfMonth 1	-.032	.045
	(.026)	(.036)
Rush Hour	.019	-.334
	(.210)	(.188)
LegalSpeed	.008**	.014*
	(.003)	(.008)
11-15 Miles Over	-.120**	.137
	(.039)	(.110)
More than 16 Miles Over	-.124	.161
	(.100)	(.138)
Weekday	-.130**	-
	(.049)	
N	266	265
ln L	-151.60	-175.09
BIC	314.35	361.33

Note: The reported values are the marginal effects, estimated using individual-level tickets. Robust standard errors, clustered by area, are in parentheses. * denotes significance at a 10% level, and ** denotes significance at a 5% level. Weekday for column II drops out since 3 observations were weekend tickets issued to women.

sources are objective there should be no difference in ticketing by race or gender merely because it is light as opposed to dark.[26]

The coefficient on *Daylight Visibility* is significant when considering police issued tickets, where ticketed drivers are more likely women (African-American), but there is no difference for automated sources. These findings coincide with results when automated cameras are used as the comparison to police-issued tickets, providing supportive evidence that the automated cameras can be used as a valid comparison group.

Table 6. Probit marginal effects: investigating the effect of daylight on automated-issued tickets

Variable	African-American	Female
Daylight	-.117	-.005
	(.145)	(.167)
70503	-.126**	.366**
	(.019)	(.018)
70506	-.170**	.274**
	(.029)	(.014)
70508	-.085	.495**
	(.052)	(.032)
HalfMonth 1	.044	.249**
	(.061)	(.077)
Rush Hour	-	.173
		(.356)
LegalSpeed	.002	-.013**
	(.003)	(.005)
11-15 Miles Over	.055	-.134**
	(.095)	(.048)
Weekday	-.011	-.093
	(.172)	(.127)
N	119	126
ln L	-68.04	-69.68
BIC	150.42	153.87

Note:The reported values are the marginal effects, estimated using individual-level tickets. Robust standard errors, clustered by zip code, are in parentheses. * denotes significance at a 10% level, and ** denotes significance at a 5% level. Only one ticket over 16, which was excluded from the sample for estimation.

[26] This analysis can be performed by zip code, but the sample size for some are too small to estimate. However, those where the sample is large enough produce similar results as when aggregated. These results are available upon request.

VII. Conclusion

This paper aims to explain whether police issue speeding tickets differently to individuals based on their race or gender. I find that in the city of Lafayette, Louisiana, the probability of a ticketed driver being a woman is significantly higher if the ticket was issued by a police officer versus an automated source. The results are mixed when considering race of a driver, suggesting non-systematic consideration of race in issuing speeding tickets, though some evidence still exists that race is not completely ignored. Conversely, the results imply that gender plays an important role when police decide whether to ticket a speeding driver. Even when controlling for additional factors like type of road where the ticket is issued, the results remain the same.

This rich dataset has not been used previously to study police behavior and differential treatment in receiving speeding tickets based on gender and race.[27] As a result of the specific type of analysis, this paper does not suffer from common issues in this realm of literature. The city implemented the automated camera system to improve safety and decrease the number of crashes caused by red light runners, and was not intended for any use involving investigation of police bias. Also, these data were not collected as a result of a lawsuit, and therefore police had no incentive to alter their behavior. Another problem in some existing literature is the use of police reported stops, where not all stops are actually recorded. However, the present data set includes all speeding tickets given during the sample time period. Every instance when a police officer wrote a ticket is included and police cannot misreport their actions.

This paper also has a large advantage over existing literature because it employs a completely objective measure of the speeding population. For the most part, vans and police officers are located either very close to each other (on the same street or city block), or they are within a few blocks. This suggests that police officers and vans are not differentially located to deliberately target different sub-populations. This provides a distinct advantage in that after controlling for incident and street characteristics, any differences between automated and police issued tickets arise from the subjective nature of police tickets.

I employ numerous techniques to illustrate that the automated sources do provide a valid population measure to speeders observed by police. Suggestive evidence using maps of Lafayette and extensive regression controls for location and driver

[27] Another study mentioned in Grogger and Ridgeway (2006), done by the Montgomery County Police Department (2002), used photographic stoplight enforcement to measure the at risk population of speeders. However, this study could not be accessed, so it is uncertain how closely their methodology relates to the current work.

behaviors, as well as manipulation of daylight visibility all provide the same conclusion: police officers ticket a larger proportion of women than automated sources. In regressions explicitly controlling for type of road (city vs. neighborhood), drivers' race is not significant, however, in all other specifications it is. The gender effect is larger, and more consistent throughout all methods.

Despite the fact that I cannot determine whether the differential treatment is a result of preference-based discrimination or statistical discrimination, the results still illustrate some type of discrimination, which has potential welfare implications. For example, assume that police ticket African-Americans at a higher rate not because of a taste for discrimination, but because police believe that African-Americans are less likely to contest a speeding ticket. This would mean that higher penalties are levied on African-Americans than whites despite the fact that they have the same offending (speeding) intensity. Given that the incomes of African-Americans are less than half that of whites in this population of speeders,[28] this would constitute a regressive tax based on unequal treatment. Further research is necessary to investigate whether differential contesting rates (or likelihood of future offenses, etc.) can explain police behavior, or if preference-based discrimination is really the cause of the disparities between tickets issued by police officers and automated sources, but in any case, the welfare implications are similar.

References

Antonovics, Kate and Brian G. Knight (2009). A new look at racial profiling: evidence from the Boston police department. *The Review of Economics and Statistics* 91: 163-177.

Anwar, Shamena and Hanming Fang (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *The American Economic Review* 96: 127-151.

Argys, Laura M. and H. Naci Mocan (2004). Who shall live and who shall die? An analysis of prisoners on death row in the United States. *Journal of Legal Studies* 33: 255-281.

Bar-Ilan, Avner and Bruce Sacerdote (2004). The response of criminals and noncriminals to fines. *Journal of Law and Economics* 47: 1-16.

Becker, Gary S. (1957). *The Economics of Discrimination*, 1st Edition. The University of Chicago Press.

[28] Based on the zip code analysis previously discussed.

Blalock, Garrick; Jed DeVaro, Stephanie Leventhal; and Daniel H. Simon (2007). Gender bias in power relationships: evidence from police traffic stops. Cornell Working Paper.

Durlauf, Steven N. (2006). Assessing racial profiling. *The Economic Journal* 116: F402-F426.

Eeckhout, Jan; Nicola Persico; and Petra E. Todd (2010). A theory of optimal random crackdowns. *The American Economic Review* 100: 1104-1135.

Garrett, Thomas A. and Gary A. Wagner (2009). Red ink in the rearview mirror: local fiscal conditions and the issuance of traffic tickets. *Journal of Law and Economics* 52: 71-90.

Grogger, Jeffrey and Greg Ridgeway (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association* 101: 878-887.

Knowles, John; Nicola Persico; and Petra Todd (2001). Racial bias in motor vehicle searches: theory and evidence. *Journal of Political Economy* 109: 203-229.

Makowsky, Michael and Thomas Stratmann (2009). Political economy at any speed: what determines traffic citations? *The American Economic Review* 99: 509-27.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynn E. Browne, and James McEneaney (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review* 86: 25-53.

Persico, Nicola and Petra E. Todd (2007). The hit rates test for racial bias in motor-vehicle searches. *Justice Quarterly* 24: 741-757.

Pierce, Walter (2010). Hardy seeks racial profiling task force. *The Independent Weekly*, June. Accessed 8/5/10 <http://www.theind.com/news/6394-hardy-seeks-racial-profiling-task-force>.

Redflex Traffic Systems Official Website. Mobile Datasheet. Accessed 5/15/2010 <http://www.redflex.com/html/usa/solutions/REDFLEXspeed/datasheet_mobile>.

Sanga, Sarath (2014). Does officer race matter? *American Law and Economics Review* 16: 403-432.

U.S. Census Bureau, Census 2010 Summary File 1, Tables P5, P6, P8, P12, P13, P17, P19, P20, P25, P29, P31, P34, P37, P43, PCT5, PCT8, PCT11, PCT12, PCT19, PCT24, H3, H4, H11, H12, and H16. Accessed: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

U.S. Census Bureau, 2007-2011 American Community Survey. Accessed: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_11_5YR_B19301&prodType=table.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 29-47

TRIPLE PENALTY IN EMPLOYMENT ACCESS: THE ROLE OF BEAUTY, RACE, AND SEX

Francisco B. Galarza and Gustavo Yamada*
Universidad del Pacífico

Submitted July 2015; accepted August 2016

This article reports the results from a first experiment specifically designed to disentangle the effect of beauty from that of race in the observed labor market discrimination, for both females and males in Peru. We randomly assigned Quechua and white surnames and (subjectively perceived) attractive or homely-looking photographs (or no photos) to 4,899 fictitious résumés sent in response to 1,247 job openings. We find that candidates who are physically attractive, have a white-sounding surname, and are males, receive 82%, 54%, and 34% more callbacks for job interviews than their similarly-qualified counterparts, thus imposing a triple penalty on homely-looking, indigenous, and female job candidates. We further find that the intensity of discrimination by race and physical appearance differs for males and females; the intensity of discrimination by physical appearance and sex differs for Quechua and white applicants; and the intensity of racial and sexual discrimination differs for beautiful and homely-looking persons.

JEL classification codes: C93, J71
Key words: discrimination, labour market, beauty, race, sex, Peru

I. Introduction

Requests for fair employment opportunities take place everywhere, but they are particularly sturdy in mixed-race emerging countries, with large groups of descendants

---

* Francisco B. Galarza (corresponding author): Universidad del Pacífico, Department of Economics and Research Center (CIUP), Av. Salaverry 2020, Jesús María, Lima, Peru; email: galarza_fb@up.edu.pe. Gustavo Yamada: Universidad del Pacífico, Department of Economics and Research Center (CIUP), Av. Salaverry 2020, Jesús María, Lima, Peru; email: yamada_ga@up.edu.pe. We would like to thank participants at several seminars for helpful comments and suggestions. André Miranda, Chris Túpac and Renzo Maldonado provided valuable research assistance. We also thank two anonymous referees for valuable comments and suggestions. The research reported herein was partially funded by the Consortium of Economic and Social Research (CIES). All errors are our own.

from indigenous people and foreign migrants, such as Peru. The outstanding macroeconomic performance attained by this country over the last decade has not sufficed to achieve a substantial reduction in income inequality and other well-being indicators, a result that has awaken claims for redistributing the benefits from economic growth to those groups traditionally excluded from it, namely people living in rural areas, largely populated by indigenous people.

While there is a widespread belief that Peru is a discriminatory society, little robust evidence exists on this matter, especially regarding the extent of discrimination in the labor market (being Galarza and Yamada 2014, and Moreno et al. 2012 the only exceptions, as far as we are aware). Official statistics are of not much help here, and can only be used to estimate gender and racial wage gaps. Thus, while Yamada, Lizarzaburu, and Samanamud (2012) find potential racial gaps in the range of 11% and 22% for indigenous versus non-indigenous workers in Peru, they cannot be wholly attributed to discrimination.

Unlike the most recent literature on labor discrimination that only analyzes the impact of race or beauty on labor market outcomes (e.g., Kantor, Shapir and Shtudiner 2015, López Boó, Rossi and Urzúa 2013), our paper exploits a large-scale résumé audit study, specifically designed to detect the role of beauty, race, *and* sex, in the hiring decisions. We constructed fictitious résumés with similar human capital, and randomly assigned surnames (our proxy variable for race), sex, and headshots (deemed as either beautiful or homely-looking). This random assignment should result in statistically similar average call-back rates for all groups, in the absence of discrimination.

Our results provide novel evidence about the extent of discrimination that can be attributed to each of the three dimensions mentioned earlier. We find statistically significant beauty (82%), racial (54%), and sexual (34%) gaps in call-backs against homely-looking, indigenous, and female job applicants in Lima, Peru, all of which imposes a sizeable triple penalty in the access to employment on the homely-looking indigenous females. This result implies that the quest for equal employment opportunities in an emerging, fast-growing country, such as Peru, has a long way to go.

The remainder of the article is structured as follows. Section II reviews the related literature. Section III describes our experimental methodology and the procedures we followed during the field work. Section IV discusses the results, and Section V concludes.

II. Related literature

In Economics, discrimination in the labor market is defined as a situation in which a person who provides a service, and is similarly productive to another person, is treated differently and unfairly (for example, with lower wages or lower call-back rates for job interviews); moreover, this differential treatment is related to an observable characteristic, such as race, ethnicity or sex (List and Rasul 2011).

There are two broad sets of methods used to measure discrimination. The first one involves the use of observational data (such as national household surveys) to compare the wage gaps of any two groups, so that the part of such a gap unexplained by observables can be, at least partially, attributed to discrimination. The second method uses tools from experimental economics to collect data, and is referred to as audit study. The audit studies can be of two types, but both involve the use of fictitious candidates. The first type consists of sending fictitious job applicants, with similar academic background and work experience, trained by the researcher, to actual interviews.[1] In the second type, researchers send fictitious applicants' résumés with similar human capital. Given the equivalence in human capital, the level of discrimination is thus measured by any statistically significant difference in the hiring rate or the average call-back rates received by different groups of job candidates.

Does beauty matter in human's interactions? According to several authors, there seems to be a strong preference in humans for people with attractive faces (Dion, Berscheid, and Walster 1972; Griffin and Langlois 2006, cited in Pallet, Link and Lee 2010; Jefferson 2004; Hamermesh 2011). This preference, they claim, is based on a similar standard of beauty across age, sex, ethnicity, and social class (Hamermesh 2011; Jefferson 2004; Cunningham et al. 1995; Jones 1996; Perrett et al. 1994, 1998), even though this statement may be mostly based on evidence for Western countries (Sorokowski, Kóscinski and Sorokowska 2013).

What is the role (if any) of beauty in the labor market? A major benefit that attractive persons can have is related to their labor market returns. For the U.S., Hamermesh (2011) shows that 'attractive' men (women) earn, on average, wages that are 17% (12%) higher than for those who are 'unattractive,' controlling for a wide set of factors that can affect earnings (primarily, education and experience).

---

[1] An obvious shortcoming of this type of audit study is that many things can happen during interviews that cannot be controlled by the candidate or the experimenter, which may explain the difference in the performance of two otherwise comparable candidates.

There is also evidence that physically attractive people receive more callbacks for job interviews than homely-looking ones in Argentina (López Boó, Rossi and Urzúa 2013), Italy (Busetta, Fiorillo, and Visalli 2013), Germany (Kraft 2012), and Israel (Ruffle and Shtudiner 2014), where the authors conducted résumé audit studies. For a review of these types of studies on labor discrimination, the reader is referred to Galarza and Yamada (2014).[2] For our purposes, it suffices to mention that Bertrand and Mullainathan (2004) is the most widely cited article in this literature. These authors analyzed only racial discrimination in the United States (Chicago and Boston), by examining the call-back rates of candidates with African American and white names. They found a 50% higher call-back rate for whites.

To the best of our knowledge, the only experimental works that have examined the role of physical appearance in labor discrimination thus far are Kantor, Shapir and Shtudiner (2015), Ruffle and Shtudiner (2014), Kraft (2012) and López Boó, Rossi and Urzúa (2013), and all used a resume audit study in the same fashion as Bertrand and Mullainathan (2004) did. López Boó, Rossi and Urzúa (2013) sent CVs in pairs (a CV with a beautiful photo and another one with a homely-looking photo) to analyze the effect of beauty in the labor market of Buenos Aires, Argentina. They find that beautiful job applicants received 36% more callbacks than homely-looking ones, and are called more quickly. It is important to remark that, unlike Kantor, Shapir and Shtudiner (2015), Ruffle and Shtudiner (2014) and Kraft (2012), López Boó, Rossi and Urzúa (2013) used an objective measure of beauty: a ratio that measures the symmetry of the eyes and nose. The unattractive photos were thus obtained by altering this symmetry.[3]

Finally, the study by Galarza and Yamada (2014) pioneered the analysis of racial and sexual discrimination in the labor market of Lima, Peru, using an experimental design similar to that of Bertrand and Mullainathan (2004). They sent 4 résumés (2 white candidates and 2 Quechua candidates) with the same level of human capital for every single selected job opening, and found that there was significant discrimination against Quechua candidates and, to a lesser extent, against women. An important point to mention about this study is that white applicants had a higher score of subjective beauty (determined by a panel of judges) than Quechua applicants. This is an important limitation of that work, which is overcome by this article.

---

[2] For extensive reviews of literature examining experimental studies on racial, ethnic and gender discrimination, see Rich (2014) and Neumark (2016), Section 8.5.

[3] Facial symmetry is essential in determining beauty. Some authors suggest that such universal pattern of beauty is based on the "divinity ratio", 1.618, which must equal the ratio of the vertical distance between the head and the chin and the horizontal distance between the ears.

In the context described earlier, this article reports the results from the first experiment specifically designed to disentangle the effect of beauty from that of race in the observed labor market discrimination, for both females and males. The analysis of this matter is particularly important in a context of a booming economy –such as the one Peru has been facing recently– where one would expect that the candidates' qualifications will prevail in the hiring decisions, while in times of recession, we would expect the discriminatory stereotypes to strengthen.

III. The experiment

A. Selecting job vacancies and constructing résumés

Our experiment was designed to study the role of beauty, race, and sex on a labor market outcome, measured by the call-backs for interviews. It was conducted in Lima between April and September of 2012. Job vacancies were selected from one of the largest job networks in Lima, *Aptitus*, which publishes, online and in a newspaper, hundreds of job ads on a weekly basis (our selection is based almost exclusively on the newspaper ads: 99.5%).[4] When selecting the job postings, we did not restrict our sample to any particular economic activity (we thus include unskilled, technical and professional jobs). Rather, our selection reflects the intensity of the labor demand from each economic activity that seeks jobs through ads publicly advertised. The quality of formal education was standardized with academic degrees granted by similarly prestigious public institutions in all cases.

We created a database of résumés using real CVs available on two large employment Web sites, http://www.bumeran.com.pe and http://www.computrabajo.com.pe. This allowed us to more quickly tailor the résumés to the specific requirements of the selected job ads during the field work. The formatting of every set of résumés sent in response to each job ad was similar. We thus sent résumés with similar quality (in terms of job experience, skills, and training) for every selected job ad. Once constructed, race, sex, and level of physical appearance were randomly assigned to the résumés. Similarly, postal addresses, and school names were assigned at random.

All fictitious résumés were sent electronically before the deadline (if specified in the job ad), and every set of four résumés for a given job vacancy typically shipped away the same day, but at different hours, in order to avoid any effect related to the

[4] Results, reported in Section IV, do not change if we only use the newspapers ads.

day/time of receipt by employers. Moreover, we focus only on entry-level jobs (requiring up to three years of experience, as shown in the Appendix Table A1), and excluded job ads asking for salary expectations or in-person delivery of the résumés.

B. Selecting surnames and photos

We used a large online database of indigenous and white surnames[5] that classifies names by their origins to construct the identities of our job applicants. The selected white-sounding surnames have a predominantly white origin (English, French, Italian, and Spanish), while the indigenous-sounding surname have a clear *Quechua* origin. Sample surnames used include Anderson, Freundt, Bresciani, Camogliano, Goicochea (for white applicants), and Achachau, Aylas, Huamancuri, Sullca and Waylla (for Quechua candidates).[6] After getting a long list of the two types of surnames, we got random combinations of those to come up with a database of 720 full names (first name + paternal surname + maternal surname) for each type of applicant (Quechua and white). We then created personalized e-mail accounts. Every résumé sent for a given job ad included a different cell phone number, which our research assistants answered.

The photos of our candidates were also collected from the Internet,[7] and were subsequently modified by a Photoshop professional, in order to stylize each group, if needed (darker skin, thicker nose and lips, and more pronounced cheekbones for indigenous applicants), standardize the style (all men wear suits and women wear formal dress, with no accessories). Every candidate from each racial group was assigned a corresponding full name.[8] We constructed a subjective attractiveness

---

[5] http://apellidosperuanos.wordpress.com/.

[6] A referee suggested us to conduct a survey with the surnames used in the experiment, in order to assess the race of the person (with the choices being "white", "Quechua" and "Other"). Our survey respondents sample is composed of undergraduate students, from ages between 17 and 23. They were given the 374 surnames used in our experiment (129 Quechua-sounding and 245 white-sounding ones). The results: most of the Quechua-sounding surnames were considered Quechua, indeed, by our respondents (with an average percent of 84.4%), while a lower percent of the White-sounding surnames were considered as such (78.5%). If we take into account the fact that the maternal surname reinforced the origin of the paternal surname on the résumés we sent, in addition to the greater knowledge of the root of the surnames that any typical recruiter must have with respect to our typical survey respondent, we can have at least a moderate confidence that the surnames used are capturing the distinctive origins of our applicants.

[7] Photos belong to young females and males, in their twenties. Photos are not suited for a particular occupation (since the selection of job ads did not target any particular economic sector); they were rather standardized in style for *any* job. In Peru, this is a usual practice among job applicants (typically, photo studios do that on their behalf).

[8] It would be interesting to test mixed-race discrimination using this method, especially because most of the Peruvians self-report as *mestizos*. This could be part of our future research.

indicator (beauty), based on the ratings made by a panel of more than 50 judges, including students and professionals with different backgrounds who have made hiring decisions at some point in their careers (human resource specialists, psychologists, anthropologists, business administrators, economists, mathematicians, and sociologists). Judges rated headshots using a 1-to-7 scale (from homely-looking to strikingly handsome), and we labeled the photos as "beautiful" if the ratings ranged between 3.5 and 6.2, and "homely-looking" if they were between 1.6 and 3.1 (the standard deviation of all ratings goes from 0.71 to 1.97, with an average of 1.17), as shown in Table 1 (where we also report the range for each category).[9] From our initial pool of 150 photos for both white and indigenous candidates, we eliminated photos ranked around the mean of the sample physical appearance distribution. The main selection criterion for the final headshots was to make coincide, as best as we could, the average scores for the groups under scrutiny (females vs. males & indigenous people vs. whites). We thus ended up using 79 headshots for indigenous people and 76 for whites. We acknowledge that such selection process is arbitrary (as it would be any other alternative one), but it responds to our interest in examining the *differential* callback rates by physical appearance in the observed callback rates.[10] As seen in Table 1, our average beautiful applicant is significantly 'more attractive' than the typical homely-looking one (reported in last two columns).

Table 1. Average rating of physical appearance and means tests

	Male		Female		Total	
	White	Quechua	White	Quechua	White	Quechua
Beautiful	4.68	4.92	4.78	4.79	4.73	4.87
	[3.76, 6.00]	[3.68, 6.00]	[3.48, 6.20]	[4.00, 6.00]	[3.48, 6.20]	[3.68, 6.00]
Homely	2.57	2.48	2.61	2.21	2.59	2.35
	[1.80, 2.92]	[1.80, 2.92]	[1.60, 3.08]	[1.60, 2.88]	[1.60, 3.08]	[1.60, 2.92]
T-test for difference in means					Homely - Beautiful	Homely - Beautiful
Null hypothesis: Difference is equal to zero					-2.14	-2.51
(p-value)					(1.000)	(1.000)

Note: Figures in brackets indicate the range of ratings for each group.

[9] We followed this procedure under the premise of a common standard of facial beauty advocated by recent works in psychology (Cunningham et al. 2005, Perret, May, and Yoshikawa 1994, Perret et al. 1998).

[10] It is worthwhile to mention that we do not know what the average job applicant's physical appearance rating would be in the Peruvian labor market. Thus, although we acknowledge that a valid critique to our arbitrary selection of photos would be that the differential call-back rates by physical appearance we intend to capture may depend on the (average) levels of physical appearance of the two comparison groups (beautiful and homely-looking), addressing such concern goes beyond the scope of this paper Future research should address such concern. An anonymous referee suggested running some robustness checks with some sections of the physical appearance ratings density. This is done in section IV.

C. Treatments and sample size

We sent 4 résumés for each vacancy selected, 2 included a white-sounding surname (for female and male) and 2 included a Quechua-sounding surname. Every set of four résumés could fall into one of the following three categories, with roughly the same probability: (i) no photo attached (1,628) (treatment 1), (ii) a beautiful photo (1,676) (treatment 2), or (iii) a homely-looking photo (1,684) (treatment 3), for a total of 4,988 résumés.[11] It is worth mentioning that, even though the Law prohibits employers to request pictures attached to the résumés, it is standard for any job applicant in Lima to enclose a picture as part of their application packet. In that sense, it is highly unlikely that our experiment suffer from a selection problem. In the absence of labor discrimination, one should observe no statistical difference in the average call-back rate for every group of job candidates (attractive/unattractive /no photo, white/Quechua, and females/males). The existence of such difference would hence suggest discrimination. We address this issue in the next section.

IV. Empirical results

We estimate the causal effect of sex, race, and facial attractiveness (beauty) on call-backs for job interviews, using the following linear equation:

$$Callback\ Dummy_i = \alpha_0 + \alpha_1 Male_i + \alpha_2 White_i + \alpha_3 Attractiveness_i + \varepsilon_i \qquad (1)$$

where *Callback Dummy* takes the value of 1 if candidate “i” received a call-back or an email response for an interview (more than 97% of the responses were via phone), and 0, otherwise. *Male* and *White* are dummy variables for sex and race. We use two *Attractiveness* indicators: a dummy variable for (subjective) beauty, and a continuous variable that reports the normalized level of subjective attractiveness (whose construction was described above).

[11] In our design, each potential employer faces a choice among candidates from different races and sexes, given a level of physical attractiveness. The preference for physical attractiveness is then captured by comparing across potential employers. An alternative design would be to make each potential employer face a choice among candidates with different levels of physical attractiveness and sexes (or races), for a given racial (o sex) group (as in Ruffle and Shtudiner 2014 and López Boó, Rossi and Urzúa 2013). The preference for a racial (sex) group would be then captured by comparing across potential employers. Since we are trying to capture the effects of three variables on call-backs, we had to choose among the abovementioned designs. We chose the former for future comparison with our previous research.

Table 2 reports the regression results from the estimated linear probability model (results from estimating binary choice models are similar). Columns 1 to 5 include a dummy variable for beautiful job candidates, while columns 6 to 10 use the standardized attractiveness indicator. As shown in column 5, including all interactions among our beauty, race, and sex dummy variables, the beauty, race, and sex gaps are statistically significant at 1%. What is the magnitude of those gaps? Taking column 3, the reported coefficients imply that males receive 34% more call-backs than females, whites receive 54% more call-backs than similarly-qualified Quechuas, and beautiful candidates receive 82% more call-backs that homely-looking job applicants (results are similar when we include company fixed effects, to control for the differences in economic activity by sector).[12] Beauty not only pays, then, in the Lima labor market, but it also pays substantially more than whiteness and being male. To give an idea about the magnitude of these gaps, our results show that, in order to get a similar chance of being called back for a job interview than a beautiful white male applicant who sends 100 résumés, a homely-looking Quechua female candidate must send 380 résumés! This represents a large display of extra effort that is imposed on the Quechua people in the labor market, which adds more hurdles to the already many obstacles that this population needs to overcome in order to complete high school and then undergraduate studies. These results remain mostly unaltered when we control for postal addresses and school names in separate regressions (unreported results).[13]

Finally, our results show that each additional standard deviation in the level of subjective attractiveness increases the probability of being called back for an interview by 3.9 percentage points (column 10). We further find that including a photo in the résumé pays off (column 11). Obviously, this last result depends heavily on the average level of attractiveness of the photos attached to the résumés. As a robustness check of our results, we use different sections of our sample density of physical attractiveness. As reported in Appendix Table A2, when we use the top 20 percentile, which may be understood as a "beauty premium" (column 2), and the lowest 20 percentile, which may be understood as a "homeliness penalty" (column

[12] Using the parameter estimates, and setting the values of the other variables at their means, the predicted call-back rates are 17.88% for males (0.0545 + 0.0458 + 0.0661*0.5 + 0.0909*0.5) versus 13.30% for females (0.0545 + 0.0661*0.5 + 0.0909*0.5), 18.90% for whites (0.0545 + 0.0458*0.5 + 0.0661 + 0.0909*0.5) versus 12.29% for indigenous people (0.0545 + 0.0458*0.5 + 0.0909*0.5), and 20.14% for beautiful (0.0545 + 0.0458*0.5 + 0.0661*0.5 + 0.0909) versus 11.05% for homely-looking candidates (0.0545 + 0.0458*0.5 + 0.0661*0.5). Figures reported for the aforementioned gaps come from these data.

[13] Only in the latter case, did the coefficient on Male turned insignificant (0.0246, with p-value of 0.262).

3), as regressors, instead of the dummy variable for beauty (column 1), our results remain qualitatively unaltered. We also find a symmetric effect of such premium and penalty on the call-back rate (column 4).[14]

Another way, suggested by a referee, to look at the call-backs is by analyzing their distribution at the level of job ad. As reported in the Appendix Table A3, 64.05% of the employers did not call any of our 4 candidates, 20.5% called just 1 of them, and only 1.2% called all of them. The remaining 14.2% called 2 or 3 of our candidates. Moreover, "equal treatment" job ads represent 72.7% of the total, while 18% of them favor whites and 9.3% favor Quechuas. In either of these last two cases, most of the employers contact only 1 candidate.

We further run auxiliary regressions to split our analysis of call-backs by firm size (using the job ad size in the newspapers as a proxy variable. Table A1 in the Appendix shows the composition of firms by size), and by whether jobs involve a direct contact with the customer or not. As shown in Table A4 in the Appendix, in the former case, we find that large and medium-sized firms strongly prefer males in the first place, and then prefer whites, and attractive people (this is different from the results Kantor, Shapir and Shtudiner 2015 find for Israel). In contrast, for small and micro enterprises, beauty is the most highly correlated characteristic with call-backs, followed by race and sex. We thus observe more heterogeneity in smaller-sized firms. On the other hand, for the jobs not involving direct contact with the customer, we find that, to our surprise, beauty is the most important variable correlated with call-backs, followed by sex and race, which suggests that a taste-based discrimination story should be in order in this case. Moreover, for jobs which do involve direct contact with the customer, only race and beauty significantly affect call-backs, and in a similar magnitude. Beauty matters, yes, but in a similar magnitude for jobs involving contact with the customer and for those without such contact (a possible explanation for this result is that the greater self-confidence and ability for social interactions that may be attributed to attractive people and/or the taste-based discrimination –employers prefer to work with attractive employees– present in office jobs, may be as important as the delegated discrimination that may be present in jobs involving contact with the public –employers may observe that physical appearance is appreciated by their customers). In contrast, race matters more in jobs involving contact with the customer. Further research is need, in order to disentangle the mechanisms behind these results.

Lastly, we examine the role of beauty and race for females and males, beauty

---

[14] We thank an anonymous referee for this suggestion.

Table 2. Regression results on the callback dummy

Variable	Beauty Indicator										With Photo
	Beautiful (dummy variable)					Attractiveness (normalized)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Beauty Index	0.0909***	0.0909***	0.0909***	0.0909***	0.0924***	0.0426***	0.0419***	0.0414***	0.0416***	0.0392***	
	(0.0124)	(0.0124)	(0.0124)	(0.0124)	(0.0135)	(0.0061)	(0.0061)	(0.0061)	(0.0061)	(0.0065)	
White		0.0661***	0.0661***	0.0631***	0.0631***		0.0643***	0.0643***	0.0567***	0.0571***	0.0545***
		(0.0124)	(0.0123)	(0.0163)	(0.0163)		(0.0124)	(0.0123)	(0.0163)	(0.0163)	(0.0127)
Male			0.0458***	0.0428***	0.0428***			0.0437***	0.0361**	0.0365**	0.0473***
			(0.0123)	(0.0159)	(0.0159)			(0.0123)	(0.0159)	(0.0160)	(0.0126)
*Male*White*				0.0059	0.0089				0.0151	0.0147	-0.0016
				(0.0247)	(0.0275)				(0.0247)	(0.0246)	(0.0194)
*Male*White*Beauty Indicator*					-0.0060					0.0122	
					(0.0312)					(0.0176)	
With photo											0.0494***
											(0.0098)
Constant	0.1105***	0.0774***	0.0545***	0.0560***	0.0553***	0.1556***	0.1235***	0.1016***	0.1054***	0.1052***	0.0557***
	(0.0076)	(0.0092)	(0.0104)	(0.0112)	(0.0113)	(0.0062)	(0.0080)	(0.0096)	(0.0104)	(0.0104)	(0.0100)
Average call-back rate	0.1395	0.1395	0.1395	0.1395	0.1395	0.1395	0.1395	0.1395	0.1395	0.1395	0.1395
Observations	3360	3360	3360	3360	3360	3360	3360	3360	3360	3360	4988
R^2	0.0157	0.0240	0.0280	0.0280	0.0281	0.0138	0.0217	0.0253	0.0254	0.0256	0.0150

Notes: Robust standard errors in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

and sex for Quechuas and Whites, and race and sex for beautiful and homely-looking job applicants. Results, reported in Table A5 in the Appendix, thus address questions like: Is beauty or race more important for males versus females? Is sex more important for Whites or beautiful people versus Quechuas and homely-looking persons? In addition to finding that beauty is, by far, the most important variable positively correlated with callbacks for each race and sex considered (columns 1, 3, 5, and 7), we find that race and sex are similarly important for homely-looking persons to get a callback (column 9), while race is twice as important as sex for beautiful persons to get a callback (column 10). Again, adding company-size and sector-of-economic-activity fixed effects does not change the results. Overall, these results suggest that the intensity with which discrimination by race and physical appearance operates in the Lima labor market differs for males and females; the intensity of discrimination by physical appearance and sex differs for Quechua and white applicants; and the intensity of racial and sexual discrimination differs for beautiful and homely-looking persons.

V. Conclusion

The present article provides novel evidence on the magnitude of labor discrimination based on looks, race, and sex using a large-scale field experiment specifically designed for that purpose. While there is an increasing number of works that study labor discrimination –with a strong emphasis on developed countries– existing studies only examine one or two of the three variables of interest, which are jointly analyzed in this article.

Our results provide quite unprecedented indicators on labor market discrimination. In particular, we find significantly different treatments to physically attractive job applicants versus homely-looking ones (“beauty gap”), to whites versus indigenous people (*Quechua*) (“racial gap”), and to males versus females (“sex gap”). Which variable triggers the most unequal treatment (i.e., the greatest gap)? We show that the beauty gap in call-backs more than doubles the sex gap, and is 1.5 times the racial gap, thus imposing a triple penalty on job candidates who are homely-looking, indigenous, *and* female. Putting our findings in perspective, this means that, in order to have an equal chance of being called-back for a job interview than a beautiful applicant, a homely-looking one needs to send over 80% more résumés. This figure is 54% for the comparison between whites and Quechuas, and 34% for that between males and females. Altogether, this means that a Quechua, homely-looking female must send 380 résumés in order to have an equal chance to be called-back for a job interview than a beautiful white male who sends 100 résumés. This large cost that the labor market imposes on the least-favored group is what we refer to as the "triple penalty in employment access".

We further find that the patterns of discrimination differ by firm size and that beauty matters in a similar magnitude for jobs involving contact with the customer and for those without such contact. This sheds some evidence suggesting that beauty-bias in hiring would arise not only on behalf of the final customer's preferences but it could also reveal deeply embedded employer and workforce tastes. In contrast, race matters more in the former type of jobs. Lastly, our analysis within each group (beautiful and homely-looking, white and Quechua, and males and females) reveals that the intensity of discrimination by race and physical appearance differs for males and females; the intensity of discrimination by physical appearance and sex differs for Quechua and white applicants; and the intensity of racial and sexual discrimination differs for beautiful and homely-looking persons.

A number of limitations are worth mentioning. First, our analysis does not account for the wide spectrum of physical appearance. Secondly, we only focus on two minority groups in Peru (whites and Quechuas), thus excluding the afro-descendants and *mestizos* (who are the vast majority of the Peruvian population). Including the latter would imply challenges at the moment of identifying who belongs to that category, since surnames as a proxy variable for race would be inconclusive (e.g., there could be a white-looking Gonzales or a mestizo one). In such a case, race could perhaps be better captured by physical appearance (photos).

Appendix

Table A1. Summary indicators

Job applicant characteristics	Race	White	2494	50.00%
		Quechua	2494	50.00%
	Sex	Male	2494	50.00%
		Female	2494	50.00%
	Treatment	No photo	1628	32.64%
		Homely	1684	33.76%
		Beautiful	1676	33.60%
Job characteristics	Job category	Professional	1576	31.60%
		Technical	1692	33.92%
		Unskilled	1720	34.48%
	Job experience	None	4084	81.88%
		Up to 1 year	428	8.58%
		Between 1 & 2 years	436	8.74%
		3 or more years	40	0.80%
	Public / Office	Contact with the public	3164	63.43%
		Office	1824	36.57%
Firm characteristics	Size of firm[1/]	1: Micro	3037	60.89%
		2	1035	20.75%
		3: Medium	496	9.94%
		4	224	4.49%
		5: Large	172	3.45%
	Economic sector	Wholesale and retail	804	16.12%
		Consultancy (professional, technical or scientific)	628	12.59%
		Manufacturing	572	11.47%
		Others	2984	59.82%

Notes: [1/] Based on the size of newspaper job ads. 24 job ads were gathered from Aptitus online.

Table A2. Auxiliary regressions using alternative definitions of beauty

	(1)	(2)	(3)	(4)
Beautiful (dummy)	0.0909***			
	(0.0124)			
White	0.0661***	0.0690***	0.0546***	0.0594***
	(0.0123)	(0.0124)	(0.0129)	(0.0131)
Male	0.0458***	0.0463***	0.0347**	0.0375**
	(0.0123)	(0.0124)	(0.0131)	(0.0131)
Top 20 percentile of beauty density		0.0535**		0.0405*
		(0.0163)		(0.0173)
Lowest 20 percentile of beauty density			-0.0596***	-0.0467**
			(0.0143)	(0.0152)
Constant	0.0546***	0.0867***	0.123***	0.108***
	(0.0105)	(0.0101)	(0.0122)	(0.0136)
Observations	3360	3360	3360	3360
R^2	0.028	0.016	0.016	0.018

Notes: Robust standard errors in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A3. Call-back Rates by Job Ad

Equal Treatment	No call-backs	1W + 1Q	2W +2Q
72.71%	64.05%	7.46%	1.20%
3,627	3,195	372	60
Whites Favored (WF)	1W + 0Q	2W + 0Q	2W + 1Q
17.98%	13.41%	2.97%	1.60%
897	669	148	80
Quechuas Favored (IF)	1Q + 0W	2Q + 0W	2Q + 1W
9.30%	7.14%	0.72%	1.44%
464	356	36	72

Notes: *Equal treatment* includes the case in which the potential employer did not call any of our 4 candidates (no call-backs), called 1 white and 1 Quechua (1W + 1Q), or called all of our applicants (2 whites & 2 Quechuas) (2W + 2Q). *Whites Favored* includes the cases in which the employer called 1 of our white applicants (1W + 0Q), 2 of them (2W + 0Q), or 2 whites and 1 Quechua (2W + 1Q). *Quechuas Favored* includes the cases in which an employer called 1 of our Quechua applicants (1Q + 0W), 2 of them (2Q + 0W), or 2 Quechuas and 1 white (2Q + 1W).

Table A4. Auxiliary regression results: by firm size and contact with customer
Dependent variable: callback dummy

Variable	Firm size		Contact with customer?	
	Large & medium	Small & micro	Yes	No
Male	0.1000***	0.0404***	0.0318	0.0536***
	(0.0371)	(0.0131)	(0.0209)	(0.0152)
White	0.0905**	0.0623***	0.0953***	0.0499***
	(0.0371)	(0.0131)	(0.0209)	(0.0152)
Beautiful	0.0775**	0.0952***	0.0857***	0.0931***
	(0.0371)	(0.0131)	(0.0208)	(0.0153)
Constant	0.0490	0.0520***	0.0537***	0.0551***
	(0.0322)	(0.0109)	(0.0182)	(0.0128)
Average call-back rate	0.1379	0.1390	0.1453	0.1362
Observations	420	2920	1196	2164
R^2	0.0404	0.0284	0.0322	0.0271

Notes: Robust standard errors in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A5. Auxiliary regression results: by sex, race, and physical appearance. Dependent variable: callback dummy

Variable	Female		Male		Quechua		White		Homely	Beautiful
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Attractiveness		0.0457***		0.0373***		0.0295***		0.0587***		
(normalized)		(0.0079)		(0.0093)		(0.0074)		(0.0104)		
Beautiful	0.0752***		0.0758***		0.0752***		0.1261***			
(dummy)	(0.0207)		(0.0241)		(0.0207)		(0.0253)			
White	0.0378**	0.0560***	0.0638***	0.0716***					0.0378**	0.0887***
	(0.0189)	(0.0163)	(0.0236)	(0.0185)					(0.0189)	(0.0266)
Male					0.0425**	0.0381**	0.0685***	0.0523***	0.0425**	0.0431*
					(0.0191)	(0.0160)	(0.0234)	(0.0188)	(0.0191)	(0.0254)
*White * Beautiful*	0.0509		0.0105							
	(0.0327)		(0.0371)							
*Male * Beautiful*					0.0006		-0.0398		0.0260	-0.0144
					(0.0318)		(0.0378)		(0.0303)	(0.0391)
Constant	0.0638***	0.1058**	0.1064***	0.1418***	0.0638***	0.1042***	0.1016***	0.1612***	0.0638***	0.1391***
	(0.0190)	(0.0105)	(0.0150)	(0.0120)	(0.0119)	(0.0105)	(0.0147)	(0.0125)	(0.0190)	(0.0170)
Average callback rate	0.1162	0.1162	0.1628	0.1628	0.1127	0.1127	0.1664	0.1664	0.1105	0.2014
Observations	1680	1680	1680	1680	1680	1680	1680	1680	1692	1668
R^2	0.0321	0.0270	0.0194	0.0174	0.0175	0.0138	0.0230	0.0224	0.0148	0.0124

Notes: Robust standard errors in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

References

Bertrand, Marianne, and Sendhil Mullainathan. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94: 991-1013.

Busetta, Giovanni, Fabio Fiorillo and Emanuela Visalli (2013). Searching for a job is a beauty contest. University of Messina, University Politecnica delle Marche. Mimeo.

Cunningham, Michael R., Alan R. Roberts, Cheng-Huan Wu, Anita P. Barbee and Perry B. Druen (1995). Their ideas of beauty are, on the whole, the same as ours: Consistency and variability in the cross-cultural perception of female attractiveness. *Journal of Personality and Social Psychology* 68: 261–79.

Dion, Karen, Ellen Berscheid and Eliane Walster (1972). What is beautiful is good. *Journal of Applied Social Psychology*, 24: 285–90.

Galarza, Francisco and Gustavo Yamada (2014). Labor market discrimination in Lima, Peru: Evidence from a field experiment. *World Development* 58: 83-94.

Griffin, Angela M. and Judith H. Langlois (20006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition* 24: 187–206.

Hamermesh, Daniel S. (2011). *Beauty pays. Why attractive people are more successful*. Princeton. NJ: Princeton University.

Jefferson, Yosh (2004). Facial beauty-Establishing a universal standard. *International Journal of Orthodontics* 15: 9-22.

Jones, Doug (1996). *Physical attractiveness and the theory of sexual selection: Results from five populations*. Ann Arbor, University of Michigan Museum.

Kantor, Jeffrey, Offer Shapir and Ze'ev Shtudiner (2015). Beauty is in the eye of the employer: Labor market discrimination of accountants. Mimeo.

Kraft, Philipp (2012). The role of beauty in the labor market. *Dissertation,* Center for Economic Research and Graduate Education, Charles University Prague.

List, John A. and Imran Rasul (2011). Field experiments in labor economics. In D. Card and O. Ashenfelter, editors, *Handbook of labor economics,* Volume 4a. Amsterdam, North Holland, Elsevier.

López Bóo, F., Martin A. Rossi and Sergio Urzúa (2013). The labor market return to an attractive face: Evidence from a field experiment. *Economics Letters* 118: 170-2.

Moreno, Martin, Hugo Ñopo, Jaime Saavedra, and Máximo Torero (2012). Detecting gender and racial discrimination in hiring through monitoring intermediation services: The case of selected occupations in Metropolitan Lima, Peru. *World Development* 40: 315-28.

Neumark, David (2016). Experimental research on labor market discrimination. Working Paper 22022. Cambridge, MA, NBER.

Pallet, Pamela, Stephen Link and Kang Lee (2010). New 'Golden' Ratios for facial beauty. *Vision Research* 50: 149-54.

Perrett, David I., Karen A. May and Sakiko Yoshikawa (1994). Facial shape and judgments of female attractiveness. *Nature* 368: 239–42.

Perrett, David I., Kieran J. Lee, Ian Penton-Voak, Duncan A. Rowland, Sakiko Yoshikawa, D. Michael Burt, Peter Henzi, Duncan L. Castles, and Shigeru Akamatsu (1998). Effects of sexual dimorphism on facial attractiveness. *Nature* 394: 826–27.

Rich, Judy (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. Discussion Paper 8584. Bonn, Institute for the Study of Labor (IZA).

Ruffle, Bradley J. and Ze'ev Shtudiner (2014). Are good-looking people more employable? *Management Science* 61: 1760-76.

Sorokowski, Piotr, Krzysztof Kóscinski and Agnieszka Sorokowska (2013). Is beauty in the eye of the beholder but ugliness culturally universal? Facial preferences of Polish and Yali (Papua) People. *Evolutionary Psychology* 11: 907-25.

Yamada, Gustavo, Adriana Lizarzaburu and Katia Samanamud (2012). La persistencia de las brechas étnicas en el mercado laboral peruano. In F. Galarza, editor, *Discriminación en el Perú. Exploraciones en el estado, la empresa, y el mercado laboral*. Lima, Universidad del Pacífico.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 49-73

CONVENTIONAL VIEWS AND ASSET PRICES: WHAT TO EXPECT AFTER TIMES OF EXTREME OPINIONS?

J. Daniel Aromí *

IIEP Universidad de Buenos Aires

Submitted April 2015; accepted March 2016

This study evaluates the performance of stock market indices after times of extreme opinions. The underlying conjecture is that extreme opinions are associated to overreactions in the perception of wealth. The analysis covers 34 countries from 1988 through 2013. In a novel approach, views regarding economic performance are approximated using content in the global economic press. Consistent with the overreaction conjecture, stock market indices are shown to under-perform following extreme optimistic views and over-perform after pessimistic views. A long-short contrarian portfolio earns 11% annually over the next five years. This persistent and predictable difference in returns cannot be explained by risk considerations and cannot be replicated using alternative strategies based on past returns or past economic growth.

JEL classification codes: G12, G17, D84

Key words: asset prices, opinions, expectations, overreaction.

I. Introduction

In November 2009, the weekly magazine *The Economist* ran a cover in which the title "Brazil takes off" was accompanied by a *statue* of Christ the Redeemer ascending like a rocket from Rio de Janeiro's Corcovado mountain. This strong sign of optimism was later reversed in September 2013 when the cover asked "Has Brazil blown it?"

---

* Daniel Aromí: Universidad de Buenos Aires, Facultad de Ciencias Económicas, Instituto Interdisciplinario de Economia Política (IIEP-BAIRES), Av. Córdoba 2122 - 2do (C1120 AAQ), Ciudad Autónoma de Buenos Aires, Argentina; email aromi.daniel@gmail.com. We would like to thank Daniel Heymann for generous and fruitful discussions. Also, we would to thank Daniel Chodos for providing data and Marcos Dal Bianco, Luciano Del Corro, and participants at the 2014 AAEP meeting and the seminars held at IIEP-Baires, UCEMA and Universidad del Sur for helpful comments.

together with a picture of a collapsing statue-rocket.[1] As eloquently exemplified by the pair of covers, opinions regarding economic prospects are regularly seen to enter stages of high hopes or, in other occasions, periods of intense gloom. Determining the accuracy of these emergent judgments is a matter of interest. This is because important economic decisions by private and public actors can benefit from a better understanding of the information conveyed by conventional views.

Conventional views are understood as a set of beliefs that are broadly shared, are known to be shared and so on. In this work, the focus is placed on broadly shared views regarding economic conditions. For example, beliefs regarding future economic prosperity or the evolution of aggregate productivity. Importantly, the public condition of this set of beliefs implies that they are observable and can be approximated, for example, analyzing contents in the economic press.

The underlying assumption behind this work is that the evolution of these public beliefs are strongly linked to the path of asset prices in an economy. More specifically, one plausible conjecture is that extreme conventional assessments are associated to excessive responses. Under this conjecture, the occasional emergence of extreme shared opinions could be linked to mispricing of broad classes of assets and predictable errors in saving and investment decisions. Despite its relevance, formal empirical evidence of this conjecture is hindered by lack of sufficiently comprehensive and precise measures of conventional views.

