# Don't Stop Me Now: Gender Attitudes in Academic Seminars Through Machine Learning

Mateo Seré*

Oct 2022

**Draft version**

**Abstract**

This paper focuses on the interactions between peers by building on a widely cited finding in the gender literature that establishes that men interrupt women more than women interrupt men. For that I use audio recordings from economic seminars and I identify all the different speakers that intervene on it and their gender. I find that (i) females are more interrupted than males when presenting and also are interrupted earlier in the seminar; (ii) this is explained to a large extent by interruptions made by women in the audience rather than by men; (iii) the way that men and women in the audience use to interrupt female speakers and the content of their interruptions is significantly different; (iv) having a female chairing the seminar does not affect the number of interruptions made by women but reduces the overall number of interruptions made by males. These results are robust when I control by affiliation, seniority and ranking of the department to which the presenter belongs as well as topic of the presentation and seminar series.

*Keywords*: gender, academic environment, machine learning, audio processing

*JEL-code*: A1, C8, C45, J4, J7

*Center for Social Policy (University of Antwerp), Department of Economics (Pompeu Fabra University), Barcelona School of Economics (BSE) - mateo.sere@uantwerpen.be

# 1 Introduction

It has been well established that men and women diverge in their attitudes towards different life dimensions such as work, education investment, family arrangements or financial decisions. This is stark on individuals in professions with high skills and effort requirements but with risky career outcomes, such as academic ones, in which women are still under-represented. One way in which those attitudes manifest are in the interactions between peers and in particular through the conversation dynamics. What is the nature of those dynamics in environments with such a large gender imbalance? In this domain, one of the findings most widely cited as well established is that men interrupt women more than women interrupt men. This paper will show evidence disputing this conclusion.

I accomplish this by exploiting the *virtualization* of the majority of the academic activity due the 2020 COVID-19 outbreak. Among those activities, academic seminars -one of the main instances for researchers to present their findings but also a place for socializing among faculty members- were not only moved to a virtual setting, but also made publicly available on YouTube later. This is used to construct a 2,154 economics seminars database, with talks that were web-streamed between 2020 and 2022 and belong to the top 320 best ranked universities worldwide.

This paper is built on those talks with goal of studying whether female presenters are interrupted more than their male counterparts. The main information gathered from the recordings is assembled in two steps. First, I use an audio processing technique known as "speaker diarisation" which consists on estimating the number of speakers in an audio stream associating each speech segment with a speaker (Dadvar 2011). Second, once the composition of "who spoke when" is done, I identify the gender of each of the speakers of the seminars based on their voice. These types of techniques are commonly applied by daily life devices like mobile phones or in domains in which audio data is needed but its use in the social science domain is rare and a novelty of this paper. [1]

I complement this with information of seminar's length, speaker's university, number of citations, seniority, academic interests and other relevant information that is available in Google Scholar and RePEc. The text transcripts of each seminar are used in the identification of the topic of the presentation and to gather information about the content of each interruption.

As here the term interruption is constructed by considering every change of speaker in a talk,

[1] Example of the use of these techniques in common life activities are the virtual assistant of our phone or when we want to distinguish between a doctor's questions and a patient's responses.

I take variations of it to complement the main results. From the number of total interruptions I also use only those that are made in an interrogative way. In addition it is considered the way in which the interruption is made (by a smooth change of voices or by overlapping an existent voice) and whether the interruption is made by the chair of the seminar or by someone else in the audience.

This paper contributes to the literature on how individuals' behavior is affected by the presence of peers by means of three key results. First, it shows that female presenters are more interrupted than males in seminars. On average they receive around one extra interruption per seminar. Second, it shows that those extra interruptions received by female presenters are in its great majority not due to men in the audience but to females. Being a female presenter significantly increases the number of overall interruptions that women in the audience do. Same occurs when from the overall interruptions it is only considered those that were labeled as proper questions. By contrast being a female presenter is not significant to explain the total number of interruptions made by males in the audience and reduces the number of questions that they do.

As a way of reinforcing this it is shown that those extra interruptions received by female presenters take places in seminars with higher share of women in the audience. As the latter is not directly observed, different proxy variables of this are used. This reads as the number of interruptions increases when the presenter is a female presenting in seminars where the share of women in the audience is larger.

The previous results are robust to whether the chair of the seminar is considered or not for the overall number of interruptions. However, in third place it is shown how the number of interruptions received by a presenter varies according to the gender of the chair of the seminar. A female chairing the seminar reduces the overall number of interruptions in a seminar and turns the gender of the presenter less significant to explain them. This reduction is due to a decrease in the number of interruptions made by males in the audience rather than by females.

way in which the interruption takes place and the relationship between the content of the interruption and content of the speech that is interrupted.

As a matter of fact, the implications of an interruption or a question are not straightforward. No interruptions may imply a captivated audience but also an uninterested one. Moreover, even if future research could distinguish for example the tone of an interruption, it should be noted that the same interruption can be perceived differently according to the age, position or the gender of the receiver.[2] However, as pointed by Zimmermann and West (1996) conversational interaction may be related to enduring problems of power and dominance in social life, this work also seeks to contribute to this literature by analyzing the "how" and the "what" of an interruption. The "how" refers to whether the interruption occurs through a smooth transition or in contrast, by a speech overlap. I find that men in the audience reduces the number of overlap interruptions when the presenter is female while women increases it. As early works on conversational interaction presented this simultaneous talk as a negative and dysfunctional act this result may sounds surprising. However, more recent research has signaled how this simultaneous talk is used to signal and promote solidarity between speakers. Furthermore, two underlying reasons have been proposed to explain this type between women: the higher legitimacy that a female may feel to attempt to take the floor from other women but also that this is used as a supportive and rapport-building function which gives a sense of cooperation (Dindia 1987).

Considering the "what" in the interruption, I study the similarity between what it is said in the interruption and what it was being said prior the interruption. The results shows a decrease in the speech similarity between male interruptions and female presenters. Even if modest this negative coefficient indicates that men interruptions are slightly less related to the content of the stream that is being interrupted when the speaker is a woman.

These results also looks to add a different perspective to the recent booming literature that has found evidence of gender discrimination at virtually every stage of the academic profession in economics, from undergraduate enrollment to tenure decisions. Paredes et al. (2020) provides evidence on how undergraduate students exhibited more gender bias after studying economics. This effect is stronger among male students and weaker in departments with more female faculty members. Wu (2018) attempts to assess the existence of an unwelcoming or stereotypical culture using evidence on how women and men are portrayed in online forums. Hengel (2022) finds that women are held to higher standards of writing and research than their male colleagues and Card et al. (2020) that female-authored papers in top economics journals are held to different standards than the male-authored ones. While this can be seen as part of a larger and more complex problem, academy economics exhibits greater disparities than those observed in the social sciences overall. Considering tenured academic jobs as example, female economists are 20 p.p less likely to have it 10 years after PhD. receipt which contrast with –3.9 percent gap that

[2] The work of Hilton and Jeong (2019) and Hilton (2018) offers a recent discussion about the systematic disparities on how individuals from different groups perceive interruptions.

favors them in engineer and 8.1 in the other social sciences fields (Ginther and Kahn 2004).

This literature is also enriched by the recent work of Dupas et al. (2021). They used a hand coded dataset from 420 seminars, mostly in the applied micro field in the top 30 US universities, to study their dynamics. In line with this paper, the two main findings of their work are that women are asked more questions than men during a seminar and the questions asked to women presenters are more likely to be patronizing or hostile.

This paper is composed of 6 sections. The next one presents the data used in this work. Section 3 introduces the machine learning algorithms used for the speaker diarisation and gender recognition. Section 4 presents the econometric model used and the text analysis tools used on the seminars transcripts. Section 5 presents the main results and Section 6 concludes.

## 2 Data

The main data used in this work comes from web-streamed economic seminars organized from 2020 until 2022, which were available on YouTube. With few exceptions these talks can still be found there.[3] To be included in the dataset the seminar has to be part of a seminar series. In addition the seminar series has to be organized or sponsored by the economic department of an American or European based university. Seminars held by leading research institutions in economics as the National Bureau of Economic Reserach (NBER), the American Economic Association (AEA) and the Centre for Economic Policy Research (CEPR) were also considered.

YouTube metadata was also gathered including seminar's title and description, comments and number of likes that the video received and date in which it was posted. Using natural language processing tools developed in Qi et al. (2020) I identify the name of the speaker from the video's title or description. The name of the speaker is used later in Google Scholar to gather affiliation, number of citations, academic interest and year in which the scholar published its first paper. As I do not observe the age of the speaker, I use the years that passed since his or her first publication as proxy of seniority.

Table 1 groups the main summary statistics of the database in four panels. As seminars with only one presenter were considered, the upper-left panel shows that the mean seminar duration is of 62.3 minutes. More than 80% of the seminars are between 45 min and 90 min long. Only

[3] These exception comes from seminars that were streamed in YouTube but deleted afterwards. For example presentations at the NBER Summer Institute are kept online for a two-week window period after each session and deleted afterwards.

3.5% of the seminars have a duration of more than 90 minutes. The upper-right shows that on average, a seminar has almost 11 interruptions, being the majority of them (6.6 interruptions) shorter than 30 seconds. The bottom-left panel reveals the relatively high concentration of of speakers affiliated to top economics departments. Around two thirds of the presenters belong to the top ten economics departments and less than 10% of them are form economics department ranked below position 100. Finally the bottom-right box presents a miscellaneous of indicators. It is worth to notice that in only 8.3% of the seminars no interruptions were registered and that in 10.1% of the seminars the unique interruptor was the host or organizer. The percentage of seminars with a female as presenter are 35.3% of the dataset.

Figure 3 shows how the distribution of the number of questions asked during seminars presented by women is slightly shifted right with respect to the males one. This is similar to the findings of Dupas et al. (2021).

Table 2 and graph 2 shows origin of the economic department to which speakers are affiliated. Around two-thirds of the speakers belong to economics departments located in the US. Of the rest, 14% belongs to European countries (excluding United Kingdom), 11% to the United Kingdom and the remaining to the rest of the world.

Additionally, when the video audio transcripts were available they were downloaded. For the cases in which that was not possible the audio was converted to text through speech recognition following Ravanelli et al. (2021). These transcripts were pre-processed for text analysis by removing punctuation, excess spaces, numbers, misspelled words, and so-called "stop words," which are common words that carry no intrinsic meaning such as "and" or "the.". From the remaining I only kept nouns, adjectives, verbs or adverbs which then were lemmatized to group all inflected forms of a word.[4] As a first step in visualizing this data and gauging a quick idea of the topics of the seminars, Figure 5 shows the word clouds derived from this transcripts for male and female speakers. In word clouds, the font size for each word is proportional to its frequency.

One of the weakness of word clouds is that they do not account for synonyms. Hence, topics for which there are many possible words to express the same thought may be artificially diluted, while niche topics that feature clear buzzwords may be inflated in importance (Sarkar 2019). The next subsection complements this initial scanning with a topic analysis.

[4]For instance, "policies" becomes "policy", "were" becomes "be".

### 2.1 Seminar's topic identification

The main way of identifying the topic of the seminar consisted on a topic extraction based on the audio transcripts of each seminar. As a robustness check I assigned to each seminar the JEL-code corresponding to the main code used by the speaker in his or her previous publications.

#### 2.1.1 Topic identification based on seminar's transcripts

Topic modeling is an unsupervised learning technique designed for extracting distinguishing concepts or topics from a large corpus that has various types of documents (in this case seminars' transcripts). For the seminar's topic extraction the pre-processed transcripts introduced in Section 2 were used as the corpus data. A "topic" consists of a cluster of words that frequently occur together. In this paper, I use MALLET which implements Gibbs sampling and the Latent Dirichlet Allocation (LDA) technique, a generative probabilistic model in which each document is assumed to have a combination of topics similar to a probabilistic (McCallum 2002). Figure 11 in the Appendix presents the coherence score for models trained with different number of topics. As the score seems to increase, even after training the model with 40 topics, I pick the model with 15 topics which gives the highest value before the curve starts to flattening out (Röder et al. 2015).

The most probable words of each topic are presented in the world-clouds of Figure 6. Once the topics are produced it is possible to determine what topic a given seminar is about by finding the topic number that has the highest percentage contribution in that seminar transcript.

**Assigning labels to the topics modeled**

The words produced in each of the topics can also be associated to words of the different categories of the JEL system classification. For that I study which JEL-category is the most similar to each of the produced topics. This will be addressed as a problem of similarity between two set of documents. One set of documents will be based on the 15 topics emerged from the topic analysis and the other set on 18 JEL categories. [5] The distance between a pair of documents will be given by the distance between the vectors formed by the TF-IDF of those documents. The distance between two vectors, each one coming from a different set of documents, is computed using the

[5] The JEL classification contains 20 different categories. However the categories "A. General Economics and Teaching", "Y. Miscellaneous Categories" and "Z. Other Special Topics" were pooled together as "Other topics", a residual category for topics that could not be assigned to the other categories.

cosine similarity score.

Once I compute the distance of each topic to all the JEL categories, I assign to a topic the JEL-code of the category with the shortest distance. Table 3 summarizes the topic modeling results. Of the 15 topics, 3 could not be associated with any JEL-code and labeled as "Not Assigned". Excluding them, the share of the remaining topics across the seminars shows a relatively homogeneous distribution. Topic 10 (F - International Economics) and Topic 4 (E - Macroeconomics and Monetary Economics) have the largest share, being the main topics in 9.8% and 9.2% of the seminars respectively. By contrast, Topic 5 (D - Microeconomics) and Topic 1 (C Mathematical and Quantitative Methods) are the ones with the lowest share, being the main topic in 4.9% and 5.0% of the seminars respectively.

### 2.1.2 Topic identification based on speaker's JEL-code

An alternative way of assigning a topic to a seminar is based on the identification of the JEL-codes which were used most times by the speaker in his or her previous publications. This was obtained matching the speaker's name with a database of more than 714,702 classified papers with author's name and corresponding JEL-code from the Research Papers in Economics (RePEc) database. A total of 1,242 names were found in the RePEc database, which implies the assignment of a JEL-code to 58.2% of the seminars. The bar plot in Figure 7 shows the total number of seminars matched by the different JEL-codes.

Assuming a paper path dependence in which the current paper being presented corresponds to the same field that the previous papers presented by the same author can be an unrealistic assumption, even when the broader JEL-code categories are used. Moreover, only 17.8% of the that were matched with the RePEc database appeared with only one category.

