Evolutionary Economic Geography: Including Gender into the Analysis.

A three-stage decomposition of Related and Unrelated Variety

Barbara Martini

University of Rome Tor Vergata

barbara.martinini@roma2.it

Key words: Related Variety, Unrelated Variety, Gender, Knowledge spillovers.

JEL Classification: R00, O35, R11

Highlights

- Gender plays a role in knowledge spillover.
- Related and Unrelated Variety does not consider gender. A three-stage decomposition to capture gender knowledge spillovers is provided.
- Theoretical results highlight that gender and the share of females and males in industries play an essential role in knowledge spillovers.
- Increasing females' employment is an important aim, but increasing females' employment in an industry where the females' share is low can be more effective in labor growth.

Abstract

Evolutionary Economic Geography (EEG). EEG aimed to understand why industries concentrate in the space, how networks evolve, and why some regions grow more than others using three key concepts: proximities, capabilities and routines, and industries relatedness. This literature does not consider gender even though gender can play an important role. Gender differences in social behaviour can impact knowledge's proximities and diffusion, especially when females and males are not equally distributed between industries and firms. Industries relatedness and knowledge spills-over are captured through the Related and Unrelated Variety. The Variety index will be decomposed into Related and Unrelated Variety using a three-stage decomposition to consider gender. The properties of these measures will be investigated using a theoretical approach. The results highlight that RV and UV measures have different behaviour when females' (males) employment increases.

Furthermore, RV and UV will depend on the females' (males) share in the industry, and they exhibit increasing returns to scale when females' share in an industry is lower than the males' share. This finding has significant consequences in terms of policies. Increasing females' participation in the labour market is essential, but it is also crucial in which industries females will be employed. Increasing females in industries in which the females' share is already high are less effective for labour growth than increasing females' share in industries in which the females' share is low.

1. Introduction

Following the definition provided by the World Health Organization, gender is used to describe the characteristics of women and men that are socially constructed. At the same time, sex refers to those that are biologically determined. For example, people are born female or male but learn to be girls and boys who grow into women and men. This learned behavior makes up gender identity and determines gender roles. The economic theory included gender in the analysis starting from the contributions of Hakim (2000), in which the author uses the preference to explain females' behavior and choices between employment and family work. Outset from this pioneering contribution, a growing body of economic literature aims to investigate the different behavior of females and males in the job market and its consequences on gender segregation[1]. Gender segregation, actual dominance of one sex in a particular occupation or the higher share of one sex relative to the expected share, can be horizontal such as vertical. The first one is generally pictured as women and men's disparate concentration across industries and occupations. In contrast, the second one refers to gender disparities in positions and roles with different statuses or employment advancement potential. Employment segregation often constraints females' labor force participants. However, higher participation rates may not imply greater gender equality or female empowerment. This unequal distribution of females and males between industries means that an increase of females in the labor market will be not equally distributed between industries. As a result, females have a greater probability of falling in some industries than others. Ngai and Petrongolo (2017) pointed out that females' have a comparative advantage in producing services. The evolution of production structure, the de-specialization process taking place in Italy since 1995 (Martini 2020), has increased the service share and, consequently, the females' employment in service sectors (Olivetti and Petrongolo 2016, Petrongolo and Ronchi 2020). Addressing employment segregation is central to reducing the gender wage gap, improving job quality and earnings, and increasing female labor force participation. Employment segregation has significant consequences for overall economic growth, household welfare, firm performance, and intergenerational social mobility. Efforts to reduce employment segregation can create a virtuous cycle in which increased female participation in high return occupations creates more extensive networks of women and changes social norms (Das and Kotikula 2018). Female participation in the job market is also conditioned by national institutions such as welfare regimes, social policies, employment protection legislation (Hall et al., 2019), and cultural norms (Alesina et al., 2011).

Despite this increasing interest in including gender in the analysis of different fields, gender has not yet been considered in Evolutionary Economic Geography (EEG). EEG aimed to understand why industries concentrate in the space, how networks evolve in the space and why some regions grow more than others (Boschma and Frenken 2015) using three key concepts: proximities, capabilities and routines, and industries relatedness. Proximities play a central role in understanding interactive learning and innovation (Boschma 2005). Following Boschma (2005), five forms of proximity can be identified: cognitive, social, institutional, organizational, and geographical. These proximities influence the diffusion of knowledge and, consequently, innovation. Gender differences in social behavior have been explained by the social role theory (Eagly 1987; Eagly and Wood 2012). Females and males have different behavior due to the roles they are engaged in are usually associated with diverse requirements (Eagly 2000). Consequently, they have different skills and capabilities and different behavior in the structure of social networks and the attitude toward them (Collischon and

---

Chapter 1 EIGE https://eige.europa.eu/thesaurus/terms/1304

Eber 2020, Emmerik 2006, Brashears 2008). For example, females focus on ties that provide friendship and emotional support, while males focus on job-related information. These different attitudes can impact skill-relatedness connectivity, proximities, and the diffusion process of knowledge, especially when females and males are not equally distributed between industries and firms.

Furthermore, EEG deals with the uneven distribution of economic activities across space, and it focuses on the historical process that produces these patterns, and the Nelson and Winter (1982) organizational routines are taken as a unit of analysis (Boschma and Frenken 2011). As Ter Wal and Boschma 2007 pointed out, firms broadly differ in their capabilities, strategies, and routines. However, Nelson and Winter's routines assume homogeneity of human capital (Eber 2018). Nevertheless, individuals are not homogeneous in organizations. They can have different values, beliefs, preferences and, different gender (Felin et al., 2012). These differences are central to understanding the organization's outcome level (Abel et al., 2008, Grant 1996, Simon 1991). Gender diversity in the organizations can enhance knowledge outcome (Maes et al., 2012, Valentine and Collins 2015) and gender diversity among teams' members can enhance the innovation process by offering various diverse ideas and mindsets. It can promote innovation and creativity within the groups generating informal advantages (Sastre 2015, Ostergaard et al., 2011, Jackson et al., 1995, O'Reilly et al., 1997, Xie et al., 2020). In addition, gender diversity yields social benefits to organizations by introducing various value systems and behavioral modes. For example, males tend to be more assertive and task-oriented, whereas females tend to be friendly, agreeable with others, and process-oriented (Karakowsky and Siegel 1999, Myaskowsky et al., 2005, Wood et al., 1987). Finally, meta-analytic studies highlighted that gender-mixed groups outperform homogeneous groups (Bowers et al., 2000, Williams and O'Relly 1988, Wood et al.,1987). Consequently, the hypothesis that individuals are homogeneous in an organization should be relaxed when gender is included in the analysis.

For which concern the industries relatedness and knowledge spills-over, the question at the core of these studies is whether firms learn more from local firms in the same industry –regional specialization; or from local firms in other industries – regional diversity. In other words, are the most innovative and fast-growing regions sectoral specialized or diversified? (Iammarino 2011). Even though regional diversification is recognized as a critical factor in creating new growth paths and offset stagnation (Boschma and Gainelle 2014), there is no univocal causal relationship between regional industrial structure and economic growth. In contrast with Marshall Arrow Romer externalities (MAR) generated by firms in the same industry, Jacobs' externalities refer to externalities generated by different industries. Under this point of view, knowledge spills over across industries are the most critical source of innovation, and economic diversity is a key to urban prosperity. As Jacobs (1969) pointed out, diversity does not imply knowledge spillovers and growth. "To reap the benefits from diversity, 'various efficient economic pools of use' are needed, and that in districts or regions lacking the ability to create such effective pools of use and economic interactions, diversity may be associated with stagnation and decline" (Jacobs, 1961, 194). Knowledge spillover exists when complementarities exist among industries in terms of shared competencies and capabilities (Jacobs, 1969). As recognized by Jacobs herself, knowledge spills over effectively only when complementarities exist among industries in terms of shared competencies and capabilities. Sufficient knowledge spillovers for innovation will not exist in cases where cognitive distance is too large (i.e., regional diversity is too high). When the cognitive distance is too high, knowledge has difficulties being reorganized (Aarstad et al., 2016). Knowledge spills over between related sectors facilitate the recombination of knowledge in entirely new ways and, thereby, innovation. Knowledge will spills-over from one industry to another only when the industries are complementary in shared competencies. The crucial point is the right balance between cognitive distance and proximity, allowing for innovation and interaction (Noteboom 2000). Literature captured industry-relatedness through the Related and Unrelated Variety (Boschma, 2005; Frenken et al., 2007; Boschma et al., 2009; Boschma and Iammarino, 2009) based on Theil index (1965, 1967). Related Variety would support innovation as it spreads within industries and captures Jacobs' externality. The Unrelated Variety captures the diversity between industries, and, according to the portfolio theory, it can be considered a strategy to protect the region from asymmetric shock and, consequently, protect the labor market from unemployment. The variety decomposition used by Frenken et al. (2007) is based on a two-level hierarchical structure where the first level represents the industries at 2-digit and the second level represents the industries at 5-digit. The literature has explored the relationship between gender, diversity, and innovation with mixed results. Some studies have found a negative link between them (Beghetto, R. 2010), while others (Hernández-Lara et al., 2021, Torchia et al., 2018) highlight a positive and significant difference contribution of women managers in R&D and organizational innovation. And that firms with gender-diverse boards have more patents, more novel patents, and higher innovative efficiency (Griffin et al., 2019). Consequently, the process of knowledge spill over and the complementarities can be hindered by gender.

