

The Duration-Based Measurement of Unemployment

Estimation Issues and an Application to Male-Female
Unemployment Differences in France

Stephen Bazen, Aix-Marseille Université, Greqam
Xavier Joutard, Aix-Marseille Université, Greqam
Mouhamadou M. Niang, Aix-Marseille Université, Greqam

WP 2012 - Nr 06

The duration-based measurement of unemployment: estimation issues and an application to male-female unemployment differences in France

Stephen Bazen

Aix-Marseille Université, AMSE, GREQAM,
stephen.bazen@univ-amu.fr
Centre de la Vieille Charité, 13002 Marseille,

Xavier Joutard,

Aix-Marseille Université, AMSE, GREQAM,
xavier.joutard@univ-amu.fr
Centre de la Vieille Charité, 13002 Marseille
and

Mouhamadou M. Niang *

Aix-Marseille Université, AMSE, GREQAM
mouhamadou.NIANG@univ-amu.fr
Centre de la Vieille Charité, 13002 Marseille

March 2012

*We are grateful to Benoît Tarroux and participants at the Journées Louis-André Gerard-Varet (LAGV), Marseille in 2010 for helpful comments. We also thank Carlo G. Camarda for his programme for modelling digit preference bias.

Abstract

This paper examines how unemployment can be measured in normative fashion - taking into account the mean and inequality of spell lengths - and how the extent of unemployment can be estimated from cross section data of the type found in labour force surveys. The issue is not straightforward since in these surveys completed durations of unemployed individuals are not observed yet they constitute the basis for calculating the kind of index that has been proposed to measure the extent of unemployment in a way that goes beyond the unemployment rate. The index proposed by Shorrocks has robust normative foundations and has an equivalent representation in terms of average complete duration and the density of completed durations. Building upon earlier work applied in the United States for estimating the first of these, we present a method that enables the index to be calculated based on an estimate of the density of completed durations. The approach is illustrated in the context of comparing male-female unemployment differences in France, where historically female unemployment has been higher than that of males.

JLE Classification: J64, C41, D63

Keywords: unemployment measurement, completed duration, synthetic cohort, gender gap.

1 Introduction

Unemployment is usually referred to in terms of the number of persons concerned or the proportion of the active labour force without work at a given time. These ‘headline’ figures are clearly relevant indicators of the extent of unemployment but they do not take into account an important dimension of unemployment experience. Labour economists have long emphasized the duration of a spell of unemployment and much effort has gone into understanding why certain individuals find it difficult to leave unemployment and thus experience longer spells than others. The duration of unemployment is regarded as an important component since it is associated with social and psychological problems and in economic terms, due to the degradation of an individual’s human capital and capacity to hold down a durable job, unemployment could become self-perpetuating.

In the early 1990s, a number of authors developed a normative approach to measuring unemployment in ways similar to those used in examining the extent of inequality or poverty (Sengupta (2008); Shorrocks (2008b,a); Paul (1992))¹. Essentially a longer duration is seen as aggravating the extent of unemployment in welfare terms, other things being equal. Furthermore, if unemployment is unequally distributed with a higher proportion of longer than shorter durations, then from society’s point of view, unemployment is more severe. The key variable that emerges in this literature is thus the duration of a spell of unemployment, and this dimension is not captured by the unemployment rate. This constitutes a normative measure since it amounts to giving as much weight to individuals with very different durations. Naturally, this approach could be extended to a longer time scale in which recurrent unemployment spells - individuals moving in and out of unemployment - could also be taken into account (Shorrocks (2008b); Disney (1979); Sengupta (2008)). In this case, instead of concentrating on spell length, the welfare evaluation of unemployment would be based on time spent in unemployment in a given period - say five years.

However, in order for these aspects of unemployment to be taken into account in an aggregate index or in some form of dominance analysis, it is necessary to have a measure of the length of a spell of unemployment. For a given population, this would normally be possible after a number of years since individuals will not remain in the same labour market state indefinitely. Most of those experiencing a spell of unemployment will exit from that state either into employment or leave the labour force altogether. However, the most readily available data on the unemployed come from periodic - these days, mainly

¹Interestingly two of these papers remained unpublished until recently being revived in 2008 in the Journal of Economic Inequality, Volume 10. Lambert (2008) speculates on why this may have occurred

quarterly - labour force surveys which record the length of ongoing or uncompleted spells at a given point in time. If a welfare-based measure of unemployment based on duration is to be made operational, the passage between the observed uncompleted duration and the definitive spell length will be need to be established. This is an issue that has featured heavily in the econometric and statistical literature on survival analysis (for example, [Nickell \(1979\)](#) and [Baker and Trivedi \(1985\)](#)). In a given cross section survey, almost all durations will be right-censored² and it is not possible to calculate a final, completed duration using these data without making a number of fairly strong assumptions.

The aim of the current paper is to make a welfare-based measure of unemployment operational. It takes as a starting point an aggregate index proposed by [Shorrocks \(2008a\)](#), which determines the extent of unemployment in a given population at a given point in time as the average of a polynomial function of individual unemployment durations. The exponent on the duration term determines the degree of unemployment aversion. This index has a number of appealing properties from a social welfare point of view: it can be constructed from a number of coherent axioms, it is additively decomposable and it has limited information requirements. It also has a very straightforward interpretation as the degree of unemployment aversion increases since it can be expressed as a product of the unemployment rate, a polynomial of the average duration and a measure of the inequality of spell lengths.

The Shorrocks index is used to assess the extent of unemployment in France using data from the Labour Force Survey. France has had a persistently high rate of unemployment since the 1980's, varying between 7 percent and 11 percent. Using the Shorrocks index for 2003 to 2008, it is found that while the underlying rate is stable until mid 2006, before falling significantly, measures based on estimated completed durations show a clear downward trend from the first quarter of 2004. Furthermore, unlike most OECD countries, female unemployment is higher than male unemployment, in terms of both the rate and the number concerned. In a recent paper, [Azmat, Güell, and Manning \(2006\)](#) were unable to identify why this may be the case. In the current paper we use the Shorrocks index to analyse the size of the gender gap in unemployment and two clear conclusions emerge. First, the source of the gap is the difference in the rate of unemployment rather than the distribution of spells. Secondly, based on a formal test, the gender gap is found to disappear after 2006:4.

The paper is organised as follows. First we present the Shorrocks index and other approaches. In the second section we address the issue of how to calculate the index using

²There may be individuals who at the time of interview are on the verge of completing a spell e.g. about to begin a job.

data on interrupted or right censored unemployment spells. In the next section, the extent and gender composition of unemployment in France are examined using the index for the period 2003:1 to 2008:3. In the final section, we point out some avenues for further research into the determinants of the gender gap in unemployment in France.

2 Welfare-based measurement of unemployment - aggregation

There have been a number of attempts to treat the measurement of unemployment from a social welfare point of view, in a similar fashion to the measurement of inequality and poverty. By setting out a number of axioms that are thought to command widespread support, it is possible to derive a measure that will reflect explicitly stated social welfare judgments when gauging the extent of inequality (Atkinson (1970)) and poverty (Sen (1976); Foster, Greer, and Thorbecke (1984)). The same kind of approach has been applied to the measurement of unemployment. The chronology of the development of such measures is complicated since three of the key contributions were developed in the early 1990s but only published in 2008³ (Sengupta (2008); Shorrocks (2008b); and Shorrocks (2008a)).

The first published paper using a rigorous welfare economic approach is Paul (1992) who proposes an aggregate measure based the concept of 'illfare'. Illfare increases with the length of a spell of unemployment (d). For a labour force of size n , of which n_u are unemployed, and for which the unemployment rate is $U^R = \frac{n_u}{n}$, Paul's measure of unemployment is given by:

$$U_P(\alpha) = \frac{U^R \times \bar{d}}{1 + z(\alpha)} \quad (2.1)$$

where $z(\alpha) = (\frac{1}{n_u} \sum_{i=1}^{n_u} d_i^\alpha)^\frac{1}{\alpha}$, $\alpha > 1$ and $\bar{d} = \frac{1}{n_u} \sum_{i=1}^{n_u} d_i$.

A second published paper by Riese and Brunner (1998) is a purely theoretical treatment in which the measurement of the severity of unemployment is associated with the establishment of dominance relations between distributions of unemployment spells. They do not provide an index measure but instead argue that in a stochastic dominance framework, the disutility of unemployment between groups, countries or points in time, can

³Sengupta (2008) proposes an index that applies to total unemployment experience in a given time window. It is not a measure of the extent of unemployment at a given point in time.

be ranked according 'severity curves'. These relate a multiple of the unemployment rate to the length of a spell and non-intersecting severity curves mean that an unambiguous ranking can be established. One important assumption made in this analysis is that unemployment is in a steady-state so that inflows equal outflows.

A third published work by [Borooah \(2002\)](#) proposes an equally-distributed equivalent measure of the distribution of unemployment spells in the same vein as [Atkinson \(1970\)](#) for income inequality measurement. Interestingly, in terms of the historical development of welfare based measures of unemployment, Borooah cites none of the references mentioned hitherto. His measure is basically a correction of the observed unemployment rate which takes into account the degree of inequality in unemployment spells and is expressed as follows:

$$U_B(\alpha) = U^R(1 + A(\alpha)) \quad (2.2)$$

$$\text{where } A(\alpha) = \left\{ \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{d_i}{d} \right)^{1+\alpha} \right]^{\frac{1}{1+\alpha}} \right\} - 1$$

Note that all n members of the labour force are included in the calculation of the inequality index which is accommodated by setting $d_i = 0$ for employed individuals. If all durations are equal then there is no inequality and so $A(\alpha) = 0$, and the actual unemployment rate adequately measures the social loss from unemployment. The higher the value of the unemployment aversion parameter, α , the greater the social loss from a given distribution of unemployment spells. If $\alpha = 0$ then society is indifferent to the distribution of unemployment spells and $A(\alpha) = 0$.

[Shorrocks \(2008a\)](#) proceeds by setting out a number of axioms concerning unemployment experience and proposes an index for measuring unemployment which is reminiscent of the poverty index proposed by [Foster, Greer, and Thorbecke \(1984\)](#), and is expressed in terms of the length of unemployment spells. Using the definition of a spell of unemployment for the employed as $d_i = 0$, the index is defined as follows:

$$U_s(\alpha) = \frac{1}{n} \sum_{i=1}^n d_i^\alpha \quad (2.3)$$

Shorrocks derives this measure using a normative approach based on axioms that have found wide command in the literature on measuring poverty. If we consider that information on unemployment is captured by a vector $d = (d_1, \dots, d_n)$ of unemployment durations in the labour force of size n , an unemployment index, according to Shorrocks, should satisfy the following six axioms:

(A1) Symmetry: $U_\alpha(d') = U_\alpha(d)$, whenever d' is obtained from d by a permutation. This axiom is the equivalent of the anonymity axiom in income inequality measurement and states that no significance is placed on the characteristics of the person associated with some particular spell length.

(A2) Replication invariance: $U_\alpha(d') = U_\alpha(d)$, whenever d' is obtained from d by a replication. This axiom is the conventional way to enable comparison between populations of different size.

(A3) Monotonicity: $U_\alpha(d') < U_\alpha(d)$, whenever d' is obtained from d by a spell reduction. This axiom states that unemployment should increase if an unemployed person experiences a longer spell of unemployment or if a previously employed person becomes unemployed. It can be formalized by saying d' is obtained from d by a spell reduction if $n(d') = n(d)$ and for some transformation $T : d \rightarrow d'$ satisfying: $0 < T_i(d) < d_i$ for some i and $T_j(d) = d_j$ for $j \neq i$.

(A4) Preference for duration equality: $U_\alpha(d') < U_\alpha(d)$, whenever d' is obtained from d by a spell equalization. This axiom is analogous to the Pigou-Dalton ‘principle of transfers’ used in inequality and welfare analysis of income distribution and is closely related to the strict concavity assumption in second order stochastic dominance. It states that the unemployment indices should favour any trend towards a more equal distribution of unemployment durations. This requirement encompasses preference for short spells and may be formalized by saying d' is obtained from d by spell equalization if $n(d') = n(d)$ and there exist i and j such that:

$$d_i > d'_i > d_j; d'_i + d'_j = d_i + d_j; d'_k = d_k \forall k \neq i, j$$

(A5) Normalization: $U_\alpha(d) = 0$ whenever $d_i = 0 \forall i$. This axiom states that unemployment indices should be equal to zero when no one is unemployed.

(A6) Homogeneity with respect to unemployment rate: $U_\alpha(d') = rU_\alpha(d)$ whenever d' is obtained from d by an r -replication of the unemployed. For a formal definition, let us consider that the duration distribution is split as follows $d = (d+, d0)$ where $d+$ stands for the durations of the unemployed and $d0 = (0, 0, \dots, 0)$ is the duration vector of the employed. This is allowed by the anonymity assumption **(A1)**. We say that d' is obtained from d by an r -replication of the unemployed if and only if one can write d' as follows $d' = (d+, d+, \dots, d+, d0)$ such that the size of the labour force is unchanged. So d' satisfies $n^+(d') = rn^+(d)$ and $n(d') = n(d)$.

Shorrocks also shows that for any unemployment index that belongs to the family of indices satisfying the above axioms, the ranking of unemployment level we obtained is the same as that obtained in a stochastic dominance framework.

The measure has a number of interesting properties. Firstly it is decomposable so

that the aggregate value of the unemployment measure is equal to the weighted sum of unemployment in any number of groups:

$$U_s(\alpha) = \sum_{j=1}^J \frac{n_j}{n} U_s^j \quad (2.4)$$

where $U_s^j = \frac{1}{n_j} \sum_{i \in J} d_i^\alpha$

Secondly, for each value of the unemployment aversion parameter, α , the index takes on a specific form. Thus :

- $\alpha = 0 : U_s = \frac{n_u}{n} = U^R$
- $\alpha = 1 : U_s = \frac{n_u}{n} \bar{d} = U^R \bar{d}$ where $\bar{d} = \frac{1}{n_u} \sum_{i=1}^{n_u} d_i$
- $\alpha = 2 : U_s = U^R \bar{d}^2 (1 + E)$ where $E = \frac{1}{n_u} \sum_{i=1}^{n_u} \left\{ \left(\frac{d_i}{\bar{d}} \right)^2 - 1 \right\}$ is defined for $d_i > 0$ and is known as the generalized entropy inequality index. This is equivalent to the Herfindahl-Hirschman index (HHI) divided by N , and the square of the coefficient of variation.

These equivalent representations will be particularly useful when making the measure operational. Furthermore, it is clear that the measure can be viewed as a generalization of the standard measure of unemployment since the unemployment rate is a special case of the index. As the aversion parameter increases, greater weight is placed on longer durations and thus on the inequality of unemployment experience.