In this work, this conjecture is empirically evaluated for the case of financial assets. The performance of stock market indices is evaluated after times of extreme optimism and extreme pessimism. The study covers 34 countries from 1988 through 2013. One distinctive aspect of this work is the approximation of conventional views using content published in the international economic press.

Consistent with the postulated conjecture, the results show that optimism is followed by lower mean returns and pessimism by higher mean returns. This difference in performance is highly persistent and economically significant. A long-short contrarian portfolio earns 11% annually over the next five years. Additionally, it is found that the performance of sentiment based portfolio strategies cannot be replicated using information on past returns or past economic growth. Finally, the findings suggest that changes in anticipated risk levels are not a good explanation of the reported return differentials.

A natural interpretation of these findings is that the occasional emergence of conventional views regarding economic prospects generates mispricing for a broad

---

[1] See *Economist* (2009) and *Economist* (2013).

set of assets. Contributions associated to social learning can help rationalize these events. The well-established literature on social learning has shown that information can be aggregated quite inefficiently.[2] Complementarily, social learning can lead to more severe inefficiencies if agents follow simple rules that ignore redundancies in public information[3] or if a subset of actors is too influential.[4] Additionally, the existence of return predictability based on publicly available information suggests the presence of individuals that process information in an incomplete and correlated manner.[5]

The recurrent, persistent and predictable overreactions documented in this study have important implications that go beyond financial markets. Given the key function of asset prices in the aggregation of information and the coordination of actions, the results are relevant for the understanding of macroeconomic dynamics and public policy that aims for stability.

One key assumption of the current study is the idea that conventional views regarding economic prospects can be approximated processing information in the economic press. This idea is supported by the dual role of the media. The economic press publishes information that reflects and, at the same time, shapes public opinion. It is worth noting that press content is selected based on journalists' or editors' beliefs regarding dominant public opinions. In this context, confirming public opinions might be a profitable marketing strategy.[6] Additionally, it is reasonable to assume that an important fraction of the content is forward looking. For example, the inspection of some distant episode might be linked to its relevance given current opinions.

The reported findings show that empirical studies that focus on subjective states can advance the understanding of economic dynamics. This is a natural outcome once it is recognized that economic processes emerge as the result of the co-evolution of structural and subjective elements. According to this perspective, subjective elements cannot be inferred through logical deductions and there is value in finding estimations of these elements that can allow for new insights.

This study is related to a fertile body of theoretical and empirical contributions that calls attention to the role of expectations, learning and coordination in financial

---

[2] See for example Banerjee (1992) and Bikhchandani et al. (1992).

[3] Eyster and Rabin (2010, 2014).

[4] Golub and Jackson (2010).

[5] Models of cognitive limits or simple responses can be found, for example, in Mullainathan (2002), Hong and Stein (1999) and Brock and Hommes (1998). Conceptual analyses in this line can also be found in Kindleberger and Aliber (2011) and Shiller (2005).

[6] See for example Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2010).

and macroeconomic dynamics. Using theoretical models, learning dynamics about structural parameters have been explored as a source of aggregate fluctuations.[7] Additional sources of non-fundamental volatility are suggested by models with strategic complementarities and multiple equilibria. In these analyses, under-determination of equilibrium can be interpreted as an indication that, inevitably, beliefs are a distinct element in the determination of economic outcomes.[8] Complementarily, the literature on herd behavior has shown that information can be aggregated inefficiently and, as a result, aggregate beliefs and behavior can be quite idiosyncratic.[9] In financial markets, models of limits to arbitrage have shown that market participants might be unable or unwilling to implement trading strategies that transmit information to asset prices.[10] Importantly, further sources of non-fundamental volatility emerge once cognitive limits and simple mental models are allowed for.[11]

In a related contribution, Dumas et al. (2011) show that some well-known anomalies observed in international equity markets can be explained by a model in which agents perceive differences in the precision of information regarding local and foreign markets.

This article is also connected to the well-known empirical literature that shows evidence consistent with inefficiencies, excess volatility and overreaction in stock markets. Most of these analyses focus on firm level returns and use historic information on asset prices, financial statements and market activity.[12] More closely related to the current study, some contributions focus on the dynamic relationship of aggregate proxies for investor sentiment and stock returns.[13] In these contributions, the proxies are based on consumer confidence surveys and financial market outcomes. In a related analysis of international equity markets, Hwang (2011) finds that sentiment toward countries, as inferred from public opinion surveys, is associated to demand for securities and distortions in asset prices.

---

[7] For example, see learning models in Sargent (1993), Heymann and Sanguinetti (1998) and Milani (2007).

[8] See Diamond (1982), Cooper and John (1988) and Obstfeld (1996).

[9] For early contributions see Banerjee (1992) and Bikhchandani et al. (1992). Angeletos et al. (2010) provide a related model applied to macroeconomic fluctuations.

[10] See De Long et al. (1990) and Shleifer and Vishny (1997).

[11] See for example models of categorical thinking (Mullainathan 2002), trend chasing (Hong and Stein 1999) and predictor selection dynamics (Brock and Hommes 1998).

[12] See, for example, Shiller (1981), Bondt and Thaler (1985), Lakonishok et alt. (1994), Jegadeesh and Titman (2001), and Baytas and Cakici (1999). For a comprehensive evaluation of these anomalies see Asness et al. (2013).

[13] Baker and Wurgler (2007), Jansen and Nahuis (2003) and Schmeling (2009).

There is a growing set of contributions that use information in the press to describe dynamics in macroeconomic and financial settings. Information in the press has been used to describe predictive content related to the economic cycle (Baker et al. 2012, Aromí 2014), to exchange rate volatility (Krol 2014) and to describe dynamics of consumer confidence (Doms and Morin 2004). Anticipation of daily stock market returns has been shown by Tetlock (2007) and Garcia (2013) for the US and Aromí (2013) for the case of Argentina. As in the case of the current study, but at higher frequencies, the evidence found in those articles is consistent with overreactions in stock prices.

The rest of the paper is organized as follows. Section II describes the data used in the study and the way in which media information is summarized. Section III presents the evidence from portfolio strategies. Investment strategies based on alternative sources of information are described in section IV. Section V presents evidence using monthly returns data. Section VI concludes.

II. Data and opinion metrics

The analysis uses two categories of yearly frequency information: financial assets returns and media content. The first set of data is given by the returns of stock market indices expressed in dollars. The main source for this data is the World Bank.[14] The data covers 34 countries over 26 years (1988-2013). For the early part of the sample (1988-1995), for some countries, this data was not available from this source. As a result, supplementary data was obtained from a private data vendor[15] and, in few cases, from the relevant stock exchange. The sample covers countries that belong to different regions and display heterogeneous levels of economic development.

Given the value of the stock market index of country i at the end of year t (SMI_{it}), the annual return in year t for country i is given by the difference of the logs of the index for years t and $t-1$: $r_{it} = \log(SMI_{it}) - \log(SMI_{it-1})$. Table 1 shows descriptive statistics for the return in the sampled countries.

The second type of data is an indicator of conventional views based on content published in the economic press. More specifically, this indicator was built based in information published in *The Wall Street Journal* (1984-2013) and *The Economist* (1992-2013). Together with *The Financial Times*, these publications are among the three main business publications in the English language.

---

[14] http://data.worldbank.org/indicator/CM.MKT.INDX.ZG.

[15] http://www.tradingeconomics.com/

Table 1. Descriptive statistics

Country	Sentiment Index				Annual Return			
	Mean	St. dev.	Min.	Max.	Mean	St. dev.	Min.	Max.
Argentina	0.067	0.008	0.047	0.082	0.076	0.541	-0.821	1.617
Austria	0.052	0.007	0.039	0.067	0.027	0.325	-1.050	0.476
Belgium	0.059	0.006	0.048	0.071	0.046	0.316	-1.079	0.495
Brazil	0.062	0.005	0.051	0.072	0.091	0.628	-1.309	1.356
Chile	0.061	0.009	0.049	0.085	0.117	0.304	-0.528	0.668
Colombia	0.083	0.011	0.065	0.114	0.085	0.380	-0.635	0.765
Czech Republic	0.053	0.010	0.038	0.085	0.046	0.291	-0.616	0.565
Denmark	0.050	0.008	0.032	0.064	0.090	0.257	-0.713	0.445
Finland	0.050	0.007	0.034	0.060	0.075	0.391	-0.844	0.871
Greece	0.068	0.010	0.051	0.089	0.060	0.499	-1.079	1.092
Hungary	0.060	0.011	0.042	0.086	0.077	0.382	-0.994	0.693
India	0.069	0.008	0.053	0.088	0.077	0.372	-1.022	0.663
Indonesia	0.065	0.013	0.046	0.086	0.065	0.571	-1.347	1.261
Ireland	0.061	0.008	0.046	0.084	0.060	0.339	-1.204	0.438
South Korea	0.062	0.005	0.054	0.073	0.077	0.490	-1.171	0.793
Malaysia	0.058	0.011	0.038	0.077	0.060	0.374	-1.309	0.621
Mexico	0.064	0.008	0.053	0.081	0.149	0.363	-0.598	0.775
New Zealand	0.051	0.005	0.044	0.061	0.023	0.279	-0.734	0.470
Norway	0.055	0.007	0.044	0.074	0.056	0.343	-1.079	0.647
Pakistan	0.088	0.008	0.070	0.105	0.071	0.425	-0.968	0.751
Peru	0.072	0.012	0.047	0.102	0.201	0.423	-0.528	1.078
Philippines	0.070	0.008	0.055	0.087	0.060	0.450	-0.968	0.859

Table 1. (continued) Descriptive statistics

Country	Sentiment Index				Annual Return			
	Mean	St. dev.	Min.	Max.	Mean	St. dev.	Min.	Max.
Poland	0.063	0.008	0.053	0.090	0.109	0.583	-0.868	2.202
Portugal	0.059	0.011	0.043	0.078	0.025	0.278	-0.755	0.399
Russia	0.071	0.006	0.058	0.084	0.130	0.765	-1.833	1.345
Singapore	0.052	0.008	0.040	0.068	0.061	0.324	-0.755	0.571
South Africa	0.071	0.010	0.056	0.090	0.075	0.264	-0.545	0.445
Spain	0.058	0.008	0.045	0.077	0.065	0.246	-0.562	0.438
Sweden	0.053	0.006	0.043	0.069	0.083	0.312	-0.755	0.536
Taiwan	0.058	0.004	0.048	0.066	-0.012	0.352	-0.821	0.610
Thailand	0.068	0.011	0.047	0.097	0.044	0.527	-1.561	0.904
Turkey	0.073	0.007	0.060	0.086	0.055	0.618	-0.968	1.267
Venezuela	0.068	0.012	0.041	0.095	-0.014	0.500	-1.079	0.779
Vietnam	0.083	0.008	0.061	0.099	-0.105	0.503	-1.139	0.385
Average	0.063	0.008	0.048	0.082	0.065	0.412	-0.948	0.802

Note: The Sentiment Index was constructed using content from *The Wall Street Journal* (1984-2013) and *The Economist* (1992-2013). Annual Return is the dollar return of each country stock market index and was provided by the World Bank (1988-2013) and, in few instances, by tradingeconomics.com.

As discussed in the introduction, there are reasons to believe that conventional views can be approximated through content published in the media. On the other hand, summarizing media content to generate sentiment indices is a challenging task. Below a description of the path followed in this work is provided.

Opinions transmitted in the press regarding different countries are summarized computing the frequency of words with negative content in a relevant subset of text. This is a simple approach that has proven successful in other contexts.[16]

The first step in the construction of the indices involves identifying a list of keywords associated to each country: name of country, capital city[17] and demonym. Next, for each year in the sample period, the set of articles in which at least one of these keywords is present is identified. For each of these articles, the portions of text that are sufficiently close to a keyword associated to any country are selected. More specifically, the selection corresponds to words that are up to 10 words before or 10 words after one of the keywords associated to any country.[18] The strings of text associated to country c and year t are merged forming a selection of text K_{ct}. This concludes the text extraction stage.

An indicator of sentiment for the relevant country for each year is generated computing the frequencies of words with negative content. Following the seminal contribution by Tetlock (2007), the list of negative words is built using the negative valence category from the Harvard IV dictionary. The dictionary was procured from General Inquirer, a website that provides tools for content analysis of textual data.[19]

It must be noted that there are other lists of negative words that can be used in this analysis. In a relevant antecedent, Loughran and McDonald (2011) develop a list of negative words for financial contexts. Their analysis shows that informational gains can be attained using a more precise list of words. On the other hand, lists of words that are generated after the date of the sample run the risk of incorporating forward looking bias. As a result, despite the potential gains that could result from

[16] In line with the findings reported in Tetlock (2007), indices constructed using words with positive content lack information on future stock returns. This can be explained by asymmetries in the information content of positive and negative words as found in the natural language processing literature (Garcia et al. 2012).

[17] In the case of Brazil, Spain, India, Philippines and South Africa big cities that can be unambiguously linked to the country are also included (e.g., Cape Town for South Africa).

[18] Following usual practice in content analyses, stop words (common words with no relevant content) were eliminated before neighboring text was extracted. Additionally, it is worth noting that variations in which 5 or 50 neighboring words were selected lead to very similar results.

[19] http://www.wjh.harvard.edu/~inquirer/homecat.htm.

considering alternative list of words, in the main part of the analysis we select the most cautious path and use the list originally employed in Tetlock (2007). Alternative lists are also evaluated as part of the sensitivity analysis.

The original list of negative words from General Inquirer includes 2291 words. In order to improve the precision of the indices, this original list was expanded to include plural noun forms, different verb tenses and adverbs. This procedure resulted in a list of 5364 words.

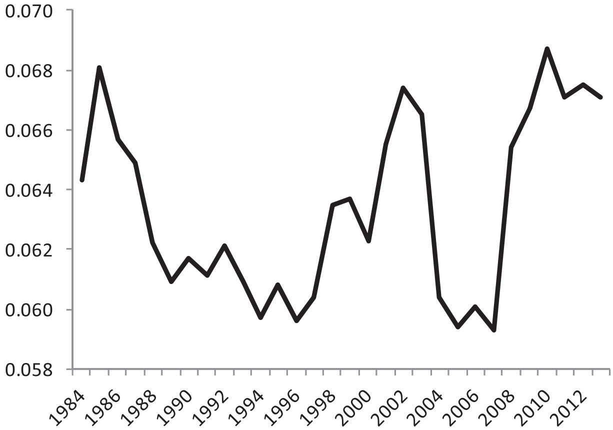
Let T_{ct} be the number of words in K_{ct}, the selected text corresponding to year t and country c and let N_{ct} be the number of times a negative word is detected in K_{ct}. Then, the corresponding sentiment index is given by $s_{ct} = N_{ct} / T_{ct}$. A higher number is associated to more pessimistic views while a lower number is associated to more optimistic assessments. Thus, the indicator could be labeled as a negative sentiment index. Throughout the document, at the risk of some confusion, the shorter expression sentiment index will be used. The construction of this indicator involves the selection of text comprising approximately 23 million words out of which more than 1.5 million correspond to words classified as negative words.

Table 1 shows important cross-country differences in the level of the sentiment index. The average value of the sentiment index ranges from 0.05 in the case of Denmark to 0.088 for the case of Pakistan. The standard deviation ranges from 0.004 in the case of Taiwan to 0.013 in the case of Indonesia. Looking more meticulously, it is found that the minimum value for the index is 0.032 and corresponds to Denmark in 1989. On the other hand, the maximum value is 0.114 and corresponds to Colombia during 2003, a year in which the country experienced weak economic performance and high levels of violence. These observations constitute a first indication suggesting that the index is able to capture important economic developments in the selected countries.

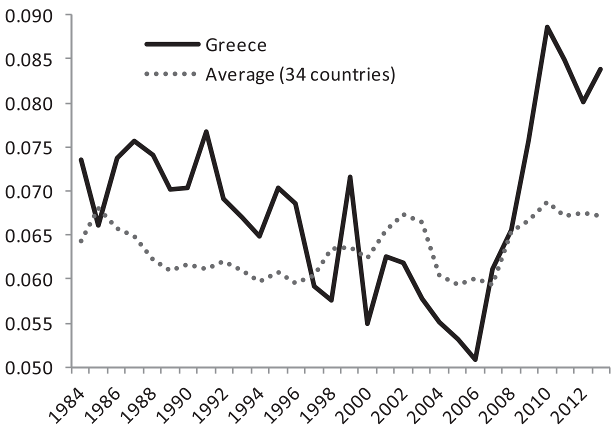
As a first step in the analysis, it could be instructive to observe the evolution of the indices during the sample period. Figure 1.A presents the average value of sentiment index for the 34 countries from 1984 through 2013. The path of the index suggests that there is a close association with the evolution of the global economy. For example, by a small margin, the minimum value is observed in 2007, on the final year of a period known as the great moderation. On the other hand, the maximum value is observed in 2010, in the aftermath of the global crisis originated in the U.S. and during a critical phase of the European crisis. In addition, a peak observed in 2001-2003 can be linked to the burst of the dot-com bubble, crises in emergent markets and the conflicts after 9/11. A period of low values is observed in the mid-nineties, a period of average economic bonanza.

Figure 1. Sentiment index

A. Average for 34 countries



B. Greece



Note: Shaded regions in panel A correspond to years with world GDP growth below 2%.

Additional preliminary insights can be gained considering the case of a single country. Figure 1.B shows the sentiment index for Greece. Particularly suggestive is the drop in the index, that is, the increasingly optimistic views observed from 2002 through 2006. In 2006 the index for Greece reaches a historic low of 0.0508. Four years after, in 2010, the index reaches a historic high of 0.0887. This extreme pattern is suggestive of overreaction in at least one of these instances. These observations based on anecdotal evidence are informative and suggests the plausibility of the overreaction conjecture. Formal empirical evaluations of this conjecture are implemented in the next sections.

III. Performance of contrarian portfolio strategies

In this section, contrarian portfolio strategies are implemented and their performance is evaluated. Each year, past values of the sentiment index are used to construct a portfolio of countries associated to past optimistic views and a portfolio of countries associated to past pessimistic views. Under the overreaction conjecture, it is expected that the first portfolio will experience inferior returns and the second portfolio will experience superior returns.

The sorting of countries is implemented using the average value of the sentiment index computed using four-year moving windows. It is expected that averaging the value of the index over multiple years will reduce noise in the identification of extreme views. Additionally, averaging is compatible with the focus of this study on low frequency, highly persistent dynamics.

A. Baseline estimates

As shown in the previous section, the mean values of the indices associated to different countries show ample variation. Hence, unless some correction is implemented, there are some countries that could be systematically associated to optimistic states or pessimistic states simply due to stable high or low values in the sentiment index. This suggests that adjusting for differences in mean values is needed if the intention is to capture changes in shared views about countries' prospects instead of reflecting countries' permanent characteristics. The indices are adjusted using the historical values of the index for the corresponding country.

Additionally, the indices associated to different countries express ample variation in terms of standard deviation. This could be due to more volatile perspectives or to differences in noise of the index due to heterogeneous levels of coverage of sampled countries. Taking into account these differences, in the exercise below the indices are adjusted using only historical information so any look-ahead bias is avoided.

More specifically, the standardized sentiment index $\hat{s}_{ct}$ is given by $\hat{s}_{ct} = \left(s_{ct} - \overline{s}_{ct}\right) / s^{v}_{ct}$ where $\overline{s}_{ct}$ is a weighted average of past values and s^{v}_{ct} is a measure of past volatility. Adjustment parameters are given by $\overline{s}_{ct} = \sum_{k=t_0}^{t-1} s_{ck}\,\omega(t-k)$ and $s^{v}_{ct} = \sum_{k=t_0}^{t-1} \left|s_{ck} - \overline{s}_{ck}\right| \omega(t-k)$, where t_0 is the first period of the sample and the weighting function $\omega(t-k)$ is a decreasing function which satisfies $\sum_{k=t_0}^{t-1}\omega(t-k)=1$, where $\omega(t+1)=(3/2)\omega(t)$, that is, the weight decreases with distance at a constant 33% rate. The weighting function captures the idea that more recent observations are more informative while the expanding window allows for more informed adjustments.[20] To secure for informed adjustment of the index value, the analysis is restricted to observations for which there exist at least four years of historic sentiment data. Finally, the indicator used to construct the portfolios for year t is the average value of the standardized index in the four most recent years $s_{ct}^{*}=\sum_{k=t-3}^{t}\hat{s}_{ck}$.

For each year, two portfolios are built identifying the top and bottom deciles using the adjusted index s_{ct}^{*}. The portfolio associated to optimism, or Portfolio 1, is composed by the stock market indices of the countries that belong to the bottom decile. Similarly, the portfolio associated to pessimism, or Portfolio 2, is composed by the stock market indices of the countries that belong to the top decile. Let $r_t^{P1(t-l)}$ be the average return in year t for the stock indices that belong to Portfolio 1 built in year $t-l$. In the same fashion, $r_t^{P2(t-l)}$ is the equivalent indicator for Portfolio 2. The analysis below will focus on the returns of these portfolios for different values of the lag parameter l. If extreme assessments are associated to excessive reactions, it is expected that Portfolio 1, the optimistic portfolio, will underperform and Portfolio 2, the pessimistic portfolio, will show superior performance.

In addition to comparing average returns, in this section a formal test for abnormal returns is implemented. A simple empirical model that includes a market factor is proposed. The statistical model is given by the following equation:

$$r_t^{P(t-l)}=\alpha+\beta r_t^{m}+\epsilon_t,$$

where $r_t^{P(t-l)}$ is the return of portfolio $P(t-l)\in\{P1(t-l),P2(t-l)\}$ in year t, r_t^m is the average return for all sampled countries in year t and ϵ_t is an error term. Following a standard procedure in the asset pricing literature, the estimate for the parameter α is interpreted as the abnormal return of the relevant portfolio. Additionally, similar calculations are computed for a long-short portfolio strategy in which Portfolio 1, the optimistic portfolio, is the short position and Portfolio 2, the pessimistic portfolio, is the long position. Standard errors and associated t-statistics are corrected for heterocedasticity.[21]

---

[20] Unreported exercises show that the results are not sensitive to changes in the weighting function.

[21] Package car in platform R was used to estimate robust standard errors.

Table 2. Performance of sentiment based portfolios

A. Return and performance for different lags in portfolio formation

Lags	Portfolio 1 (Optimism)				Portfolio 2 (Pesssism)				Portfolio 2 - Portfolio 1			
	Mean	St. Dev.	Alpha	t-stat.	Mean	St. Dev.	Alpha	t-stat.	Mean	St. Dev.	Alpha	t-stat.
0	0.087	0.302	0.038	0.9	0.121	0.435	0.049	0.9	0.034	0.337	0.011	0.1
1	0.045	0.367	-0.035	-1.3	0.210	0.372	0.134	2.7	0.165	0.245	0.169	2.7
2	0.011	0.340	-0.055	-2.2	0.058	0.387	-0.011	-0.2	0.047	0.249	0.044	0.7
3	-0.002	0.297	-0.062	-2.2	0.074	0.401	-0.001	-0.0	0.077	0.298	0.061	0.8
4	-0.016	0.310	-0.079	-2.5	0.126	0.295	0.068	1.7	0.142	0.167	0.147	2.9
5	0.002	0.348	-0.067	-1.7	0.155	0.336	0.088	2.7	0.153	0.258	0.155	2.5
6	0.007	0.324	-0.040	-1.3	0.072	0.379	0.022	0.4	0.066	0.279	0.022	0.4

B. Return and performance of mean sentiment based-portfolios (1 through 5 lags)

	Portfolio 1	Portfolio 2	Portfolio 2 - Portfolio1
Mean	0.010	0.121	0.111
St. dev.	0.324	0.337	0.176
Alpha	-0.059	0.051	0.110
t-stat.	-2.7	1.8	2.5

Notes: Panel A: Portfolio 1 has equal sized long positions in the stock indices of the countries of the most optimistic decile. Portfolio 2 has equal sized long positions in the stock indices of the countries in the most pessimistic decile. For the model: $r_t^{P(t-l)} = \alpha + \beta r_t^m + \epsilon_t$ for $l = 0,1,\ldots,6$, alpha is the estimated constant, $r_t^{P(t-l)}$ is the return of the corresponding sentiment-based portfolio (Portfolio 1, Portfolio 2 or Portfolio 1 – Portfolio 2), r_t^m is the market return and ϵ_t is an error term. Panel B: mean sentiment-based portfolio corresponds to the average of the portfolios with 1 through 5 lags in portfolio formation year. The associated return is: $\bar{r}^{P(t)} = \sum_{l=1}^{5} r_t^{P(t-l)} / 5$. The t-statistics are corrected for heterocedasticity.

Table 2.A shows the results for multiple values of the lag parameter l. One notable feature is the poor performance of Portfolio 1. From two years and up to six years after portfolio formation, the mean return of this portfolio is below 1.1%. More dramatically, for the case of three and four year lags, the mean return is negative. Similar conclusions emerge from observing the estimation of the market factor model. For example, four years after portfolio formation, the estimated abnormal return is -7.9%.

On the other hand, the performance of Portfolio 2, the portfolio associated to pessimism, takes the opposite direction. For example, one year after portfolio formation, the mean return is 21%. High values are also observed for the mean returns four and five years after portfolio formation. The estimations of factor models provide similar results. The abnormal returns one, four and five years after portfolio formation are estimated to be above 6.8% and statistically different from zero. In contrast, two and three year lags result in estimated abnormal returns that are slightly negative but statistically null.

The last panel in Table 2.A describes the return associated to the long-short portfolio strategy. Lags of one, four and five years result in positive abnormal returns of at least 15%. These estimations are suggestive of overreactions that are gradually corrected years after the extreme opinions are identified. Lags of two, three and six years show positive but statistically insignificant abnormal returns. At this stage, it is not clear whether the differences in performance for different number of lags are simply noise or reflect a stable property of the return reversal process. Another interesting observation is that the results suggest a weak contemporaneous relationship between conventional views and stock index returns.

In terms of summarizing the results of this exercise, it is convenient to provide a description of the performance of the portfolio strategies when the returns associated to different values of the lag parameter l are combined. With this objective, the average return $\bar{r}^{P(t)} = \sum_{l=1}^{5} r_t^{P(t-l)} / 5$ is computed for portfolios $P(t-l) \in \{P1(t-l), P2(t-l)\}$. In other words, the new portfolios are built combining, with equal weights, the portfolios that exploit information with one through five lags. It is expected that this portfolio will allow for a more precise assessment of the association between lagged sentiment and return differentials.

As shown in Table 2.B, the mean return of Portfolio 1 is 1% while the mean return of Portfolio 2 is 12%. Suggesting that risk considerations are unlikely to provide a satisfactory explanation, the estimated standard deviations for each portfolio are very similar. The estimated abnormal returns are -5.9% for the case of Portfolio

1 and 5.1% in the case of Portfolio 2. The associated long-short portfolio strategy results in a highly statistically significant abnormal return equal to 11%.

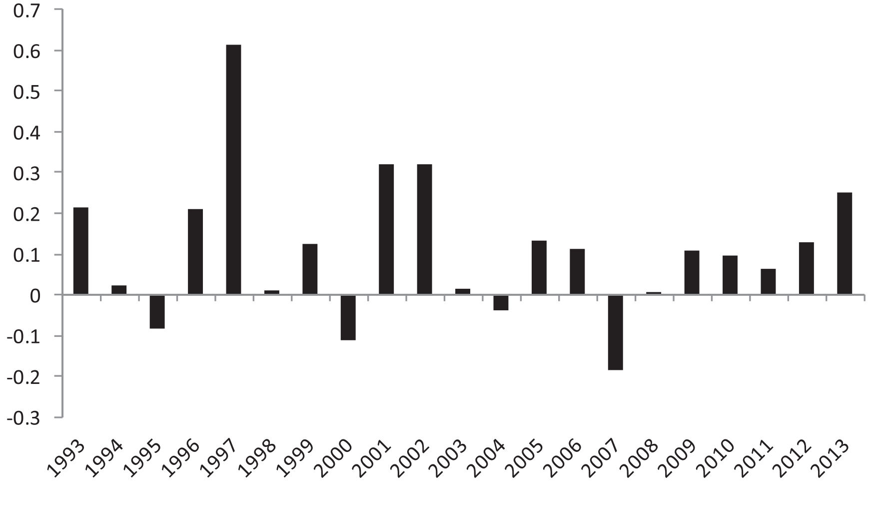
Figure 2 shows the returns of this long-short portfolio strategy. The figure shows a difference in performance that seems to be quite stable throughout the sample period. As can be observed, there are only four years in which the return is negative and in none of these years the return is below -20%. In contrast, the annual return is above 20% in six occasions.

The results presented above are consistent with the proposed conjecture of excessive responses associated to extreme opinions. Optimism is followed by underperformance and pessimism is followed by over-performance. This difference is highly persistent and stable over the sampled period. Risk considerations are unlikely to provide a satisfactory explanation.

B. Sensitivity analysis

In this subsection we provide information on several exercises in which aspects of the original exercise are modified. It is understood that this alternative exercises might shed light on the robustness of the previous results and the directions that

Figure 2. Returns of the long-short portfolio (based on four years of publications' data)



Notes: The bar corresponding to year t represents the average return on year t of the portfolios built using the sentiment indices of years t-5 through t-1.

can allow for informational gains. In general, the exercises show that the results are quite robust. Also, there seems to be little room for informational gains.

Table 3 provides information on the returns of the average contrarian portfolio under four types of modifications of the original exercise. The first analysis evaluates alternative sizes for the moving windows over which average sentiment is computed. As panel A shows, little changes are observed when the original size of 4 years is modified in the direction of shorter or longer periods.

The second examination deals with the selection of words. In the original exercises, words that are at a distance equal or lower than 10 from relevant keywords

Table 3. Characterization of returns in sensitivity analysis exercises

A. Size of moving window (years)

	3	4	5
Mean	0.110	0.111	0.119
St. dev.	0.177	0.176	0.172
Alpha	0.108	0.110	0.113
t-stat.	2.4	2.5	2.8

B. Maximum distance of selected words

	5	10	50
Mean	0.122	0.111	0.107
St. dev.	0.140	0.176	0.161
Alpha	0.122	0.110	0.112
t-stat.	3.0	2.5	2.6

C. Sentiment dictionary

	Harvard IV Dictionary	Loughran & McDonald (2011)	Mohammad & Turney (2013)
Mean	0.111	0.065	0.110
St. dev.	0.176	0.187	0.154
Alpha	0.110	0.063	0.112
t-stat.	2.5	1.3	2.5

D. Weight adjustment parameter

	1	1.5	2
Mean	0.110	0.111	0.092
St. dev.	0.175	0.176	0.165
Alpha	0.105	0.110	0.092
t-stat.	2.5	2.5	2.1

Notes: Return of the mean portfolios computed averaging the returns of the portfolios associated to 1 through 5 lags in formation year: $\overline{r}^{P(t)} = \sum_{l=1}^{5} r_t^{P(t-l)}$. Alpha is the estimated constant for the model: $r^{P(t)} = \alpha + \beta r_t^m + \epsilon_t$ for $l = 0,1,\ldots,6$. Where r_t^m is the market return and ϵ_t is an error term. t-statistics are corrected for heterocedasticity.

were selected. In the modified implementation presented in panel B of Table 3, distances of 5 and 50 are considered. It is verified that minor informational gains result from imposing a smaller distance. But the results are not significantly altered in any of the two cases.

Following Tetlock (2007), the sentiment indices were built exploiting the negative category in Harvard IV dictionary. Alternative lists of words could have been exploited. For example, Loughran and McDonald (2011) developed a list of negative words for financial contexts, while Mohammad and Turney (2013) constructed a dictionary of words employing a novel form of online collaboration. Panel C shows that our original exercise displays results that dominate those associated to the list of negative terms generated by Loughran and McDonald (2011) and are similar to the ones associated to the list generated by Mohammad and Turney (2013).

Finally, the standardization of individual country sentiment indices is considered. As described in the previous section, each country index was standardized using historic values of the average index and a historic metric of variability. In that exercise, the weight allocated to past values decreased with distance at a rate of 50% according to the rule: $\omega(t+1) = (3/2)\,\omega(t)$. In other words, the weight adjustment parameter was set equal to 3/2. According to the results shown in panel D, the results still hold when the rate at which the weight drops with distance is altered. It is observed that doubling the rate of increment in weights and keeping the weights constant result in similar properties for the distribution of returns of the contrarian sentiment long-short portfolio.

IV. Alternative sources of return prediction: past returns or economic growth

In this work, a novel source of persistent return predictability is identified. Extreme negative and positive opinions are found to predict expected returns. Shared views are computed using information distributed in the economic press. One relevant question is the extent to which similar return predictability could be achieved using more standard forms of information such as past returns or recent performance in terms of economic growth. This is a sensible consideration since periods of intense optimism are likely to be associated to positive returns and growth accelerations while periods of intense pessimism are commonly associated to negative returns and poor economic growth. Additionally, conventional opinions are measured with noise, hence it is plausible that alternative sources of information can result in more predictability.

In this section, two alternative sources of information are considered. A portfolio strategy that takes into account past economic growth is implemented. This portfolio strategy can be thought to evaluate the extent to which return predictability is explained by naïve projection of recent economic growth performance. Also, the case of portfolio strategies based on cumulative returns in the most recent years is considered. This strategy can be linked to the well-known literature that evaluates the performance of strategies that bet on previous losers and against previous winners.[22] In each case, the strategy is implemented following the algorithm described in the previous section.

Economic growth data from the World Bank is used to compute, for each year, the economic growth in the most recent four years.[23] Portfolio 1 is associated to the stock indices of countries in the decile with the largest economic growth. Portfolio 2 is the portfolio associated to the stock indices of countries in the decile with the lowest economic growth.

Similarly, in the case of strategies that exploit past returns, Portfolio 1 is associated to the stock indices that show the largest cumulative returns for the previous four years. Portfolio 2 is the portfolio associated to the stock indices with the lowest cumulative returns. The stock indices selected correspond to the top and bottom deciles respectively.

Table 4 describes the returns associated to the two portfolio strategies described above. For the strategy based on previous economic growth, the mean return associated to high growth is below the growth return associated to low economic

Table 4. Return and performance of alternative portfolio strategies

	GDP growth sorts			Past return sorts		
	High (H)	Low (L)	L-H	High (H).	Low (L)	L-H
Mean	0.024	0.079	0.055	0.019	0.074	0.055
St. dev.	0.404	0.385	0.353	0.345	0.408	0.252
Alpha	-0.052	0.009	0.061	-0.036	0.015	0.051
t-stat.	-1.1	0.2	0.7	-1.6	1.2	0.8
Correl. w/sentiment portfolio	0.53	0.66	0.71	0.75	0.57	0.70

Notes: Return of the mean portfolios computed averaging the returns of the portfolios associated to 1 through 5 lags in formation year : $\overline{r}^{P(t)} = \sum_{l=1}^{5} r_t^{P(t-l)}$. Alpha is the estimated constant for the model: $r^{P(t)} = \alpha + \beta r_t^m + \epsilon_t$ for $l = 0,1,\ldots,6$. Where r_t^m is the market return and ϵ_t is an error term. t-statistics are corrected for heterocedasticity.

[22] See Bondt and Thaler (1985) for the seminal contribution and Asness et al. (2013) for a recent comprehensive evaluation.

[23] The data is available at http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG.

growth. The estimated return is 6.1% when the market factor model is estimated. Nevertheless, the estimated parameter is not significantly different from zero. Despite the difference in terms of return predictability, it must be noted that there is a strong association between the return of the long short portfolio strategy based on extreme opinions and the return of the long short portfolio strategy based on economic growth. More precisely, the correlation coefficient for the returns of the respective long-short portfolios is 0.71.

Similar results are observed in the case of portfolio strategies that exploit information on past returns. The mean return for the portfolio of past winners is 1.9% while the mean return for past losers is 7.4%. The difference has the expected signed but is significantly smaller than the difference observed in the case of sentiment portfolios. According to the estimated market factor model the abnormal return is 5.1% but statistically it is not significantly different from zero. Despite the difference in performance, it is clear that the sentiment based strategy and the strategy based on past returns are strongly associated. The correlation coefficient for the returns of the long-short portfolios is 0.70.

In summary, it has been found that strategies based on past returns and economic growth are strongly linked to portfolio strategies based on sentiment indices. On the other hand, these alternative portfolio strategies are unable to replicate the performance of the sentiment based strategies. This suggests that the sentiment indices are able to reflect information on subjective states that cannot be captured with similar precision by alternative simpler indicators.

V. Performance of contrarian strategies at higher frequencies

So far the analysis has been carried out using annual returns data. This is consistent with the focus of this work on associations between data with multiple year lags. As shown in section III, sentiment indices are found to anticipate returns differentials up to five years after sentiment levels are measured. In favor of annual return data analyses, it must be noted that they exclude high frequency noise that might hide long term associations. In addition, comparable stock market data for long periods is more easily available at annual frequencies.

On the other hand, one shortcoming of low frequency analyses is given by the relatively small number of observations. Statistical tests are more reliable under a larger set of observations. Additionally, a monthly analysis can inform about the short term risk associated to exploiting this long term patterns in the data. In this section, the analysis of section III is replicated using monthly returns data.

As indicated, available data covers a shorter time span and a smaller set of countries. The sample period goes from January 1999 through July 2014. The number of sampled countries drops from 34 in the annual analysis to 30 in the monthly evaluation.[24] Dollar returns were computed using the stock index denominated in local currency and the dollar exchange rate for the last day of the month. The source for monthly stock market indices is Bloomberg. For most countries, exchange rate data corresponds to the daily series provided by the Federal Reserve Bank of St. Louis.[25] For eight countries, this data was not available from that source and was obtained from a private data provider.[26]

Table 5 shows results for the monthly frequency analysis. The average returns presented in panel A show that the returns following periods of high sentiment (high pessimism) are, on average, higher than the returns that follow periods of low sentiment (low pessimism). As in the annual analysis, the difference is the largest for five year lags. In this case, monthly average returns differ by 0.95%. This gap is similar to that observed in the annual returns analysis. The findings are also replicated when differences in volatility are evaluated. For example, portfolios built using five-year lagged sentiment metrics show no difference in terms of the standard deviation. In both cases, the standard deviation of monthly returns is approximately 7.4%. Importantly, despite the differences in data coverage and high frequency noise, the estimated abnormal returns and associated t-statistics are also in line with the annual return analyses.

Panel B in Table 5 shows the performance of the average portfolio strategy that combines portfolios built using one through five-year lagged indices. The monthly return of the portfolio associated to lagged optimism is 0.15%. In contrast, the monthly return of the portfolio associated to lagged pessimism is 0.96%. For the portfolio associated to optimism, the estimated monthly abnormal return is above 0.5% and highly significant in statistical terms. The third column shows that the average monthly return of a long-short portfolio strategy is 0.8%. The associated abnormal monthly return is 0.79% and statistically significant. This abnormal return is approximately 10% in annual terms which is similar to the value observed for a different sample in the annual frequency analysis of section III.

---

[24] The list of countries 30 covered by the monthly returns dataset is given by: Argentina, Austria, Belgium, Brazil, Chile, Colombia, Denmark, Finland, Greece, India, Indonesia, Ireland, Malaysia, Mexico, New Zealand, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Singapore, South Africa, South Korea, Spain, Sweden, Taiwan, Thailand, Turkey and Vietnam.

[25] The series can be found at: https://research.stlouisfed.org/.

[26] The data from these countries (Argentina, Chile, Indonesia, Pakistan, Peru, Philippines, Turkey and Vietnam) was provided by OANDA (http://www.oanda.com/).

Table 5. Performance monthly of sentiment based portfolio strategies

A. Return and performance for different lags in portfolio formation

Lags	Portfolio 1 (Optimism)				Portfolio 2 (Pesssism)				Portfolio 2 - Portfolio 1			
	Mean	St. Dev.	Alpha	t-stat.	Mean	St. Dev.	Alpha	t-stat.	Mean	St. Dev.	Alpha	t-stat.
1	0.0015	0.0692	-0.005	-2.4	0.0108	0.0776	0.004	1.2	0.0093	0.0583	0.009	2.1
2	0.0012	0.0703	-0.006	-2.7	0.0079	0.0780	0.000	0.1	0.0067	0.0534	0.006	1.5
3	0.0012	0.0708	-0.006	-2.7	0.0080	0.0746	0.001	0.3	0.0068	0.0513	0.007	1.7
4	0.0005	0.0699	-0.007	-2.8	0.0081	0.0713	0.001	0.4	0.0076	0.0506	0.008	1.9
5	0.0033	0.0740	-0.004	-1.5	0.0128	0.0734	0.006	2.0	0.0095	0.0528	0.010	2.5
6	0.0050	0.0728	-0.002	-0.8	0.0110	0.0728	0.004	1.4	0.0060	0.0514	0.006	1.5

B. Return and performance of mean sentiment based-portfolios (1 through 5 lags)

	Portfolio 1	Portfolio 2	Portfolio 2 - Portfolio1
Mean	0.0015	0.0096	0.0080
St. dev.	0.0718	0.0770	0.0612
Alpha	-0.0056	0.0024	0.0079
t-stat.	-3.6	1.0	2.5

Notes: Panel A: Portfolio 1 has equal sized long positions in the stock indices of the countries of the most optimistic decile. Portfolio 2 has equal sized long positions in the stock indices of the countries in the most pessimistic decile. Alpha is the estimated constant for the model: $r_t^{P(t-l)} = \alpha + \beta r_t^m + \epsilon_t$ for $l = 0,1,\ldots,6$. Where, $r_t^{P(t-l)}$ is the return of the corresponding sentiment-based portfolio (Portfolio 1, Portfolio 2 or Portfolio 1 – Portfolio 2), r_t^m is the market return and ϵ_t is an error term. Panel B: mean sentiment-based portfolio corresponds to the average of the portfolios with 1 through 5 lags in portfolio formation year. The associated return is: $\overline{r}^{P(t)} = \sum_{l=1}^{5} r_t^{P(t-l)}$. t-statistics are corrected for heterocedasticity.

In summary, the analyses that use monthly return data are consistent with return predictability. Despite the high frequency noise that characterizes monthly returns and the reduced time and country sample coverage, the differences in return are statistically significant. The results serve as a robustness check of the exercises developed using annual return data.

VI. Conclusions

This study proposes a novel metric of conventional views to evaluate instances of overreaction in financial and macroeconomic contexts. More specifically, the metric uses content of the international economic press. The analysis focuses on the relative performance of stock market indices following periods of extreme opinions. The results show that strong positive views are associated to subsequent lower returns and strong negative views are associated to subsequent higher returns. The difference in returns is highly persistent; for example, returns for portfolios constructed using five year-old information show differences in means that are both statistically and economically significant.

The evidence is consistent with the occasional emergence of extreme shared views that result in mistaken valuations of a broad class of financial assets that are corrected in subsequent years. According to the results, predictable differences in risk cannot explain the differences in returns.

This evidence is relevant for the interpretation of dynamics in financial and macroeconomic contexts. The existence of recurrent, persistent and predictable overreactions has implications for both private actors' decision making and the design of public policies that aim for stability. Additionally, this evidence suggests that there is value in empirical analyses that exploit estimations of subjective states. Strategies based on past returns or economic growth records are not able to replicate the results attained when the sentiment indices are used.

References

Angeletos, George, Guido Lorenzoni, and Alessandro Pavan (2010), Beauty contests and irrational exuberance: A neoclassical approach. Working Paper 15883, NBER.

Aromí, J. Daniel (2013), Media news and the stock market: evidence for Argentina. IIEP-Baires. mimeo.

Aromí, J. Daniel (2014), Estados de opinión y ciclo económico: un estudio empírico para Argentina, IIEP-Baires, mimeo.

Asness, Clifford, Tobias Moskowitz, and Larry Pedersen (2013). Value and momentum everywhere. *Journal of Finance* 68: 929-85.

Baker, Michael, and Jeffrey Wurgler (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives* 21: 129-52.

Baker, Scott, Nicholas Bloom and Steven Davis (2012). Measuring economic policy uncertainty. Research Paper No. 13-02, University of Chicago Booth School of Business.

Banerjee, Abhijit (1992), A simple model of herd behavior. *Quarterly Journal of Economics* 107: 797-817.

Baytas, Ahmet and Nusret Cakici (1999). Do markets overreact: International evidence. *Journal of Banking & Finance* 23: 1121-44.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* 100: 992-1026.

Brock, William, and Cars Hommes (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control* 22: 1235-74.

Cooper, Russell, and Andrew John (1988). Coordinating coordination failures in Keynesian models. *Quarterly Journal of Economics* 103: 441-63.

De Bondt, Werner and Richard Thaler (1985). Does the stock market overreact? *Journal of Finance* 40: 793-805.

De Long, J. Bradford, Andrei Shleifer, Lawrence Summers and Robert Waldmann (1990). Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance 98:* 379-95.

Diamond, Peter (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy* 90: 881-94.

Doms, Mark, and Norman Morin (2004). Consumer sentiment, the economy, and the news media. Working Paper 2004-09, Federal Reserve Bank of San Francisco.

Dumas, B., K. Lewis, and E. Osambela (2011). Differences of opinion and international equity markets. Working Paper No. 16726, NBER.

Economist (2009). Brazil takes off. November 14.

Economist (2013). Has Brazil blown it? September 28.

Eyster, Erik, and Matthew Rabin (2010). Naive herding in rich-information settings. *American Economic Journal: Microeconomics* 2: 221-43.

Eyster, Erik, and Matthew Rabin (2014). Extensive imitation is irrational and harmful. *Quarterly Journal of Economics* 129: 1861-1898.

Garcia, Diego (2013). Sentiment during recessions. *Journal of Finance*. 68: 1267-300.

Garcia, David, Antonios Garas, and Frank Schweitzer (2012). Positive words carry less information than negative words. *EPJ Data Science* 1: 1-16.

Gentzkow, Matthew, and Jesse Shapiro (2010. What drives media slant? Evidence from US daily newspapers. *Econometrica* 78: 35-71.

Golub, Benjamin, and Matthew Jackson (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics* 2: 112-49.

Heymann, Daniel, and Pablo Sanguinetti (1998). Business cycles from misperceived trends. *Economic Notes* 27: 205-32

Hong, Harrison, and Jeremy Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance* 54: 2143-84.

Hwang, Byoung-Hyoun (2011). Country-specific sentiment and security prices. *Journal of Financial Economics* 100: 382-401.

Jansen, W. Jos, and Niek Nahuis (2003). The stock market and consumer confidence: European evidence. *Economics Letters* 79: 89-98.

Jegadeesh, Narasimhan, and Sheridan Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance* 56: 699-720.

Kindleberger, Charles, and Robert Aliber (2011). *Manias, panics and crashes: a history of financial crises*. Palgrave Macmillan.

Krol, Robert (2014). Economic policy uncertainty and exchange rate volatility. *International Finance* 17: 241-56.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance* 49: 1541-78.

Loughran, Tim, and Bill McDonald (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66: 35-65.

Milani, Fabio (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics* 54:2065-82.

Mohammad, Saif M., and Peter Turney (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29: 436-465.

Mullainathan, Sendhil (2002). Thinking through categories. MIT. mimeo.

Mullainathan, Sendhil, and Andrei Shleifer (2005). The market for news. *American Economic Review* 95: 1031-53.

Obstfeld, Maurice (1996). Models of currency crises with self-fulfilling features. *European Economic Review* 40: 1037-47.

Schmeling, Maik (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance* 16: 394-408.

Shiller, Robert (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71: 421-36.

Shiller, Robert (2005). *Irrational exuberance*, 2nd ed. Princeton University Press.

Shleifer, Andrei, and Robert Vishny (1997). The limits of arbitrage. *Journal of Finance*. 52: 35-55.

Tetlock, Paul, (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62: 1139-68.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 75-103

DETERMINANTS OF INFLATION DIFFERENTIALS IN THE EURO AREA: IS THE NEW KEYNESIAN PHILLIPS CURVE ENOUGH?

Sérgio Lagoa*
Instituto Universitário de Lisboa (ISCTE-IUL), DINÂMIA'CET-IUL

Submitted December 2014; accepted May 2016

In the euro area, inflation rates diverged after the creation of the single currency, and started to converge again from mid-2002. It is against this background that the paper studies the determinants of inflation differentials in the euro area. We start by using the New Keynesian Phillips Curve (NKPC) to explain inflation differences for a panel of countries. Exchange rate movements and expected inflation in particular play an important part in bringing about diverging inflation dynamics, while lagged inflation does not. The Incomplete Competition Model (ICM) adds explanatory power to the NKPC in describing inflation dynamics across countries. The latter model does not encompass ICM, and the variables proposed by the ICM are statistically significant: the growth in nominal Unit Labour Cost and the long-run disequilibrium between prices and costs explain inflation differentials.