A more realistic approach could consider for example the real JEL-code used for the specific paper presented. However this seems unfeasible at the moment given that an important part of the papers presented in the seminars used for this database are likely of being still not published.

## 3 Machine learning algorithms for audio processing

The processing of the audio data can be divided in two major steps. First it is necessary to construct a map of all the speakers participating in an audio and of the moment in which they intervene, commonly called speaker diarisation. Once this is done, the voice of each of the speakers identified is used to predict their gender.

Both in the speaker diarisation and in the gender prediction I make use of the Mel Frequency Cepstral Coefficients (MFCC) (as well as its first and second derivative). This is one of the most commonly used short-term acoustic signal features for getting information about speaker's vocal tract characteristics (Müller 2021; Anguera et al. 2012). Additional details on how this feature is constructed are provided in Appendix III.

## 3.1 Speaker diarisation

Speaker diarisation is the process of detecting the turns in speech because of the changing of speaker and clustering the speech from the same speaker together, and thus provides useful information for the structuring and indexing of the audio document. The diarisation system used here performs three basic tasks. First, it discriminates speech segments from the non-speech ones. Second, it detects speaker change points to segment the audio data. Finally, it groups these segmented regions into speaker homogeneous clusters (Pulkki et al. 2017). Figure 9 in the Appendix III provides a visual example of this process.

Voice activity detection (VAD) is a binary classification task of inferring which segments of input audio contain speech versus which segments are background noise or silence. This activity improves the quality of the output by masking the effect of silent frames and noise as well as accelerate the signal processing by avoiding extra runs for silent frames. Commonly, to detect and trim off non-speech segments it is possible to relay on the assumption that voiced frames has more energy than silent. Speech is a time-varying and non-stationary signal, but, in a short segment, for example 10-20 milliseconds, the speech signal is nearly stationary. So speech, signal can he split many short segments to be processed (Enqing et al. 2002). As speech adds energy to the signal, high-energy regions of the signal can be associated with voice activity. As the audio signals analyzed for this paper are relatively *clean* this was the process followed to remove non speech frames. A visual representation of this process is presented in Figure x of Appendix Appendix III. Alternatively, recent research effort has been devoted to finding efficient deep-learning-based VAD model architectures. For example Jia et al. (2021) proposed a neural network model constructed with a stack of blocks with residual connections. There, each block is composed from 1D time-channel separable convolutions, batch normalization, ReLU, and dropout layers. This alternative was followed as robustness check without meaningful differences in the final output.

To determine the change points in the audio signal I use a Gaussian Mixture Model (GMM), one of the most popular ways of modeling speech audio data (Moattar and Homayounpour 2012; Cettolo et al. 2005). The number of segments that composes the signal is determined by the Bayesian Information Criterion (BIC). This technique segments the audio signal within a window using a penalized likelihood ratio test of whether the data in the window is better modeled by a single distribution or by two different distributions. The null hypothesis states that there is no speaker change point at time $t_j$. The data Z = X + Y, is modeled by a multivariate Gaussian probability density function with a set of parameters $\theta_Z$ and a log likelihood $L_0$ in the following way:

$$L_0 = \sum_{i=1}^{n_X} logN(x_i|\theta_Z) + \sum_{i=1}^{n_X} logN(y_i|\theta_Z), \tag{1}$$

being $n_X$ and $n_Y$ the numbers of data samples in analysis windows X and Y respectively. Under the alternative hypothesis a speaker change point exists at $t_j$ and the windows X and Y are modeled by two multivariate Gaussian densities. In this case each density has its own set of parameters, $\theta_X$ and $\theta_Y$. The log likelihood $L_1$ is obtained by

$$L_1 = \sum_{i=1}^{n_X} logN(x_i|\theta_X) + \sum_{i=1}^{n_X} logN(y_i|\theta_Y). \tag{2}$$

The set of parameters $\Theta$ is estimated via the Expectiation Maximization (EM) algorithm. This procedure works with two adjacent sliding windows on the audio data, compute a distance between them, then decide whether the two windows originate from the same speaker. The dissimilarity between the two neighboring is estimated here by ΔBIC, defined as:

$$\Delta BIC = L_1 - L_0 - \lambda R, \tag{3}$$

where R represents the penalty term to compensate the excess of parameters under the alternative hypothesis model with respect to the null hypothesis and $\lambda$ a fine tuning or penalty factor. If ΔBIC is positive, a local maximum is found and time $t_j$ is considered to be a speaker change point. In other case there is no speaker change point at time $t_j$. This process is repeated along multiple samples per analysis windows as candidate boundaries.

The third step consists in clustering the identified audio segments. Here speaker homogeneous segments obtained from the speaker change detection step are grouped according to the hypothesized identity of the speaker. Traditionally this is accomplished by hierarchical clustering algorithms. However, given the relatively homogeneous sequence of audio signals that are analyzed, in which one main intervention is normally followed by short and not overlapped small set of interruptions, another approach is used here. The strategy followed consists in comparing the speaker similarity between all the different possibles pairs of clusters. When the voice at segment $n$ is different to the ones of segments $\{1, ..., n-1\}$, a new id number is assigned to that speaker. In other case, the speaker of that segment receives the id number of the speaker whose voice it resembled when compared to the previous segments. This process is done following Ravanelli et al. (2021) and Desplanques et al. (2020) in which a system composed of an ECAPA-TDNN model is used to compute voice similarities. This is done by means of a combination of convolutional and residual blocks in which embeddings are extracted using attentive statistical pooling. The system is trained using data from Nagrani et al. (2020) which provides short audio clips of human speech, extracted from interview videos uploaded to YouTube. Speaker similarity is performed using cosine distance between speaker embeddings.

### 3.2 Gender recognition

To predict gender I use Mozilla's Common Voice Dataset (Ardila et al. 2020), the largest publicly available corpus of speech data in which gender of the speaker is annotated on each audio track. After cleaning and filtering audios, there were 6,995 male audio files and 5,662 female audio files in the dataset.

In addition to the already mentioned MFCC, previous works on gender recognition based on audio speech relies on the MEL Spectrogram Frequency, another widely used audio feature used for audio processing for which additional details are provided in Appendix II.

As common in the literature (Alnuaim et al. 2022; Chachadi and Nirmala 2022) I use a deep feed-forward neural network with five hidden layers. As a regularization I use a 30% dropout rate, which is one of the most effective and commonly used techniques in neural networks for regularization (Chollet 2021). The pretrained model achieved an accuracy of 90.95% similar to previous works reported with the same dataset (see Chachadi and Nirmala 2022 for an overview).

## 4 Econometric model

I first examine the relationship between the number of interruptions received by the presenters in a seminar and their gender. For that I use the linear specification

$$Y_i = \beta_0 + \beta_1 \text{FemalePresenting}_i + X\gamma + Z\lambda + \epsilon_{i,YT}. \tag{4}$$

in which the variable $Y_i$ is the number of interruptions in a given seminar after the presenter started the presentation. $\beta_1$ shows the effect of being a female presenting in the seminar. A positive value of $\beta_1$ would indicate that being a female presenter leads to a higher number of interruptions. Additionally $X$ is a vector of presenter related characteristics as citations, ranking of the affiliated university, research interests and year of his or her first publication, which is used as a proxy of seniority. The vector $Z$ included seminar related variables as its duration, gender of the interrupters and topic of the presentation. The stochastic error term is $\epsilon_{i,YT}$.

This same specification is used to explain how much time passes before the first interruption in a given seminar takes place. In this case the variable $Y_i$ is the number of minutes from the moment in which the presenter starts to speak until the first interruption. In this case a negative value of $\beta_1$ would imply that being a female is associated with a shorter time in minutes before the first interruption.

Standard errors are cluster mainly at the YouTube channel level that upload the video which in most of the cases corresponds to the seminar series.[6] Alternatively economic department to which the presenter belongs, country of location of the department and presenter are also used for robustness as cluster variables. In the case of presenter it should be noted that from the 845 different presenters identified in the dataset, 76 presented in more than one seminar series.

Making use of the transcripts of each seminar, two strategies were followed to determine the topic of the presentation. The first was based on the count of words used by the presenter that matched JEL categories. For example if the presenter mentions "household behavior" or "life cycle model" the seminar will be coded under the code D1, which corresponds to the category "Household Behavior and Family Economics". When the presenter mentions words from different JEL categories, a majority rule was used. The second strategy consisted on the topics produced by a topic analysis performed on the transcripts of the presenter. The following section presents

[6]Exceptions to this can be found for example in the NBER or CEPR YouTube channel where videos from different seminar series are posted.

details on how the topic analysis was produced.

### 4.1 Inferring type of interruption by audio transcripts

The procedure explained in Section 3 does not identify the nature of the interruption. This implies that under the mask of interruption, it is being pooled together different type of interactions as questions, clarifications, criticisms or suggestions. Making use of the audio transcripts I can predict which of the interruptions are interrogatives. For that I run a gradient boosting algorithm which was trained using a corpus with more than 10,000 human annotated posts made in online forums available in Forsythand and Martell (2007). Among others, the posts contains dialogue-act tagged information about whether a sentence is interrogative or not.[7]

Once interrogative interruptions are identified it is possible to define a variable based on how many questions a speaker received during the presentation. This variable can be used as dependent variable in Equation 4. In this case, a positive value of $\hat{\beta}_1$ would imply that being a female presenter increases the number of questions received.

## 5 Results

Results are grouped in three parts. The first explores the relationship between being a female presenter and the number of interruptions received during a seminar. Also this part discusses how interrogative interruptions varies according to the gender of the presenter. The second part discusses what drives the gap in interruptions received by male and female presenters. On the last part additional results and robustness checks are presented.

### 5.1 The effect of being a female presenter on seminar interruptions

The baselines results of this paper are presented in Table **??**. There Equation 4 is by OLS with the number of interruptions as dependent variable. In this case the number of interruptions is used independently of who makes it. Section 5.3 shows results considering whether the interruption is made by the same or different individuals in the audience.

In all the specifications, the variable of interest, gender of the presenter, is significantly positive. This implies that female presenters receive on average between 0.9 and 1.4 more inter-

[7] The tags includes “Wh” questions (questions that begin with “what”, “when”, “where”, “who”, “whom”, “which”, “whose”, “why” and “how”) and closed question which can be only answered with yes or not.

ruptions in seminars compared to males. Interestingly this coefficient seems relatively constant across the different specifications. As expected, seminar duration has a significant effect on the number of interruptions and its inclusion in the model also increases the effect of female presenter on the number of interruptions. On the other hand, the effect on the coefficient of female presenter does not have a significant change when I control by seniority or citations in in the presenter's Google Scholar profile. A presenter with more citations (column 3) or more senior (column 4), receives on average fewer interruptions in a seminar. The direction of this effect persists when the 15 topics identified in the topic analysis described in Section 2 are used as control variables in the last column. However in this case the gender of the presenter does not appear to be significant. As it will be discussed in Section 5.2, the relatively smaller coefficient associated to being a female presenter masks an important heterogeneity across the different topics.

Results are less conclusive when interrogative interruptions are used as dependent variable. Even if the effect is positive, Table **??** shows how in only two, out of six specifications, being a female presenter has a significant effect on the number of questions received. While for the overall population of seminar presenters, being a female has a positive and significant effect on the number of interruptions received, this effect can not be established when only interrogative interruptions are considered.

In both specifications, standard errors were clustered at the YouTube channel level, the proxy of seminar series used in this work. Results remain unchanged when the cluster variable is the speaker, the economic department of affiliation or the country of location of the department.

### 5.2 What drives this gap?

The previous results are meaningful and at the same time consistent with for example the findings of Dupas et al. (2021). However with the available data it is possible to shed some light on what drives this behavior by combining the gender of the individuals that makes the interruptions and the topic of the presentation.

Table 5 presents the average number of interruptions received by male and female speakers in total numbers and controlling by the gender of the interrupter. The first column shows that on average female presenters receive one more interruption per seminar than male presenters. The second and third column shows on average how many of those interruptions were driven by males and females interrupters in the audience. The fact that on average, males make more interruptions than females which emerges from comparing the second and third column should not be surprising as it is expected a higher presence of males in the audience than of females. However the comparison between how males and females interrupt to the speakers it is worthy of notice. While men attending to the seminars make an average of 0.2 extra interruptions to female presenters in comparison to male presenters, women make almost one extra interruption to female than to male presenters.

Table **??** presents results when Equation 4 is estimated considering the gender of the interrupter. Panel A on the top, considers interruptions made by males in the audience and the bottom panel, B, does it when interruptions came by females in the audience.

In Panel A, when the number of interruptions made by male participants is used as dependent variable in Equation 4, the gender of the presenter is not statistically significant from zero only in most of the specifications. This seems to indicate that being a female presenter does not increase the number of interruptions by males in the audience. This contrasts with Panel B in which, even if in a modest magnitude, gender of the presenter seems to be significant to explain the number of interruptions made by females in the audience. In other words, a seminar in which the presenter is a woman increases the number of interruptions made by women in the audience.

Even more notable differences are found when interrogative interruptions and the gender of who does it is considered. Panel A of Table **??** show results when questions made by males in the audience is the explained variable and Panel B when questions made by females is the variable to explain. In both panels being a female is significant in order to explain the number of questions received by males and females respectively. However they do it with opposite signs: being a female presenter leads to fewer questions by males and to more questions by females in the audience. In contrast to the previous results, for both set of specifications gender of the presenter is significant also when topic of the presentation is included as control variable.

So far, the topic of the presentation has been used only as a control variable in regressions in which the main target variable was the gender of the presenter. However it is possible to observe how this effect varies according to the topic of the presentation. Results are presented in Table 7, and Table 8. Table 7 shows how male and female presenters face a disparate treatment according to the topic of the seminar. At a first glance, it does not emerges a clear pattern of the interruptions received by males and females by topic. Excluding Topics 3, 9 and 15 in which it was not possible to identify a clear JEL topic, female presenters receive at least one more extra interruption than male presenters in 4 of the 15 topics while male presenters receive one or more interruptions than females in 3 topics. In the remaining topics, differences between male and female presenters is less than one interruption.

Table 8 explores whether this pattern is driven by male or female attendees. In Topic 1 (Math and Quant) and Topic 10 (Int Econ) female presenters receive on average close to two extra interruptions compared to male presenters. In both cases, those extra interruptions are driven almost entirely by males in the audience. In Topic 1 men make an average of 1.9 extra interruptions to female presenters while in Topic 10 they make 1.4 extra interruptions (and in this last topic women make 0.3 extra interruptions to female presenters). By contrast in Topic 6 (Labour) the extra interruptions received by female presenters is driven mostly by women in the audience (men make 0.9 extra interruptions while women make 2.4 extra interruptions to female presenters).

