Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

MPIDR WORKING PAPER WP 2001-021
JULY 2001

Fertility and female employment reconsidered: A macro-level time series analysis

Henriette Engelhardt (engelhardt@demogr.mpg.de)
Tomas Kögel (koegel@demogr.mpg.de)
Alexia Prskawetz (fuernkranz@demogr.mpg.de)

Fertility and female employment reconsidered: A macro-level time series analysis*

Henriette Engelhardt, Tomas Kögel and Alexia Prskawetz

Max Planck Institute for Demographic Research,

Doberaner Str. 114, D-18057 Rostock, Germany

July 24, 2001

Abstract

Our paper examines causality and parameter instability in the long-run relation between fertility and female employment. This is done by cross-national comparison of macro-level time series data. By applying error-correction models – a combination of Granger-causality tests with recent econometric time series techniques – we find causality in both directions. This finding is consistent with simultaneous movements of both variables brought about by common exogenous factors such as social norms, social institutions and financial incentives. We find a significant negative correlation until the 1970s, respectively 1980s (depending on the country under investigation) and an insignificant or weaker correlation afterwards. This result is consistent with a recent hypothesis in the demographic literature according

*For helpful comments and suggestions we would like to thank Juha Alho, Horst Entorf, Ulf Christian Ewert, Tomas Frejka, Jan Hoem, Johannes Huinink, Stefan Klasen, Michaela Kreyenfeld, Thomas Lindh, Michael Murphy and Warren Sanderson. In addition, we thank participants of workshops in Heidelberg and Rostock, and at the Annual Meeeting of the Population Association of America 2001 in Washington, DC and the Annual Conference of the European Society for Population Economics 2001 in Athens. We are especially grateful to Tom DiPrete who essentially contributed to improve the exposition of the paper. The views expressed in this paper are the authors' own views and do not necessarily represent those of the Max Planck Institute for Demographic Research.

to which changes in the institutional context like childcare availability and attitudes towards working mothers might have reduced the incompatibility between childrearing and female employment.

1 Introduction

The relationship between fertility and female labor force participation is a long-standing question in demography. One has generally argued that a negative association between these two variables is evidence for the incompatibility of raising children and staying in the workforce in today's society, where the place of work and home are normally separated spatially. Decreasing fertility is thus associated with increasing female employment, and rising female employment is associated with falling fertility. It remains unclear whether these mutual relations are causal in one direction or the other.

The question "What causes what?" has received renewed attention in the demographic literature in recent years. This renewed interest resulted from recent studies which have shown that a simple cross-country correlation coefficient between the total fertility rate and the female labor force participation rate switched from a negative value before the 1980s to a positive value thereafter. The question then arises as to whether there is any causal relationship at all.

Several studies go beyond calculating the correlation and explicitly attempt to test for the existence and direction of causality between fertility and female employment. Due to substantive and methodological shortcomings these studies have found conflicting results. Our paper aims to clarify the question of a relationship between fertility and female employment in three specific ways. First, we apply methods that are designed to avoid the problem that is referred to as "spurious regression" in the time series literature. This problem frequently plagues the analysis of variables with trends, and it arguably afflicts existing efforts to estimate the causal relation between female labor force participation and fertility. Second, we estimate what are called "error-correction models," which are the appropriate models to test for causality between stochastic trending time series. Third, we explicitly test for "parameter instability," i.e., the possibility that the causal relation between the total fertility rate and female labor force participation has changed over time.

Our results are twofold: (i) We find that there was a long-run causal relation linking the total fertility rate and female employment until the end of the 1970s or the begin of the 1980s (depending on the country of investigation). The results suggest that the causation was in both directions, but in any case, we statistically rule out the possibility that the well-known correlation between these variables was spurious. We argue that this bidirectional causality is consistent with the view that comovements of fertility and female employment are caused by common exogenous factors. These factors can be social norms and social institutions, but also opportunity costs of childrearing and family income. (ii) Our empirical results show a significant negative correlation between fertility and female employment from the 1960s to about the beginning of the 1970s for all countries in our study. However, for the non-Mediterranean countries in our sample the causal relation becomes weaker (though it is still significant and negative) at about the end of the 1970s. Finally, we find considerable country-level heterogeneity in the relationship between fertility and female employment in the 1980s and 1990s. In some countries, the relation becomes weaker to the point of insignificance, while in others the relation remains significantly negative.

Our results can be seen as supporting a recent hypothesis in the demographic literature according to which *societal level responses* have eased the incompatibility between childrearing and female employment in most developed countries (Brewster and Rindfuss 2000; Rindfuss et al. 2000; Rindfuss and Brewster 1996). For the Mediterranean country in our sample, Italy, we find the opposite result. That is, the negative correlation between fertility and female employment becomes even stronger across time. This finding coincides with the argument in Rindfuss et al. (2000) and Brewster and Rindfuss (2000) that we are only likely to see increasing female employment not leading to a decrease in fertility in countries that have succeeded in minimizing the incompatibility between childrearing and female work. In Mediterranean countries this incompatibility between female employment and childrearing still persists.

The set up of our paper is as follows. In Section 2 we first discuss the possible relationships between fertility and female employment from a micro-theoretical point of view. We then discuss the gap in the existing macro-level time series literature which we aim to close with our paper. In Section 3 we describe the data, and in Section 4 we explain the econometric methodology that is applied in the paper. We present the results in Section 5 and our conclusions in Section 6.

2 Theoretical and methodological considerations

2.1 Micro explanations

At the individual level, numerous studies have shown a negative association between fertility and female labor force participation (e.g., Lehrer and Nerlove 1986; Brewster and Rindfuss 2000). On average, women in gainful employment tend to have fewer children, and women with children spend less time in the labor market. Weller (1977: 43) lists four possible explanations for this negative association between fertility and female labor force participation:

1. women's fertility affect their labor force participation;
2. women's labor force participation affect their fertility;
3. both women's fertility and their labor force participation affect each other; and
4. the observed negative relationship is spurious and is caused by common antecendents of both variables.

According to the above mentioned "role incompatibility hypothesis" both women's fertility and their labor force participation affect each other reciprocally because of the strain between the roles of mother and employee. Nothing in this hypothesis suggests causality in one direction rather than the other (Lehrer and Nerlove 1986).

From the point of view of economic theory, fertility and female employment are simultaneously determined by the same basic economic variables. More specifically, female labor market participation and fertility are both choice variables which households choose simultaneously given their exogenous constraints. If both variables fluctuate to some extent synchronously, then – according to the logic of economic theory – this must be caused *entirely* by external variables that determine both variables exogenously. Examples of such external variables are the real female wage, the unemployment rate and – according to recent work by some economists – social norms.

Many researchers would not go as far as economic theory and would argue that at least part of the correlation between fertility and female employment is not determined by external variables. Some of these researchers view fertility and female employment more as the result of a sequential decision process rather than as the result of a simultaneous decision problem. If these variables are indeed the result of a sequential decision process, then it is quite possible that one variable exogenously causes the other variable.

2.2 Macro studies

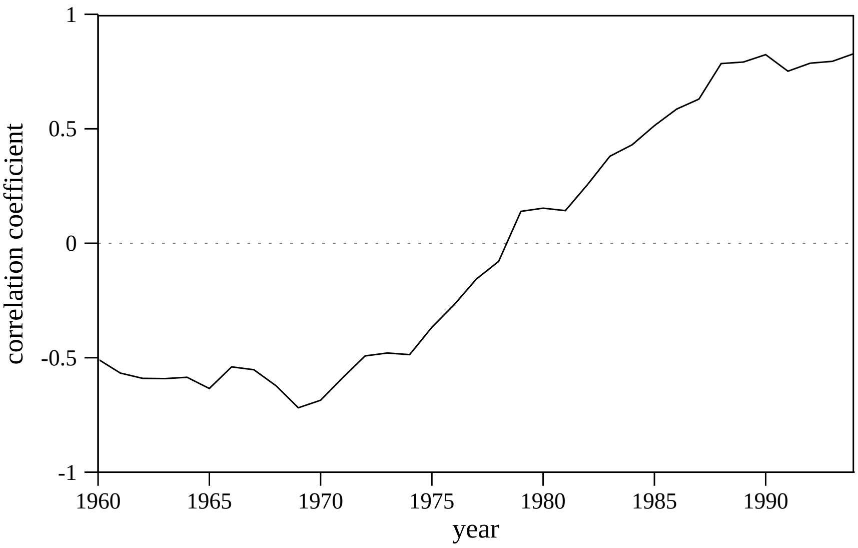
Given the explanations of the fertility/employment nexus on the micro level mentioned above, it is no wonder that previous empirical research has concentrated mainly on micro-level data (for an extensive review of the micro literature see Cramer 1980; Lehrer and Nerlove 1986; Spitze 1988). However, it is very difficult to resolve the fertility/employment question with existing cross-sectional as well as longitudinal survey data since the empirical findings depend on the underlying decision model (sequential or simultaneous) and thereby on the applied statistical model. Furthermore, work intentions may cause actual fertility behavior and fertility intentions may cause actual work behavior. That is, future events may cause present behavior (Bernhardt 1993; Ní Bhrolcháin 1993). Macro-level studies – and especially cross-national comparisons – are an alternative way to resolve the fertility/employment puzzle because they do not require detailed individual-level data (see Rindfuss and Brewster 1996: 262, who also stress the value of a cross-national assessment). However, relationships at the individual and the aggregate level may be different (cf., e.g., Ní Bhrolcháin 1993).