3 The definition of a spell of unemployment - the identification issue

These different welfare-based approaches to the measurement of unemployment are all defined in terms of the length of a spell of unemployment experienced by each unemployed individual. If measurement is being undertaken for historical purposes, then there is no ambiguity concerning the definition of this spell. An individual who experiences a period of unemployment will not remain in that labour market state forever and using retrospective data, a completed duration will be observed for all of the unemployed. However, the usual reference to unemployment is the extent of unemployment in the previous month, quarter or year. National labour force surveys now enable the calculation of the

rate of unemployment on the basis of the internationally recognized definition in terms of the concept of a person without a job and actively seeking work on the basis of a number of search criteria and in terms of the availability of the individual to start in the imminent future. However, in these surveys the completed duration is rarely recorded. What is observed is the interrupted or right-censored duration. This poses questions for the use of the indices developed on the basis of the length of a spell of unemployment.

The authors cited above naturally address this identification issue but settle on the uncompleted or censored duration. However a cross section of uncompleted durations does not correspond to the welfare foundations of the measures proposed. Firstly, it is not representative picture of actual unemployment experience since for an inflow at a given date, those with shorter durations will have already left the sample⁴. Furthermore, of those remaining, the observed duration is not an accurate measure of the true (completed) duration⁵. However, in order to calculate the Shorrocks index (for $\alpha > 0$) it is not necessary to have precise information on each individual's completed spell length. The equivalent expressions of the Shorrocks index when $\alpha = 1$ and $\alpha = 2$ can be calculated using data on incomplete durations in order to obtain an estimate of the density (or related) function of completed durations.

There have been several attempts to estimate the *average completed duration* from data on censored unemployment spells. Completed durations can be determined either directly or indirectly. Often this is obtained for a given inflow or cohort - the group of persons entering unemployment at a same given date. In terms of the calculation of the extent of unemployment at a given point in time, however, the unemployed population will comprise persons having entered unemployment at *different* dates. The stock of unemployed will contain a mix of different cohorts. In order to take this into account, an important distinction is made between stationary and non-stationary distributions of unemployment spells. The former is where the unemployed stock is constant over time. This entails the same constant number entering unemployment in each month (say) and exactly the same number exiting. This in turn means that the distribution of completed spells is the same for each or cohort that enters unemployment.

If stationarity is assumed, there are a number of easy and straightforward ways of calculating the average completed spell length and other aspects of the extent of unemployment from a sample of unemployed persons with incomplete spells. The hypothesis

⁴This is referred to as 'length bias' in [Salant \(1977\)](#) and causes the mean of interrupted durations to be higher than the mean of completed durations

⁵[Salant \(1977\)](#) call this the 'interruption bias' and causes the opposite effect a lower mean interrupted duration.

of a stationary distribution of unemployment gives rise to the following features. First, the number of persons in the stock of unemployed at any time is equal to the inflow multiplied by the average completed duration (see [Baker and Trivedi \(1985\)](#), for a demonstration). The latter is obtained by dividing through by the inflow into unemployment. Second, for a given unemployed person, the expected complete duration is equal to twice the length of current uncompleted spell. Thirdly, the density of completed spells can be obtained directly from the density of uncompleted durations. [Salant \(1977\)](#) shows that if escape rates from unemployment are drawn from a gamma distribution, the implied density of uncompleted spells is:

$$g(t) = (r-1)a^{r-1}(a+t)^{-r} \quad (3.1)$$

and the corresponding density of completed durations is given by :

$$f(d) = ra^r(a+d)^{r-1} \quad (3.2)$$

The mean spell duration can be straightforwardly estimated. The values of the parameters a and r can be estimated using maximum likelihood. [Nickell \(1979\)](#) shows how a parametric hazard function can be estimated from censored durations using the likelihood approach.

The stationary case is useful as a reference situation, but this assumption will need to be dispensed with when measuring the extent of unemployment at any given point in time. In the 1980s, studies of the US experience developed an approach based on the notion of a *synthetic cohort* in order to circumvent the absence of uncensored, completed spells (see [Baker \(1992\)](#); [Sider \(1985\)](#)) without having recourse to the stationarity assumption. The basic idea is that the sample of persons who currently have been unemployed for s months are in the same cohort as those who are unemployed for $s+1$ months in the following month. Given that there will be fewer unemployed persons after one more month, using two independent consecutive cross sections enables one to estimate the survival rate for the group in question for that month. These observations can then be used to estimate the average completed duration for the currently unemployed and obtain a non parametric estimate of the survivor function. Recently [Guell and Hu \(2006\)](#) have used the same sampling scheme to estimate parametric models of the hazard rate that allow time-varying covariates and unobserved heterogeneity to be taken into account. Importantly, this method does not require the steady state or stationary assumption to hold.

In the context of the Shorrocks index, the information requirements in the absence of the observed completed duration for each individual are: n_u the number of unemployed, the number of employed $n - n_u$, the average completed duration $\bar{d}$ and the density of completed durations $f(d)$. The latter two can be estimated non-parametrically using the synthetic cohort approach. The method employed does not require the stationarity assumption, but is nevertheless based on the hypothesis that current economic conditions prevail into the future.

The method is based on rewriting the survivor function for an uncompleted spell t at the survey date. First, we decompose the time interval from 0 to t in the following way: from 0 to t_1 of length a_1 , from t_1 to t_2 of length $a_2, \dots$, from t_{k-1} to t . Each sub-interval is of length a_k . These intervals need not necessarily be equally spaced, even if with the standard surveys the length of sub-interval is a quarter. Next we apply the conditional decomposition.

$$\begin{aligned} S_t &= \Pr(D > t) \\ &= \Pr(D > t | D > t_{k-1}) \Pr(D > t_{k-1}) \\ &= \Pr(D > t | D > t_{k-1}) \Pr(D > t_{k-1} | D > t_{k-2}) \Pr(D > t_{k-2}) \end{aligned} \tag{3.3}$$

where D is the random completed duration variable.

This, finally yields:

$$\begin{aligned} S_t &= \prod_{j=1}^k S_{t_j | t_{j-1}} \\ &= \prod_{j=1}^k S_{t_j | t_j - a_{j-1}} \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} S_{t_j | t_{j-1}} &\equiv \Pr(D > t_j | D > t_{j-1}) \\ &= \Pr(D > t_j | D > t_j - a_{j-1}) \\ &\equiv S_{t_j | t_j - a_{j-1}} \end{aligned}$$

corresponds to the conditional survival on t_j , i.e. remaining unemployed for at least t_j months given having been unemployed for $t_j - a_{j-1}$ months. For purpose of simplification, we set $t_1 = a_1$ and $t_k = t$.

Using two surveys at dates $\tau - a$ and τ , an estimate of the (conditional) survival rate on t between $\tau - a$ and τ can be obtained by using the following assumption: with

two representative surveys, the unemployed individuals for t months at date τ should be drawn from the same population as those unemployed individuals for $t - a$ months at time $\tau - a$.

By defining the following numbers :

$N(t, \tau)$ which counts the number of unemployed individuals for t months at date τ

$N(t - a, \tau - a)$ which counts the number of unemployed individuals for $t - a$ months at date $\tau - a$

we deduce an estimate of the conditional survival rate:

$$\begin{aligned}\widehat{S}_{t|t-a} &= \frac{N(t, \tau)}{N(t - a, \tau - a)} \\ &\equiv r_t\end{aligned}\tag{3.5}$$

where $0 < r_t < 1$

By using these ratios for each time interval, we can deduce the empirical survivor function for t :

$$\widehat{S}_t = \prod_{j=1}^k r_{t_j}\tag{3.6}$$

The expected completed duration can also be calculated by applying a discrete version of the theoretical result ($E(D) = \int_0^\infty S_t dt$) and by retaining the maximum of the observed uncompleted duration, t_K with a time decomposition of K intervals:

$$\bar{d} = 1 + a_1 \widehat{S}_{t_1} + a_2 \widehat{S}_{t_2} + \dots + a_{K-1} \widehat{S}_{t_{K-1}} + a_K \widehat{S}_{t_K}\tag{3.7}$$

where a_k is the length of the k th duration interval (number of months) $[t_k, t_{k-1}]$.

This is sufficient for estimating the Shorrocks index for $\alpha = 1$. Using the same data, the value of the index for $\alpha = 2$ can be calculated by using the density estimate for t :

$$\widehat{f}(t) = \widehat{S}_{t-1} - \widehat{S}_t\tag{3.8}$$

The index is then calculated as across duration intervals:

$$\frac{1}{n} \sum_{i=1}^n d_i^2 = U^R \bar{d}^2 (1 + \widehat{E})\tag{3.9}$$

where $\widehat{E} = \sum_{k=1}^K \widehat{f}(k) \left[\left(\frac{d_k}{\bar{d}} \right)^2 - 1 \right]$

In the current paper, the focus is on obtaining a value of the Shorrocks index of the extent of unemployment at any given point in time, using quarterly labour force surveys. Data are currently used to obtain standardised internationally comparable *unemployment rates* on the basis of the ILO definition of an unemployed person. Given that to qualify as the latter requires certain responses to a series of questions, the labour force survey is the appropriate data source for determining the rate of unemployment. However, the data on *unemployment durations* contained therein are not of the same quality. They refer to incomplete and therefore right-censored spells and are self-reported rather than observed durations. Thus while the current unemployment status of the respondent is established on the basis of the ILO criteria, the same tests cannot be applied to each of the months of unemployment that constitute the duration declared by the respondent. Furthermore, such recall data are subject to bias - not simply because of the vagueness of memories but also due to approximate answers. The latter give rise to digit preference in reporting the number of months and more importantly to a tendency to round up or down to the nearest quarter or sixth months. The density of declared durations used below has large spikes at 12, 18, 24 etc months-see Figure 20 in appendix for data from French Labour Force Survey 2005:1. The procedure used to correct for this bias is set out in Appendix A.

4 Data and Results

The data are taken from the French Labour Force Survey 2003-2008, which are the statistical basis for the calculation of the national unemployment rate using the ILO definition. Each survey covers around 60,000 households and the information is gathered by questionnaire interview in person in first appearance in the survey and subsequently by telephone in the four subsequent quarters. Each household remains in the survey for six quarters. The quality of the data is higher for the first quarter i.e. at the time of entry into the survey.

The information on labour market status and therefore whether an individual is unemployed is based on responses to questions about search activity and availability for work. Information about unemployment duration is provided by the respondent and refers to a personal estimate of the time spent unemployed in the current spell in months. It is not possible to establish whether the person was 'available and actively seeking work' (that is unemployed according the ILO definition) during the period of unemployment declared by the respondent.

The data used are for economically active persons aged between 20 and 60 and whose durations are less than 24 months. The majority of incomplete spells (more than

half) are of less than 7 months, and there are spikes at 12, 18, 24 (and subsequent multiples of 12) suggesting that after a year the estimates of spell durations are approximate. In what follows, the intervals are of three months : <4 months, 4 to 6 months, 7 to 9 months, 10 to 12 months and so forth. These are matched to corresponding cohorts in subsequent surveys in order to estimate the conditional survival rate using equation 3.5. The mean completed duration is obtained using equation 3.7, which is used to calculate the unemployment index for $\alpha = 1$ (which is simply the product between the mean of completed durations and the unemployment rate). Finally the index is calculated using the estimated density of completed spells using equation 3.8. These are referred to as estimates based on the synthetic cohort method.

The index is also computed using the observed uncompleted durations of unemployment and using Salant's method for obtaining the density of completed durations. Data on interrupted durations are used to estimate the parameters r and a in 3.1 and 3.2 by maximum likelihood. The resulting estimated density of completed durations can be used to obtain the Shorrocks index of unemployment based on the assumption of stationarity.

In what follows, we first compare the different methods used to compute the extent of unemployment in France based to Shorrocks' index. When then proceed to a comparison of the gender gap in unemployment in France.

4.1 Unemployment in France between 2003 and 2008

The quaterly rate of unemployment over 20-60 year olds in France according to the International Labour Organisation (ILO) defined over the period 2003-2008 oscillates between 7.5 percent and 8.5 percent upto 2006:4, with clear sesonal features. Thereafter it falls in two clear jumps to around 6.5 percent in 2008⁶.

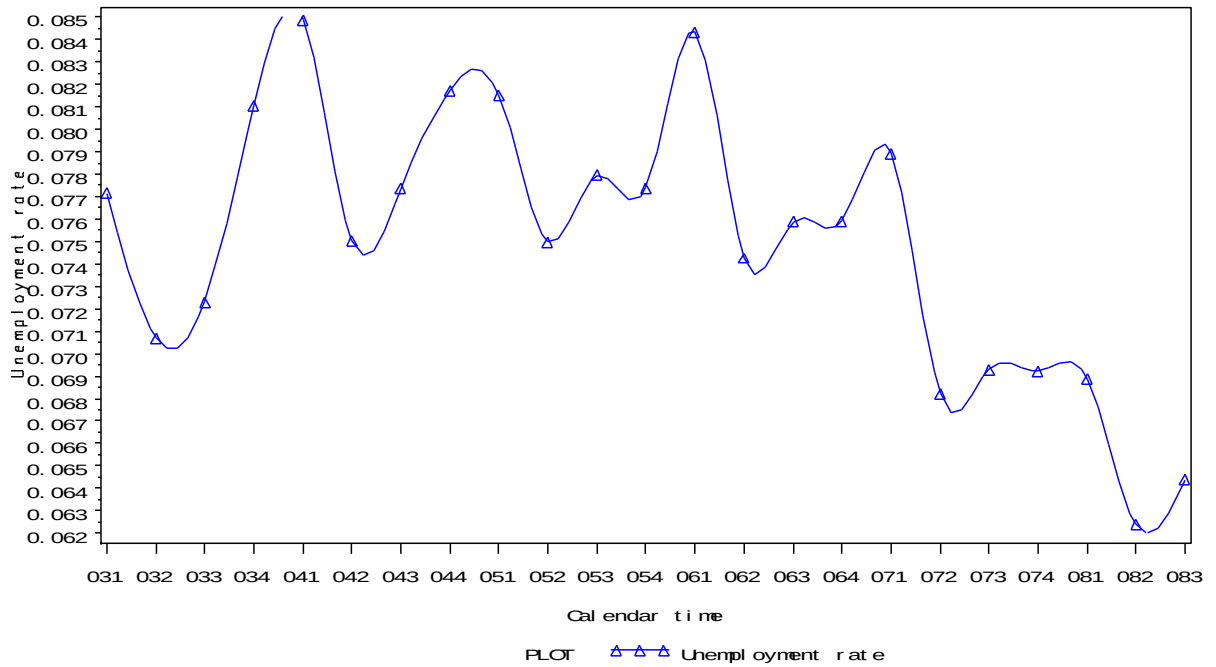
The proportion of labour force participants who are unemployed - or the unemployment rate - is widely used as an indicator of the extent of unemployment, since it is easy to compute and interpret. Figure 1 presents this rate for the sample used here for those aged 20 to 60.