In conclusion adding the gender dimension in EEG means rethinking the literature from its micro-foundations relaxing the assumption that individuals are homogeneous in an organization. Furthermore, proximities can be affected not only by gender but also by gender segregation. Females and males are not employed in the same industries, consequently the proximities can be influenced by differences in terms of gender. Finally, the industry relatedness needs to be rethanked. Innovation and knowledge spillovers can be affected by gender and the growth can be faster or slower also by the presence of a more equal or unequal females labor force participation or a lower gender segregation.

In this contribution, we examine the concept of Related and Unrelated Variety when gender dimension is added. The entropy measures proposed by Frenken et al., (2007) need to be modified to include gender in the analysis. The novelty of the contribution is a three-stage decomposition which allows us to investigate regional knowledge spillovers also in terms of gender. As result five entropy measures will be obtained. The first one will capture the diversity between gender in the whole economy while the other four are Related and Unrelated Variety by gender. These measures are a decomposition by gender of Unrelated and Related Variety measure proposed by Frenken et al., (2007) and they capture the gender differences between and within industries. To our knowledge, this is the first contribution in literature exploring the relationship between gender, knowledge spillovers, and labor growth.

The paper is organized as follows. The next section provides a conceptual framework while the next summarizes the Related and Unrelated Variety obtained when a one-stage decomposition is used. A two-stage variety decomposition, including the gender dimension into the analysis, is presented next, followed by the study of the results obtained. The last section concludes and discusses policy implications.

2. Conceptual framework

EEG literature captured the proximity between and within industries through the Related and Unrelated Variety. However, the approach followed by Frenken et al., (2007) is based on total employment in 5-digit and 2-digit levels and, it does not include gender in the analysis. Nevertheless, females and males can have uneven distribution and the whole economy, as shown in Figure 1.

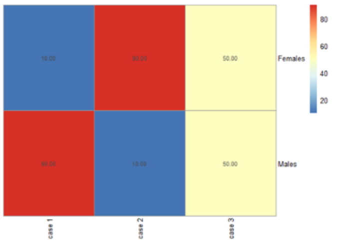
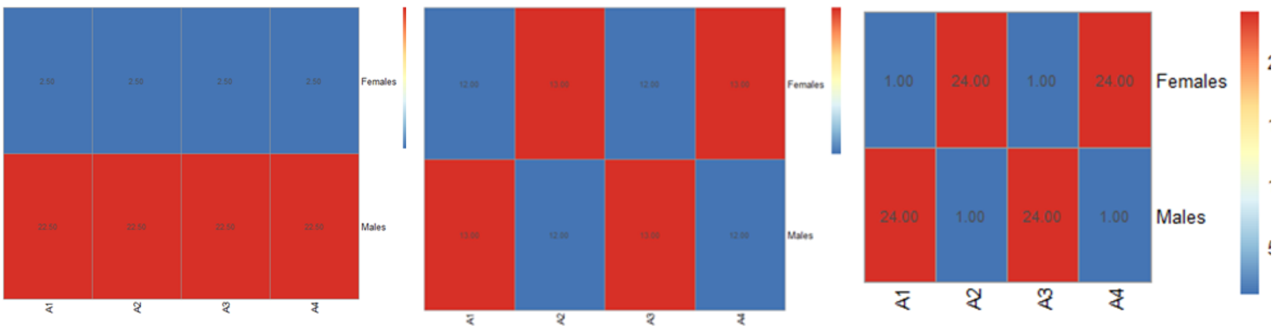


Figure 1: Share of females and males in the whole economy

Figure 1 depicts the females' and males' share in the whole economy. In case1, the males' share is 90% while the females' share is 10%. In case 2, females' share is 90%, and males' share is 10%. Finally, in Case 3, females and males are equally distributed. All three cases have the same number of employed, but they describe very different situations in terms of gender. Following the previous discussion, gender can be a source of differences in terms of knowledge spillovers and growth. Therefore, a diversity measure able to capture the diversity between gender in the whole economy is needed. Furthermore, there may also be differences in terms of gender between 2-digit industries, as shown in Figure 2:



Case 1 Case 2 Case 3

Figure 2: Shares of females and males between the 2-digit industries

Figure 2 describes three different cases. In Case 1, the females' share in the whole economy is 10%, while the males' share is 90%. Females and males are equally distributed between 2-digits industries. In Case 2, the females' and males' share in the whole economy equals 50%. Furthermore, females and males are equally distributed between 2-digits industries. Finally, in Case 3, the females' share in the whole economy is equal to the males' share across the entire economy (50% males and 50% females), but females and males are not equally distributed between 2-digits industries. Cases 1-3 depicted a different situation in terms of gender and, following the literature previously described, knowledge spillover can differ from case to case due to gender. Therefore, the Unrelated Variety between industries needs to be modified. The new measure should be able to capture the diversity between the 2-digit industries considering gender. Finally, the same example can be applied to the 5-digit industries. In conclusion, Related and Unrelated Variety (Frenken et al., 2007) need to be modified to include gender dimension.

3. From Variety to Related and Unrelated Variety: one stage decomposition

Related and Unrelated Variety (Boschma, 2005; Frenken et al., 2007; Boschma and Iammarino, 2009) are based on Theil index (1967) that is an inequality measure related to Shannon (1948) entropy. The antilog of Shannon's entropy possesses three properties that support its use as an index of industrial Variety (Straathof 2006). First, the index reduces to the number of industries if all industries have the same weight. Second, the index can be decomposed. Third, it ensures that adding or removing an industry with a zero weight does not change the value of the variety index. Variety has the characteristics of being decomposable. It can be decomposed into Variety between sectors -Unrelated Variety-, at a higher level of aggregation (usually at 2-digits), and variety within the sectors -Related Variety- at a lower level of aggregation (commonly at 5-digits). The higher is the Related and Unrelated Variety, the lower is the industry concentration. Related Variety would support innovation as it spreads within industries and captures Jacobs' externality. The Unrelated Variety captures the diversity between industries, and, according to the portfolio theory, it can be considered a strategy to protect the region from asymmetric shock and, consequently, protect the labor market from unemployment. By Frenken et al., (2007), economic activities can be classified as a two-level hierarchical structure that is a tree structure with two layers: the roots, level A, and level B. The first level, the root node, includes only one node. It represents all the economic activities. Level A represents the industries at 2-digit (division by NACE Rev.2)[2] while level B represents the industries at 5-digit (the category by Ateco 2007) as shown in Figure 3:

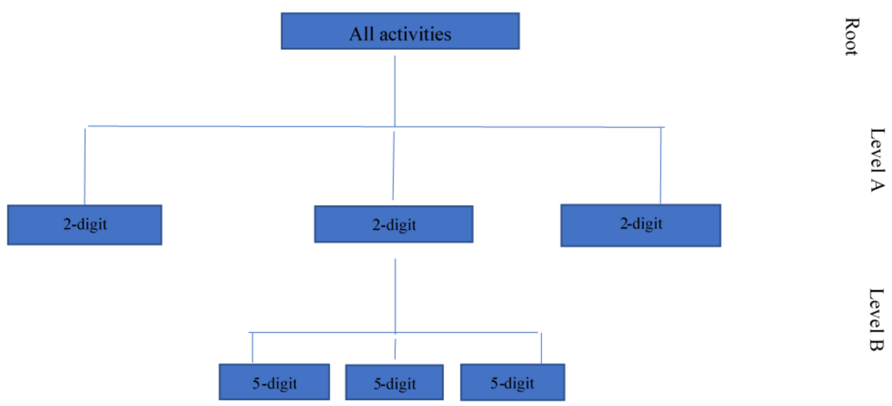


Figure 3: Two levels hierarchical structure

[2] https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF

Following Attran (2006), the variety measure -VAR- will compare the regional employment distribution against a uniform employment distribution where the employment is equiproportional in all the sectors. The measure of industrial Variety in each region is:

$$VAR(P_1, P_2, \ldots\ldots P_n) = -\sum_{i=1}^{N} P_i \, log_2(P_i) \quad (2.1)$$

where $P_i = \frac{E_{ij}}{E_j}$, E_{ij} represents the number of employed at 5-digit in industry *i* in region *j*, E_j is the total employment in region *j*, and *N* is the number of industries at the 5-digit level. The equation (2.1) can be rewritten as[3]:

$$VAR(P_1, P_2, \ldots\ldots P_n) = -\sum_{i=1}^{N} P_i \, ln(P_i) \quad (2.2)$$

Moreover, $0 \leq VAR \leq ln(N)$ where VAR=0 corresponds to the highest concentration, and it occurs when $E_{ij} = 1$ i.e when all the employment is concentrated in only one industry while $VAR = ln(N)$ represents the greatest level of dispersion and corresponds to an equi-proportional distribution in all the sectors. VAR depends only on N, $P_1, P_2, \ldots\ldots P_n$, and is a symmetric continuous function of the p's depending only on their relative magnitude and not on their order. If one industry, the n^{th}, was subdivided into two sub-industries with relative share q_1 and q_2, the new measure of Variety is the original measures plus the conditional diversity within the industries (Hackbart and Anderson 1975). Nevertheless, VAR does not highlight which is the optimal combination i.e. different combinations can have the same Variety.
Using a pre-given hierarchical classification, as provided by the Nace_2 Rev classification, the Variety index can be decomposed into Related and Unrelated Variety. The one-stage decomposition is provided in Appendix 1 and the Variety in (2.2) becomes:

$$VAR = RV + UV \quad (2.3)$$

UV varies from 0, all employment is concentrated in only one two-digit industry, to $ln(G)$, all the industries employ an equal number of employees. RV can range from 0 (employment in each two-digit sector is concentrated in only one of its five-digit industries) to $ln(I) - ln(G)$ when all five-digit industries within a two-digit industry have an equal employment share (Castaldi et al., 2015, Fritsch and Kublina 2016).

4. Including gender into the analysis: the two-stage variety decomposition.

Variety and its decomposition in Related and Unrelated Variety previously analyzed do not consider gender. Including gender in the analysis means including another dimension. Consequently, entropy decomposition will be based on a three-level hierarchical structure (Akita 2001, Wu et al., 2018). The hierarchical structure previously described needs to be modified, and Figure 2 previously showed becomes:

[3] Shannon builds its index taking into consideration 2, 10 and *e* as logarithm bases.

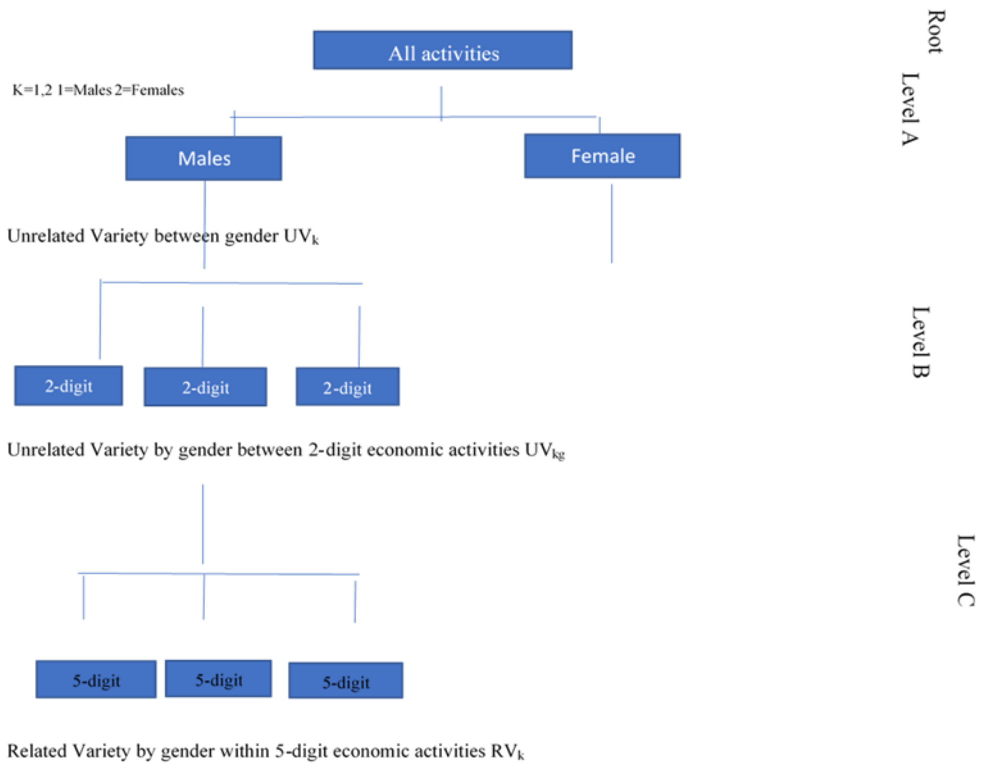


Figure 4: Three levels hierarchical structure

Economic activities are now classified as a three-level hierarchical structure that is a tree structure with three layers: the roots, level A, level B, and level C. Under Figure 4, the total employment in region j, E_j, will be divided into K subgroups with K=1,2; where K=1 represents the numbers of males employed in region *j* and K=2 represents the number of females employed in region *j*; $P_k = \frac{E_{kj}}{E_j}$ is the share of females/males employed in region *j* on the total employment in region *j* with $\sum_{k=1}^{2} P_k = E_j$. Furthermore, $P_{kg} = \frac{E_{kgj}}{E_j}$ represents the share of females/males employed in 2-digit industry *g* in region *j* on total employment in region *j* with $\sum_{g=1}^{G} P_{kg} = P_k$. The employment in 2-digit industry *g* is distributed between 5-digit industries *i*. $P_{ki} = \frac{E_{kij}}{E_j}$ represents the share of females/males employed in the 5-digit industry *i* in region *j* on total employment in region *j* with $\sum_{i=1}^{N} P_{ki} = P_k$. By the decomposition provided in Appendix 2 the Variety measure can be decomposed as follows:

$$VAR = \underbrace{UV_k + UV_{kg}}_{UV} + \underbrace{RV_F + RV_M}_{RV} \quad (3.1)$$

This new decomposition provides five different measures. UV_k represents the diversity by gender between all the economic activities, and it captures the disparities between females and males in the whole economy. UV_K varies from 0 to ln(K). UV_k=0 when $\frac{E_k}{E} = 1$ i.e when all employed belong to the same gender. When UV_K= ln(K) females and males are equally distributed between economic activities. The term UV_k is symmetric to the females' (males') and it reaches its maximum in correspondence of ln(2). The term UV_{kg}, with K=M;F, represents the Unrelated Variety between industries at a 2-digit level considering the females/males' share in the industry at a 2-digit level. The sum of UV_k and UV_{kg} gives the UV as in Frenken at al., (2007). Finally, the term RV_k, K=M;F, is the Related Variety (RV) at a 5-digit level in each industry, considering the females/males' share in the industry at a 5-digit level.