Existing macro studies can be divided into studies which analyze macro-level data on a cross-country basis and studies which apply methods of time series analysis. Various authors (Ahn and Mira 2000; Brewster and Rindfuss 2000; Esping-Andersen 1999; Rindfuss et al. 2000) find that, in OECD countries, the cross-country correlation between the total fertility rate (TFR) and the female labor market participation rate (FLP) turned from a negative value before the 1980s to a positive value thereafter. The countries that now have the lowest levels of fertility are those with relatively low levels of female labor force participation, and the countries with higher fertility levels tend to have relatively high female labor force participation rates. Following the graphical presentation in the literature (e.g., Rindfuss et al. 2000), Figure 1 illustrates this change for 12 selected OECD countries.[1]

Several recent papers have suggested a weakening link between fertility and

[1]Countries included are Belgium, Canada, Finland, France, Italy, Japan, Norway, Sweden, Switzerland, United Kingdom, United States and West-Germany. The data sources can be found in Appendix A.

Figure 1: Correlation between the total fertility rate and female labor force participation rate, 1960-1994



female employment due to greater availability of market child care, family policies (such as state mandated maternity leave) and changing attitudes towards working mothers (Brewster and Rindfuss 2000; Rindfuss et al. 2000; Rindfuss and Brewster 1996). For that reason, they argue that changes in the institutional context at the macro-level must have enabled women in some countries to better combine work and childrearing.

The cross-sectional studies do not, however, explicitly address the causality question. This is done in studies that apply formal Granger causality tests to aggregate time series in different countries (Cheng 1996; Klijzing et al. 1988; Michael 1985; Zimmermann 1985). The basic logic underlying these tests is quite simple: "We say, that [some time series] Y_t is causing X_t if we are better able to predict X_t using all available information than if the information apart from Y_t had been used" (Granger 1969: 428). Causality between two time series can be either unidirectional (X causes Y *or* the reverse), bidirectional (X causes Y *and* the reverse – this is termed feedback) or non existent in which case the two time series move independently or contemporaneously.

Table 1 provides a summary of empirical results of macro studies with time series data. Analyzing German time series from 1960-1979, Zimmermann (1985)

Table 1: Summary of macro-level time series studies

author	method	data	results
Zimmermann (1985)	modified Granger / first differences	German time series 1960-79	$TFR \rightarrow FLP$
Michael (1985)	standard Granger / levels	US time series 1948-80	$TFR \leftarrow FLP$ $FER \rightarrow FLP$
Klijzing et al. (1988)	indirect Granger / first differences	Dutch survey data 1977-84	$BIRTHS \rightarrow LFP$ $(BIRTHS \leftarrow LFP)$
Cheng (1996)	modified Granger / first differences	US time series 1948-93	$TFR \rightarrow FLP$

Notes: TFR – total fertility rate; FLP – female labor force participation rate; FER – age-specific fertility rate; $BIRTHS$ – lagged (by 10 months) number of children born (per month); LFP – percentage of women participating in the labor market (per month).

concludes from a modified Granger-causality test applied to first differences of all variables that increasing female employment does not cause decreasing fertility; rather, the reduction in birth rate causes the increase in female labor force participation. Applying standard Granger-causality tests to the levels of US time series data from 1948-1980, Michael (1985) finds that female labor force participation positively causes fertility and not the other way around. However, this result seems to be sensitive to the definition of fertility. With age-specific fertility rates, Michael finds that fertility negatively affects female labor force participation and not the other way around. Klijzing et al. (1988) use monthly individual data from a Dutch survey for a seven-year period (1977-1984). In a first step they calculate for each month the average number of children of all women in this survey and the percentage of all women in this survey which participate in the labor market. In a second step they apply Sims' indirect Granger-causality test to the first differences of these data. They find that labor force participation has no influence on subsequent fertility decision-making and that fertility decisions do have an impact on female labor force participation. However, with standard Granger tests they find causality in both directions. Cheng (1996) applies a modified version of the Granger-causality method to first differences of aggregate US data for 1948-1993. He finds unidirectional negative causality running from fertility to female employment.

Obviously, the time series literature has not yet come to an agreement on the presence or the direction of causality between fertility and female employment. In our view, this might be due to two issues that have not yet been addressed in the literature. First, the literature has not yet taken into consideration several important recent developments in the econometric time series literature (see Section 4). In particular, Michael (1985) does not consider non-stationarity of the time series, that is, whether the mean and/or the variance of the time series change over time. This is problematic because, if the time series are non-stationary, then there is the possibility of "spurious" causality results. Cheng (1996), Klijzing et al. (1988) and Zimmermann (1985) account for non-stationarity. However, in applying only Granger-causality tests to first differences, they refrain from the use of valuable information about a possible long-run relationship between the variables.[2] As a consequence, the results in these studies might be wrong.

Our paper contains two advances over these earlier attempts to determine causality. First, we use more recent data, which is important because the relationship between TFR and FLP may have shifted in recent years. Second, we employ more sophisticated econometric methods to overcome deficiencies in earlier efforts. Apart from the methodological issues related to the stationarity assumption of the time series, we allow for a further methodological correction. We consider the possibility of parameter instability in the long-run relation between fertility and female employment (as suggested by inspection of Figure 1) and structural breaks in the trend of the variables. Clearly, it would be desirable to include in the regressions socio-demographic variables that caused this change in behavior. However, in our view this would be too complex – if not impossible – and, for that reason, we approximate this change in behavior with the inclusion of dummy variables. The use of such a "minimal" approach is a standard procedure in applied econometrics.

3 Data

We assembled annual time series of the total fertility rate (TFR) and female labor force participation rate (FLP).[3] We assembled annual data from 1960-94 for five

[2] Cheng (1996) applies a unit root test to the OLS residuals and finds absence of cointegration, that is, absence of a long-run relation between fertility and female employment. Given his result, he uses the correct Granger-causality method. However, we have shown that Cheng's result of absence of cointegration is not robust. We applied an error-correction model to the same data and found cointegration. We believe this to be due to the well-known fact that error-correction models are superior to the residual based approach, which tends to reject cointegration too often (see Kremers et al. 1992).

[3] See Appendix A for definition and source of data. In a few time series the values for very few single years were missing. In those cases we calculated the missing value as the average of the

European countries (France (FRA), West-Germany (FRG), Italy (ITA), Sweden (SWE), and the United Kingdom (UK)). Moreover, we were able to assemble annual data from 1948-1995 for the USA. Longer time series are preferable because they increase the power of cointegration tests (Otero and Smith 2000). For the purpose of comparison of the results of the USA with the results of the European countries we also show the results for a truncated time series for the USA that extends over 1960-94. For FLP we choose the female labor force of women of age 15 to 64 divided by the female population of age 15 to 64. Hence, women of age 65 and older are not included in our measure of FLP.

There are two possible measurement problems with our data. First, while the TFR constitutes an age standardized measure, the measure of FLP we apply is not age standardized. Second, our measure of FLP includes women aged 15 to 65. It would be more convincingly to exclude women aged above 50 since their fertility rates are almost zero. To check whether these two possible problems affected our empirical results, we applied the same regressions also to the fertility rate of women of age 20-24 (respectively 25-34) and the female labor force participation rate of women of age 20-24 (respectively 25-34) with annual US time series from 1948-1995.[4] With age specific data one can avoid the two possible problems and check whether age structure changes affect the empirical results.