The extent of unemployment depends on inflows and outflows in a mechanical way. In a steady state in which the number (or rate) of unemployed (U) is constant:

$$U = E \times \bar{d} \tag{4.1}$$

⁶This follows the published figure which applies to those aged 15-64

Figure 1. Quarterly unemployment rate between 2003 and 2008 in France(20-60 year olds)



Source: French Labour Force Survey

or alternatively:

$$U_R = \frac{U}{L} = E_R \times \bar{d} \quad (4.2)$$

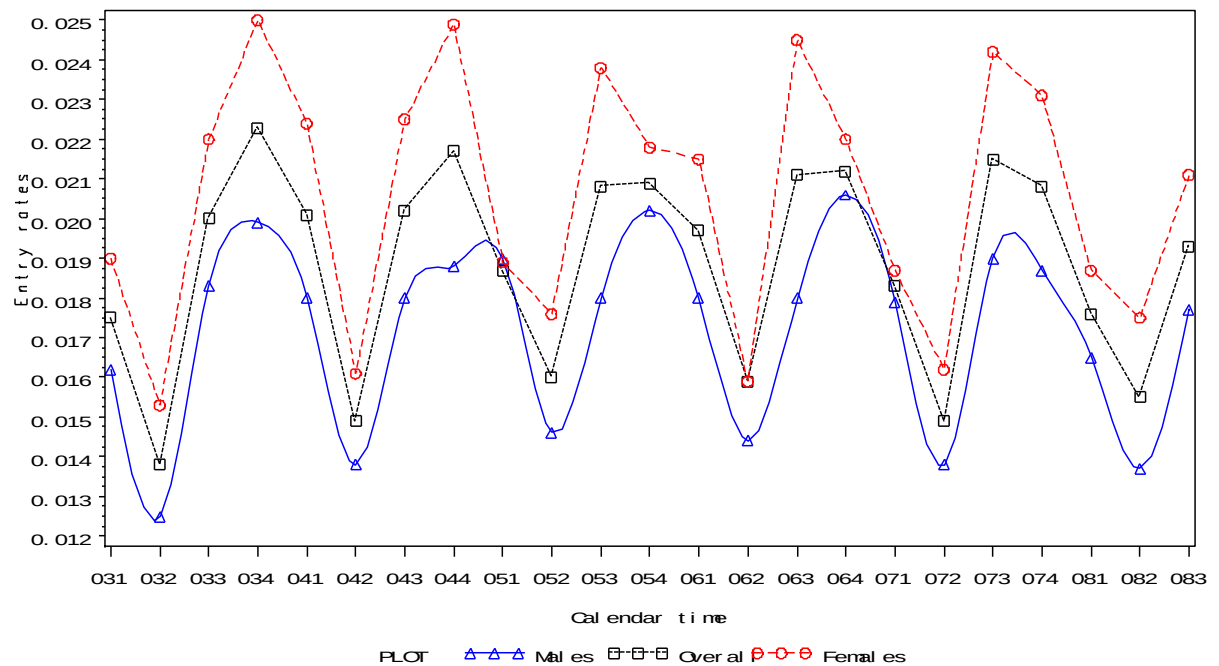
where E_R is the inflow rate and $\bar{d}$ the mean duration.

Between two points in time (two steady states) the variation in the unemployment rate can be due to variations in these two components. The quarterly inflow rates⁷ into unemployment by sex are shown in figure 2. Interestingly the inflow rates vary less than the unemployment rate, and there is little evidence of a decline from 2007 on. However the female inflow rate is higher and the seasonal pattern is not regular.

Figure 3 shows the time path of different estimates of the mean duration. Given that we only have data on interrupted durations, we use the Salant's method and the synthetic cohort method to estimate the mean completed duration. The results in Figure 3 show that using interrupted durations given in labour force surveys lead to an overestimation of the mean duration. This is due to what Salant (1977) calls the 'length bias' which is a sampling bias stemming from the fact that unemployed with long duration have a higher

⁷These are computed as the proportion of unemployed having a duration lower than 3 months (the length of time between two consecutive surveys). The base is the number of unemployed plus the number of employed.

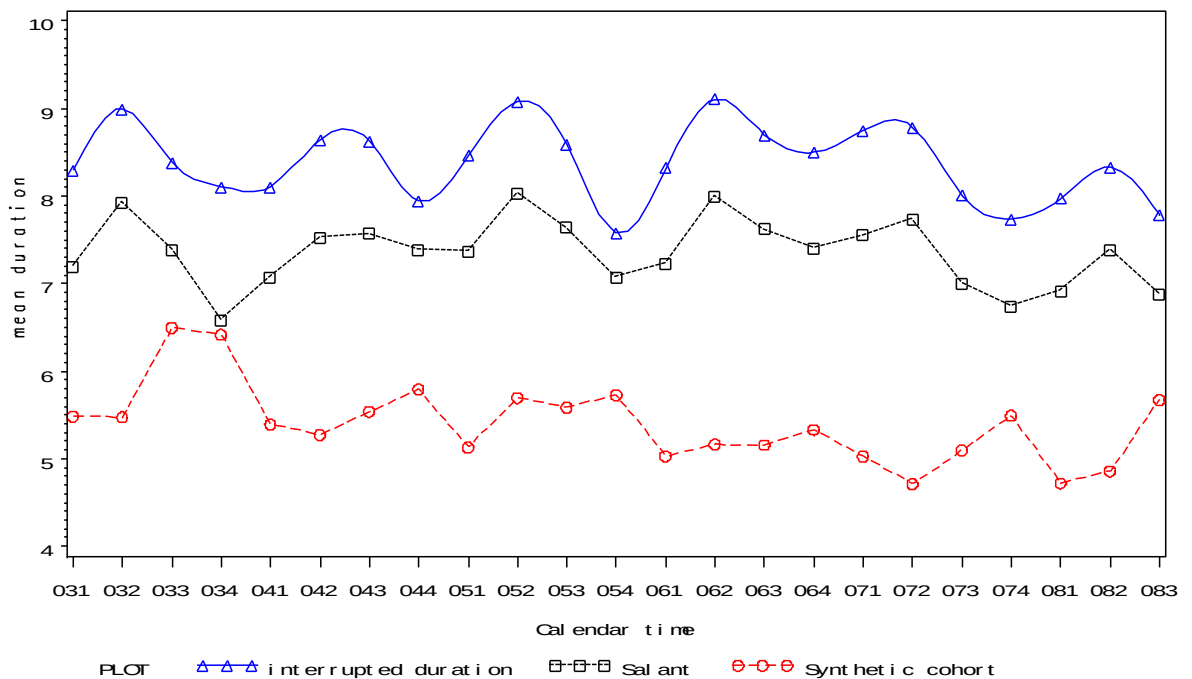
Figure 2. Quarterly entry rate in France between 2003 and 2008 (20-60 year olds)



Source: French Labour Force Survey

chance of being unemployed at the time of the survey. In the current case, length bias is more important than the interruption bias.

Figure 3. Mean duration of unemployment using different definition of durations

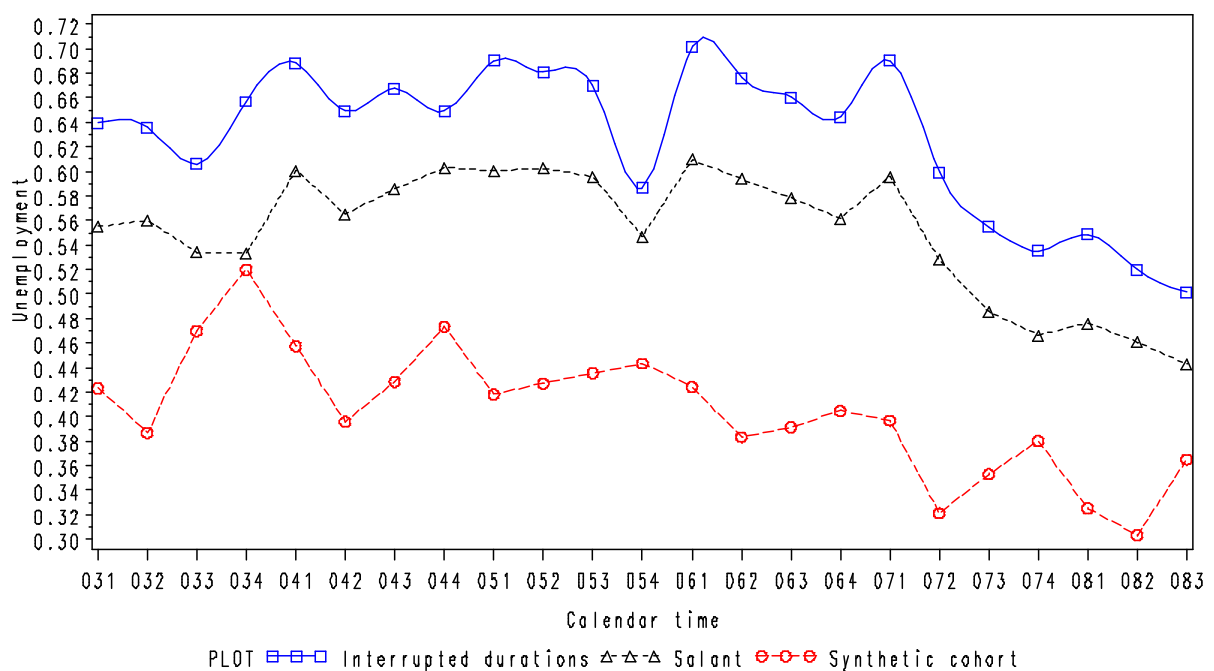


Source: Authors' own calculations

In order to correct for this, [Salant \(1977\)](#) assumes that the labour market is in steady state (inflows=outflows) and obtains a simple relationship between the distribution of interrupted durations and completed durations. The mean duration obtained using Salant's approach (7 months) is lower than the mean interrupted spell (9 months). Over time the two estimates are very highly correlated and exhibit seasonal features. The estimated mean duration based on the synthetic cohort approach is much smaller (between five and six months) and apart from seasonal variation is on a downward trend over the period. The latter estimates, unlike the Salant's approach, do not depend on the stationarity condition according to which inflows to unemployment are constant.

The unemployment is in fact the value of the Shorrocks index when $\alpha = 0$. Multiplying this by the mean duration gives the Shorrocks index for $\alpha = 1$; values of which are presented in Figure 4 using the different methods of estimation. Clearly the declining rate of unemployment is the key contribution to the reduction in the extent of unemployment registered after 2007:1.

Figure 4. Shorrocks' index of unemployment for $\alpha=1$



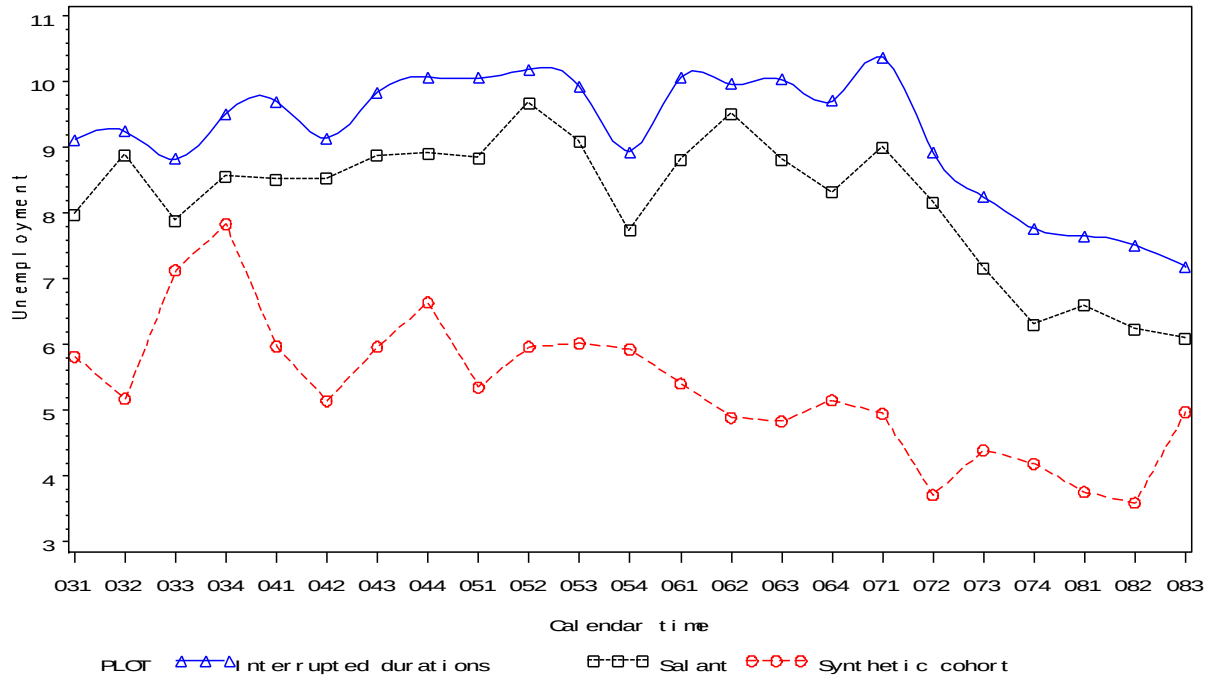
Source: Authors' own calculations

The synthetic cohort based estimates however suggest that declining mean duration has also played a part after 2003:4. Furthermore based on the estimated variances of the two methods (Salant and synthetic cohort), Salant's approach provides less precise estimates due to the restrictive stationarity assumption.

Finally, the distribution of unemployment spells can be incorporated in the mea-

surement of the extent of unemployment by setting $\alpha = 2$. The pattern of estimates over time is nearly identical to the case of $\alpha = 1$ whatever the statistical method used. The synthetic cohort method indicates a downward trend beginning 2003:4, while the others exhibit a certain stability until 2006:4 after which a significant reduction in unemployment is registered.