JEL classification codes: C23, E12, E31, F41
Key words: inflation, business cycles, convergence, New Keynesian Phillips Curve, Incomplete Competition Model

I. Introduction

Since the creation of the European Exchange Rate Mechanism (ERM) in 1979, monetary and financial convergence in the euro area has been accompanied by inflation convergence. However, some inflation divergence did occur after the introduction of the euro (Lane 2006; Busetti et al. 2007), as can be observed in

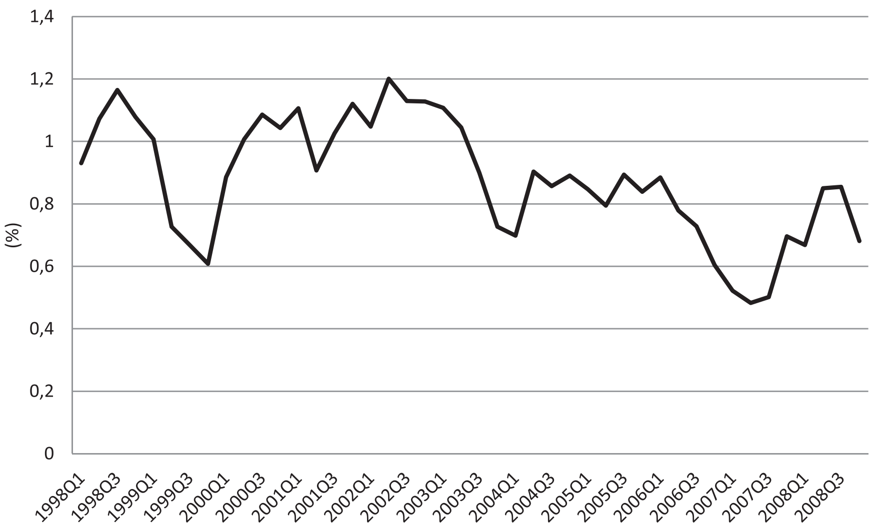
---

* ISCTE-Instituto Universitário de Lisboa, Avenida das Forças Armadas, 1649-026 Lisboa, Portugal; telephone 00351217650318; email sergio.lagoa@iscte.pt. I thank Stephen G. Hall, Jorge M. Streb (the Editor), and two anonymous referees for valuable comments. This paper has benefited from the financial support of *Fundação para a Ciência e a Tecnologia, Ministério da Ciência e Tecnologia*, Portugal.

Figure 1. Due to the nominal convergence effort before the creation of the euro, the cross section standard deviation of inflation rates in the euro area decreased to 0.6% in the fourth quarter of 1999.[1] Subsequently, inflation differentials increased to 1.2% in the second quarter of 2002. After this peak, inflation dispersion decreased again to the lowest level ever observed of 0.48% in the second quarter of 2007. In the first years of the euro (1999-2002), Greece, Ireland, the Netherlands, Portugal and Spain had the highest inflation rates.

As highlighted by the Optimum Currency Area literature, large inflation differentials may undermine the success of a monetary union. Differences in inflation rates can be caused by temporary asymmetric shocks, such as demand shocks, but the ability to deal with these impacts is limited in the absence of a national monetary policy. Inflation differentials cannot be corrected by nominal currency depreciation of high-inflation countries. Although countries may use expansionary fiscal policy to solve the problem of deflationary shocks, this can lead to a violation of the

Figure 1. Cross section standard deviation of inflation rates after 1998



Source: authors' calculations based on Eurostat data.

[1] In the empirical results of this paper, "euro area" only refers to 12 countries, the original 11 plus Greece: Austria, Belgium, Finland, France, Greece, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. Data: annual inflation rates based on quarterly CPIs: $(p_t/p_{t\text{-}4}-1)*100$. For each quarter, we obtained the standard deviation for the group of 12 countries.

Stability and Growth Pact with negative effects on the euro area financial markets (Honohan and Lane 2003).

Inflation differentials in the euro area are larger and more structural than in the US because the mechanisms for adjustments to asymmetric impacts are less effective (Zdárek and Aldasoro 2009). It is well known that the euro area has less wage flexibility and wages are influenced by national labour market institutions, there is lower labour mobility across countries, and the EU budget does not permit significant transfers to countries in crisis. European Treaties also limit national fiscal policies when they try to act as effective adjustment mechanisms, and in some cases these policies even create inflation differentials (ECB 2003). Households' financial portfolios are poorly diversified across euro area countries, thus making financial markets a weak mechanism for adjusting to asymmetric shocks (Lane 2006).

Persistent economic mechanisms put further restrictions of dealing with asymmetric shocks. If the labour market is not perfectly flexible, with current rather than future inflation determining wages growth, higher inflation may lead to higher wage growth and thus start an upward spiral of wage growth and inflation. Therefore, if inflation persists, temporary asymmetric shocks may produce persistent differences in inflation (Hofmann and Remsperger 2005).

Additionally, the creation of the euro produced inflation differentials with destabilising macroeconomic consequences. Convergence to the euro meant a bigger decline in real interest rates in peripheral countries, causing faster growth in credit, house prices, aggregate demand, and therefore inflation. This one-off expansionary shock dissipated over time, notably because higher inflation led to the real appreciation of currencies.

Higher than average inflation rates in a monetary union produce lower than average real interest rates, which may lead to both excessive debt accumulation and growth of property prices along with the subsequent painful adjustment process. This can then exacerbate the differences in business cycles among European countries, further widening inflation differentials in a cycle of divergence (Honohan and Lane 2003; Dullien and Fritshe 2008). Indeed, Vines, Kirsanova and Wren-Lewis (2006)'s theoretical model shows that when there is significant persistence of inflation, countries in a monetary union may be subject to large business cycles after asymmetric shocks.

Inflation differentials are sometimes the result of equilibrium mechanisms due to the convergence of long-run price levels. Inflation differences can also replace nominal exchange rate adjustments, as countries with low levels of economic growth, inflation and wage growth gain external competitiveness (Lane 2006).

The euro area has two empirically relevant stabilising mechanisms (Hofmann and Remsperger 2005). Firstly, GDP growth in one country has positive output spillover effects on the others, reducing inflation differentials. Secondly, the real exchange rate acts as a correcting mechanism: countries with higher than average inflation rates will face real appreciation of currencies that reduces demand and inflationary pressures.

Summing up, despite equilibrium mechanisms, inflation differentials are potentially dangerous for a monetary union. This constitutes the motivation for studying the determinants of inflation differences and the respective correcting mechanism in the euro area before the Euro Sovereign Debt Crisis. We do not include this crisis period, which in some economies was dominated by strong financial constraints on private and public sectors, in order to study inflation adjustments outside a crisis situation. Our study presents two distinctive features. Firstly, we test how inflation and exchange rate expectations affect inflation divergence. Expectations have been ignored in the literature despite their importance in explaining national inflation rates. Next, the New Keynesian framework is tested to see whether it provides a complete description of inflation differentials, and is assessed against the Imperfect Competition Model (ICM). As a by-product of the convergence analysis, we estimate the New Keynesian Phillips Curve (NKPC) for the euro area using panel data. This contributes to the scarce literature on the NKPC using panel data (some examples are Bjørnstad and Nymoen 2008; Paloviita 2008).

Our empirical evidence shows that expectations of both inflation and exchange rates are statistically significant for inflation differences and their introduction changes the significance of other variables. Moreover, the only relevant business cycle indicator for the explanation of inflation divergence is labour cost growth. The equilibrium conditions for prices suggested by the ICM also help explain differences in inflation rates, and that model is not encompassed by the NKPC when explaining inflation differences. Lastly, our panel data evidence supports the NKPC for national inflation rates and the existence of the cost channel.

The remainder of the paper is organised as follows. In Section II we revise the main determinants of inflation differentials in the euro area. Section III estimates a model for inflation differentials using the NKPC. Section IV assesses whether the ICM can explain inflation differentials better than the NKPC. Finally, Section V concludes.

II. Literature on the determinants of inflation differentials in the euro area

There are several possible determinants of inflation differentials in a monetary union such as the euro area. Firstly, differences may emerge in countries' business cycles due to supply shocks (e.g., oil price) or domestic demand shocks, which can result from differences in fiscal policy, country-specific non-policy demand shocks (e.g., taste shocks), or asymmetric effects of common demand shocks. These demand shocks can be induced by monetary policy or exchange rate movements. In fact, the common policy interest rate may have different impacts on each country due to distinctions in financial and economic structures (Hofmann and Remsperger 2005). Moreover, exchange rate evolution can cause inflation differentials even though countries share a common currency because differences in trade partners may mean that national effective exchange rates respond differently to the euro's evolution. The weight of imported consumption goods and inputs from outside the euro zone also differs from country to country.

It should be noted that the exchange rate not only affects inflation through aggregate demand, but also through its direct impact on import prices and inflation. The same could be said for the nominal interest when the cost channel is relevant: it has a direct and indirect impact on inflation (Ravenna and Walsh, 2006). The interest rate affects inflation both directly due to its effect on total wage costs, and indirectly because of its effect on the output gap.

Asymmetric demand shocks may occur due to differences in consumption patterns, which also imply that the weight of each sub-index of products in the Harmonized Index of Consumer Prices (HICP) differs across countries. As a result, symmetric changes in the prices of products across the monetary union imply different inflation rates measured by the HICP. However, this effect has not played a relevant role in explaining inflation differentials in the euro area (Hofmann and Remsperger 2005; ECB 2003).

On the structural side, a monetary union may have inflation differentials due to price level convergence resulting from both converging tradable and non-tradable prices (ECB 2003; Hofmann and Remsperger 2005). Prices of tradable goods converge because of the integration of markets, probably boosted by a single currency. In turn, the euro has raised GDP per capita (Conti 2014) and fostered real income convergence, which may lead to the convergence of prices of non-tradable goods, as explained by the Balassa-Samuelson effect.

Turning now to empirical models of determinants of inflation differences in the euro area, Honohan and Lane (2003) found that the output gap, the change in the nominal effective exchange rate (NEER) and price level convergence were significant in explaining inflation differentials in the euro area for the period 1998-2001. Hofmann and Remsperger (2005) examined national inflation rates as opposed to inflation differentials and found that the real exchange rate may act as a correcting mechanism because of its direct effect on inflation. The currency in countries with high inflation experiences real appreciation that has the direct effect of lowering inflation.

In line with Honohan and Lane (2003), Rogers (2002) conclude for the EMU-11 that CPI inflation differentials in the 1997-2000 period were fundamentally explained by the lagged price level, output gap and trade openness;[2] the price level had the expected negative coefficient and the two other variables had positive coefficients. The lagged per capita GDP also had a negative effect on inflation differentials at a 10% level of significance. The significance of the price level was not sufficiently robust to withstand more substantive analyses and most inflation differences were accounted for by factors other than convergence of prices.

Data used by Angeloni and Ehrmann (2007) covered one year more than that of Honohan and Lane (2003); taking into account data revisions, they confirm that the exchange rate is a determinant of inflation differentials albeit with a weak statistical significance. In contrast, the significance of the output gap and the lagged price level increases in their estimates.

An updated study by Honohan and Lane (2004) covering two additional years of annual data to obtain a sample for the 1999-2003 period is unable to confirm their previous result for the significance of the change in NEER to explain CPI inflation differentials, although the output gap remains significant. It seems that NEER affects inflation through the output gap in particular. To make evidence even more complex, when using *quarterly* data for 1999Q1-2004Q1, they conclude that the level of NEER explains CPI inflation differentials, but that output gap does not. Regarding exchange rates, they argue that in a monetary union national inflation rates act to correct misalignments in exchange rates: when the euro is under-valued, the increase in inflation acts as a correction mechanism, reducing external competitiveness, especially for countries more exposed to extra-euro trade.

It is clear from the above discussion that there is no consensus in the literature on the significance of output gap or NEER in explaining inflation differences in

[2] It can be argued that the relevance of trade openness occurs because of the exchange rate.

the euro area. To contribute to the clarification of the relationship between inflation and the business cycle, we will use real ULC as an alternative to output gap. The New Keynesian literature proposes that this is the correct driver of inflation; it also stresses that inflation is forward-looking, with lagged inflation playing a secondary role. Assessing the importance of lagged inflation is of relevance because if inflation is sufficiently persistent, temporary asymmetric shocks of demand and supply may cause persistent inflation differentials (Hofmann and Remsperger 2005). Besides its contemporaneous value, we also use forward and lagged exchange rate terms to clarify the variable's impact on inflation differentials. The nominal interest rate may also play a direct role in inflation divergence if the cost channel is relevant.

Finally, although fiscal deficits and the real interest rate may also have contributed to inflation differentials, this was probably due essentially to the output gap. Along this line, Honohan and Lane (2003) found that after controlling for output gap, fiscal positions did not have a statistically significant effect on inflation divergence in the euro area between 1999 and 2001.

III. Inflation differentials and the NKPC

In order to highlight the difference between factors influencing inflation and inflation differentials, we start by explaining national inflation rates and then analyse inflation differentials. The open economy NKPC to describe national inflation rates is:

$$\pi_{i,t} = \gamma_f E_t \pi_{i,t+1} + \gamma_b \pi_{i,t-1} + \gamma_{mc} \widehat{mc}_{i,t} + \gamma_e \Delta e_{i,t} + \gamma_{ef} E_t \Delta e_{i,t+1} + \gamma_{eb} \Delta e_{i,t-1} + \varepsilon_{i,t} \,, \tag{1}$$

where $\pi_i = p_{i,t} - p_{i,t-1}$ is CPI inflation in t, $p_{i,t}$ the log of CPI, $\widehat{mc}_{i,t}$ is the marginal cost in percentage deviation from the steady state, $mc_{it} - mc^{ss}$ (with both marginal costs defined in logs), and $\Delta e_{i,t}$ is the change in the log of the nominal effective exchange rate. The marginal cost is $mc_{it} = i_{i,t} + s_{i,t} - \log(\alpha_n)$, where $i_{i,t}$ is the log of the nominal interest rate of country i, $s_{i,t}$ the log of labour income share, and α_n is the elasticity of labour in the Cobb-Douglas production function.

Expected inflation affect current inflation because agents are forward-looking and prices are rigid. In addition, previous studies (for example, Galí and Gertler 1999; Galí, Gertler and López-Salido 2001) have shown that a proportion of agents have backward-looking expectations, justifying the introduction of lagged inflation. The business cycle's effect in inflation is captured by the labour share, which takes into account wages (w_t) and labour productivity (pr_t): $s_t = ulc_t - pd_t = w_t - pr_t - pd_t$, with pd_t as the GDP deflator. This Phillips Curve also includes open-economy variables in the spirit of Batini, Jackson and Nickell (2005), with the change in the nominal exchange rate expressing the impact of import prices on CPI inflation. We expect that an appreciation of the euro, i.e., an increase in $\Delta e_{i,t}$, to have a negative impact on inflation ($\gamma_e < 0$). The expected and lagged values of change in the exchange rate are introduced due to the assumption that expected and lagged domestic inflation affect present domestic inflation (Kara and Nelson 2003). The lag and lead exchange rate coefficients are expected to be positive (γ_{ef} and $\gamma_{eb} > 0$). Finally, the variable $i_{i,t}$ affects inflation because it is assumed that firms have to pay wages before receiving the income from sales, and so they resort to credit (Barth and Ramey 2001). As a result, the nominal interest rate affects the marginal cost, which is commonly called the cost channel (Ravenna and Walsh 2006).

If marginal cost is not expressed in deviations from the steady state, the Phillips curve expressed by equation (1) can be written as:

$$\pi_{i,t} = \alpha + \gamma_f E_t \pi_{i,t+1} + \gamma_b \pi_{i,t-1} + \gamma_s s_{i,t} + \gamma_{in} i_{i,t} + \gamma_e \Delta e_{i,t} + \gamma_{ef} E_t \Delta e_{i,t+1} + \gamma_{eb} \Delta e_{i,t-1} + \varepsilon_{i,t} \, , \quad (2)$$

with $= -\gamma_{mc}\,[mc^{ss} + \log(\alpha_n)]$. Then, the constant α includes the common steady-state marginal cost. The NKPC can also be defined using the output gap, $x_{i,t}$, to measure the business cycle's impact on inflation:[3]

$$\pi_{i,t} = \alpha + \gamma_f E_t \pi_{i,t+1} + \gamma_b \pi_{i,t-1} + \gamma_x x_{i,t} + \gamma_{in} i_{i,t} + \gamma_e \Delta e_{i,t} + \gamma_{ef} E_t \Delta e_{i,t+1} + \gamma_{eb} \Delta e_{i,t-1} + \varepsilon_{i,t} \, . \quad (3)$$

When estimating equations (2) and (3), data poolability was assumed, i.e., that the equation's coefficients are the same for all countries; Bjørnstad and Nymoen (2008), stress that this has the advantage of bringing efficiency gains. In the euro area, the poolability assumption makes sense as countries are relatively homogeneous: they have been converging in nominal and real terms and share similar monetary and fiscal policy frameworks. The use of panel data with the poolability assumption is also advisable because inflation convergence is an aggregate phenomenon, simultaneously involving the dynamic evolution of a group of countries. In addition, there is no need to explicitly measure common factors when using a panel (which

[3] Rotemberg and Woodford (1997) conclude for $\widehat{mc}_{i,t} = x_{i,t}$ using sticky prices, complete nominal wages flexibility and absence of variable capital. Ravenna and Walsh (2006) assume the existence of the cost channel and express the marginal cost as depending on output gap and the nominal interest rate: $\widehat{mc}_{i,t} = \theta x_{i,t} + i_{i,t}$, where θ is a parameter dependent on other structural parameters.

always involves some aggregation problems), because they can be captured by time dummies (as we will see below).

Turning now to inflation differentials, if equation (2) is valid for each country, it can also be applied to the euro area inflation rate, $\pi_{euro,t}$. The inflation differential for country i is simply the difference between its inflation rate and the euro area inflation rate, $\pi_{i,t} - \pi_{euro,t}$. Taking into account the determinants of national and euro area inflation rates suggested by equation (2), the inflation difference for country i can be expressed as:

$$\begin{aligned}\pi_{i,t}-\pi_{euro,t} &= \gamma'_p\left(pl_{i,t-1}-pl_{euro,t-1}\right)+\gamma'_f E_t\left(\pi_{i,t+1}-\pi_{euro,t+1}\right)+\gamma'_b\left(\pi_{i,t-1}-\pi_{euro,t-1}\right)\\ &+\gamma'_s\left(s_{i,t}-s_{euro,t}\right)+\gamma'_{in}\left(i_{i,t}-i_{euro,t}\right)+\gamma'_e\left(\Delta e_{i,t}-\Delta e_{euro,t}\right)\\ &+\gamma'_{ef}E_t\left(\Delta e_{i,t+1}-\Delta e_{euro,t+1}\right)+\gamma'_{eb}\left(\Delta e_{i,t-1}-\Delta e_{euro,t-1}\right)+u_{i,t}.\end{aligned} \tag{4}$$

Here, the price level pl_{t-1} is introduced to capture the price convergence effect (Honohan and Lane 2003), as countries with higher price levels are expected to have lower inflation ($\gamma_p < 0$).

The euro area variables can be combined linearly in a time dummy ϕ_t (Honohan and Lane, 2003). Thus, inflation differentials can be expressed as:

$$\begin{aligned}\pi_{i,t} &= \phi_t+\gamma'_p pl_{i,t-1}+\gamma'_f E_t\pi_{i,t+1}+\gamma'_b\pi_{i,t-1}+\gamma'_s s_{i,t}+\gamma'_{in}i_{i,t}\\ &+\gamma'_e\Delta e_{i,t}+\gamma'_{ef}E_t\Delta e_{i,t+1}+\gamma'_{eb}\Delta e_{i,t-1}+u_{i,t}.\end{aligned} \tag{5}$$

The model was estimated using a panel of 12 euro area countries: the 11 founders and Greece. The panel comprises the period 1999Q1-2008Q4 and is unbalanced only when interest rate on loans or the price of imports are used. [4] Data is described in the Appendix. It is worth noting that in the empirical application $pl_{i,t-1}$ is a dummy that takes value one when the price level is above one, in which case the country's price level is above the European average. This procedure is justified by the fact that the price index is non-stationary.

The estimation was made using Panel GMM due to the presence of expectations. In order to estimate equation (2), expectations are replaced by observed values under the assumption of rational expectations. This assumption implies that agents'

[4] Before 2003Q1 there were no data available on the interest rate for Luxembourg and before 2000Q1 there were no data available on imports price for Ireland (see data description in the Appendix).

forecast errors are not correlated with information available when they form expectations. As a result, we obtain orthogonality conditions to apply GMM.

It is worth mentioning that we do not introduce country fixed effects for two reasons. First, expectations for inflation can accommodate differences in inflation rates that remain constant for the entire sample, without it being necessary to include a constant for this purpose. Second, fixed effects with a lagged dependent variable produce bias in the results.[5]

It is important to choose good instruments in GMM estimations. We follow the convention in the literature of using at least past information on the endogenous and forcing variables (Binder and Pesaran 1995). Two additional reasons justify the use of lag variables as instruments. Firstly, as the exchange rate may be endogenous, only lags of this variable can be used as instruments, as the lags are exogenous to period t. Secondly, information for period t may not yet be available when agents form expectations. Therefore, we used two lags of inflation as instruments (t-1 and t-2), one lag of the change in exchange rate, output gap, interest rate and price level.[6] We added some further instruments that proved to have a strong explanatory power in the first stage regression: CPI lagged two periods, one lag of both the change in import prices and real exchange rate, and a dummy variable for 1999Q1-2002Q4.[7,8]

When choosing instruments, their weakness should be tested because weak instruments are common in forward-looking models with rational expectations (Mavroeidis 2004). We performed this test using the first stage regression of $t+1$ inflation on the instruments. Then, we retained the F-statistic of the joint significance of instruments. The same was done for the variation of the exchange rate in $t+1$. The rule of thumb is that when the F-statistic is larger than 10, the existence of weak instruments can be ruled out (Stock, Wright and Yogo, 2002), which is almost always the case in our estimations.

As the models estimated are overidentified (i.e. the number of instruments is larger than the number of regressors), the correlation between the error term and the instruments can be tested with the J-test. When this test is applied in all of the following regressions, it indicates that instruments are not correlated with the error term.

---

[5] Omitting the unobserved fixed effects may also lead to bias. As a robustness check, we perform an estimation with fixed effects (Table 1, equation 4).

[6] The inclusion of an additional lag of inflation (t-2) is because lag t-1 is an explanatory variable in the hybrid NKPC.

[7] The dummy variable was introduced to accommodate the fact that retail interest rates were not harmonised before 2003Q1 (see Appendix).

[8] There are some small changes in instruments depending on the exact specification of the estimated equation. See notes to the tables.

Table 1. GMM estimation of the NKPC for a panel of 12 euro area countries

	(1) – Output gap	(2) - RULC	(3) – Output gap	(4) - RULC
C	-0.00088	-0.010	-0.00033	-0.011
	(0.00057)	(0.023)	(0.00063)	(0.037)
$\pi_{i,t+1}$	0.83***	0.89***	0.68***	0.84***
	(0.014)	(0.098)	(0.111)	(0.121)
$\pi_{i,t-1}$	0.13*	0.092	0.14**	0.117
	(0.073)	(0.070)	(0.068)	(0.080)
$x_{i,t}$	0.013	-	0.011	-
	(0.014)		(0.015)	
$s_{i,t}$	-	0.0020	-	0.0021
		(0.0050)		(0.0079)
$\Delta e_{i,t+1}$	0.057	0.044	0.059	0.062
	(0.058)	(0.057)	(0.060)	(0.058)
$\Delta e_{i,t}$	-0.045	-0.017	-0.057	-0.041
	(0.007)	(0.066)	(0.075)	(0.072)
$\Delta e_{i,t-1}$	0.051***	0.049***	0.049**	0.049**
	(0.018)	(0.018)	(0.017)	(0.019)
$i_{i,t}$	0.00026**	0.00022**	0.00030**	0.00044**
	(0.00010)	(0.00010)	(0.00012)	(0.00018)
$i_{i,t} \cdot D_t$	-0.00014**	-0.00011*	-0.00020**	-0.00017
	(0.000065)	(0.000064)	(0.000088)	(0.00011)
$pi_{i,t} - pd_{i,t}$	-	-	0.0084**	-
			(0.0034)	
Country fixed effect	No	No	No	Yes
F-stat 1st stage reg.:				
$\pi_{i,t+1}$	11.72	11.53	12.37	6.85
$\Delta e_{i,t+1}$	18.33	16.63	18.46	8.87
J-statistic	2.48	3.45	3.78	3.25
	[0.28]	[0.32]	[0.15]	[0.35]
Q (1) stat.	80.678	78.809	82.857	78.808
	[0.00]	[0.00]	[0.00]	[0.00]

Note: Panel GMM with period SUR weights and robust standard deviations. Instruments: Eq. (1): constant, $\pi_{i,t-1}$, $\pi_{i,t-2}$, $x_{i,t-1}$, $\Delta e_{i,t-1}$, $\Delta pi_{i,t-1}$, $q_{i,t-1}$, $i_{i,t-1}$, $pl_{i,t-1}$, $p_{i,t-2}$ and one dummy, D_t, that takes the value one for the period 1999Q1-2002Q4. The variable q is the real exchange rate. Eq. (2): the same as eq. (1) plus $s_{i,t-1}$.Eq. (3): the same as eq. (1) plus $pi_{i,t-1} - pd_{i,t-1}$. Eq. (4): the same as eq. (1) plus country dummies. (...) contain standard errors robust to arbitrary serial correlation and time-varying variances of the errors. [...] contain p-values.*** significance at 1%, ** at 5%, and * at 10%. Q(1) is the Ljung-Box statistics to test zero autocorrelation in the residuals up to lag 1.

Finally, we used a weighting matrix and standard deviations robust to arbitrary serial correlation and time-varying variances of the errors (White period method).[9]

Next, estimation results are analysed. Table 1, column (1), shows that we can replicate the traditional features of the hybrid NKPC. Coefficients of both lead and lag inflation are statistically different from zero, and their sum is less than one but not statistically different from one. Also, the forward component of inflation is larger than the backward component. Output gap has a positive but statistically insignificant effect on inflation. The cost channel is present due to the positive and significant effect of the nominal interest rate on inflation.[10]

Finally, the coefficients of the change in the NEER have the right signs and the coefficient associated with the lagged rate is statistically significant. If we replace the change in NEER with changes in Real Effective Exchange Rate or in import price deflator, we do not obtain more significant results for these variables.[11]

Lagoa (2014) suggests that the correct identification of the cost channel requires the introduction of the relative price of imports ($pi_{i,t} - pd_{i,t}$) in the Phillips Curve. Even after introducing this variable, the cost channel continues to be significant after 2002Q4, but it becomes insignificant before this date (Table 1, column 3).[12]

The results including country fixed effects are presented for comparison purposes in Table 1, column (4). All the main results remain valid, confirming their robustness.

It can be seen that residuals of the models are autocorrelated. This is because replacing the expectations for variables with observed values induces a first order moving-average structure in the error term (Pesaran 1987). We tackle this problem by using standard errors robust to autocorrelation.

As in the present paper, several studies confirm the statistical insignificance or the wrong sign of output gap in the NKPC (e.g., Galí and Gertler 1999; Galí, Gertler and López-Salido 2001); however, they obtain good results when using the real ULC.[13] When we use the real ULC instead of output gap, results remain unchanged and real ULC does not have a statistically significant effect on inflation (Table 1,

[9] The White estimator for the weighting matrix is based on the Panel Corrected Standard Error methodology (Beck and Katz 1995; Eviews 2007), where residuals are replaced by moment estimators of the unconditional variance.

[10] The p-value of the null hypothesis of "no interest rate effect on inflation between 1999Q1-2002Q4" is 0.0346. So we do not reject the null hypothesis at a 1% level of significance. The weaker evidence on the cost channel before 2003Q1 probably occurs because interest rate data before this quarter are not fully harmonised across countries.

[11] These results are available upon request from the authors.

[12] Before 2003Q1 the interest rate is only significant at a 10% level. The p-value of the null hypothesis of "no interest rate effect on inflation between 1999Q1-2002Q4" is 0.0776.

[13] Using non-standard measures of output gap, studies by Neiss and Nelson (2005) and Garrat, Lee, and Shields (2009) are among the few that obtain good empirical results when using the output gap in the NKPC.

column 2); nevertheless, it should be mentioned that the statistical insignificance of the real ULC in the NKPC is not unusual in the literature. In Bjørnstad and Nymoen (2008), which uses panel data, the real ULC has a negative sign and is not statistically significant. In a time series context, Bårdsen, Jansen and Nymoen (2004) show that the significance of wage share in Galí, Gertler and López-Salido (2001) for the euro area is not robust to small changes in the estimation methodology.

The fact that we are able to reproduce the basic characteristics of the Phillips curve found in estimations for individual countries constitutes evidence in favour of the poolability of the data.

Turning now to inflation differentials, estimating equation (5) shows that expected inflation is statistically significant and its coefficient is larger than in the equation for national inflation rates (Table 2, column 2). In contrast, lagged inflation is not statistically significant. Even though the coefficients of exchange rate are also not statistically significant, they have the right signs in t and t+1. However, the coefficient of the lagged change in exchange rate has the wrong sign, confirming that past dynamics do not seem to explain differences in inflation. In turn, output gap has a positive effect but is not statistically significant.[14]

Finally, the nominal national interest rate and the lagged price level dummy are not statistically significant. In Hofmann and Remsperger (2005) proxies of price level convergence are also not significant in explaining national inflation rates. Likewise, in Rogers (2002), the lagged price level is insignificant in explaining inflation differences when the Arelano-Bond GMM estimator was used. The fact that this variable is also statistically insignificant in our estimates probably means that the level of price convergence in the euro area was already high enough over the sample period. Indeed, Rogers (2007) shows that much of the price level convergence in Europe took place close to the completion of the Single Market in January 1993.

From the above results, we can conclude that the lagged inflation rate and nominal interest rate play a role in explaining national inflation rates, but not in explaining inflation differences across countries.

It should be noted that expectations play a central role in our results. If they are ignored, our results are similar to Honohan and Lane (2003), with output gap, the level of the nominal exchange rate and the lagged price level having a statistically significant impact on inflation differentials (Table 3). The presence of the level of exchange rate can be interpreted as national inflation rates acting to correct

[14] We also made an estimation (available upon request) with the relative price of imports, $pi_{i,t} - pd_{i,t}$, which had a positive but insignificant coefficient.

disequilibrium in the exchange rate. Another possible interpretation is that with imported inputs in production, the level of exchange rate directly affects marginal cost (Kara and Nelson 2003).

Table 2. Determinants of inflation differentials for a panel of 12 euro area countries. GMM estimation

	(1) – Output gap	(2) – RULC	(3) – Change in ULC	(4) – Change in ULC
c	0.00072	0.0142	0.00062	0.00059
	(0.00058)	(0.015)	(0.00060)	(0.00059)
$\pi_{i,t+1}$	0.877***	0.882***	0.775***	0.787***
	(0.122)	(0.108)	(0.128)	(0.127)
$\pi_{i,t-1}$	0.023	0.0190	0.050	0.050
	(0.064)	(0.062)	(0.066)	(0.066)
$x_{i,t}$	-0.018	-	-	-
	(0.025)			
$s_{i,t}$	-	-0.0029	-	-
		(0.0033)		
$\Delta ulc_{i,t}$	-	-	0.076**	0.074**
			(0.030)	(0.029)
$\Delta e_{i,t+1}$	0.251	0.283	0.386**	0.375**
	(0.178)	(0.160)	(0.165)	(0.161)
$\Delta e_{i,t}$	-0.046	-0.149	-0.282	-0.323*
	(0.229)	(0.205)	(0.210)	(0.25)
$\Delta e_{i,t-1}$	-0.032	-0.028	-0.025	-
	(0.050)	(0.050)	(0.053)	
$i_{i,t}$	0.000031	0.000061	0.000087	0.000081
	(0.00065)	(0.000056)	(0.000055)	(0.000053)
$dummy\ pl_{i,t-1}$	-0.00043	-0.0004	-0.00042	0.00037
	(0.00033)	(0.0002)	(0.00035)	(0.00033)
Time dummies	Yes	Yes	Yes	Yes
F-stat 1st stage reg.:				
$\pi_{i,t+1}$	18.67	18.36	18.07	18.07
$\Delta e_{i,t+1}$	111.76	109.46	111.60	111.60
J-statistic	3.59	3.69	2.91	4.14
	[0.30]	[0.44]	[0.40]	[0.53]
Q (1) stat.	113.69	110.80	103.95	109.25
	[0.00]	[0.00]	[0.00]	[0.00]

Notes: See notes to Table 1. Instruments: eq. (1): constant, $\pi_{i,t-2}$, $\Delta e_{i,t-1}$, $\Delta pi_{i,t-1}$, $q_{i,t-1}$, $i_{i,t-1}$, $pl_{i,t-1}$, $dummy\ pl_{i,t-1}$, $p_{i,t-2}$, $p_{i,t-3}$, $difp_{i,t-1}$ and time dummies. Eq. (2): the same as eq. (1) plus $s_{i,t-1}$. Eq. (3) and (4): the same as eq. (1) plus $\Delta ulc_{i,t-1}$.

Table 3. Determinants of inflation differentials ignoring expectations. GMM estimations for a panel of 12 euro area countries.

	c	$x_{i,t}$	$e_{i,t}$	$pl_{i,t-1}$
Coeff.	0.54***	0.11***	-0.117***	-0.031***
s.e.	(0.18)	(0.04)	(0.039)	(0.0082)
Time dummies:	Yes			
J-statistic:	2.59	Q (2) stat.:	291.12	
	[0.62]		[0.00]	

Notes: See notes to Table 1. Instruments: constant, $\pi_{i,t-1}$, $x_{i,t-1}$, $e_{i,t-1}$, $i_{i,t-1}$, $p_{i,t-2}$, $p_{i,t-3}$, $difp_{i,t-1}$ and time dummies.

Returning to the regressions with expectations, the statistical insignificance of output gap is an intriguing result, raising the question of whether the use of an alternative measure of business cycle would produce more significant results. We therefore replace output gap with real ULC, but this variable also proves statistically insignificant (Table 2, column 2). Overall, inflation expectations is the only element of the New Keynesian approach that is valid for inflation differences, and no significant role is found for interest rate, exchange rate and business cycle indicators.

There is some preliminary empirical evidence that wage growth is associated with different inflation dynamics in the euro area (ECB 2003). In addition, Lown and Rich (1997) were able to track inflation in the 1990s using a traditional Phillips curve augmented with nominal ULC growth. Following this evidence, nominal ULC growth was used instead of output gap or real ULC, and a positive and statistically significant coefficient was obtained for that variable (Table 2, column 3). Most other coefficients of variables remained roughly the same as when output gap was used; the expected change in the exchange rate is the exception as it becomes statistically significant. We can then conclude that the cyclical position influences inflation differentials if it affects the growth in nominal ULC; as this is not a proposed variable for the standard form of the NKPC, it is evidence of the curve's weakness in explaining inflation differences.

In the last estimation, the lagged change in NEER again has the wrong sign. When removed, the expected and current NEER become significant (Table 2, column 4). As in Honohan and Lane (2003), a depreciation of the euro in *t* tends to increase inflation differentials. This can be explained by the fact that countries with more imports from outside the euro area suffer higher imported inflation when the exchange rate depreciates. Different velocities of exchange rate pass-through can also account for the temporary impact of movements in the euro on inflation differentials (Honohan and Lane 2003). However, the exchange rate effect on inflation differentials will probably tend to decrease with time (Honohan and Lane 2003).

IV. Inflation differentials, imperfect competition model and the NKPC

Given the empirical relevance of nominal ULC, let us look at the imperfect competition model (ICM) of inflation, which defines a role for nominal ULC. Here, we add the cost channel to the ICM presented by Bjørnstad and Nymoen (2008). This model assumes that the price of domestically produced goods, pd_t, is set as a mark-up over nominal ULC and the nominal interest rate, and the mark-up depends on the relative price of domestic goods in terms of foreign goods, pi_t, (all variables are in logs):

$$pd_t = m_0 + m_1(pi_t - pd_t) + i_t + ulc_t, \tag{6}$$

where i_t is the gross nominal interest rate and m_0 is the steady-state mark-up. In equilibrium, there is a relationship between domestic prices on one hand, and ULC, nominal interest rate and import prices on the other hand. The nominal interest rate affects domestic prices because firms have to pay salaries in advance, i.e., due to the cost channel.

With a constant share of imports in consumption, $1 - \gamma$, the CPI is by definition:

$$p_t = \gamma pd_t + (1-\gamma) pi_t . \tag{7}$$

If we solve (7) for pd_t and replace the expression obtained in (6), after some manipulations we obtain:

$$p_t = \mu_0 + \mu_1(i_t + ulc_t) + (1-\mu_1) pi_t,$$

with $\mu_0 = m_0\mu_1$ and $\mu_1 = \gamma/(1 + m_1)$. Given that prices are often not in equilibrium, the model should be expressed in an equilibrium correction form, where:

$$\begin{aligned} \pi_t &= \mu_0\beta_1 + \alpha^f E_t\pi_{t+1} + \alpha^b \pi_{t-1} + \beta_1(ulc_{t-1} + i_{t-1} - p_{t-1}) \\ &+ \beta_2(ulc_{t-1} + i_{t-1} - pi_{t-1}) + \beta_3 \Delta ulc_t + \beta_4 \Delta pi_t + \beta_5 \Delta i_t , \end{aligned} \tag{8}$$

with all coefficients α and β positive, except β_2 that is negative.[15] When the last period ULC plus nominal interest rate is larger than the consumer price index, $ulc_{t-1} + i_{t-1} > p_{t-1}$, the disequilibrium is corrected with an increase in inflation in

[15] We do not test for cointegration for the two error correction terms due to the small number of years in the sample.

the current period. This occurs because ULC and nominal interest rate are excessively high compared with prices charged by firms. In turn, if in t-1 the ULC plus nominal interest rate is larger than imports price, $ulc_{t-1} + i_{t-1} > pi_{t-1}$, then in t inflation decreases.[16] In the last equation, it was assumed that the dynamic part of the NKPC is valid: α^f and α^b are different from zero.

The open economy NKPC can be expressed in an error correction model of the price level, similar to (8). The initial equation is:

$$\pi_t = \alpha^f E_t \pi_{t+1} + \alpha^b \pi_{t-1} + b\widehat{mc_t} + cz_t \,, \tag{9}$$

where $\widehat{mc_t} = \left(s_t + i_t - log(\alpha_n) - mc^{ss}\right)$, z_t is a vector containing open economy variables, like for example the change in real price of imports, $\Delta(pi_t - p_t)$; and s_t is the wage share, defined as

$$s_t = ulc_t - pd_t \,. \tag{10}$$

Using (7), (9) and (10), after some manipulations, we obtain:

$$\pi_t = \alpha + \frac{\alpha^f}{1+\frac{b}{\gamma}} E_t \pi_{t+1} + \frac{\alpha^b}{1+\frac{b}{\gamma}} \pi_{t-1} - \beta_1\left(p_{t-1} - \gamma ulc_{t-1} - (1-\gamma) pi_{t-1}\right)$$

$$+\beta\gamma\Delta ulc_t + \beta(1-\gamma)\Delta pi_t + \beta\gamma\Delta i_t + \psi z_t \,,$$

with $\alpha = -b\left(\log(\alpha_n) + mc^{ss}\right)$, $\beta = b/(\gamma + b)$ and $\psi = (c\gamma)/(\gamma + b)$. The last equation can be expressed as

$$\pi_t = \alpha + w^f E_t \pi_{t+1} + w^b \pi_{t-1} + \beta_1\left(ulc_{t-1} + i_{t-1} - p_{t-1}\right) + \beta_2\left(ulc_{t-1} + i_{t-1} - pi_{t-1}\right)$$

$$+\beta_3 \Delta ulc_t + \beta_4 (1-\gamma)\Delta pi_t + \beta_5 \Delta i_t + \psi z_t \,,$$

with $w^f = \dfrac{\alpha^f}{1+\frac{b}{\gamma}}$ and $w^b = \dfrac{\alpha^b}{1+\frac{b}{\gamma}}$, $\beta_1 = \beta$, $\beta_2 = -\beta(1-\gamma)$, $\beta_3 = \beta\gamma$, $\beta_4 = \beta(1-\gamma)$, and $\beta_5 = \beta\gamma$. This equation imposes three restrictions on the ICM: $H_0^a: \beta_1 + \beta_2 = \beta_3$, $H_0^b: \beta_4 = -\beta_2$ and $H_0^c: \beta_5 = \beta_3$. If z_t includes the change in imports price, then H_0^b

---

[16] Intuitively, in this situation the import price is low and the mark-up decreases, leading to a fall in inflation.

is no longer an imposition arising from the NKPC. The significance of expected inflation is also fundamental for the validity of the NKPC.

In relation to national inflation rates, Bjørnstad and Nymoen (2008) show with an annual panel of 20 OECD countries, from 1960 to 2004, that: (1) the ICM model encompasses the NKPC (H_0^a is rejected), and (2) the expected rate of inflation serves as a replacement for the ICM specific equilibrium correction terms. In other words, when equilibrium terms are included, the coefficient of expected inflation is no longer significant. This means that the omission of equilibrium correction terms creates an upwards bias in the estimate of α^f, explaining why the lead coefficient of inflation is significant in many estimates of the Phillips curve. Also, for the UK, Bårdsen, Jansen and Nymoen (2004) show that the introduction of two equilibrium correction terms, deviations from a long-run wage curve and an open economy price mark-up, makes forward inflation insignificant.

Based on the previous discussion, ICM is an alternative to the NKPC to explain inflation differentials. Therefore, for this purpose, we augmented the ICM in equation(8) with the lagged price level dummy and estimated it (Table 4, column 1). The null hypothesis H_0^a is rejected,[17] which means that the ICM model is not encompassed by the NKPC. In other words, it is better to use the ICM model than the NKPC because the former is an unrestricted version of the latter; moreover, the error correction variables are significant.

In addition, previous results obtained in this paper are confirmed: the relevance of expected inflation, the change in the nominal ULC, and change in import prices. We observe that even though the coefficient of expected inflation decreases with the introduction of the error correction variables, it continues to be statistically significant. This confirms the importance of forward inflation in explaining differences in inflation dynamics, a result that to some extent contradicts Bjørnstad and Nymoen (2008).

Although we found above that national inflation rates are persistent, lagged inflation does not explain inflation differentials. However, inflation depends on past economic conditions via the two error correction terms, notably when it includes the CPI ($ulc_{i,t-1} + i_{i,t-1} - p_{i,t-1}$); additionally, inflation differentials are explained by the variation in ULC, and if this is persistent (and it is well know that wages in Europe are sticky), then inflation depends on past economic evolution.

It should also be noted that even though the change in nominal interest rate does not have a significant effect on inflation differentials, its level is present in the long term marginal cost, which has a significant impact on inflation. In addition, note

---

[17] P-value of 0.0088.

that in the regressions of Table 2, columns (3) and (4), the expected change of the exchange rate is statistically significant. Nevertheless, if by analogy we introduce the expected change in import prices in equation (8), this term is statistically non-significant.[18]

Table 4. GMM estimation of the ICM for inflation differentials of 12 euro area countries, 1999Q1-2008Q4

	(1)	(2) - Periphery	(3) - Core
c	0.00050	0.0046***	-0.0015
	(0.0011)	(0.0017)	(0.0025)
$ulc_{i,t-1}+i_{i,t-1}-p_{i,t-1}$	0.0089**	0.0138***	0.0084
	(0.0035)	(0.0050)	(0.0060)
$ulc_{i,t-1}+i_{i,t-1}-pi_{i,t-1}$	-0.0087*	-0.0149***	-0.0073
	(0.0035)	(0.0049)	(0.0055)
$\pi_{i,t+1}$	0.7114***	0.4798**	0.7610***
	(0.1273)	(0.1860)	(0.1276)
$\pi_{i,t-1}$	0.0614	0.0319	0.0744
	(0.0674)	(0.0583)	(0.0796)
$\Delta ulc_{i,t}$	0.0547***	0.0847**	0.0241
	(0.0226)	(0.0330)	(0.0416)
$\Delta pi_{i,t}$	0.0690**	0.0178	0.0728
	(0.0317)	(0.0267)	(0.0480)
$\Delta i_{i,t}$	0.00065	-0.0014	0.0038
	(0.00096)	(0.0009)	(0.0028)
$\Delta i_{i,t}$	-0.00029	-0.00035	-0.00014
	(0.00017)	(0.00022)	(0.00037)
Time dummies	Yes	Yes	Yes
F-stat 1st stage reg.:			
$\pi_{i,t+1}$	17.18	7.48	9.95
J-statistic	9.62	9.75	4.5
	[0.21]	[0.20]	[0.72]
Q (1) stat.	57.09	35.88	44.50
	[0.00]	[0.00]	[0.00]
	[0.00]	[0.00]	[0.00]

Notes: Periphery includes Greece, Portugal, Ireland, Spain and Italy, while the Core includes Austria, Belgium, Finland, France, Germany, Luxembourg, and the Netherlands. Notes: See notes to Table 1. Instruments: constant, $\pi_{i,t-2} x_{i,t-1}, q_{i,t-1}, i_{i,t-1}, \Delta i_{i,t-1}, p_{i,t-2}, p_{i,t-3}$, $difp_{i,t-1}, \Delta pi_{i,t-1}, \Delta e_{i,t-1}, ulc_{i,t-1}+i_{i,t-1}-p_{i,t-1}, ulc_{i,t-1}+i_{i,t-1}-pi_{i,t-1}, \Delta ulc_{i,t-1}, pl_{i,t-1}, dummy\ pl_{i,t-1}$ and time dummies.

---

18 Results available upon request.

In order to analyse the impact of countries with a specific and crisis prone evolution, namely Greece, Portugal and Ireland, we removed each from the sample one at a time and re-estimated equation(8), and found that results basically remain unchanged.[19]

Table 5. GMM estimation of the ICM for inflation differentials with different measures of inflation, 1999Q1-2008Q4

	(1) - services	(2) – core inflation	(3) – non-adm. prices
c	-0.00055	-0.00073	0.00038
	(0.00095)	(0.0012)	(0.00169)
$ulc_{i,t-1}+i_{i,t-1}-p_{i,t-1}$	-0.00055	0.0063	0.0121***
	(0.00206)	(0.0040)	(0.0044)
$ulc_{i,t-1}+i_{i,t-1}-pi_{i,t-1}$	0.00106	-0.0055	-0.0118**
	(0.00213)	(0.0040)	(0.0046)
$\pi_{i,t+1}$	0.8667***	0.7259***	0.6119***
	(0.0995)	(0.1528)	(0.0866)
$\pi_{i,t-1}$	0.0980	0.0508	0.1903***
	(0.0658)	(0.0882)	(0.0593)
$\Delta ulc_{i,t}$	0.0299	0.0666***	0.0422*
	(0.0236)	(0.0223)	(0.0238)
$\Delta pi_{i,t}$	-0.0196	0.0643*	0.0707**
	(0.0214)	(0.0359)	(0.0276)
$\Delta i_{i,t}$	0.00048	0.0011	0.00039
	(0.00069)	(0.0011)	(0.00127)
$dummy\ pl_{i,t-1}$	-0.00014	-0.00035*	-0.00032*
	(0.00010)	(0.00019)	(0.00032)
Time dummies	Yes	Yes	Yes
F-stat 1st stage reg.:			
$\pi_{i,t+1}$	8.56	10.56	22.37
J-statistic	5.86	9.09	8.05
	[0.55]	[0.26]	[0.42]
Q (1) stat.	107.45	107.63	68.26
	[0.00]	[0.00]	[0.00]

Notes: see notes to Table 1. Instruments: constant, $\pi_{i,t-2}$, $x_{i,t-1}$, $q_{i,t-1}$, $i_{i,t-1}$, $\Delta i_{i,t-1}$, $p_{i,t-2}$, $p_{i,t-3}$, $difp_{i,t-1}$, $\Delta pi_{i,t-1}$ $ulc_{i,t-1}+i_{i,t-1}-p_{i,t-1}$, $ulc_{i,t-1}+i_{i,t-1}-pi_{i,t-1}$, $\Delta ulc_{i,t-1}$, $pl_{i,t-1}$, $dummy\ pl_{i,t-1}$ and time dummies. Variables $\pi_{i,t}$, $p_{i,t}$ and $difp_{i,t}$ were computed for columns (1), (2) and (3) using the index of services, index excluding energy prices and unprocessed food, and the index excluding the fully administered prices, respectively. Eq. (3) includes $\pi_{i,t-1}$ a additional instrument, and it was only estimated for the period 2001Q1-2008Q4 due to lack of data on inflation excluding administered prices.

[19] Results available upon request.