In Figure 2 and Figure 3 we plot the time series of TFR and FLP for each country for 1960-94. As is well-known, the time series of TFR show for most developed countries a kink in the 1960s. Some researchers argue that this kink represents the diffusion of the use of the contraceptive pill (e.g. Goldin and Katz 2000). Other researchers argue the kink is the result of changing social norms. One can see that Italy is an exceptional case since its fertility decline was very slow in comparison to most countries in the developed world. The Swedish TFR shows a small hill around 1990. The demographic literature offers some explanations for this hill which are, however, outside of the scope of our paper (see, e.g., Andersson 1999, and Hoem 1990, who explain the increase of the Swedish total fertility rate at the end of the 1980s with newly enacted leave and wage compensation policies).

---

year that preceded the missing year and the year that followed the missing year.

[4]We neglected fertility of women of age below 20 and above 34 since the magnitude of their fertility is relatively small. Unfortunately, data on fertility of women of age 25-29 and 29-34 were only available since 1976. Therefore, we restricted attention to the sum of these two age groups. In addition, this data limitation made it impossible to construct an age standardized measure of the female labor force participation rate (which would be an alternative to avoid the first mentioned possible problem).

Figure 2: Time series of the total fertility rate for six countries, 1960-94

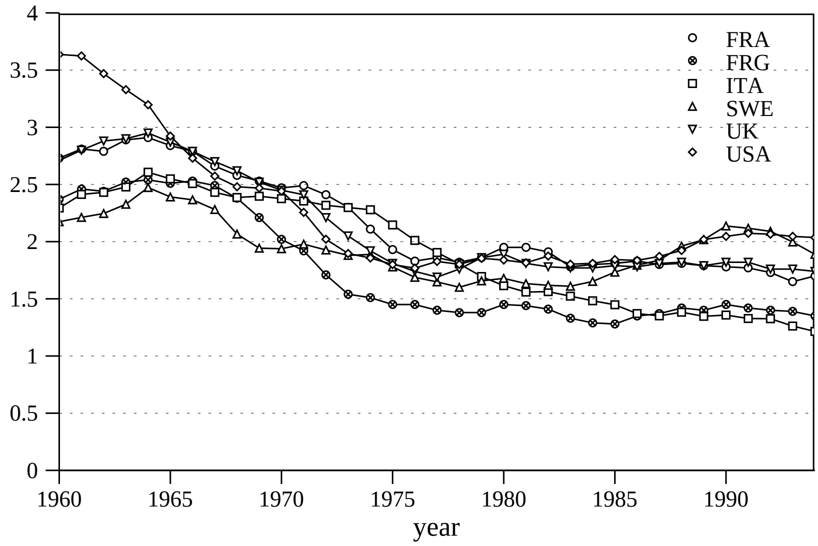
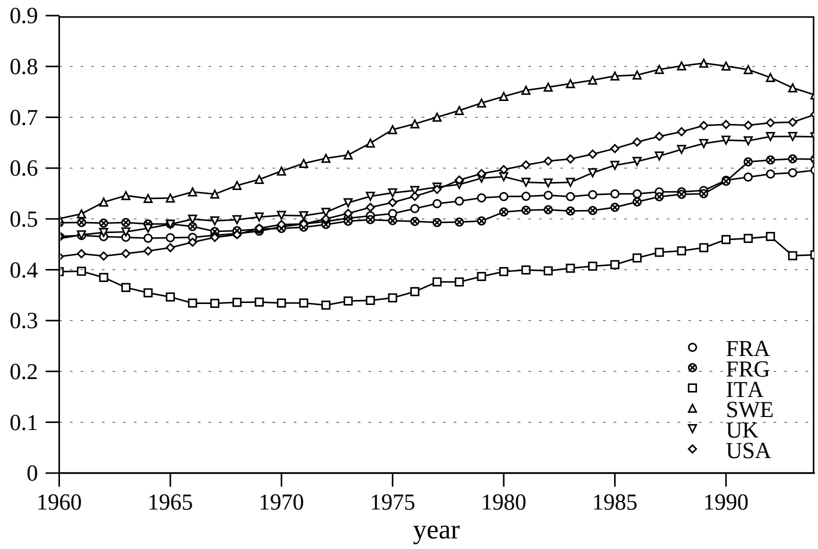


Figure 3: Time series of female labor force participation rate for six countries, 1960-94



As is also well-known, the time series of FLP show a clear upwards trend in most developed countries. Again Italy is an exception. There, the rise of female employment is rather modest. However, high education levels of younger Italian women (not shown) seem to indicate a future change in the FLP even in Italy. The Swedish FLP is distinguished by its relatively higher level over the entire time period from 1960 to 1994.

We have opted to include Italy and Sweden in our set of countries since each of them represents an extreme position in the spectrum of family policies and norms that may influence the fertility/employment relation. The exceptional behavior of the Italian time series is often explained with traditional norms and the view of the family as a private domain in which the government does not intervene with many state services. On the other extreme of a spectrum of family policy and norms is clearly Sweden. The policy in the 'nation of individuals' (Chesnais 1996; Hantrais 1997) tend to be both supportive of women's desires and concerned with children's care. France and the United Kingdom provide a somewhat weaker illustration of 'nations of individuals'. West-Germany as a 'nation of families' shares a strong commitment toward families, backed by monetary allowances for housing, child benefit packages, and well-paid maternal leave. To summarise: The time series in Figure 2 and 3 show that our set of six countries is representative for most developed countries.

4 Econometric methodology

Before introducing the specific method we apply, we shortly review some of the recent econometric time series techniques that need to be taken into account when testing for causality.

Cointegration

When estimating with time series data the relation between two trending variables one often gets *spurious regression* results, that is, a seemingly significant effect even though the variables are actually unrelated. An illustrative example is the fertility rate and the number of storks in a region. Both variables have mostly a downward trend. Due to their trends, when regressing the fertility rate on the number of storks with OLS one might find a significant positive effect even though it would be absurd to argue that such a causal relation exists in reality. Often detrending (that is, including a smooth function of calendar time as a further element on the right hand side of the regression equation) helps to eliminate spurious regression results. But as a recent econometric literature (started by Granger and Newbold 1974) shows, detrending does not help in case the variables are difference-stationary, also labeled I(1). A series is difference-stationary if its mean and its variance are constant over time after first differencing, but not in levels. We found that FLP and TFR series appear to be difference-stationary. Consequently, analyses that attempt to estimate causation and that control for trends by including a function of calendar time on the right side of the equation will not generally give unbiased standard errors for inference of the causal relation.

If two time series are trending, the important question to answer is whether the relation between these trends is "tight" enough so as to warrant the conclusion that either there is a direct causal relationship between them, or that there is some external force that is jointly determining both variables. *Cointegration tests* can be applied to detect whether a relation between two I(1) variables is 'true' or spurious. These tests aim to detect synchronous movements in deviations from the trend of both variables. More technically, cointegration tests aim to detect whether there exists a linear combination of two I(1) variables which is stationary (possibly after detrending). In case such a stationary linear combination exists, the variables are said to be cointegrated and the possibility of a spurious correlation due to trends can be excluded.

Granger-causality

Granger (1969) introduced tests to detect causality between time series (henceforth labeled as *Granger-causality test*). The Granger causality test is typically based on the estimation of a vector autoregressive (VAR) model with variables in levels or in first differences:

$$\Delta Y_t = \alpha_0 + \sum_{i=1}^{m_Y} \alpha_{Y,i} \Delta Y_{t-i} + \sum_{j=1}^{n_Y} \beta_{Y,j} \Delta X_{t-j} + u_{Y,t}, \qquad (1)$$

$$\Delta X_t = \beta_0 + \sum_{k=1}^{m_X} \alpha_{X,k} \Delta Y_{t-k} + \sum_{l=1}^{n_X} \beta_{X,l} \Delta X_{t-l} + u_{X,t}, \qquad (2)$$

where $\Delta Z_t = Z_t - Z_{t-1}, \forall Z = Y, X$, with all variables being in logarithms. The α's and β's are constants, the m's and n's are the optimal numbers of lags of the series Y and X and the u_t's are serially uncorrelated random disturbances with zero mean.[5] For given values of the lag lengths (i.e, the m's and n's) it can, for

[5]Henceforth the u_t's denote any disturbance term with zero mean.

example, be tested, whether X *Granger-causes* Y by testing the hypothesis H_0: $\beta_{Y,1} = \beta_{Y,2} = ... = \beta_{Y,n_Y} = 0$ against the alternative H_a: not H_0. The joint significance of the coefficients can be tested by a Wald test with the F-statistic.