Additionally, differences in the number of interruptions received by male and female presenters are narrower in Topic 8 (Dev & Growth), Topic 12 (Labour) and Topic 13 (Pub. Economics). However in these three topics it is also observed that female presenters receive a higher number of interruptions by women than by men in the audience. The reason why the difference between interruptions to female and male presenters is smaller than the observed in Topic 6 is because in these cases, men in the audience compensate this by making more interruptions to male than to female presenters.

The variation by gender and topic in the number of interruptions received by the presenters seems not to be at random and it can be explained by the share of females present in each topic. As recent literature has pointed out, there is a wide variation in the share of females across topics in economics (Card et al. 2020; Chari and Goldsmith-Pinkham 2017). The available data does not contain information on the gender composition of the attendees to certain seminar. Only the gender of those participants that speak on it can be predicted. However it is possible to take a proxy of this variable, as for example, the percentage of presenters that are females in each field: topics in which there is a higher percentage of females presenters are likely to also have a larger composition of females attending to it. Accepting this variable as proxy, Figure **??** shows how in topics with a higher presence of females, the number of interruptions made by women over the total number of interruptions increases. A similar pattern is depicted in Figure **??** when Card et al. (2020) data, who uses the fraction females in EconLit, is used instead.

The first 6 columns of Table 9 re-estimates the Equation 4 using the percentage of female presenters by topic as an additional control variable. Additionally in Column 7, it is used the share of females in EconLit as measure of presence of females in the field. As the values of this variables are computed by topic, to avoid multicollinearity I re-group topic in four groups: microeconomics, macroeconomics, quantitative methods and others.[8] Given everything else, 1% extra in the presence of females in the field leads to an increases in the number of interruptions by between 0.09 and 0.33 according to the specification. Interestingly, when I control by the presence of females in the field, the gender of the presenter is not longer significant to explain the number of interruptions received by the presenter. This supports the idea that, contrary to Dupas et al. (2021), it is not begin a female presenter which explains the higher number of interruptions in a seminar, but the share of females in the audience what does it.

### 5.3 Other specifications and robustness checks

In Table 15 results are presented when Equation 4 is estimated having how long it passes before the first interruption takes place as explained variable. The negative sign in the first column suggests that on average female presenters receive their first interruption around 6 minutes before their male counterpart. As in the previous results, this findings are robust when seminar's duration, citations, seniority and topic of the seminar are included as control variables. Conclusions also remains unchanged when standard errors are clustered at speaker, economics department and its location level.

In Table **??** the same specifications are proposed but in this case the dependent variable is the number of different seminar attendees that make an interruption. This implies that an attendant asking several questions will contribute to the dependent variable in the same way that an attendant asking a single one. The positive and significant coefficient associated to female presenter suggests that, even if in a small magnitude, more different individuals engage to participate in a seminar when the presenter is a woman.

Finally, in many seminar series there is a moderator that presents the seminar and to the speaker. Additionally in some cases this moderator organizes and reads questions posted in the chat of the seminar. Several rules were used to identify the moderator of a seminar in case there was. For example a moderator has to be the individual who speaks at the beginning of the seminar and at the same time do it for a short time period. The results presented above does

[8] Group 1: Topic 2 (IO), Topic 5 (Micro) and Topic 14 (Micro). Group 2: Topic 4 (Macro), Topic 10 (Int. Economics), Topic 11 (Pub. Economics) and Topic 13 (Pub. Economics), Group 3: Topic 6 (Labor), Topic 7 (Environmental Economics), Topic 8 (Development and Growth) and Topic 12 (Labor). Group 4: Topic 1 (Math and Quantitative Methods).

not change substantially when the interruptions of the moderator are not considered.

## 6 Conclusions & Future Avenues

Female presenters in economic academic seminars are more interrupted than males. On average a female presenter receives between 1 and 2 more interruptions, which means around 15% of extra interruptions than their male counterpart. These findings goes in line with the previous existent. Less clear is the role that the gender of the presenter plays when only interrogative interruptions are considered. Only two out of the six proposed specifications statistically supports the idea that female presenters receives more questions than male presenters.

In addition, they are interrupted earlier in the presentation and by a larger number of individuals in the audience. Controlling by the duration of the seminar, the citations and seniority of the presenter, female presenters are interrupted between 6 and 4 minutes on average earlier than males. Using the same set of control variables I also find a small but significant effect of in the number of different individuals interrupting the seminar when the presenter is a female.

On top, the interruptions that they receive normally takes a longer time than the ones received by males. Among other things this implies that a male presenter is speaking for around 80% of the seminar duration while a female speaker does it for less than 75%.[9]

Yet, this work provides evidence of a less widespread idea: those extra interruptions received by female presenters are not entirely due to men in the audience but to female attendees. As a matter of fact, being a female presenter is significant to explain the number of interruptions made by women in the audience but is not (only two out of the six proposed specifications were significant) to explain the number of interruptions made by males in the audience. Furthermore, when considering interrogative interruptions, significant evidence supports that being a female presenters increases the number of questions received by women but reduces the number of questions received by men.

Even if these results are significant when the topic of the presentation is used as control variable, the effect of being a female presenter on the number of interruptions varies across the different economic fields. As an example, in seminars assimilated to the field of labor economics and international economics a female presenter receives an average of two extra interruptions compared to male presenters. However while in labor economics the extra interruptions are

[9]This result has not been included yet in the main analysis.

driven to a large extent by females in the audience, in the seminars of international economics, males in the audience are the responsible for most of them.

As pointed out by the existent literature, the type of social behavior in which women engage, for example while speaking up in meetings, is affected by the presence of peers in a different way than the one of men. In that sense this paper provides evidence on how rather than the gender of the presenter, the share of females in the field is the key variable to explain the number of interruptions that a speaker receives. In the different proposed specifications a higher share of females in the audience increases the number of interruptions received overall and makes the gender of the presenter non significant to explain the number of interruptions. Including this variable to explain the number of interruptions received only by male presenters has a modest and sometimes not significant effect. By contrast it has a significant and larger effect when the variable to explain is the number of interruptions received by female presenters.

As stated in the introduction, most of the results presented in this paper take an agnostic view about the role played by interruptions in a seminar. In a medium-term time horizon further research should also explore for example whether papers presented in seminars with more interruptions, succeeded differently in the academic world than others with less interventions during its presentation.

## References


Alnuaim, Abeer Ali, Mohammed Zakariah, Chitra Shashidhar, Wesam Atef Hatamleh, Hussam Tarazi, Prashant Kumar Shukla, and Rajnish Ratna (2022). "Speaker Gender Recognition Based on Deep Neural Networks and ResNet50". In: *Wireless Communications and Mobile Computing* 2022.

Anguera, Xavier, Simon Bozonnet, Nicholas Evans, Corinne Fredouille, Gerald Friedland, and Oriol Vinyals (2012). "Speaker diarization: A review of recent research". In: *IEEE Transactions on audio, speech, and language processing* 20.2, pp. 356–370.

Ardila, R., M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber (2020). "Common Voice: A Massively-Multilingual Speech Corpus". In: *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pp. 4211–4215.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri (2020). "Are referees and editors in economics gender neutral?" In: *The Quarterly Journal of Economics* 135.1, pp. 269–327.

Cettolo, Mauro, Michele Vescovi, and Romeo Rizzi (2005). "Evaluation of BIC-based algorithms for audio segmentation". In: *Computer Speech & Language* 19.2, pp. 147–170.

Chachadi, Kavita and SR Nirmala (2022). "Voice-based gender recognition using neural network". In: *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Springer, pp. 741–749.

Chari, Anusha and Paul Goldsmith-Pinkham (2017). *Gender representation in economics across topics and time: Evidence from the NBER summer institute*. Tech. rep. National Bureau of Economic Research.

Chollet, Francois (2021). *Deep learning with Python*. Simon and Schuster.

Dadvar, Maral (2011). *Who spoke when? Audio-based speaker location estimation for diarization*. LAP LAMBERT Academic Publishing.

Desplanques, Brecht, Jenthe Thienpondt, and Kris Demuynck (2020). "Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification". In: *arXiv preprint arXiv:2005.07143*.

Dindia, Kathryn (1987). "The effects of sex of subject and sex of partner on interruptions". In: *Human Communication Research* 13.3, pp. 345–371.

Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, et al. (2021). *Gender and the dynamics of economics seminars*. Tech. rep. National Bureau of Economic Research.

Enqing, Dong, Liu Guizhong, Zhou Yatong, and Cai Yu (2002). "Voice activity detection based on short-time energy and noise spectrum adaptation". In: *6th International Conference on Signal Processing, 2002.* Vol. 1. IEEE, pp. 464–467.

Forsythand, Eric N and Craig H Martell (2007). "Lexical and discourse analysis of online chat dialog". In: *International Conference on Semantic Computing (ICSC 2007)*. IEEE, pp. 19–26.

Ginther, Donna K and Shulamit Kahn (2004). "Women in economics: moving up or falling off the academic career ladder?" In: *Journal of Economic perspectives* 18.3, pp. 193–214.

Hengel, Erin (2022). "Publishing while Female. Are women held to higher standards? Evidence from peer review." In.

Hilton, Katherine (2018). "Social meaning in a shifting grammatical landscape: The perception of nonagreement in existential there constructions". In: *Journal of Sociolinguistics* 22.2, pp. 233–249.

Hilton, Katherine and Sunwoo Jeong (2019). "The role of context in sociolinguistic perception". In: *Linguistics Vanguard* 5.s1.

Jia, Fei, Somshubra Majumdar, and Boris Ginsburg (2021). "Marblenet: Deep 1d time-channel separable convolutional neural network for voice activity detection". In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 6818–6822.

McCallum, Andrew (2002). "MALLET: A Machine Learning for Language Toolkit". http://mallet.cs.umass.edu.

Moattar, Mohammad Hossein and Mohammad M Homayounpour (2012). "A review on speaker diarization systems and approaches". In: *Speech Communication* 54.10, pp. 1065–1103.

Müller, Meinard (2021). *Fundamentals of Music Processing: Using Python and Jupyter Notebooks*. Springer Nature.

Nagrani, Arsha, Joon Son Chung, Weidi Xie, and Andrew Zisserman (2020). "Voxceleb: Large-scale speaker verification in the wild". In: *Computer Speech & Language* 60, p. 101027.

Paredes, Valentina A, M Daniele Paserman, and Francisco Pino (2020). *Does Economics Make You Sexist?* Tech. rep. National Bureau of Economic Research.

Pulkki, Ville, Symeon Delikaris-Manias, and Archontis Politis (2017). *Parametric time-frequency domain spatial audio*. John Wiley & Sons.

Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning (2020). "Stanza: A Python natural language processing toolkit for many human languages". In: *arXiv preprint arXiv:2003.07082*.

Ravanelli, Mirco, Titouan Parcollet, Peter Plantinga, Aku Rouhe, Samuele Cornell, Loren Lugosch, Cem Subakan, Nauman Dawalatabad, Abdelwahab Heba, Jianyuan Zhong, Ju-Chieh Chou, Sung-Lin Yeh, Szu-Wei Fu, Chien-Feng Liao, Elena Rastorgueva, François Grondin, William Aris, Hwidong Na, Yan Gao, Renato De Mori, and Yoshua Bengio (2021). *SpeechBrain: A General-Purpose Speech Toolkit*. arXiv: `2106.04624 [eess.AS]`.

Röder, Michael, Andreas Both, and Alexander Hinneburg (2015). "Exploring the space of topic coherence measures". In: *Proceedings of the eighth ACM international conference on Web search and data mining*, pp. 399–408.

Sarkar, Dipanjan (2019). *Text analytics with Python: a practitioner's guide to natural language processing*. Springer.

Wu, Alice H (2018). "Gendered language on the economics job market rumors forum". In: *AEA Papers and Proceedings*. Vol. 108, pp. 175–79.

Zimmermann, Don H and Candace West (1996). "Sex roles, interruptions and silences in conversation". In: *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*. JOHN BENJAMINS BV, pp. 211–236.

# Appendix I

Table 1: Sample characteristics (standard error in parenthesis)

| **Duration** | | **Interruptions** | |
|---|---|---|---|
| Duration (mean, in min) | 62.3 | Interruptions (mean) | 10.8 |
| Less than 45 min | 14.2% | Inter. 30 sec or less (mean) | 6.6 |
| Between 45min and 70 min | 56.5% | Inter. between 30 and 60 sec (mean) | 2.6 |
| Between 70min and 90 min | 25.8% | Inter. between 60 and 120 sec (mean) | 1.2 |
| More than 90 min | 3.5% | Inter longer than 120 sec (mean) | 0.4 |
| **Ranking** | | **Other** | |
| Dept. ranking <10 | 68.2% | Sem. without inter. | 8.3% |
| Dept. ranking between 10 and 20 | 7.4% | Sem. only w/host inter. | 10.1% |
| Dept. ranking between 20 and 50 | 10.5% | Sem. w/inter. from one pers. | 22.0% |
| Dept. ranking between 50 and 100 | 5.2% | Female presenters | 35.3% |
| Dept. ranking >100 | 8.6% | Female hosts | 28.2% |
| Presenters | 1,547 | Seminars | 2,131 |

Table 2: Location of the department of seminar presenters

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Australia | 2.5% | France | 3.1% | Japan | 0.4% | Singapore | 0.7% |
| Austria | 0.2% | Germany | 1.6% | Luxembourg | 0.9% | South Korea | 0.4% |
| Belgium | 1.1% | Hong Kong | 0.2% | Netherlands | 1.4% | Spain | 1.3% |
| Brazil | 0.2% | Ireland | 0.7% | New Zealand | 0.2% | Sweden | 0.7% |
| Canada | 2.7% | Israel | 0.5% | Norway | 0.2% | USA | 66.2% |
| Colombia | 0.2% | Italy | 2.0% | Portugal | 0.4% | United Kingdom | 12.3% |