As the first years of the euro were a transition period, it is interesting to test whether our results hold if we exclude these years; in fact, estimating equation (8) starting in 2001Q1 does not change the main results.[20]

Next, we make separate regressions for the peripheral countries more pressured by the sovereign debt crisis (Greece, Portugal, Spain, Ireland and Italy) and for the others (the core countries) –Table 4, columns (2) and (3). The results of both regressions have less statistical significance, which may be explained by the smaller samples, but the signs of the coefficients are broadly similar. It is worth mentioning that for peripheral countries the coefficient of expected inflation decreases relative to the regression including all countries, and the coefficients of the error correction terms increase. Expected inflation is the only statistically significant determinant of inflation for the core countries. These countries have a history of lower and more stable inflation, which may explain why expected inflation is more important and the error correction terms are not statistically significant (see Hofmann and Remsperger 2005 for a similar result).

In addition, we apply equation (8) to three sub-indexes of the HICP: index excluding fully administered prices, index excluding energy prices and unprocessed food (used to compute core inflation), and the index for services. For the first index, the existence of administered prices that do not respond to market mechanisms is a source of inflation differentials in Europe (ECB 2003; Égert 2007). In turn, core inflation removes the most volatile components of HICP. Energy prices are also an important source of inflation differentials as a result of different levels of oil dependence across European countries (ECB 2003). Regarding the index for services, differences across countries in inflation for non-traded goods (notably services) are larger than for traded goods (Zdárek and Aldasoro 2009).

The main determinant of inflation differences in services is expected inflation, and this has a large coefficient – 0.85 (Table 5, column 1). Core inflation's differentials are explained by a larger number of variables, notably expected inflation, variation in unit labour costs, variation in import prices and lagged dummy for the price level (Table 5, column 2). This is one of the few regressions where price convergence is statistically significant, albeit with a small coefficient. We do not find complete support for the ICM in explaining core inflation convergence because the error correction terms are statistically insignificant; however, the New Keynesian framework is not enough to explain inflation differences due to the significance of the change in the nominal ULC.

---

[20] Results available upon request.

Divergence in inflation excluding fully administered prices is essentially explained by the same factors as divergence in overall inflation, with one major exception: lagged inflation is statistically significant (Table 5, column 3).

Estimating the equation exclusively based on the NKPC for national inflation rates (equation 2) –i.e. excluding the ICM elements– for services, core inflation and inflation without the administrative prices yields very similar results to those obtained in Table 2: statistical significance of expected inflation and the irrelevance of the other variables, notably of both real ULC and output gap.[21] There is however one difference: the lagged inflation is relevant for both inflation in services and excluding the administered prices.

In sum, robustness checks essentially confirm our previous analysis, supporting the importance of expectations as a driver of inflation differentials. However, the robustness analysis does not confirm the relevance of the ICM to understand divergence in inflation dynamics in some cases, namely for core countries and inflation in services.

The identified relevance of the nominal ULC for inflation differentials may create destabilising macroeconomic effects because inflation differentials can lead to differences in wage growths that will have a further effect on inflation differentials. This feedback effect may not have a significant impact because, as suggested by Hofmann and Remsperger (2005), the euro area has mechanisms to correct inflation differentials.

V. Conclusion

The main goal of this paper was to identify the determinants of inflation differences in the euro area. For a panel of twelve euro area countries, the estimation of the NKPC with quarterly data for the period 1998Q1-2008Q4 produces similar results to other studies with time-series and panel data. Inflation has both forward- and backward-looking components, but the former is more important. Exchange rates also play a role in price changes, with lagged exchange rate having a statistically significant impact. The cost channel is present and the output gap or real ULC has a positive effect on inflation though it is not statistically significant.

Regarding inflation differentials, we observe that the expected inflation rate and exchange rate movements are important determinants of differences in inflation

---

[21] The expected exchange rate change is significant at 10% for inflation excluding the administered prices. Results available upon request.

rates. On the other hand, past dynamic of inflation does not play a very relevant role. The usual measures of business cycle (output gap and real ULC) and interest rate are not statistically significant in causing differences in inflation dynamics. Price convergence is only statistically significant in some estimations, and its economic significance is small. Finally, our results indicate that there is not a direct correspondence between determinants of national inflation rates and determinants of inflation differentials.

It should be noted that expected inflation plays a fundamental role in the results. When this variable was introduced, the lagged price level and the output gap lost their statistical significance, suggesting they were only significant because they forecast inflation.

The growth in nominal ULC also plays a significant role in explaining inflation differentials. This means that the business cycle affects inflation differentials when it causes differences in wages evolution across countries, as predicted by the ICM. Inflation rate differences are also affected by the lagged disequilibrium in the long-run relationship proposed by the ICM, which involves domestic prices on one hand, and the ULC, nominal interest rate and imports price on the other. Furthermore, the introduction of the error correction terms proposed by the ICM reduced the coefficient of expected inflation but did not eliminate its statistical significance. Also, the ICM model is not encompassed by the NKPC when explaining inflation differences. In sum, our results show that the NKPC is an insufficient framework to explain inflation differentials: its greatest strength is the key role given to expectations. Inflation divergence is explained better using central aspects of the ICM together with expected inflation.

This paper contributes to a better understanding of the functioning of a monetary area, notably to the origin of the euro area sovereign debt crisis. The inflation differentials in the euro area is one of the explanatory factors for the crisis. The inception of the euro meant a significant reduction in interest rates in peripheral economies and thus a marked increase in debt accumulation (Higgins and Klitgaard 2011). As a result, these economies had positive inflation differentials that weakened their external competitiveness and economic growth. Divergences in inflation rates are linked to differences in the growth in unit labour costs, which according to Barbosa and Alves (2011) explain dissimilar real exchange rate growth rates. Competitive wage restraint in Germany was also responsible for differences in wage evolution and created balance of payment imbalances (Bibow 2013). In addition, inflation differentials result from the lack of policy coordination and of mechanisms to address asymmetric shocks.

It is clear that inflation differentials present a challenge for the ECB because a common monetary policy becomes sub-optimal and national economies with positive inflation differentials lose external competitiveness. In this context, our results show that managing expectations and controlling labour costs are fundamental to ensure inflation convergence. The ECB should also take the impact of the exchange rate on inflation differentials into account. Given the relevance of labour costs, further work is necessary to assess the empirical relevance of a diverging inflationary cycle arising from the interaction between labour costs and inflation.

Turning now to the immediate response to the euro area debt crisis, peripheral countries and indeed almost all countries followed contractionary fiscal policies to improve fiscal fundamentals. Similarly, wage growth was contained to reduce public expenses and regain external competitiveness.

According to our results, depressed inflation expectations and reduction in unit labour costs led to a decline in inflation and inflation differentials, which reached an all-time low in the euro area. The ECB cut interest rates and resorted to unconventional monetary policy measures due to a scenario of deflation, the approaching of the liquidity trap, a dysfunctional monetary policy transmission mechanism, and the need to finance countries affected by the crisis. This policy included buying covered bonds from banks, conducting longer-term refinancing operations (LTROs) with three years' maturity, the increase of collateral accepted to lend money to banks, buying securities issued by governments (quantitative easing), and forward guidance that will maintain interest rates at low levels for a long period. Forward guidance is an attempt to influence inflation expectations due to its importance for inflation as we have seen above.

The depreciation of the euro caused by the monetary policy measures helped boost aggregate demand and inflation prospects. Although the ECB policy has fostered growth and inflation has regained momentum, more needs to be done in light of the risks of deflation, especially in high-debt countries.

Fiscal policy is known to be an effective tool near the zero lower bound interest rate, when monetary policy becomes less effective. However, peripheral countries cannot use expansionary fiscal policy due to strong market pressures. It should be the central economies that implement this policy as they can afford it and it would create positive spillovers in the more depressed economies. Higher inflation and wage growth in some countries (notably in the countries with a trade surplus such as Germany) would therefore be beneficial (Moro 2014). In other words, inflation differentials in the short term are useful as they help reduce the balance of payment imbalances within the euro area. In the current architecture of this area, Germany cannot permanently have a trade surplus (Bibow 2013).

In order to reduce debt in some countries, Antzoulatos (2012) argues for the need of a higher long term inflation target for the ECB. Likewise, Caraballo and Dabús (2013) claim that the optimal annual inflation rate for Spain would be 4%. Other measures are required to tackle the problems in the euro area and some are already being implemented: banking union (Shambaugh 2012), structural reforms targeting non-price components of external competitiveness (Estrada, Galí and López-Salido 2013), and monitoring macroeconomic imbalances (including credit growth and external imbalances) to avoid future crisis (Lane 2012).

Appendix

Data used are described below.

Quarterly inflation ($\log(p_{i,t})$-$\log(p_{i,t-1})$) was measured using the seasonally adjusted Harmonized Index of Consumer Prices (HICP) from Eurostat ($p_{i,t}$).

Difference of CPI indexes ($difp_{i,t}$: $\log(p_{i,t})$-$\log(p_{euro12,t})$). HICP for euro area 12 ($p_{euro12,t}$) was obtained from Eurostat. Both indexes are seasonally adjusted and have value 100 in 2005.

The price level ($pl_{i,t}$) is a dummy variable computed based on the price level index of household final consumption expenditure from Eurostat (Honohan and Lane 2003, also uses this price index). The dummy takes value one if that price level is above one, in which case the country's price level is above the European average. The price index is obtained by Eurostat as the purchasing power parity of consumption over current nominal exchange rate, with the initial 12 members of euro area as a reference. The quarterly data was obtained interpolating the original annual data with local quadratic polynomial.

Real seasonally adjusted GDP was obtained from OECD Quarterly National Accounts and from International Financial Statistics of IMF (IFS/IMF) for Ireland, Luxembourg and Portugal. The output gap ($x_{i,t}$) for each country was calculated as the difference between the log of output and the log of the output trend, with series starting in 1979Q1 or 1980Q1. To calculate the output trend, we used the HP filter with lambda fixed at 1600.

Real ULC ($s_{i,t}$), or wage share, was obtained dividing the nominal ULC (2005=100) by the GDP deflator (2005=100).

Nominal unit labour costs ($ulc_{i,t}$) refer to the trend-cycle series for the entire economy obtained from Main Economic Indicators/OECD. As ULC for the entire economy for Portugal were not available, we used ULC for the business sector.

Seasonally adjusted GDP deflator ($pd_{i,t}$) was obtained from OECD National Accounts, except in the cases of Portugal, Ireland and Luxembourg, for which IFS/IMF data was used.

Nominal effective exchange rate ($e_{i,t}$) was obtained from IFS/IMF, base year 2005. This measure uses weights from the trade of manufactured goods. An increase in $e_{i,t}$ corresponds to an appreciation of the euro.

Real effective exchange rate ($q_{i,t}$) is based on relative Consumer Prices, 2005=100, and is from IFS/IMF.

Imports price ($pi_{i,t}$) seasonally adjusted are measured by import price deflator from Quarterly National Accounts OECD (Imports of goods and services, 2005=100).

Retail interest rate ($i_{i,t}$): loans to corporations up to one year from Eurostat. Before 2003Q1, data are not harmonised. To accommodate this, we used a dummy for the period 1998Q1-2002Q4. Note that as there was no data available for Luxembourg and Finland before 2003Q1, we used the interest rate of loans to firms above one year.

References

Angeloni, Ignazio, and Ehrmann, Michael (2007). Euro area inflation differentials. *The B.E. Journal of Macroeconomics* 7: 1935-1690.

Antzoulatos, Angelos A. (2012). Policy responses to the European Debt Crisis: Treating the "symptoms" or the "Disease"? *Panoeconomicus* 529-552.

Barbosa, João R., and Alves, Rui H. (2011). The Euro area ten years after its creation: (Divergent) competitiveness and the optimum currency area theory. *Panoeconomicus* 605-629.

Bårdsen, Gunnar, Jansen, Eilev, and Nymoen, Ragnar (2004). Econometric evaluation of the New Keynesian Phillips curve. *Oxford Bulletin of Economics and Statistics* 66: 671-686.

Barth, Marvin, and Ramey, Valerie (2001). The cost channel of monetary transmission. *NBER Macroeconomics Annual* 16: 199-240.

Batini, Nicoletta, Jackson, Brian, and Nickell, Stephen (2005). An open-economy New Keynesian Phillips Curve for the UK. *Journal of Monetary Economics* 52: 1061-1071.

Beck, Nathaniel, and Katz, Jonathan (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review* 89: 634-647.

Bibow, Jörg (2013). The Euroland crisis and Germany's euro trilemma. *International Review of Applied Economics* 27: 360-385.

Binder, Michael, and Pesaran, M. Hashem (1995). Multivariate rational expectations models: a review and some new results. In M. Hashem Pesaran and Michael R. Wickens, editors, *Handbook of Applied Econometrics - Volume Macroeconomics*. Oxford, Blackwell Publishing.

Bjørnstad, Roger, and Nymoen, Ragnar (2008). The New Keynesian Phillips Curve tested on OECD panel data. *Economics E-Journal* 2: 1-18.

Busetti, Fabio, Forni, Lorenzo, Harvey, Andrew, and Venditti, Fabrizio (2007). Inflation convergence and divergence within the European Monetary Union. *International Journal of Central Banking* 3: 95-121.

Caraballo, Ma A., and Dabús, Carlos (2013). Price dispersion and optimal inflation: The Spanish case. *Journal of Applied Economics* 16: 49-70.

Conti, Maurizio (2014). The introduction of the euro and economic growth: Some panel data evidence. *Journal of Applied Economics* 16: 199-212.

Dullien, Sebastian, and Fritshe, Ulrich (2008). Does the dispersion of unit labor cost dynamics in the EMU imply long-run divergence? *International Economics and Economic Policy* 5: 269-295.

ECB (2003). Inflation differentials in the Euro Area: Potential causes and policy implications. Frankfurt, ECB.

Égert, Balázs (2007). Real convergence, price level convergence and inflation differentials in Europe. Working Paper No. 895. Michigan, William Davidson Institute.

Estrada, Ángel, Galí, Jordi, and López-Salido, David (2013). Patterns of convergence and divergence in the Euro area. *IMF Economic Review* 61: 601-630.

Eviews (2007). *User's guide*. California, Quantitative Micro Software.

Galí, Jordi, and Gertler, Mark (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics* 44: 195-222.

Galí, Jordi, Gertler, Mark, and López-Salido, Jordi D. (2001). European inflation dynamics. *European Economic Review* 45: 1237-1270.

Galí, Jordi, and López-Salido, J. David (2001). A New Phillips Curve for Spain. Paper No 3. Basel, BIS.

Garrat, Anthony, Lee, Kevin, and Shields, Kalvinder (2009). Measuring the natural output gap using actual and expected output gap. Working Papers 09/21. Leicester, University of Leicester Department of Economics.

Higgins, Matthew, and Klitgaard, Thomas (2011). Saving imbalances and the euro area sovereign debt crisis. *Current Issues in Economics and Finance* 17: 1-11.

Hofmann, Boris, and Remsperger, Hermann (2005). Inflation differentials among the Euro area countries: Potential causes and consequences. *Journal of Asian Economics* 16: 403-419.

Honohan, Patrick, and Lane, Philip R. (2003). Divergent inflation rates in EMU. *Economic Policy* 18: 357-394.

Honohan, Patrick, and Lane, Philip R. (2004). Exchange rates and inflation under EMU: An update. Discussion Paper No. 31. Dublin, IIIS.

Kara, Amit, and Nelson, Edward (2003). The exchange rate and inflation in the UK. *Scottish Journal of Political Economy* 50: 585-608.

Lagoa, Sérgio (2014). Inflation dynamics in open economies: Empirical evidence for G7 countries on the role of import prices and the cost channel. *Research in Economics* 68: 354-371.

Lane, Philip R. (2006). The real effects of European Monetary Union. *Journal of Economic Perspectives* 20: 47-66.

Lane, Philip R. (2012). The European sovereign debt crisis. *Journal of Economic Perspectives* 26: 49-68.

Lown, Cara S., and Rich, Robert W. (1997). Is there an inflation puzzle? *Federal Reserve Bank of New York Economic Policy Review* December: 51-69.

Mavroeidis, Sophocles (2004). Weak identification of forward-looking models in monetary economics. *Oxford Bulletin of Economics and Statistics* 66: 609- 632.

Moro, Beniamino (2014). Lessons from the European economic and financial great crisis: A survey. *European Journal of Political Economy* 34: S9-S24.

Neiss, Katharine S., and Nelson, Edward (2005). Inflation dynamics, marginal cost, and the output gap: evidence from three countries. *Journal of Money, Credit, and Banking* 37: 1019-1045.

Paloviita, Maritta (2008). Comparing alternative Phillips curve specifications: European results with survey-based expectations. *Applied Economics* 40: 2259-2270.

Pesaran, M. Hashem (1987). The limits of rational expectations. Oxford, Basil Blackwell.

Ravenna, Federico, and Walsh, Carl E. (2006). Optimal monetary policy with the cost channel. *Journal of Monetary Economics* 53: 199-216.

Rogers, John H. (2002). Monetary union, price level convergence, and inflation: How close is Europe to the United States? International Finance Discussion Papers No. 740. Washington, Board of Governors of the Federal Reserve System.

Rogers, John H. (2007). Monetary union, price level convergence, and inflation: How close is Europe to the USA? *Journal of Monetary Economics* 54: 785-796.

Rotemberg, Julio, and Woodford, Michael (1997). An optimization-based econometric framework for the evolution of monetary policy. *NBER Macroeconomics Annual* 12: 297-346.

Shambaugh, Jay C. (2012). The euro's three crises. *Brookings Papers on Economic Activity* Spring: 157-211.

Stock, James H., Wright, Jonathan H., and Yogo, Motohiro (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20: 518-529.

Vines, David, Kirsanova, Tatiana, and Wren-Lewis, Simon (2006). Fiscal Policy and Macroeconomic Stability within a Currency Union. Discussion Paper No. 5584. London,. CEPR.

Zdárek, Václav, and Aldasoro, Juan I. (2009). Inflation differentials in the euro area and their determinants – an empirical view. Advanced Studies Working Paper No. 450. Kiel, Kiel Institute for the World Economy.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 105-120

THE GROWTH-INFLATION NEXUS FOR THE U.S. FROM 1801 TO 2013: A SEMIPARAMETRIC APPROACH

Mehmet Balcilar
Eastern Mediterranean University, University of Pretoria and Montpellier Business School

Rangan Gupta*
University of Pretoria and IPAG Business School

Charl Jooste
University of Pretoria

Submitted July 2015; accepted August 2016

We study the existence of a threshold level of inflation for the U.S. economy over 1801-2013, beyond which it has a negative effect on economic growth. A combination of nonparametric (NP) and instrumental variable semiparametric (SNP-IV) methods obtain inflation thresholds for the United States. The results suggest that the relationship between growth and inflation is hump shaped –that higher levels of inflation reduce growth more compared to low inflation or deflation. The strongest result to emerge from the study consistently shows that inflation above two per cent negatively affects growth. Two additional parametric methods confirm this finding. Another important result is that high or very low levels of inflation are undesirable and are associated with lower growth - hinting that a growth maximizing value of inflation exists.

JEL classification codes: C14, E31
Key words: inflation, growth, nonparametric, semiparametric

---

* Rangan Gupta (corresponding author): Professor, Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; IPAG Business School, Paris, France; email: rangan.gupta@up.ac.za. Mehmet Balcilar: Professor, Department of Economics, Eastern Mediterranean University, Famagusta, Northern Cyprus, via Mersin 10, Turkey; Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; Montpellier Business School, Montpellier, France; email: mehmet@mbalcilar.net. Charl Jooste: Graduate Student, Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; email: Jooste.Charl@yahoo.co.za. We would like to thank an anonymous referee for many helpful comments. However, any remaining errors are solely ours.

I. Introduction

What rate of inflation is unfavorable for economic growth? This paper shows that inflation above two per cent reduces GDP growth. We analyse the inflation-growth trade-off in a semiparametric model using 213 years of United States GDP and inflation data. This paper contributes to the existing literature with a specific focus on the United States and avoids assumptions regarding explicit functional relationships.

The inflation-growth trade-off is interesting for various reasons. Its policy relevance relates to making decisions regarding interest rates and understanding the endogenous response of output growth to inflation. With interest rates at the zero lower bound (ZLB), it is natural to ask what rate of inflation is optimal. This is equally important when the Federal Reserve (Fed) considers the rate of inflation that reduces the probability of hitting the ZLB. However, rising inflation expectations have economic costs –a decrease in GDP growth. The literature on the inflation–growth nexus shows mixed results on the level of inflation that subtracts from economic growth. If certain rates of inflation increase growth, then one might make the argument for raising the inflation target. This is a Mundell-Tobin effect where inflation expectations shift investments away from money balances into other types of capital that have growth inducing effects.

The so-called point at which inflation has a negative effect on economic growth has policy consequences. Consequently, many studies attempt to pinpoint the exact inflation threshold. The empirical methods used are diverse and cover nonlinear functional forms (to test possible hump-shaped relationships) to linear regressions estimated over certain break points to dynamic models that are estimated over certain periods (i.e., rolling window regressions) and panel models that take account of cross-sectional heterogeneity.

The large body of research that analyses inflation thresholds imposes a functional form (see Barro 1995; Fischer 1993 and Rousseau and Wachtel 2001 as an example) –a practice that might be wrong. Our paper comes perhaps closest to Vaono and Shiavo (2007), who use nonparametric methods in a panel framework to study possible inflation thresholds. They use a panel data set that includes 85 countries from 1960 to 1999 over five year period intervals. We depart from their work by focusing on obtaining results for the U.S. using a novel data set with a much longer and recent time series. We use instrumental variable (IV) methods to control for possible endogeneity between growth and inflation in a semiparametric model. This gives us an unbiased model free from assuming a functional form that imposes a nonlinear relationship or one that assumes inherent structural breaks, i.e., we let the data speak for itself. In addition to the nonparametric approach, we use various nonlinear parametric methods for robustness. Specifically, we test for inflation thresholds and its impact on economic growth using a time varying parameter and a smooth transition regression.

II. Literature review

It is generally accepted that inflation reduces overall welfare. The welfare costs of inflation have been extensively studied for the U.S. The general consensus is that inflation reduces overall welfare irrespective of discretionary or committed monetary policy. As an example, Miller et al. (2014) estimate the time-varying welfare costs of inflation in a money demand model using a time-varying cointegrated model. Their estimate of the welfare cost of 10 per cent inflation in terms of GDP range between 0.025 and 0.75 per cent, with an average of 0.27 per cent. This is well in line with other studies (see Fischer 1981; Serletis and Yavari 2004 and Ireland 2009).

From a theoretical perspective, Billi (2011) studies the optimal long-run rate of inflation in a New-Keynesian model that counteracts the negative economic effects of hitting the ZLB. Billi (2011) obtains estimates of the long-run inflation rate under three regimes (discretion, commitment and a Taylor rule) and implicitly accounts for misspecification in all cases. Under commitment policy, the optimal inflation rate is 0.2% (under no misspecification) and 0.9% (extreme misspecification) whereby the government hits the zero-lower-bound only occasionally. The government can stimulate the economy by creating inflation expectations and commitment implies a lowering of real interest rates. The consumption (welfare) loss is lower in the model with commitment compared to the model with discretion. The optimal long run inflation rate is very high if a government re-optimises (discretion) and extremely high accounting for misspecification. This causes high inflation expectations that could lead to high real interest rates that harm the economy. The optimal inflation rate can be as high as 13.4% and 16.7% (under extreme misspecification). Finally, under a standard Taylor rule the optimal long run inflation rate is between 8% and 9.8% (this rule does not account for inertial interest rates). Consumption loss is lowest in the model with an inertial Taylor rule (as opposed to a standard Taylor rule, commitment, and re-optimizing government) - this allows for slightly higher optimal long run inflation target. Billi (2011) shows that discretionary policy makers are so averse to deflation, that they are willing to tolerate massive inflation bias. This model, however, assumes that the monetary policy rate is the only policy measure available.

Reifschneider and Williams (2000) find a 2% inflation goal to be a sufficient level to counter the adverse economic effects of the ZLB. The frequency of mild recessions, due to the ZLB on the interest rate, are reduced for a two per cent inflation target compared to a zero per cent inflation target. However, the frequencies of severe recessions are higher in a model of two per cent inflation target relative to a zero inflation target.

Krugman (1998) interestingly states that committed inflation policy is a reason that economies remain at the ZLB: agents perceive expansionary monetary policy as temporary and believe the central bank's commitment to low inflation. A higher inflation commitment or alternative government policies might help an economy escape from the ZLB. Unfortunately, the Krugman study ignores the cost of higher inflation. However, Billi (2011) shows that the welfare costs of inflation are low when escaping the ZLB.

Khan and Senhadji (2001) estimate the inflation threshold of about 11% for developing countries and 1% for industrialized countries. They use a data set that covers 140 countries from 1960-1998. They use an indicator function to test the statistical significance of various thresholds. The inflation value that minimizes the sum of squared residuals of their specification is the inflation threshold. Burdekin et al. (2004) show that inflation (as low as single digit inflation) has a negative impact on economic growth. They estimate a spline equation that allows for structural breaks in inflation. This allows them to study the impact of inflation on growth at various thresholds. The growth cost of inflation becomes more pronounced at inflation levels in excess of 50%. Furthermore, they show that the inflation impact on growth is biased downwards when not accounting for nonlinearities.

Barro (1995) shows that a ten per cent increase in inflation reduces growth in real GDP per capita by about 0.2-0.3 percentage points per year –a significant reduction of GDP over a long period, which highlights the importance of price stability from a GDP perspective. This is in line with a plethora of panel studies that obtain similar results (see Roubini and Sala-i-Martin 1992; Fischer 1993 and Chari et al. 1995)

Bruno and Easterly (1998) show, however, that the negative association between inflation and economic growth is difficult to establish for low to moderate levels of inflation in a panel of 31 countries –they emphasise that pooled cross-country datasets are not informative about what happens to growth at low levels of inflation; something that this paper tries to remedy.

Vaona and Schiavo (2007) shows that the relationship between inflation and growth is nonlinear. They use nonparametric and semi-parametric IV methods to estimate thresholds. They obtain a 12 per cent threshold for developed countries while no clear indication of a threshold is found for developing countries.

Vaona (2012) finds a different result than Vaona and Schiavo (2007) using a similar methodology on a different dataset of 85 countries: inflation has a negative linear impact on growth; inflation is simply growth reducing.

III. Methodology

We use annual real GDP growth and inflation data from 1801 to 2013 calculated as percentage change of the real GDP (at constant 2009 prices) and the consumer price index (with a base period of 1982-1984). Real GDP data comes from the Global Financial Database, the CPI data comes from the website of Professor Robert Sahr.[1] The start and endpoints of our sample are purely driven by data availability; though the CPI data goes as far back as 1774, real GDP data is only available from 1800. Since we use growth rates of the two variables, we lose the observations corresponding to 1800. We control for different lags of inflation and GDP. This gives us 213 observations over different regimes and possible structural breaks. Inflation over the period averaged 1.37 per cent while GDP growth averaged 3.65 per cent (see Table 1).

Table 1. Descriptive statistics

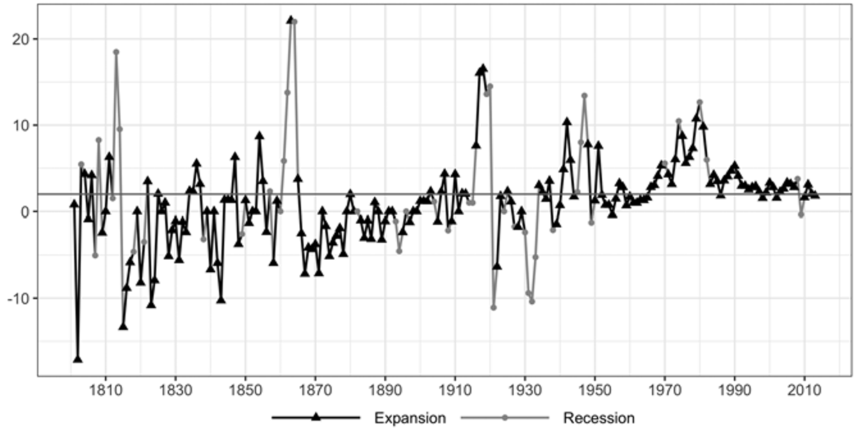
	Inflation	Growth rate
N	213	213
Mean	1.3713	3.6559
S.D.	5.4928	5.3335
Min	-17.1358	-13.9291
Max	22.1161	16.9925
Skewness	0.5057	-0.3137
Kurtosis	2.4409	0.6985
JB	64.3380***	8.3180**
Q(1)	59.5541***	7.6934***
Q(4)	80.8971***	9.4057*
ARCH(1)	45.9068***	13.5161***
ARCH(4)	62.6250***	20.7283***

Notes: In addition to the mean, the standard deviation (S.D.), minimum (min), maximum (max), skewness, and kurtosis statistics, the table reports the Jarque-Bera normality test (JB), the Ljung-Box first [*Q*(1)] and the fourth [*Q*(5] autocorrelation tests, and the first [ARCH(1)] and the fourth [ARCH(5)] order Lagrange multiplier (LM) tests for the autoregressive conditional heteroskedasticity (ARCH). The asterisks ***, ** and * represent significance at the 1%, 5%, and 10% levels, respectively.

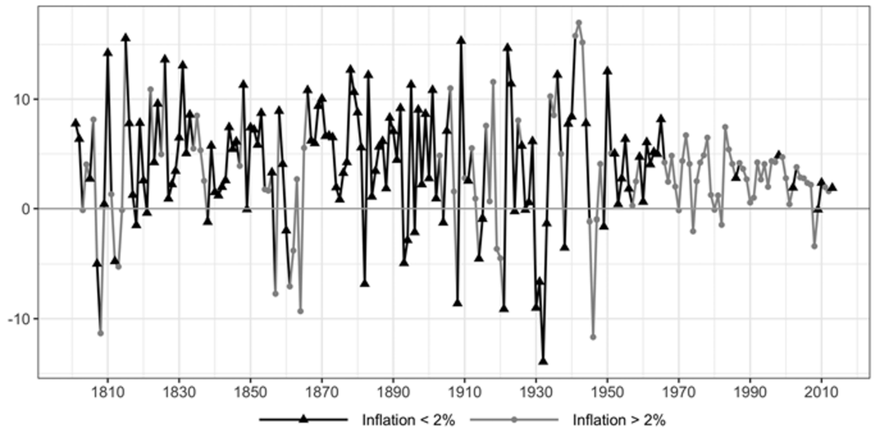
[1] http://oregonstate.edu/cla/polisci/sahr/sahr.

Figure 1. Inflation and GDP growth: 1801-2013

(a) Inflation



(b) Growth



Notes: Part (a) presents the annual inflation rate with recession and expansion periods indicated by different symbols and colors. A light colored horizontal line is draw at the 2% threshold inflation rate. Part (b) presents the GDP growth rate with periods corresponding to less than and greater than the 2% threshold inflation rate market with different colors and symbols.

Figure 1(a) presents the annual inflation rate with recession and expansion dates indicated with gray and symbols. Analogously, Figure 1(b) presents the annual GDP growth rates with the periods corresponding to less than 2% and greater than 2%

inflation rates marked with different colors and symbols. The 2% inflation rate corresponds approximately to the threshold inflation rate estimated nonparametrically in Section IV. Inflation rates above the 2% threshold reduce the GDP growth rate. Figure 1 shows that the inflation threshold and business cycle are not likely to be related, and inflation rates above the threshold corresponds to both expansion and recession periods.

We use a semiparametric model. In the case of the inflation-growth relationship, conditional expectation restrictions might not be satisfied –especially if there are strong feedback loops between inflation and GDP growth. We use an IV approach to control for endogeneity. The semiparametric model allows the data to uncover a more realistic functional form. Our relatively large sample size reduces the possibility of misspecification bias.

The selected IV models are based on the Wu-Hausman *F*-test and the Sargan *J*-test. Wu-Hausman tests the null that the regressors are not correlated with the disturbance term. The *J* statistic tests the null hypothesis that all instruments are exogenous.

In a first stage, we determine which instruments are relevant. At this point, an *F*-test is used to determine whether or not the instruments should enter the first stage regression. Weak instruments imply a small first-stage *F*-test. The auxiliary instrumental variables regressions take the following form:

$$\pi_t = \mu + \theta' z_t + \varepsilon_t \, , \tag{1}$$

where $\pi_t =$ inflation rate, $z_t = \left[\pi_{t-1}, \pi_{t-2}, \ldots, \pi_{t-p}, g_{t-1}, g_{t-2}, \ldots, g_{t-q}\right]'$ (instrumental variables), g_t = real GDP growth rate and $\varepsilon_t \sim iid\,(0,\sigma^2)$ is the error term.

Then, we estimate the OLS and IV regression models of the following form:

$$g_t = \alpha + \beta\pi_t + \varepsilon_t \, , \tag{2}$$

and the OLS-lagged model:

$$g_t = \alpha + \beta\pi_{t-1} + \varepsilon_t . \tag{3}$$

Finally, the semiparametric specification can be expressed as follows:

$$g_t = \phi' x_t + f\left(\pi_t\right) + \varepsilon_t \, , \tag{4}$$

where $f(\pi_t)$ is a nonlinear function and x_t is a set of exogenous variables. We account for the possibility that $E[\varepsilon_t | \pi_t] \neq 0$ by estimating (4) using those models, whose instrument validity is not rejected by the Sargan *J*-test. Following Vaona and Schiavo (2007), we estimate the model in equation (4) using the semiparametric IV estimation approach of Park (2003). The degree of complexity, or optimal data driven method of bandwidth selection, is determined using the least-square cross validation method of Li et al. (2013). A Gaussian kernel is used for all nonparametric and semiparametric models.

IV. Results

Our instruments and exogenous variables only include lagged inflation and lagged GDP. Unfortunately, we do not have enough observations going back until 1802 to control for other variables such as investment or terms of trade. Table 2 and 3, however, report that our instruments are reliable. We estimate nine specifications with different instruments. The specifications in Table 2 indicate that all instruments are adequate (*F*-test is rejected).

The Sargan *J*-statistic from Table 3 show that only models 1, 2, 3, 4, 8 and 9 should be used in the semiparametric model. The coefficients of inflation and lagged inflation on GDP growth are all negative: Inflation reduces economic growth in a linear model. The non-instrumental regressions seem to underestimate the effects of inflation on GDP growth, while the estimates of the IV regressions are all very similar, falling in the interval [-0.274, -0.323].

Table 2. Estimates of the IV auxiliary regressions

	Instruments	R^2	Adj. R^2	$\hat{\sigma}$	F	$Q(20)$
Model 1	π_{t-1}	0.3197	0.3164	4.4506	97.2871***	25.1592
Model 2	π_{t-1}, π_{t-2}	0.3214	0.3149	4.4558	48.7916***	25.5407
Model 3	$\pi_{t-1}, \ldots, \pi_{t-3}$	0.3241	0.3142	4.4577	32.7724***	26.9195
Model 4	$\pi_{t-1}, \ldots, \pi_{t-4}$	0.3350	0.3220	4.4325	25.6969***	24.1412
Model 5	π_{t-1}, g_{t-1}	0.3219	0.3153	4.4542	48.9004***	24.3469
Model 6	$\pi_{t-1}, \pi_{t-2}, g_{t-1}$	0.3232	0.3133	4.4608	32.6359***	24.8406
Model 7	$\pi_{t-1}, \pi_{t-2}, g_{t-1}, g_{t-2}$	0.3253	0.3121	4.4647	24.5929***	25.6369
Model 8	$\pi_{t-1}, \ldots, \pi_{t-3}, g_{t-1}, \ldots, g_{t-3}$	0.3281	0.3082	4.4774	16.4424***	26.8569
Model 9	$\pi_{t-1}, \ldots, \pi_{t-4}, g_{t-1}, \ldots, g_{t-4}$	0.3395	0.3131	4.4614	12.8523***	23.3202

Notes: R^2 is the coefficient of determination Adj. R^2 is the adjusted coefficient of determination $\hat{\sigma}$ is the standard error of the regression F is the regression *F* statistic *Q*(20) is the Ljung-Box portmanteau tests of autocorrelation for order up to 20. *, **, and *** denote significance at 10%, 5%, and 1%, respectively.

Figure 2 plots the relationship between inflation and economic growth. This relationship is nonlinear for all the specifications. The "hump-shaped" relationship almost delivers what looks like a growth maximizing effect. However, it should be noted that the plots generated do not come from any functional form. A functional form that tests for growth maximizing effects of inflation would include a squared term for inflation (with the sign being negative). The results, however, suggest that deflation and high levels of inflation consistently reduces GDP growth across various specifications. The inflation threshold is close to zero and as high as 2 percent.

Table 3. OLS and IV estimates of the linear growth and inflation relationship

	Instruments	Intercept [α]	Inflation [β]	Wu-Hausman *F*-test	Sargan *J*-Test
OLS		3.8468***	-0.1451**		
		(0.3810)	(0.0686)		
OLS lagged		3.8983***	-0.1796***		
		(0.3790)	(0.0681)		
IV Model 1	π_{t-1}	4.0940***	-0.3180***	3.0120*	
		(0.4130)	(0.1230)		
IV Model 2	π_{t-1},π_{t-2}	4.0930***	-0.3180***	3.0290*	0.001
		(0.4130)	(0.1230)		
IV Model 3	$\pi_{t-1},\ldots,\pi_{t-3}$	4.1020***	-0.3230***	3.2840*	0.281
		(0.4130)	(0.1220)		
IV Model 4	$\pi_{t-1},\ldots,\pi_{t-4}$	4.0670***	-0.2990**	2.573	1.495
		(0.4100)	(0.1200)		
IV Model 5	π_{t-1},g_{t-1}	4.0560***	-0.2920**	2.179	6.7700***
		(0.4110)	(0.1220)		
IV Model 6	$\pi_{t-1},\pi_{t-2},g_{t-1}$	4.0590***	-0.2940**	2.264	6.8570**
		(0.4110)	(0.1220)		
IV Model 7	π_{t-1},π_{t-2}	4.0520***	-0.2880**	2.119	7.1830*
	g_{t-1},g_{t-2}	(0.4110)	(0.1210)		
IV Model 8	$\pi_{t-1},\ldots,\pi_{t-3}$,	4.0690***	-0.3010**	2.529	8.953
	$g_{t-1},\ldots,g_{t-3}$	(0.4110)	(0.1210)		
IV Model 9	$\pi_{t-1},\ldots,\pi_{t-4}$,	4.0310***	-0.2740**	1.832	10.481
	$g_{t-1},\ldots,g_{t-4}$	(0.4080)	(0.1190)		

Notes: OLS model is the estimate of $g_t=\alpha+\beta\pi_t+\varepsilon_t$, while OLS-lagged estimates $g_t=\alpha+\beta\pi_{t-1}+\varepsilon_t$ using non-instrumental OLS estimation. IV models are estimated by two stage least squres using the corresponding instruments given in the second column. Wu-Hausman *F* statistic tests for the endogeneity of inflation in Equation (2) given the instruments. Sargan *J* statistic tests for the overidentfying restrictions. *, **, and *** denote significance at 10%, 5%, and 1%, respectively.

Anything higher than two per cent reduces GDP growth. The nonlinear nature shows that high inflation decreases growth proportionally more compared to low inflation. The slope of Figure 2 is fairly flat for deflation. It is important to note that the confidence bands suggest that the results become insignificant with very high or low inflation (i.e. the growth effect of inflation becomes indiscernible. There exists, however, a corridor or significance that varies between -10 percent and 10 percent of inflation. There exists also a curl at the 20 percent mark. This is simply an artifact of the data - very few data points exist where inflation exceeds 20 percent (only 0.9% of the data is above 20 percent) and this is associated with GDP growth between 2.69 and -9.31 percent, respectively.

While most of the plots suggest that the relationship between inflation and economic growth is "hump-shaped", subplot (b) shows that the relationship between past inflation and current economic growth is monotonic. This is just one incidence where using a misspecified model can lead to strikingly different conclusions. The tests of Table 2 and Table 3 suggest that a more nuanced model is the correct model (i.e. one with a different set of controls).

A parametric model approach is also used to study the evolution of the inflation threshold on GDP growth. Two models are used: time varying parameter (TVP) and smooth transition regression (STR) models.[2] These models can help detect any nonlinearity between inflation and growth and specifically the inflation effect on growth past a certain threshold.

The TVP model is written as:

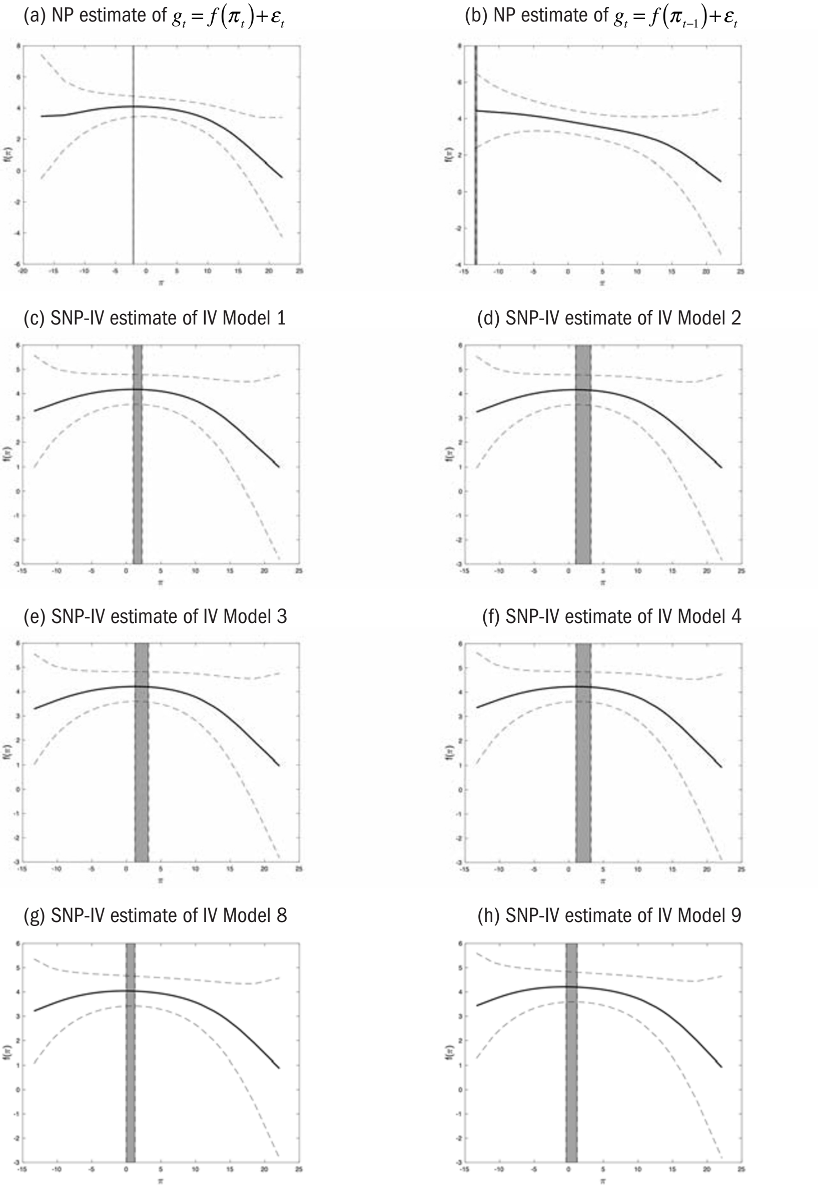
$$g_t = \alpha_t + \beta_t \pi_t + \varepsilon_t . \tag{5}$$

The Kalman filter produces the parameters in equation (5). The evolution of β_t illustrates the periods where inflation has a negative effect on growth. Figure 3(a) presents the estimates of β_t between 1801 and 2013. Inflation had a negative effect on growth the majority of time, barring two periods. The estimates in Figure 3(a) show that inflation has a negative impact on GDP growth during periods 1801-1832, 1850-1892, and 1950-2013. Inflation did not decrease growth in 1833-1849 (this was part of the industrial revolution age for the U.S. and of inflows of immigrants) and a longer period between 1893 and 1949.

The functional approach of the TVP regression assumes that the relationship between inflation and growth is linear –as such, it is difficult to assess whether an

[2] We thank an anonymous reviewer for suggesting the TVP model.

Figure 2. NP and SNP-IV estimates



Notes: Start and end of the maximum of nonparametric curve estimate is marked with vertical dashed lines and sheded with gray color. Upper-end of the maximum of the curve designates the inflation threshold.

inflation threshold exists. A STR model explicitly controls for thresholds and may yield a linear or non-linear result, where nonlinearity is a smooth logistic function over some threshold. We estimate the following STR model:[3]

$$g_t = [\alpha_1 + \beta_1 \pi_t] + [\alpha_2 + \beta_2 \pi_t] G(\pi_t; \gamma, c) + \varepsilon_t, \quad (6)$$

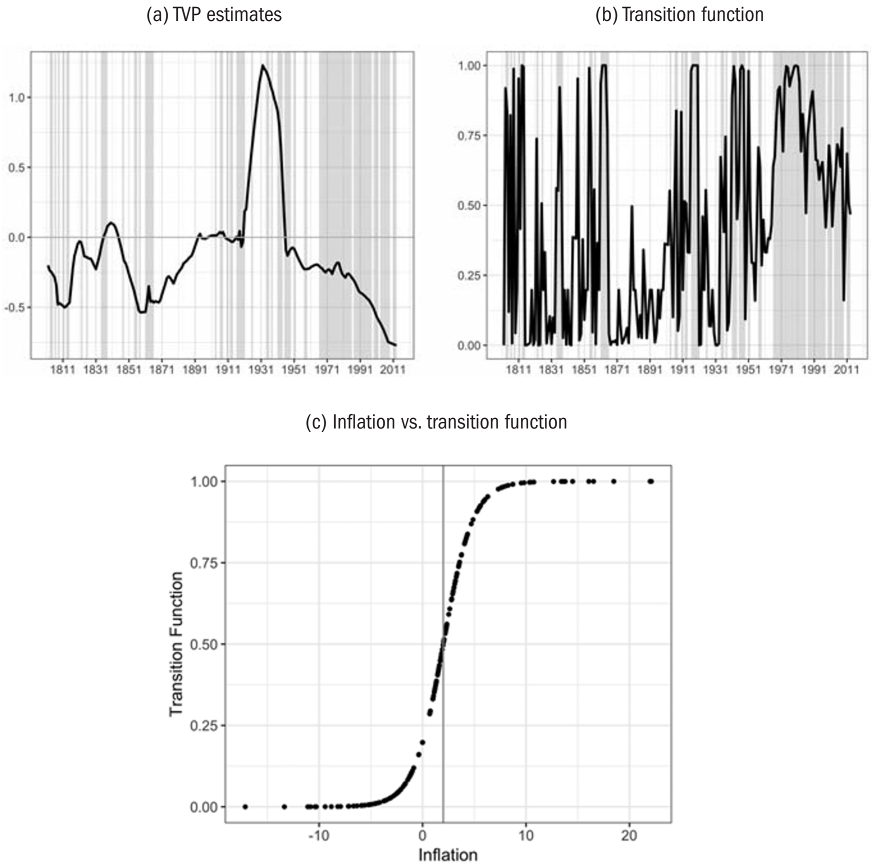
where c is the threshold inflation rate, $G(\pi_t;\gamma,c) \in (0,1)$ is a continuous function and allows a smooth transition between the below threshold regime ($\pi_t \leq c$) and the above threshold regime ($\pi_t > c$), and $\gamma > 0$ is the slope parameter that controls the speed of adjustment from one regime to the other. In our application $G(\pi_t;\gamma,c)$ is specified with a logistic function and takes the following form:

$$G(\pi_t;\gamma,c) = \left[1 + \exp\{-\gamma(\pi_t - c)\}\right]^{-1}. \quad (7)$$

In order to be consistent with the nonparametric estimate of the inflation threshold we set $c = 2\%$ in equation (7) and estimate the remaining parameters conditional on the pre-specified threshold. Estimates of the logistic STR model are shown in Table 4. The effect of inflation on GDP growth is positive but insignificant in the below threshold inflation regime ($\pi_t \leq 2\%$) while it is negative and significant in the above threshold inflation regime ($\pi_t > 2\%$). This result is in perfect agreement with our finding from the nonparametric analysis, where the estimated curve between growth and inflation is flat up to 2% for inflation, but significantly negatively sloped above the 2% inflation level. Figure 3(b) presents the estimates of the transition function given in equation (7). The shaded region in Figure 3(b) corresponds to the regime with above threshold inflation periods, i.e., $\pi_t > 2\%$. We also shade the same periods in Figure 3(a) to compare the STR and TVP estimates. The periods where the inflation has negative impact on growth correspond to periods 1803-1920, 1941-1950, and 1965-2013. These periods closely correspond to the negative impact estimates from the TVP model, showing that both the TVP and STR uncover similar results. Figure 3(c) is a scatterplot of the estimates of the $G(\pi_t;\gamma,c)$ in equation (7) against the threshold variable inflation. The shape of the transition function in Figure 3(c) show that the transition between the regimes is moderately slow as indicated by the estimate of γ, which 3.8543 as seen from Table 4, and more importantly points to nonlinearity.