Engle and Granger (1987) have shown that the Granger-causality test can be seriously wrong if the time series are I(1) and cointegrated. For that reason, the literature suggests to test first (with so-called unit root tests) whether the variables are I(1). If the variables turn out to be I(1), the recent literature suggest then to use a single error-correction model (ECM) to test for cointegration and causality upon application of, e.g., a method of Boswijk (1994).

Testing for cointegration in a single error-correction model

When trends are cointegrated, it follows that there is a long-run relation between them. This long-run relation can be expressed in terms of what is called an error-correction model. An error-correction model explains changes in one series in terms of three components: (1) random shocks, (2) lagged changes of the two series, and (3) an "error-correction" term that pushes the series back into its long-run relation with the other series. The significance of the parameter for the error-correction term amounts to a statistical test for the existence of a long-run causal relation between the two series. We implement this error-correction model below and we further allow for this long-run relation to change historically, in order to test the hypothesis that the employment-fertility conflict has changed in recent years at least for some countries in our sample. An ECM can be tested with OLS and is written in the following form:[6]

$$\Delta Y_t = \alpha_0 + \lambda_Y Y_{t-1} + \psi_Y X_{t-1} + \varphi_Y t + \sum_{i=1}^{m_Y} \alpha_{Y,i} \Delta Y_{t-i} + \sum_{j=1}^{n_Y} \beta_{Y,j} \Delta X_{t-j} + u_{Y,t}, \tag{3}$$

where the α_Y's, β_Y's, α_0, λ_Y, ψ_Y and φ_Y are constants, t is a trend term and m_Y and n_Y are again the optimal number of lags of the first differences (we choose m_Y and n_Y so that the Schwarz criterion was minimized, but limited m_Y and n_Y to a maximum of 2). The purpose of the lags is to correct for autocorrelation in the random disturbance term. Henceforth we include the trend term in the ECM only if its coefficient is significant according to the t-statistic (the t-statistic is reliable if Y and X are cointegrated).

[6] An analogous equation to (3) exist for ΔX_t as the dependent variable.

Boswijk shows that one can test in (3) for cointegration between Y and X with a joint Wald test of the hypothesis H_0: $\lambda_Y = \psi_Y = 0$. In this Wald test one has to compare the χ^2-statistic that results from a test of H_0 with the critical values tabulated in Boswijk in Table B.1-B.5 (the tables are for various cases, that is, with and without trend term in (3) and so forth). Cointegration can not be rejected if the χ^2-statistic is larger than the relevant critical value.

If there is absence of cointegration, then it is appropriate to test for causality upon application of the standard Granger-causality test in first differences, that is, to test for joint significance of the lags in (1)-(2).

Testing for causality in a single error-correction model

In case of cointegration, absence of weak exogeneity of X for Y implies long-run causality of X for Y (Hall and Milne 1994). A test for weak exogeneity in an ECM can be applied in two stages (Boswijk 1994). In the first stage one may estimate the following equation with OLS:

$$Y_t = \gamma_Y + \delta_Y X_t + \theta_Y t + u_{Y,t}, \tag{4}$$

$$X_t = \gamma_X + \delta_X Y_t + \theta_X t + u_{X,t}, \tag{5}$$

where the γ's, δ's and θ's are constants. The relation in (4) and (5) is labeled as long-run relation between Y and X. Note that this relationship involves three elements: a slope coefficient, a time trend, and a random error. This relation is "long-term" in the sense that $Y_t - \delta_Y X_t$ is stationary once a smooth function of time (in our case, a linear time trend) is controlled for. The crucial point here is that a specific linear combination of these variables obeys an equilibrium relation in the long run. The residuals of the estimates of (4) (that is, $\hat{\mu}_{Y,t} = Y_t - \hat{\gamma}_Y - \hat{\delta}_Y X_t - \hat{\theta}_Y t$ – where a hat on top of a variable denotes the estimate of the corresponding parameter) and (5) (that is, $\hat{\mu}_{X,t} = X_t - \hat{\gamma}_X - \hat{\delta}_X Y_t - \hat{\theta}_X t$) are saved and labeled as error-correction terms. In the second stage the following equations are tested with OLS:

$$\Delta Y_t = \alpha_0 + \lambda_Y \hat{\mu}_{Y,t-1} + \sum_{i=1}^{m_Y} \alpha_{Y,i} \Delta Y_{t-i} + \sum_{j=1}^{n_Y} \beta_{Y,j} \Delta X_{t-j} + u_{Y,t}, \tag{6}$$

$$\Delta X_t = \beta_0 + \lambda_X \hat{\mu}_{X,t-1} + \sum_{k=1}^{m_X} \alpha_{X,k} \Delta Y_{t-k} + \sum_{l=1}^{n_X} \beta_{X,l} \Delta X_{t-l} + u_{X,t}, \qquad (7)$$

where the α's and λ's are constants, the $\hat{\mu}_{t-1}$'s denote the lagged error-correction term, the m's and n's are again the optimal numbers of lags of the first differences chosen according to the Schwarz criterion. The variable X (resp. Y) is *weakly exogenous* for Y (resp. X), if λ_Y (resp. λ_X) is insignificant according to the t-statistic. Otherwise *long-run causality* of X for Y cannot be rejected. Intuitively, long-run causality implies that a deviation in the long-run relation between X and Y, that is, $\hat{\mu}_{Y,t} \neq 0$ for some t, will impact on the value of ΔY_t (and analogously for $\hat{\mu}_{X,t} \neq 0$).

If λ_Y as well as λ_X are significant – which means that X long-run causes Y as well as Y long-run causes X –, then one can interpret this as evidence for that there exists a Z vector containing exogenous variables which is common to Y and X and which "long-run causes" the cointegration relation of Y and X.

Contrary to long-run causality, *short-run causality* of X for Y prevails if in (6) (resp. in (7)) the lags of ΔX (resp. of ΔY) are jointly significant.[7] However, in the subsequent analysis we only test for long-run causality (henceforth just labeled "causality"). This also agrees with Granger (1988) who questions the usefulness of the concept of short-run causality for difference-stationary time series.[8]

In addition to merely testing for causality between two time series one may also test for parameter instability in the long run relation between TFR and FLP.

Testing for parameter instability in the long-run relation between two variables within a single error-correction model

Parameter instability occurs when the long-run relation between X and Y changes. We allow for two forms of parameter instability in our model. First, the parametric relationship between X and Y (as represented by the coefficient for X) can change. Second, the linear time trend can change. To test for parameter instability one may estimate the following equation with non-linear least squares (NLS) (the estimation procedure – but not the test for parameter instability – follows Boswijk 1994):

[7] X is termed *strongly exogenous* for Y, if there is absence of short-run causality and long-run causality of X for Y and, in addition, $u_{Y,t}$ in (6) and $u_{X,t}$ in (7) are uncorrelated with each other (and analogously for Y in relation to X).

[8] In case of stationary time series one can apply system (1)-(2) to test for causality without the need to distinguish between short-run and long-run causality.

$$
\begin{aligned}
\Delta Y_t &= \kappa(Y_{t-1} - \gamma_Y - \delta_{Y,1} X_{t-1} - \delta_{Y,2}\, DUM_{t_B}\, X_{t-1} - \theta_Y t) \quad (8) \\
&\quad + \sum_{i=1}^{m_Y} \alpha_{Y,i} \Delta Y_{t-i} + \sum_{j=1}^{n_Y} \beta_{Y,j} \Delta X_{t-j} + u_{Y,t}, \\
with\ DUM_{t_B} &= \begin{cases} 1 & if\ t > t_B \\ 0 & otherwise \end{cases},
\end{aligned}
$$

where $\kappa, \gamma_Y, \delta_{Y,1}, \delta_{Y,2}, \theta_Y$ are constants. t_B denotes the possible date of a structural break in the slope parameter of X. The hypothesis of a break in this slope parameter can not be rejected if $\delta_{Y,2}$ is significant according to the t-statistic. Inspired by a literature that suggests methods to test whether a time series is difference-stationary (see Maddala and Kim 1998), we suggest to choose the date of a possible break in the slope endogenously. More specific, we suggest to estimate (8) for various values of t_B and compare the corresponding t-statistics for testing $\delta_{Y,2} = 0$. The value of t_B for which the absolute value of the t-statistic is maximized may be chosen as the date of a possible break in the slope. In case a break in the slope can not be rejected, we suggest to test the sign and the significance of the correlation between X and Y before the break and after it by testing the following equation with NLS:[9]

$$
\begin{aligned}
\Delta Y_t &= \kappa[Y_{t-1} - \gamma_Y - \delta_{Y,1}(DUM_0 - DUM_{t_B}) X_{t-1} \quad (9) \\
&\quad - \delta_{Y,2}(DUM_{t_B} - DUM_T)\, X_{t-1} - \theta_Y t] \\
&\quad + \sum_{i=1}^{m_Y} \alpha_{Y,i} \Delta Y_{t-i} + \sum_{j=1}^{n_Y} \beta_{Y,j} \Delta X_{t-j} + u_{Y,t}, \\
where\ \frac{\partial Y_{t-1}}{\partial X_{t-1}} &\mid_{long-run} = \begin{cases} \delta_{Y,1} & for\ t < t_B \\ \delta_{Y,2} & otherwise \end{cases},
\end{aligned}
$$

and $DUM_0 = 1,\ \forall t \epsilon [0, T]$ and $DUM_T = 0,\ \forall t \epsilon [0, T]$, with the subscript 0 and respectively T as the dates at which the sample starts and respectively ends. The

[9]Equation (9) is nothing else than equation (8) with the dummies rearranged. This rearrangement does not change the content of the estimation results, but – in our view – simplifies the interpretation of the results. To give an example, in (9) the coefficient in front of the term $(DUM_0 - DUM_{t_B})X_{t-1}$ represents the coefficient of the slope before the break (that is, from the date of the start of the sample to the date of the break).