Figure 1: Location of the department of seminar presenters

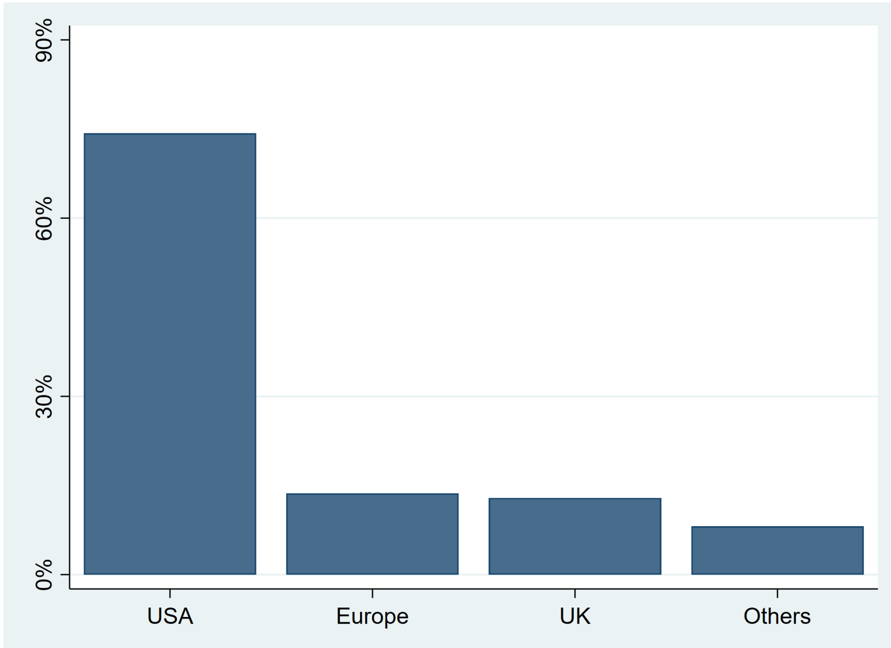



Figure 2: Ranking of the department of seminar presenters

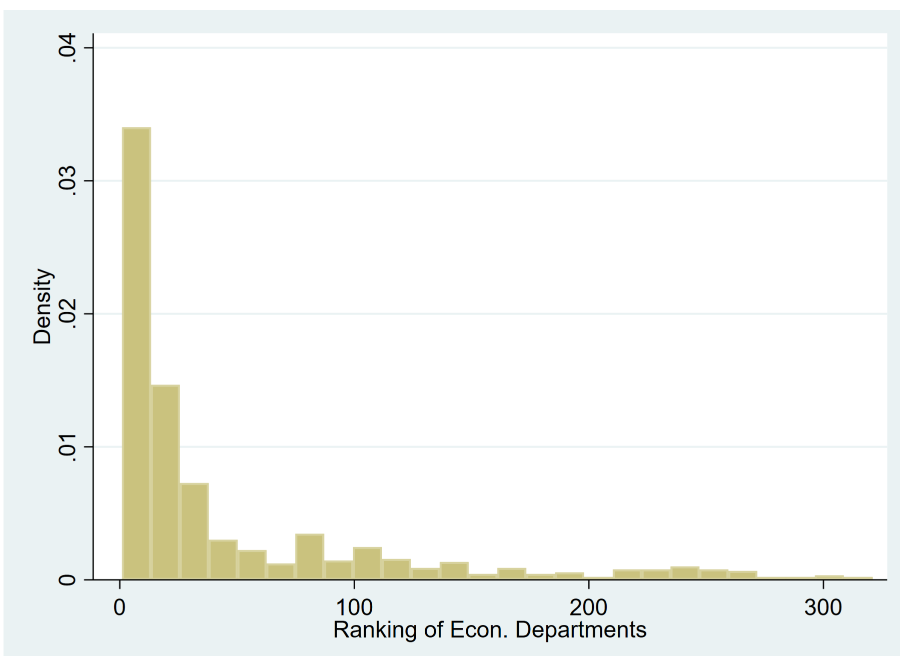


Figure 3: Density of number of interruptions by gender of the presenter

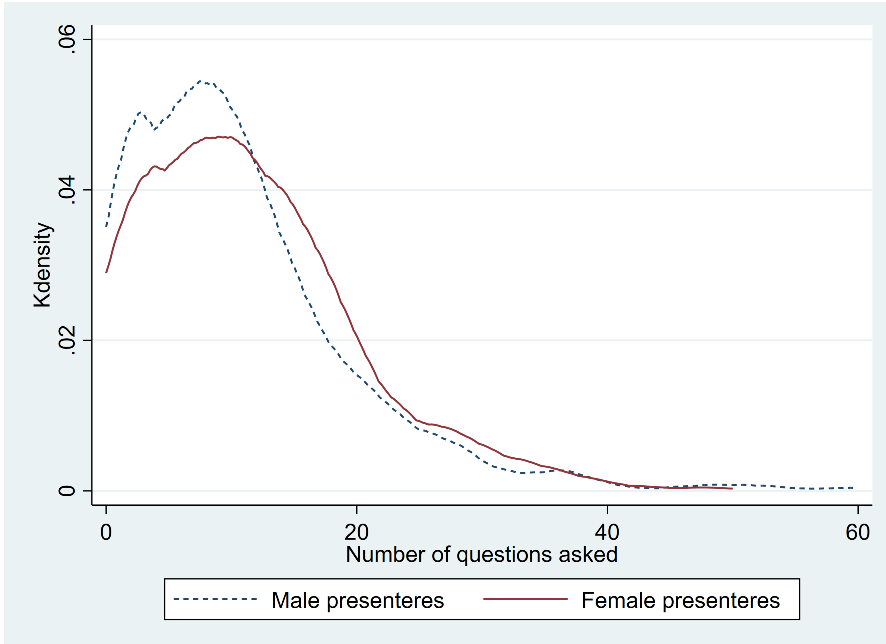



Figure 4: Density of number of interruptions by gender of the interrupter

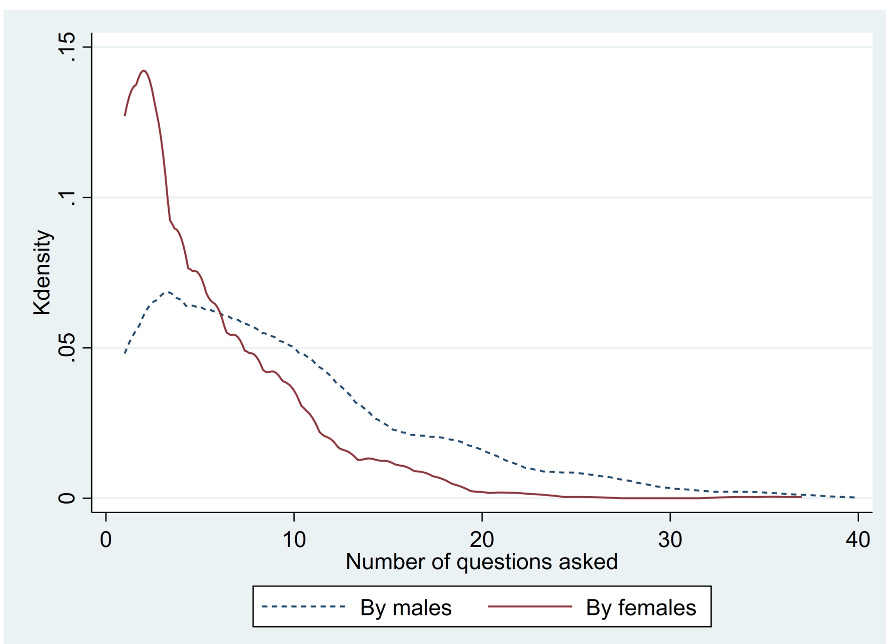
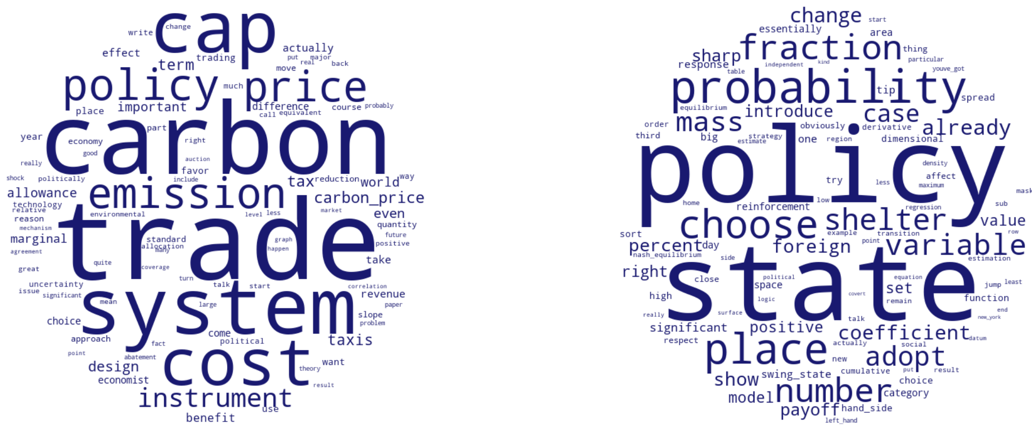


Figure 5: Word clouds based on seminars' audio transcripts




(a) Female's transcripts

(b) Male's transcripts

Figure 6: Topic analysis results

Table 3: Interruptions by topic

| Topic | Share of topics | Female presenters | First 5 words of the topic | JEL - Category |
|---|---|---|---|---|
| 1 | 5.0% | 23.7% | model, estimate, datum, variable, sample | C - Math & Quant. Methods |
| 2 | 6.5% | 30.3% | price, market, platform, consumer, pay | L - Industrial Organization |
| 3 | 6.2% | 43.1% | effect, datum, year, paper, find | Z - Not Assigned |
| 4 | 9.2% | 29.6% | market, bank, asset, price, risk | E - Macro. and Monet. Eco. |
| 5 | 4.9% | 22.8% | model, rate, cost, low, policy | D - Microeconomics |
| 6 | 7.9% | 40.9% | country, political, immigrant, state, migration | J - Labor and Demog. Economics |
| 7 | 6.4% | 28.0% | change, climate, technology, world, policy | Q - Enviromental & Ecological Econ. |
| 8 | 7.7% | 43.3% | group, experiment, treatment, social, study | O - Dev. & Growth. |
| 9 | 5.5% | 46.9% | datum, research, system, information, patient | Z - Not Assigned |
| 10 | 9.8% | 39.1% | firm, trade, country, sector, market | F - International Econ. |
| 11 | 5.8% | 27.9% | policy, number, crisis, pandemic, year | H - Public Economics |
| 12 | 7.9% | 44.6% | worker, job, labor, child, high | J - Labor and Demog. Economics |
| 13 | 5.5% | 28.1% | income, household, tax, city, datum | H - Public Economics |
| 14 | 8.2% | 35.4% | agent, information, set, game, state | D - Microeconomics |
| 15 | 3.5% | 31.7% | kind, thing, sort, lot, talk | Z - Not Assigned |

Figure 7: Number of speakers by JEL-code

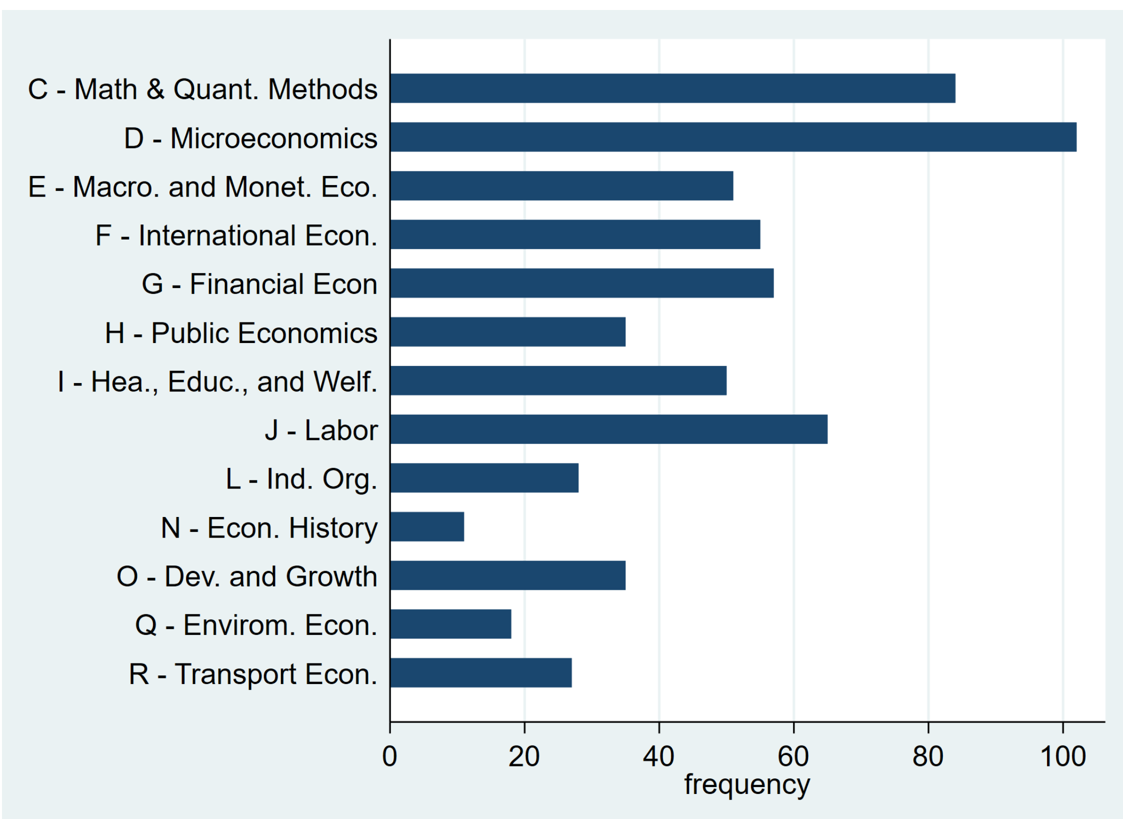


Table 4: Interruptions in a seminar

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Female presenter | 0.90** | 1.39*** | 1.18** | 1.22** | 1.05* | 1.06* | 1.53** |
| | (0.43) | (0.39) | (0.55) | (0.52) | (0.56) | (0.53) | (0.60) |
| Duration (in hs) | | 0.21*** | 0.22*** | 0.22*** | 0.22*** | 0.21*** | 0.17*** |
| | | (0.05) | (0.05) | (0.06) | (0.05) | (0.05) | (0.05) |
| Citations | | | -0.25*** | | -0.19*** | -0.28*** | -0.32*** |
| | | | (0.06) | | (0.06) | (0.07) | (0.10) |
| Seniority | | | | -0.11*** | -0.08** | -0.09** | -0.07* |
| | | | | (0.03) | (0.03) | (0.03) | (0.04) |
| Topic | | | | | | Yes | Yes |
| Speaker's Dept. Locat. | | | | | | | Yes |
| Seminar Series | | | | | | | Yes |
| Constant | 10.48*** | -2.92 | -2.33 | -1.66 | -1.54 | -1.60 | 1.06 |
| | (1.11) | (2.21) | (2.80) | (2.86) | (2.84) | (3.24) | (3.68) |
| $R^2$ | 0.00 | 0.17 | 0.17 | 0.16 | 0.17 | 0.22 | 0.23 |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note: Standard errors in parentheses and clustered at the YouTube channel level. * p<0.05, ** p<0.01, ***p<0.001.