Figure 5. Shorrocks' index of unemployment for $\alpha=2$

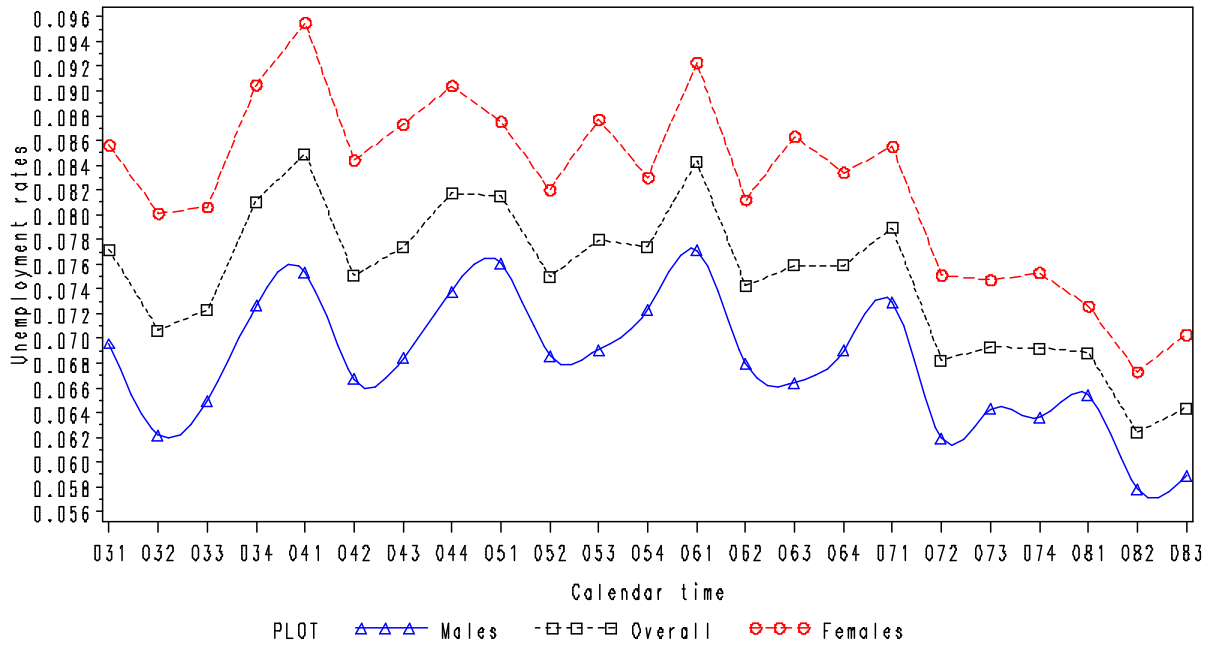


4.2 Analyzing gender gap in unemployment

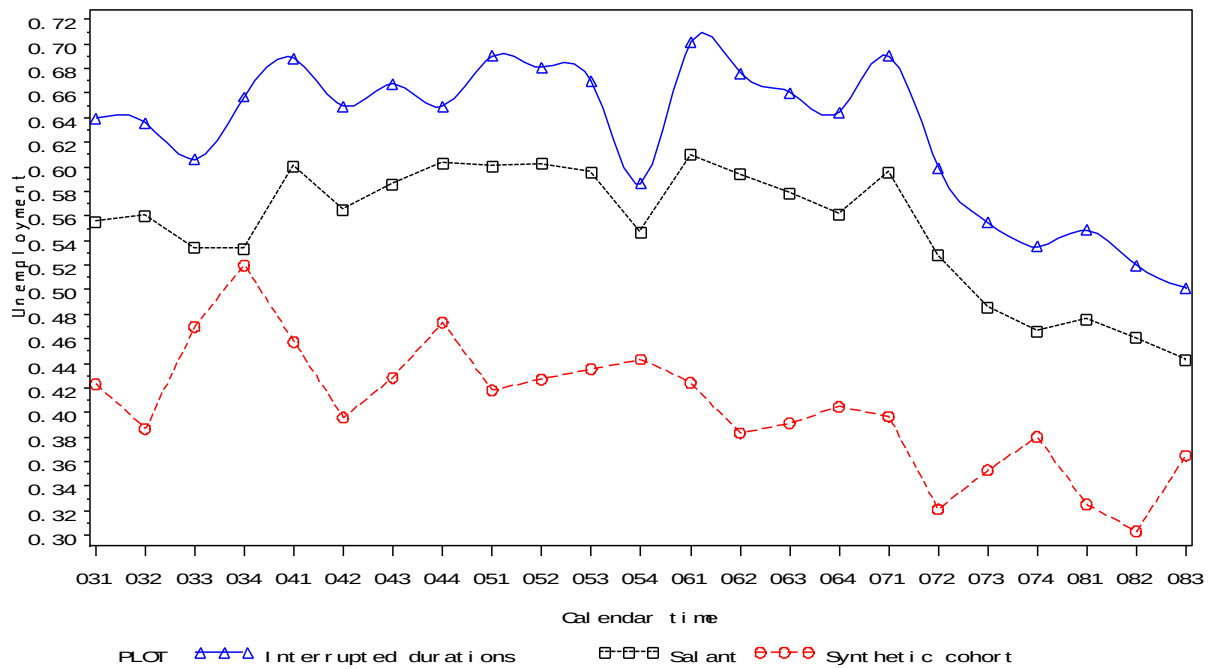
In this section, we examine the gender gap in unemployment in France over the period 2003-2008. We go beyond considering the gap in unemployment rates by using the Shorrocks' index. Figures 6 to 10 (and Figures 21 and 22 in the appendix) show the evolution of the gender gap in unemployment for different values of α for each of methods used to estimate the extent of unemployment.

For both sexes, the unemployment rate ($\alpha = 0$) is stable apart from seasonal variation upto 2006 and thereafter declines for both males and females. The vertical distance between the plots suggests that the gap declined from 2 percentage points (a proportional difference of 30 percent) to less than half a point (or about 10 percent) at the beginning of 2008.3. When the mean duration is incorporated ($\alpha = 1$), slightly different conclusions emerge. Using raw interrupted spell lengths, the gender gap remains fairly constant until

Figure 6. Quarterly unemployment rate (20-60 year olds)

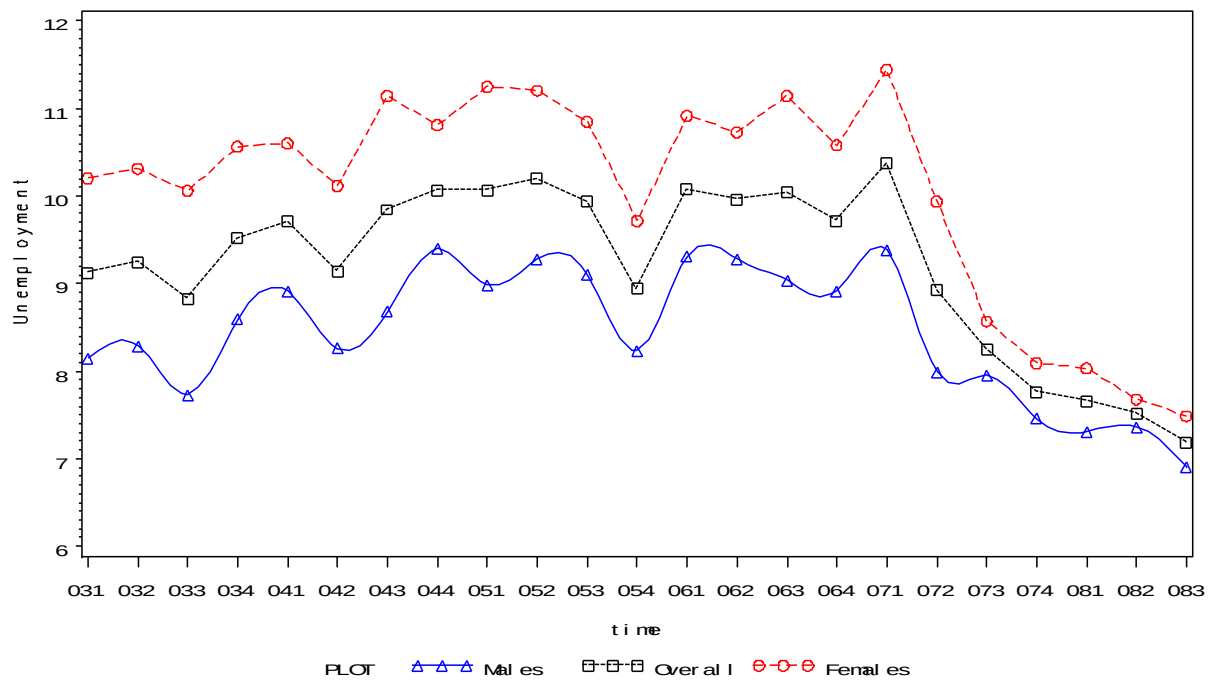


Source: French Labour Force Survey

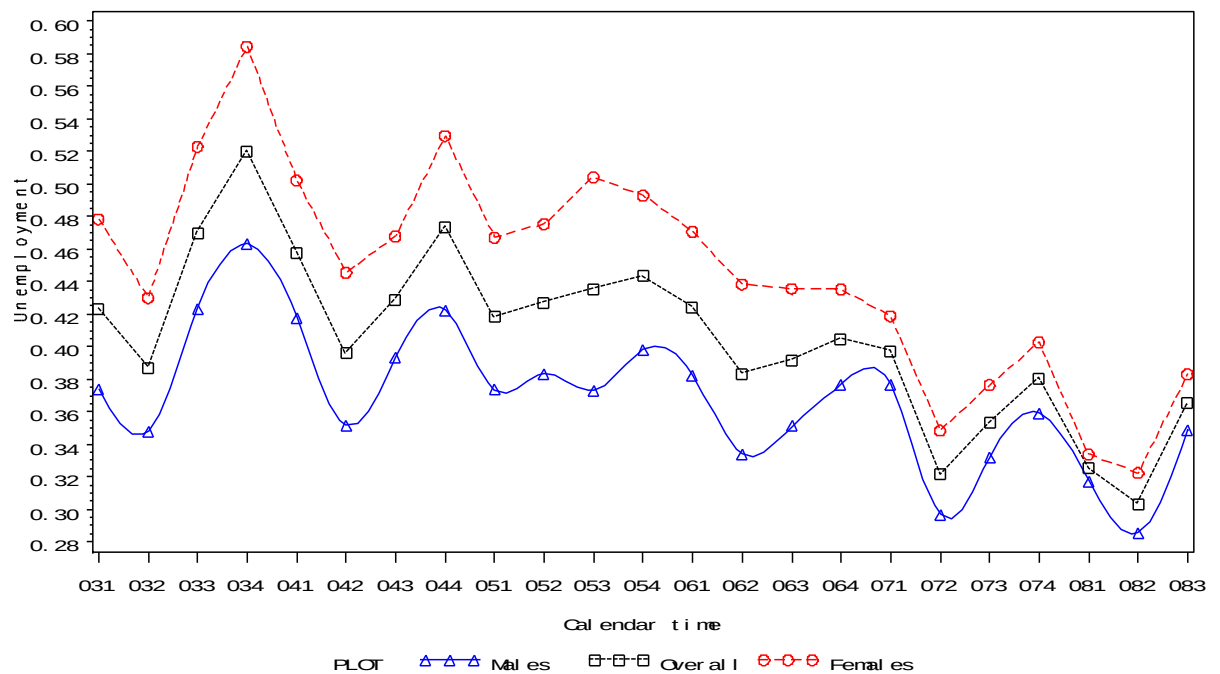
Figure 7. Shorrocks index of unemployment for $\alpha=1$ using interrupted durations

Source: Authors' own calculations

2006 and the indices for females and males converge thereafter (Figure 7). A very similar pattern is found when Salant's method is used-see Figure 21 in the appendix. In contrast, using the synthetic cohort approach, the index for the two sexes begins to converge from

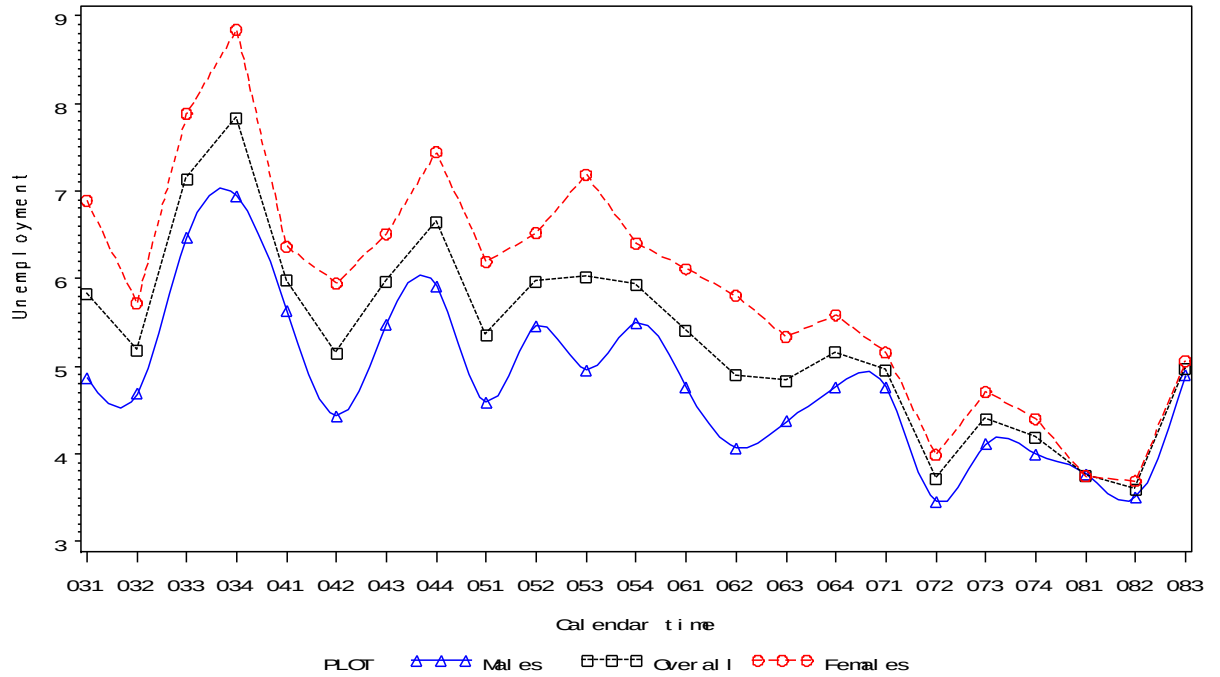
Figure 8. Shorrocks index of unemployment for $\alpha=2$ using interrupted durations

Source: Authors' own calculations

Figure 9. Shorrocks index of unemployment for $\alpha=1$ using synthetic cohort method

Source: Authors' own calculations

Figure 10. Shorrocks index of unemployment for alpha=2 using synthetic cohort method



Source: Authors' own calculations

2006 and through 2007-8 is hardly existent. Finally, taking into account the distribution of spell lengths ($\alpha = 2$) produces non-monotonic, downward changes and convergence in the gap in the extent of unemployment to the case when $\alpha = 1$.