(long-run) correlation between X and Y is significant before the break (resp. after the break) if $\delta_{Y,1} = 0$ (resp. $\delta_{Y,2} = 0$) is rejected according to the t-statistic.

We have, for didactical reasons, deferred the exposition on how to test for parameter instability to the end of this section. However, in practice one needs to test for parameter instability in the first place, in order not to distort the results of the cointegration and causality tests.

5 Empirical application

Prior to any test of causality we applied the Augmented Dickey-Fuller unit root test to the time series of TFR and FLP. These tests showed that each time series is difference-stationary. Therefore, we rule out the use of time series models that attempt to estimate causal relation by simply regressing one series on the other series, controlling for a time trend. Instead, an error-correction model is called for. Therefore, we estimated equation (8) for (i) ΔTFR_t and alternatively (ii) ΔFLP_t as the dependent variable. We found a significant trend in the long-run relation between TFR and FLP for each country when we allow for breaks in the trend. Therefore, in the first step we determined the dates and the number of breaks in the trend for each equation. To determine the exact date of a break in the trend, we followed the procedure for determining endogenously the break in the slope as was explained in the last section. That is, we compared the t-statistics of dummies for various possible dates of the break and chose the date with the highest absolute value of the t-statistic. Following Kim (1997), we chose the number of breaks in the trend that minimized the Schwarz criterion. According to our estimates, there were breaks in the trends in the long-run relation, with the timing of the breaks varying somewhat by country. In the next step we tested for instability in the long-run relation between TFR and FLP and tested for an endogenous break in the slope parameter as explained in the previous section.

Our estimates of the stability of the slope coefficient that links TFR and FLP into a cointegrated trend has changed once for the countries under investigation. The statistically optimal estimate for this break varies somewhat, depending upon whether we model the long-run relation as an effect of TFR on FLP or as an effect of FLP on TFR. We report the estimates of these break points in Table 2. In our view, the important issue here is not the exact timing of the break (we do not actually believe that the break occurred at a single distinct point in history). Rather, the important point is that, at least for some of the countries in our analysis, the long-run relation between TFR and FLP has changed in recent history. Table

2 summarizes the dates of breaks in the slope and suggests that for each country the correlation between TFR and FLP has changed across time.

Table 2: Endogenous dates of break in slope in time series regressions for six countries, 1960-94, and additionally for the US, 1948-95 and US age specific time series.

	Dependent Variable	
Country	ΔTFR_t	ΔFLP_t
FRA	1976	1982
FRG	1985	1991
ITA	1981	1980
SWE	1970	1973
UK	1977	1986
USA	1970	1984
USA (1948-95)	1970	1977
USA20-24 (1948-95)	1971	1981
USA25-34 (1948-95)	1970	1967

Note: USA20-24 (respectively USA25-34) indicates that we apply age specific time series of fertility and female labor force participation for women of age 20-24 (respectively 25-34).

Next we tested for cointegration between TFR and FLP by estimating equation (3) for (i) ΔTFR_t and alternatively (ii) ΔFLP_t as the dependent variable (we include the breaks as determined in the step before). Table 3 summarizes the cointegration test results.[10]

Table 3 shows cointegration between TFR and FLP (mostly at the 1% significance level) for all countries with the only exception being the case of Sweden when ΔTFR is the dependent variable. Hence, there is strong evidence for absence of spurious regression results for the association between TFR and FLP in aggregate time series data.

In order to test for the direction of causality between TFR and FLP we used the two stage procedure as explained in the previous section. For illustration, assume there is one break in the slope and in the trend of each time series. In this

[10] As suggested by the results in Table 2 the coefficient ψ_Y in equation (3) had in fact to be split up into two coefficients $\psi_{Y,1}$ and $\psi_{Y,2}$ with the former (latter) one referring to periods before (after) the structural break. Therefore, we tested in (3) for cointegration upon application of a Wald test of the joint hypothesis: H_0: $\lambda_Y = \psi_{Y,1} = \psi_{Y_Y 2} = 0$. Alternatively, in case of insignificance of the slope after the break, one could estimate $H_0' : \lambda_Y = \psi_{Y,1} = 0$. However, the results of testing H_0' are the same as the results of testing H_0 and are therefore not shown in the paper.

Table 3: Test for cointegration between TFR and FLP in time series regressions for six countries, 1960-94, and additionally for the US, 1948-95 and US age specific time series.

Country	Dependent Variable ΔTFR_t	ΔFLP_t
FRA	yes***	yes***
FRG	yes**	yes***
ITA	yes***	yes***
SWE	no	yes***
UK	yes***	yes***
USA	yes**	yes***
USA (1948-95)	yes***	yes***
USA20-24 (1948-95)	yes***	yes***
USA25-34(1948-95)	yes***	yes***

Notes: *, **, *** 15%, 5%, 1% significance level. "yes" means rejection of H_0: no cointegration.

case we estimated in a first stage the following equations with OLS:[11]

$$\begin{aligned} TFR_t \;=\; & \gamma_{TFR}+\delta_{TFR,1}(DUM_0-DUM_{t_{B,TFR}})FLP_t + \\ & \delta_{TFR,2}(DUM_{t_{B,TFR}}-DUM_T)FLP_t + \\ & \theta_{TFR,1}(DUM_0-DUM_{t_{\tilde{B},TFR}})(t/100)+ \\ & \theta_{TFR,2}(DUM_{t_{\tilde{B},TFR}}-DUM_T)(t/100), \end{aligned} \qquad (10)$$

$$\begin{aligned} FLP_t \;=\; & \gamma_{FLP}+\delta_{FLP,1}(DUM_0-DUM_{t_{B,FLP}})TFR_t + \\ & \delta_{FLP,2}(DUM_{t_{B,FLP}}-DUM_T)TFR_t + \\ & \theta_{FLP,1}(DUM_0-DUM_{t_{\tilde{B},FLP}})(t/100)+ \\ & \theta_{FLP,2}(DUM_{t_{\tilde{B},FLP}}-DUM_T)(t/100)\ , \end{aligned} \qquad (11)$$

where the γ's, δ's and θ's are constants, $DUM_{t_{B,i}}$, $\forall i = TFR, FLP$ denotes the break in the slope, $DUM_{t_{\tilde{B},i}}$, $\forall i = TFR, FLP$ denotes the break in the trend with the subscript $t_{B,i}$ and $t_{\tilde{B},i}$ indicating the date of the break in the slope and the

[11] In these equations the dummies for the trend are rearranged in the same way as are the dummies of the slope in (9). E.g. in (10) the coefficient in front of the term $(DUM_0 - DUM_{t_{\tilde{B},TFR}})(t/100)$ represents the coefficient of the trend between the start of the sample and the date of the break in the trend $t_{\tilde{B},TFR}$.

trend respectively. As was the case in the cointegration tests, we included in the estimations of equations (10) and (11) the breaks which were significant according to the estimations of (8). The residuals that resulted from estimation of (10) (resp. (11)) were saved and used in the estimation of (6) (resp. (7)) with ΔTFR_t (resp. ΔFLP_t) as the dependent variable. Table 4 summarizes the results of the tests for causality.

Table 4: Testing for causality between FLP and TFR for six countries, 1960-94, and additionally for the US, 1948-95 and US age specific time series.