5. Exploring Related and Unrelated Variety measures

The previous analysis highlighted that knowledge spillovers between and within industries could be influenced by gender. Furthermore, the shares of females and males in different industries differ. Therefore, we aim to investigate the RV and UV behavior when the females' share increases to examine if and how RV and UV measures can be affected by the gender share. Three different cases will be considered. The first one analyzes the RV changes when the females' increase in only one 5-digit industry, considering all the others unchanged. The second one will investigate the change of UV at the 2-digit level when the females' employment increases in only one 2-digit industry. Finally, the UV between females and males will be inspected when the females' employment increases in the whole economy. Our analysis is based on the hypothesis that the males' employment is constant and remains unchanged.

Question 1: How does RV change when the numbers of females increase in only one 5-digit industry?

The results obtained in Appendix 3 highlight that if females' employment increases in only one 5-digit industry, all the other things being equal, this change will impact the RV measure. Nevertheless, RV, under our decomposition, is the sum of RV_F and RV_M where RV_F captures the differences within industries at the 5-digit level for females and RV_M captures the differences within industries at the 5-digit level for males. Furthermore, our results highlight that RV_F and RV_M display different behavior. When the females' employment in one 5-digit industry increases- EF_{11}- the RV_F increases only if the share of females in the industry is lower than the equal distribution. Once the equal distribution is reached, an increase in females' employment will cause a decrease in RV_F, as shown in Figure 5.1. This result highlights that spillovers effects depend not also on the gender share. Increasing the females' employment in one industry will positively affect Related Variety, but this effect can exhibit decreasing return depending on the females' share in the industry.

Furthermore, as male employment is constant, an increase in females' employment will cause a decrease in RV_M, as depicted in Figure 5.2. RV_M exhibits a trend represented by a branch of hyperbole. Finally, RV behavior, the sum of RV_F+RV_M, is shown in Figure 5.3. This investigation allows us to draw two conclusions. First, gender has an impact on Related Variety. Second, the change in RV will depend on initial conditions that matter.

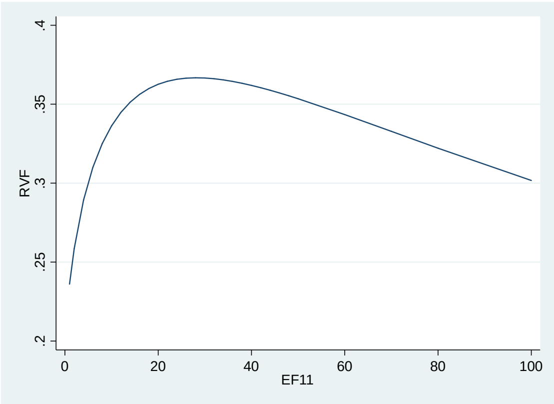


Figure 5.1: changes in RV_F when the numbers of females increase in only one 5-digit industry

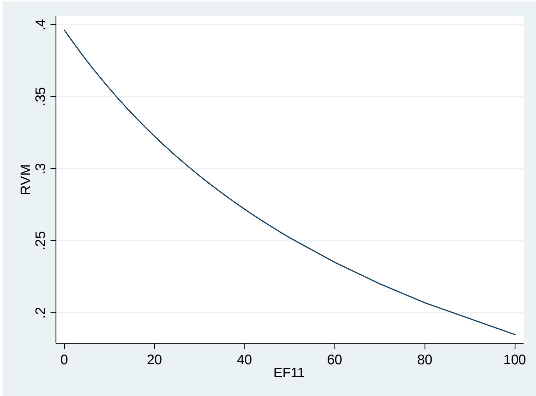


Figure 5.2: changes in RV_M when the numbers of females increase in only one 5-digit industry

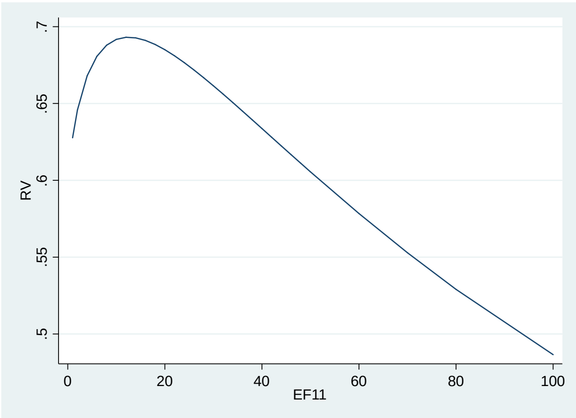


Figure 5.3 : changes in RV when the numbers of females increase in only one 5-digit industry

Question 2: How does UV_{Fg} changes when the numbers of females increase in only one 2-digit industry?

The results obtained for the UV_{kg}, proofed in Appendix 4, are similar to the ones obtained for the RV. When the females' employment in one 2-digit industry increases the UV_{Fg}, the share of females in the industry at 2-digit is lower than the equal distribution, as shown in Figure 6.1. After this point, the UV_F starts to decrease. Thus, UV_M exhibits a trend represented by a branch of hyperbole. Finally, UV behavior, the sum of UV_F+UV_M, is depicted in Figure 6.3. as previously proofed gender has an impact not only on Related Variety but also on Unrelated Variety. This impact will depend on the females' share in the 2-digit industry.

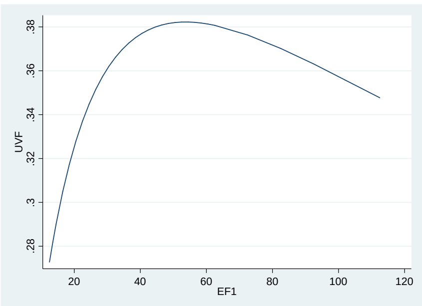


Figure 6.1: changes in UV_{Fg} when the numbers of females increase in only one 2-digit industry

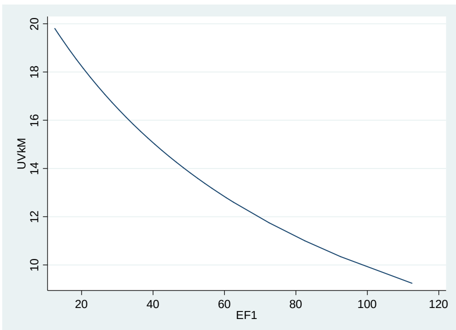


Figure 6.2: changes in UVM_g when the numbers of males increase in only one 2-digit industry

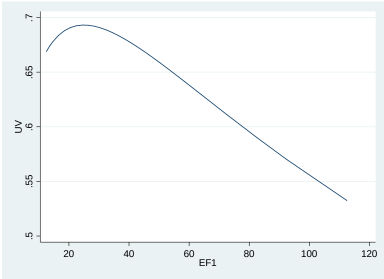


Figure 6.3: changes in UV when the numbers of females increase in only one 2-digit industry

Question 3: How does UV_F changes when the numbers of females increase in the whole economy?

As proofed in Appendix 5, UV_F has the same behaviors as RV_F and UV_{Fg}. It increases when EF<EM and decreases when EF>EM.

The relationship previously explored can be summarized as follows.

Assume an increase of females' employment: $EF\ \uparrow\rightarrow UV_k\ changes$. EM remains unchanged.