[3] See Teräsvirta and Anderson (1992) and Granger and Teräsvirta (1993) for the details of the STR model.

Figure 3. Time varying parameter and threshold regression estimates



Notes: Part (a) presents the estimates of the slope parameter β_t from a time varying parameter (TVP) model $g_t = \alpha_t + \beta_t \pi_t + \varepsilon_t$. Part (b) presents the transition function $G(\pi_t;\gamma,c)$ of the two-regime smooth transition regression (STR) model $g_t = [\alpha_1 + \beta_1 \pi_t] + [\alpha_1 + \beta_1 \pi_t] G(\pi_t;\gamma,c) + \varepsilon_t$ with the threshold inflation c set to 2%. Part (c) presents the scatter plot of the transition function $G(\pi_t;\gamma,c)$ against the switch variable inflation (π_t). A vertical line at 2% threshold inflation is drawn on Part (c). The shaded regions in Parts (a) and (b) corresponds to the periods where $G(\pi_t;\gamma,c) > 0.50$, the case where the inflation is above the threshold level 2%.

The effects of deflation on growth should be analysed with caution. The adverse economic effects of the ZLB, such as the inability of monetary policy to stimulate aggregate demand in recessions (Krugman 1998), might outweigh the growth benefits of deflation. It would seem prudent for the FED to keep inflation expectations anchored at a very low level, given the consequences of the ZLB and balancing it with the economic effects of higher inflation; or in the model specification of Billi (2011) –commitment seems to be the appropriate policy response. An analysis of the economic benefits and adverse effects of deflation is beyond the scope of this

Table 4. Estimates of the logistic STR model

Parameter	Estimate	Standard error
Low regime parameters ($\pi_t \leq 2\%$)		
α_1	4.2830***	1.0650
β_1	0.1134	0.1558
High regime parameters ($\pi_t > 2\%$)		
α_2	4.8847**	2.4271
β_2	-0.3573**	0.1734
Nonlinear parameters		
γ	3.8543	2.7383
c	2.0000	–

Notes: Table reports the estimates of the STR model $g_t = [\alpha_1 + \beta_1\pi_t] + [\alpha_1 + \beta_1\pi_t]G(\pi_t;\gamma,c) + \varepsilon_t$ given in equation (6) with the threshold inflation c set to 2%. The transition function G(ptg,c) is specified as a logistic function given in equation (7). *, **, and *** denote significance at 10%, 5%, and 1%, respectively.

paper, but offers an interesting avenue for further study. The results, in this study, however, suggest that deflation, although not as strong as high inflation, also subtracts from economic growth.

The policy implications of these results are also quite difficult to advance. One might be tempted to gauge from the results that inflation much higher or lower than 2 percent permanently reduces growth. However, the empirical results rely on growth rates (i.e., they are stationary) and inference on the long-run is thus not possible. The result from the welfare literature on inflation mentioned earlier suggests that high inflation may permanently reduce economic growth. The results do not control for inflation expectations. As a consequence no inference can be made regarding the role of inflation expectations on economic growth or how agents substitute one type of asset class for another because of inflation expectations. We argue that the long time series data, and the implicit relationship between growth and inflation in this data, controls for changes in expectations. The paper does not advance a position on what the FED target for inflation should be. It only describes the relationship between inflation and growth and suggests that inflation roughly equal to 2 percent does not subtract from growth.

V. Conclusion

The effects of inflation on economic growth for the United States are studied using a novel time series dataset that spans over a century. A semiparametric instrumental variables method controls for endogeneity and allows the data to uncover results without imposing any functional forms or restrictions. The results consistently suggest that the inflation-growth relationship is nonlinear, with a threshold of about two percent. Two additional parametric specifications confirm the results. The hump shaped relationship implies that high inflation reduces economic growth proportionally more relative to low inflation. An interesting finding of this paper is that it shows that deflation can also reduce growth, contrary to some research findings. This paper does not suggest that the Fed should follow a deflation policy –the obvious consequence being that monetary policy becomes ineffective at the zero lower bound. This result does however warrant further investigation of a comparison of the economic benefits of deflation and the adverse effects of the zero lower bound on growth.

References

Billi, Roberto M (2011). Optimal inflation for the U.S. economy. *American Economic Journal:Macroeconomics* 3: 29–52.

Barro, Robert J (1995). Inflation and economic growth. Working Paper 5326. Cambride, MA, NBER.

Bruno Michael, and William Easterly (1998). Inflation crises and long-run growth. *Journal of Monetary Economics* 41:3–26.

Budekin, Richard C.K., Arthur T. Denzau, Manfred W. Keil, Thitithep Sitthiyot, and Thomas D. Willett (2004). When does inflation hurt economic growth? Different nonlinearities for different economies. *Journal of Macroeconomics* 26: 519–532.

Chari, V.V., Larry E. Jones, and Rodolfo E. Manuelli (1995). The growth effects of monetary policy. *Quarterly Review* 19: 18–32. Federal Reserve Bank of Minneapolis.

Fischer, Stanley (1981). Towards an understanding of the costs of inflation: II. *Carnegie-Rochester Conference Series on Public Policy* 15: 5–41.

Fischer, Stanley (1993). The role of macroeconomic factors in growth. Working Paper 4565. Cambridge, MA, NBER.

Granger, Clive W. J. and Timo Teräsvirta (1993). *Modelling nonlinear economic relationships*. Oxford University Press, Oxford.

Ireland, Peter N. (2009). On the welfare cost of inflation and the recent behavior of money demand. *American Economic Review* 99: 1040-1052.

Khan, Moshin S.. and Abdelhak S. Senhadji (2001). Threshold effects in the relationship between inflation and growth. *IMF Staff Papers* 48: 1–21.

Krugman, Paul.R. (1998). It's baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity* 29: 137–206.

Li, Qi, Juan Lin, and Jeffrey S. Racine (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business and Economic Statistics* 31: 57–65.

Miller, Stephen M., Luis Filipe Martins, and Rangan Gupta (2014). A time-varying approach of the U.S. welfare cost of inflation. Department of Economics, University of Connecticut, Working Paper No. 2014-11.

Park, Sanjin (2003). Semiparametric instrumental variable estimation. *Journal of Econometrics* 112: 381–399.

Reifschneider, David, and John C. Williams (2000). Three lessons for monetary policy in a low-inflation era. *Journal of Money, Credit and Banking* 32:936–966.

Roubini, Nouriel, and Xavier Sala-i-Martin (1992). A growth model of inflation, tax evasion, and financial repression. National Bureau of Economic Research. Working Paper 4062. Cambridge, MA, NBER.

Rousseau, Peter L., and Paul Wachtel (2001). Inflation, financial development and growth. In T. Negishi, R. Ramachandran, K. Mino, editors, *Economic theory, dynamics and markets: Essays in honor of Ryuzo Sato*. Kluwer. Available at SSRN:http://ssrn.com/abstract=252929.

Serletis, Apostolos, and Kazem Yavari (2004). The welfare cost of inflation in Canada and the United States. *Economics Letter* 84: 199–204.

Teräsvirta, Timo, and Heather M. Anderson (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7: S119–S136.

Vaona, Andrea, and Stefano Schiavo (2007). Nonparametric and semiparametric evidence on the long-run effects of inflation on growth. *Economics Letters* 94: 452–458.

Vaona, Andrea (2012). Inflation and growth in the long run: A new Keynesian theory and further semiparametric evidence. *Macroeconomic Dynamics* 16: 94–132.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 121-140

ROMER WAS RIGHT ON OPENNESS AND INFLATION: EVIDENCE FROM SUB-SAHARAN AFRICA

Faqin Lin and Dongzhou Mei*
Central University of Finance and Economics (CUFE)

Huanhuan Wang
East China Normal University

Xi Yao
Peking University

Submitted December 2015; accepted July 2016

Romer (1993) documents a negative relation between trade openness and inflation and offers an explanation based on time-inconsistency of monetary policy, but subsequent research casts doubt on the negative relationship and the explanation. This paper contributes to this debate by estimating the effect of openness to international trade on inflation with panel data from Sub-Saharan Africa. Employing instrumental variable techniques that correct for endogeneity bias of trade openness, the empirical evidence suggests that within-country variations in trade openness restrict inflation: a 1 percentage point increase in the ratio of trade over gross domestic product is associated with a decrease in inflation of approximately 0.08 percentage points per year. These results are robust to additional controls, different measurements of trade openness and alternative instruments. Finally, we inspect the time-inconsistency mechanism of the negative-relationship between trade openness and inflation.

JEL classification codes: C23, E31, E52, F41
Key words: trade openness, inflation, instrumental variables

---

* Faqin Lin (corresponding author): School of International Trade and Economics, Central University of Finance and Economics (CUFE), 39 South College Road, Haidian District, Beijing, P.R.China, 100081; linfaqin@126.com. Dongzhou Mei: School of International Trade and Economics, CUFE; meidongzhoupku@126.com. Huanhuan Wang: School of Law, East China Normal University, No.3663, Zhongshan North Road, Shanghai, P. R. China, 200062; ahwanghuanhuan@163.com Xi Yao: Guanghua School of Management, Peking University, No.5 Yiheyuan Road Haidian District, Beijing, P.R.China, 100871; yaoxi_mail@163.com. We thank the Editor Jorge Streb and and the anonymous referees for their insightful comments and suggestions. Faqin Lin acknowledges the support of the National Natural Science Foundation of China (71503281). Faqin Lin highly appreciates the support provided by the Program for Innovation Research, CUFE, and by the Young Elite Teacher Project, CUFE, to this work.

I. Introduction

Whether or not trade openness restricts inflation has been a widely debated question. In his seminal paper, for the first time, Romer (1993) observes that there is a negative correlation between trade openness and inflation. Subsequently, the negative relationship is called into question by several scholars. For example, Terra (1998) argues that the negative relation between trade openness and inflation in Romer (1993) largely depends on his data sample during the debt crisis period when most countries are heavily indebted. Temple (2002) even cautions that the direct evidence for Romer's negative correlation is not at all strong.

The above-mentioned empirical studies are based on cross-section data. One weakness of the cross-sectional regression design is the omitted variable bias. For instance, Terra (1998) shows that both the time period and the unobservable country-specific characteristics have an important influence on the relation between openness and inflation. While the effects of these characteristics of the countries and time period may be purged by including country fixed effects and year fixed effects, the cross-sectional regression design makes it infeasible to do so. The panel data methodology appears to be the most appropriate for dealing with this type of problem, because it allow us to detect effects that are either typical of certain countries or changing over time. Hence, we apply rigorous panel-data estimation techniques.

Sachsida, Carneiro and Loureiro (2003) are the first to use panel data to study the relation between openness and inflation. They find that the negative relation still holds, but they do not control for time effects and other important variables that may determine inflation. Controlling for time and country fixed effects, Alfaro (2005) finds that openness does not play a role in restricting inflation in the short run, which runs counter to Romer (1993). In this paper, we empirically examine how trade openness affects inflation. Using panel data on a sample of 46 Sub-Saharan Africa (SSA) countries during the period 1985-2012, we find that there is a negative causal relationship between trade openness and inflation.[1] Therefore, our paper contributes to the debate by reinforcing the view of Romer (1993) that trade openness restricts inflation.

Even though we can control for country-specific and year effects in panel data, the causal effect of trade openness on inflation is still not easy to identify empirically. The previous studies do not find a good way to solve the endogeneity issue, failing

[1] Table A in the Appendix presents the list of 46 SSAs.

to rule out the omitted variable bias or the reverse causality problem and therefore cannot convincingly isolate the causal effect of trade openness on inflation. Romer (1993) already cautions that there is an endogeneity problem in the relation between openness and inflation because trade openness is itself endogenous, and makes two IV estimates using different proxies for country size to instrument openness.

To examine the causal relationship between trade openness and inflation, in this paper, we adopt an instrumental variable strategy by using the GDP growth rates from the economies of OECD + China + India (SSA's most important trade partners) as an instrument for international trade openness of SSA economies.[2] Our instrumental approach is related to Brückner and Lederman (2015) who firstly use a similar instrument. There is a reason of such an IV. Higher real GDP growth of the OECD + China + India economies can increase trade openness of SSA countries through two main channels: the supply channel (higher OECD + China + India GDP growth increases their exports of goods and services) and the demand channel (higher OECD + China + India GDP growth increases the consumption of goods and services produced by Sub-Saharan African countries).[3] The fact that Sub-Saharan African countries' GDP is only a tiny fraction of the OECD + China + India economies' GDP ensures that variations in the OECD + China + India economies' GDP growth are plausibly exogenous to within-country variations of Sub-Saharan African countries' openness to international trade. For instance, the total GDP of the 49 SSA economies accounts for around 1% of world GDP in 2012 and averages 0.85% over the period 1985-2012.

In addition, we test the validity of the OECD + China + India growth as an instrument of trade openness by using rainfall as well as the Baltic Dry Index (BDI) cost as the instruments. In the literature, rainfall is a well-accepted instrument as an exogenous shock to income in SSA economies, and so possibly to trade openness. BDI cost has proved to be a good instrument for trade in poor countries, such as SSA, since they export a lot of primary goods but their economic scale is too small to drive BDI (Lin and Sim 2013; 2014). Though rainfall and BDI cost appear to be weak instruments for trade openness in SSA, on one hand, we find that our OECD + China + India growth variable enters insignificantly in the second stage using

[2] For example, SSA's trade with the OECD + China + India economies accounts for around 75 percent of SSA's total trade with the world over the period 1985-2012. The OECD + China + India economies account for 80 percent of SSA's total imports and 70 percent of SSA's total exports.

[3] It is worth noting that when we use the import GDP ratio of SSAs as a trade openness measure, as done in the literature, real GDP growth of the OECD + China + India economies mainly increase trade openness of SSA countries through the supply channel.

rainfall and BDI cost as instruments for trade openness. On the other, the OECD + China + India growth variable has a statistically and quantitatively significant effect on inflation from the reduced-form estimates. Hence, conditional on trade openness, we find that the OECD + China + India growth variable has no direct effect on inflation. In other words, the OECD + China + India growth variable can only affect inflation through the trade openness channel and our benchmark results of trade openness on inflation using the OECD + China + India growth variable as an instrument are valid.

We apply our instrumental variable strategy and show that trade openness has statistically significant and quantitatively large negative effects on inflation. The instrumental variable estimates suggest that, on average, a one percentage point increase in the ratio of imports over GDP in SSA countries is associated with a decrease in inflation of about 0.08 percentage points. The negative causal correlation between trade openness and inflation is robust to a number of robustness checks that include additional controls, such as the independence of the central bank, the exchange-rate policy, the government budget balance, the use of different measurements of trade openness and alternative instruments. Thus, from an identification perspective, our paper contributes to the line of research on identifying the causality of trade openness on inflation. More importantly, our instruments are country specific, thus, we can apply rigorous panel data estimation techniques that account for both unobservable cross-country heterogeneity and common year shocks and identify in our empirical analysis the effect of trade openness on inflation from, exclusively, the within country variation of the data.

Romer (1993) links the negative relationship between inflation and openness to the problem of time inconsistency of monetary policy, which he shows is more moderate in more open economies. Romer's explanation relies on the specific channel of negative terms-of-trade effects that counterbalance the inflationary bias. Lane (1997) holds that for countries that are not large enough to affect the structure of international relative prices, the terms-of-trade channel is not relevant. He offers a complementary explanation of the moderating effect of openness on the time-inconsistency problem that is based on a lower domestic incentive to engineer surprise inflation, so the effect is still present in small economies that are price-takers on international markets. Lane's explanation for the time-inconsistency problem applies to our sample of SSA countries because they can be considered to be small.

In order to show why there is a negative causal relationship between trade openness and inflation, we rely on this time-inconsistency-based hypothesis to inspect the mechanism more closely. Using a monetary dependence index, we find that the more independent the central bank, the weaker the negative relation between openness and inflation, which supports the time-inconsistency explanation.

The explanation of the time-inconsistency problem can be traced back to the Kydland and Prescott (1977) and Barro and Gordon (1983) time-inconsistency models. In these models, in the absence of an independent monetary authority, the more benefit the monetary expansion brings, the higher the incentives the government has to make surprise inflation.[4] However, there are other logics for the time-inconsistency problem. For example, Ruge-Murcia (2003) offers an alternative explanation where the time-consistency problem does not have to do with a positive inflationary bias, as in Barro and Gordon (where the monetary authority aims at an unemployment level lower than the natural rate), but rather with asymmetric reactions to positive and negative shocks.

There are several reasons why we focus on Sub-Saharan African countries in this paper. The first is that the IV strategies are particularly suitable for SSA. As already mentioned before, Sub-Saharan African GDP is only a tiny fraction of OECD GDP. Thus, GDP growth of OECD countries is plausibly exogenous to variations in Sub-Saharan African countries' openness to international trade. Secondly, it is important to note the significant policy interest in factors associated with inflation in Sub-Saharan Africa. Inflation is not good for the national welfare, especially high inflation, which leads to important redistributive effects across households and usually benefits rich people but harms poor ones. For instance, Lucas (2000) show that the loss from increasing the annual inflation rate from zero to 10 percent is equivalent to a decrease in real income of around one percent.[5]

Unfortunately, the inflation in SSA is very high compared with the rest of the world average and OECD countries. In 2012, the inflation for SSA was 6.54 percent, while for the OECD countries it was only 1.53 percent, and the world average was 3.51 percent.[6] For some SSA's country, like, Sudan and Ethiopia, the inflation rate

---

[4] Lohmann (1998) shows that with asymmetric information on policy decisions, time-consistency problems are exacerbated by political business cycles because incumbents have an incentive to produce monetary expansions in election years to increase their chances of winning. In the SSA region, most of the countries (over 70%) are not democracies so this explanation would not apply.

[5] Lagos and Wright (2005) show such effect of the loss from increasing the annual inflation rate from zero to 10 percent is equivalent to a decrease in real income of 1.3 percent and Chiu and Molico (2011) show such effect is around 0.6 percent.

[6] The average inflation rate over the period 1985-2012 is 6.81, 4.87, and 2.23 percent, for SSA, the world and the OECD, respectively.

in 2012 was as high as 40 and 34 percent. If trade openness can help to restrict inflation, it will be another important benefit to SSA besides income and productivity (Frankel and Romer 1999; Feyrer 2009a, 2009b, Van Biesebroeck 2005; Brückner and Lederman 2015; Lin and Sim 2013, 2014).[7]

Thirdly, due to the insular economy in SSAs, there is a significant policy interest in factors driving trade growth in Sub-Saharan Africa. See, for example, World Economic Forum (2011), IMF (2011), World Bank (2011, 2012), or the African Growth and Opportunity Act, which enabled African countries to export over 4000 products, including hundreds of apparel products, quota-free and duty-free to the US, has been the object of much economic research.[8]

The remainder of the paper is structured as follows. Section II introduces data, in particular, our instrumental variable, the growth of OECD + China + India. The empirical strategy is introduced in Section III. Section IV presents the results of our IV regression of trade openness on inflation. Section V discusses several robustness checks of our benchmark results and provides a possible mechanism based on the ideas in Romer (1993). Section VI concludes the paper.

II. Data

Our data spans from 1985 to 2012. The key variables for the subsequent econometric analyses are trade openness and inflation in Sub-Saharan Africa countries. Trade and GDP in developing and developed countries are also important variables which are later used as an IV. In addition, in the main regressions and robustness checks we also include variables related to inflation such as GDP per capita, budget balance, government debt ratio to GDP, financial openness, and exchange rate flexibility. To check the quality of our instrument, we also use other IVs in the literature such as rainfall and Baltic Dry Index. To explain our finding, we use monetary dependence index. The following paragraphs describe the relevant data, their sources and the variable definitions (see Table 1).

---

[7] Van Biesebroeck (2005) demonstrates that exporting increases firms' productivity in SSA. Brückner and Lederman (2015) show that in SSA economies such effect in short-run is 0.5 percent and in long-run is about 0.8 percent. Lin and Sim (2014) show that a one percentage point expansion in trade raises the GDP per capita of the SSA countries by approximately 0.6–0.7 percentage points and Lin and Sim (2013) show such effect for LDCs, 33 of which are located in SSA, is around 0.5.

[8] Collier and Venables (2007) and Frazer and Biesebroeck (2010) both find that AGOA trade preferences had a positive and significant impact on exports from Africa to the US.

To be consistent with the literature, openness is measured as the average share of imports (including goods and services) in GDP. When we use the share of both exports and imports in GDP, the results are robust. The results are not sensitive to using only imports of goods.[9] Inflation is measured as the average annual change in the log GDP deflator, as in the literature. The openness variables come from the WDI and the UNCTAD trade database.[10] The GDP deflator is from WDI too.

We construct a bilateral-trade-weighted GDP growth rate of OECD + China + India (*OECDCI*) trading partners for each country in our SSA sample. For country

Table 1. Data variables, definitions and sources

Variables	Definition	Source
Inflation	Annual change in the log of the GDP deflator	WDI
Import	Share of imports in GDP	WDI
Export	Share of exports in GDP.	WDI
Trade	Share of exports and imports in GDP.	WDI
GDP per capita	Real GDP per capita at PPP (Purchasing Power Parity)	WDI
GDP growth	GDP annual growth rate	WDI
Bilateral trade	Imports by SSA country from OECD + China + India	UNCTAD and Feenstra et al. (2005)
Budget balance	Government budget balance as a percentage of GDP	WDI
Government debt	Government debt as a percentage of GDP	WDI
BDI cost	General indicator of shipment rates for dry bulk cargoes	The Baltic Exchange and Lin and Sim (2013)
Financial openness	A country's degree of capital account openness	Chinn and Ito (2006)
Rainfall	Log of annual rainfall	GPCP and Bruckner's paper
Exchange rate flexibility	Exchange rate flexibility index	IMF
Monetary dependence	Monetary dependence index	Aizenman, Chinn and Ito (2008)

Note: *Inflation*, *Import*, *Export*, *Trade*, *GDP growth*, *Bilateral trade*, and *Financial openness* have the same unit, where a 1 percentage point increase of these variables is counted as 0.01.

[9] See the robustness check in section V.

[10] See http://data.worldbank.org/news/world-development-indicators-2012- update and http://www.unctad.org/Templates/Page.asp?intItemID=1584&lang=1.

i and year t, the trade-weighted GDP growth rate of OECD + China + India trading partners is constructed as:

$$OECDCI_{it} = \sum_{i=1}^{n} \theta_{ij} \ GDP \ growth_{jt} \tag{1}$$

where n is the number of OECD + China + India economies, and θ_{ij} is the imports in goods by country i from j as a share of country i's GDP in 1985, which is the first year during our sample period (the service bilateral trade data is unavailable).[11] Its predetermined nature motivates the use of the period's starting-year's trade share as the interaction term in (1). The bilateral trade data are from UNCTAD and Feenstra et al. (2005). Data on the real GDP growth rate of *OECDCI* economies are from the WDI. To check the quality of the instrument, we also use rainfall and Baltic Dry Index. Our data on year-to-year variations in rainfall are from the National Aeronautics and Space Administration (NASA) Global Precipitation Climatology Project (GPCP).[12] The BDI data is drawn from the London-based Baltic Exchange office.

As to other controls, data on income (real GDP per capita), budget balance, government debt ratio to GDP, and financial openness are also from WDI. The exchange rate flexibility classification is taken from the International Monetary Fund's (IMF) *Annual Report on Exchange Arrangements and Exchange-Rate Restrictions*: the greater the value, the more flexible the exchange rate. We use the monetary dependence index from Aizenman, Chinn and Ito (2008) to test the Romer (1993) independent monetary authority hypothesis; the smaller the value, the more independent the monetary authority. For summary statistics on these variables, see Table B in the online appendix.

III. Empirical strategy

Our main estimating equation relates inflation, measured by the first difference of the log of the GDP deflator for country i in year t, as follows:

$$Inflation_{it} = C_y + \beta \ Import_{it} + \gamma Z_{it} + \mu_i + \mu_t + \delta^i Trend + \vartheta_{i,t}, \tag{2}$$

---

[11] Our results are robust to using openness variable only considering goods, as shown in Section III.

[12] We thank Markus Bruckner's generosity for sharing the data with us.

where $Import_{it}$, the share of imports in GDP, is the main causal variable of interest. C_y is the constant term. Z includes other controls such as GDP per capita, budget balance, government debt ratio to GDP, financial openness, and exchange rate flexibility in our main regressions and robustness checks. We let μ_i be the country fixed effects that represent time invariant permanent differences across countries, μ_t the common time effects, and $\delta^i Trend$ a country-specific linear time trend that captures additional within-country time series variation. Finally, $\vartheta_{i,t}$ is the idiosyncratic error term clustered at the country level.

The extent of how openness affects inflation is summarized by β. This cannot be consistently estimated by OLS regression as openness is likely to be endogenous in the inflation equation, in spite of controlling for country-specific characteristics such as country and year fixed effects and common time trend to solve the omitted variable bias. Firstly, decisions on whether to trade, and how much to trade, are not randomly assigned. Secondly, the regression analysis may be confounded by the reverse causal effect going from inflation to imports. For instance, high inflation leads to depreciation of domestic currency, and this will decrease imports; hence the OLS regression is susceptible to reverse-causality problems.

This paper uses the variable *OECDCI* to obtain the exogenous variation in the import openness of the SSAs. The estimating equation that relates imports to the *OECDCI* is given by:

$$Import_{it} = C_I + \alpha\ OECDCI_{\mathrm{it}} + \rho Z_{it} + \mu_i + \mu_t + \delta^i Trend + \epsilon_{i,t}, \tag{3}$$

where C_I is a constant term and $\epsilon_{i,t}$ is the idiosyncratic error term clustered at the country level. This strategy relies on the variable we labeled $OECDCI_{it}$ –the trade-weighted real GDP growth rate of *OECDCI* trading partners– which we use as an instrument for trade openness (for robustness purposes, we also present the results excluding China and India). The assumption is that changes in economic conditions of non-*OECDCI* economies have negligible effects on the GDP growth rate of *OECDCI* economies. For the group of Sub-Saharan African countries this exogeneity assumption is plausible: for all years since 1985, the GDP of the group of Sub-Saharan African countries is less than 2 percent of *OECDCI* economies' GDP (WDI 2013).

The exclusion restriction is that $OECDCI_{it}$ should only affect Sub-Saharan African countries' inflation through trade openness. With regard to this exclusion restriction, the year fixed effects are an important ingredient in our estimating framework since they pick up Sub-Saharan Africa-wide variation in inflation that is due to *OECDCI*-wide growth in GDP. The instrumental variables regressions that use $OECDCI_{it}$ as an instrumental variable therefore only use variations in *OECDCI* economies' GDP growth that is specific to each Sub-Saharan African country. That is, these are country-specific variations in GDP growth. They arise precisely because bilateral trade flows between each *OECDCI* and Sub-Saharan African country are country specific. It is important to note that we use 1985 bilateral trade flows to construct the trade weighted *OECDCI* GDP growth instrument. Using time-invariant bilateral trade flows ensures that within-country variations in Sub-Saharan African countries' inflation do not affect the *OECDCI* growth instrument.

To make the argument more solid, we test the exclusion restriction formally. As Lin and Sim (2013) have proclaimed, small economic scale and trade participation will not affect BDI, which makes BDI cost a powerful instrument for SSA trade (especially for exports).[13] In addition, the agricultural sector in SSA economies is large: roughly one-third of GDP comes from agriculture, and over two-thirds of the population is employed in agriculture, which makes rainfall a powerful instrument as a determinant of SSA income.[14]

Our main causal variable is import openness. Though these two instruments are not directly determining imports, BDI cost and rainfall may affect imports indirectly through, for example, the income channel. Thus, we rely on these two variables to test the validity of $OECDCI_{it}$ as an instrument for imports by using rainfall and BDI cost as the instrument of imports.[15] We show that in the second stage, $OECDCI_{it}$ has a statistically insignificant and quantitatively small effect on inflation after adopting BDI cost and rainfall as instruments for imports, but we find a significant effect of $OECDCI_{it}$ on inflation from the reduced-form estimates in Section IV. Hence, conditional on imports we find that $OECDCI_{it}$ has only very small effects on inflation.

After checking the validity of our instrument, equation (2) is estimated using two-stage least squares in conjunction with (3) as the first-stage regression. We also estimate the effect of $OECDCI_{it}$ on income by looking at the reduced form equation:

$$Inflation_{it} = C_G + \sigma\ OECDCI_{it} + \tau Z + \mu_i + \mu_t + \delta^i Trend + \varepsilon_{i,t}. \quad (4)$$

---

[13] See Lin and Sim (2013, 2014) and Lin et al. (2014) for more details about BDI cost, which is constructed by the primary products share and the BDI.

[14] Rainfall is a well-accepted instrument as an exogenous shock to income in SSA economies in the literature. See, e.g., Miguel et al. (2004), Barrios et al. (2010), Brückner (2010), Brückner and Ciccone (2011), Miguel and Satyanath (2011), Ciccone (2011), Arezki and Brückner (2012) as well as Brückner (2013).

[15] Though they are weak instruments, see section IV for details.

Equation (4) allows us to investigate directly the within-country effect that *OECDCI* has on inflation that is facilitated by the trade openness channel.

IV. Results

We begin our empirical analysis by estimating the response of within-country variations in inflation to trade openness. Openness is not significant in any specification with OLS regressions. Thus, we look into the 2SLS results to see how trade openness affects inflation using an instrumental approach. In addition, we discuss the quality of our instrumental variable $OECDCI_{it}$ when estimating the 2SLS regression.

A. IV regression results

Table 2 presents the 2SLS estimates of the causal effect of trade openness on inflation by exploiting plausibly exogenous variation of imports that is driven by $OECDCI_{it}$. Columns I reports the 2SLS results where the endogeneity test rejects the hypothesis that trade openness is exogenous.[16] We show that there is a statistically significant (at 0.01 levels) and quantitatively large negative effect of import openness on inflation, which is very different from our OLS result (see online appendix). The estimated coefficient implies that on average a one percentage point higher trade openness level is associated with around 0.08 percentage points lower inflation level. As to its economic significance, a one standard deviation change in openness from its mean value of 0.440 affects inflation by 0.08 x 0.327 = 0.02616. This impact of 2.6 percentage points is economically significant: given an average annual inflation of is 5.7% in the 46 SSA countries in the sample, inflation falls from from 8.3% to 3.1% as openness rises from one standard deviation below the mean to one standard deviation above it.

Concerning identification, the first stage results in Column II suggest that the OECD + China + India economic growth variable ($OECDCI_{it}$) is a strong, positive determinant of trade openness: a one percentage point increase in economic growth leads to about 1 percentage point increase in trade openness. Column III reports the least squares (reduced form) estimates of the within-country effect that the OECD + China + India economic growth variable ($OECDCI_{it}$) has on inflation through imports. The increase of the economic growth decreases the inflation levels in SSAs.

[16] Before we discuss the coefficient of greatest interest to us, we briefly discuss the other determinants of inflation. The empirical model seems to work well. It delivers precisely estimated coefficients that are sensible and similar to those estimated by others. For instance, income is negatively related to inflation (Romer 1993; Alfaro 2005).

Table 2. 2SLS results of trade openness on inflation

	I	II	III	IV
	2SLS	First stage	Reduced form	System GMM
Dependent variables	*Inflation*	*Import*	*Inflation*	*Inflation*
Import	-0.081***			0.0004
	(0.019)			(0.0012)
Log(*GDP per capita*)	-0.017*	-0.099	-0.0035	-0.0033*
	(0.0091)	(0.0104)	(0.0033)	(0.0019)
Budget balance	0.013	0.185	-0.0017	-0.001
	(0.010)	(0.134)	(0.0011)	(0.0012)
OECDCI		1.066***	-0.0713***	
		(0.160)	(0.0144)	
$Inflation_{-1}$				0.0034***
				(0.001)
Angrist Pischke F test		44.67		
Endogeneity test (P)	0.000			
Hansen J statistic (p)				1.000
AR(1)				0.052
AR(2)				0.631
Country effect	yes	yes	yes	yes
Year effect	yes	yes	yes	yes
Country trend	yes	yes	yes	yes
Observations	976	976	1001	1204
R-squared	-5.40	0.018	0.056	

Note: Clustered robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

For comparison purposes, we report in Column IV the system-GMM estimates. The estimate of trade openness is positive in sign and statistically insignificant and quantitatively, they are near zero as the OLS results. As already analyzed, one reason for this difference in size is reverse causality bias: if inflation has a positive effect on imports, the system-GMM estimate could have upward bias. The instrumental variable estimates that use of $OECDCI_{it}$ as an instrument for imports do not suffer from this reverse causality bias because year-to-year variations in $OECDCI_{it}$ are exogenous to economic conditions of Sub-Saharan African countries. Another reason for the quantitatively larger (in absolute value) instrumental variable estimates is measurement error. It is well known that national accounts statistics of Sub-Saharan African countries are plagued by measurement error. To the extent that this measurement error is classical, it will attenuate the least squares and system-GMM estimates towards zero but not the instrumental variable estimates.

Our absolute values of the negative estimates are much larger than previous estimates in the literature. Two observations can be made. Firstly, endogeneity issues could lead to such a bias; for example, classical measurement error would bias the estimate towards zero. Secondly, because our paper focuses on the SSA countries while the previous literature looks at both developed and developing countries, our 2SLS estimates indicate that in low-income countries there could be a larger response of inflation to the opening of trade.

B. Instrument quality

The quality of $OECDCI_{it}$ as an instrument for imports in terms of first-stage fit is reasonable. The first-stage estimates on $OECDCI_{it}$'s effect on import openness are positive and significant at the 1 percent level. Also, the joint first-stage F-statistic on $OECDCI_{it}$ in Column I of Table 2 is above the rule-of-thumb threshold of 10 suggested by Staiger and Stock (1997). What about the exclusion restriction? The assumption in the instrumental variable regression is that $OECDCI_{it}$ only affects inflation through its effect on imports. To examine empirically whether there are any significant effects of $OECDCI_{it}$ on inflation beyond trade openness, we rely on two well-accepted instruments for trade and income in SSA in the literature for this purpose. One is BDI cost for trade and income raised by Lin and Sim (2013, 2014), which has proved to be a powerful instrument for SSA trade (especially for exports). Another is rainfall used for income by, for instance, Miguel et al. (2004) and Brückner and Ciccone (2011) since agricultural sector in SSA economies is large.

Though these two instruments are not directly determining imports, BDI cost and rainfall may affect imports indirectly through, for example, income channel. Thus, we rely these two instruments to test the validity of the $OECDCI_{it}$ as an instrument for imports by using rainfall and BDI cost as the instrument for import openness. To make it more intuitive, we include $OECDCI_{it}$ directly in the second stage. Table 3 reports the results. The first stage result suggests that BDI cost reduces imports by increasing trade costs while rainfall increases imports through the income channel. Second stage result shows that, $OECDCI_{it}$'s effect on inflation beyond imports is statistically insignificant. For the reader's convenience and comparison purpose, we repeat the results from our reduced regression in Column III, where we find the effect of $OECDCI_{it}$ on inflation is statistically significant

Table 3. Test of exclusion restriction for the IV: OECDCI

	I	II	III
	2SLS	First stage	Reduced form
Dependent variables	*Inflation*	*Import*	*Inflation*
Import	-0.02**		
	(0.01)		
OECDCI	-0.068		-0.0713***
	(0.056)		(0.0144)
BDI cost		-0.111*	
		(0.060)	
Rainfall		0.086**	
		(0.037)	
Angrist Pischke F test		3.28	
Log(*GDP per capita*)	yes	yes	yes
Budget balance	yes	yes	yes
Country effect	yes	yes	yes
Year effect	yes	yes	yes
Country trend	yes	yes	yes
Observations	811	811	1001
R-squared	-0.188	0.529	0.056
First stage Angrist Pischke F test for combined instruments			
OECDCI & BDI cost		0.569	
OECDCI & Rainfall		1.444	
OECDCI & BDI cost & Rainfall		1.070	

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

at the 1 percent level. Hence, conditional on import openness we find that $OECDCI_{it}$ has no significant effects on inflation.

Since BDI cost and rainfall could be instruments of imports, readers may argue: why we do not rely on multiple instruments for identification but just on one instrument $OECDCI_{it}$? The reason is that Column II shows that the instruments are weak since the joint first-stage F-statistic for BDI cost and rainfall is 3.28, which is far below the rule-of-thumb threshold of 10 suggested by Staiger and Stock. Furthermore, in Table 3, we check the weak instruments issue by using other combinations between BDI cost, rainfall and $OECDCI_{it}$, and all the joint first-stage F-statistic are very low. Thus, we use $OECDCI_{it}$ for identification, since the first-stage F-statistic is far above 10.

We do several robustness checks. Firstly, we look at the sensitivity of the baseline estimates to the inclusion of additional explanatory variables that might be relevant in determining inflation. Our second robustness check relates to using different trade openness indicators. The third robustness check looks at the sensitivity of our benchmark results to using slightly different instrument. The baseline 2SLS results are robust to all these checks and are reported in the online appendix.

C. Mechanism: why is there a negative causal trade openness-inflation relation?

In this subsection, we try to offer supporting evidence to explain the negative relationship between trade openness and inflation. Following Romer (1993), if the openness-inflation relationship arises from the dynamic inconsistency of discretionary monetary policy, the relationship should be weaker in countries that have more independent central banks, since one would expect these countries to have had more success in overcoming the dynamic inconsistency problem. Using the central bank independence index from Aizenman, Chinn and Ito (2008), we investigate interactions between monetary dependence measures and openness.

Column I in Table 4 shows the effect of adding the measure of monetary dependence interacted with trade openness as a further control variable. The estimated impact of openness on inflation is moderately reduced by including the interaction measure of lack of monetary independence in Column I (the interaction term generates the expected negative sign but is statistically insignificant).

In Column II, after including the interactions as well as the monetary dependence index, the openness-inflation effect is over-turned, becoming positive. As expected, the monetary dependence index is strongly associated with average inflation: the less independent, the higher inflation. More importantly, the interaction term enters with the expected negative sign and is quantitatively large (the absolute value is far larger than the positive trade openness estimate). This pattern is consistent after including other controls, as we do in the robustness check in Column III and IV. Unlike Alfaro (2005), we do not see a significant effect of exchange rate flexibility and the interaction between openness and exchange rate flexibility in Column V.

Thus, the negative relationship between openness and inflation is much stronger in countries that have a less independent monetary authority (higher value of the index). To see more evidence on such pattern, we provide the results in Table 5 by splitting the sample into two groups: high central bank independenc (less the mean value) and low central bank independence (above the mean value). We can see that trade openness significantly reduces inflation when there is low independence of the monetary authority, while there is no significant effect when there is high independence. To summarize, openness plays a significant role in restricting inflation even in the short-run based on our panel data analysis. Hence, our results are consistent with the Romer (1993) explanation of this negative link through the time-inconsistency logic.

Table 4. Mechanism investigation results

	I	II	III	IV	V
Dependent variable: *Inflation*					
Import	-0.07***	0.11*	0.10*	0.122**	0.121
	(0.011)	(0.06)	(0.058)	(0.061)	(0.164)
Import × *Monetary*	-0.047	-0.364***	-0.358***	-0.394***	
dependence	(0.037)	(0.135)	(0.129)	(0.145)	
Import × *Exchange*					-0.204
rate flexibility					(0.193)
Monetary dependence		0.137***	0.140***	0.148***	
		(0.05)	(0.06)	(0.057)	
Exchange rate flexibility					0.075
					(0.072)
Log(*GDP per capita*)	yes	yes	yes	yes	yes
Budget balance	yes	yes	yes	yes	yes
Government debt			yes	yes	
Financial openness			yes	yes	
GDP growth				yes	
Exchange rate flexibility				yes	
Angrist Pischke F test	35.28	8.012	8.012	6.733	0.632
Country effect	yes	yes	yes	yes	yes
Year effect	yes	yes	yes	yes	yes
Country trend	yes	yes	yes	yes	yes
Observations	907	907	907	880	779
R-squared	-7.630	-9.811	-9.810	-10.331	-130

Note: Clustered robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Mechanism investigation results: splitting the sample

	I	II
Dependent variable: *Inflation*	High central bank independence	Low central bank independence
Import	0.0036	-0.179***
	(0.0044)	(0.051)
Other controls	yes	yes
Country effect	yes	yes
Year effect	yes	yes
Country trend	yes	yes
First stage Angrist Pischke F test	9.734	14.69
Observations	349	627
R-squared	-0.024	-22.65

Note: Clustered robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

VI. Conclusion

Estimates of the effect of trade openness on inflation remain contentious. The literature has raised concerns about the validity and interpretation of cross-country evidence. In particular, since Romer (1993)'s seminal paper, it remains unclear whether a negative causal correlation between trade openness and inflation withstands the test. This paper tackled the debate with panel data and new IV strategies applied mainly to the case of Sub-Saharan Africa.

Our approaches entailed an instrumental-variable identification strategy. The strategy relied on the GDP growth of the OECD + China + India economies as an IV for trade openness in Sub-Saharan Africa. The results appear to be robust and the IVs seem to be both valid and relevant based on our IV quality test techniques. The results show that, on average, trade openness appears to have a significant negative effect on inflation in Sub-Saharan Africa. A 1 percentage point increase in the ratio of trade over gross domestic product is associated with a decrease in inflation of approximately 0.08 percentage points per year, while the OLS estimate is small and biased towards zero. These results are robust to the inclusion of additional controls such as government debt, financial openness and exchange rate flexibility and are not sensitive to the use of different measurements of trade openness and alternative instruments.

The comparison between our 2SLS results and OLS estimates and previous estimates in the literature highlights the importance of addressing the endogeneity issue. Furthermore, because our paper focuses on SSA countries, our 2SLS estimates indicate that in low income countries there could be a large response of inflation to the opening of trade. Thus, our result suggests some significant policy interest in the function of trade openness on inflation in Sub-Saharan Africa since inflation is high in SSA and it is not good for the national welfare.

Finally, we inspect the mechanism of the negative-relationship between trade openness and inflation and find that our results are consistent with Romer (1993). Romer was right on the fact that the time-inconsistency logic provides an explanation for this negative link.

References

Aizenman, Joshua, Menzie D. Chinn, and Hiro Ito (2008). Assessing the emerging global financial architecture: Measuring the trilemma's configurations over time. Working Paper 14533, Cambridge, MA, NBER.

Alfaro, Laura (2005). Inflation, openness, and exchange-rate regimes: The quest for short-term commitment. *Journal of Development Economics* 77: 229-49.

Arezki, Rabah, and Markus Brückner (2012). Rainfall, financial development, and remittances: Evidence from sub-Saharan Africa. *Journal of International Economics* 87:377-85.

Barrios Cobos, Salvador, Luisito Bertinelli, and Eric Strobl (2010). Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. *Review of Economics and Statistics* 92:350-66.

Barro, Robert J., and David B. Gordon (1983). A positive theory of monetary policy in a natural rate model. *Journal of Political Economy* 91:589-610.

Brückner, Markus (2013). On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics* 28:126-50.

Brückner, Markus, and Antonio Ciccone (2011). Rain and the democratic window of opportunity. *Econometrica* 79:923-47.

Brückner, Markus, and Daniel Lederman (2015). Trade openness and economic growth: Panel data evidence from sub-Saharan Africa. *Economica* 82:1302-23.

Chinn, Menzie D., and Hiro Ito (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics* 81:163-192.

Chiu, Jonathan, and Molico Miguel (2011). Uncertainty, inflation, and welfare. *Journal of Money, Credit and Banking* 43: 487-512.

Ciccone, Antonio (2011). Economic shocks and civil conflict: A comment. *American Economic Journal: Applied Economics* 3:215-27.

Collier, Paul, and Anthony J. Venables (2007). Rethinking trade preferences: How Africa can diversify its exports. *The World Economy* 30:1326-45.

Deaton, Angus S. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics* 87:1-19.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo (2005). World trade flows: 1962-2000. Working Paper 11040. Cambridge, MA, NBER.

Frazer, Garth, and Johannes Van Biesebroeck (2010). Trade growth under the African Growth and Opportunity Act. *The Review of Economics and Statistics* 92: 128-44.

Feyrer, James (2009a). Trade and income - exploiting time series in geography. Working Paper 14910. Cambridge, MA, NBER.

Feyrer, James (2009b). Distance, trade and income - The 1967 to 1975 closing of the Suez Canal as a natural experiment. Working Paper 15557. Cambridge, MA, NBER.

Frankel, Jeffrey A.,and David Romer (1999). Does trade cause growth? *American Economic Review* 89: 379-99.

IMF (2011), *Sub-Saharan Africa: Sustaining the expansion*. Washington, D.C.

Kydland, Finn E., and Edward C. Prescott (1977). Rules rather than discretion: the inconsistency of optimal plans. *Journal of Political Economy* 85:473-91.

Lagos, Ricardo, and Wright Randall (2005). A united framework for monetary theory and policy analysis. *Journal of Political Economy* 113: 463-84.

Lane, Philip R. (1997). Inflation in open economies. *Journal of International Economics* 42: 327-47.

Lin, Faqin, and Nicholas C.S. Sim (2013). Trade, income and Baltic Dry Index. *European Economic Review* 59:1-18.

Lin, Faqin, and Nicholas C.S. Sim (2014). Baltic Dry Index and the democratic window of opportunity. *Journal of Comparative Economics* 42:143-59.

Lohmann, Susanne (1998). Rationalizing the political business cycle: A workhorse model. *Economics and Politics* 10:1-17.

Lucas, Robert E. (2000). Inflation and welfare. *Econometrica* 68: 247-74.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti (2004) Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*112:725-53.

Miguel, Edward, Shanker Satyanath (2011). Re-examining economic shocks and civil conflict: A comment. *American Economic Journal: Applied Economics* 3:228-32.

Romer, David (1993). Openness and inflation: Theory and evidence. *Quarterly Journal of Economics* 108: 869-903.

Ruge-Murcia, Francisco J. (2003). Does the Barro-Gordon model explain the behavior of U.S. inflation? A reexamination of the empirical evidence. *Journal of Monetary Economics* 50:1375-90.

Sachsida, Adolfo, Francisco Galrão Carneiro, and Paulo R.A. Loureiro (2003). Does greater trade openness reduce inflation? Further evidence using panel data techniques. *Economics Letters* 81: 315-19.

Staiger, Douglas, James H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65:557-86.

Terra, Cristina (1998). Openness and inflation: A new assessment. *Quarterly Journal of Economics* 113:641-648.

Temple, Jonathan (2002). Openness, inflation, and the Phillips curve: A puzzle. *Journal of Money, Credit, and Banking* 34:450-68.

Van Biesebroeck, Johannes (2005). Exporting raises productivity in sub-Saharan African manufacturing firms. *Journal of International Economics* 67: 373-91.

World Bank (2011). Africa's future and the World Bank's support to it. Washington, D.C.

World Bank (2012). De-fragmenting Africa: Deepening regional trade integration in goods and services, Washington, D.C.

World Economic Forum (2011). *Africa competitiveness report*. Geneva.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 141-168

IS GERONTOCRACY HARMFUL FOR GROWTH? A COMPARATIVE STUDY OF SEVEN EUROPEAN COUNTRIES

Vincenzo Atella and Lorenzo Carbonari*
Università di Roma "Tor Vergata"

Submitted October 2014; accepted July 2016

We study the relationship between gerontocracy and aggregate economic performance in a simple model where growth is driven by human capital accumulation and productive government spending. We show that less patient élites display the tendency to underinvest in public education and productive government services, and thus are harmful for growth. The damage caused by gerontocracy is mainly due to the lack of long-term delayed return on investments, originated by the lower subjective discount factor. An empirical analysis using public investment in Information and Communication Technologies (ICT) is carried out to test theoretical predictions across different countries and different economic sectors. The econometric results confirm our main hypotheses.