Country	$FLP \rightarrow TFR$	$TFR \rightarrow FLP$
FRA	yes***	yes**
FRG	yes**	no
ITA	yes*	yes**
SWE	no	no
UK	yes**	yes***
USA	no	yes**
USA (1948-95)	yes***	yes**
USA20-24 (1948-95)	yes***	yes***
USA25-34 (1948-95)	yes***	yes*

Notes: *, **, *** 15%, 5%, 1% significance level. “yes” means rejection of H_0: weak exogeneity.

The table shows clear evidence of causality in both directions for almost all cases. The only country were we found no causality at all is Sweden. Note that in case of Sweden when ΔTFR is the dependent variable we applied the standard Granger-causality test in first differences because in this case cointegration was rejected.

In our view Table 4 is, by and large, consistent with simultaneous movements of both variables brought about by common exogenous third variables. These third variables are presumably financial incentives, social norms and/or social institutions which help to combine work and family.

Table 5 summarizes the findings from the estimates of the long-run relation in (9), that is, – for our illustrative example with one break in the trend and for the case with TFR as the dependent variable:

$$\begin{aligned} TFR_{t-1} &= \hat{\gamma}_{TFR} + \hat{\delta}_{TFR,1}(DUM_0 - DUM_{t_{B,TFR}})FLP_{t-1} \\ &\quad + \hat{\delta}_{TFR,2}(DUM_{t_{B,TFR}} - DUM_T)FLP_{t-1} \\ &\quad + \hat{\theta}_{TFR,1}\,(DUM_0 - DUM_{t_{\tilde{B},TFR}})(t/100) \\ &\quad + \hat{\theta}_{TFR,2}(DUM_{t_{\tilde{B},TFR}} - DUM_T)(t/100) \end{aligned} \tag{12}$$

(and similar for the case with FLP as the dependent variable). The quantitative estimation results are shown in Appendix B.

Table 5: Relation between TFR and FLP after the break in the slope for six countries, 1960-94, and additionally for the US, 1948-95 and US age specific time series.

	ΔTFR_t			ΔFLP_t		
Country	–	0	+	–	0	+
FRA		x			x	
FRG		x				x
ITA	x			x		
SWE					x	
UK		x			x	
USA		x			x	
USA (1948-95)		x			x	
USA20-24 (1948-95)		x			x	
USA25-34 (1948-95)		x			x	

Notes: – indicates that the relation between TFR and FLP after the break is significant negative and its magnitude exceeds the one before the break; 0 indicates that the relation between TFR and FLP is either significant negative and weaker as compared to the period before the break, or insignificant; + indicates a significant and positive relation between TFR and FLP after the break. The significance level is in any case the 15% level. For each country we indicate with a cross the valid case.

Our estimation results (Appendix B) show a significant negative correlation between TFR and FLP before the break in the slope for all countries in all equations with the exception of Sweden in case with ΔTFR as the dependent variable. Furthermore, for all countries the magnitude of this relation in the equation with ΔTFR as the dependent variable exceeds the one in the equation with ΔFLP as the dependent variable. Table 5 shows that in almost all cases (that is, for all countries almost always no matter whether ΔTFR or ΔFLP is the dependent variable) the correlation between TFR and FLP becomes smaller in magnitude or even insignificant after the break. However, the opposite holds for Italy where the negative relation between TFR and FLP becomes even stronger after the break. Note that in case of Sweden when ΔTFR is the dependent variable we found no cointegration and according to econometric theory there is no long-run relation when there is no cointegration.

To conclude, we find for all non-Mediterranean countries evidence for a significant negative relation between TFR and FLP before the break in the slope and an insignificant or significant negative, but weaker relation after the break. This finding is consistent with the view in Rindfuss and Brewster (1996), Rindfuss et al. (2000), and Brewster and Rindfuss (2000) that changes in childcare availability and attitudes towards working mothers might have reduced the incompatibility between childrearing and female employment. Moreover, our empirical results show that for Italy the relation between TFR and FLP did neither become insignificant nor weaker – actually it became even bigger in magnitude as compared to the period before the break. This latter result is consistent with the view in Brewster and Rindfuss (2000) and Rindfuss et al. (2000) that in the Mediterranean countries there were not such changes of family policies and/or social norms which reduced the incompatibility between childrearing and employment for women.[12]

A glance at the time plot of the Italian FLP might even give an explanation for our findings. The time series clearly reveals that in Italy the FLP for women fell from 1960 to 1965. It might be that in the year 1960 (and possibly earlier) in Italy men's wage income was not enough to support the family and so often women had also to participate in the labor market, but could not restrict their fertility (for example, because at that time the contraceptive pill was not introduced). As in the early 1960s the wage income of the men grew, often women had not anymore to work to support the family and, as a consequence, reduced their employment. Clearly, such a mechanism would reduce the correlation between TFR and FLP in the initial periods of our sample.

6 Conclusion

In this study we have applied recent econometric time series techniques to test for causality between fertility and female employment with macro-level time series

[12]We would expect for other Mediterranean countries (such as Greece, Spain and Portugal) the empirical results to be similar to the result for Italy.

data from six developed countries. Compared to previous research we introduced two new methodological elements: (i) we allowed for parameter instability, and (ii) we used an error-correction model.

The existing literature mostly found unidirectional causation and conflicting results on the direction of causality between fertility and female employment. In light of our empirical results – which show causality in both directions – we suggest that previous research tended to reject causality too often. The failure to account for parameter instability (either in the long-run relation between fertility and female employment and/or the trend of each time series) may partly explain why this was the case (see Table 1). Moreover, most previous research either ignored difference-stationarity or applied Granger-causality tests to the first differences of the time series without testing for cointegration.

Our conclusion that there is causation in both directions indicates that ignoring difference-stationarity probably did not lead to wrong causality results since this failure tends to find "spurious" causality rather than to reject causality too often (indeed Michael 1985, who applies Granger-causality tests to the levels and ignores difference-stationarity as well as parameter instability, finds some evidence for causality in both directions). Nevertheless, in the case of difference-stationary time series, an error-correction model is by far the most appropriate model, and testing for cointegration is imperative in this case.

The application of a Granger-causality test to the first differences of the time series is a special case of the application of an error-correction model. More specifically, it is a Granger-causality test without the inclusion of the error-correction term. Not including the error-correction term might lead to serious problems, however, because it contains the information about a possible long-run relation between fertility and female employment. For that reason, Granger-causality tests applied to first differences can only test for "instantaneous causation", and the information on long-run causation is lost in this framework. From a statistical point of view, this loss of information might be an additional explanation why the existing literature has tended to reject causality too often. From a substantive point of view, if the correlation between fertility and female employment is the result of intentions (as, for example, economists would argue), then testing for instantaneous causation is not useful. Clearly, if individuals "decide" about their number of children, their timing of childbearing and their labor market participation, then a response to shocks takes time, which makes a test of long-run causality seem much more useful.

Allowing for parameter instability not only prevented us from rejecting causality too often. Omission of a break in the trend can also lead to biased estimates of the coefficient of the relationship between fertility and female employment. Obviously, the same is true for the failure to account for instability in the correlation between fertility and female employment. Michael (1985), for example, finds a positive effect from female employment on fertility with macro-level data. We also found in our empirical applications a positive coefficient for some countries when we did not account for structural breaks (the results are not shown). However, when we accounted for parameter instability, we found for all countries a negative coefficient before the break. Empirical results that show a positive correlation between fertility and female employment already in the 1960s and the 1970s might – in our view – indicate biased estimates due to the failure to account for parameter instability (it is possible that for Nordic countries the situation is different).

Besides the improvements in the estimation results, accounting for parameter instability is also an interesting finding in its own right. Most importantly, allowing for instability in the correlation between fertility and female employment revealed that, for all non-Mediterranean countries, the correlation became weaker over time. This result with time series data complements and supports recent studies which found for OECD countries that the cross-country correlation between fertility and female employment turned from a negative value before the 1980s to a positive value thereafter. Our empirical finding thus supports the view expressed in recent demographic literature which argues that societal level responses reduced the incompatibility between fertility and female employment.

While the latter studies as well as our study do not distinguish between full and part-time employment, the availability of part-time employment is clearly a further element of societal level responses which might also have reduced the incompatibility between childrearing and female labor market participation. Since more data about the availability of part-time employment will become accessible in the future, we suggest for future work that one should estimate the contribution of availability of part-time employment to the weakening association between fertility and female employment.