The decomposability characteristic of Shorrocks index enables one to examine each group's contribution to total unemployment-see Figure 11 and 12. These show the contribution of females to total unemployment. Using notations in equation 2.4, It can be expressed as follows:

$$\frac{n_{female}}{n} \frac{U_s^f(\alpha)}{U_s(\alpha)} \text{ for } \alpha = 0, 1, 2.$$

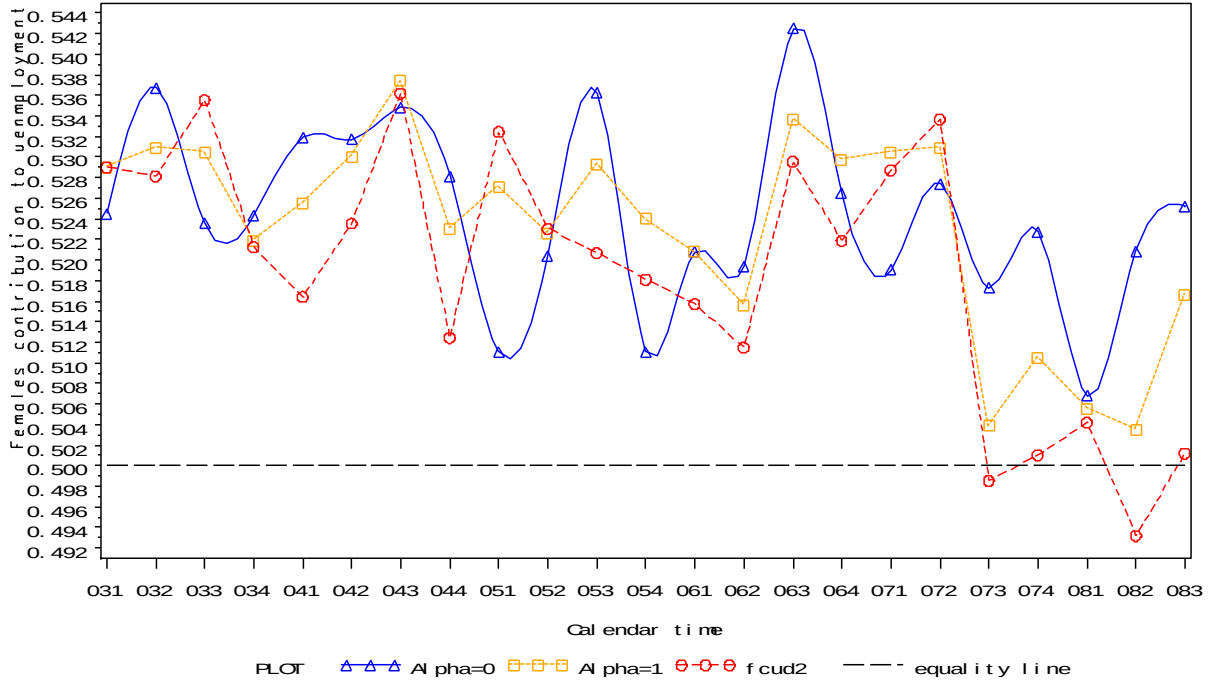
where U_s^f is the Shorrocks index computed over active females and n_f is their number.

These confirm that females are more heavily present in aggregate unemployment than males.

4.3 A test of the convergence of male and female unemployment

In view of the narrowing of the gender gap in unemployment, especially towards the end of the 2003-2008 period, it is appropriate to undertake a formal statistical test. This will

Figure 11. Females' contribution to unemployment for $\alpha = 0, 1, 2$ using interrupted durations



Source: Authors' own calculations

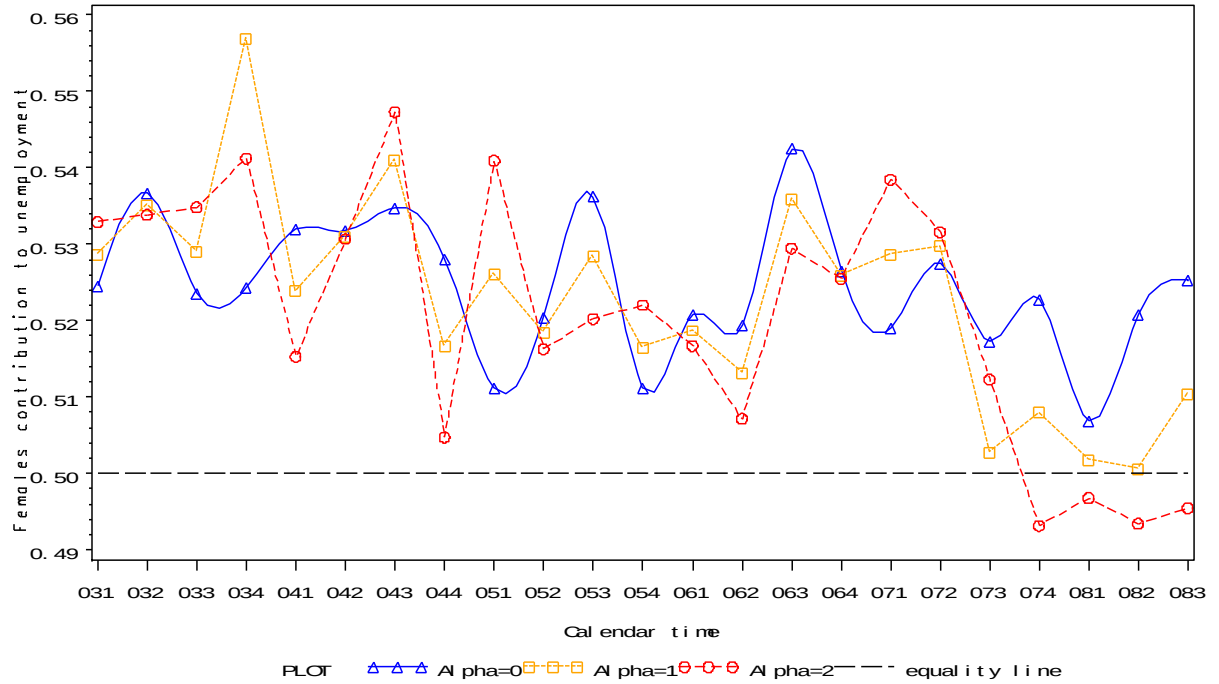
enable us to distinguish first whether the actual gender gap in unemployment is significantly different from zero at any point in the period covered and second, if this is the case, whether the gap disappears towards the end of the period. Since Shorrocks index is analogous to the [Foster, Greer, and Thorbecke \(1984\)](#) index for poverty, the statistical basis for hypothesis testing developed by [Bishop, Chow, and Zheng \(1995\)](#) for the FGT index can also be applied here. The null hypothesis to be tested is:

$$H_0 : U_{\alpha,F} - U_{\alpha,M} = 0 \text{ for } \alpha = 0, 1, 2.$$

Using the value of the Shorrocks index obtained from a sample, $\hat{U}_{\alpha,i}$, the hypothesis can be tested using the following statistic:

$$S = \frac{\hat{U}_{\alpha,F} - \hat{U}_{\alpha,M}}{\sqrt{(\text{Var}(\hat{U}_{\alpha,F}))/N_F + (\text{Var}(\hat{U}_{\alpha,M}))/N_M}} \hookrightarrow ANormal(0,1) \quad (4.3)$$

where $\text{Var}(\hat{U}_{\alpha}) = \hat{U}_{2\alpha} - \hat{U}_{\alpha}^2$.

Figure 12. Females' contribution to unemployment for $\alpha = 0, 1, 2$ using Synthetic cohort

Source: Authors' own calculations

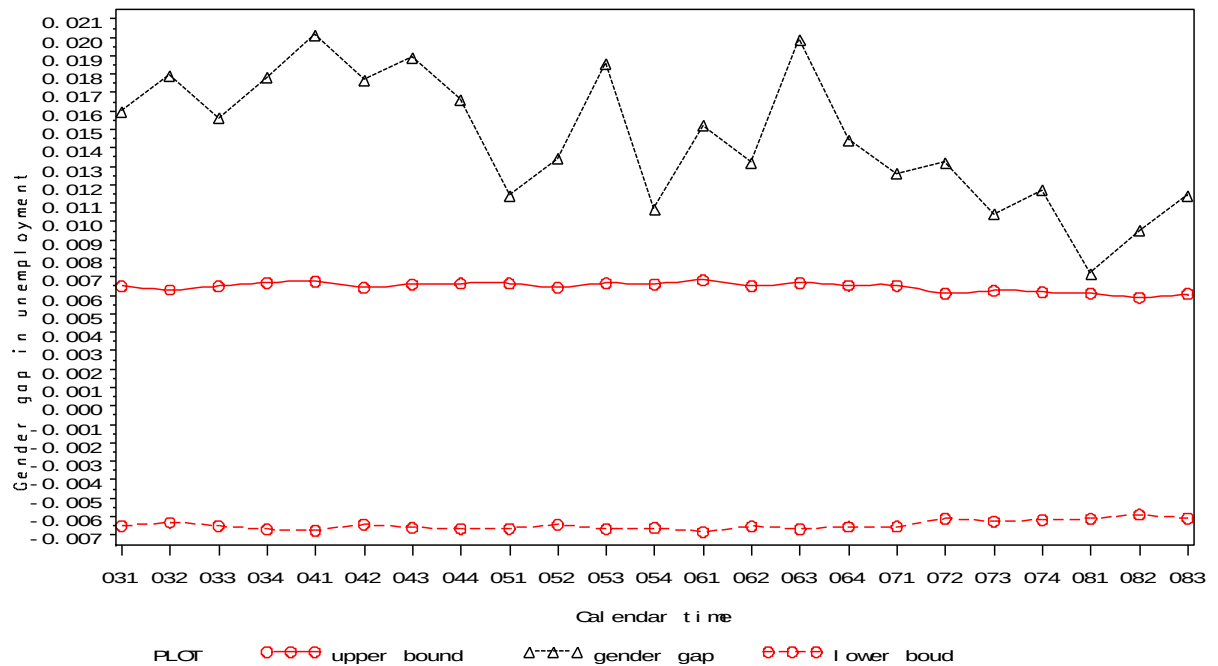
In order to establish if and when the gender gap in unemployment disappears, we present the numerator of the statistic S and the standard error (the denominator of S) multiplied by $+1.96$ and -1.96 in Figures 13 to 17⁸. Thus for $\alpha = 0$, the gap in unemployment rate certainly declines but does not disappear. Using the interrupted spell data to estimate mean durations, the gender gap for $\alpha = 1$ is not different from zero from 2007:3 onwards, but the estimated gap runs along the upper confidence limit.

A similar picture emerges for the synthetic cohort estimate of mean completed durations (Figure 16). When the distribution of spells is taken into account both methods indicate that the gender gap is not significantly different from zero- from 2007:2 for interrupted spell durations and from 2006:3 for the synthetic cohort estimates.

The gender gap in unemployment using Shorrocks' index can be decomposed into the gender gap in rates, the gap in mean durations and that in the inequality index. Figures 18 and 19 (and Figure 26 in the appendix) present the gender ratio in term of these three components. Interestingly, the gender ratio in mean durations and the gender ratio in inequality fluctuate around the equality line while the gender ratio in unemployment rate remains high. These findings mean that the gender gap in unemployment rate is more

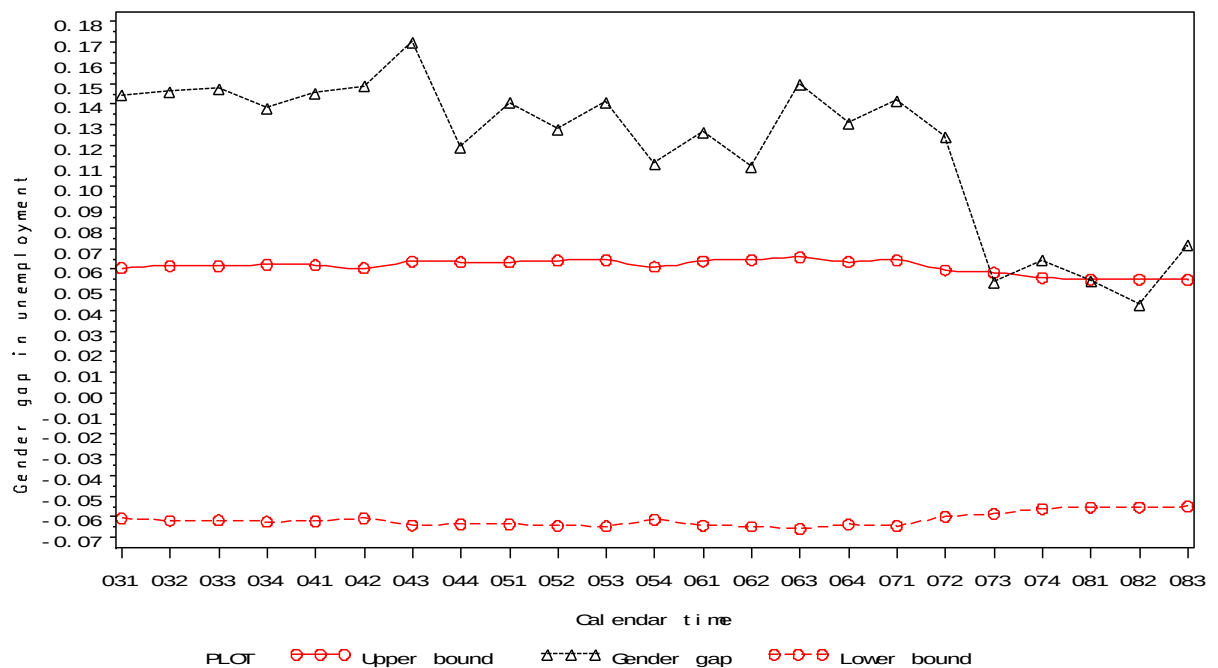
⁸If the gap lies inside the confidence interval, the null hypothesis cannot be rejected