JEL classification codes: J1, O4
Key words: gerontocracy, economic growth and aggregate productivity, education, ICT

I. Introduction

Over the last twenty years, per capita income growth rates have ceased to converge across OECD countries, and there has been a surge of academic research and policy attention about the causes underlying differences in economic growth performance

---

* Lorenzo Carbonari (corresponding author): Dipartimento di Economia e Finanza, Università di Roma "Tor Vergata" and CEIS, Via Columbia 2, Italy; +39 06 72595708, e-mail: lorenzo.carbonari@uniroma2.it. Vincenzo Atella: Università di Roma "Tor Vergata", CEIS and CHP-PCOR Stanford University, Via Columbia 2, Italy; +39 06 72595601, e-mail: atella@uniroma2.it. The authors are grateful to Federico Belotti, Stefano Gagliarducci, Nigar Hashimzade, Joseph Zeira for their helpful comments and suggestions, and the session participants at the 2011 International Economic Association meeting, the 2011 Italian Economic Association meeting and the 2013 Structural Change, Dynamics, and Economic Growth Conference for their comments. Francesca Vinci, Francesco P. Parabita, Fabrizio Purpura and Francesco Rinaldi provided excellent research assistance.

across these countries. While productivity has accelerated in emerging economies and, most notably, in the United States, it has substantially slowed down in continental Europe and Japan (OECD 2004). Focusing on Europe, it is easily observed that, since the mid-1990s, economic performance has experienced a significant contraction compared to earlier periods. The economic literature developed so far has provided various explanations for such a sclerosis (Blanchard 2002; Gordon 2004). The most commonly cited causes of the slow growth concern the rigidity of the European economic model, the burden of taxation, the strict dependency of citizens on the welfare system, and the evidence that Europe has used some of the past productivity improvements to increase leisure rather than income. In particular, a wide consensus has been reached among researchers regarding the "European model", which, despite its successes during the post-war era, is proving to be inadequate now that economic development is increasingly based on innovation, and national firms can no longer be protected from foreign competition. Moreover, several researchers point out that the adoption of important general purpose technologies associated with the Information and Communication Technologies (ICT) revolution has been hindered or impeded in Europe by an excessively regulated labor market and an insufficient level of competition (van Ark et al. 2008). Although this productivity crisis is a common feature of a number of European economies, remarkable differences emerge from cross-country comparisons.[1]

Most recently, a new strand of literature has emerged, proposing that a large share of the heterogeneity in productivity growth across countries (and within Europe in particular) could be attributed to the economic and political élites' capacity of managing a country (Caselli and Morelli 2004; Mattozzi and Merlo 2007). Along these lines of thinking, the élites' responsibilities, with respect to the institutional, social, and technological delays accumulated in the recent past, have become an issue in the European economic panorama.

In contrast to this literature, our claim in this paper is that the élites' responsibility does not exclusively derive from their simple tendency to maintain the status quo. It is also due to their inability to seize the opportunity offered by new technologies and to implement the best choice for the economy as a whole, which is a direct consequence of the obsolescence of their personal human capital. Indeed, as pointed out by Messner and Polborn (2004), many political or economic reforms resemble

[1] For example, OECD (2004) reports that, compared with the previous decade, hourly labor productivity picked up in a group of economies, including Norway, Portugal, Germany, Finland, and Sweden, while it remained stable or reduced in the others.

investment projects in their return streams; initially, there is a cost to be borne, but eventually there will be benefits. In this frame, young people will be able to enjoy the benefits longer and hence will be more inclined to favor reforms than older people. It then follows that, among individuals of different ages, the oldest ones will not be in favor of change because they suffer the costs without being able to reap much of the benefits. Therefore, we define a gerontocratic society as a place where the decision-making process and the political environment are dominated by the oldest individuals, with negative consequences on economic performance in periods of rapid change and instability, when innovation and flexibility are at a premium.

The inability of an older ruling class in managing innovation is therefore a key feature of our research. Existing literature on labor economics provides further support in favor of this idea. Several studies show that a negative link between size and productivity exists, and it is even more pronounced in the ICT sector (see Daveri and Maliranta 2007). Indeed, workforce ageing is known to entail skill deterioration and lessened ability to adapt and learn new things. One possible explanation relies on the tendency of cognitive abilities to deteriorate with age. Although this decline is not uniform across abilities, after a certain age threshold, further advancements in age are seemingly associated with lower productivity at work. Beyond that threshold, further increases of experience add little or nothing to the working ability of a given worker. There are no reasons to believe that power élites are excluded from this process.

Along this line of reasoning, our work is also related to the literature on interest group politics, where existing powerful interest groups may impede the introduction of new technologies in order to protect their economic rents (Acemoglu and Robinson 2006; Alesina and Rodrik 1994, Fernandez and Rodrik 1991). In these contributions, political élites block technological and institutional developments because of a political replacement effect. Innovations often erode élites' incumbency advantages, increasing the likelihood that they will be replaced. Fearing replacement, political élites are unwilling to initiate change and may even block economic development. Moreover, the theoretical model we develop belongs to the broad literature that studies the links between different political variables and economic growth (Bellettini et al. 2009, Hashimzade and Davis 2006, Hopenhayn and Muniagurria 1996, Krusell and Rios-Rull 1999, Krusell et al. 1997). In particular, Hashimzade and Davis (2006) provide an interesting example of how political uncertainty might impede economic growth. The main conclusion of their theoretical work is that an increase in political instability produces growth-reducing policies that lead governments to invest less in activities that support human capital accumulation. Along the same line of reasoning, through a simple model very close to the one developed by these authors, we argue that gerontocracy, involving an elder ruling class with a shorter life horizon, results in lower investments in human capital and in productive public services and thereby depresses economic development.

Our aim with this paper is to study whether the aggregate economic performance of a country can be negatively affected by the age of its political élite. We propose then a "toy" model in which we study how élite's patience influences the adoption of growth promoting policies. We conjecture that older élites are more impatient, i.e., have a lower subjective discount. The goal of the model is to highlight the mechanism that originates investments in education and productive government spending.[2] For the same reason, we do not allow for any other differences (e.g., ideologies, organization, etc.) between the élites' that alternate in the office. Although we are aware that such an approach could sound oversimplified, we are confident that this parsimonious conceptualization is suitable to isolate the effect that politicians' preferences (induced by their age) has per se on growth.

Through our simple model, we show that a more impatient ruling class, whose interests are less devoted to long-term delayed return on investment, may weaken the human capital accumulation process because of inadequate public education policies and may hinder private sector productivity growth because of poor expenditure in productive services. We conjecture that the term structure of élites subjective discount rates displays a decreasing pattern. In this sense, we argue that gerontocracy is harmful for growth. Using standard comparative statics analysis, we derive testable restrictions on the growth reduced form equation that we test in the empirical part of the paper. To measure the impact of politicians' age on economic growth, we combine information from a group of European countries on socio-economic characteristics and background of the political élites —which we identify with the parliamentarians— with information from a rich industry-level data set. Our main goal is to exploit differences in politicians' age across countries to estimate the

[2] This paper is not an attempt to explain the origin of a gerontocratic society. This implies that, for us, the (average) age of the ruling class is not an endogenous variable. The analysis of this phenomenon is on our research agenda but is out of the scope of this exercise. Here we conjecture that a link between gerontocracy and subjective discount exists and try to explain the effect of this on productivity growth. There are several ways to endogenize the politicians' age structure. In models of representative democracy, legislation and policy are responsive to the preferences of the median voter in the legislature (see, e.g., McCubbins, Noll, and Weingast 2007). Insofar, as the median representative that decides policy is representative of the median voter in the population at large, gerontocracy is perhaps the endogenous result of an ageing population.

effect that gerontocracy exerts on the allocation of public spending on productive investments and thus on economic growth.

The plan of the article is the following. Section II lays out the baseline model and discusses the links among the élites' patience, public investments and economic growth. Our main conclusion is that a gerontocracy, which is characterized by a higher impatience, is an important source of innovation-retarding policies and therefore depresses economic development. Therefore, it can be seen as plausible explanations of the growth differentials across countries. Section III discusses the data. The countries involved in our study are Denmark, Finland, France, Italy, Germany, Netherlands and the UK. Due to limitations on the availability of political data, we have not been able to extend the analysis to all EU countries. Our sample, however, represents a large share of the European economy and population. Section IV presents the empirical analysis, which corroborates our theoretical predictions while section V concludes the work.

II. Theoretical model

In this section, we present a simple theoretical model that extends the framework proposed by Hashimzade and Davis (2006) by taking into account the role of public productive services, along with the public investment in education, as engine of the human capital accumulation.

Demography. In a discrete-time $t \in \{0, 1, \ldots, \infty\}$ economy, a continuum of measure 1 of consumers/workers, who lives forever, produces a single homogenous good. Similar to Glomm and Ravikumar (1997), in each period agents allocate their time between education (e) and production ($1-e$).

Technology. Production function requires the use of human capital and government purchases and can be written as follows:

$$Y_t = AG_t^{\eta}\left[\left(1-e\right)H_t\right]^{1-\eta}, \tag{1}$$

where $A>0$ is the constant social marginal return of human capital, $(1-e)H_t$ is the stock of human capital at time t (i.e., efficiency of labor hour), G_t is the productive government spending (e.g., the provision of productive services, the roll-out and adoption of broadband, antitrust legislation, etc.) available at the beginning of period t, and $0<\eta<1$.[3]

[3] The public factor in equation (1) is a common external input, i.e., G is a pure public good.

Human capital accumulation is determined according to the following production function:

$$H_{t+1} = H_t + \phi(H_t, E_t), \tag{2}$$

where –without loss of generality– no depreciation is assumed. E_t is the public investment in education and ϕ is the learning technology described by the following homothetic function:

$$\phi(H_t, E_t) = e\zeta H_t^{\alpha} E_t^{1-\alpha}, \tag{3}$$

with $\zeta > 0$ and $0 < \alpha < 1$. Output is taxed at fixed rate τ. This implies that the following condition, representing the government budget constraint, must hold:

$$\tau Y_t = G_t + E_t + \left(R_t^g + R_t^r\right) = \sigma_{gt}\tau Y_t + \sigma_{et}\tau Y_t + \left(1 - \sigma_{gt} - \sigma_{et}\right)\tau Y_t, \quad \text{with} \left(\sigma_{gt} + \sigma_{et}\right) < 1 \;\forall t, \tag{4}$$

where σ_{gt} and σ_{et} are the share of revenues allocated to finance productive government spending, and public education, respectively. It then follows that the share $(1-\sigma_{gt}-\sigma_{et})$ is used to finance expenditure that produces no benefit for the community, and it can be seen as private benefit (or appropriation of tax revenues) enjoyed by the élites. We denote with R_t^g the *government rent*, enjoyed by the politicians in charge, and with R_t^r the *retirement rent*, received in the case of electoral loss. We assume that the retirement rent is constant and lower than R_t^g, with $R^r < R_t^g - R^r$. Finally, $C_t^p = (1-\tau)Y_t$ is consumed by the consumers/workers.

Political environment. We consider an environment where two parties randomly alternate in office. To keep matter simple, we assume that the two parties face the same exogenous probability π of being voted out and replaced. At each time t the government in charge chooses σ_{gt} and σ_{et}. At time zero, political élite knows their status $\epsilon_0 \in \{l, w\}$. When $\epsilon = l$ the incumbent government has lost the election. We assign at this event a positive probability π. In addition, with probability (1 − π), $\epsilon = w$, and the incumbent government remains in charge. In the former case ($\epsilon = l$), the political élite receives the retirement rent R^r, while in the latter ($\epsilon = w$) it allocates again tax revenues between productive government spending, public education, and its own (unproductive) rent.

A. The optimization process

The political élite, belonging to the "government in charge", maximizes:

$$\theta U\left(R_t^g\right)+(1-\theta)U\left(C_t^p\right), \tag{5}$$

where U is the strictly concave twice differentiable per-period utility,

$$R_t^g \equiv \left(1-\sigma_{gt}-\sigma_{et}\right)\tau Y_t - R^r \tag{6}$$

is the government rent, C_t^p is the private consumption, and θ is the weight the government assigns to government rent and private consumption. Therefore, θ can be interpreted as a measure of politicians' "selfishness" (i.e., the higher θ, the higher the degree of "selfishness"). Notice that, in this environment, the controls σ_g and σ_e at date t depend only on the current state H, so that

$$\sigma_{gt}=\sigma_g\left(H_t\right) \text{ and } \sigma_{et}=\sigma_e\left(H_t\right). \tag{7}$$

This implies that any given policy generates a stochastic law of motion for the state:

$$H_{t+1}=\Xi\left(H_t,\sigma_{gt},\sigma_{et}\right), \tag{8}$$

which will be stationary if σ_g and σ_e are stationary.

Following the standard notation used in the literature, we denote the variables at time t and t+1 as those without and with primes. The functional equation associated to the maximization problem faced by a government in charge at the beginning of period t is

$$V(H,\epsilon)=\max_{\{\sigma_e,\sigma_g\}_{t=0}^{\infty}}\left\{\left[\theta U\left(R^g\right)+(1-\theta)U\left(C^p\right)\right]+\beta\mathbb{E}\left[V\left(H',\epsilon'\right)\mid\epsilon\right]\right\}, \tag{9}$$

subject to

$$Y=AG^{\eta}\left[(1-e)H\right]^{1-\eta}, H_0>0,\ H'=\Xi\left(H,\sigma_g,\sigma_e,\epsilon\right), C=(1-\tau)Y,$$

$$R^g = \begin{cases} \left(1-\sigma_g-\sigma_e\right)\tau Y - R^r \ \ if \ \epsilon = w \\ \qquad\qquad\qquad\qquad R^r \ \ if \ \epsilon = l \end{cases}, \tag{10}$$

where β is the subjective discount factor ($\beta = \frac{1}{1+\rho}$, where ρ is the rate of time preference); H_0 is pre-determined at time $t = 0$, R_0^g and H_1 are chosen, and the uncertainty is due to the risk of an electoral loss in the subsequent period. It than follows that associated with the solution (9) is a *policy vector* defined by $\Psi = \left\{\left(\sigma_{g1}, \sigma_{e1}\right), \left(\sigma_{g2}, \sigma_{e2}\right), \ldots\right\}$. Notice that the value function (9) is the present discounted value of the incumbent ruling class evaluated along the optimal program.

The following assumptions are maintained for the remainder of this section.

Assumption 1. $H \in \mathcal{H} \subset \mathcal{R}$, $\left(\sigma_g + \sigma_e\right) \in (0,1)$ *and* $E, G \in \mathcal{A} \subset \mathcal{R}$.

Assumption 2. $U : X \to \mathcal{R}$ *is a strictly increasing, twice continuously differentiable and concave utility function, with* $U'(0) = \infty$ *and* $U'(\infty) = 0$.

Assumption 3. *Retirement rent* $R^r < \frac{1}{2} R_t^g$.

B. Equilibrium and comparative statics

Here we are interested in analyzing the long-run effects of gerontocracy. Therefore, we focus on the stationary equilibrium which involves time-invariant decision rules in the infinite horizon. This concept uses a recursive representation of the political élites' problem.

Definition 1. *Given the initial* H_0 and $H_t \in \Gamma\left(H_{t-1}\right) \subset \mathcal{H}$ *, with* Γ *continuous and compact-valued, a Balanced Growth Path (hereafter* BGP*) for the economy is a collection of sequences* $\left\{H, Y, C^p, R^g, \sigma_g, \sigma_e, G, E, e\right\}_{t=0}^{\infty}$ *such that:*

(i) H evolves according to (8)*;*

(ii) the government budget is balanced: $\tau Y_t = G_t + E_t + R_t^g + R^r$;

(iii) politicians solve problem (9–10).

Let now V_l denote the value of an electoral loss, which occurs with probability π, and V_w the value of being (re)elected, which occurs with probability $(1-\pi)$. Then, the optimal value function V for the political élites' optimization problem (9–10) is the solution to the following Bellman equation, subject to (10):

$$\max_{\{\sigma_e,\sigma_g\}_{t=0}^{\infty}} \theta U\left(R^g(H)\right)+(1-\theta)U\left(C^p(H)\right)+\beta\upsilon\left[\pi V_l(H')+(1-\pi)V_w(H')\right]. \tag{11}$$

With interior equilibrium, the first order conditions and the envelope condition for the political élites' problem are respectively:

$$[FOC] \qquad \frac{\partial V}{\partial \sigma_g}=0 \Rightarrow \frac{\partial U}{\partial \sigma_g}+\beta\left[\pi\frac{\partial V_l}{\partial H'}\frac{\partial H'}{\partial \sigma_g}+(1-\pi)\frac{\partial V_w}{\partial H'}\frac{\partial H'}{\partial \sigma_g}\right]=0\ , \tag{12}$$

$$[FOC] \qquad \frac{\partial V}{\partial \sigma_e}=0 \Rightarrow \frac{\partial U}{\partial \sigma_e}+\beta\left[\pi\frac{\partial V_l}{\partial H'}\frac{\partial H'}{\partial \sigma_e}+(1-\pi)\frac{\partial V_w}{\partial H'}\frac{\partial H'}{\partial \sigma_g}\right]=0, \tag{13}$$

$$[ENV] \qquad \frac{\partial V}{\partial H}=\frac{\partial Y}{\partial H}\left[\theta U'\left(R^g\right)\left(1-\sigma_g-\sigma_e\right)\tau+(1-\theta)U'\left(C^p\right)(1-\tau)\right]$$
$$+\beta\left[\pi\frac{\partial V_l}{\partial H'}\Xi'+(1-\pi)\frac{\partial V_w}{\partial H'}\Xi'\right] \tag{14}$$

Conditions (12)–(14), together with the transversality condition

$$\lim_{t\to\infty}(\beta)^t\frac{\partial U(\cdot)}{\partial H}H_t=0 \tag{15}$$

and the initial condition of the economy fully characterize the solution of the political élites' problem.

Finally, the assumption of identical governments implies that they choose the same optimal level of σ_e and σ_g, which is constant along the BGP, where all the per capita variables grow at the same rate given by

$$\gamma=\zeta e\left[A^{1/(1-\eta)}\ \ \sigma_e\sigma_g{}^{\eta/(1-\eta)}\ \ \tau(1-e)\right]^{1-\alpha}. \tag{16}$$

Simple algebra provides the following proposition.

Proposition 1. *Along the* BGP*, the growth rate of per capita variables is increasing in the amount of tax revenues used to finance education, and productive services:*

$$\left.\frac{\partial\gamma}{\partial\sigma_e}\right|_{BGP}>0 \text{ and } \left.\frac{\partial\gamma}{\partial\sigma_g}\right|_{BGP}>0.$$

Proof 1. *See the online appendix.*
Recalling that along BGP, $H' = H(1+\gamma)$, Proposition 1 also implies:

$$\frac{\partial H'}{\partial \sigma_e} = H\left(\frac{1-\alpha}{\sigma_e}\right)\gamma, \tag{17}$$

$$\frac{\partial H'}{\partial \sigma_g} = H\left(\frac{1-\alpha}{1-\eta}\frac{\eta}{\sigma_g}\right)\gamma. \tag{18}$$

Finally, in order to obtain explicit solutions for σ_e and σ_g, we assume that the politicians' preferences are logarithmic. Solving (12–14) with respect σ_g and σ_e yields:

$$\sigma_g^* = \eta \, \frac{\beta(1-\pi)(1-\alpha)}{\theta + \beta(1-\pi)\,(1-\alpha)}, \tag{19}$$

$$\sigma_e^* = (1-\eta)\frac{\beta(1-\pi)(1-\alpha)}{\theta + \beta\,(1-\pi)(1-\alpha)}. \tag{20}$$

Proposition 2. *Along the* BGP*, the optimal government spending in productive services* σ_g^**, and education* σ_e^* *declines with politicians' impatience. Thus, the more impatient is the political élite the lower is the equilibrium growth rate* γ*.*

Proof 2. *See the online appendix.*

Overall, the main task of our toy model is to isolate the optimizing behavior of the political élites. A political élite behaves as a single agent and solves an optimization problem over an infinite horizon. In order to be able to analyze our main question in a meaningful way, we first solve the élites' optimization problem. This allows to identify a link between the subjective discount rate of the cabinet in charge (β) and the policies implemented. Then, we add an aggregate technology that ensures perpetual growth, driven by productive government services, and investment in education. The provision of both government services and public education is financed by a tax on income, whose revenues are also used to finance the élites' unproductive rent R^g. As will be more clear in the following paragraph, this assumption is crucial to highlight the trade-off faced by the policy maker and the role of gerontocracy. Each rational government will choose the amount of tax revenues to invest in innovation and education that will guarantee a rent R^g as large as possible, under the uncertainty of being re-elected in the subsequent election.

The way we bring the toy model to the data is the following. We conjecture that the patience (which negatively affects the subjective discount rate) declines with age.[4] This implies that an older élite weights future returns less and, therefore, is the more reluctant to adopt potential growth enhancing policies. If this conjecture on the linkage between politicians' age and their discount factor is correct, then public investments do respond to changes in the ruling class age structure, which affect the size of the unproductive rent enjoyed by the élite. The empirical content of Proposition 2 is then that the older the political élite, the lower the public resources devoted to productive services and education, human capital accumulation declines and economic growth slows considerably.

III. The data

The data used in the empirical analysis have been collected from different sources. In the following, we provide a description of the data and discuss the procedures adopted to merge data from different sources into a single dataset.

The first source is the DataCube dataset, obtained from the EURELITE network, that contains information on personal characteristics of national parliamentarians in several European countries from 1983 to 2004.[5] DataCube includes about fifty variables related to the social and political background of national parliaments members. Unfortunately, this dataset does not provide any information on government' member ages. Therefore, in our empirical exercise, we proxy gerontocracy with the average age of the members of national parliaments of each country.[6] Beyond some basic socio-demographic variables (i.e., occupation, education, age, and sex), the dataset includes information on politicians' background, with particular attention to the pre-parliamentary political experience, including political and administrative

---

[4] Laboratory and field studies of time preference identify a decreasing slope in the term structure of subjective discount rates. The interested reader can refer, among the others, to : Angeletos et al. (2001), Laibson, (1997) and O'Donoghue and Rabin (1999).

[5] For more information on the EURELITE network see http://www.eurelite.uni-jena.de/eurelite/portrait/introduction.html.

[6] There may be other ways to proxy for gerontocracy. Given our data, we do not think that using the (average) age of the parliamentarians rather than that of the government members matters for our exercise. In fact, we did not find any argument to support the idea that, at the country level, there exists a significant difference in terms of age between these two groups of politicians. However, we are aware that, for this to be true, the members of the cabinet must be a randomly selected subset (in terms of age) of the Parliament.

appointments at the local level (town, county, and region), parliamentary career (i.e., age at entry into parliament and the number of successful elections), leading party functions, and government appointments.

The second source is represented by the EU-KLEMS dataset, which contains variables measuring output, productivity, employment (skilled and unskilled), physical capital, ICT investments, and technological change at the industry level, for all European Union member states from 1970 onwards.[7] At the lowest level of aggregation, data were collected for 71 industries. The industries are classified according to the European NACE revision 1 classification. Since the level of detail varies across countries, industries and variables, due to data limitations, we choose a level of aggregation that produces 25 industries, which for our purposes have been further grouped into 6 “macro” sectors (Manufacturing, Electrical machinery and telecommunication, Finance and business services, Retail and distribution services, Personal and social services, and Other goods producing industries).[8] The availability of data at the industry level is extremely important for our analysis, as we believe that the relationship between the level of gerontocracy, investments in ICT, and economic growth may be quite heterogeneous across the many sectors of the economy. Industry level data will then be able to capture such heterogeneity better than aggregate measures, such as the per capita GDP. EU-KLEMS also provides information on the so-called “non-market economy”. This aggregate includes data from the following sub-sectors: public administration, education, and health and social services. In our regressions, we proxy public ICT investment by the sum of the ICT investments undertaken by those sub-sectors. All variables and definitions are provide in Table 1.

As the number of countries covered and the time span of the EU-KLEMS are both larger than those available in the EURELITE dataset, the merging procedure of these two sources has produced a sample that includes 7 countries (Denmark, Finland, France, Germany, Italy, the Netherlands and the UK) and 25 industries, with a time span ranging from 1983 to 2004, for a total of 3,500 potential observations. However, as for some variables — like gross operating surplus — data are missing in the early years in some countries, the actual sample consists of 3,416 observations. Finally, we have added a variable accounting for public expenditure on education

[7] For more information on the EU-KLEMS dataset see http://www.euklems.net/.

[8] We decided to keep the electrical machinery and telecommunication sector separated from the aggregated manufacturing sector because we believe that in this sector the correlation between investment in ICT and TFP growth could be particularly relevant.

at the country level, as provided by EUROSTAT and UNESCO.[9] For our purposes, this variable has been standardized with respect to GDP. However, since we do not have information on the German public education expenditure before the pre-unification period, in our empirical analysis we split our sample in two sub-samples. The first sub-sample, comprising 2,916 observations, spans the whole period from 1984 up to 2004 and includes data from all country but Germany. The second sub-sample, comprising 1,485 observations, spans the sub-period from 1995 up to 2004, but includes data from Germany. Finally, we obtain a sub-sample of control,

Table 1. Data definitions and sources

Variables	Source
Gerontocracy related variables	
log(*gerontocracy*) = log of the politicians' mean age	EURELITE
log(*newcomers*) = log of the newcomers' mean age	EURELITE
background = % of politicians with local/national political backbround	EURELITE
female = % of female politicians	EURELITE
Growth accounting variables	
log(*tfp*) = log of TFP (value added based) growth (1995=100)	EU-KLEMS
log(*ict*) = log of ICT capital services (1995=100)	EU-KLEMS
log(*nict*) = log of non-ICT capital services (1995=100)	EU-KLEMS
log(*gict*) = log of non-market sector ICT capital services (1995=100)	Authors w/EU-KLEMS
log(*hhs*) = log of hours worked by high-skilled persons engaged (share in total hours)	EU-KLEMS
log(*hms*) = log of hours worked by medium-skilled persons engaged (share in total hours)	EU-KLEMS
log(*hls*) = log of hours worked by low-skilled persons engaged (share in total hours)	EU-KLEMS
log(h_{29}) = log of hours worked by persons engaged aged 15-29 (share in total hours)	EU-KLEMS
log(h_{49}) = log hours worked by persons engaged aged 29-49 (share in total hours)	EU-KLEMS
log(h_{+50}) = log of hours worked by persons engaged aged 50 and over (share in total hours)	EU-KLEMS
gos = Gross operating surplus (in millions of local currency)	EU-KLEMS
log(*marketopenness*)= log of exports plus imports as % of GDP	PWT 6.1
Education variables	
pexpedupe = public expenditure on education as a percentage of total public expenditure	EUROSTAT
pexpedugdp = public expenditure on education as a percentage of GDP	EUROSTAT

[9] Data source: http://appsso.eurostat.ec.europa.eu for the period 1995-2004 and http://www.uis.unesco.org/Education/Pages/default.aspx for the period 1983-1994.

Table 2. Summary statistics

Variable	Definition	Sample 1	1983:2004	Sample 2	1995:2004	Sample 3	1995:2004
		Mean	*Std. dev.*	*Mean*	*Std. dev.*	*Mean*	*Std. dev.*
background	% of politicians with local/regional political background	60.27	17.71	59.33	15.5	60.19	16.55
female	% of female politicians	20.54	12.11	24.72	11.12	23.8	11.78
gerontocracy	politicians' mean age	48.32	1.92	48.74	1.99	44.58	2.11
newcomers	newcomers' mean age	43.63	3.03	44.6	2.59	44.65	2.78
hhs	hours worked by high-skilled workers (% of total hours)	10.37	9.47	11.38	9.72	12.21	10.18
hms	hours worked by medium-skilled workers (% of total hours)	64.17	17.39	67.7	14.7	68.13	15.68
hls	hours worked by low-skilled workers (% of total hours)	25.45	16.52	21.02	13.01	19.66	13.4
h_{29}	hours worked by persons engaged aged 15-29 (% total hours)	27.63	7.82	24.44	7.26	24.92	7.6
h_{49}	hours worked by persons engaged aged 29-49 (% total hours)	53.99	7.83	55.58	7.27	55.28	7.61
h_{+50}	hours worked by persons engaged aged 50 and over (% total hours)	18.38	6.4	19.98	6.72	19.8	7
tfp	TFP (value added based) growth, 1995=100	100	57.1	104.84	17.76	104.63	17.21
gos	gross operating surplus	0.18	0.12	0.18	0.11	0.18	0.12
ict	ICT capital services, volume indices, 1995 = 100	117.59	105.52	191.08	109.9	197.89	114.84
nict	non ICT capital services, volume indices, 1995 = 100	98.41	18.39	109.22	16.21	110.08	16.63
gict	public ICT capital services, volume indices, 1995 = 100	119.93	110.3	208.28	102.82	213.92	107.57
tax	taxes (minus subsidies on production) over gross output	0.01	0.02	0.01	0.03	0.01	0.02
marketopenness	exports/gdp (constant US$)	0.34	0.13	0.33	0.13	0.37	0.14
$\log(pexpedu)_{gdp}$	public expenditure on education as a % of GDP	5.8	1	5.66	1.25	5.85	1.26
Observations		2916		1485		1269	

Note: Sample 1 refers to the period 1983-2004 and it includes only DNK, FIN, FRA, ITA, NLD and UK Sample 2 includes all countries, but it spans a shorter period from 1995 to 2004. Finally, Sample 3 contains data from all the countries included in Sample 1 but with respect the sub-period period 1995-2004.

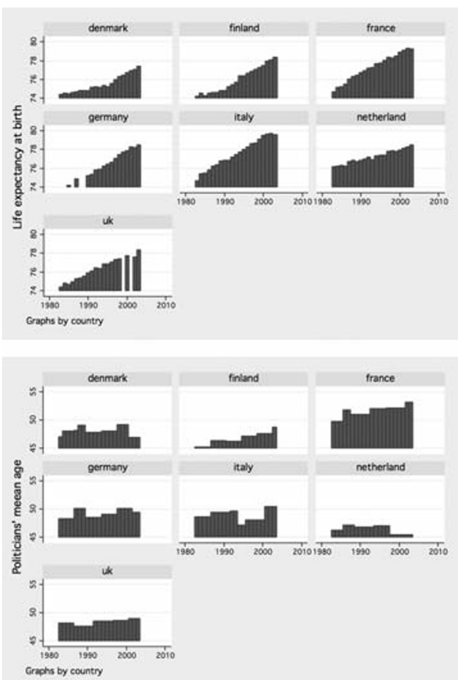
comprising 1,269 observations, that spans the sub-period from 1995 up to 2004, but does not include Germany.

Table 2 reports all summary statistics. According to our data, only 21% of the national representatives are female, and the average age is about 48 years. About 60% of parliamentarians in the sample has a university degree. Furthermore, about 60% of them had previous local/regional background activity in terms of being a representative elected by citizens, and about 60% have been elected in the same place of origin. Figure 1 shows a detailed picture of the cross-country differences in life expectancy and politicians' age in the sample, with France showing the oldest parliament and the Netherlands the youngest.

Concerning the economic data, we see that workers with average skills account for about two-thirds of total hours worked, followed by low skilled and high skilled workers. In particular, high skilled workers account for only 13% of total hours (for each sector, see Table A1 in the online appendix).

For a better understanding of our data and the relationships between them, we have also computed unconditional correlation coefficients between gerontocracy

Figure 1. Life expectancy at birth and politicians' mean age, our sample



and TFP growth and between ICT (both private and public) and TFP growth. In Table A2, the (*n*,*m*) cell shows the average correlation between the TFP growth of industry *n* and the level of gerontocracy attributed to country *m*. The general negative impact exerted by gerontocracy is quite transparent when looking at the last row of the table, which reports the correlation column average by country. In particular, this detrimental effect seems to be stronger and larger in the Electrical machinery and telecommunication sector. Notice that (on average) the older the politicians are, the larger the negative correlations.

As suggested by our theoretical model, a positive correlation between public ICT and TFP growth should emerge from the data, with the former positively affecting the latter and being complement with the private ICT. The unconditional correlation coefficients reported in Table A3 seem to confirm our theoretical predictions, with public ICT and TFP strongly correlated, and with public and private investments in ICT being complementary (see Figure A1, in the online appendix, where we plot the (log of) public and private ICT).

In Table A4 we observe the correlation between private ICT and TFP. Even in this case, the positive correlation seems to hold, and it is stronger in those sectors where we expect ICT to be a major driver for TFP.

Finally, a different picture emerges if we look at the correlation between the (log of the) age of the newcomers (which provides a measure of the demographic dimension of the political turnover) in each national Parliament and the TFP growth. The results reported in Table A5 suggest that the problem is not the politicians' age *sic et simpliciter*. In comparison with the previous table, correlations are much more tenuous and are often positive. A possible explanation could be that older newcomers, during their working life (presumably in the private sector), have acquired skills and competences that (partially or completely) compensate for the human capital obsolescence due to ageing.

IV. The empirical analysis

A. The econometric strategy

In this section, we present the econometric strategy used to test the main hypothesis of our theoretical model, namely that gerontocracy negatively affects economic growth due to its incapacity to provide sufficient investments in innovation (public and private) and education. However, as we lack adequate information on education expenditure, we limit the empirical analysis to the study of the effect of gerontocracy on innovation, thus assuming that the level of expenditure in education is given.[10] Therefore, our empirical model will be specified to estimate the impact of gerontocracy on public productive investments and, only indirectly, on TFP growth.

The primary econometric strategy discussed here is based on a reduced form three-equation system, while alternative approaches will be assessed in subsection C. The toy model is used to provide restrictions on the number of equations, the choice of the endogenous variables, and the set of regressors. First of all, since the model describes an economy where growth is driven by productive government spending (which we identified with the public investment in ICT) and public education, the TFP growth equation can be expressed as follows:

$$\begin{aligned}\log\left(tfp_{ijt}\right) = &+\alpha_1 \log(pexpedu_{jt-1}) + \alpha_2 \log(ict_{ijt-1}) + \alpha_3 \log(gict_{jt-1}) + \alpha_4 du95 \\ &+\alpha_5 du95 \bullet \log(ict_{ijt-1}) + \alpha_6 S_{ijt-1} + \alpha_7 X_{jt} + \eta_{ijt},\end{aligned} \tag{21}$$

where i is the sector, j is the country, and t is time; *tfp* is the TFP growth index; *pexpedu* is the public expenditure on education; *ict* is the private ICT capital service, while *gict* is the public ICT capital service.

Furthermore, S is a vector of sector-specific variables (share of labor input with different skills and share of workers with different age) and X is a vector of other controls at the country level, such as market openness and country dummies. Following the empirical evidence reported in van Ark et al. (2008) and Dahl et al. (2011), we include in our TFP equation the dummy variable (*du*95) and its interaction with *ict* to captures a structural break that could have changed the productivity trend from 1995 onward.

Since the toy model has showed that older politicians are more willing to pursue an (unproductive) rent rather than public investment, because of their shorter lifespan, and hence lower incentive to accumulate public capital, we added the following equation to link this kind of productive public spending with the set of gerontocracy-related variables:

[10] Unfortunately, homogeneous and comparable data on education expenditure at country level is available only in the aggregate, thus preventing us from distinguishing expenditures at different levels of education. In fact, we expect that expenditure at lower levels of education, although important for economic growth, may be positively related with gerontocracy and, in fact, could be aligned with vested interests of teacher unions for preserving a status quo where insiders obtain all the benefits, without caring about quality. On the contrary, the financing of higher education and research activities may be much less correlated with gerontocracy, as it usually leads to breakthroughs and innovations that are not in line with the idea of maintaining the status quo of a gerontocratic system. Based on simple descriptive statistics, our data do not show any correlation between gerontocracy and public expenditure on education.

$$\log\left(gict_{jt}\right)= \gamma_0+\gamma_1\log(pexpedu_{jt-1})+\gamma_2\log(gerontocracy_{jt-1})$$
$$+\gamma_3\log(newcomers_{jt-1)}+\gamma_4 background_{it-1}+\gamma_5\log(female_{jt-1})+\gamma_6 S_{ijt}+\gamma_7 X_{jt}+\xi_{ijt}\,, \tag{22}$$

where *gerontocracy* is the politicians' mean age, *newcomers* is the mean age of the politicians who are in office for the first time, *background* is the percentage of politicians with local/regional political background, and log is the percentage of female politicians.

Finally, the interaction between private ICT and public ICT is captured by:

$$\log\left(ict_{ijt}\right)=\beta_0+\beta_1\log(pexpedu_{jt-1})+\beta_2\log(gict_{jt})+\beta_3 S_{ijt}+\beta_4 X_{jt}+\varepsilon_{ijt}\,. \tag{23}$$

To avoid potential endogeneity problems between TFP growth and ICT variables, whenever reasonable, regressors have been lagged one period, while the potential endogeneity between ICT variables and gerontocracy has been controlled through the use of country dummies, which should wipe out all the time invariant unobserved heterogeneity at the country level.

Given our system of equations (21)–(23), we can easily see that gerontocracy affects private ICT only through public ICT (*gict*). At the same time, gerontocracy affects TFP through both private and public ICT. Therefore, the total effect of gerontocracy on TFP is given by the following relationship:

$$\frac{\partial TFP}{\partial gerontocracy}=\left(\alpha_3\gamma_2\right)+\left(\alpha_2+\alpha_5 du95\right)\beta_2\gamma_2\,,$$

where the first term on the right side of the equation reflects the (direct) effect of gerontocracy on TFP through the public ICT investment and the second term is the (indirect) effect through the private ICT investment.

As we assume a recursive structure for our empirical model, the parameters have been estimated using the SUR technique (Zellner 1962, Zellner and Huang 1962 and Zellner 1963). In the following, we present the results obtained by pooling all countries and sectors and later we discuss the results obtained fitting our model by sector or country.

B. Results

Here, we present and comment on the empirical results of our analysis. We first discuss the results obtained with the pooled data (all sectors and countries). Then, we introduce and compare the results by sector and country. Finally, we present some robustness checks that should help reinforce the conclusion of our study. All analyses have been carried out using the three different samples discussed in section III.

Estimates from the pooled data

Table 3 presents the estimates of the parameters in equations (21)–(23) for the pooled data, using the three samples. Overall, the results clearly corroborate our theoretical predictions, with the gerontocracy variable that negatively affects public ICT, that in turn affects TFP. This result is robust across sub-samples. Furthermore, coherently with our theoretical predictions, gerontocracy affects TFP through the public ICT investment channel. In fact, as can be seen in the top panel of Table 4, using the pooled data, a 1% increase in the level of gerontocracy reduces the TFP index by an amount ranging from 0.314% to 0.438%, depending on the sample employed. By disentangling the total effect into its direct and indirect components, we note that the direct effect is what really drives the result. Finally, by comparing the different samples, we notice in Table 3 that the negative effect of gerontocracy has decreased over time (by comparing the countries of sample 1 across the two periods) and it seems to have an important effect in Germany (by comparing samples 2 and 3 across the same period).

Consistently with the idea that the attitude to innovate declines with the politicians' age, from Table 3 we see that past experience of political government at the local/regional level (*background*) seems to be negatively related to *gict*. In particular, it negatively and significantly affects the TFP growth index in sub sample 2 and 3 (with elasticity equal to -0.435 and -0.855 respectively), i.e., when the role of public ICT capital is stronger, while its effect is positive but light (with elasticity 0.196) in sample 1, when the impact of *gict* on TFP is relatively smaller. This may be partly explained by thinking that being elected to national parliament can be seen as the culmination of a political career spent largely at the local or regional level. Under this perspective, *background* proxies politicians' age: therefore, the same argument used for *gerontocracy* can be applied to explain its effect on productive public spending.

Our estimates document that public ICT capital is a main determinant of the TFP growth index. The parameter of *gict* is positive and significant in each sample; it is definitely greater than that of the private ICT. In particular, the contribution of the private ICT is positive and not significant when the time horizon is longer

Table 3. SUR parameter estimates - pooled data

	Sample 1 - 1983:2004			Sample 2 - 1995:2004			Sample 3 - 1995:2004		
	log(*tfp*)	log(*ict*)	log(*gict*)	log(*tfp*)	log(*ict*)	log(*gict*)	log(*tfp*)	log(*ict*)	log(*gict*)
$\log(hhs)_{t-1}$	0.0123	0.0317**	0.0524***	0.00398	0.0109	0.00646	0.0153	-0.012	0.0033
$\log(hms)_{t-1}$	0.135***	0.165***	0.135***	0.0856**	0.0134	0.0378*	0.0544	0.0812	-0.000403
$\log(hls)_{t-1}$	-0.0112	0.0739***	0.131***	0.0264**	0.0117	-0.0055	0.0294**	0.00366	0.00756
$\log(h_{29})_{t-1}$	0.1393***	0.196***	-0.167***	0.0975***	0.206***	0.0151	0.112***	0.162**	-0.00295
$\log(h_{49})_{t-1}$	0.2074***	0.311***	-0.313***	0.371***	0.310**	-0.0159	0.356***	0.320**	0.00908
$\log(h_{50+})_{t-1}$	0.0784***	0.0302	-0.126***	0.0221	0.122**	0.0469***	0.0399	0.093	0.00338
$\log(gerontocracy)_{t-1}$			-4.5578***			-5.0014***			-3.294***
$\log(newcomers)_{t-1}$			0.9068***			0.6866***			0.386***
$\log(background)_{t-1}$			0.1963***			-0.4346***			-0.855***
$\log(female)_{t-1}$			0.0123			0.0290*			0.0275*
$pexpedu_{t-1}$	-0.0211**	0.0229	0.184***	-0.0633***	-0.0692*	-0.004	-0.0623***	-0.0786**	0.00332
$\log(ict)_{t-1}$	0.0104			-0.0720***			-0.0580***		
$du95_{t-1}$	0.0178***								
$du95 \cdot \log(ict)_{t-1}$	-0.0083**								
$\log(nict)_{t-1}$	-0.0981***	0.287***		-0.0744**	0.702***		-0.0452	0.601***	
$\log(gict)_{t-1}$	0.0783***	0.606***		0.196***	0.647***		0.183***	0.714***	
gos_{t-1}	0.2194***	0.169**		0.287***	-0.0969		0.307***	-0.190*	
$\log(marketopeness)_{t-1}$	0.2020***			0.0623			-0.0511		
trend		0.0468***	0.118***		0.0309***	0.110***		0.0217**	0.121***
constant	2.9171***	-95.96***	-218.8***	2.437***	-64.89***	-196.7***	2.200***	-46.41**	-221.5***
Observations	2,803	2,803	2,803	1,336	1,336	1,336	1,144	1,144	1,144
R - squared	0.150	0.831	0.929	0.161	0.685	0.967	0.156	0.685	0.979

Note: Country dummies included in all estimates. Sample 1 includes DNK, FIN, FRA, ITA, NLD and UK from 1983 to 2004. In sample 2 we add GER but limit the time period from 1995 to 2004. Sample 3 includes countries of Sample 1 but spans from 1995 to 2004. *** Indicates significance at the 1% level ** significance at 5%, * significance at 10%.

Table 4. Elasticities of TFP growth with respect to gerontocracy: pooled data and by sector

Sample	Direct effect via *gict*	Indirect effect via *ict*	Total effect
Pooled data (2,803 obs., 1,144 obs., 1,336 obs.)			
Sample 1 -1983:2004	-0.481***	0.043	-0.438***
Sample 3 -1995:2004	-0.341***	-0.008	-0.349***
Sample 2 -1995:2004	-0.313***	-0.002	-0.314***
Electrical machinery and TC (249 obs., 96 obs., 112 obs.)			
Sample 1 -1983:2004	-0.685***	0.026	-0.659***
Sample 3 -1995:2004	-0.217	-0.101	-0.116
Sample 2 -1995:2004	-0.483***	0	-0.483***
Manufacturing (1,290 obs., 480 obs., 560 obs.)			
Sample 1 -1983:2004	-0.548***	0.031	-0.516***
Sample 3 -1995:2004	-0.14	0.006	-0.134
Sample 2 -1995:2004	-0.094	0.001	-0.093
Finance and business services (238 obs., 88 obs., 104 obs.)			
Sample 1 -1983:2004	-0.826***	0.495***	-0.330**
Sample 3 -1995:2004	-0.113	0.021	-0.092
Sample 2 -1995:2004	-0.147*	0.008	-0.137*
Retail services (468 obs., 192 obs., 224 obs.)			
Sample 1 -1983:2004	-0.940***	0.269***	-0.671***
Sample 3 -1995:2004	-0.200*	0.003	-0.197*
Sample 2 -1995:2004	-0.218**	-0.075	-0.294**
Personal and social service (258 obs., 96 obs., 112 obs.)			
Sample 1 -1983:2004	-0.133*	-0.012	-0.145**
Sample 3 -1995:2004	-0.253**	-0.088	-0.341**
Sample 2 -1995:2004	-0.592***	-0.119	-0.711***

Note: Sample 1 includes DNK, FIN, FRA, ITA, NLD and UK from 1983 to 2004. In Sample 2 we add GER but limit the time period from 1995 to 2004. Sample 3 includes countries of Sample 1 but spans from 1995 to 2004. *** Indicates significance at the 1% level ** significance at 5%, * significance at 10%.

(i.e., sample 1), while it is negative and significant when we focus on the last decade of our dataset, irrespective of the presence of Germany in the dataset. This result is consistent with the literature on TFP growth in the European countries. In fact, along a time span similar to the one taken into account in the present analysis, van Ark et al. (2008) show that the effect of private ICT on TFP growth for the continental European countries is zero up to the mid-1980s, significantly negative during 1991-1996 and zero after that, leading the authors to conclude that ICT has at best had no effect on TFP index.

Estimates document the substitutability between ICT (public and private) capital and non-ICT capital (*nict*), which enters into the TFP equation with a negative (and also significant) parameter in all the (two of the three) samples employed. Furthermore, they show that, over the whole period, the TFP growth index increases with the share of medium skilled workers (*hhms*), while all employees contribute to the investment in private and public ICT (with some differences). On the contrary, when we consider the shorter samples, highly skilled workers never play a role.

Similarly to what happens in the political arena, our estimates suggest that age affects the contribution of the workforce (i.e., the labor productivity) to the TFP and private ICT, given that the parameter associated to lower ages (h_{29} and h_{49}) is generally greater than the one associated with h_{+50}. The worker age does not seem to have an effect on the public ICT equation.

Finally, looking at education (the second channel through which gerontocracy may affect economic performance according to our theoretical model), our results do not support the idea that public expenditure on education –whose limits we have previously described– unambiguously enhances TFP. Regressions ran with alternative aggregate measures (i.e., the share of the total public expenditure, TPE) confirm that, regardless of the proxy employed, the final impact of *pexpedu* on TFP growth is inconclusive.

Estimates using data by sector and countries

The results presented so far, although interesting, provide only an aggregate picture of the relationship between gerontocracy, ICT, and TFP. However, we know that it can be highly heterogeneous across the many sectors of the economy and/or by country. As already discussed in the previous sections, some of the relationships between ICT and TFP may be stronger or weaker depending on the specific sector/country where they apply. Therefore, in what follows, we first present elasticity results obtained by splitting our pooled samples by sector, and later we comment on the results by country.[11]

Table 4 provides the elasticities of TFP growth with respect to *gerontocracy* by sector. Estimates show that *gerontocracy* mainly hampers ICT intensive sectors, such as "Electrical machinery and telecommunication" and "Manufacturing", and the "Retail services" (with significant elasticities in the range from -0.516 up to -0.671, in sample 1). Moreover, in contrast with the other sectors, the impact on

[11] The full set of parameter estimates by sector and country are available upon request from the authors.

"Personal and Social services" has grown along the period under observation (with elasticities moving from -0.145 in sub sample to -0.711 in sub sample 2).

We don't find evidence of the indirect effect "via private ICT", which is almost always non-significant or positive.

Finally, elasticities computed by country, reported in Table 5, show that the loss in terms of TFP growth has been particularly relevant in the UK (-1.611) and Italy (-4.160), and dramatic in Germany and Finland, where the estimated elasticities of TFP growth with respect to our measure of gerontocracy have been greater than 14% and 17%, respectively.[12]

Table 5. Elasticities of TFP growth with respect to Gerontocracy: by country

Country	Sample	Period	Obs.	Direct effect via *gict*	Indirect effect via *ict*	Total effect
Denmark	Sample 1	1983:2004	480	0.254	0.079	0.333
Finland	Sample 1	1983:2004	480	-15.026***	-2.501	-17.528***
France	Sample 1	1983:2004	480	0.038	-0.02	0.172
Germany[a]	Sample 2	1995:2004	192	-12.273**	-2.26	-14.533**
Italy[b]	Sample 1	1983:2004	415	-4.160***	0	-4.160***
Netherland	Sample 1	1983:2004	468	-1.312***	0.919***	-0.393
UK	Sample 1	1983:2004	480	-1.646***	0.035	-1.611***

Notes: ab Due to constancy over time, some variables have not been included as controls in the gict equation for Germany and Italy: therefore, they are slightly different from those of other countries. *** Indicates significance at the 1% level, ** significance at 5%, * significance at 10%.

C. Robustness checks

In order to check the robustness of our results to different model specifications, in this section, we briefly present all the alternatives we have estimated and compare the results with the baseline specification presented in the previous section.[13]

Our first robustness check has been devoted to analyzing the role of gerontocracy variables as regressors in the private ICT equation (23). In fact, although according to our theoretical model the set of gerontocracy-related variables should not affect

[12] These results must be interpreted with caution since, in contrast to pooled and sector estimates, country estimates are not robust to changes in the definition of the sample.