$UV_k \uparrow$	if	EF<EM
$UV_k \downarrow$	if	EF>EM

By hypothesis, EF will increase only the employment in one 2-digit industry:

$$\Delta EF = \Delta E_{Fg;g=1}$$

$UV_{Mg} \downarrow$		
$UV_{Fg} \uparrow$	if	$P_{Fg;g=1} = \frac{E_{Fg;g=1}}{E} < P_{Mg;g=1} = \frac{E_{Mg;g=1}}{E}$
$UV_{Fg} \uparrow$ but with decreasing returns	if	$P_{Fg;g=1} = \frac{E_{Fg;g=1}}{E} > P_{Mg;g=1} = \frac{E_{Mg;g=1}}{E}$

By hypothesis $\Delta E_{Fg;g=1}$ will increase only the employment in one 5-digit industry:

$$\Delta E_{Fg;g=1} = \Delta E_{Fi;i=1}$$

$RV_{Mg} \downarrow$		
$RV_{Fg} \uparrow$	if	$P_{Fi;i=1} = \frac{E_{Fi;i=1}}{E} < P_{Mi;i=1} = \frac{E_{Mi;i=1}}{E}$
$RV_{Fg} \uparrow$ but with decreasing returns	if	$P_{Fi;i=1} = \frac{E_{Fi;i=1}}{E} > P_{Mi;i=1} = \frac{E_{Mi;i=1}}{E}$

6. Discussion and policy implications

EEG considers knowledge spillovers as a source of regional economic growth. This approach does not consider gender. Social and economic literature highlighted that females and males have different approaches to innovation, inter-industry collaboration, sharing knowledge and social relation, job preferences. These differences impact knowledge transmission and are amplified when females are males are concentrated in different industries. Our findings highlight that when gender is added to the analysis, the traditional RV and UV measures proposed by Frenken et al., (2007) fail because they cannot capture gender differences between and within industries. Based on Frenken (2007), our contribution proposed a new decomposition in which Related and Unrelated Variety are decomposed by gender and within and between industries. The theoretical analysis highlights that these measures have different behavior when females' (males) employment increases. RV_F and UV_F (as well as RVM and UVM) exhibit increasing returns to scale when females' share in an industry is lower than the males' share, while when females' share in an industry is higher than the males' share, they exhibit decreasing returns to scale. This result opens a new scenario in terms of policies devoted to increasing females' participation in the labor market. Females are more concentrated in some industries than others. Increasing the females' employment in an industry in which the females' share is already high is less effective than increasing the females' employment in an industry in which the females' share is low.

References

Aarstad , J., Kvitastein, O. A., & Jacobsen, S. E. (2016). Related and Unrelated Variety as Regional Drives of Enterprise Productivity and Innovation. A Multilevel Study. *Research Policy, 44*(4), 844-856. doi:https://doi.org/10.1016/j.respol.2016.01.013

Akita, T. (2003). Decomposing Regional Income Inequality in China and Indonesia using Two-stage Nested Model. *The Annals of Regional Science, 37*, 55-77.

Attran, M. (1986). Industrial Diversity and Economic Performance in U.S. Areas. *The Annals of Regional Science, 20*, 44-54.

Azara-Caro, J., Guiterrez-Gracia, A., & Fernandez De Lucio, I. (2007). Faculty Support for the Objectives of University-Industry relations Versus Degree of R&D Cooperation: The Importance of Regional absorptive Capacity. *Research Policy, 35*(1), 37-55.

Beghetto, R. (2010). Creativity in the Classroom. In J. C. Kaufman, & R. J. Stenberg, *Cambridge Handbook of Creativity* (pp. 447-459). NY: Cambridge University Press.

Boschma , R., & Iammarino, S. (2009). Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography, 85*(3), 289-311. doi:10.1111/j.1994.8287.2009.01034.x

Boschma , R., & Martin, R. (2007). Constructing an Evolutionary Economic Geography. *Journal of Economic Geography, 7*(5), 537-548.

Boschma, R. (2005). Proximity and innovation: A Critical Assessment. *Regional Studies, 39*(1), 61-74. doi:10.1080/0034340052000320887

Boschma, R., & Lambooy, J. G. (1999). Evolutionary Economics and Economic Geography. *Journal of Evolutionary Economics, 9*(4), 411-429.

Bozeman, B., & Gaugham, M. (2007). Impacts of Grants and Contracts on academic Researcher. *Research Policy, 36*(5), 694-707.

Brashears, M. E. (2008). Gender and Homophily: Differences in male and Female Association in Blue Space. *Social Science Research*, 400-415.

Castaldi, C., Frenken, K., & Los, B. (2015). Related Variety, Unrelated Variety, and Technological Breackthroughs: an Analysis of the U.S. State-Level Patenting. *Regional Studies, 49*(5), 767-781.

Collishon, M., & Eberl, A. (2021). Social Capital as a Partial Explanation for Gender wage Gaps. *The British Journal of Sociology, 0*, 1-17.

Eagly, A. H., & Wood, W. (1999). Explaining Sex Differences in Social Behaviour. A Meta Analytic Prospective. *Personality and Social Psycology, 17*(5), 306-315.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies, 41*(5), 685-697.

Fritsch, M., & Kublina, S. (2018). Related Variety, Unrelated Variety, and Regional Growth: the Role of Absorbitive Capacity and Entrepreneurship. *Regional Studies, 52*(10), 1360-1371.

Griffin, D., Li, K., & Xu, T. (2012). Board gender Diversity and Corporate Innovation: International Evidence. *Journal of financial and Quantitative Analysis, 56*(1), 123-154.

Hackbart, M. M., & Anderson, D. A. (1975). OnMeasuring Economic Diversification. *Land economics, 51*(4), 374-378.

Hernandez-Lara, A. B., Gonzales-Bustos, J. P., & Alarcon-Alarcon, A. (2021). Social Sustainability on Corporate Boards: The Effects of Female Family Members on R&D. *Sustainability, 13*(4), 1-14.

Jacobs, J. (1969). *The Economy of Cities.* New York: Random House.

Lin, H. F. (2007). Knowledge Sharing and Firms Innovation Capability: an empirical Study. *International Journal of Manpower, 28*(3), 315-332.

Martin, R., & Sunley, P. (2006). Path Dependency and Regional Economic Evolution. *Journal of Economic Geography, 6*(4), 395-437.

Messina, L., Chapman, G., & Hewitt-Dundas, N. (2020). Gender Diversity in R&D Teams and Its Impact on Firm Openess. In H. Lawtonsmith, C. Henry, H. Etzkowitz, & A. Poulovassilis, *Gender, Science and Innovation. New Prospective.* Cheltenham-UK: Edwar Elgar Publisher.

Noteboom, B. (2000). Learning by Interaction: Absorptive Capacity,Cognitive Distance and Governance. *Journal of Management and Governance, 4*, 69-92.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, XXVII*(3), 379-423.

Straathof, S. M. (2007). Shannon's Entropy as an Index of Product Variety. *Economic Letters, 94*(2), 297-303.

Tartari, V., & Salter, A. (2015). The Engagment Gap: Exploring Gender Differences in University-Industry Collaboration Activities. *Research Policy, 44*(6), 1176-1191.

Theil, H. (1967). *Economics and Information Theory.* Amsterdam: North Holland.

Torchia, M., Calabrò, A., Gabaldon, P., & Kanadli, S. B. (2018). Women Directors Contributions to Organizational Innovation: A Behavioral Approach. *Scandinavian Journal of Management, 34*, 215-224.
Van Emmeik, I. J. (2006). Gender Differences in the Creation of Different Type of social Capital: a Multilevel Study. *Social Network*, *28*, 24-37.
Wu, D., Yuan, L., Li, R., & Li, J. (2018). Decomposing Inequality in research Funding by University-Institute Sub-Group. *Journal of Informetrics*, *12*, 1312-1326.