Figure 13. Gender gap in unemployment rate



Source: Authors' own calculations

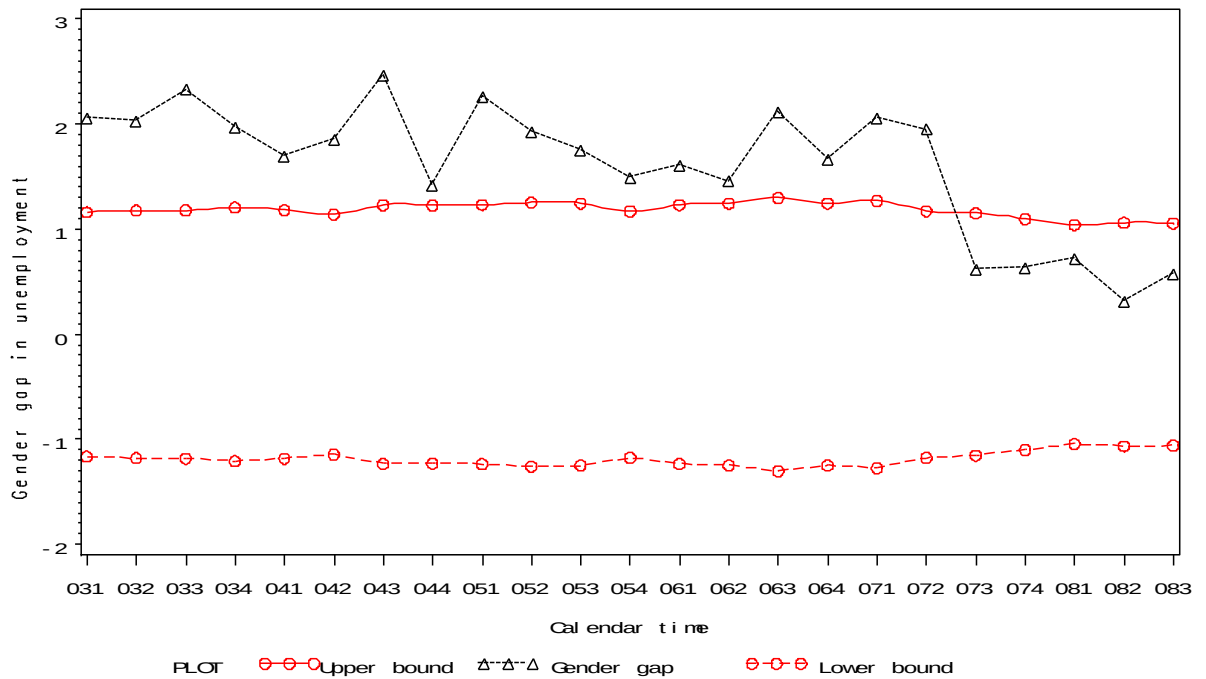
Figure 14. Gender gap in unemployment using Shorrocks' index of unemployment based on interrupted durations for $\alpha=1$



Source: Authors' own calculations

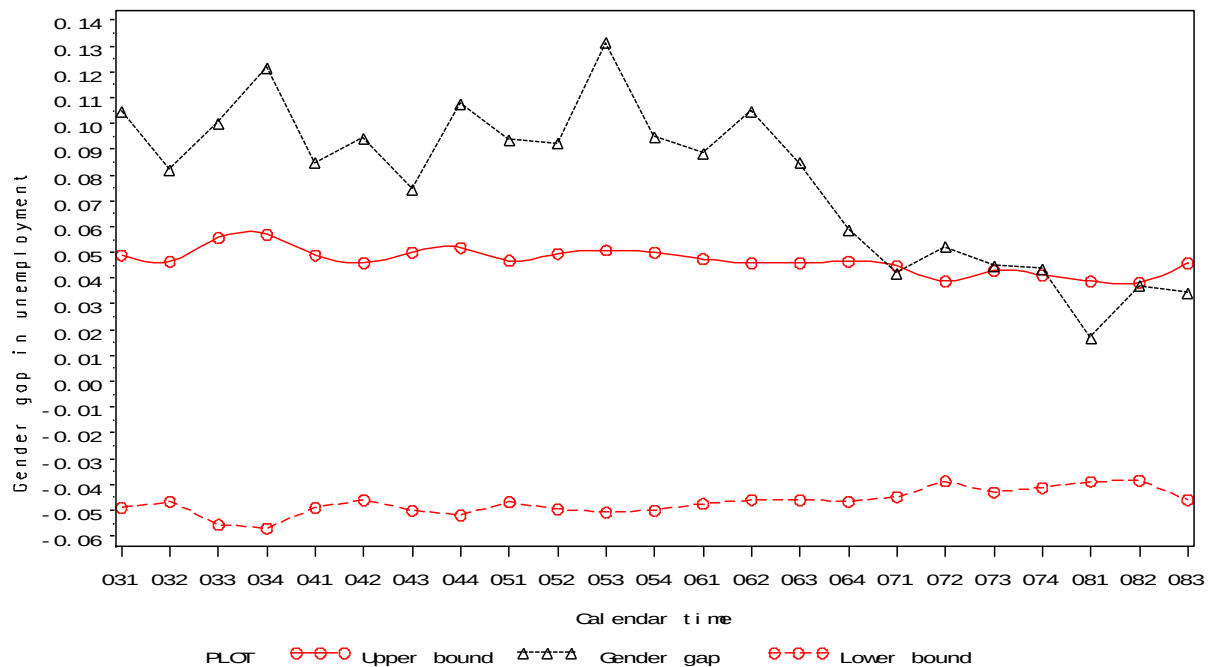
important than the gender gap in mean durations and in inequality of durations. This is true whatever the definition of durations and the method of estimation used.

Figure 15. Gender gap in unemployment using Shorrocks' index of unemployment based on interrupted durations for $\alpha=2$



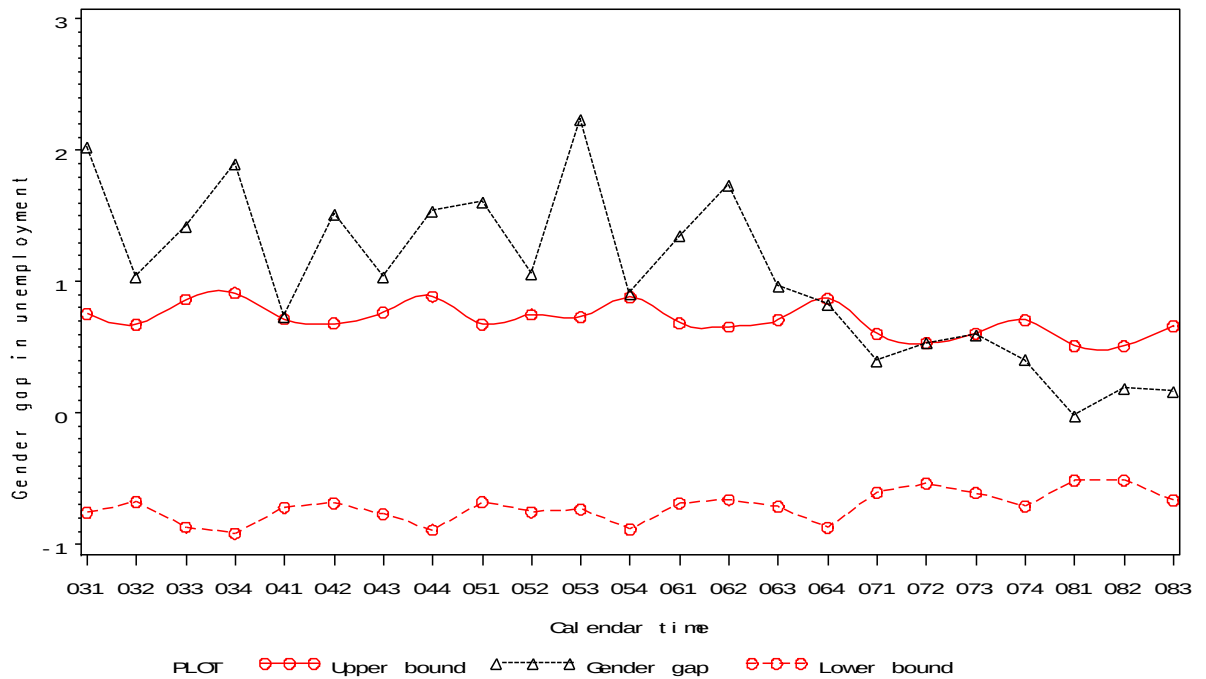
Source: Authors' own calculations

Figure 16. Gender gap in unemployment using Shorrocks' index of unemployment based on synthetic cohort method for $\alpha=1$



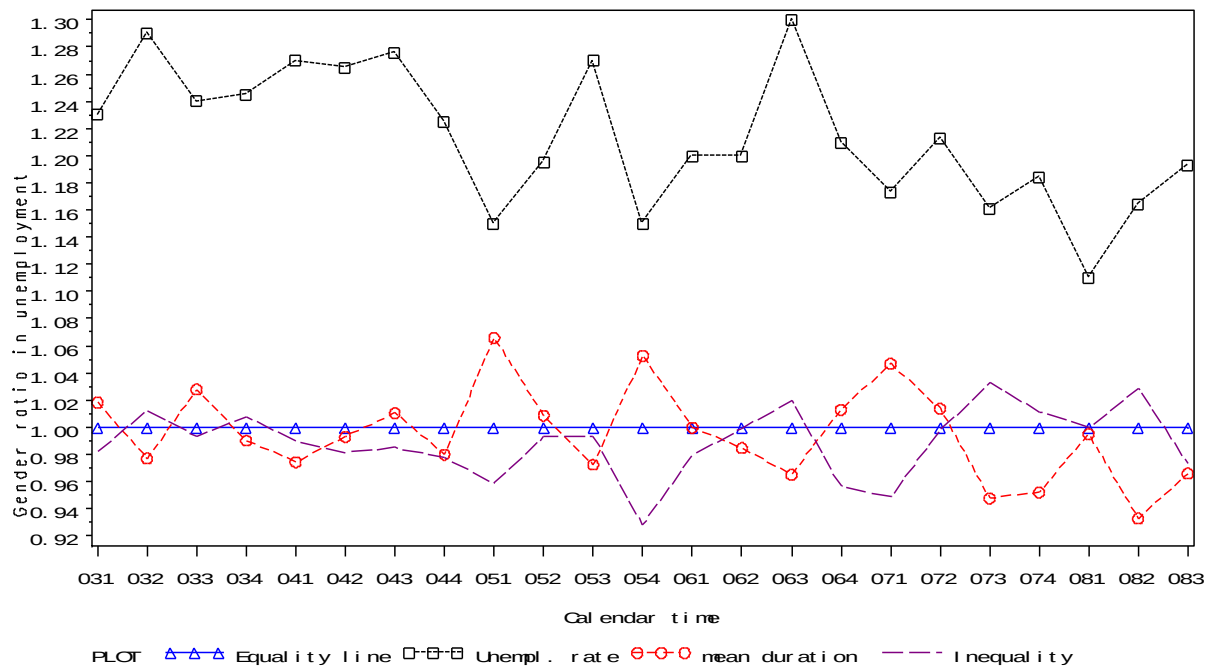
Source: Authors' own calculations

Figure 17. Gender gap in unemployment using Shorrocks' index of unemployment based on synthetic cohort method for $\alpha=2$



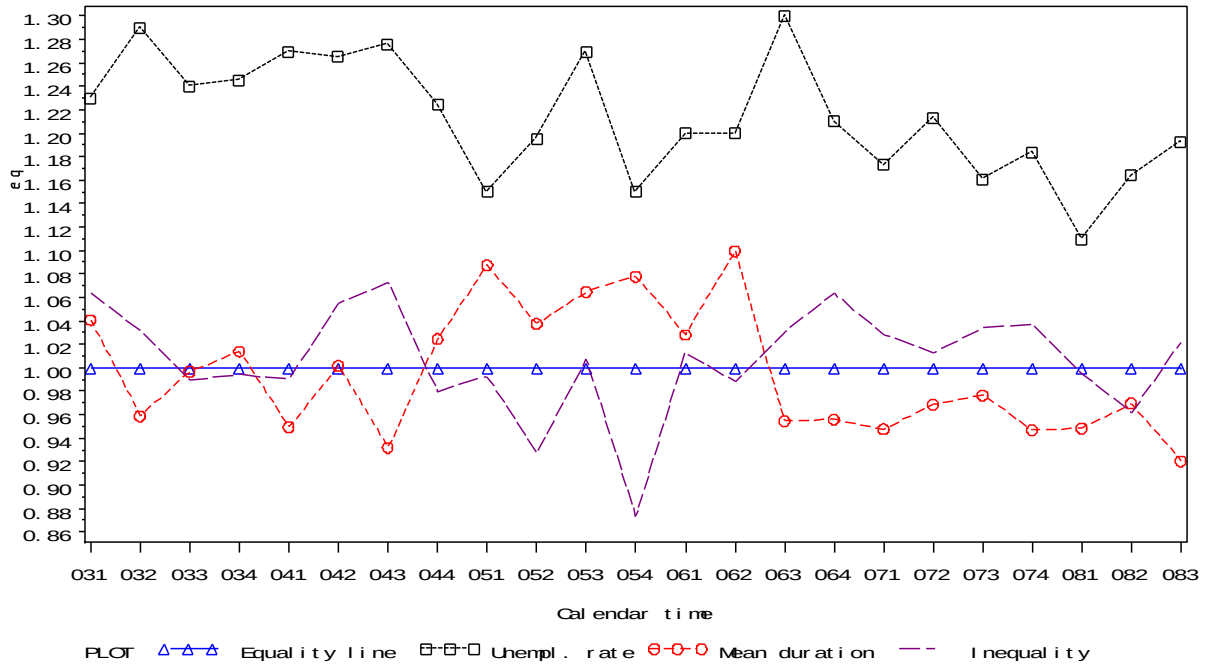
Source: Authors' own calculations

Figure 18. Gender ratio in unemployment using interrupted durations



Source: Authors' own calculation

Figure 19. Gender ratio in unemployment using synthetic cohort method



5 Conclusions

This paper examines how unemployment can be measured in normative fashion - taking into account the mean and inequality of spell lengths - and estimated from cross section data of the type found in labour force surveys. The issue is not straightforward since completed durations of unemployed individuals are not observed yet they constitute the basis for calculating the kind of index that has been proposed to measure the extent of unemployment that goes beyond the unemployment rate. The index proposed by Shorrocks has robust normative foundations and has an equivalent representation in terms of average complete duration and the density of completed durations. Building upon earlier work applied in the United States for estimating the first of these, we present a method that enables the index to be calculated based on an estimate of the density of completed durations. The approach is illustrated in the context of comparing male-female unemployment differences in France.

Several clear-cut conclusions emerge from the empirical application of these methods. First, the stationarity assumption is too restrictive: using interrupted spell lengths or Salant's approach leads to an overestimation of the mean completed duration and the Shorrocks index. Secondly, the synthetic cohort method has a lower variance (empirically) than the Salant's approach. Third, the gender gap in unemployment in France is

statistically significant in the the mid 2000s but disappears towards the end of the decade. Fourth, while most of the gender gap is due to a higher rate of unemployment, the distribution of durations plays a role in the compression of the gap and the convergence observed in the final years of the period analysed.