[13] As our results are robust to the alternative specifications used, for sake of brevity we do not present and discuss in detail all the parameter estimates. However, they are available upon request from the authors.

private ICT, we have run a model specification that includes them. Results have shown that these variables are never statistically significant, and, in any case, the magnitude of the parameter estimates has always been very low across samples, sectors, and countries.

We have checked if alternative specifications, involving gerontocracy variable interactions and politicians' background variables could have had an effect on the overall results. According to our results, adding these interactions produces slightly less accurate estimates, but the main results do not change significantly with respect to those reported in the previous section. This effect has been noticed in particular in the estimates by sector and by country, and in our view this should simply reflect a problem of efficiency (due to small sample size in presence of an increased number of parameters to be estimated).

As a further robustness check, we have also estimated a model in which the lagged logarithm of private ICT enters as regressor in the *gict* equation. While the overall results and economic conclusions do not change, it is interesting to note that with this new specification there is a strong feedback effect between *ict* and *gict*, self reinforcing each other. No change is observed in terms of a gerontocracy effect on TFP.

We have also adjusted *gerontocracy* and *newcomers* for country specific life expectancies, in order to account for different interpretations of the politicians' age according to country specific social norms imposed by different countries average ages. Actually, these are further ways to control for endogeneity. All results are fully confirmed in terms of sign, magnitude and significance. Alternatively, we add the (log of the) population's median age to the set of regressors in equation (22). The results are in line with those obtained in the baseline specification, and the elasticity of the TFP growth with respect to gerontocracy varies between -0.397 and -0.382. We also run regressions replacing our measure of gerontocracy with the population median age. Even in this case, the signs and the magnitudes of the elasticities are consistent with the theory, and the elasticity of the TFP growth with respect to this other proxy for gerontocracy is equal to -0.429 in sample 1, -0.919 in sample 2, and -0.462 in sample 3 (see Table A6 in the online appendix).[14]

We attempted to control for the cabinet's political orientation by adding a dummy variable, which is equal to 1 if the Prime Minister belongs to a conservative party (e.g., the Tories in UK, Democrazia Cristiana or Forza Italia in Italy, Christlich Demokratische Union in Germany, Rassemblement pour la République in France,

[14] Data for life expectancy and population median age are obtained by EUROSTAT.

etc.) and 0 otherwise. The effect of this dummy on the TFP growth is slightly negative (i.e., a right-wing government is correlated to a decrease of 0.007% of the TFP index) in sample 1 and slightly positive in the other two samples (0.009% and 0.005%, respectively) while the elasticities of TFP with respect to gerontocracy do not change significantly (see Table A7 in the online appendix).[15]

Moreover, as our education variable does not produce convincing results, we have estimated our model using a measure of education expenditure obtained as ratio to TPE rather than to GDP. Even in this case, education appears to affect private and public ICT not in an unambiguous way, while the results in terms of gerontocracy remain in line with those presented in the previous section.

V. Conclusions

In this paper, we argue that when relatively young people cease to be the engine of an economy, long-run economic growth is endangered. Over the last three decades, many European economies have fallen into an old-age trap, a self-reinforcing mechanism whereby élites, generally the most aged individuals, have used control of the political system to exclude new generations, who are reasonably the most dynamic and innovative part of the population, from the access to power.

While we do not analyze this mechanism formally (i.e., we do not explain what the determinants of gerontocracy are), nor we do focus on some possible "positive" consequences that gerontocracy may have on a society as a whole, for example in reducing inequalities, we focus our effort on exploring the possible linkages between the age of the ruling class and the long-run growth rates both theoretically and empirically.

To achieve this goal, we have developed a simple endogenous growth model where the long-run growth rate is directly affected by public productive services and public investment in education. Moving from the conjecture that an older élites displays a higher impatience rate, the main testable hypothesis derived from our theoretical model is that the older the ruling class, the lower the public investment in education and productive services.

---

[15] This result is rather obvious. In fact, our sample includes six high-income liberal democracies along the period 1983-2003. These countries are quite "homogenous" with respect to the characteristics of the political environment, i.e., average electoral volatility, party closeness, etc. Our estimates show that in these countries the swing of the *political pendulum* (left to right and vice versa) does not matter a lot. On the contrary, it may matter in the comparison across countries. In our empirical analysis, however, this feature is already captured by the country dummies.

The empirical analysis corroborates these findings. Estimates indicate that, on average, a decrease of gerontocracy unambiguously increases TFP, with elasticities ranging between -0.314% and -0.438%, depending on the sample employed. Furthermore, we find gerontocracy affects TFP via *gict*, i.e., what we called the "direct effect" is always negative, and this result holds using both the pooled data or the data by country and sector. In addition, the negative effect of *gerontocracy* on TFP growth is stronger in those sectors, such as electrical machinery and telecommunication, retail service and manufacturing, where ICT is expected to be essential. Our estimates indicate that the consequences of gerontocracy have been more severe in Germany, Finland, and Italy compared to the other European countries included in our sample. These results are robust to different alternative model specifications and estimators, although sometimes the magnitudes of the effect have changed.

Finally, in terms of our future agenda, there are several extensions to our approach that are worth pursuing. In the theoretical model, for instance, we introduce several assumptions aimed at obtaining an analytically friendly framework. The next step will be to test how robust these results are when these simplifications are relaxed. In particular, we plan to address in a subsequent work the formal attempt to endogenize the gerontocracy. Moreover, from an empirical standpoint, we delegate to a further paper the extension of our data set in order to include information on the managers employed in the private sector.

References

Acemoglu, Daron, and James A. Robinson (2006). Economic Backwardness in Political Perspective. *American Political Science Review*, 100: 115-131.

Alesina, Alberto, and Dani Rodrik (1994). Distributive Politics and Economic Growth. *Quarterly Journal of Economics* 59: 465-90.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman and Stephen Weinberg (2001). The hyperbolic consumption model: Calibration, simulation and empirical valuation. *Journal of Economic Perspective* 15: 47–68.

Barro, Robert (1990). Government spending a simple model of endogenous growth. *Journal of Political Economy* 98: 103-125.

Barro, Robert, and Xavier Sala-i-Martin (1992). Public finance in models of economic growth, *Review of Economic Studies* 59: 645-661.

Barro, Robert, and Xavier Sala-i-Martin (1995). *Economic Growth*, The MIT Press, Cambridge, Massachusetts.

Bassanini, Andrea, and Roamin Duval (2009). Unemployment, institutions, and reform complementarities: re-assessing the aggregate evidence for OECD countries. *Oxford Review of Economic Policy* 25: 40-59.

Bellettini, G., Ceroni, C., Prarolo, G., (2009). Political Persistence, Connections and Economic Growth. CESifo Working Paper Series 2553, CESifo Group Munich.

Blanchard, Olivier (2004). The Economic Future of Europe. *Journal of Economic Perspectives* 18: 3-26.

Brunetti, Aymo (1997). Political variables in cross-country growth analysis. *Journal of Economic Surveys* 11: 163-190.

Carboni, Carlo (2007). *Elite e Classi dirigenti in Italia*, Editori Laterza.

Caselli, Francesco, and Massimo Morelli (2004). Bad politicians. *Journal of Public Economcs* 88: 759-782.

Chari, V.V., and Hugo A. Hopenhayn (1991). Vintage Human Capital, Growth, and the Diffusion of New Technology. *The Journal of Political Economy* 99: 1142-1165.

Dahl, Christian L., Hans C. Kongsted, and Anders Sørensen (2011). ICT and productivity growth in the 1990s: panel data evidence on Europe. *Empirical Economics* 40:141-164.

Daveri, Francesco, and Mika Maliranta (2007). Age, Seniority and Labor Costs; lessons from the finnish IT revolution. *Economic Policy* 49:117-175.

Fernandez, Raquel, and Dani Rodrik (1991). Resistance to reform: status quo bias in presence of individual specific uncertainty. *American Economic Review* 81:1146-55.

Gordon, Robert J. (2004). Why was Europe left at the station when America's productivity locomotive departed? Working Paper 10661. Cambridge, MA, NBER.

Glomm, Gerhard, and B. Ravikumar (1997). Productive government expenditure and long-run growth. *Journal of Economics Dynamics and Control* 21: 183-2004.

Hashimzade, Nigar, and George Davis (2006). Human capital and growth under political uncertainty. *Economics Bulletin* 15: 1-9.

Hopenhayn, Hugo A., and Maria E. Muniagurria (1996). Policy variability and economic growth. *Review of Economic Studies* 63: 611-625.

Kaas, Leo (2003). Governernment spending, growth and sequential voting. *European Journal of Political Economy* 19:227-246.

Krusell, Per, Jose-Victor Rios-Rull (1999). On the size of U.S. government: political economy in the neoclassical growth model. *American Economic Review* 89:1156-1181.

Krusell, Per, Vincenzo Quadrini, and Jose-Victor Rios-Rull (1997). Politico-economic equilibrium and economic growth. *Journal of Economic Dynamics and Control* 21: 243-272.

Laibson, David (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Econ*omics 112: 443-497.

Mattozzi, Andrea, and Antonio Merlo (2007). Mediocracy. Working Paper 12920. Cambridge, MA, NBER.

McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast (2007). The Political-Economy of Law, in A. Mitchell Polinsky and Stephen Shavel, editors, *Law and Economics Handbook*, 2, Elsevier Press 2007.

Messner, Matthias, and Mattias K. Polborn (2004). Voting on Majority Rules. *Review of Economic Studies* 71: 115-132.

Nicoletti, Giuseppe, Stefano Scarpetta, and Olivier Boylaud (2000). Summary Indicators of Product Market Regulation with an Extension to Employment Protection Legislation. Working Papers 226. OECD Economics Department, OECD, Economics Department.

O'Donoghue, Ted, and Matthew Rabin (1999). Doing it now or later. *American Economic Review* 89: 103-124.

OECD, 2004, *The Sources of Economic Growth in OECD Countries*, Paris, France.

Tabaka Feliz, and Steve H. Barr (1999). Propensity to adopt technological innovations: the impact of personal characteristics and organizational context. *Journal of Engineering and Technology Management* 16: 247-270.

Persson, T., and Tabellini, G., 1994, Is Inequality Harmful for Growth? *American Economic Review* 84: 600-621.

van Ark, Bart, Mary O'Mahoney, and Marcel P. Timmer (2008). The Productivity Gap between Europe and the United States: Trends and Causes. *Journal of Economic Perspectives* 22: 25-44.

Zellner, Arnold (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57: 348-368.

Zellner, Arnold, and David S. Huang (1962). Further properties of efficient estimators for seemingly unrelated regression equations. *International Economic Review* 3: 300-313.

Zellner, Arnold (1963). Estimators for seemingly unrelated regression equations: some exact finite sample results. *Journal of the American Statistical Association* 58: 977-992.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 169-192

A DEA-LOGISTICS PERFORMANCE INDEX

Luisa Martí
Universidad Politécnica de Valencia

Juan Carlos Martín
Universidad de Las Palmas de Gran Canaria

Rosa Puertas*
Universidad Politécnica de Valencia

Submitted October 2015; accepted December 2016

Logistics and transport increasingly play a pivotal role in international trade relations. The Logistics Performance Index (LPI) measures the on-the-ground efficiency of trade supply chains or logistics performance. The aim of this paper is to propose a data envelopment analysis (DEA) approach to compute a synthetic index of overall logistics performance (DEA-LPI) and benchmark the logistics performance of the countries with LPI. Dealing with the six dimensions of LPI, the proposed approach uses DEA as a tool for multiple criteria decision making (MCDM). Furthermore, the paper also analyses the potential differences observed when using different variables, namely income and geographical area. Our findings suggest that the logistics performance depends largely on income and geographical area. High income countries are in the group of best performers, which is highly dominated by the EU.

JEL classification codes: C5, F1, O52, R4
Key words: logistics performance, freight transport, data envelopment analysis, logistics performance index

---

* Rosa Puertas (corresponding author): Universidad Politécnica de Valencia, Grupo de Economía Internacional, 46022 Valencia, Spain; email: rpuertas@esp.upv.es. Luisa Martí: Universidad Politécnica de Valencia, Grupo de Economía Internacional, 46022 Valencia, Spain; email: mlmarti@esp.upv.es. Juan Carlos Martín: Universidad de Las Palmas de Gran Canaria, Institute of Tourism and Sustainable Economic Development, 35017 Las Palmas de Gran Canaria, Spain: email: jcarlos.martin@ulpgc.es. We thank the editor and two anonymous referees for insightful comments and suggestions.

I. Introduction

International trade has recently been affected by globalization and increased competitiveness of lagged regions that in the past did not play such an important role in the world. This trend has had an extreme impact on logistics as one of the key elements facilitating the mobility of products, ensuring their safety and speed as well as providing cost reductions when international trade among countries is growing. De Souza et al. (2007) define logistics as part of the value chain that plans, implements and controls the efficient flow of goods, services and information from the source to the consumer. The importance of the key components of logistics –transport, inventory and warehousing– has been recognized in the last 20 years although these elements have been fundamental in the industrial and economic life of nations for countless years (Rushton et al. 2014).

In late 2013, after more than 10 years of negotiation, the World Trade Organization (WTO) approved the *Agreement on Trade Facilitation*, providing crucial guidance on trade policies. This Ministerial Declaration deals with three key issues: trade facilitation, agriculture, and commercial impetus to help developing countries mitigate existing differences. Specifically, it contains provisions to speed up and improve the efficiency of customs procedures and border management (Sanz 2014). However, as outlined by Arvis et al. (2014), it introduces only minimum common standards, and in no way guarantees success. Only if countries are truly prepared to implement the advances that commercial globalization requires, can they benefit from the advantages of improved logistics performance. Hence, a suitable quantitative instrument is clearly needed in order to measure and compare the role of logistics in different parts of the world.

The *Logistics Performance Index* (LPI) was established with a view to bridging this gap. Its main objective is to measure the efficiency of logistics supply chains based on survey feedback from export companies. The LPI was first published in 2007 and led to a global debate on the importance of logistics in world economic growth. At the same time it revealed the need to implement concrete policies to improve future performance. By comparing the results obtained by the LPI for the four years analysed, the enormous value of the trade facilitation policies (i.e. the international distribution of production) can be appreciated. This index and its components can help countries (governments and corporations) to get to know their business partners more closely and anticipate any possible adjustments that could harm their competitiveness.

LPI is sometimes compared to other indicators such as the *Doing Business* ranking. However they differ in a number of respects, and so are not interchangeable.

Specifically, the *Doing Business* ranking makes use of data on regulations that are "on the books", while the LPI draws on surveys of logistics professionals who answer questions about their experiences in different countries. In this way it seeks to capture the day-to-day reality facing the private sector much more accurately. Moreover, the *Global Competitiveness Index* published by the World Economic Forum measures the ability of countries to provide high levels of prosperity to their citizens based on 12 pillars, and therefore only in the area of quality of transport infrastructure could it be considered comparable with one of the LPI components.

A synthetic index has been accepted as a useful tool for performance comparisons, benchmarking, policy analysis and public communication in various fields such as economy, environment and society. In this context, the paper proposes this type of index to measure the logistics performance of the countries. According to the OECD, 'a composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model of the multidimensional concept that is being measured'.[1] It is a mathematical aggregation of a set of sub-indicators for measuring multi-dimension concepts that cannot be captured by a single indicator (OECD 2008).The literature includes research on synthetic indexes in technological and social capabilities of countries (Mahlberg and Obersteiner 2001; Osberg and Sharpe 2002; Filippetti and Peyrache 2011). Nevertheless, the literature on countries' logistics performance is certainly scant with the exception of Markovits-Somogyi and Bokor (2014). Thus, this paper contributes with a second empirical application to this understudied research field providing a new synthetic logistics performance index based on DEA and using the LPI database.

In this study, we propose another objective measure for countries´ logistics. Therefore, the aim of this paper is fourfold: (1) to compute a synthetic index of overall logistics performance (DEA-LPI) using a DEA method that could be used to benchmark the logistics performance of the countries; dealing with the six dimensions of LPI, the proposed approach uses DEA as a tool for multiple criteria decision making (MCDM) under three different scenarios considering a different selection of inputs and outputs; (2) since DEA measures the relative efficiency of Decision Making Units (DMUs), in our case, one hundred forty one countries in the sample, a set of corresponding efficient DMUs called a reference set will be identified; this group of countries can be used as benchmarks for improvement of inefficient DMUs, providing clear guidelines for benchmarking of national

[1] OECD, *Glossary of Statistical Te*rms.

logistics performance; (3) a comparison of the results with the different methods will be performed analysing to what extent there exists a positive association between all the methods; furthermore, (4) the paper will also analyse the potential differences observed using income and geographical area as determinants of logistics performance.

The remainder of the paper is organized as follows: Section II offers some insights from the literature, Section III details the methodology, Section IV describes the data section, Section V presents and discusses the results, and Section VI concludes.

II. Literature review

Major trade reforms have been successfully implemented all over the world. These range from trade liberalization policies that have fostered bilateral and multilateral international trade agreements to more ambitious international integration treaties. Nevertheless these reforms do not always compensate the myriad of non-tariff barriers that are often more important for trade than actual tariffs. As Blyde and Iberty (2012) contended, even though the underlying theoretical models do not make a distinction between developed and developing countries when it comes to the predictions of decreasing trade costs, it is possible that the effects could differ for various reasons. For instance, developing countries tend to exhibit additional non-tariff barriers associated with traded goods, like higher transport costs, less efficient port infrastructures or more cumbersome custom procedures, than developed countries. The existence of such additional barriers might detain the full effects of a trade liberalization process.

Trade facilitation measures have been a central issue in the WTO negotiations on trade facilitation since 2005 when the OECD Trade Committee analysed the costs of introducing and implementing trade facilitation measures, based on the experience of fifteen developing countries. The WTO defines trade facilitation measures as: 'the simplification and harmonization of international trade procedures, including the activities, practices and formalities involved in collecting, presenting, communicating and processing data and other information required for the movement of goods in international trade'. Wilson et al. (2005) define trade facilitation using four indicators: port efficiency, customs, regulations and use of e-commerce. Soloaga et al. (2006) apply the same definition to analyse the impact of changes in the trade facilitation of Mexican industrial goods flows, suggesting that trade reform could boost total Mexican exports by 22.4%.

Möisé (2013) analyses the three main areas of trade facilitation measures –transparency and predictability; procedural simplification and streamlining; and coordination and cooperation between border agencies– finding that equipment and infrastructure seem to be the most expensive elements of trade facilitation, in particular the introduction and use of information technologies and the establishment of single window mechanisms. However, countries themselves reported that the most important area was training, given its fundamental role in bringing about sustained change in the business practices of border agencies.

Notwithstanding, other studies have proposed a sole indicator to estimate trade facilitation and ascertain its impact on exports (UNDP 2001, OECD 2003, Dennis 2006, Decreux and Fontagne 2006). In the same vein, Behar and Manners (2008) and Puertas et al. (2014) use the LPI published by the World Bank to explore the relationships that exist between bilateral exports and logistics. Hoekman and Nicita (2011) and Korinek and Sourdin (2011) include the LPI using a gravity equation for exports as an indicator of trade costs, together with others such as Doing Business Costs, concluding that domestic costs are quantitatively important and that the LPI has the largest effect on trade.

Many organizations such as the United Nations, the European Commission, and the OECD have developed and used an ample panoply of composite indicators (CIs) in different areas such as energy, environment, logistics, and quality of life, among others, in which sub-indicators are transformed mathematically into one synthetic indicator, with a view to provide comparisons of countries in complex policy issues. These measures are gaining more acceptance as a tool for policy making and, especially, benchmarking analysis on countries' relative performance (Cherchye et al. 2008).

The construction methodology that is used in the present study is based on Data Envelopment Analysis (DEA). DEA was originally designed to measure the performance of a firm on a context of production economics. This methodology was initially proposed by Charnes et al. (1978) to evaluate the performance of different DMUs —a set of some decision-making units. The authors described the DEA methodology as a mathematical programming model applied to observed data that provides a new way of obtaining empirical estimates of extremal relationships such as the production functions and/or efficiency production possibility surfaces that are the cornerstones of modern economics. This was the origin of a discipline that deals with how one could measure each decision-making unit's relative efficiency, given observations on input and output quantities in a sample of peers (Charnes and Cooper 1985). Mathematically, we will see below that DEA is a linear programming-based methodology whose main advantage is that it does not require any assumption on the shape of the frontier surface.

A well-known feature of DEA is that it looks for endogenous weights that can be constrained, which maximize the overall score for each DMU given a set of other observations. For this reason, it gained its acceptance in real policy-related settings in different fields, such as education, health care, banking, armed forces, sports, transportation, agriculture, and electricity among others, and there has been a continuous explosion of sectorial studies using conventional or more sophisticated DEA models. Some authors remark that there are inherent benefits of applying DEA in the context of countries' performance analysis as the method is based on the most favorable and country-specific weights (Atkinson et al. 2002; Cherchye et al. 2008). Thus, the controversy on the subjective judgments regarding the weights for the sub-indicators that are needed in other methodologies does not exist. There exist a number of papers that analyze under different perspective the previous research that have appeared in the DEA literature (Charnes et al. 1994; Emrouznejad et al. 2008; Cook and Seiford 2009; Cooper et al. 2011; Zhu 2014).

These reviews show that the applications which deal directly with DEA evaluations of countries' logistics performance are inexistent. To our best knowledge, there are only two studies that analyze the logistics performance of cities (Jiang 2010) or regions (Jiang and Fu 2009), and there is only one study which deals with the countries' logistics performance (Markovits-Somogyi and Bokor 2014). In this last paper, the authors analyzed the logistics efficiency of 29 European countries using a methodology where DEA is combined with an analytic hierarchy process to fully rank all the countries included in the analysis. The authors compared also the results with a DEA-PC (pairwise comparison) methodology and with the 'Logistics quality and competence' index of the LPI.

III. Methodology

In DEA analysis, it is generally assumed that there are n production units to be evaluated, using amounts of m different inputs to produce quantities of s different outputs. Specifically, the o'th production unit consumes x_{io} units of input i (I = 1 to m) and produces y_{ro} units of output r (r =1 to s). The o'th production unit can be described more compactly with the vector (X_o,Y_o), which denote, respectively, the vectors of input and output values for DMU_o.

Next, it is necessary to determine a potential set of possible dominant or non-dominant comparisons for each production unit considered in the analysis. DEA usually considers the dominance of all the possible linear combinations of the n DMUs, i.e. , with the scalar restricted to be non-negative.[2] The production unit o is dominated, in terms of inputs, if at least one linear combination of production units shows that some input can be decreased without worsening off the rest of inputs and outputs. In the same way, it is dominated in terms of outputs if at least one linear combination of production units shows that some output can be increased without worsening off the rest of inputs and outputs.[3]

In our case, policy makers can affect the logistics performance of their country making policies that improve some of the dimensions considered in the LPI such as the efficiency of customs and border clearance, the quality of trade and transport infrastructure, the ease of arranging competitively priced shipments, the competence and quality of logistics services –trucking, forwarding, and customs brokerage, the ability to track and trace consignments or the frequency with which shipments reach consignees within scheduled or expected delivery times. It is out of the scope of the present paper to give some guidelines about what trade facilitation measures should be implemented as we do not have the costs of such policies. Möisé (2013) contended that some measures need to be evaluated taking a long term perspective as these may be expensive to introduce but not costly to operate. Other actions require political commitment rather than funds, and some institutional barriers act as real impediments to achieve any gain. In any case, the author concluded saying that an increasing amount of technical and financial assistance to implement some trade facilitation measures has been made available to developing countries over the last decade.

In this paper, countries' logistics performance is going to be based on a Constant Returns to Scale (CRS) input orientation model. In this sense, the problem is resolved for each country through the following linear programming specification:

$$\begin{aligned} &\max_{v\mu} \sum_{r=1}^{s} \mu_r y_{r0} \quad ; \\ &\text{s.t.} \ \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0 \quad (j = 1 \cdots n), \sum_{i=1}^{m} v_i x_{i0} = 1, \text{where } v_i, \mu_r \geq 0. \end{aligned} \tag{1}$$

---

[2] Different envelopment surfaces may be obtained considering additional constraints about the scalars. For example, VRS models are obtained imposing convexity constraint N1´λ=1; and non-increasing return to scale models (NIRS) are characterized by the restriction of the sum of scalars being less or equal to one.

[3] This discussion is very close to the definition of Pareto-Koopmans efficiency. The unit o is considered fully efficient if and only if the performance of other DMUs does not provide evidence that some of the inputs or outputs of the unit o could have been improved without worsening off some of its other inputs or outputs. This definition of relative performance has its origin in Farrell (1957).

A country is in the frontier only if $\sum_{r=1}^{s}\mu_r y_{ro}=1$, this is optimality. The constraint $\sum_{i=1}^{m}v_i x_{io}=1$ is known as a normalization constraint, and the weighted input and output are called virtual input and virtual output, respectively. See Seiford and Thrall (1990) for a detailed discussion of these models. The efficiency ratio ranges from 0 to 1. Thus, for each country under analysis the weights will be chosen so as to maximize self-efficiency, given the constraints. This intrinsic characteristic of the model explains partly the appeal of DEA-based CIs in real policy-related exercises. It is unarguable that several policy issues should balance adequately different regional interests taking into account supranational, regional or country-specific policy priorities. For this reason, a fixed set of weights to compare the multidimensional performance of countries may prevent the acceptance of the evaluation.

IV. Data

The LPI is a good indicator of trade facilitation for a broad group of countries. The logistics index values differ between countries and provide a general picture of customs procedures, logistics costs and the quality of the infrastructure necessary for overland and maritime transport. The World Bank has published this index for 4 years (Arvis et al. 2007, 2010, 2012 and 2014), ranking 150 countries and providing an extensive explanation of logistics performance in these countries (43 from Africa, 42 from Europe, 41 from Asia, 22 from South America, 5 from the Pacific, and 2 from North America). The first edition contains data compiled in 2005; the second edition contains data processed between 2008 and 2009; the third edition contains information for the year 2010, following the same sequence for 2012. The index makes an important statistical contribution by establishing a harmonized scale for all countries to identify the difficulties faced by bilateral trade, together with their requirements in terms of logistics associated with existing facilities. From the information obtained, the LPI is constructed using the Principal Component Analysis (PCA), a statistical technique used to reduce the dimensionality of a dataset. Thus, using inputs corresponding to each of the six components, and then averaging out scores for each country, the PCA ultimately provides a single indicator - the LPI - thereby establishing a logistics ranking for the countries analysed.

The LPI is built on the basis of a worldwide survey carried out on companies responsible for the transport of goods and for the facilitation of trade globally. Specifically, it was developed with the assistance of over 800 professionals involved across the different areas of the sector's lines of activity.[4] Each respondent of the survey was asked for data pertaining to the eight countries they most traded with at the international level.

The aggregate index is calculated by analysing six main components using the following indicators: customs, infrastructure, international shipments, logistics quality and competence, tracking and tracing, and timeliness.[5] None of these independently guarantee a good level of logistics performance, and their inclusion is conditioned to empirical studies and extensive interviews carried out with specialists in international freight transport. All the indicators have been aggregated and duly weighted. Scores range from 1 to 5, the highest score representing the best logistics performance. Each component is defined as follows:

- Customs: measures agility clearance processes, in terms of speed, simplicity and predictability of formal issues conducted by customs control bodies.
- Infrastructure: evaluates the quality of maritime, land, rail and air transport infrastructure. The perception held by respondents about this infrastructure is valuated in terms of the modes of transport together with storage and moving goods.
- International shipments: measures the ease of negotiating competitive prices for sending.
- Logistics quality and competence: indicates the quality of logistical services, such as transport operators or customs agents.
- Tracking and tracing: measures the follow-up and location of shipments. Identifying the exact location and route followed by each good is relevant up to the moment of delivery to the final client. In this component, all agents of the good's supply chain are involved; therefore, traceability is the result of global action.
- Timeliness: refers to the exact time of shipment delivery. It is important to consider this factor because due to the high degree of existing competition, not meeting the established times is unacceptable.

These indicators can be divided into two main areas: (1) regulatory policies (Customs, Infrastructure and Logistic quality and competence), and (2) service delivery performance outcomes (Timeliness, International shipments, and Tracking and tracing). The first concerns the distribution chain, while the second determines

---

[4] The questionnaire is available at www.worldbank.org/lpi

[5] The LPI published in 2010, 2012 and 2014 only take six indicators into consideration (they exclude the domestic logistics costs included in 2007).

the efficiency of the service. Each component is key to determining competitiveness in international trade within each country. Any changes to these components has important repercussions. For example, an improvement in Customs and Infrastructure would lead to an increase of 4.7% and 14.5% of GDP and global trade respectively. If tariffs were completely eliminated worldwide, GDP would increase by 0.7% and trade by 10.1% (400 billion and 1.1 trillion dollars), or at any rate would improve the efficiency of the international transport of goods. In the literature studies tend to follow this approach, concluding that frequency, time flexibility, development of infrastructure, and on-time delivery are all key factors in international competitiveness variables.[6]

In general, low-income countries, with little development or geographical impediments as far as market access goes, occupy the last places of the ranking (countries from Africa and Central Asia). However, it should be clarified that when trade has been a factor in accelerating their growth, logistical performance is also significantly better than in other locations with similar income levels (India and Vietnam, both low income, are ranked 46 and 53, respectively, in 2010).

According to the index published in 2014 (Arvis et al., 2014), higher-income countries occupy the top 10 positions in the ranking (Germany, Netherlands, Belgium, United Kingdom, Singapore, Sweden, Norway, Luxembourg, USA and Japan). These countries are well positioned logistically, and play a key role in supply chains at both global and regional levels. In a similar way, at the bottom of the ranking lie lower income countries, mainly African nations or countries where conflicts have undermined their development.

On the other hand, the distance between the highest and lowest countries has narrowed progressively. The LPI expressed as a percentage of the highest country in the ranking reveals that the LPI for Somalia represents 25% of the highest performer (Germany), while previously it was 19% in 2012, 11% in 2010 and 7% in 2007. At the same time, the gap between countries at the top of the ranking is narrowing. This might be explained by the improvement in infrastructure to foster trade in low and middle-income countries, and to a lesser extent by their logistics performance and customs clearance. Hence, the same progressive strategies clearly cannot be applied equally to all countries.

In this paper, three different DEA scenarios are proposed (Table 1). The first scenario is characterized by considering customs, infrastructure, and international shipments as inputs, applying a monotone decreasing transformation (five minus

---

[6] For example see, Boske (2001) and Herrera (2005)

the original values). This first scenario is based on the first category proposed by Arvis et al (2007). The authors claimed that this category contains mainly inputs to the supply chain such as customs, infrastructure, and ease of arrangement for international shipments. The rest of components that belong to the second category are left as outputs with their original values. The second scenario is based on the DEA method in which all the LPI components are considered as outputs jointly with a single constant input variable. In the third scenario, the role of inputs and outputs between the original LPI components is reversed. These three scenarios are used to perform a sensitivity analysis as a way to assess the robustness of the final results.[7]

Table 2 shows the descriptive statistics for the inputs and outputs of the LPI components that were included in our analysis under the first scenario. As it can be observed, there are great differences between the minimum and maximum values of almost all variables. However, the standard deviation and average figures do not

Table 1. Inputs and outputs used in the study under three different scenarios

A. First scenario (base)

Inputs	Outputs
Customs	Logistics quality and competence
Infrastructure	Tracking and tracing
Ease of arrangement shipments	Timeliness

B. Second scenario

Inputs	Outputs
A single constant variable (1)	Customs
	Infrastructure
	Ease of arrangement shipments
	Logistics quality and competence
	Tracking and tracing
	Timeliness

C. Third scenario

Inputs	Outputs
Logistics quality and competence	Customs
Tracking and tracing	Infrastructure
Timeliness	Ease of arrangement shipments

Source: Own elaboration.

[7] The robustness analysis was suggested by two anonymous reviewers.

show any particular pattern with the exception that timeliness is the only variable that presents an average figure higher than three. This means that timeliness is the most positively valued by the logistics professionals from the companies responsible for moving goods around the world who answered the structured online survey administered by the World Bank. Looking at those countries which present the best and worst performance values, it can be seen that good performers (Norway, Germany and Luxembourg) are Western European countries. Regarding the worst performers, there are only two countries (Somalia and Yemen) that present the lowest figures in the whole set of dimensions.

The poor results of Somalia and Yemen can be partly explained by the thousands of attacks on cargo ships perpetrated in the Somalian coast during the last decade that have caused a significant burden to maritime trade in the area. Burlando et al. (2015) found that cargo passing through pirate waters has been reduced by 4.1% per year in the period 2000-2010 and that this reduction is not evenly distributed in all the groups of goods that are shipped by sea. They also found that five countries and the EU shouldered 70% of the total costs. The Somalian and Yemeni results are a consequence of a combination of sources such as weak governmental institutions, a natural bottleneck in the area, and a significant flow of merchant ships through the Gulf as more than a 10% of the cargo use the Suez Canal and is potentially affected by this threat. Recent reports indicate that piracy is on the decline in Somalia (Saul 2013). The ongoing slowdown in attacks might be due to the presence of navy patrols and enhanced on-board security (World Bank 2013). In any case, even in the absence of a significant number of attacks, pirates would have increased cargo tariffs affecting international trade.

Table 2. Descriptive statistics

Variables	Mean	SD	Min	Country	Max	Country
Inputs						
Customs	2.24	0.60	0.79	Norway	3.38	Yemen
Infrastructure	2.19	0.67	0.68	Germany	3.50	Somalia
International shipments	2.10	0.49	1.18	Luxembourg	3.25	Somalia
Outputs						
Logistics quality and competence	2.90	0.58	1.75	Somalia	4.19	Norway
Tracking and tracing	2.94	0.59	1.75	Somalia	4.17	Germany
Timeliness	3.30	0.59	1.88	Somalia	4.71	Luxembourg

Source: Own elaboration.

V. Results

As discussed earlier, we use a multiplier DEA input model to analyze the logistics performance for a group of 141 countries. Table 3 shows the results for the twenty best and worst countries in the world included in our analysis under the first scenario. We find that the group of best performers are mainly characterized by high-income countries that belong to Europe and Asia together with the US and Canada. However, the list of the twenty worst countries is highly biased to the Africa continent and some other low-income countries of other regions like Bhutan, Myanmar, Haiti and Afghanistan.[8]

Table 3. The 20 best and worst countries in the world according to the DEA-LPI. 1st scenario.

Rank	Country	VDEA	Rank	Country	VDEA
20 Best Countries					
1	Belgium	1.00000	11	Denmark	0.85666
2	Germany	1.00000	12	USA	0.82392
3	Norway	1.00000	13	Japan	0.81430
4	Luxembourg	1.00000	14	Switzerland	0.80654
5	Sweden	0.94133	15	China. Hong Kong SAR	0.80531
6	Singapore	0.93740	16	New Zealand	0.80391
7	Netherlands	0.91997	17	Ireland	0.79227
8	United Kingdom	0.89631	18	Malaysia	0.78170
9	France	0.86466	19	Australia	0.77113
10	Taiwan, China	0.85707	20	Canada	0.76995
20 Worst Countries					
122	Bhutan	0.27736	132	Haiti	0.25536
123	Lesotho	0.27704	133	Sudan	0.25501
124	Zimbabwe	0.27650	134	Kyrgyz Republic	0.25125
125	Azerbaijan	0.27380	135	Mozambique	0.24209
126	Zambia	0.27243	136	Mauritania	0.23931
127	Gabon	0.27225	137	Djibouti	0.22814
128	Tanzania	0.27083	138	Eritrea	0.22584
129	Cameroon	0.26782	139	Syrian Arab Republic	0.22284
130	Yemen, Rep.	0.26521	140	Afghanistan	0.21587
131	Myanmar	0.25619	141	Somalia	0.15952

Source: Own elaboration.

[8] We note here that all the DEA-LPI figures are lower than or equal to one. 1. The values have been calculated according to the formulation of DEA-LP program described by equation 2.

An examination of the Table 3 reveals that, according to the efficiency DEA-LPI score, Belgium, Germany, Norway and Luxembourg are the most competitive countries in the world regarding their logistics performance. In fact, they form the peers in the frontier according to DEA parlance. It is interesting to remark that these countries are all located in the European continent, and although Norway is not a member of the European Union (EU), the country has a long established good relationship through the Agreement on the European Economic Area (EEA) which facilitates that Norway takes part in the EU internal market. Norway also signed the Schengen Agreement and cooperates with the EU on foreign and security policy issues. Regarding the ten best performers, all the countries are considered as high income according to the PPP-GNI[9] index for the year 2011 elaborated by the World Bank. Most of them belong to the OECD and only Taiwan and Singapore are non-OECD countries. On the other hand, it can be seen that Djibouti, Eritrea, Syrian Arab Republic, Afghanistan and Somalia are the least competitive countries of the world. The majority of the countries in the lower end of the ranking are located in Africa. The freight logistics systems in Afghanistan are exploited for a variety of illicit activities, in particular for trafficking of prohibited and restricted goods.[10] For example, the heroin annual flows into the global market are assessed to be between 430-450 tons, and Afghanistan is the main source followed by Myanmar and Laos (UNODC 2010). Djibouti and Eritrea share part of the coast of the Red Sea but very near to Somali routes where the pirates' conflicts of the last decade have reduced the cargo trade passing through the Gulf of Aden. All the countries belong to the groups of low or lower middle income.

Comparing the groups of worst and best performers countries according to the DEA-LPI and the LPI, it can be seen that the four countries that belong to the frontier using our empirical results are ranked as the first (Germany), the third (Belgium), the seventh (Norway) and the eighth (Luxembourg).

---

[9] GNI per capita based on purchasing power parity (PPP). PPP GNI is gross national income (GNI) converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GNI as a U.S. dollar has in the United States. GNI is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad.

[10] It is out of the scope of the current paper to analyze to what extent there exist a negative relationship between this illicit trade and the logistics performance. Hintsa and Mohanty (2014) prepared a literature-based qualitative framework for the assessment of socio-economic negative impacts on six commonly occurring illegal trade flows: (1) trafficking in cocaine and heroin; (2) counterfeit products; (3) ozone depleting substances; (4) firearms; (5) stolen cultural products; and (6) endangered species.

Focusing on the logistics performance of the ten best countries according to these two methodologies, we observe that there are four main mismatches in the following set: Taiwan (10, 12, 17, 19), France (9, 13, 14, 13), United States (12, 6, 6, 9) and Japan (13, 14, 11, 10) (Table 4). The first three figures in parenthesis show the rank obtained by our DEA-LPI method under the three different scenarios and the last figure gives the rank obtained by the LPI methodologym.[11] Regarding the other extreme, the five worst performers according to our methodology are also located in the set of the seven worst performers of the LPI method. There are only three mismatches looking at the group of the ten worst performers using both methods, namely Yemen (131, 132, 131, 136), Mozambique (136, 138, 136, 132), and Haiti (133, 135, 134, 129). Using a Spearman correlation coefficient to estimate a rank-based measure of association between these four ranking methods, we can conclude that there is a positive association between all the four methods (ϱ lies in

Table 4. Mismatches between DEA-LPI ranks and LPI rank. Spearman correlation coefficients

	DEA-LPI			LPI-Rank
	1st scenario	2nd scenario	3rd scenario	
Best dountries				
Taiwan	10	12	17	19
France	9	13	14	13
USA	12	6	6	9
Japan	13	14	11	10
Worst countries				
Yemen	131	132	131	136
Mozambique	136	138	136	132
Haiti	133	135	134	129

	Spearman correlation coefficients		
	DEA-LPI (1)	DEA-LPI (2)	DEA-LPI (3)
LPI-Rank	0.9819	0.9655	0.9870
DEA-LPI (2)		0.9891	0.9907
DEA-LPI (3)			0.9821

Source: Own elaboration.

[11] The LPI is constructed using PCA in which the normalized scores for each of the six original indicators are multiplied by their component loadings and then summed. The component loadings represent the weight given to each original indicator in constructing the international LPI. Since the loadings are similar for all six, the international LPI is close to a simple average of the Indicators.

the range between 0.9655 and 0.9907). The values of ϱ show that these four methods do not obtain the same ranking logistics performance as discussed above. Nevertheless, the robustness of DEA results to different selection of inputs and outputs has been proven. Thus for the rest of the paper, DEA-LPI results are referred to the first scenario under consideration.

By analyzing the group of worst and best performers, it seems that income and geographical area might influence the DEA-LPI score. For this reason, one-way analysis of variance is going to be used in order to examine whether there are significant differences that can be accrued to these particular factors. Table 5 shows the standard ANOVA table, which divides the variability of the DEA-LPI performance into two parts: variability due to the differences among the factor groups means (variability between groups); and variability due to the differences between the individual country performance in each group and the group mean (variability within groups).

The results of the ANOVA analysis show that the null hypothesis, i.e., the average performance of the DEA-LPI is equal independently of the geographical area location or income, may be rejected. The p-value, shown in the sixth column, casts doubt on the null hypothesis and suggests that at least the logistics performance in some group of countries is significantly different from other groups. We compare the

Table 5. One-way analysis of variance. TTCI performance by income and geographical area

	Df	SumSq	MeanSq	Fvalue	Pr(>F)	
Income	4	4.140	1.034	66.6	<2e-16	***
Residuals	136	2.113	0.015			
Geographical Area	6	2.362	0.3926	13.55	5.54e-12	***
Residuals	134	3.891	0.029			
Grand mean: 0.4747						

Income factor means

Low Income	Lower Middle Income	Upper Middle Income	High Income Non OECD	High Income OECD
0.308 (29)*	0.351 (36)	0.447 (32)	0.574 (15)	0.774 (29)

Geographical area factor means

East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa
0.58 (19)	0.61 (41)	0.39 (21)	0.42 (15)
North America	South Asia	Sub-Saharan Africa	
0.79 (2)	0.35 (7)	0.32 (36)	

Notes: *** 1% significance codes.* The number of countries appears between parentheses for each of the factor means.
Source: Own elaboration.

performance of the groups of countries according to their geographical area and income and we test the hypothesis that the average DEA-LPI score is the same, against the general alternative that some significant differences exist. However, as we accept the alternative hypothesis and it is too general, we would like to obtain more particular information about which pairs of means are significantly different, and which are not. For this reason, we study pair wise mean differences to assess in what sense a group can be characterized by its better or lower performance.

To do this, we need to use some multiple comparison procedure. In our case, we use the Tukey-Kramer test in order to determine whether the DEA-LPI performance is significantly different according to each of the factors under analysis. As we want to compare every group to each other, we can form ten and twenty-one different pairwise comparisons to obtain their mean differences attending their income and geographical area, respectively. Differences and 95% confidence interval for these differences are presented in Table 6.

As shown in Table 6, we find that that the difference between the High Income OECD countries and High Income non OECD countries is 0.2004 and a 95% confidence interval for the true mean is [0.0908, 0.3100]. In this example the confidence interval does not contain 0, so the difference is significant at the 0.05 level,[12] and we can conclude that the performance of logistics of High Income OECD countries is better than those that belong to the group of High income non OECD countries.

Table 6. Tukey multiple comparisons of means. 95% family-wise confidence level

Comparison (Income)	Difference	Lower	Upper	Probability
High income OECD/High income non OECD	0.2004	0.0908	0.3100	0.0000134 s
Low income/High income non OECD	-0.2663	-0.3759	-0.1567	0.0000000 s
Lower middle income/High income non OECD	-0.2232	-0.3291	-0.1173	0.0000004 s
Upper middle income/High income non OECD	-0.1266	-0.2345	-0.0188	0.0125630 s
Low income/High income OECD	-0.4667	-0.5572	-0.3762	0.0000000 s
Lower middle income/High income OECD	-0.4236	-0.5096	-0.3377	0.0000000 s
Upper middle income/High incom OECD	-0.3270	-0.4154	-0.2387	0.0000000 s
Lower middle income/Low income	0.0430	-0.0429	0.1290	0.6381660
Upper middle income/Low income	0.1397	0.0513	0.2280	0.0002335 s
Upper middle income/Lower middle income	0.0966	0.0128	0.1803	0.0149684 s

Notes: s Differences are statistically significant for the comparison between the groups under consideration at least at 95 per cent of confidence level. Source: Own elaboration.

[12] In fact, the probability shown in the last column of the table can be used to obtain the exact p-confidence value.

If the confidence interval contains the zero value, then we conclude that the difference is not statistically significant at the 0.05 level (see, for example, the eighth row in Table 6). In this case, we can conclude that the performance of the Lower Middle Income countries is not significantly different from the Low Income countries. However, it can be seen that the rest of the rows show a statistical significant difference between the countries that belong to different income groups. In all the cases, the expected conclusion that says that higher income countries are better logistics performers is observed.

In a similar way, Table 7 shows the relative performance of the countries focusing now in the geographical area. In this case, it can be seen that the differences can be accrued to different areas that include Latin American & Caribbean, South Asia,

Table 7. Tukey multiple comparisons of means. 95% family-wise confidence level

Comparison (geographical area)	Difference	Lower	Upper	Probability
Europe & Central Asia/East Asia & Pacific	0.03565	-0.10592	0.17723	0.98872
Latin America & Caribbean/East Asia & Pacific	-0.18551	-0.34703	-0.02399	0.01344 s
Middle East & North Africa/East Asia & Pacific	-0.15567	-0.33187	0.02052	0.12105
North America/East Asia & Pacific	0.21386	-0.16537	0.59309	0.62508
South Asia/East Asia & Pacific	-0.22528	-0.45083	0.00028	0.05050 s
Sub-Saharan Africa/East Asia & Pacific	-0.25987	-0.40453	-0.11522	0.00001 s
Latin America & Caribbean/Europe & Central Asia	-0.22116	-0.35806	-0.08427	0.00007 s
Middle East & North Africa/Europe & Central Asia	-0.19133	-0.34527	-0.03739	0.00529 s
North America/Europe & Central Asia	0.17821	-0.19121	0.54762	0.77675
South Asia/Europe & Central Asia	-0.26093	-0.46956	-0.05231	0.00488 s
Sub-Saharan Africa/Europe & Central Asia	-0.29553	-0.41204	-0.17901	0.00000 s
Middle East & North Africa/Latin America & Caribbean	0.02984	-0.14262	0.20229	0.99856
North America/Latin America & Caribbean	0.39937	0.02186	0.77688	0.03067 s
South Asia/Latin America & Caribbean	-0.03977	-0.26241	0.18287	0.99828
Sub-Saharan Africa/Latin America & Caribbean	-0.07436	-0.21444	0.06571	0.68938
North America/Middle East & North Africa	0.36954	-0.01448	0.75355	0.06757
South Asia/Middle East & North Africa	-0.06960	-0.30311	0.16391	0.97318
Sub-Saharan Africa/Middle East & North Africa	-0.10420	-0.26097	0.05257	0.42604
South Asia/North America	-0.43914	-0.84816	-0.03012	0.02671 s
Sub-Saharan Africa/North America	-0.47374	-0.84434	-0.10313	0.00366 s
Sub-Saharan Africa/South Asia	-0.03460	-0.24532	0.17613	0.99893

Notes: s Differences are statistically significant for the comparison between the groups under consideration at least at 95 per cent of confidence level. Source: Own elaboration.

Sub-Saharan Africa, Middle East & North Africa as low performance regions; and East Asia & Pacific, Europe & Central Asia and North America as high performers.

Sub-Saharan Africa region presents the lowest DEA-LPI performance of the world in spite of receiving a lot of funding and attention in the last years for potential infrastructure development that addressed this area's massive deficiencies in transport provision (Gwilliam 2010; Gwilliam et al. 2010). Foster and Briceño-Garmendia (2010) found the following Decalogue: (1) infrastructure has been responsible for more than half of Africa's recent improved growth performance and has the potential to contribute even more in the future; (2) Africa's infrastructure networks increasingly lag behind those of other developing countries and are characterized by missing regional links; (3) Africa's economic geography presents a particular challenge for the region's infrastructure development; (4) Africa's infrastructure services are twice as expensive as elsewhere, and lack of competition is one of the main causes; (5) Power is by far Africa's largest infrastructure challenge; (6) The costs of addressing Africa's infrastructure needs is around \$93 billion a year; (7) The infrastructure challenges are very heterogeneous among different countries; (8) A large share of Africa's infrastructure is still domestically financed mainly by the central government budget; (9) Africa would still face an infrastructure funding gap of \$31 billion a year; and (10) Africa's institutional, regulatory, and administrative reforms are only halfway along.

VI. Concluding remarks

Our DEA-LPI has aimed to contribute to the literature strand on the ranking of countries regarding the performance in logistics. To our knowledge, there are only two studies with a similar aim: the DEA-PC method proposed by Markovits-Somogyi and Bokor (2014) and the LPI method proposed by Arvis et al. (2014). The approach adopted in the present study is an hybrid of both of the methods as it used DEA as the methodological approach and the LPI database in terms of the variables and the countries included in the analysis. Our method is based on an input orientation DEA efficiency model under three different scenarios which presents some major advantages over other traditional ranking and benchmarking models. In particular, it could be used to rank all the countries unambiguously except the four countries that were part of the frontier.