To illustrate the importance of applying more sophisticated econometric methods, as suggested in our paper, we conclude by showing the results of a simple OLS regression of TFR on FLP with a trend and a constant included on the right hand side.[13] We applied this regression to all six countries in our sample and for the time interval 1960-1994.

[13]Results from simple OLS regressions of FLP on TFR with a trend and a constant included on the right hand side are very similar.

Table 6: OLS regressions of TFR on FLP

Country	Coefficient of FLP	P-value of coefficient of FLP
FRA	–2.47	0.00
FRG	2.38	0.00
ITA	–0.85	0.00
SWE	–1.73	0.00
UK	–0.40	0.73
USA	–4.11	0.00

Note: All regressions include a constant and a trend.

Although for five of our six countries the OLS coefficients on FLP are negative (cf. Table 6), a closer look at the level and the significance of the estimates reveals several counterintuitive results. The positive coefficient on FLP for West-Germany seems clearly to go against the general predictions of demographic theories. Though recent demographic findings suggest a reduction in the incompatibility between childrearing and female employment, none of these theories actually predicts a positive causal effect of FLP on TFR. Comparing the estimates for Italy, Sweden, and the UK, the latter has the lowest level of incompatibility between childrearing and female employment (where the coefficient on FLP is insignificant) followed by Italy, with Sweden ranking last. Again, these results go against the general findings in the demographic literature so far, where Sweden is always regarded as the country with the highest level of compatibility between childrearing and female employment. Finally, according to the OLS estimates, the USA turns out to be the country with by far the highest level of incompatibility between childrearing and female employment in our sample. Again, one would rather expect Italy to be a country that ranks lowest here. In conclusion, the empirical results in our paper are much more intuitive than those of Table 6 and are supported by the view expressed in the recent demographic literature which argues that societal level responses reduced the incompatibility between fertility and female employment. The importance of applying more sophisticated econometric methods that go beyond simple OLS regressions is therefore evident.

References

Ahn N. and P. Mira. 2000. “A note on the changing relationship between fertility and female employment rates in developed countries.” *Journal of Population Economics*, forthcoming.

Andersson, G. 1999. “Trends in childbearing and nuptiality in Sweden, 1961-1997.” Paper presented at the Annual Meeting of the Population association of America, New York.

Bernhardt, E.M. 1993. “Fertility and employment.” *European Sociological Review* 9: 25-42.

Brewster, K.L. and R.R. Rindfuss. 2000. “Fertility and women’s employment in industrialized nations.” *Annual Review of Sociology* 26: 271-296.

Boswijk, H.P. 1994. “Testing for an unstable root in conditional and structural error correction models.” *Journal of Econometrics* 63: 37-60.

Cheng, B.S. 1996. “An investigation of cointegration and causality between fertility and female labour force participation.” *Applied Economics Letters* 3: 29-32.

Chesnais, J.C. 1996. “Fertility, family, and social policy in contemporary western Europe.” *Population Development Review* 22: 729-739.

Cramer, J.C. 1980. “Fertility and female employment: Problems of causal direction.” *American Sociological Review* 45: 167-190.

Engle, R.F. and C.W.J. Granger. 1987. “Co-integration and error-correction: representation, estimation and testing.” *Econometrica* 55: 251-276.

Esping-Andersen, G. 1999. *Social foundations of postindustrial economies.* Oxford: Oxford University Press.

Goldin C. and L.F. Katz. 2000. “The power of the pill: Oral contraceptives and women’s career and marriage decisions.” National Bureau of Economic Research working paper no. 7527, Cambridge MA.

Granger, C.W.J. 1969. “Investigating causal relations by econometric models and cross-spectral methods.” *Econometrica* 37: 424-438.

Granger, C.W.J. 1988. “Some recent developments in a concept of causality.” *Journal of Econometrics* 39: 199-211.

Granger, C.W.J. and P. Newbold. 1974. "Spurious regression in econometrics", *Journal of Econometrics* 26: 1045-1066.

Hall, S.G. and A. Milne. 1994. "The relevance of p-star analysis to UK monetary policy." *Economic Journal* 104: 597-604.

Hantrais, L. 1997. "Exploring relationships between social policy and changing family forms within the European Union." *European Journal of Population* 13: 339-379.

Hoem, J. 1990. "Social policy and recent fertility change in Sweden." *Population and Development Review* 16: 735-748.

Kim, I.-M. 1997. "Detecting the number of structural breaks", *Economics Letters* 57: 145-148.

Klijzing, E., J. Sieger, N. Keilman and L. Groot. 1988. "Static versus dynamic analysis of the interaction between female labour force participation and fertility." *European Journal of Population* 4: 97-116.

Kremers, J., N.R. Ericsson and J.D. Juan. 1992. "The power of cointegration tests." *Oxford Bulletin of Economics & Statistics* 54: 325-48.

Lehrer, E. and M. Nerlove. 1986. "Female labor force behavior and fertility in the United States." *Annual Review of Sociology* 12: 181-204.

Maddala, G.S. and I.-M. Kim. 1998. *Unit roots, cointegration, and structural change.* Cambridge, UK: Cambridge University Press.

Michael, R.T. 1985. "Consequences of the rise in female labor force participation rates: questions and probes." *Journal of Labor Economics* 3: 117-146.

Ní Bhrolcháin, M. 1993. "How to anser the fertility/employment question (and how not to)." In: Máire Ní Bhrolcháin (ed.), New perspectives on fertility in Britain. *Studies on Medical and Population Subjects* 55: 151-157.

Otero, J. and J. Smith. 2000. "Testing for cointegration: Power versus frequency of observation-further Monte Carlo results." *Economics Letters* 67: 5-9.

Rindfuss, R.R., K. Benjamin and S.P. Morgan. 2000. "The changing institutional context of low fertility." Working paper, University of North Carolina.

Rindfuss, R.R. and K.L. Brewster. 1996. "Childrearing and fertility." *Population and Development Review* 22 (Supplement): 258-289.

Spitze, G. 1988. “Women’s employment and family relations: A review.” *Journal of Marriage and the Family* 50: 595-618.

Weller, R.H. 1977. “Wife’s employment and cumulative family size in the United States, 1970 and 1960.” *Demography* 14: 43-65.

Zimmermann, K.F. 1985. *Familienökonomie.* Berlin: Springer.

Appendix A: The data

1. TFR: Total fertility rate.

   Definition: Sum of age specific fertility rates (in the estimations of section 5 always in logarithm).

   Data sources: German Federal Statistical Office, section population and labor force, Fachserie 1, Reihe 1 (for West-Germany), New Cronos 1998, Eurostat Database, CD-Rom (for France), Recent Trends in Demographic Statistics, Council of Europe (for UK 1960-82) and UK-Office for National Statistics (ONS), Population Trends (for UK 1983-94), U.S. Bureau of the Census, International Data Base (for Belgium 1989-94) and UN Demographic Yearbook 1948-1997 (2000), CD-Rom (for the remaining countries and Belgium 1960-88).

2. FLP: Female labor force participation in female population of age 15 to 64.

   Definition: Female labor force of women of age 15 to 64 including unemployed women of that age divided by female population of age 15 to 64 (in the estimations of section 5 always in logarithm).

   Data source: Comparative welfare states data set http://www.lis.ceps.lu/ (except West-Germany after 1989 and USA where the sources are: German Federal Statistical Office, microcensus and US Bureau of Labor Statistics http://stats.bls.gov/).