Appendix

Appendix 1: Related and Unrelated Variety using a one-stage decomposition

Following Frenken et al., (2007) the *i*-5-digit industry can fall exclusively under 2-digit industry g, and P_g can be defined as:

$$P_g = \sum_{i \in g} P_i \quad \text{(A1.1)}$$

where $P_g = \frac{E_{gj}}{E_j}$ is the employment at 2-digit in region *j* on the total employment in region *j*. Using equation (A1.1), equation (2.2) can be rewritten as:

$$VAR = \sum_{g=1}^{G} \sum_{i \in g} P_i \ln\left(\frac{1}{P_i}\right) \quad \text{(A1.2)}$$

And, multiplying for $\frac{Pg}{P_g}$, equation (A1.2) becomes:

$$VAR = \sum_{g=1}^{G} \left[\sum_{i \in g} \frac{P_g}{P_g} * P_i \left(ln \frac{P_g}{P_g} \frac{1}{P_i}\right)\right]$$

Applying the log properties, and rearranging:

$$VAR = \sum_{g=1}^{G} \left[\sum_{i \in g} \frac{P_g}{P_g} * P_i \left(ln\left(\frac{P_g}{P_i}\right) + ln\left(\frac{1}{P_g}\right)\right)\right] = \sum_{g=1}^{G} \left[\sum_{i \in g} \frac{P_i}{P_g} * P_g ln\left(\frac{P_g}{P_i}\right) + \sum_{i \in g} P_i ln\left(\frac{1}{P_g}\right)\right] \text{(A1.3)}$$

Using equation (A1.1), equation (A1.3) can be re-written as:

$$VAR = \sum_{g=1}^{G} P_g \left[\sum_{i \in g} \frac{P_i}{P_g} ln\left(\frac{P_g}{P_i}\right)\right] + \left[\sum_{g=1}^{G} P_g ln\left(\frac{1}{P_g}\right)\right] \quad \text{(A1.4)}$$

Defining:

$$H_g = \left[\sum_{i \in g} \frac{P_i}{P_g} ln\left(\frac{P_g}{P_i}\right)\right] \quad \text{(A1.5)}$$

$$\text{RV} = \sum_{g=1}^{G} P_g H_g \quad \text{(A1.6)}$$

$$\text{UV} = \left[\sum_{g=1}^{G} P_g ln \frac{1}{P_g}\right] \quad \text{(A1.7)}$$

The Variety in equation (A1.4) becomes:

$$VAR = RV + UV \text{ (A1.8)}$$

Appendix 2: Related and Unrelated Variety using a two-stage decomposition

By this new classification equation (A1.1) becomes:

$$P_{kg} = \sum_{i \in g} P_{ki} \quad \text{(A2.1)}$$

and equation (A1.2) becomes:

$$VAR = \sum_{k=1}^{2} \sum_{g=1}^{G} \sum_{i \in g} P_{ki} \ln\left(\frac{1}{Pk_i}\right) \quad \text{(A2.2)}$$

Multiplying for $\frac{P_{kg}}{P_{kg}}$ equation (A2.2) can be rewritten as:

$$VAR = \sum_{k=1}^{2}\sum_{g=1}^{G}\sum_{i\in g} P_{ki}\frac{P_{kg}}{P_{kg}}\ln\left(\frac{1}{P_{ki}}\frac{P_{kg}}{P_{kg}}\right)=\sum_{k=1}^{2}\sum_{g=1}^{G}\sum_{i\in g} P_{ki}\frac{P_{kg}}{P_{kg}}\left[\ln\left(\frac{1}{P_{kg}}\right)+ln\left(\frac{P_{kg}}{P_{ki}}\right)\right] =$$

$$VAR = \sum_{k=1}^{2}\sum_{g=1}^{G}\left[\sum_{i\in g} P_{ki}\ln\left(\frac{1}{P_{kg}}\right)+\sum_{i\in g} P_{kg}\frac{P_{ki}}{P_{kg}}ln\left(\frac{P_{kg}}{P_{ki}}\right)\right] \quad (A2.3)$$

Using equation (A2.1) equation (A2.3) can be rewritten as:

$$VAR = \sum_{k=1}^{2}\left(\sum_{g=1}^{G} P_{kg}\ln\left(\frac{1}{P_{kg}}\right)+\sum_{g=1}^{G} P_{kg}\sum_{i\in g}\frac{P_{ki}}{P_{kg}}ln\left(\frac{P_{kg}}{P_{ki}}\right)\right) \quad (A2.4)$$

Equation (A1.5) now becomes:

$$H_{kg} = \left[\sum_{i\in g}\frac{P_{ki}}{P_{kg}}ln\left(\frac{P_{kg}}{P_{ki}}\right)\right] \quad (A2.5)$$

where H_{kg} represents the Variety by gender within groups with $\sum_{k=1}^{2} H_{kg} = H_g$ as in equation (A1.5). Using equation (A2.5) the equation (A2.4) can be rewritten as:

$$VAR = \sum_{k=1}^{2}\left(\sum_{g=1}^{G} P_{kg}\ln\left(\frac{1}{P_{kg}}\right)+\sum_{g=1}^{G} P_{kg}H_{kg}\right) = \sum_{k=1}^{2}\sum_{g=1}^{G} P_{kg}\ln\left(\frac{1}{P_{kg}}\right)+\sum_{k=1}^{2}\sum_{g=1}^{G} P_{kg}H_{kg} \quad (A2.6)$$

Equation (A2.6) is the sum of two components. The second one:

$$RV= \sum_{k=1}^{2}\sum_{g=1}^{G} P_{kg}H_{kg}=\sum_{k=1}^{2} RV_k$$

i.e. the sum of Related Variety by gender, while the first term

$$UV=\sum_{k=1}^{2}\sum_{g=1}^{G} P_{kg}\ln\left(\frac{1}{P_{kg}}\right)=\sum_{k=1}^{2} UV_k \quad (A2.7)$$

i.e the sum of Unrelated Variety by gender.

Equation (A1.8) becomes:

$$VAR= \sum_{k=1}^{2} UV_k + \sum_{k=1}^{2} RV_k \quad (A2.8)$$

Where $\sum_{k=1}^{2} UV_k = UV$ and $\sum_{k=1}^{2} RV_k = RV$.

The equation (A2.8) is equal to equation (A1.8) used by Frenken al., (2007). The novelty is represented by the decomposition in females and males. This new measure allows us to explore if and how gender affects the innovation process.

The UV in equation (A2.7) can be furtherly decomposed. Indicating

$$P_k = \sum_{g\in k} P_{kg} \quad (A2.9)$$

and multiplying equation (A2.7) by $\frac{P_{kg}}{P_{kg}}$ we obtain:

$$UV = \sum_{k=1}^{2}\left[\sum_{g\in k}\frac{P_k}{P_k} * P_{kg}ln\left(\frac{P_k}{P_k}\frac{1}{P_{kg}}\right)\right]= \sum_{k=1}^{2}\left[\sum_{g\in k}\frac{P_k}{P_k} * P_{kg}ln\left(\frac{P_k}{P_{kg}}\right)+\sum_{g\in k}\frac{P_k}{P_k} * P_{kg}ln\left(\frac{1}{P_k}\right)\right]$$

Using equation (A2.1) we obtain:

$$UV = \sum_{k=1}^{2} P_k\sum_{g\in k}\frac{P_{kg}}{P_k}ln\left(\frac{P_k}{P_{kg}}\right)+\sum_{k=1}^{2} P_k ln\left(\frac{1}{P_k}\right) \quad (A2.10)$$

and indicating H_k as:

$$H_k = \sum_{g \in k} \frac{P_{kg}}{P_k} ln\left(\frac{P_k}{P_{kg}}\right)$$

Equation (A2.10) becomes:

$$UV = \sum_{k=1}^{2} P_k H_K + \sum_{k=1}^{2} P_k ln\left(\frac{1}{P_k}\right)$$

where H_k represents the unrelated Variety by gender between industries at a 2-digit level. It varies from 0 to $ln(G)$-$ln(K)$. H_k=0 when $P_k = P_{kg}$ i.e all the employment of a given gender belongs to the same 2-digit industry. When H_k= $ln(G)$-$ln(K)$ genders are equidistributed between 2-digit industries. The first term:

$$UV_{kg} = \sum_{k=1}^{2} P_k H_K$$

represents the weighted unrelated Variety by gender between industries at a 2-digit level while the second term:

$$UV_k = \sum_{k=1}^{2} P_k ln\left(\frac{1}{P_k}\right)$$

represents the diversity by gender between all the economic activities and it captures the disparities between females and males in the whole economy. UV_K varies from 0 to ln(2). UV_k=0 when $\frac{E_k}{E} = 1$ i.e when all employed belong to the same gender. When UV_K= ln(2) females and males are equi-distributed between economic activities. The term UV_k is symmetric to the females' (males') and it reaches its maximum in correspondence of ln(2). The sum of UV_k and UV_{kg} gives the UV as in Frenken at al., (2007). Finally, equation (A2.8) can be rewritten as:

$$VAR = \underbrace{UV_{kg} + UV_k}_{UV} + \underbrace{RV_F + RV_M}_{RV} \quad \text{(A2.11)}$$

Appendix 3: RV changes when the numbers of females increase only in one 5-digit industry.