Table 5: Interruptions in total numbers and be gender of the interrupter

| | Total interruptions | Interruptions by males | Interruptions by females |
|---|---|---|---|
| Male Presenters | 10.3 | 7.9 | 2.3 |
| Female Presenters | 11.3 | 8.1 | 3.2 |

Table 6: Interruptions made by males and females in the audience

| | Coefficient of female presenter | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Interruptions by males | 0.21 | 0.60* | 0.58 | 0.71* | 0.54 | 0.20 | 0.78 |
| | (0.36) | (0.33) | (0.43) | (0.40) | (0.43) | (0.48) | (0.57) |
| Interruptions by females | 0.69** | 0.79*** | 0.60** | 0.51* | 0.52* | 0.86** | 0.78* |
| | (0.28) | (0.27) | (0.29) | (0.28) | (0.29) | (0.40) | (0.57) |
| Duration (in hs) | | Yes | Yes | Yes | Yes | Yes | Yes |
| Citations | | | Yes | | Yes | Yes | Yes |
| Seniority | | | | Yes | Yes | Yes | Yes |
| Topic | | | | | | Yes | Yes |
| Speaker's Dept. Locat. | | | | | | | Yes |
| Seminar Series | | | | | | | Yes |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note: Standard errors in parentheses and clustered at the YouTube channel level.
* p<0.05, ** p<0.01, ***p<0.001.

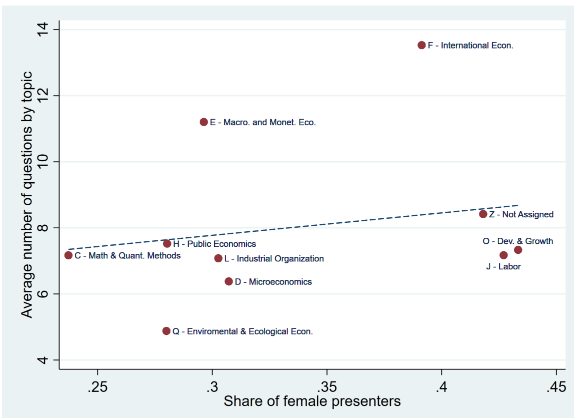
Table 7: Total interruptions to male and female presenters

| | Topic 1: **Math & Quant** | Topic 2: **IO** | Topic 3: N/A | Topic 4: **Macro** | Topic 5: **Micro** |
|---|---|---|---|---|---|
| Male presenter | 8.0 | 9.4 | 12.9 | 13.9 | 10.3 |
| Female presenter | 9.6 | 7.3 | 15.6 | 12.1 | 10.8 |
| | Topic 6: **Labor** | Topic 7: **Envirom.** | Topic 8: **Dev.&Grow.** | Topic 9: N/A | Topic 10: **Int. Econ.** |
| Male presenter | 9.0 | 7.1 | 9.6 | 8.8 | 16.9 |
| Female presenter | 12.4 | 6.8 | 10.1 | 9.6 | 18.7 |
| | Topic 11: **Pub. Econ** | Topic 12: **Labor** | Topic 13: **Pub. Econ** | Topic 14: **Micro** | Topic 15: N/A |
| Male presenter | 9.8 | 11.5 | 10.6 | 6.2 | 14.3 |
| Female presenter | 10.2 | 10.9 | 9.5 | 8.0 | 10.2 |

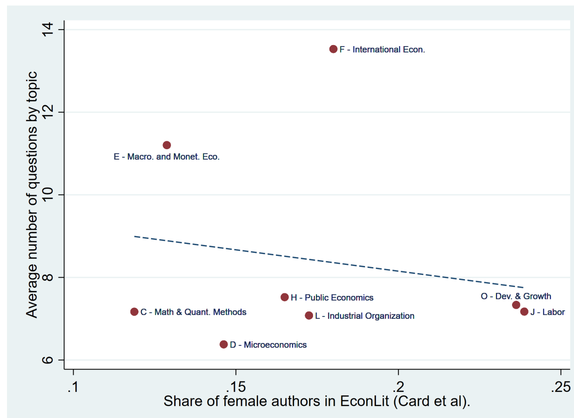
Table 8: Total interruptions to male and female presenters (by gender of the interrupter)

| | **Topic 1**: **Math & Quant** | | **Topic 2**: **IO** | | **Topic 3**: N/A | | **Topic 4**: **Macro** | | **Topic 5**: **Micro** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females |
| Male pres. | 6.7 | 1.2 | 7.7 | 1.7 | 8.0 | 4.9 | 11.3 | 2.5 | 8.0 | 2.3 |
| Female pres. | 8.6 | 0.9 | 5.7 | 1.7 | 10.5 | 5.1 | 10.9 | 1.2 | 8.5 | 2.3 |
| | **Topic 6**: **Labor** | | **Topic 7**: **Enviorm.** | | **Topic 8**: **Dev & Growth** | | **Topic 9**: N/A | | **Topic 10**: **Int. Econ** | |
| | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females |
| Male pres. | 7.0 | 1.9 | 5.2 | 1.9 | 7.5 | 2.1 | 6.6 | 2.3 | 13.0 | 4.0 |
| Female pres. | 8.1 | 4.3 | 4.1 | 2.7 | 7.1 | 3.0 | 7.7 | 1.9 | 14.4 | 4.3 |
| | **Topic 11**: **Pub. Econ** | | **Topic 12**: **Labor** | | **Topic 13**: **Pub. Econ** | | **Topic 14**: **Micro** | | **Topic 15**: N/A | |
| | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females | Int. by males | Int. by females |
| Male pres. | 7.7 | 2.1 | 7.7 | 3.7 | 8.0 | 2.7 | 5.0 | 1.2 | 10.3 | 4.0 |
| Female pres. | 8.0 | 2.2 | 5.8 | 5.1 | 5.4 | 4.1 | 6.0 | 2.0 | 7.4 | 2.8 |

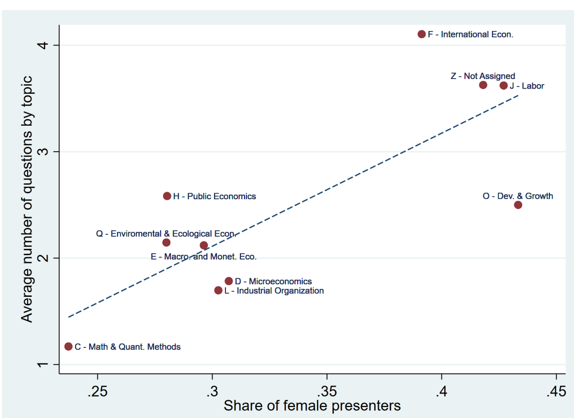
Figure 8: Presence of females by topic and interruptions



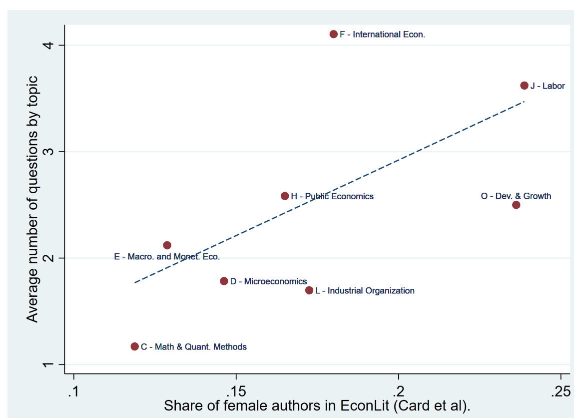

(a) Male's interruptions and share of female presenters




(b) Male's interruptions and EconLit (Card et al. 2020)




(c) Female's interruptions and share of female presenters




(d) Female's interruptions and EconLit (Card et al. 2020)

Table 9: Interruptions in a seminar (controlling by presence of females)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7*) |
|---|---|---|---|---|---|---|---|
| Presence of females | 0.08 | 0.08 | 0.08 | 0.08 | 0.09 | 0.16 | 0.07 |
| | (0.04) | (0.04) | (0.05) | (0.05) | (0.05) | (0.07) | (0.24) |
| Female presenter | -2.14 | -1.59 | -2.03 | -1.93 | -2.25 | -4.87 | -0.26 |
| | (2.70) | (2.47) | (3.18) | (3.19) | (3.18) | (3.84) | (3.64) |
| Presence of fem x Fem. Present | 0.07* | 0.09** | 0.09* | 0.09* | 0.09** | 0.24*** | 0.27** |
| | (0.04) | (0.04) | (0.05) | (0.05) | (0.05) | (0.11) | (0.16) |
| Duration (in hs) | | 0.21*** | 0.22*** | 0.22*** | 0.22*** | 0.21*** | 0.21*** |
| | | (0.01) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Citations | | | -0.24*** | | -0.20*** | -0.28*** | -0.29*** |
| | | | (0.06) | | (0.07) | (0.08) | (0.09) |
| Seniority | | | | -0.10*** | -0.06* | -0.06 | -0.07* |
| | | | | (0.03) | (0.03) | (0.04) | (0.04) |
| Topic (mod) | | | | | Yes | Yes | Yes |
| Speaker's Dept. Locat. | | | | | | Yes | Yes |
| Seminar Series | | | | | | | Yes |
| Constant | 8.17*** | -6.34*** | -5.61*** | -4.53** | -4.47** | -11.82*** | -13.33** |
| | (1.52) | (1.63) | (2.09) | (2.18) | (2.18) | (3.20) | (6.20) |
| R2 | 0.01 | 0.17 | 0.17 | 0.17 | 0.18 | 0.25 | 0.22 |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note 1: Standard errors in parentheses and clustered at the YouTube channel level. * p<0.05, ** p<0.01, ***p<0.001. Specification (7*) uses Card et al. 2020 data for computing Presence of females.

Table 10: Interruptions in a seminar (controlling by presence of females and by who asks the question)

| | Interr. made by males | | | Interr. made by females | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| Presence of females | 0.01 | 0.19 | 0.27 | 0.07** | 0.05* | 0.22 |
| | (0.05) | (0.11) | (0.24) | (0.02) | (0.03) | (0.14) |
| Female presenter | -1.77 | -2.84 | 1.88 | -0.48 | -2.03 | -2.14 |
| | (3.03) | (2.91) | (3.41) | (1.52) | (1.55) | (1.78) |
| Presence of fem x Fem. Present. | 0.06 | 0.08 | -0.09 | 0.02** | 0.08** | 0.16** |
| | (0.08) | (0.08) | (0.18) | (0.01) | (0.04) | (0.09) |
| Duration (in hs) | -0.18*** | -0.24*** | -0.25*** | 0.04*** | 0.04*** | 0.04*** |
| | (0.07) | (0.07) | (0.09) | (0.01) | (0.01) | (0.01) |
| Citations | -0.02 | -0.03 | -0.04 | -0.01 | -0.03 | -0.04 |
| | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.05) |
| Seniority | | -2.60** | 1.99 | -0.04** | -0.03** | -0.03 |
| | | (1.23) | (2.12) | (0.02) | (0.01) | (0.02) |
| Topic | | Yes | Yes | | Yes | Yes |
| Speaker's Dept. Locat. | | | Yes | | | Yes |
| Seminar Series | | | Yes | | | Yes |
| Constant | -2.86 | -5.20* | -8.62 | -1.62 | -0.91 | -4.71 |
| | (2.07) | (3.06) | (5.81) | (1.04) | (1.35) | (3.02) |
| $R^2$ | 0.13 | 0.20 | 0.18 | 0.05 | 0.07 | 0.06 |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note 1: Standard errors in parentheses and clustered at the YouTube channel level.
* p<0.05, ** p<0.01, ***p<0.001.

Table 11: Number of questions

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Panel A: All Questions | | | | | | |
| Female presenter | 0.29 | 0.40** | 0.33 | 0.37* | 0.29 | 0.35 | 0.52** |
| | (0.18) | (0.18) | (0.21) | (0.21) | (0.22) | (0.23) | (0.20) |
| | Panel B: Males' Questions | | | | | | |
| Female presenter | -0.93*** | -0.84*** | -0.99*** | -0.94*** | -1.01*** | -1.04*** | -0.91*** |
| | (0.14) | (0.14) | (0.21) | (0.21) | (0.22) | (0.24) | (0.21) |
| | Panel C: Females' Questions | | | | | | |
| Female presenter | 1.21*** | 1.24*** | 1.31*** | 1.30*** | 1.30*** | 1.39*** | 1.43*** |
| | (0.19) | (0.19) | (0.19) | (0.19) | (0.19) | (0.19) | (0.20) |
| Duration (in hs) | | Yes | Yes | Yes | Yes | Yes | Yes |
| Citations | | | Yes | | Yes | Yes | Yes |
| Seniority | | | | Yes | Yes | Yes | Yes |
| Topic | | | | | | Yes | Yes |
| Speaker's Dept. Rank. | | | | | | | Yes |
| Seminar Series | | | | | | | Yes |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note 1: Standard errors in parentheses and clustered at the YouTube channel level.
* p<0.05, ** p<0.01, ***p<0.001.

Table 12: Text Similarity between presentation and interruption

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | | | Panel A: All Interruptions | | | | |
| Female presenter | 0.007 | 0.008 | 0.000 | 0.000 | 0.000 | -0.001 | -0.002 |
| | (0.005) | (0.005) | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) |
| | | | Panel B: Males' Interruptions | | | | |
| Female presenter | -0.005 | -0.004 | -0.013*** | -0.014*** | -0.014*** | -0.013** | -0.013* |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.007) |
| | | | Panel C: Females' Interruptions | | | | |
| Female presenter | 0.000 | 0.001 | -0.005 | -0.005 | -0.006 | -0.009 | -0.013 |
| | (0.007) | (0.007) | (0.008) | (0.008) | (0.008) | (0.010) | (0.011) |
| Duration (in hs) | | Yes | Yes | Yes | Yes | Yes | Yes |
| Citations | | | Yes | | Yes | Yes | Yes |
| Seniority | | | | Yes | Yes | Yes | Yes |
| Topic | | | | | | Yes | Yes |
| Speaker's Dept. Rank. | | | | | | | Yes |
| Seminar Series | | | | | | | Yes |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note 1: Standard errors in parentheses and clustered at the YouTube channel level.
* p<0.05, ** p<0.01, ***p<0.001.