References

- AABERGE, R., AND T. WENNEMO (1990): "Non-stationary inflow and duration of unemployment," *Discussion paper-Central Bureau of Statistics Oslo*.
- AKERLOF, G. A., AND B. G. M. MAIN (1980): "Unemployment spells and unemployment experience," *American Economic Review*, 70(5), 885–93.
- (1981): "An experience-weighted measure of employment and unemployment durations," *American Economic Review*, 71(5), 1003–11.
- ATKINSON, A. B. (1970): "On the measurement of inequality," *Journal of Economic Theory*, pp. 244–63.
- AZMAT, G., M. GÜELL, AND A. MANNING (2006): "Gender gaps in unemployment rates in OECD countries," *Journal of Labor Economics*, 24(1), 1–38.
- BAKER, G. M., AND P. K. TRIVEDI (1985): "Estimation of unemployment duration from grouped data: a comparative study," *Journal of Labor Economics*, 3(2), 153–74.
- BAKER, M. (1992): "Unemployment duration: compositional effects and cyclical variability," *American Economic Review*, 82(1), 313–21.
- BISHOP, J. A., K. V. CHOW, AND B. ZHENG (1995): "Statistical inference and decomposable poverty measures," *Bulletin of Economic Research*, 47(4), 0307–3378.
- BOROOAH, V. K. (2002): "A duration-sensitive Measure of the unemployment rate: theory and application," *LABOUR*, 16(3), 453–468.
- CAMARDA, C. G., P. H. C. EILERS, AND J. CAMPE (2008): "Modelling general patterns of digit preference," *Statistical Modeling society*, 8(4), 485–401.
- DISNEY, R. (1979): "Recurrent spells and the concentration of unemployment in Great Britain," *Economic Journal*, 89, 109–119.
- FOSTER, J., J. GREER, AND E. THORBECKE (1984): "A class of decomposable poverty measures," *Econometrica*, 52, 761–776.
- GUELL, M., AND L. HU (2006): "Estimating the probability of leaving unemployment using uncompleted spells from repeated cross-section data," *Journal of Econometrics*, 133(1), 307–341.

- KAITZ, H. (1970): "Analyzing the length of spells of unemployment," *Monthly Labor Review*, 15, 11–20.
- NICKELL, S. J. (1979): "Estimating the probability of leaving unemployment," *Econometrica*, 47(5), 1249–66.
- PAUL, S. (1992): "An illfare approach to the measurement of unemployment," *Applied Economics*, 24, 739–743.
- RIESE, M., AND K. BRUNNER (1998): "Measuring the severity of unemployment," *Journal of Economics*, 67(2), 167–180.
- SALANT, S. W. (1977): "Search theory and duration data: a theory of sorts," *The Quarterly Journal of Economics*, 91(1), 39–57.
- SEN, A. (1976): "Poverty: an ordinal approach to measurement," *Econometrica*, 44(2), 219–231.
- SENGUPTA, M. (2008): "Unemployment duration and the measurement of unemployment," *Journal of Economic Inequality*, 7, 273–294.
- SHORROCKS, A. (2008a): "On the measurement of unemployment," *Journal of Economic Inequality*, 7, 311–327.
- (2008b): "Spell incidence, spell duration and the measurement of unemployment," *Journal of Economic Inequality*, 7, 295–310.
- SIDER, H. (1985): "Unemployment duration and incidence: 1968-82," *American Economic Review*, 15(3), 461–472.

A Appendix

A.1 The treatment of digit preference

The synthetic cohort method to estimate the survivor function is based on frequencies of observed durations. This method of estimation is based on frequencies of observations on durations. While this kind of estimation is theoretically and intuitively correct, there may be problems due to sampling when it is applied⁹, since the conditional survival could be greater than one if there are errors in the reporting of the durations. Unfortunately, this kind of problem is likely to occur for the duration variable because the duration is not observed; it is obtained by asking how long they have been unemployed. It is clear that such a procedure involves reporting errors as in the case of earnings and other variables that are measured from information given by individuals. Individuals are likely to give a digit number which is close to the right one. This phenomenon is known as *digit preference*.

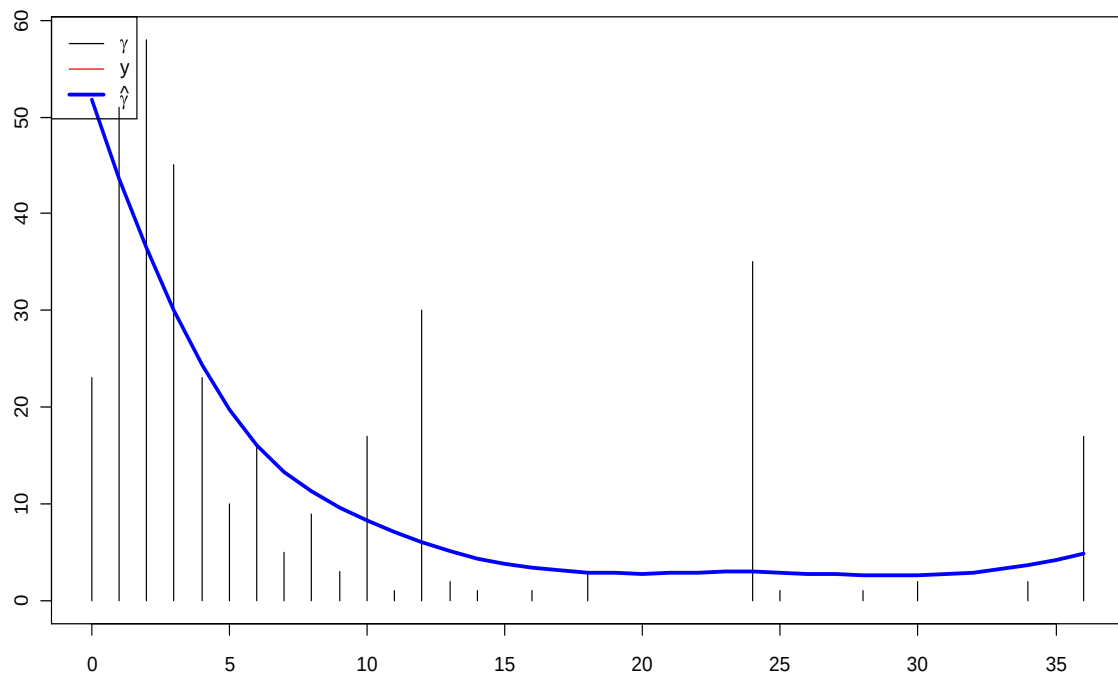
Figure 20 gives an illustration of this phenomenon using a sample of durations drawn from the French Labour Force survey in which we omit observations with duration greater than 36 months and where age is not in the interval 20 to 60, in order to have a stable sample. As it can be seen there is spikes in round numbers such as 12, 24, 36 months. The spikes are explained by the preference of individuals to give round numbers instead of their correct length of ongoing spells. This deficiency in the data causes problems for the implementation of the synthetic cohort method using French data.

Although we cannot retrieve individuals' precise interrupted durations, we can correct the frequencies using a method proposed in [Camarda, Eilers, and Campe \(2008\)](#). The method is based on the Composite Link Model (CLM) to retrieve the right frequencies by a smoothing exercise consisting of splitting frequencies assumed to present misreporting, into the neighbouring interval (as in Figure 20).

In order to see how this method works, assume that the correct frequencies, in the absence of digit preference, are given by $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{36})$ for the case where there are 36 frequencies (as in Figure 1). The data generation process of the frequencies is given by $\mu = C\gamma$. The observed frequencies are realizations from a Poisson process y with parameter $\mu = E(y)$. C is a composite matrix that ensures the digit preference mechanism with a redistribution of some of the 36 elements of γ to the neighbouring intervals. They assume that $\gamma = \exp(\beta)$ to ensure non negative γ . The estimation of the vector γ is undertaken using iteratively reweighted least-square (IWLS). They propose also an estimation

⁹See [Baker and Trivedi \(1985\)](#) for further information on drawbacks of this method.

Figure 20. Correction of data affected by digit preference

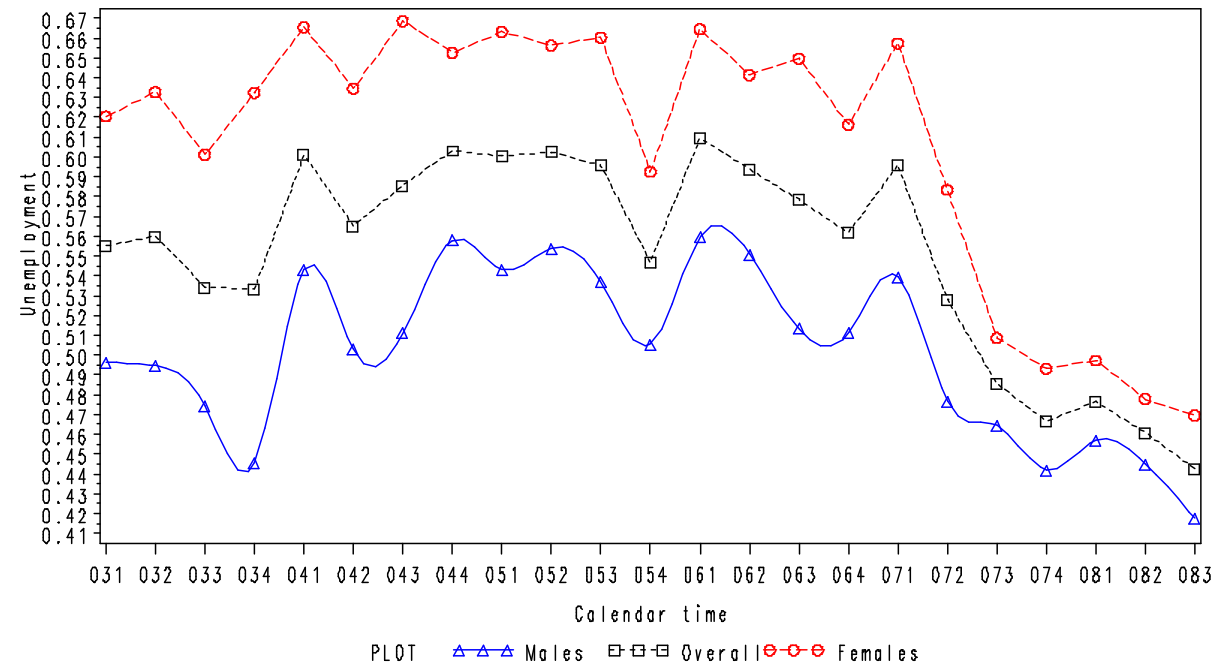


Note: corrected (blue line) and uncorrected (bar) frequencies of the distribution of durations

of the matrix C that creates the allocation mechanism that brings about a redistribution. The results of this procedure are represented by the continuous line in Figure 20. The smoothing results from the penalization contained in the likelihood function used.