In accordance with the previous decomposition the RV is given by:

$$\text{RV}= \sum_{k=1}^{2} \sum_{g=1}^{G} P_{kg}\, H_{kg} = \sum_{k=1}^{2} RV_k \text{ (A3.1)}$$

Where H_{kg} is given by equation (A2.5). RV can be rewritten as:

$$RV = \sum_{g=1}^{G} P_{Fg}\, H_{Fg} + \sum_{g=1}^{G} P_{Mg}\, H_{Mg} = \sum_{g=1}^{G} P_{Fg} \left[\sum_{i \in g} \frac{P_{Fi}}{P_{Fg}} ln\left(\frac{P_{Fg}}{P_{Fi}}\right)\right] + \sum_{g=1}^{G} P_{Mg} \left[\sum_{i \in g} \frac{P_{Mi}}{P_{Mg}} ln\left(\frac{P_{Mg}}{P_{Mi}}\right)\right] \quad \text{(A3.2)}$$

Where:

$$P_{Fi} = \frac{E_{Fi}}{E};\ P_{Fg} = \frac{E_{Fg}}{E};\ P_{Mi} = \frac{E_{Mi}}{E};\ P_{Mg} = \frac{E_{Mg}}{E};$$

Equation (A3.2) can be rewritten as:

$$\text{RV} = \underbrace{\sum_{g=1}^{G} \frac{E_{Fg}}{E}\left[\sum_{i \in g} \frac{E_{Fi}}{E_{Fg}} ln\left(\frac{E_{Fg}}{E_{Fi}}\right)\right]}_{RV_F} + \underbrace{\sum_{g=1}^{G} \frac{E_{Mg}}{E}\left[\sum_{i \in g} \frac{E_{Mi}}{E_{Mg}} ln\left(\frac{E_{Mg}}{E_{Mi}}\right)\right]}_{RV_M} \quad \text{(A3.3)}$$

Remembering that: $E = E_F + E_M$; $E_F = \sum_{i=1}^{N} E_{Fi}$; $E_{Fg} = \sum_{i\in g} EF_i$ equation (A3.3) can be rewritten as:

$$RV_F = \frac{1}{(E_M+E_F)}\left\{\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Fi} ln\left(\frac{E_{Fg}}{E_{Fi}}\right)\right]\right\} \quad \text{(A3.4.1)}$$

$$RV_M = \frac{1}{(E_M+E_F)}\left\{\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Mi} ln\left(\frac{E_{Mg}}{E_{Mi}}\right)\right]\right\} \quad \text{(A3.4.2)}$$

To investigate the RV changes when the females increase in only one 5-digit industry the RV derivative with respect to the females' change needs to be calculated. We suppose that females' employment increase while males' employment remains unchanged. Consequently, the term $\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Mi} ln\left(\frac{\sum_{i\in g} E_{Mi}}{E_{Mi}}\right)\right] = cost$. Our aim is to investigate the changes in RV_F when the females' employment changes only in one of the i industries belonging to g with g=1. Indicating with x the females' employment in industry *i=1* (E_{F1}), equations (A3.4.1) (A3.4.2) can be rewritten as:

$$RV_F = \frac{1}{x+\sum_{i=2}^{N} E_{Fi}+E_M}\left\{\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Fi} ln\left(\frac{\sum_{i\in g} E_{Fi}}{E_{Fi}}\right)\right]\right\} \quad \text{(A3.5.1)}$$

$$RV_M = \frac{cost}{x+\sum_{i=2}^{N} E_{Fi}+E_M} \quad \text{(A3.5.2)}$$

The terms $\sum_{i=2}^{N} E_{Fi} + E_M$ in equation (4.4.1) are constant with respect to the change of E_{F1}.

$$c_1 = \sum_{i=2}^{N} E_{Fi} + E_M$$

$$RV_F = \frac{1}{x+c_1}\left\{\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Fi} ln\left(\frac{\sum_{i\in g} EF_i}{E_{Fi}}\right)\right]\right\} \quad \text{(A3.5.1 bis)}$$

By hypothesis industry i=1 can fall exclusively in one g. Suppose that i=1∈g=1. The content of the square brackets in equation A.1 can be rewritten as:

$$\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Fi} ln\left(\frac{\sum_{i\in g} E_{Fi}}{E_{Fi}}\right)\right]=\sum_{g=2}^{G}\left[\sum_{i\in g;g\neq 1} E_{Fi} ln\left(\frac{\sum_{i\in g;g\neq 1} E_{Fi}}{E_{Fi}}\right)\right]+\left[\sum_{i\in g;g=1} E_{Fi} ln\left(\frac{\sum_{i\in g;g=1} E_{Fi}}{E_{Fi}}\right)\right]$$

Moreover, term $\sum_{g=2}^{G}\left[\sum_{i\in g;g\neq 1} E_{Fi} ln\left(\frac{E_{Fg}}{E_{Fi}}\right)\right]$ is constant with respect to E_{F1}, so:

$$c_2 = \sum_{g=2}^{G}\left[\sum_{i\in g;g\neq 1} E_{Fi} ln\left(\frac{\sum_{i\in g} E_{Fi}}{E_{Fi}}\right)\right]$$

$$RV_F = \frac{1}{x+c_1}\left\{c_2 + \left[\sum_{i\in g;g=1} E_{Fi} ln\left(\frac{\sum_{i\in g;g=1} E_{Fi}}{E_{Fi}}\right)\right]\right\}$$

The term can be rewritten as:

$$\left[\sum_{i\in g;g=1} E_{Fi} ln\left(\frac{\sum_{i\in g;g=1} E_{Fi}}{E_{Fi}}\right)\right]=$$
$$x * ln\left(\frac{x+\sum_{i\neq 1\in g;g=1} E_{Fi}}{x}\right) + \sum_{i\neq 1\in g;g=1} E_{Fi} * ln\left(\frac{x+\sum_{i\neq 1\in g;g=1} E_{Fi}}{E_{Fi}}\right)$$

Where $\sum_{i\neq 1\in g;g=1} E_{Fi}$ is constant with respect to x:

$$c_3 = \sum_{i \neq 1 \in g; g=1} E_{Fi}$$

Consequently:

$$RV_F = \frac{1}{x + c_1}\left\{c_2 + \left[x * ln\left(\frac{x + c_3}{x}\right) + c_3 ln\left(\frac{x + c_3}{E_{Fi}}\right)\right]\right\}$$

Applying the log properties to the term:

$$c_3 ln\left(\frac{x + c_3}{E_{Fi}}\right)$$

We obtain:

$$c_3[ln(x + c_3) - ln(E_{Fi})] = c_3 ln(x + c_3) - c_3\ ln(E_{Fi})$$

Moreover, the term $c_3\ ln(E_{Fi})$ is constant with respect to x. We indicate it as $c_4 = c_3\ ln(E_{Fi})$;

Finally,

$$c_3 ln\left(\frac{x+c_3}{E_{Fi}}\right) = c_3 ln(x + c_3) - c_4$$

RV_F is equal to:

$$RV_F = \frac{1}{x + c_1}\left\{c_2 + \left[x * ln\left(\frac{x + c_3}{x}\right) + c_3 ln(x + c_3) - c_4\right]\right\}$$

$$RV_F = \frac{1}{x+\sum_{i=2}^{N} E_{Fi}+E_M}\left\{\sum_{g=2}^{G}\left[\sum_{i\in g; g\neq 1} E_{Fi} ln\left(\frac{\sum_{i\in g} E_{Fi}}{E_{Fi}}\right)\right] + \left[x * ln\left(\frac{x+\sum_{i\neq 1\in g; g=1} E_{Fi}}{x}\right) + \left[ln\left(x + \sum_{i\neq 1\in g; g=1} E_{Fi}\right)\right]\sum_{i\neq 1\in g; g=1} E_{Fi} - \left(\sum_{i\neq 1\in g; g=1} E_{Fi}\right) ln(E_{Fi})\right]\right\} \quad \text{(A3.6)}$$