This paper has offered interesting insights into the benchmark position of the countries regarding the logistics performance. Our findings reveal striking differences among the best and the worst performers, as well as among different geographical areas. We have shown that the DEA-LPI results are robust to different input and output selection. We have also compared our results with those provided by the original LPI, finding that there is a significant positive association between all the four analyzed methods. Nevertheless, as it was discussed, the methods did not give the same results.

We have also found that the logistics performance depends largely on income and geographical area. On one hand, our findings suggest that high income countries are in the group of best performers. In particular, we found that the group of the ten best performers is highly dominated by the EU. It is difficult to predict additional related strategic re-location of logistics and production platforms in specific industries that would result in a deterioration of Europe's role as a main production/industrial world region. However, the recent financial crisis that has affected the euro zone and the Greek situation could affect this leadership in the near future. On the other hand, in spite of all the efforts that have been made in the recent past there is still a big gap between this developed region and the Sub-Saharan region. More innovative logistics programs need to be developed in the lagged regions of the world.

This new method maximizes the radial distance for those variables considered as outputs taking into account that all the countries are relatively dominated by those countries that form the technological frontier. Furthermore, the new model could be adapted to reflect realistic conditions in an efficiency-improvement projection taking into account different layers of projection conformed by a different set of countries. Thus, the stepwise projection allows all the countries to incorporate more realistic levels of potential improvement that take into account their own characteristics and those from the area where they are located. To summarize, this stepwise DEA-LPI model would be able to present a more realistic direction and intensity for efficiency-improvement regarding logistics performance, and may thus provide a valid tool for planners and policy makers for implementing adequate logistics programs.

We consider that DEA-LPI is a promising tool for ranking countries logistics performance. In this paper, the analysis has been focused on the tool properties and the effects of income and geographical area. However, an interesting area for future research that needs to be addressed could be based on the individual country performance regarding the relative efficiency progress or regress in two different periods of time. Thus, a policy analysis of some specific actions could be analyzed to obtain important practical implications for different stakeholders.

References

Arvis, Jean-François, Mónica Mustra, Lauri Ojala. Ben Shepherd. and Daniel Saslavsky (2010, 2012). *Connecting to compete: Trade logistics in the global economy*. The World Bank, Washington DC.

Arvis, Jean-François, Mónica Mustra, John Panzer, Lauri Ojala, and Tapio Naula (2007). *Connecting to compete: Trade logistics in the global econom*y. The World Bank, Washington DC.

Arvis, Jean-François., Daniel Saslavsky., Lauri Ojala., Ben Shepherd., Cristina Busch., and Anasuya Raj (2014). *Connecting to compete 2014: Trade logistics in the global economy – The Logistics Performance Index and its indicators*. The World Bank, Washington DC.

Atkinson, Tony, Bea Cantillon, Eric Marlier, and Brian Nolan (2002). *Social indicators. The EU and social inclusion*. Oxford, Oxford University Press.

Behar, Alberto, and Phil Manners (2008). Logistics and exports. Mimeo, CSAE, Oxford.

Blyde, Juan, and Gonzalo Iberti (2012). Trade costs, resource reallocation and productivity in developing countries. *Review of International Economics* 20: 909–923.

Boske, Leigh (2001). Maritime transportation in Latin America and Caribbean. Policy Research Project Report No. 138. School of Public Affairs, The University of Texas at Austin.

Burlando, Alfredo, Anca Cristea, and Logan M. Lee (2015). The trade consequences of maritime insecurity: Evidence from Somali piracy. *Review of International Economics* 23: 525–557.

Charnes, Abraham, and William. W. Cooper (1985). Preface to topics in data envelopment analysis. *Annals of Operations Research* 2: 59–94.

Charnes, Abraham, William W. Cooper, Arie Y. Lewin, and Lawrence M. Seiford (1994). *Data envelopment analysis. Theory, methodology and applications*. Boston, Kluwer Academic.

Charnes, Abraham, William W. Cooper, and Edwardo Rhodes (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research* 2: 429–444.

Cherchye, Laurnes, Wim Moesen, Nicky Rogge, Tom Van Puyenbroeck, Michaela Saisana, Andrea Saltelli, Roman Liska, and Stefano Tarantola (2008). Creating composite indicators with DEA and robustness analysis: The case of the Technology Achievement Index. *Journal of the Operational Research Society* 59: 239–251.

Cook, Wade D., and Lawrence M. Seiford (2009). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research* 192: 1–17.

Cooper, William W, Lawrence M. Seiford, and Joe Zhu (2011). *Handbook on data envelopment analysis*. Springer Science & Business Media.

De Souza, Robert, Mark Goh, Sumeet Gupta, and Luo Lei (2007). An investigation into the measures affecting the integration of ASEAN's priority sectors: Phase 2: the case of logistics. REPSF Project No. 06/001d Regional Economic Policy Support Facility, Association of Southeast Asian Nations, Manila.

Decreux, Yvan, and Lionel Fontagne (2006). A quantitative assessment of the outcome of the Doha development agenda. Working Paper 2006-10, CEPII.

Dennis, Allen (2006). The impact of regional trade agreements and trade facilitation in the Middle East and North Africa region. Policy Research Working Paper 3837, World Bank.

Emrouznejad, Ali, Barnett R. Parker, and Gabriel Tavares (2008). Evaluation of research in efficiency and productivity:A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences* 42: 151–157.

Farrell, Michael J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society* A120: 253–290.

Filippetti, Andrea, and Antonio Peyrache (2011). The patterns of technological capabilities of countries: A dual approach using composite indicators and data envelopment analysis. *World Development* 39: 1108–1121.

Foster, Vivien, and Cecilia Briceno-Garmendia (2010). *Africa's infrastructure: A time for transformation*. Washington, DC, World Bank and Infrastructure Consortium for Africa.

Gwilliam, Kenneth. (2010). Transport: More than the sum of its parts. In V. Foster and C. Briceno- Garmendia, editors, *Africa's infrastructure: A time for transformation*. Washington, DC, World Bank and Infrastructure Consortium for Africa.

Gwilliam, Kenneth, Kavita Sethi, Alberto Nogales, and Vivien Foster (2010). Roads: Broadening the agenda. In V. Foster and C. Briceno-Garmendia, editors, *Africa's infrastructure: A time for transformation*, Washington, DC, World Bank and Infrastructure Consortium for Africa.

Herrera, Alicia (2005). *Equipment for intermodal transportation*. U.S. Commercial Service

Hintsa, Juha, and Sangeeta Mohanty (2014). A Literature-based qualitative framework for assessment of socio-economic negative impacts of common illicit cross-border freight logistics flows. In T. Blecker, W. Kersten, and C. M. Ringle, editors, *Innovative methods in logistics and supply chain management. Current issues and emerging practices*. Berlin, epubli GmbH.

Hoekman, Bernard, and Alessandro Nicita (2011). Trade policy, trade costs, and developing country trade, *World Development* 39: 2069–2079.

Jiang, Changbing. (2010). Research on logistics network infrastructure based on HCA and DEA-PCA approach. *Journal of Computers* 5: 533–540.

Jiang, Changbing, and Peihua Fu (2009). Evaluating efficiency and effectiveness of logistics infrastructure based on PCA-DEA approach in China. In *Second International Conference on Intelligent Computation Technology and Automation, 2009 (ICICTA'09)*. IEEE.

Korinek, Jane, and Patricia Sourdin (2011). To what extent are high-quality logistics services trade facilitating? Trade Policy Working Papers 108, OECD.

Mahlberg, Bernhand, and Michael Obersteiner (2001). Remeasuring the HDI by data envelopment analysis. Interim Report 01-069, International Institute for Applied Systems Analysis, Laxenburg.

Markovits-Somogyi, Rita, and Zoltan Bokor (2014). Assessing the logistics efficiency of European countries by using the DEA-PC methodology. *Transport* 29: 137–145.

Moïsé, Evdokia (2013). The costs and challenges of implementing trade facilitation measures. Trade Policy Papers 157, OECD.

OECD. (2003). Quantitative assessment of the benefits of trade facilitation. TD/TC/WP31, Paris, OECD.

OECD. (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris, OECD Publications.

Osberg, Lars, and Andrew Sharpe (2002). An index of economic well-being for selected OECD countries. *Review of Income and Wealth* 48: 291–316.

Puertas, Rosa, Luisa Martí, and Leandro García (2014). Logistics performance and export competitiveness: European experience. *Empirica* 41: 467–80.

Rushton, Alan, Phil Croucher, and Peter Baker (2014). *The handbook of logistics and distribution management: Understanding the supply chain*. London, Kogan Page Publishers.

Sanz, Alberto (2014). IX Conferencia Ministerial de la Organización Mundial de Comercio. *Boletín Económico del ICE* 3.047: 3–9.

Saul, Jonathan. (2013). Sea piracy falls to 5-year low as Somali gangs retreat. Reuters, 16 January.

Seiford, Lawrence M., and Robert M. Thrall (1990). Recent developments in data envelopment analysis: The mathematical programming approach to frontier analysis. *Journal of Econometrics* 46: 7–38.

Soloaga, Isidro, John S. Wilson, and Alejandro Mejía (2006). Trade facilitation reform and Mexican competitiveness. Policy Research Working Paper 3953, World Bank.

UNDP. (2001). *Human development report*. New York, Oxford University Press

UNODC. (2010). *World drug report*. United Nations Publication.

Wilson, John S, Catherine Mann, and Tsunehiro Otsuki (2005). Assessing the potential benefit of trade facilitation: a global perspective. In P. Dee and M. Ferrantino, editors, *Quantitative Methods for assessing the effects of non-tariff measures and trade facilitation*. Singapore, APEC Secretariat and World Scientific.

World Bank (2013). Pirates of Somalia: Ending the threat, rebuilding the nation. Washington, DC, Regional Vice-Presidency for Africa,World Bank.

Zhu, Joe (2014). *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheet*. International Series in Operations Research & Management Science, Springer.

Journal of Applied Economics. Vol XX, No. 1 (May 2017), 193-210

IMPOSING CONCAVITY AND THE NULL-JOINTNESS PROPERTY ON THE PRODUCTION POSSIBILITIES FRONTIER IN CASE OF POLLUTING TECHNOLOGIES

Alexandre Repkine*

Konkuk University

Submitted: February 2015; accepted December 2015

Economic theory requires the directional distance functions used to study the properties of production possibility sets of polluting technologies to be concave in both outputs, while the implied production possibilities frontier (PPF) is required to be concave with respect to the bad output. However, existing estimation frameworks do not preclude the estimation of convex PPFs. We analyze geometrical properties of the quadratic approximation to the directional output distance functions to derive a constraint that guarantees PPF concavity and consider the issue of imposing the property of null-jointness on the production possibilities set, which is also required by theory. We simulate a dataset corresponding to a concave PPF and show that in case concavity and null-jointness constraints are not imposed, it is possible that the conventional estimation framework may lead to erroneous conclusions with respect to the type of curvature of both the directional output distance function, and the PPF.

JEL classification codes: D24, Q53, O44, R11

Key words: CO_2 emissions, marginal abatement cost, distance function

I. Introduction

It is convenient to model polluting production processes in terms of the multi-output technologies with at least one output, e.g., CO_2 emissions, being an undesirable bad. Output distance functions suggested by Shephard (1970) are a useful analytical tool allowing one to quantify such technologies without having to aggregate multiple outputs into a single output index. In its essence, the output distance function is

* Alexandre Repkine: Associate Professor, Economics Department, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 143-701, Korea. E-mail: repkine@konkuk.ac.kr I thank an anonymous referee for the many constructive remarks that helped to greatly improve this paper.

representing the distance between observed output combinations, and their efficient projections on the production possibilities frontier (PPF) with greater values of the distance function corresponding to less productive efficiency. Distance to the PPF in this context is measured along a vector emanating from the observed output mix, for example a vector (1,1) corresponding to a simultaneous increase in the amounts of both outputs.

Directional output distance functions developed by Fare et al. (2005) are well-suited for the analysis of polluting technologies since the directional vector along which the efficient projection is computed implies an increase in the good output(s), as is the case with the conventional output distance function, along with the *reduction* in the bad output, which corresponds to the idea of reducing pollution levels. In this study we deal with the two-output polluting technology with one good output y, such as GDP, and one bad output b, such as the CO2 emissions levels.

Given the global importance of reducing pollution levels, computing the costs of such reduction is a necessary task. One of the more popular approaches undertaken in the literature is to exploit the duality between the output distance function and the revenue function to compute the shadow price of an undesirable output as a slope of the PPF at the efficient projection point for each observed combination of outputs, often referred to as the marginal abatement costs (MAC) of reducing the bad output.

Most empirical studies that estimate output distance functions, directional or not, employ a quadratic approximation to the true distance function whose parameters are estimated by minimizing the sum of individual values of the distance function subject to a number of constraints that reflect the desired properties of the underlying production possibilities set (PPS). The translation property constraint, for instance, makes sure that the estimated quadratic function in two outputs actually has the distance function properties, i.e., it turns into zero if the good output is increased, and the bad output is decreased by the value of the approximating quadratic function times the corresponding component of the directional vector. Monotonicity constraints require the (directional) output distance function to increase in the bad output, and to decrease in the good one. If monotonicity constraints hold, the resulting PPF is upward sloping, reflecting the desirable property of the positive MACs.

An important theoretical property of the PPS of the polluting technology is that its production possibilities frontier, viewed as a function $y = f^{PPF}(b)$ of the level of bad output, has to be concave, i.e., $y''(b) \leq 0$, see, e.g., Fare et al. (1993). In addition, the distance function should inherit the properties of the underlying PPS by being a concave function of both outputs. In general, concavity of a quadratic function does not necessarily imply the concavity of any of its level curves, including the PPF defined by $D(\bullet)=0$, where $D(\bullet)$ stands for a distance function. In this study, we formally prove that due to the translation property imposed on the quadratic approximation to the directional output distance function, concavity of the latter in both outputs implies the PPF concavity, and formally derive the concavity constraints on the PPF and the distance function. We demonstrate that it is the translation property imposed on the parameters of the estimated approximation curves that constrains them to be either parabolas, or straight lines. In the latter case, the PPF concavity is satisfied automatically. However, in the former case the PPF may be approximated by convex parabolas, but we suggest that this can be avoided by imposing concavity constraints.

It is somewhat surprising that the overwhelming majority of existing empirical studies limit the set of constraints to the ones that guarantee positivity of the MACs and the fulfillment of the translation property, along with a few technical constraints. We provide an empirical example of a simulated polluting technology characterized by a concave production possibilities frontier that is estimated to be convex by applying the conventional directional distance function approach. In this way we are demonstrating the danger of not imposing concavity constraints on the estimated parameters of the directional output distance function and, as a consequence, on the PPF.

The empirical study that is the closest to ours appears to be O'Donnell and Coelli (2005) where the authors impose concavity on the PPF by employing a Bayesian estimation framework. Ours is a much simpler approach to resolving the issue of PPF concavity. We also consider the imposition of the null-jointness property on the implied production possibilities set, which precludes technologies allowing "clean" production of a good output, i.e., the possibility of producing positive amounts of a good output with no pollution at all. The null-jointness constraints are not normally imposed in the empirical literature either, which is why in addition to the concavity constraints, we suggest imposing a necessary condition for the null-jointness property.

Here, we strongly argue against interpreting the individual approximating curves as the individual PPFs for each observation since it is only a small part of the (directional) distance function's range, namely the vicinity of the estimated efficient projection point on the PPF that is used as an approximation to the true PPF. The rest of the approximating curve is unrelated to the true PPF, and as such cannot be used for inference on its properties. This implies, for instance, that the null-jointness property can only be imposed locally, i.e., at the observed combinations of outputs, rather than in terms of the parameters of the true PPF that are unknown.

This paper is organized as follows. In the next section we discuss the theoretical framework underlying the directional output distance function approach to the estimation of marginal abatement costs. In Section III we discuss the geometrical properties of the quadratic approximation to the distance function, and derive the constraints ensuring the directional output distance function and PPF concavity, and the necessary null-jointness constraints. In Section IV we discuss our empirical results. Section V concludes.

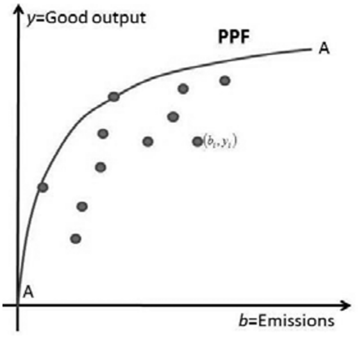
II. Theoretical framework

A. Modeling polluting technologies

Consider $P(\vec{x})=\{(b,y):\vec{x}$ can produce $(b,y)\}$ be the set of all output vectors (b,y) that can be produced using input vector $\vec{x}$ where *b* and *y* are the amounts of bad and good output, respectively. Assume all inputs are strongly disposable: if $\vec{x}'\geq\vec{x}$ then $P(\vec{x})\subseteq P(\vec{x}')$, i.e., increasing the inputs from $\vec{x}$ to $\vec{x}'$ will keep the output mix produced with $\vec{x}$ feasible. Weak disposability assumption is imposed on both outputs, namely, $(b,y)\in P(\vec{x})\Rightarrow(\theta b,\theta y)\in P(\vec{x})$ for any real $\theta\in[0,1]$: it is possible to proportionally scale down the production of both outputs simultaneously. $P(\vec{x})$ is assumed to be compact to make sure infinite amounts of outputs may not be produced with the finite amounts of inputs.

The next two assumptions are specific to the modeling of production sets with bad outputs. Null jointness is imposed in order to preclude the possibility of pollution-free production, i.e., $(y,0)\in P(\vec{x})$ implies *y*=0. Strong disposability is assumed for the desirable outputs, but is denied in case of the undesirable ones. That is, if $(b,y)\in P(\vec{x})$ then $(b,y-\Delta y)\in P(\vec{x}),\Delta y\geq 0$, e.g., it is always possible to reduce the amount of a good output by $\Delta y\geq 0$ while keeping that of a bad output intact, but the converse is not allowed, reflecting the idea of a costly reduction of the amount of bad output. The production possibilities frontier (PPF) function is defined as the maximum amount of good output *y* that can be produced given a particular amount of bad output *b*. The PPF function $y=y^{PPF}(b)$ is required to have a positive first-order derivative to reflect costly reduction of the bad output, and a non-negative second-order derivative. The latter property is based on the assumption of the non-decreasing marginal costs of pollution *reduction*. The property of null-jointness implies that the only way to produce any positive amount of good output *y* is to produce it along with some amount of the bad output b, i.e., $(b,y)\notin P(\vec{x})\forall(b,y)\in\{(b,y):b=0,y>0\}$. The assumptions on the production set outlined above result in a specific shape of the PPF, illustrated in Figure 1 below.

Figure 1. PPF of the production possibilities set with good and bad output



In Figure 1, curve AA is the production possibilities frontier. The PPF is positively sloped, with the magnitude of the slope interpreted as the marginal abatement cost of reducing bad output in terms of the good one. The positively-sloped PPF is the one feature that creates the difference with the PPF shape in case both outputs are good ones, reflecting the fact that, if producing efficiently, producing more of the good output entails creating more pollution. Marginal abatement costs are non-increasing in the amount of bad output, which is equivalent to saying that marginal costs of pollution reduction are non-decreasing in the reduction volume. Alternatively, higher production levels of good output involve creating increasingly more pollution. If that were not the case, i.e., if the PPF as a function of bad output were convex as a function of bad output, rather than concave, we would not have to deal with the problem of increasing environmental damage since after a certain threshold level of good output has been transcended, further increases in good output would only involve minor growth in pollution levels. For that reason, it is crucial that the estimated production set be convex, i.e., that the PPF as a function of bad output be concave.

B. Directional output distance functions

Fare et al. (2005) provided an operational way of parameterizing the production possibilities set described above by introducing the concept of a directional output distance function based on the idea by Shephard (1970) of the output distance function that can be shown to be dual to the revenue function. The key idea is to estimate a quadratic approximation of the directional distance function that maps actually observed output combinations (b_i,y_i) to the distances between (b_i,y_i) and the true PPF. Those distances are measured along a directional vector (hence the name) $(-g_b,g_y), g_b>0, g_y>0$, which we normalize to be a unit vector, i.e., $\sqrt{g_b^2+g_y^2}=1$. The directional distance function $\vec{D}_i\left(b_i,y_i|\vec{x}_i,g_b,g_y\right)$ is defined as follows:

$$\vec{D}_i\left(b_i,y_i|\vec{x}_i,-g_b,g_y\right)=\max\left\{\tau:\left(b_i-\tau g_b,y_i+\tau g_y\right)\in P\left(\vec{x}_i\right)\right\}, \tag{1}$$

where $\vec{x}_i$ is an input vector employed by production unit i, and the vector sign in $\vec{D}_i$ emphasizes the fact that (1) is representing a *directional* output distance function as opposed to just an output distance function. The subscript i in $\vec{D}_i(\bullet)$ above is to emphasize the fact that parameters of the relationship implicitly defined by $\vec{D}_i(\bullet)=0$ will be different depending on the input vector $\vec{x}_i$ so that an implicit function $y = y(b)$ defined by $\vec{D}_i(\bullet)=0$ in the vicinity of a particular observation (b_i,y_i), in case this implicit function exists in that vicinity, will have different parameters compared to the similar implicit function in the vicinity of (b_j,y_j), $j \neq i$.

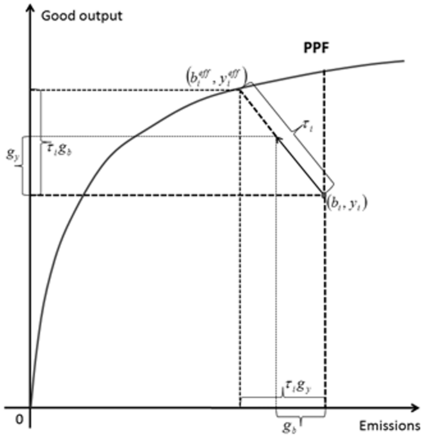
Denoting τ_i to be the solution of the maximization problem in the right-hand side of (1), the directional distance function can be interpreted to project the (inefficient) output mix (b_i,y_i) to its efficient projection $\left(b_i^{eff},y_i^{eff}\right)$, which is a point on the PPF that pointed at by the directional vector $(-g_b,g_y)$ that emanates from the observed (b_i,y_i), and where $b_i^{eff}=b_i-\tau_i g_b, y_i^{eff}=y_i+\tau_i g_y$ (see Figure 2 below).

Consider a movement from the observed output mix (b_i,y_i) towards the PPF along the directional vector $(-g_b,g_y)$ by simultaneously decreasing bad output by τg_b, while increasing the amount of good output by τg_y, where $\tau \geq 0$. Geometrically that means that we will get closer to the PPF by the distance of $\sqrt{\tau^2 g_b^2+\tau^2 g_y^2}=\tau$ since the directional vector $(-g_b,g_y)$ is assumed to be normalized to unity. By definition of the directional distance function, the latter then must satisfy the translation property:

$$\vec{D}_i\left(b_i,y_i|\vec{x}_i,g_b,g_y\right)-\vec{D}_i\left(b_i-\tau g_b,y_i+\tau g_y|\vec{x}_i,g_b,g_y\right)=\tau,\forall\tau\geq 0, i=1,...T, \tag{2}$$

where T is the number of observations. Since negative τ's correspond to the directional vector collinear with $(-g_b,g_y)$, but looking in the opposite direction, the directional distance function will assume negative values for all output combinations that do not belong to the underlying production possibilities set. Figure 2 illustrates the idea of the directional output distance function.

Figure 2. Directional output distance function



C. Empirical estimation of directional output distance functions

It became common in the empirical literature to estimate a second-order approximation to the distance function (1) subject to a set of constraints ensuring the fulfillment of the translation property and the positive sign of the implied PPF slope, which is linked to the distance function as follows:

$$\left.\frac{\partial y}{\partial b}\right|_{(b_i-\tau_i g_b, y_i+\tau_i y_b)} = -\left.\frac{\partial \vec{D}_i \big/ \partial b}{\partial \vec{D}_i \big/ \partial y}\right|_{(b_i, y_i)}, \tag{3}$$

where $\vec{D}_i = \vec{D}\left(b_i, y_i \middle| \vec{x}_i, g_b, g_y\right)$. This approach has been adopted in studies such as, e.g., Lee (2011), and Maradan and Vassiliev (2005). Equation (3) says that the negative of the ratio of the two partial derivatives of the distance function at the observed output mix (b_i, y_i) is equal to the PPF slope at the efficient projection of (b_i, y_i). This slope can be interpreted as the marginal abatement cost of reducing bad output by one unit incurred at (b_i, y_i). Fare et al. (2005) used duality theory to demonstrate that the right-hand side of (3) is equal to the ratio of the shadow prices of the two outputs.

A substantial number of empirical studies assuming the directional output distance function approach to the estimation of marginal abatement costs of reducing bad output employ the following empirical specification:

$$Min\sum_{i=1}^{T}\vec{D}_i - 0,$$
$$\vec{D}_i = \alpha_0 + \sum_{n=1}^{K}\alpha_n x_{in} + \beta_1 y_i + \gamma_1 b_i + \frac{1}{2}\sum_{n=1}^{K}\sum_{n'=1}^{K}\alpha_{nn'}x_{in}x_{in'} + \frac{1}{2}\beta_2 y_i^2 + \frac{1}{2}\gamma_2 b_i^2 + \sum_{n=1}^{K}\nu_n x_{in} b_i + \sum_{n=1}^{K}\delta_n x_{in} y_i + \mu b_i y_i, \quad (4)$$
$$\vec{D}_i \geq 0, \frac{\partial \vec{D}_i}{\partial y} \leq 0, \frac{\partial \vec{D}_i}{\partial b} \geq 0, \vec{D}\left(\vec{x}, b-\tau g_b, y+\tau g_y \middle| g_b, g_y\right) = \vec{D}\left(\vec{x}, b, y \middle| g_b, g_y\right) - \tau, \tau \in \Re, \alpha_{nn'} = \alpha_{n'n}, n = 1..K,$$

where T is the number of observations and K is the number of inputs.

The first constraint requires that the observed output mixes belong to the production set, the next two constraints ensure the positive slope of the PPF at the observed output mix, the next to last constraint ensures the translation property is satisfied, and the last constraints are the technical constraints on the coefficients' symmetry. The constraints for the estimated directional output distance function to satisfy the translation property for an arbitrary directional output vector were derived in Fare et al. (2006). We reproduce them below for further reference:

$$\begin{aligned} &\beta_1 g_y - \gamma_1 g_b = -1, \\ &\delta_n g_y - \nu_n g_b = 0, n = 1,...,K, \\ &\beta_2 g_y = \mu g_b, \\ &\gamma_2 g_b = \mu g_y. \end{aligned} \quad (5)$$

III. The geometry of second-order approximation to the directional output distance function

A. Quadratic approximation to the distance function as parabolas or straight lines

Solving the linear program (4) results in a quadratic function in K+2 variables, namely, the two outputs (b,y) and the K inputs $\left(x_1,...,x_K\right) = \vec{x}$, which implies that observed input vectors $\vec{x}_i$ for each observation determine a second-order curve $\vec{D}_i\left(b, y \middle| g_b, g_y, \vec{x}_i\right) = 0$ in two outputs of the following form:

$$\begin{aligned} &\vec{D}_i\left(b, y \middle| g_b, g_y, \vec{x}_i\right) = \left(\frac{1}{2}\beta_2\right)y^2 + \left(\frac{1}{2}\gamma_2\right)b^2 + \mu b y + \left(\gamma_1 + \sum_{n=1}^{K}\nu_n x_{in}\right)b \\ &+ \left(\beta_1 + \sum_{n=1}^{K}\delta_n x_{in}\right)y + \left[\alpha_0 + \sum_{n=1}^{K}\alpha_n x_{in} + \frac{1}{2}\sum_{n=1}^{K}\sum_{n'=1}^{K}\alpha_{nn'}x_{in}x_{in'}\right] = 0, \\ &i = 1,...,T. \end{aligned} \quad (6)$$

The quadratic function (6) may define several second- and first-order curves in the space of two variables b and y such as, e.g., a parabola, an ellipse, a hyperbola, or a straight line. In the Online Appendix we prove the following proposition:

Proposition 1.

(a) Due to the translation property, for each input vector $\vec{x}_i$ *corresponding to the observed* (b_i,y_i) *the relationship* (6) *is implicitly defining a relationship between good output y and bad output b* $\vec{D}_i\left(b,y|g_b,g_y,\vec{x}_i\right)=0$ *that is either a parabola or a straight line.*

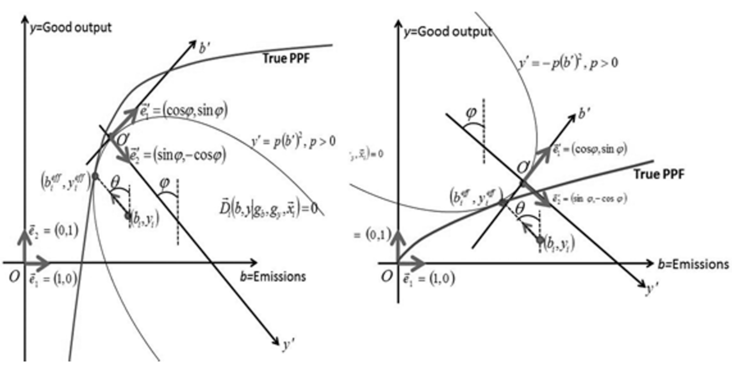
(b) In case $\vec{D}_i\left(b,y|g_b,g_y,\vec{x}_i\right)=0$ *is a parabola, the axis of symmetry of each suchparabola for all i* $=1,\ldots,T$*is forming the same angle* φ *with the vertical axis, i.e.,* $\varphi_i=\varphi_j \forall i,j=1,\ldots,T$.

(c) The angle θ *between the directional vector* $(-g_b,g_y)$ *and the vertical axis is linked to the angle* φ *as follows:* $\theta=\frac{\pi}{2}-\varphi$.

(d) Concavity: For the approximated PPF at the projection point $\left(b_i^{eff},y_i^{eff}\right)$ *to be concave, it is sufficient to require for each observation that:* $sign(\mu)=-sign(\gamma_1 g_b-\beta_1 g_y)$.

Since in the linear case (i.e., the one when $\beta_2=\gamma_2=0$), concavity of the approximating curve $\vec{D}_i\left(b,y|g_b,g_y,\vec{x}_i\right)=0$ is guaranteed automatically, let us consider the parabolic case, i.e., the one for which $\beta_2 \neq 0$ or $\gamma_2 \neq 0$. In the Online Appendix we prove that due to the translation property the relationship (6) implicitly defines either a parabola of the form $y'=p(b')^2, p>0$ or a parabola $y'=-p(b')^2, p>0$,

Figure 3. Approximating parabolas: concave and convex cases



Note: Angles θ and j are linked by the relationship $\varphi=\frac{\pi}{2}-\theta$.

where (b', y') are the parabolas' coordinates in a new coordinate system, and p is a positive number that is a function of the distance function's parameters *and* the input levels for a particular observation. The new coordinate system is obtained from the original coordinate system by a parallel shift and rotation of the latter, as demonstrated by Figure 3 above. As we demonstrate below, the failure to impose the concavity constraints in part (d) of Proposition 1 may result in a convex approximation to the concave underlying PPF, which is undesirable on theoretical grounds.

B. Concavity of the directional output distance function and ppf concavity

It follows from the PPF concavity conditions in part (d) of Proposition 1 that the PPF concavity is equivalent to concavity of the directional output distance function in both outputs. Indeed, the distance function's Hessian at the observed combination of outputs is identically equal to zero for all possible directional vectors $(-g_b, g_y)$. The Hessian $H = \begin{vmatrix} \beta_2 & \mu \\ \mu & \gamma_2 \end{vmatrix} = \beta_2\gamma_2 - \mu^2 \equiv 0$, since by translation property in (5), $\beta_2\gamma_2 \equiv \mu^2$ irrespectively of the directional vector. It follows that the distance function will be concave (but not necessarily strictly concave) at all of the observed output combinations if $\beta_2 \leq 0$ or, equivalently, if $\gamma_2 \leq 0$. The latter two conditions are equivalent to the condition on PPF concavity since by the translation property, $\beta_1 g_y - \gamma_1 g_b = -1$, and thus the PPF concavity condition $\text{sign}(\mu) = -\text{sign}(\gamma_1 g_b - \beta_1 g_y)$ reduces to $\text{sign}(\mu) = -\text{sign}(1)$, or $\mu \leq 0$. Since by the translation property the coefficients μ, β_2 and γ_2 are of the same sign, concavity of the directional distance function in both outputs implies the PPF concavity, and the other way round.

It is worthwhile noticing that in general, the concavity of a quadratic function does not guarantee the concavity of any of its level curves, of which the PPF is one. In case the polluting technology produces one good and one bad output, the translation property ensures the equivalence of concavity conditions imposed on the parameters of the directional output distance function, and the PPF. It is the subject of our further research to see whether the two concavity requirements remain equivalent in the more general case of more good and bad outputs.

C. Implications for inference on the true PPF parameters

Irrespectively of the estimated parameters, in general the quadratic function (6) is implicitly defining more than one curve linking b and y. This happens since the coefficients for b and y, i.e., the coefficients $\gamma_1 + \sum_{n=1}^{K} v_n x_{in}$ and $\beta_1 + \sum_{n=1}^{K} \delta_n x_{in}$, as well as the term $\alpha_0 + \sum_{n=1}^{K} \alpha_n x_{in} + \frac{1}{2}\sum_{n=1}^{K}\sum_{n'=1}^{K} \alpha_{nn'} x_{in} x_{in'}$, depend on the observation-specific values of production inputs $x_{i1}, x_{i2}, x_{i3}, i = 1..T$. Equation (6) in this way is defining a family of T curves, one for each observed combination of outputs (b_i,y_i).

It is important to notice that it is only the small part of the approximating parabola defined by (6) in the vicinity of the efficient projection $\left(b_i^{eff}, y_i^{eff}\right)$ that is approximating the true PPF parameters and whose slope at that projection is used for the estimation of MACs. Outside of this vicinity this parabola cannot be considered to be an efficient boundary of the production possibilities set for observation i. As a result, the estimated parameters of the directional output distance function in (6) cannot be used for inference on the true PPF parameters.

D. Implications for the null-jointness property

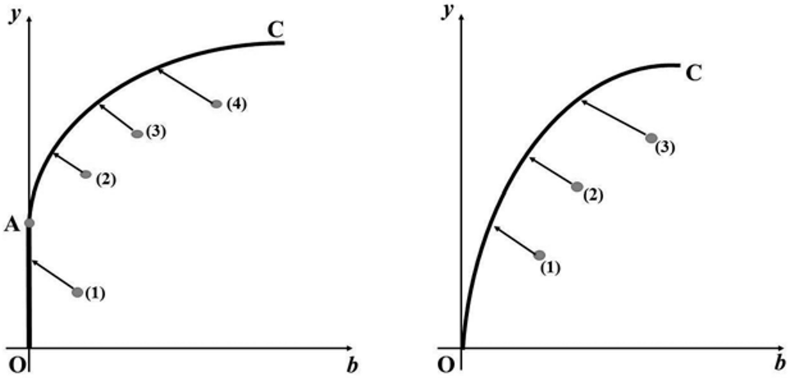
The null-jointness property is imposed on the PPS in order to preclude the possibility of producing a positive amount of good output while emitting zero pollution levels. Figure 4.A below depicts a PPS whose efficient frontier *OAC* does not satisfy the null-jointness property since it includes segment *OA*, which implies it is possible to produce positive levels of good output up to level *A* without emitting any pollution at all. The efficient frontier *OC* in Figure 4.B, on the other hand, satisfies the null-jointness property because pollution-free production of good output is impossible.

If it were possible to estimate the parameters of the directional distance function for all combinations (b,y), imposing the null-jointness property would boil down to requiring the value of the directional output distance function to be negative for all output combinations that satisfy $b = 0$ and $y = 0$. However, as discussed in the previous sub-section, the quadratic function in (6) is only approximating the true distance function in the close vicinity of $\left(b_i^{eff}, y_i^{eff}\right)$, which is not necessarily close to the locus of output combinations characterized by $b = 0$ and $y = 0$. Hence, the estimated parameters in (6) cannot be used to impose the null-jointness property.

Nevertheless, it appears possible to impose a *necessary* condition for the null-jointness property. Namely, as illustrated by Figure 4.B, if a PPS satisfies the null-jointness property, its efficient frontier must be such that all of the projections on it of the observed output combinations must satisfy $b_i^{eff} > 0$. This condition is not sufficient since, as can be inferred from Figure 4.A, if the sample of observations were limited to (2), (3) and (4), the efficient levels of bad output would be positive, while the PPS itself would not satisfy the null-jointness property. We suggest adding the set of constraints $b_i^{eff} > 0$ in addition to the concavity constraints as a second-best solution to the problem of ensuring the satisfaction of the null-jointness property.

Figure 4. Necessary condition for null-jointness property

A. Null-jointness property violated B. Null-jointness property satisfied



IV. Empirical exercise

A. Simulation

Consider an industrial plant producing two outputs: a good output y and a polluting output b. Let there be three factors of production: capital k, labor l, and fuel f. Assuming a Cobb-Douglas production function for the good output and a linear production function for the bad output, the PPF $y^* = f(b)$ will be a solution to the following optimization problem:

$$\begin{aligned} &\underset{k,l,f}{Max}\, y = k^{\alpha} l^{\beta} f^{\gamma} \\ &s.t. \;\; \delta k + \varphi l + \psi f = b, \end{aligned} \tag{7}$$

where the Greek letters stand for the real positive-valued parameters. Maximizing the Lagrangian results in the following solutions for the optimal levels of capital, labor and fuel as functions of the two production functions' parameters and the pollution level b:

$$k^* = \frac{\alpha}{\delta}\frac{b}{\alpha+\beta+\gamma},$$
$$l^* = \frac{\beta}{\varphi}\frac{b}{\alpha+\beta+\gamma}, \tag{8}$$
$$f^* = \frac{\gamma}{\psi}\frac{b}{\alpha+\beta+\gamma}.$$

Substituting these optimal levels of factor inputs back into the Cobb-Douglas production function in (7), one obtains a parametric expression for the PPF:

$$y^* = \left(\frac{\alpha}{\delta}\right)^{\alpha}\left(\frac{\beta}{\varphi}\right)^{\beta}\left(\frac{\gamma}{\psi}\right)^{\gamma}\frac{b^{\alpha+\beta+\gamma}}{(\alpha+\beta+\gamma)^{\alpha+\beta+\gamma}} = Z\times b^{\alpha+\beta+\gamma}, \tag{9}$$

where Z is a positive real number. The PPF in (9) will be concave if $\alpha+\beta+\gamma<1$.

In performing our simulation exercise, we set $\alpha=\beta=\gamma=0.25$, and $\delta=\varphi=\psi=0.33$, which according to (9) results in the PPF function $y^* = f(b) \approx 1.01\times b^{0.75}$ that is obviously concave in b. We generate a sample of uniformly distributed values of bad output $b\in(0,1)$ and use (8) and (9) to compute optimal values of the three production factors and the PPF value y^*. To introduce random variation into the dataset, we added a random noise to k^*, l^*, f^* and y^*. Table 1 below contains summary statistics for the simulated dataset.

In Table 2 below we present the results of our estimations of the parameter μ in the directional output distance function specification (6) whose sign determines the type of the approximating parabola. The estimation was done by linear programming in the SAS environment. In the column "Benchmark Model" we report estimation results of the model in (4) without imposing concavity and null-jointness constraints. In the next column we report results for a model with these

Table 1. Summary statistics for the simulated dataset

	Mean	Standard deviation	Minimum	Maximum
Good output, y	1.07	0.31	0.32	1.56
Bad output, b	0.55	0.29	0.003	0.99
Capital, k	0.78	0.34	0.08	1.41
Labor, l	0.81	0.35	0.04	1.44
Fuel, f	0.81	0.34	0.11	1.45

Note: the number of observations is 77, the simulation was performed in Excel using the uniform random number generator function rand().

Table 2. Concavity of the estimated distance function and PPF depending on the type of imposed constraints

	Parameter μ estimates	
Angle φ, degrees	Benchmark model	Concavity and null-jointness constraints
95	0.000	0.000
90	0.000	0.000
85	0.000	0.000
80	0.000	0.000
75	0.003	0.000
60	0.023	-0.005
55	0.027	-0.018
50	0.023	-0.023
45	0.016	-0.280
40	0.002	-0.325
35	-0.451	-0.220
30	-0.521	-0.182
25	-0.135	-0.883
20	-0.009	-0.146
15	-0.009	-0.021
10	0.000	0.000
5	0.000	0.000

Note: Angle φ is formed by the directional output vector with the vertical axis. Parameter μ in the directional output distance function specification is negative for the concave PPF.

constraints added to the benchmark model. We run our estimation for a series of the directional output vectors. Since we know that the underlying true PPF for this dataset is concave in b, according to Proposition 1 and the discussion in Section III, parameter μ should be estimated to be negative.

We observe that for the directional output vectors that are either too flat or too steep (i.e., for $\varphi > 75°$ and for $\varphi < 15°$) both the distance function and the PPF are estimated to be of zero curvature. The case of the most "popular", 45-degree directional output vector, is rather instructive in that if we do not impose any concavity constraints, the PPF is estimated to be convex. Imposing concavity and null-jointness constraints, the estimates are consistent with the known properties of the simulated dataset.

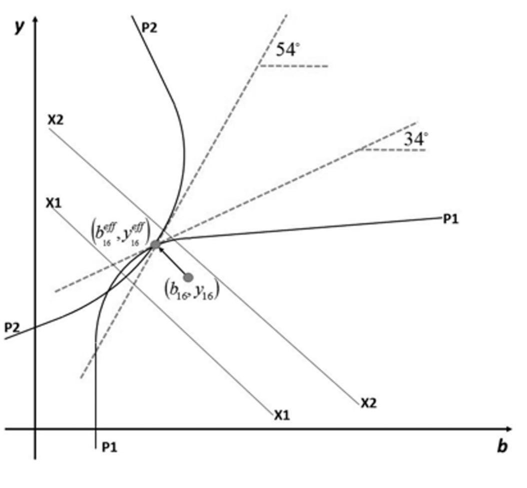
For a range of directional vectors for which $35° < \varphi < 75°$, estimating the parameters of the directional output distance function without imposing concavity and null-jointness constraints results in the wrong inference about the curvature type of the underlying PPS. Imposing these constraints, however, allows one to avoid concluding that the directional distance function is convex and its PPF is convex in b.

We illustrate in Figure 5 below what happens when concavity and null-jointness constraints are added to the estimation procedure (4). We analyze observation number 16 in the simulated dataset for the case of the "conventional" directional vector $\varphi = 45°$.

(b_{16},y_{16}) is the combination of the values of the good and bad output in our simulated sample. The true PPF is approximated at the efficient projection $\left(b_{16}^{eff},y_{16}^{eff}\right)$ taken along the 45° directional vector by the convex part of the parabola *P2-P2* in case the concavity and null-jointness constraints are not imposed, see Table 2. In this case the slope of the PPF is estimated to be 54°. The imposition of concavity and null-jointness constraints results in the approximating parabola's branches looking in the South-Eastern direction with the concave part of the parabola *P1-P1* approximating the true PPF at $\left(b_{16}^{eff},y_{16}^{eff}\right)$. The PPF slope in this case is estimated to be equal to 34°. By Proposition 1, the symmetry axes *X2-X2* and *X1-X1* of the two parabolas are parallel to each other and intersect the vertical axis at the angle $\varphi = 45°$, which is the same with the directional vector angle.

In case the concavity and null-jointness constraints are not imposed, not only the curvature type of the approximating parabola may be estimated wrongly, but also the MACs are likely to be different in the two cases. Indeed, the MAC for the

Figure 5. The effects of imposing concavity and null-jointness constraints



convex parabola is estimated at 1.36, while the MAC is 0.68 when concavity and null-jointness constraints are imposed, which is also closer to the actual MAC value of 0.71 which is known because the exact shape of the true PPF is known.

B. Suggestions for the choice of directional output vectors

It is interesting to notice that in the overwhelming majority of the empirical studies applying the directional output distance function approach to the estimation of marginal abatement costs, the 45-degree directional output vector is used almost "by default." Our simulation exercise demonstrates that the choice of this and, in fact, any other vector without imposing any additional constraints may result in the wrong inference on the curvature of the PPF and the directional output distance function. We therefore suggest choosing the directional output vector based on the prior knowledge of the PPS properties, such as the type of curvature of its PPF. We would therefore reject the choice of those directional output vectors for which the directional output distance function and, consequently, the PPF, are estimated to be convex when there are strong reasons to believe that they are concave.

V. Conclusion

In this study we explored the geometric properties of quadratic approximation to the directional output distance function commonly used in empirical research on the marginal abatement costs of the undesirable outputs. The translation property imposed on the estimated parameters of the quadratic approximation ensures that the true PPF is approximated at the observed combinations of outputs either by parabolas, or by the straight lines. While economic theory implies that both the directional output distance function and the PPF for a polluting technology $y = f^{PPF}(b)$ have to be concave, the corresponding concavity constraints are usually not imposed. We demonstrate the danger of omitting such constraints by simulating a sample of observations based on a concave PPF and running a conventional estimation procedure on it to demonstrate that the true PPF that is known to be concave in b may be approximated by the convex, rather than the concave, parabolas, which is contradicting economic theory. By adding additional constraints ensuring concavity of the approximating second-order curve and a necessary condition for the null-jointness property, we observe a significant change in both the parameters of the estimated approximation to the directional output distance function, and the marginal abatement costs.

We believe an important insight in this paper is that it is erroneous to interpret the second-order curve in b and y obtained by plugging in the observed values of production inputs into the estimated quadratic specification of the directional output distance function as an individual PPF. Indeed, the behavior of the approximating curve outside of the small vicinity around the efficient projection of the observed output mix provides little inference on the behavior of the true PPF.

Although the choice of a particular directional vector is not the main subject of this study, we would suggest rejecting those vectors for which the estimated parameters of the directional output distance function imply properties of the underlying PPS that are in conflict with one's prior knowledge about it. Thus, in our simulation example, since we know in advance that the PPF for the simulated PPS is concave in the bad output b, it would be rational not to employ vectors corresponding to $\varphi \geq 75°$, i.e., the relatively flat vectors, because for those vectors the implied PPF is estimated to be convex rather than concave.

Finally, it is worthwhile asking whether our analysis will change in the multi-output case that involves several bad outputs, which is not uncommon since polluting technologies tend to produce more than a single type of harmful emission. While the directional output distance function can be easily extended to that case with the translation property constraints already derived for the multi-output case in Fare et al. (2006), the properties of the Hessian corresponding to the multi-output directional output distance function are not obvious. For instance, it is not clear whether the translation property will ensure the Hessian to be identically equal to zero, like we show is the case of the two-output polluting technology. It is not clear either whether translation property in the multi-output case will guarantee the equivalence between concavity of the directional output distance function and that of the PPF. If this equivalence is not guaranteed, the implication for the empirical work will be to impose two groups of concavity constraints: one ensuring the concavity of directional output distance function, and the other one providing for the PPF concavity. The analysis of these issues is outside the scope of this study, being the subject of our ongoing research.

References

Fare, Rolf, Shawna Grosskopf, C.A. Knox Lovell, and Suthathip Yaisawarng (1993). Derivation of shadow prices for undesirable outputs: A distance function approach. *Review of Economics and Statistics* 75: 374–380.

Fare, Rolf, Shawna Grosskopf, and William Weber (2006). shadow prices and pollution costs in U.S. agriculture. *Ecological Economics*, 56: 89–103.

Fare, Rolf, Shawna Grosskopf, Dong-Woon Noh, and William Weber (2005). Characteristics of a polluting technology: Theory and practice. *Journal of Econometrics* 126: 469–492.

Fedorchuk, Vitaly V. (1990). *A course in analytical geometry and linear algebra*. Moscow State University Press, Moscow.

Lee, Myunghun (2011). Potential cost savings from internal/external CO_2 emissions trading in the Korean electric power industry. *Energy Policy* 39: 6162–6167.

Maradan, David, and Anatoly Vassiliev (2005). Marginal costs of carbon dioxide abatement: Empirical evidence from cross-country analysis. *Swiss Journal of Economics and Statistics* 141: 377–410.

O'Donnell, Christopher J., and Tim J. Coelli (2005). A Bayesian approach to imposing curvature on distance functions. *Journal of Econometrics* 126: 493–523.

Shephard, Ronald W. (1970). *Theory of cost and production functions*. Princeton University Press, Princeton.