Table 13: Interruption made by overlapping voices

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | | | Panel A: All overlapping interruptions | | | | |
| Female presenter | 0.07 | 0.11** | 0.12 | 0.12 | 0.11 | 0.11 | 0.13 |
| | (0.05) | (0.06) | (0.07) | (0.08) | (0.08) | (0.09) | (0.09) |
| | | | Panel B: Overlapping interruptions made by males | | | | |
| Female presenter | -0.24*** | -0.20*** | -0.16** | -0.16** | -0.17** | -0.18** | -0.14* |
| | (0.05) | (0.05) | (0.07) | (0.07) | (0.07) | (0.07) | (0.08) |
| | | | Panel C: Overlapping interruptions made by females | | | | |
| Female presenter | 0.64*** | 0.67*** | 0.67*** | 0.67*** | 0.67*** | 0.68*** | 0.65*** |
| | (0.06) | (0.06) | (0.09) | (0.09) | (0.09) | (0.10) | (0.11) |
| Duration (in hs) | | Yes | Yes | Yes | Yes | Yes | Yes |
| Citations | | | Yes | | Yes | Yes | Yes |
| Seniority | | | | Yes | Yes | Yes | Yes |
| Topic | | | | | | Yes | Yes |
| Speaker's Dept. Rank. | | | | | | | Yes |
| Seminar Series | | | | | | | Yes |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note 1: Standard errors in parentheses and clustered at the YouTube channel level.
* p<0.05, ** p<0.01, ***p<0.001.

Table 14: Number of interruptions controlling by chair's gender

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | | | | Total interruptions | | | |
| Female presenter | 0.73 | 1.03 | 0.96 | 0.95 | 0.84 | 1.15 | 0.71 |
| | (0.69) | (0.63) | (0.80) | (0.79) | (0.80) | (1.02) | (1.16) |
| Female chair | -2.85*** | -2.68*** | -3.08*** | -3.05*** | -3.07*** | -3.29*** | -4.53*** |
| | (0.82) | (0.75) | (0.96) | (0.96) | (0.96) | (1.22) | (1.44) |
| Fem. presenter x Fem. chair | 0.75 | 1.24 | 1.00 | 1.05 | 1.04 | 1.22 | 2.60 |
| | (1.42) | (1.30) | (1.70) | (1.69) | (1.70) | (2.15) | (2.46) |
| | | | | Interruptions made by males | | | |
| Female presenter | 0.14 | 0.37 | 0.12 | 0.21 | 0.09 | 0.51 | 0.34 |
| | (0.63) | (0.59) | (0.75) | (0.75) | (0.75) | (0.95) | (1.08) |
| Female chair | -7.14*** | -7.02*** | -7.43*** | -7.43*** | -7.43*** | -7.15*** | -8.64*** |
| | (0.75) | (0.70) | (0.90) | (0.90) | (0.90) | (1.14) | (1.34) |
| Fem. presenter x Fem. chair | -0.11 | 0.27 | 0.96 | 1.01 | 0.98 | 1.18 | 2.03 |
| | (1.29) | (1.21) | (1.59) | (1.59) | (1.59) | (2.00) | (2.29) |
| | | | | Interruptions made by females | | | |
| Female presenter | 0.60** | 0.66** | 0.83** | 0.74** | 0.75** | 0.64 | 0.36 |
| | (0.30) | (0.29) | (0.34) | (0.33) | (0.34) | (0.43) | (0.50) |
| Female chair | 4.30*** | 4.33*** | 4.36*** | 4.38*** | 4.36*** | 3.86*** | 4.10*** |
| | (0.36) | (0.35) | (0.41) | (0.40) | (0.41) | (0.52) | (0.62) |
| Fem. presenter x Fem. chair | 0.86 | 0.97 | 0.03 | 0.04 | 0.06 | 0.04 | 0.58 |
| | (0.62) | (0.60) | (0.72) | (0.71) | (0.72) | (0.91) | (1.06) |
| Duration (in hs) | | Yes | Yes | Yes | Yes | Yes | Yes |
| Citations | | | Yes | | Yes | Yes | Yes |
| Seniority | | | | Yes | Yes | Yes | Yes |
| Topic | | | | | | Yes | Yes |
| Speaker's Dept. Locat. | | | | | | | Yes |
| Seminar Series | | | | | | | Yes |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Table 15: Time at which occurs the first interruption

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Female presenter | -4.02*** | -3.49*** | -3.23** | -3.22*** | -2.90** | -3.79** |
| | (0.94) | (0.91) | (1.21) | (1.11) | (1.18) | (1.47) |
| Duration (in hs) | | 0.17* | 0.10 | 0.10 | 0.09 | 0.09 |
| | | (0.09) | (0.10) | (0.10) | (0.10) | (0.10) |
| Citations | | | 0.53*** | | 0.38* | 0.41* |
| | | | (0.18) | | (0.19) | (0.24) |
| Seniority | | | | 0.25*** | 0.19* | 0.23** |
| | | | | (0.09) | (0.10) | (0.10) |
| Topic | | | | | Yes | Yes |
| Speaker's Dept. Locat. | | | | | | Yes |
| Seminar Series | | | | | | Yes |
| Constant | 26.16*** | 14.87*** | 17.52*** | 15.68*** | 15.66*** | 19.57*** |
| | (1.82) | (5.12) | (5.66) | (5.68) | (5.66) | (6.10) |
| R2 | 0.01 | 0.04 | 0.04 | 0.04 | 0.05 | 0.08 |
| Observations | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 | 2,131 |

Note: Standard errors in parentheses and clustered at the YouTube channel level.
* p<0.05, ** p<0.01, ***p<0.001.

# Appendix II

## List of all speakers identified (in parenthesis the number of seminars in which they took part).

Markus Brunnermeier (5), Volker Wieland (4), Luigi Zingales (4), Stephen Redding (4), Simone Bertoli (4), Daron Acemoglu (4), Esteban Rossi-Hansberg (3), Melissa Dell (3), Harald Uhlig (3), Jean Tirole (3), Christian Krekel (3), Stefanie Stantcheva (3), Oleg Itskhoki (3), Michel Beine (3), Matthew Gentzkow (3), Shengwu Li (3), Susan Athey (3), Beata Javorcik (3), Dave Donaldson (2), Guido Imbens (2), Edward Glaeser (2), Janet Currie (2), Luis Cabral (2), Swati Dhingra (2), Douglas Bernheim (2), Jesse Schreger (2), Brett Falk (2), Avi Goldfarb (2), Jonathan Athow (2), Doireann Fitzgerald (2), Dmitry Taubinsky (2), Ricardo Reis (2), Maria Cotofan (2), Branko Milanovic (2), Olivier Blanchard (2), Stefano DellaVigna (2), Ruslan Salakhutdinov (2), Sarah Flèche (2), Esther Duflo (2), Adrien Auclert (2), Piotr Dworczak (2), Leah Boustan (2), Ying Nian Wu (2), Hillel Rapoport (2), Juliet Schor (2), Monica Morlacco (2), Leonard Wantchekon (2), Raj Chetty (2), Ana Maria Santacreu (2), Elhanan Helpman (2), Kevin Fox (2), Yuliy Sannikov (2), Mike Waugh (1), Neil Thompson (1), Mikkel Plagborg-Møller (1), Navin Kartik - Improving (1), Nellie Liang (1), Mo Salah (1), Nicholas Ashford (1), Mike Brewer (1), Monica Bell (1), Nachi Subramanian (1), Nava Ashraf (1), Nathaniel Hendren (1), Monica de Bolle (1), Natalie Lee (1), Myrna Wooders - Non-Cooperative (1), Moshe Tennenholtz (1), Natalia Ramondo (1), Motohiro Yogo (1), Mushfiq Mobarak (1), Natalia Fabra (1), Mushtaq Khan (1), Nash (1), Nancy Qian (1), Nicholas Z. Muller (1), Myra Samuels (1), Morgan Frank (1), Abhijit Banerjee (1), Nicola Fuchs-Schundeln (1), P. Koundouri (1), Pamela Medina Quispe (1), Paola Giuliano (1), Paola Manzini (1), Paolina Medina (1), Parag Pathak (1), Paschal Donohoe (1), Patricia Cortes (1), Paul Collier (1), Paul Ekins (1), Paul Elhorst (1), Paul Krugman (1), Paul Milgrom (1), Paul Novostad (1), Paul Romer (1), Pawel Adrjan (1), Pedro Souza (1), Pengpeng Xiao (1), Pete Klenow (1), Peter Buhlmann (1), Peter Cramton - Lessons (1), Peter Feldhutter (1), Pablo D'Erasmo (1), Ovanes Petrosian (1), Nicolas Morales (1), Otmar Issing (1), Nicolas Vieille (1), Nicolas Ziebarth (1), Nicole Immorlica (1), Nicolás Ajzenman (1), Nikhil Vellodi (1), Nikita Gaponiuk (1), Nikolay A. Krasovskii (1), Michela Giorcelli (1), Nina Balcan (1), Nina Pavcnik (1), Nora Lustig (1), Oded Galor (1), Odilon Câmara (1), Ole (1), Oleksiy Kryvstov (1), Oliver Hart - Prosocial (1), Olivier Darmouni (1), Olle Hammar (1), Omer Tamuz (1), Ori Heffetz (1), Oster (1), Nimmi Patel (1), Michael Greenstone (1), Micheala Giorcelli (1), Luisa Hammer (1), Linda Goldberg (1), Linda Schilling (1), Ling Zhou (1), Lingfei Wu (1), Lloyd Dean (1), Lones Smith (1), Lorenzo Caliendo: (1), Luciano Pomatto (1), Lucie Gadenne (1), Lukas Delgado-Prieto (1), Li (1), M. Clemens (1), M. Spence (1), Maarten Lindeboom (1), Madhuparna Ganguly (1), Maggie Jones (1), Majid M. Al-Sadoon (1), Manoj Pradhan (1), Manuel Adelino (1), Manuel Tong (1), Lin Tian (1), Lester T. Chan (1), Marc Meryon (1), Laura Parisi (1), Kyungmin Kim (1), L. Christensen (1), Larry Katz (1), Larry Summers (1), Lars Vilhuber (1), Laszlo Tetenyi (1), Laura Castillo Martinez (1), Laura Doval (1), Laura Gati (1), Laura Pilossoph (1), Leslie Marx (1), Laurent Clerc (1), Lawrence Carin (1), Leah Platt Boustan (1), Leandro Navarro (1), Leeat Yariv (1), Leigh Shaw-Taylor (1), Leon Musolff (1), Leonardo Bursztyn (1), Leonhard Lades (1), Marc Melitz (1), Marcel Fratzscher (1), Michaela Kreyenfeld (1), Michael Barnett (1), Matthew Gentzkow - Ideological (1), Matthieu Gomez (1), Mattia Fochesato (1), Maureen O'Hara (1), Max Winkler (1), Maximilian Kasy (1), Maya Rossin-Slater (1), Meredith Crowley (2), Peter Ingram (1), Matteo Gamalerio (1), Michael Grubb (1), Michael Hallsworth (1), Michael Jordan (1), Michael Kearns (1), Michael Keen (1), Michael Kremer (1), Michael Marder Upheaval (1), Michael Richards (1), Michael Woodford (1), Matthew Clair (1), Matt Lasmanis (1), Marcin Pęski (1), Martin Eichenbaum (1), Marco González-Navarro (1), Margaret Meyer (1), Maria Balgova (1), Maria Sole Pagliari (1), Mariana Mazzucato (1), Marie Claire Villeval (1), Mark Lowcock (1), Marshall Burke (1), Martha Justus (1), Martin Ravallion (1), Matilde Bombardini (1), Martin Weale (1), Martina Björkman Nyqvist (1), Martina Kirchberger (1), Martín Fernández-Sánchez (1), Mary Amiti (1), Mary Barra (1), Marzena Rostek - Decentralized (1), Masao Fukui (1), Massimo Anelli (1), Peter Hull (1), Rachel Cummings (1), Peter Taylor-Gooby (1), Stefan G. Hofmann (1), Tayfun Sonmez (1), Ted Miguel (1), Thies Lindenthal (1), Thomas Crossley (1), Thomas Philippon (1), Thomas Rivera (1), Thomas Schmitz (1), Thomas Thévenin (1), Thomas de Haan (1), Tianyi Wang (1), Tijan Bah (1), Tilman Börgers (1), Tim Roughgarden (1), Tito Boeri (1), Tobias Klein – 09/09/20 (1), Tobias Salz