Appendix B: Estimation results of long-run relation[14]

FRA

$$TFR_{t-1} = -0.19 - \underset{(-8.42)}{1.66^{***}}(DUM_{59} - DUM_{76})FLP_{t-1} - \underset{(-5.41)}{1.32^{***}}(DUM_{76} - DUM_{94})FLP_{t-1} - \underset{(-3.30)}{0.55^{***}}(DUM_{59} - DUM_{73})(t/100)$$
$$\underset{(-4.67)}{-1.01^{***}}(DUM_{73} - DUM_{76})(t/100),$$

$$FLP_{t-1} = -0.68 - \underset{(-8.34)}{0.17^{***}}(DUM_{59} - DUM_{82})TFR_{t-1} - \underset{(-1.15)}{0.05}(DUM_{82} - DUM_{94})TFR_{t-1} + \underset{(9.64)}{0.59^{***}}(DUM_{59} - DUM_{82})(t/100)$$
$$\underset{(4.97)}{+0.32^{***}}(DUM_{82} - DUM_{89})(t/100) + \underset{(7.94)}{0.43^{***}}(DUM_{82} - DUM_{94})(t/100).$$

FRG

$$TFR_{t-1} = -0.05 - \underset{(-2.26)}{1.30^{**}}(DUM_{59} - DUM_{85})FLP_{t-1} - \underset{(-0.96)}{0.69}(DUM_{85} - DUM_{94})FLP_{t-1} - \underset{(-8.69)}{1.67^{***}}(DUM_{68} - DUM_{85})(t/100),$$

$$FLP_{t-1} = -0.68 - \underset{(-2.49)}{0.03^{***}}(DUM_{59} - DUM_{91})TFR_{t-1} + \underset{(1.64)}{0.11^{*}}(DUM_{91} - DUM_{94})TFR_{t-1} - \underset{(-2.41)}{0.15^{**}}(DUM_{65} - DUM_{67})(t/100)$$
$$\underset{(5.64)}{+0.14^{**}}(DUM_{79} - DUM_{85})(t/100) \underset{(8.63)}{+0.25^{***}}(DUM_{85} - DUM_{89})(t/100) \underset{(8.25)}{+0.55^{***}}(DUM_{89} - DUM_{90})(t/100) \underset{(6.40)}{+0.32^{***}}(DUM_{90} - DUM_{94})(t/100).$$

[14] As mentioned before, in our application we find for some countries more than one break in the trend. The notation for the case with multiple breaks in the trend is a straightforward extension of the notation used in the text (for example, the coefficient in front of the term $(DUM_{64} - DUM_{78})(t/100)$ represents the coefficient of the trend from 1965 to 1978). Moreover, we excluded in the estimation equation the trend for a particular time interval if in prior estimations the trend was insignificant for that time interval (according to at least the 15% significance level).

ITA

$$TFR_{t-1} = 0.30 - \underset{(-2.26)}{0.85^{**}}(DUM_{59} - DUM_{81})FLP_{t-1} - \underset{(-2.72)}{1.05^{**}}(DUM_{81} - DUM_{94})FLP_{t-1} - \underset{(-3.66)}{1.78^{***}}(DUM_{64} - DUM_{74})(t/100)$$

$$\underset{(-3.52)}{-2.69^{***}}(DUM_{74} - DUM_{87})(t/100) - \underset{(-3.57)}{2.19^{***}}(DUM_{87} - DUM_{94})(t/100),$$

$$FLP_{t-1} = -0.67 - \underset{(-8.48)}{0.49^{***}}(DUM_{59} - DUM_{80})TFR_{t-1} - \underset{(-6.40)}{0.70^{***}}(DUM_{80} - DUM_{94})TFR_{t-1} + \underset{(-3.62)}{0.30^{***}}(DUM_{74} - DUM_{92})(t/100)$$

$$\underset{(-2.13)}{-0.28^{**}}(DUM_{92} - DUM_{93})(t/100).$$

SWE

$$FLP_{t-1} = -0.26 - \underset{(-4.10)}{0.35^{***}}(DUM_{59} - DUM_{73})TFR_{t-1} - \underset{(-1.58)}{0.17^{*}}(DUM_{73} - DUM_{94})TFR_{t-1} + \underset{(5.13)}{0.38^{***}}(DUM_{67} - DUM_{91})(t/100).$$

UK

$$TFR_{t-1} = 0.13 - \underset{(-7.15)}{1.25^{***}}(DUM_{59} - DUM_{77})FLP_{t-1} - \underset{(-3.34)}{0.97^{***}}(DUM_{77} - DUM_{94})FLP_{t-1} - \underset{(-3.27)}{0.46^{***}}(DUM_{64} - DUM_{71})(t/100)$$

$$\underset{(-8.97)}{-1.35^{***}}(DUM_{71} - DUM_{77})(t/100) - \underset{(-2.59)}{0.26^{**}}(DUM_{77} - DUM_{86})(t/100),$$

$$FLP_{t-1} = -0.75 - \underset{(-2.95)}{0.10^{***}}(DUM_{59} - DUM_{86})TFR_{t-1} - \underset{(-1.40)}{0.04}(DUM_{86} - DUM_{94})TFR_{t-1} + \underset{(8.18)}{0.80^{***}}(DUM_{59} - DUM_{94})(t/100).$$

USA(1960-94)

$$TFR_{t-1} = 0.21 - \underset{(-5.25)}{1.10^{***}}(DUM_{59} - DUM_{70})FLP_{t-1} - \underset{(-2.86)}{0.80^{***}}(DUM_{70} - DUM_{94})FLP_{t-1} - \underset{(-3.42)}{0.77^{***}}(DUM_{64} - DUM_{76})(t/100)$$
$$\underset{(-3.42)}{-0.45^{***}}(DUM_{86} - DUM_{94})(t/100),$$
$$FLP_{t-1} = -0.43 - \underset{(-6.69)}{0.30^{***}}(DUM_{59} - DUM_{84})TFR_{t-1} - \underset{(-5.52)}{0.21^{***}}(DUM_{84} - DUM_{94})TFR_{t-1} + \underset{(2.09)}{0.20^{**}}(DUM_{64} - DUM_{76})(t/100)$$
$$\underset{(5.34)}{+0.51^{***}}(DUM_{76} - DUM_{94})(t/100).$$

USA(1948-95)

$$TFR_{t-1} = 0.39 - \underset{(-7.46)}{0.79^{***}}(DUM_{47} - DUM_{70})FLP_{t-1} - \underset{(2.34)}{0.35^{**}}(DUM_{70} - DUM_{95})FLP_{t-1} + \underset{(-3.72)}{1.28^{***}}(DUM_{52} - DUM_{61})(t/100)$$
$$\underset{(-3.24)}{-0.53^{***}}(DUM_{64} - DUM_{73})(t/100) + \underset{(3.79)}{0.32^{***}}(DUM_{86} - DUM_{95})(t/100),$$
$$FLP_{t-1} = -0.91 - \underset{(-4.09)}{0.09^{***}}(DUM_{47} - DUM_{77})TFR_{t-1} - \underset{(-0.68)}{0.02}(DUM_{77} - DUM_{95})TFR_{t-1} + \underset{(19.91)}{1.30^{***}}(DUM_{47} - DUM_{89})(t/100)$$
$$\underset{(20.72).}{+1.23^{***}}(DUM_{89} - DUM_{95})(t/100).$$

USA20-24(1948-95)

$$FER_{t-1} = -2.21 - \underset{(-9.64)}{0.91^{***}}(DUM_{47} - DUM_{71})FLP_{t-1} - \underset{(-0.16)}{0.03}(DUM_{71} - DUM_{95})FLP_{t-1} + \underset{(4.58)}{1.09^{***}}(DUM_{53} - DUM_{61})(t/100)$$
$$\underset{(-5.22)}{-0.99^{***}}(DUM_{64} - DUM_{71})(t/100),$$

$$FLP_{t-1} = -1.27 - \underset{(-14.65)}{0.31^{***}}(DUM_{47} - DUM_{81})FER_{t-1} - \underset{(-13.87)}{0.30^{***}}(DUM_{81} - DUM_{95})FER_{t-1} - \underset{(-1.91)}{0.94^{**}}(DUM_{51} - DUM_{53})(t/100)$$
$$\underset{(10.55)}{+0.73^{***}}(DUM_{53} - DUM_{89})(t/100) + \underset{(9.41).}{0.60^{***}}(DUM_{89} - DUM_{95})(t/100).$$

USA25-34(1948-95)

$$FER_{t-1} = -1.88 - \underset{(-8.30)}{0.56^{***}}(DUM_{47} - DUM_{70})FLP_{t-1} - \underset{(2.89)}{0.32^{***}}(DUM_{70} - DUM_{95})FLP_{t-1} + \underset{(4.23)}{1.28^{***}}(DUM_{50} - DUM_{61})(t/100)$$
$$\underset{(-4.80)}{-0.64^{***}}(DUM_{64} - DUM_{76})(t/100) + \underset{(3.73)}{0.29^{***}}(DUM_{84} - DUM_{95})(t/100),$$

$$FLP_{t-1} = -2.40 - \underset{(-7.31)}{1.06^{***}}(DUM_{47} - DUM_{67})FER_{t-1} - \underset{(-7.58)}{0.96^{***}}(DUM_{67} - DUM_{95})FER_{t-1} + \underset{(7.79)}{1.24^{***}}(DUM_{47} - DUM_{90})(t/100).$$

Notes: *,**,*** 15 %, 5 %, 1% significance level, t-statistics in parenthesis and
$DUM_i = \begin{cases} 1 & if\ t > i \\ 0 & otherwise \end{cases}$, $\forall\ i\epsilon[60, 94]$, FER represents the age specific fertility rate.