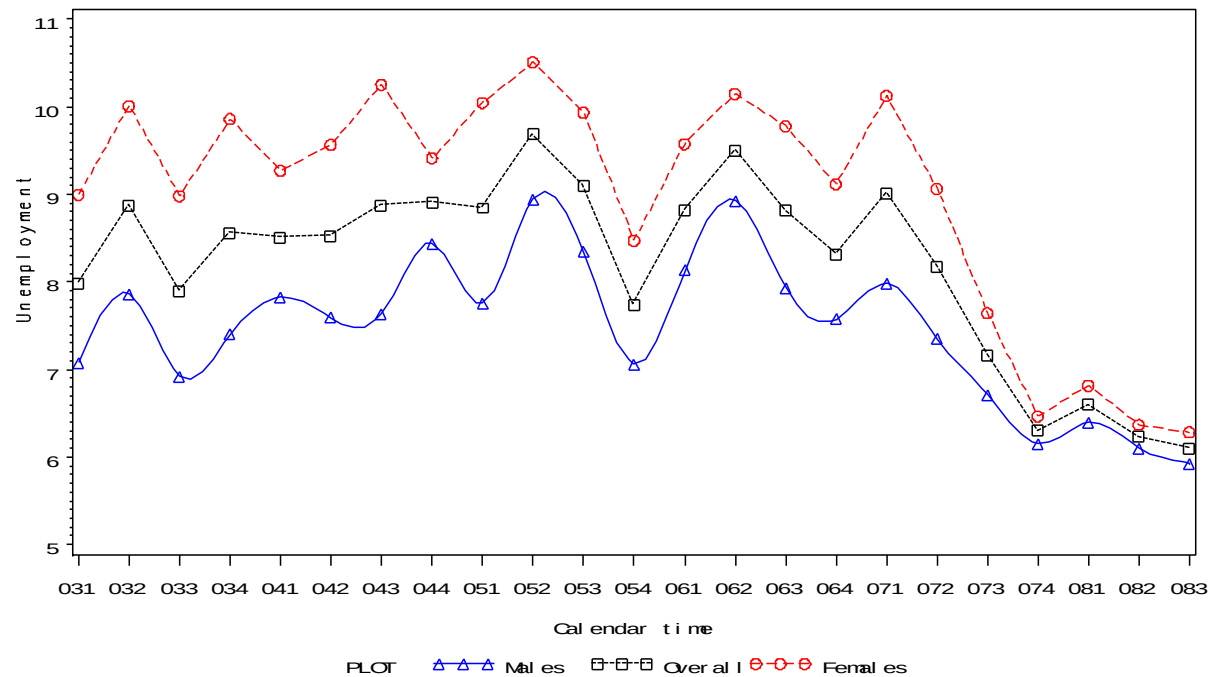
A.2 Graphs: Further graphical analysis

Figure 21. Shorrocks index of unemployment for $\alpha=1$ using Salant's method



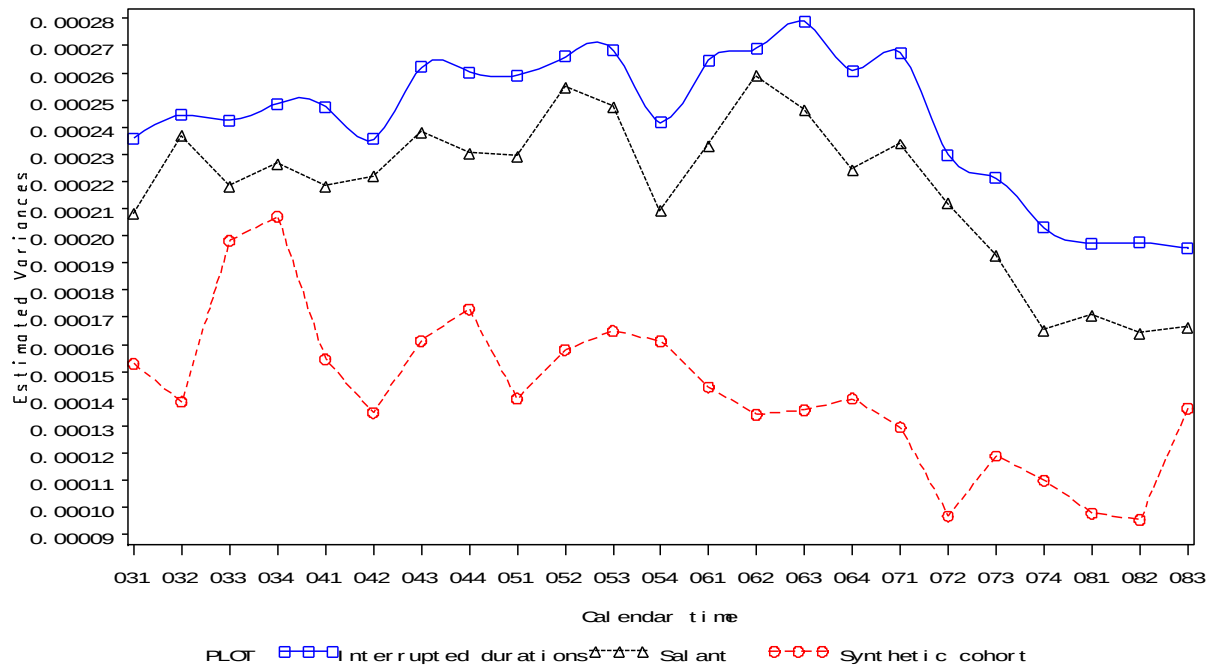
Source: Author's own calculations

Figure 22. Shorrocks index of unemployment for $\alpha=2$ using Salant's method



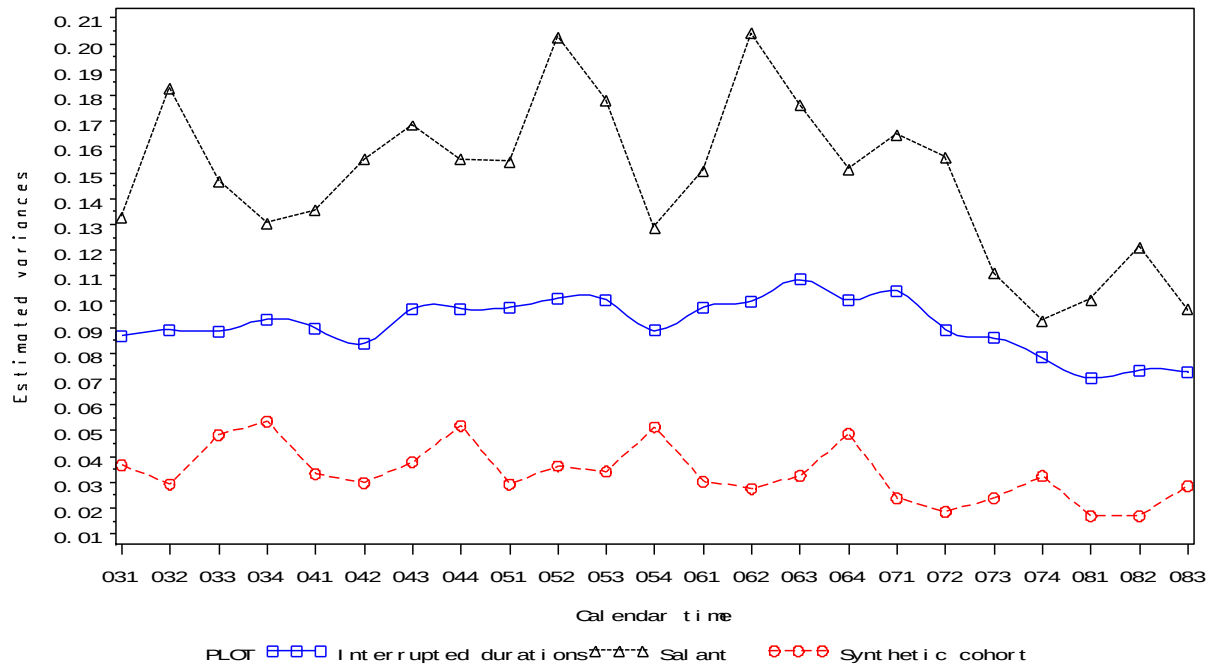
Source: Authors' own calculations

Figure 23. Quality of fit



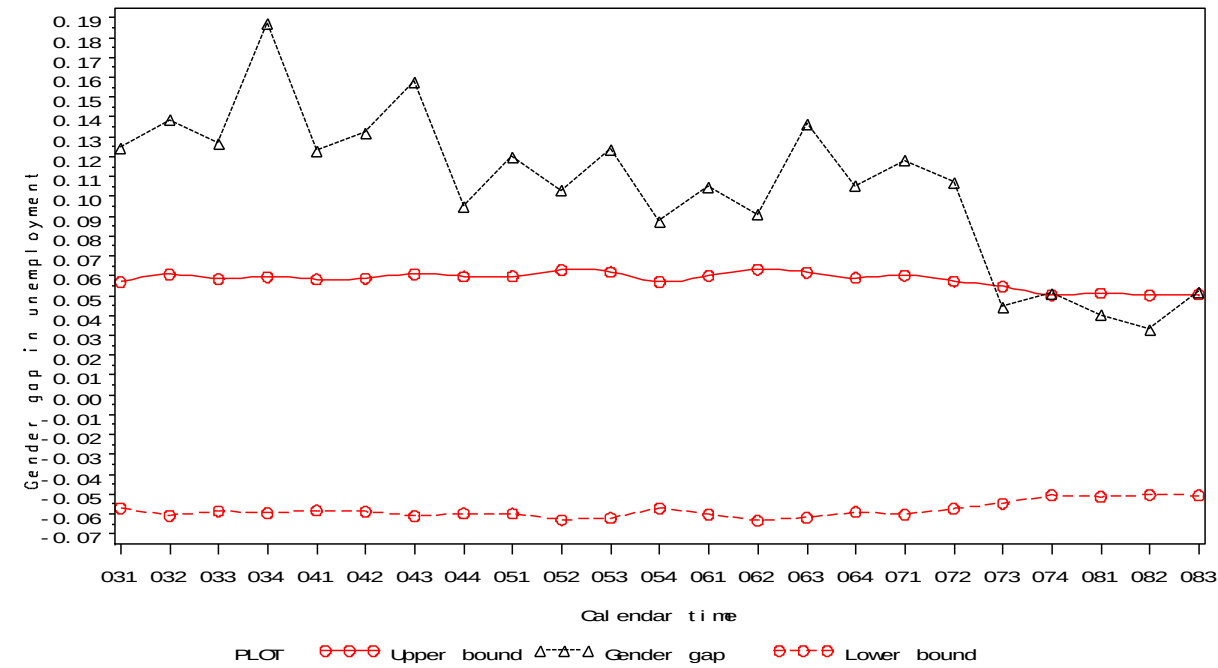
Source: Authors' own calculations

Figure 24. Quality of fit



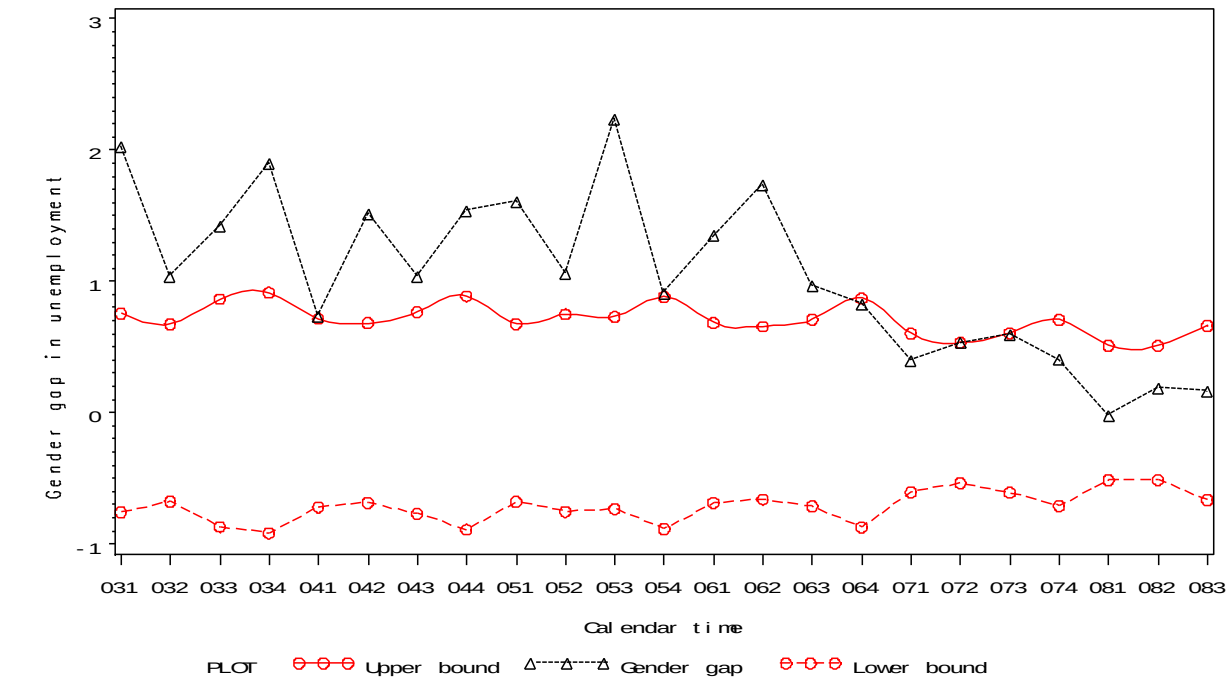
Source: Authors' own calculations

Figure 25. Gender gap in unemployment using Shorrocks' index of unemployment based on Salant method for $\alpha=1$



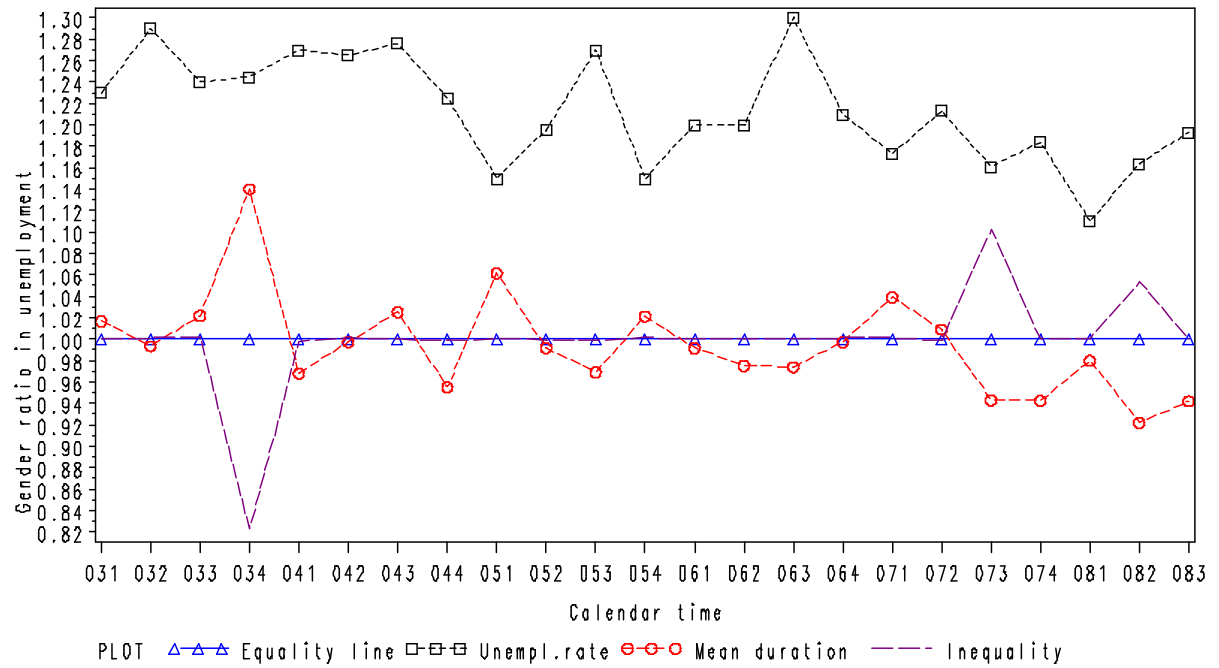
Source: Authors' own calculations

Figure 26. Gender gap in unemployment using Shorrocks' index of unemployment based on Salant method for $\alpha=2$



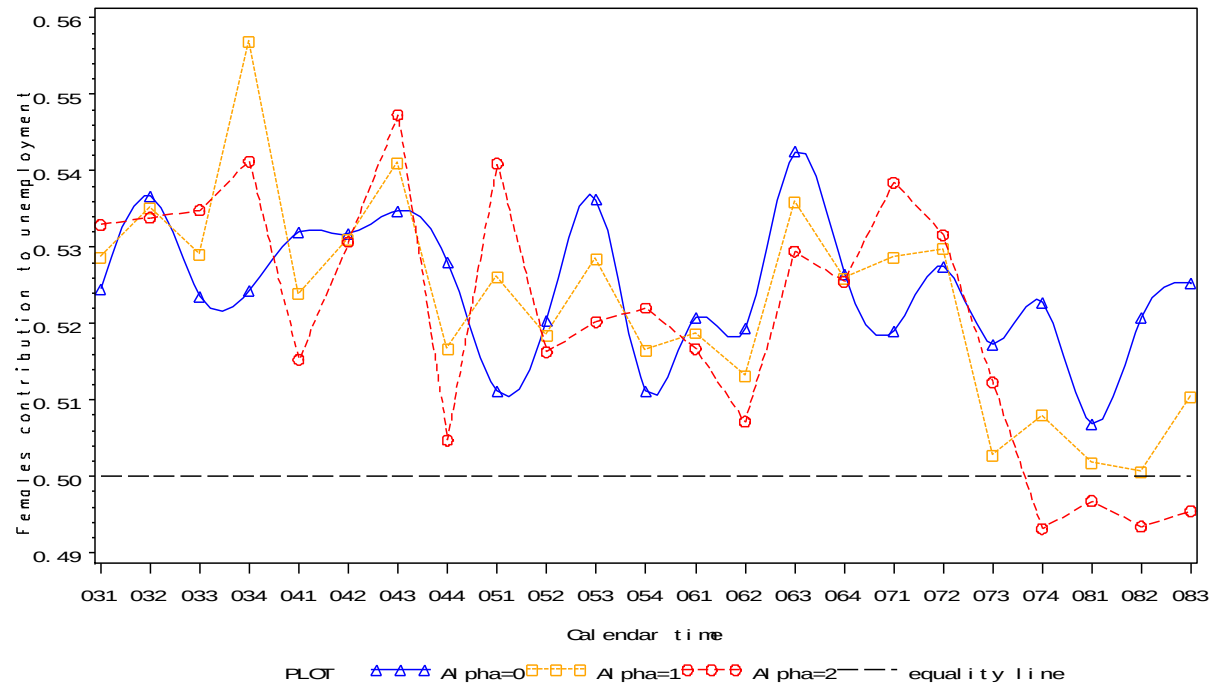
Source: Authors' own calculations

Figure 27. Gender ratio in unemployment using Salant method



Source: Authors' own calculations

Figure 28. Females' contribution to unemployment for $\alpha = 0, 1, 2$ using Salant method



Source: Authors' own calculations