(1), Tony Cookson (1), Trang Hoang (1), Trish Greenhalgh (1), Tyler Muir (1), Ulrich Laitenberger (1), Tarun Ramadorai- (1), Tarun Kabiraj (1), Tamer Başar (1), Steve Callander (1), Stefania Albanesi (1), Stefania Garetto (1), Stefano Caria (1), Stefano Giglio (1), Steffan Mau (1), Stephan Meier (1), Stephane Hallegatte (1), Stephen Machin (1), Stephen Morris (1), Steve Redding (1), Tamar Oostrom (1), Steven Ruggles (1), Steven Stillman (1), Suanna Oh (1), Sukwoong Choi (1), Sumit Agarwal (1), Sven Rady - Overcoming Free-Riding (1), Swapnika Rachapalli (1), Sydney Ludvigson (1), Syngjoo Choi (1), Ulrich Volz (1), Ulrike Malmendier (1), Utsav Sadana - Nash Equilibria (1), Yuen Yuen Ang (1), Yang Zhou (1), Yannay Spitzer (1), Yeon-Koo Che (1), Yeon-Koo Che - Weak (1), Yiling Chen - Cursed (1), Yingni Guo (1), Yingni Guo - Project (1), Yoram Halevy (1), Yossef Rapoport (1), Yuhei Miyauchi (1), Yajna Govind (1), Yvonne Giesing (1), Zhijun Chen (1), Zhou Yu (1), Ziad Obermeyer (1), Zlatko Bodrozic (1), Zoe Cullen (1), Éva Tardos (1), Éva Tardos - Virtues (1), Ömer Karaduman (1), Yan Chen (1), Xiaotie Deng - A (1), VMACS Jr. - Victoria Gregory (1), Vincent Meisner (1), Vadim Elenev (1), Van der Ploeg (1), Vasco Carvalho (1), Vassilis Zikas (1), Verena Weber (1), Vernon Henderson (1), Veronica Guerrieri (1), Victor Chernozhukov (1), Vili Lehdonvirta (1), Vivian Lee (1), Xiaosheng Mu - Privacy (1), Viviane Sanfelice (1), Vladimir Smirnyagin (1), W. Nordhaus (1), Wagner F. Oliveira (1), Warwick McKibbin (1), Willemien Kets (1), Willi Mutschler (1), Xiang Ding (1), Xiaolan Fu (1), Stefan Nagel (1), Sonia Jaffe (1), Peter Wendell (1), Sonia Bhalotra (1), Rebecca Henderson (1), Rebecca Myerson (1), Rebecca Sachs (1), Rema Hanna (1), Renato Faccini (1), Renato Gomes (1), Renato Paes Leme (1), Ricardo Reyes-Heroles (1), Richard Blundell (1), Richard S.J. Tol (1), Rick van del Ploeg (1), Robert Hill (1), Robert Inklaar (1), Robert J. Aumann (1), Robert Mendelsohn (1), Robert Pindyck (1), Robert Pollin (1), Robert Reich (1), Robert Stavins (1), Robert Wilson (1), Robert Zymek (1), Rebecca Dizon-Ross (1), Raymond Fisman (1), Raoul van Maarseveen (1), Rabah Amir - Profit- (1), Petra Moser (1), Petra Todd (1), Philip Lane (1), Pierre Yared (1), Pierre-François Weber (1), Pol Antras (1), Pol Antràs (1), Ponce Del Castillo (1), R. Gerlagh (1), Krusell (1), Rann Smorodinsky - Reaping (1), Rachel E. Kranton (1), Rachel Griffith (1), Rachid Laajaj (1), Rahul Deb (1), Raissa Fabregas (1), Ramon Faulí Oller - Fee (1), Ran Spiegler (1), Ran Spiegler - Cheating (1), Randall Akee (1), Roberto Weber (1), Robin Allen (1), Rod Garrat (1), Sharon Traiberman (1), Sebastian Heise (1), Seema Jayachandran (1), Sergei Guriev (1), Sergiu Hart (1), Seth Benzell (1), Sevgi Yuksel (1), Sharad Goel (1), Sharada Davidson (1), Sharat Ganapati (1), Sheri Berman (1), Schumpeter (1), Shota Ichihashi (1), Siddharth Suri (1), Silvia Peracchi (1), Silvio Micali (1), Simon Deakin (1), Simon Grant (1), Simon Loertscher (1), Sinan Aral (1), Soeren Henn (1), Sean Higgins (1), Schoar (1), Romer (1), Sabrina Howell (1), Rosemarie Nagel (1), Rotman School (1), Russell Cooper (1), Rutger Hoekstra (1), Ruth (1), Ryan Monarch (1), Ryan Oprea (1), Ryland Thomas (1), S. Nageeb Ali - Reselling (1), Saleemul Huq (1), Sascha Becker (1), Saleh (1), Salvatore Carrozzo (1), Samina Raja (1), Sandra Sequeira (1), Sanna Ojanpera (1), Sara Giunti (1), Sara Signorelli (1), Sarah Eichmeyer (1), Sarah Hawkes (1), Kwabena Baah Donkor (1), Katherine Eriksson (1), Kose John (1), Colin Green (1), Conor Walsh (1), Constantin Charles (1), Costas Arkolakis (1), Costas Meghir (1), Cristina Arellano (1), Cristina Bicchieri (1), Cristina Cattaneo (1), Cynthia Osborne (1), Cynthia Zhang - A (1), Damon Centola (1), Damon Silvers (1), Dani Rodrik (1), Danial Lashkari (1), Daniel Reck (1), Daniel Rock (1), Daniel Yi Xu (1), Daniela Saban (1), Daniele Nosenzo (1), Danny Dorling (1), Danny Quah (1), Darrell Duffie (1), Conor Lennon (1), Clément Bellet (1), Eric Bettinger (1), Clodomiro Ferreira (1), Charly Porcher (1), Chenuyan Liu (1), Chistoph Boehm (1), Chris Knittel (1), Chris Roth (1), Chris Warhurst (1), Christian Catalini (1), Christina Gathmann (1), Christoph Rothe (1), Christoph Trebesch (1), Christopher Giancarlo (1), Christopher James (1), Christopher Pissarides (1), Cian Ruane (1), Claire Celerier (1), Clare Short (1), Claudia Custodio (1), Claudia Steinwender (1), Claudio Mezzetti (1), Clemens Hetschko (1), Cliff Robb (1), Dashun Wang (1), Dave Rand (1), David Atkin (1), David Delacrétaz (1), Donald Rubin (1), Dorothea Kuebler (1), Drew Fudenberg (1), Duncan Gallie (1), Duncan Thomas (1), Duo Qin (1), EHEC Finkelstein (1), Edoardo Cefalà (1), Edoardo Gallo (1), Edward Miguel (1), Ekaterina Gromova (1), Ekaterina Oparina (1), Ekaterina Smetanina (1), Elena Manresa (1), Elina Ribakova (1), Elizabeth Stuart (1), Elliot Lipnowski (1), Elliott Ash (1), Emi Nakamura (1), Enrico Spolaore (1), Eran Shmaya - Disentangling (1), Dmitry Mukhin (1), Dirk Bergemann - Search (1), Dirk Bergemann (1), Davide Furceri (1), David F. Hendry (1), David Hesmondhalgh (1), David Kohn (1), David Lagakos (1), David Laibson (1), David Rand (1), David Thesmar (1), David Yanagizawa-Drott (1), David Yermack (1), Dean Eckles (1), Dimitra Petrakaki (1), Dean Yang (1), Debra Howcroft (1), Denis Tkachenko (1), Dennis Novy - Trade (1), Derya Guer-Seker (1), Devaki Ghose (1), Diane Coyle (1), Diego Aycinena (1), Diego Känzig (1), Charles Manski (1), Charles Calomiris (1), Catia Batista (1), Anke Hassel (1), Amy O'Hara (1), Ana Beatriz Galvao (1), Ana Cecilia Fieler (1), Anatole Cheysson (1), Andrea Civelli (1), Andreas Blume (1), Andreas Kleiner (1), Andreas Moxnes

(1), Andreas Veneris (1), Andrei Hagiu (1), Andrei Levchenko (1), Andrei Simonov (1), Andres Rodriguez Clare (1), Andrew Atkeson (1), Andrew Caplin (1), Andrew Clark (1), Andrew Foster (1), Andrew Hinkes (1), Andrew Patton (1), Andrew Rhodes (1), Andy Charlwood (1), Amit Seru- (1), Alyson Plumb (1), Alexey Onatskiy (1), Akosua Adomako Ampofo (1), Abi Adams-Prassl (1), Abigail Adams Prassl (1), Abigail Marks (1), Adam Dutton (1), Adam Posen (1), Adam Spencer (1), Adi Sunderam (1), Adrien Bilal (1), Ahmad Lashkaripour (1), Alan Blinder (1), Alexandra Mousavizadeh (1), Alan Davidson (1), Alan Manning (1), Alessandro Ferrari (1), Alessandro Pavan (1), Alessandro Ruggieri (1), Alessandro Sforza (1), Alex Hollingsworth (1), Alex Teytelboym (1), Alexander Frankel - Information Hierarchies (1), Angus Deaton (1), Anna Gassman-Pines (1), Catherine Eckel (1), Anna Maria Mayda (1), Bengt Holmstrom (1), Benjamin A. Olken (1), Benjamin Bernard (1), Benjamin Brooks - A (1), Benjamin Golub (1), Benoit Mojon (1), Bergemann (1), Bhramar Mukherjee (1), Bo Cowgill (1), Britta Rude (1), Bruce D. Meyer (1), Bruno Biais (1), Bruno Strulovici (1), Campbell Harvey (1), Carmen Jacqueline Ho (1), Carol Graham (1), Caroline Le Pennec (1), Carolyn Barnes (1), Caspar Kaiser (1), Catalina Amuedo Dorantes (1), Catalina Franco Buitrago (1), Ben Olken (1), Ben Moll (1), Ben Bridgeman (1), Arnar Buason (1), Anna Rettieva (1), Anna Tur (1), Annie Liang - Data (1), Anthony Lepinteur (1), Antoinette Schoar (1), Antoinette Schoar- (1), Anton Kolotilin (1), Ariel Burstein (1), Arman Sabbaghi (1), Arvind Krishnamurthy (1), Barton Lee (1), Ashvin Gandhi (1), Atheendar Venkataramani (1), Atul Gupta (1), Aviad Heifetz - Liberal Parentalism (1), Avinandan Chakraborty (1), Barbara Petrongolo (1), Barbora Šedová (1), Barry Eichengreen (1), Barry Ritholtz (1), Eran Shmaya - Project (1), Eric Ghysels (1), Kosali Simon (1), Jesús Fernández-Huertas Moraga (1), Jiangtao Li (1), Jie Cai (1), Jimi Adesina (1), Jing Cai (1), Jing Chen - Maximal (1), Jishnu Das (1), Joachim de Weerdt (1), Joan Llull (1), Joe Bergera (1), Joe Kaboski (1), Joe Perkins - 4th (1), Joe Steinberg (1), Johanna Mollerstrom (1), Johannes Boehm (1), Johannes Matschke (1), Johannes Spinnewijn (1), John Asker (1), John Beshears (1), John Cawley (1), John Geanakoplos - Money (1), John Ifcher (1), Jesús Marín Solano (1), Jessica Simes (1), Eric Melander (1), Jessica Hug (1), Itai Arieli - Persuasion (1), Itay Goldstein (1), Ivan Canay (1), Ivan Petrella (1), Ivan Werning (1), J. Aislinn Bohren (1), Jacek Barszczewski (1), Jacob Engwerda - Mini-max (1), Jacob Leshno (1), James Mitchell (1), James Robins (1), James Stock (1), Janelle Scott (1), Janine Berg (1), Jared Gars (1), Jeff Larrimore (1), Jeff MacKie-Mason (1), Jeff Small (1), Jens Ludwig (1), Jens van 't Klooster (1), Jeremy Lebow (1), John Quiggin (1), John Van Reenen (1), John Vickers (1), John Y Campbell (1), Jules Gazeaud (1), Julia Seither (1), Jérôme Valette (1), Kareen Rozen (1), Karen (1), Karlis Kanders (1), Kartik Hosanagar (1), Kate Casey (1), Kate Ho (1), Kate Laffan (1), Katherine A Eriksson (1), Abi Adams Prassl (1), Kathryn Spier (1), Katrin Millock (1), Katy Milkman (1), Kelly Shue (1), Ken Benoit (1), Kenji Wada (1), Kenneth Ahern (1), Kim Ruhl (1), Kirsten Slungaard Mumma (1), Judy Chevalier (1), Juan Ortner (1), Juan Mateos (1), Joop Adema (1), Johns Hopkins (1), Jon Gruber (1), Jon Kleinberg - Algorithmic (1), Jonathan Coslet (1), Jonathan Dingel (1), Jonathan Levy (1), Jonathan Libgober (1), Jonathan Wright (1), Jonathan de Quidt (1), Jordi Gali (1), Juan Diego Luksic (1), Jorge Guzman (1), Jorge Tamayo (1), Jose Marquez Merino (1), Jose Vasquez (1), Joseph-Simon Görlach (1), Josh Blumenstock (1), Josh Martin (1), Josh Ryan-Collins (1), José Daniel López Barrientos (1), Ismael Gálvez Iniesta (1), Isabelle Mejean (1), Isabela Manelici (1), Gianluca De Nard (1), Francesca Caselli (1), Francesca Sobande (1), Francesco D'Acunto (1), Francesco Giovanardi (1), Francisco Buera (1), François Libois (1), Friederike Mengel (1), Fuhito Kojima (1), G. Metcalf (1), Gabriel Ahlfeldt (1), Gabriela Mundaca (1), Gaelle Simard-Duplain (1), Gary Lyn (1), Gautam Rao (1), Geoffrey Heal (1), Geoffroy de Clippel (1), George J. Mailath (1), George Karystianis (1), George Marios Angeletos (1), Georgios Piliouras (1), Gerardo Ceballos (1), Florian Zimmerman (1), Flavio Cunha (1), Filippo Santi (1), Fabio Galeotti (1), Esther Arenas-Arroyo (1), Esther Delesalle (1), Eugenio Proto (1), Eunhee Lee (1), Evan Calford (1), Evgenii Monastyrenko (1), F. Capie (1), Fabian Waldinger (1), Fabien Accominotti (1), Fabrizio Zilibotti (1), Fernando Leibovici (1), Facundo Piguillem (1), Federico Echenique (1), Federico Esposito (1), Federico Etro (1), Federico Huneeus (1), Felipe Brugues (1), Felipe Schwartzmann (1), Felix Tintlenot (1), Fernanda Cross (1), Giancarlo Corsetti (1), Gianluca Violante (1), Isabel Treviño (1), Gijsbert Zwart (1), Haris Aziz (1), Hashem Pesaran (1), Hector Chade - Screening (1), Heiwai Tang (1), Helene Rey (1), Heng Huang (1), Herakles Polemarchakis (1), Horacio Larreguy (1), Hundanol Kebede: (1), Hunt Allcott (1), Hyun Song Shin (1), Hélène Rey (1), Hülya K.K. Eraslan (1), Ignacio Esponda - Asymptotic (1), Immanuel Kant (1), Imran Rasul (1), Ina Simonovska (1), Ina Taneva (1), Ingela Alger (1), Irene Lo - Simple (1), Isa Hafalir (1), Hari Govindan (1), Hans Werner Sinn (1), Hannes Ullrich (1), Gregor Jarosch (1), Gilles Duranton (1), Giorgia Menta (1), Giorgos Stamatopoulos (1), Giselle Montamat (1), Giuseppe Moscelli (1), Glen Weyl (1), Glenn Rudebusch (1), Gloria H.Y. Wong (1), Gollier (1), Gregor Semieniuk (1), Hanna Halaburda (1), Gregory Acs (1), Gregory Mankiw (1), Gregory Thwaites (1), Guillaume Blanc (1), Guillaume Gueguen (1), Ha-Joon Chang (1), Haifeng Xu (1), Haishi Li (1), Hamid Firooz (1), Özlem Bedre-Defolie (1).