To study how RV changes when EF_1 changes, the first derivative needs to be investigated. RV_M derivatives respect to EF_1 is:

$$\frac{\partial RV_M}{\partial E_{F1}} = -\frac{\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Mi} ln\left(\frac{\sum_{i\in g} E_{Mi}}{E_{Mi}}\right)\right]}{(x + \sum_{i=2}^{N} E_{Fi} + E_M)^2}$$

$$\lim_{x\to\infty} \frac{\partial RV_M}{\partial E_{F1}} = 0$$

$$\text{for x=0} \quad RV_M = -\frac{\sum_{g=1}^{G}\left[\sum_{i\in g} E_{Mi} ln\left(\frac{\sum_{i\in g} E_{Mi}}{E_{Mi}}\right)\right]}{\left(\sum_{i=2}^{N} E_{Fi}+E_M\right)^2}$$

RV_F derivatives respect to EF_1 is:

$$\frac{\partial RV_F}{\partial E_{F1}} = \frac{ln\left(\frac{x+\sum_{i\neq 1\in g; g=1} E_{Fi}}{x}\right)(\sum_{i=2}^{N} E_{Fi}+E_M) - ln\left(x+\sum_{i=2}^{N} E_{Fi}+E_M\right)(\sum_{i=2}^{N} E_{Fi}+E_M) - \sum_{g=2}^{G}\left[\sum_{i\in g; g\neq 1} E_{Fi} ln\left(\frac{\sum_{i\in g} E_{Fi}}{E_{Fi}}\right)\right] + \sum_{i\neq 1\in g; g=1} E_{Fi}\ ln(E_{Fi})}{\left(x+\sum_{i=2}^{N} E_{Fi}+E_M\right)^2}$$

$$\lim_{x \to \infty} \frac{\partial RV_F}{\partial E_{F1}} = 0$$

$$\lim_{x \to 0} \frac{\partial RV_F}{\partial E_{F1}} = \infty$$

$$\frac{\partial RV_F}{\partial EF_1} = 0 \text{ for x=}x^*.$$

For $0 < x < x^*$ $\frac{\partial RV_F}{\partial EF_1}$ increases while for $x^* < x < \infty$ $\frac{\partial RV_F}{\partial EF_1}$ decreases. x^* represent the point in which females and males are equidistribuited between gender.

Appendix 4: UV_{kg} changes when the numbers of females increase in only one 2-digit industry

In accordance with the previous results UV_{kg} is given by:

$$UV_{kg} = \sum_{k=1}^{2} P_k H_K$$

$$H_k = \sum_{g=1}^{G} \frac{P_{kg}}{P_k} ln\left(\frac{P_k}{P_{kg}}\right)$$

$$UV_{kg} = \underbrace{P_F * \sum_{g=1}^{G} \frac{P_{Fg}}{P_F} ln\left(\frac{P_F}{P_{Fg}}\right)}_{UVF_g} + \underbrace{P_M * \sum_{g=1}^{G} \frac{P_{Mg}}{P_M} ln\left(\frac{P_M}{P_{Mg}}\right)}_{UV_M}$$

Where $P_F = \frac{E_F}{E}$; E=E_F+E_M; $E_F = \sum_{g=1}^{G} E_{Fg}$. UV_{kg} can be rewritten as:

$$UV_{kg} = \underbrace{\frac{E_F}{\sum_{g=1}^{G} E_{Fg} + E_M} * \frac{1}{E_F} \sum_{g=1}^{G} E_{Fg} ln\left(\frac{E_F}{E_{Fg}}\right)}_{UV_{Fg}} + \underbrace{\frac{E_M}{\sum_{g=1}^{G} E_{Fg} + E_M} * \frac{1}{E_M} \sum_{g=1}^{G} E_{Mg} ln\left(\frac{E_M}{E_{Mg}}\right)}_{UV_{Mg}} \quad \text{(A4.1)}$$

The term $\sum_{g=1}^{G} E_{Mg} ln\left(\frac{E_M}{E_{Mg}}\right)$ is constant with respect to E_{F1}, consequently,

$c_1 = \sum_{g=1}^{G} E_{Mg} ln\left(\frac{E_M}{E_{Mg}}\right)$, furthermore, the term $\sum_{g=1}^{G} E_{Fg}$ can be rewritten as $E_{F1} + \sum_{g=2}^{G} E_{Fg}$. Equation (A4.1) becomes:

$$UV_{kg} = \underbrace{\frac{E_{F1} * ln\left(\frac{E_F}{E_{F1}}\right) + \sum_{g=2}^{G} E_{Fg} ln\left(\frac{E_F}{E_{Fg}}\right)}{E_{F1} + \sum_{g=2}^{G} E_{Fg} + E_M}}_{UV_{Fg}} + \underbrace{\frac{c_1}{E_{F1} + \sum_{g=2}^{G} E_{Fg} + E_M}}_{UV_{Mg}} \quad \text{(A4.2)}$$

The terms $\sum_{g=2}^{G} E_{Fg} ln\left(\frac{E_F}{E_{Fg}}\right)$ and $\sum_{g=2}^{G} E_{Fg} + E_M$ are constant with respect to E_{F1}.

$c_2 = \sum_{g=2}^{G} E_{Fg} + E_M$; $c_3 = \sum_{g=2}^{G} E_{Fg} ln\left(\frac{E_F}{E_{Fg}}\right)$

Equation (A4.2) becomes:

$$UV_{kg} = \underbrace{\frac{E_{F1} * ln\left(\frac{E_F}{E_{F1}}\right) + c_3}{E_{F1} + c_2}}_{UV_{Fg}} + \underbrace{\frac{c_1}{E_{F1} + c_2}}_{UV_{Mg}} \quad \text{(A4.3)}$$

$$\frac{\partial UV_{Mg}}{\partial EF_1} = \frac{-c_1}{(c_2 + E_{F1})^2}$$

$$\lim_{x\to\infty} \frac{\partial UV_{Mg}}{\partial EF_1} = 0$$

$$\frac{\partial UV_{Mg}}{\partial EF_1} = \frac{-c_1}{(c_2)^2} \text{ when } EF_1 = 0$$

$$\frac{\partial UV_{Fg}}{\partial EF_g} = \frac{-c_3 + c_2 * ln\left(\frac{E_F}{E_{F1}}\right) - E_{F1} - c_2}{(E_{F1} - c_2)^2}$$

$$\lim_{x\to\infty} \frac{\partial UV_{Fg}}{\partial EF_1} = 0$$

$$\lim_{x\to 0} \frac{\partial UV_{Fg}}{\partial EF_1} = \infty$$

Finally, $\frac{\partial UV_{Fg}}{\partial EF_1} = 0$ for x=x^*.

Appendix 5: UVF changes when the numbers of females increase in the whole economy

$$UV_k = \sum_{k=1}^{2} P_k ln\left(\frac{1}{P_k}\right)$$

$$UV_k = \frac{E_M}{E} ln\left(\frac{E}{E_M}\right) + \frac{E_F}{E} ln\left(\frac{E}{E_F}\right)$$

$$E = E_M + E_F$$

$$UV_k = \frac{E_M}{E_M + E_F} ln\left(\frac{E_M + E_F}{E_M}\right) + \frac{E_F}{E_M + E_F} ln\left(\frac{E_M + E_F}{E_F}\right)$$

$E_M = c; E_F = x$

$$UV_k = \frac{c}{c + x} ln\left(\frac{c + x}{c}\right) + \frac{x}{c + x} ln\left(\frac{c + x}{x}\right)$$

$$\frac{\partial UV_k}{\partial x} = -\frac{c\left[ln\left(\frac{c + x}{c}\right) - ln\left(\frac{c + x}{x}\right)\right]}{(c + x)^2}$$

$$\lim_{x\to\infty} \frac{\partial UV_F}{\partial EF} = 0$$

$$\lim_{x \to 0} \frac{\partial UV_F}{\partial EF} = \infty$$

Finally, $\frac{\partial UV_F}{\partial EF} = 0$ when EF=EM i.e when females and males are equi-distributed in the whole economy.