## List of all university's affiliation identified (in parhentesis the number of seminars in which they took part).

Department of Economics, Harvard University (58), Economics Department, Massachusetts Institute of Technology (MIT) (33), Department of Economics, Stanford University (25), Economics Department, London School of Economics (LSE) (22), Department of Economics, Princeton University (22), Department of Economics, University of Chicago (21), Department of Economics, Northwestern University (18), Department of Economics, University of California-Berkeley (15), Department of Economics, University of Pennsylvania (15), Department of Economics, School of Arts and Sciences, Columbia University (15), Department of Economics, University of California-Los Angeles (UCLA) (14), Department of Economics, Oxford University (13), Economics Department, Yale University (13), Department of Economics, Boston University (11), Department of Economics, University of Toronto (10), Paris School of Economics (8), Department of Economics, Duke University (8), Department of Economics, New York University (NYU) (7), Department of Economics, Cornell University (7), Faculteit Economie en Bedrijfswetenschappen, KU Leuven (6), Department of Economics, University College London (UCL) (6), Fachbereich Wirtschaftswissenschaft, Goethe Universität Frankfurt am Main (6), Economics Department, University of Michigan (5), Economics Department, Georgetown University (5), Département d'Économie et de Management, Faculté de droit, d'économie et de finance, Université du Luxembourg (5), Economics Department, University of Wisconsin-Madison (5), Department of Economics, University of Southern California (5), Toulouse School of Economics (TSE) (5), Department of Economics, European University Institute (5), Department of Economics, National University of Singapore (NUS) (4), Faculty of Economics, University of Cambridge (4), Department of Economics, University of Warwick (4), School of Economics, UNSW Business School, UNSW Sydney (4), École d'Économie, Université Clermont Auvergne (4), Facultat d'Economia i Empresa, Universitat de Barcelona (3), School of Economics, University of Manchester (3), Departamento de Economía, Universidad Carlos III de Madrid (3), Department of Economics, London Business School (LBS) (3), Economics Department, Brown University (3), Dipartimento di Economia "Ettore Bocconi", Università Commerciale Luigi Bocconi (3), Department of Economics, Indiana University (3), Department of Economics, Boston College (3), Division of Social Sciences, California Institute of Technology (3), Department of Economics, Washington University in St. Louis (3), Department of Economics, Tufts University (3), College of Business and Economics, Australian National University (3), Economics Section, Cardiff Business School, Cardiff University (3), Department of Economics, Hebrew University of Jerusalem (3), Division of Economics, Seoul National University (2), Department of Economics, University of Washington (2), Faculteit Economie en Bedrijfskunde, Rijksuniversiteit Groningen (2), Faculty of Economics, University of Tokyo (2), School of Economics, University of Queensland (2), Economics Department, Dartmouth College (2), School of Economics, University College Dublin (2), Faculteit Economie en Bedrijfskunde, Universiteit van Amsterdam (2), Economics Department, University of California-Davis (2), Nationalekonomiska institutionen, Stockholms Universitet (2), School of Economics, University of Kent (2), Department of Finance and Business Economics, Marshall School of Business, University of Southern California (2), Department of Economics, Trinity College Dublin (2), Department of Economics, University of Colorado (2), Department of Economics, Krannert School of Management, Purdue University (2), Department of Economics, University of Pittsburgh (2), Department of Economics, Pennsylvania State University (2), Department of Economics, Sussex Business School, University of Sussex (2), Department of Economics, Graduate Center, City University of New York (CUNY) (2), Department of Economics, University of California-Santa Barbara (UCSB) (2), Department of Economics, Monash Business School, Monash University (2), School of Economics, University of Nottingham (1), Department of Economics, Auburn University (1), Department of Economics, Lerner College of Business and Economics, University of Delaware (1), Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität Berlin (1), Vakgroep Algemene Economie, School of Business and Economics, Maastricht University (1), Faculty of Business and Economics, University of Hong Kong (1), Department of Economics, Johns Hopkins University (1), Unidade de Estudos sobre Complexidade e Economia (UECE), Research in Economics and Mathematics (REM), Instituto Superior de Economia e Gestão (ISEG), Universidade de Lisboa (1), Department of Economics, Iowa

State University (1), Gordon Lang School of Business and Economics, University of Guelph (1), Institut d'Économie Appliquée, HEC Montréal (École des Hautes Études Commerciales) (1), Institutt for Økonomi, Universitetet i Bergen (1), Nationalekonomiska Institutionen, Uppsala Universitet (1), Department Volkswirtschaft, WU Wirtschaftsuniversität Wien (1), Staatswissenschaftliches Seminar, Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität zu Köln (1), Department of Economics, George Washington University (1), School of Business and Economics, Universidade Nova de Lisboa (1), School of Business and Economics, Vrije Universiteit Amsterdam (1), School of Economics and Management, Universiteit van Tilburg (1), School of Economics, Faculty of Arts and Social Sciences, University of Sydney (1), Department of Economics, Maxwell School, Syracuse University (1), Department of Economics, Faculty of Business and Economics, University of Melbourne (1), School of Economics, Universiteit Utrecht (1), Department of Economics, Emory University (1), School of Economics, University of East Anglia (1), School of Economics, University of Edinburgh (1), School of Economics, University of Surrey (1), Department of Economics, Carleton University (1), Department of Economics, Business School, University of Auckland (1), Department of Economics, University of Texas-Austin (1), Facoltà di Economia, "Sapienza" Università di Roma (1), Facultad de Economía, Universidad de los Andes (1), Department of Economics, University of California-Merced (1), Department of Economics, University of Notre Dame (1), Department of Economics, W.P. Carey School of Business, Arizona State University (1), Department of Economics, University of North Carolina-Chapel-Hill (1), Argyros School of Business and Economics, Chapman University (1), Department of Economics, University of Minnesota (1), Dipartimento di Scienze Economiche, Alma Mater Studiorum - Università di Bologna (1), Department of Economics, University of Massachusetts-Amherst (1), Department of Economics, University of Illinois at Urbana-Champaign (1), Département de Sciences Économiques, Université de Montréal (1), EPGE Escola Brasileira de Economia e Finanças, Fundação Getúlio Vargas (FGV) (1), Department of Economics, University of California-San Diego (UCSD) (1), Economics Department, Eller College of Management, University of Arizona (1), Economics Department, George Mason University (1), Department of Economics, Tulane University (1), Facultad de Economía, Universidad de València (1), Economics Department, Michigan State University (1), Economics Department, Queen's University (1), Department of Economics, Texas Aandamp;M University (1), Economics Department, University of California-Santa Cruz (UCSC) (1), Economics Department, University of Essex (1), Economics Department, University of Florida (1), Economics Department, University of Strathclyde (1), Department of Economics, Stockholm School of Economics (1), Economics Department, Williams College (1), Department of Economics, Rice University (1), Economics Discipline Group, Business School, University of Technology Sydney (1), Department of Economics, Ohio State University (1), Facoltà di Economia / Wirtschaftswissenschaftliche Fakutät, Libera Università di Bolzano / Freie Universität Bozen (1), Abteilung für Volkswirtschaftslehre, Universität Mannheim (1).

# 7 Appendix III

Figure 9: Speaker diarisation

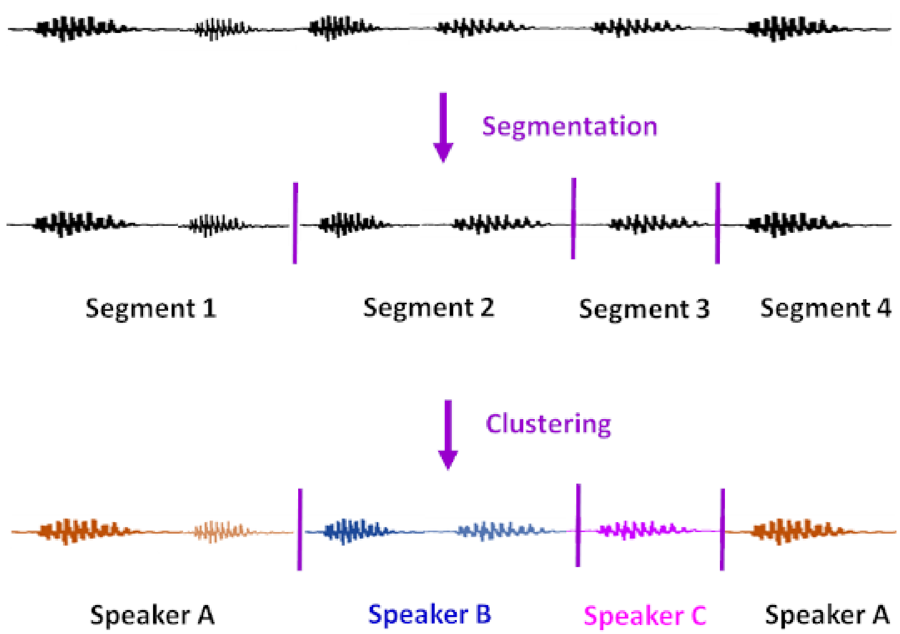


Figure 10: VAD example using an energy based criterion

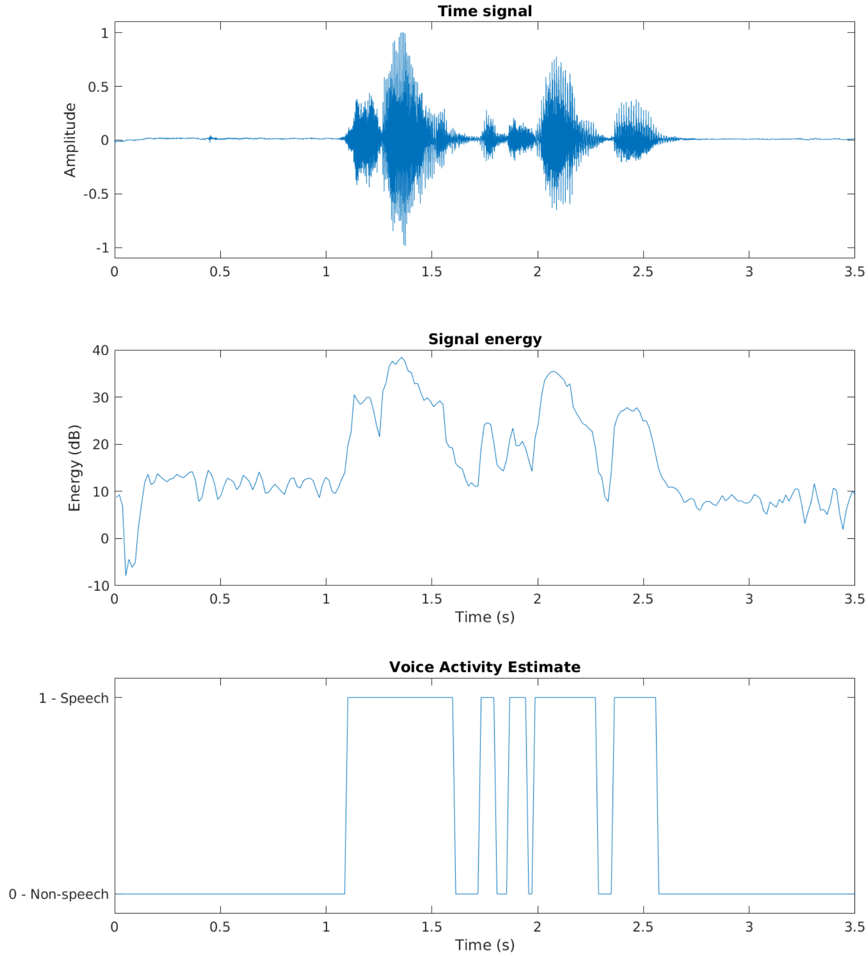


Figure 11: Choosing optimal model with coherence score

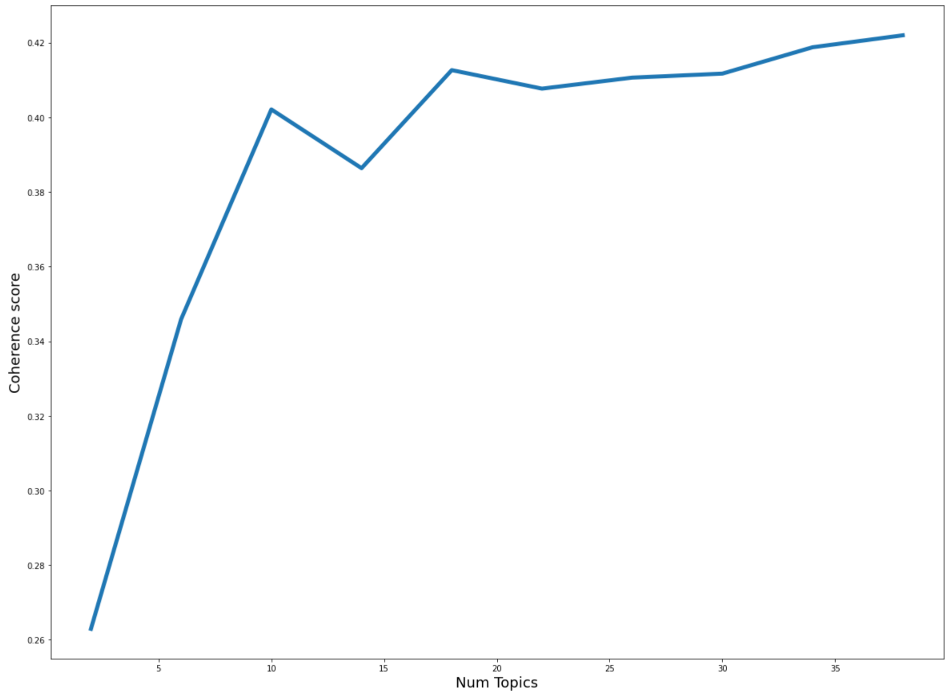



Figure 12: Example of Mel Frequency Cepstral Coefficients

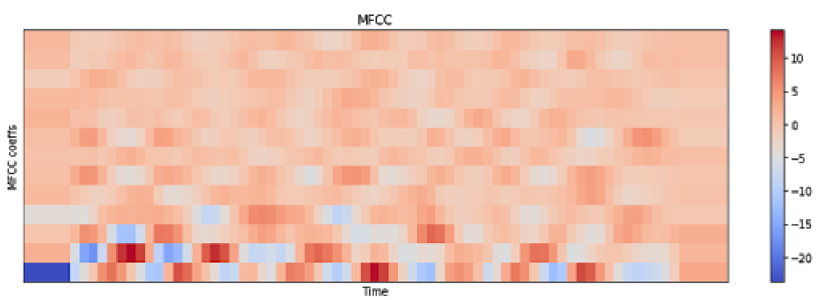



Figure 13: Example of MEL Spectogram