A new estimation method for employment trend

Andrea Tafuro, Stefano Scalone
July 25, 2016

1 Introduction

Estimating the cyclical component of the labor market variables is of crucial importance for economic research. Recent studies have tried to quantify the effectiveness of policy intervention on the process of job creation (Monacelli et al. 2010, Bruckner and Pappa 2012, and Turrini 2013), and have focused on the ability of discretionary policies (both fiscal and monetary) to reduce the cyclical component of unemployment, which is usually estimated as the deviation of the unemployment rate from the NAIRU.

In our opinion, such an empirical investigation may suffer of several short-comings, which can lead to unreliable results. First, the unemployment rate is directly affected by the behavior of GDP, since it is calculated as the ratio of unemployment to labor force: for example, a prolonged recession not only increases the absolute number of unemployed, but also makes some of them exit the labor force. Therefore, the responsiveness of the labor force to the cycle should lead to underestimate the effect of fiscal shocks on unemployment.

Second, as described by Blanchard and Summers (1986)'s seminal paper, the presence of hysteresis in the job creation process would make the estimations of the cyclical unemployment unreliable. In fact, considering only the cyclical component of unemployment, the impact of fiscal shocks on the labor market tends to be underestimated if a fraction of cyclical unemployment becomes structural. Moreover, we can better evaluate the persistency of the effect of the policy change by taking into account the long-term component of unemployment.

Third, statistical approaches that try to filter out the cyclical component of (un)employment do not capture the demographic factors that are strongly modifying the composition of the population. Population aging, for example, is creating structural breaks in the composition of the labor force and causes bias in the estimations, often leading the policy maker to misinterpret these results. For instance, one half to two thirds of the decline in labor force participation from 2007 to 2014, usually addressed as caused by cyclical factors, seems to be

attributable to population aging (Cline and Nolan, 2014).

The goal of this paper is to address the latter of these drawbacks by presenting an innovative two-step procedure: first, we create a state-space model including in the estimation method some conditioning variables to bettere detect the *cycle free* component of the labor force. Second, we want analyse the labor force by age cohort and gender, to calculate average partecipation rates and overcome the issue of the evolving composition of the population.

Our belief is that the business cycle can have heterogeneous effects across different cohorts: for instance, a contractionary phase will affect more the "marginal workers", those more likely to leave the labor force - the youngest, the oldest, or females, for which the effects can also be stronger (Krusell et al., 2010, Elsby et al., 2013). Therefore, the cyclical component of the labor force should be filtered using an *ad hoc* procedure: the total equilibrim level of population and labor force is obtained computing the weitghted average of the equilibrium level of de variables obtained in each gender/age cohorts. By doing so, we account for the structural dynamics of these two variables, such as fertility rates, net migration, schooling and education.

The paper is organized as follows: section 2 introduces the main issues that undermine the typical estimation methods, section 3 describes our model, section 4 presents our results and section 5 concludes.

2 Population Aging

When analyzing the labor market, nowadays researchers tend to underestimate the role of the structural breaks we are experiencing in the composition of the population. This clearly undermines the reliability of the results, in particular if they are focused on the medium-to-long run effect of economic policies. In order to correctly investigate the relation between labor market variables and the economic system, one has to take into account three of the main causes of these breaks: population aging, schooling, and growth of female partecipation rate.

The process of population aging is modifying deeply the composition of the labor force, and is probably the main cause of the structural breaks in the composition of the population. It is the result of a lower fertility rate (especially in developed countries, see figure (1), combined with a higher life expectancy (in particular, that of men is rapidly catching up with that of women, see figure (2). Moreover, with "baby boomers" reaching the age of retirement, the ratio between workers and pensioned is falling down, putting increasing pressure on pension systems in most Western countries.

Another important factor affecting the dynamics of the labor force is the

Figure 1: Fertility Rates. Source: OECD Data

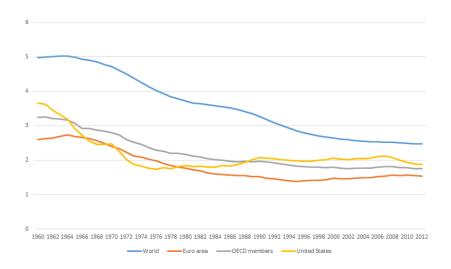
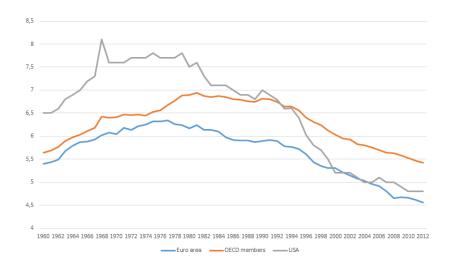


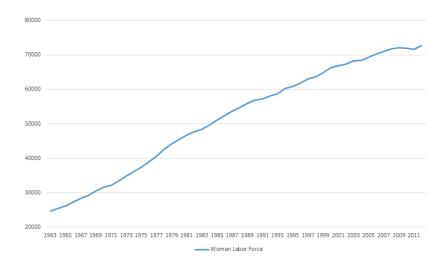
Figure 2: Life Expectancy Difference, Female-LE - Male-LE. Source: OECD Data



steadily increasing participation of women in the job market. This has had at least two important consequences: the first is the direct increase of labor demand, which contributed to enlarge the labor force. The second and more recent phenomenon is strongly related to periods of crisis: when one member of the family unit remaines unemployed, the other - often a woman - is likely to enter the labor force (Eltsby et al., 2013). Not surprisingly, the women labor force kept on increasing during the recent global economic downturn (figure (3)).

The increasing level of education is also contributing to modify the com-

Figure 3: Women Labor Force, United States. Source: OECD Data



position of the labor force, in particular for the youngest cohorts: secondary education and university delay the entrance of young workers into the labor force. We must take this relatively new phenomenon into account when creating the new labor force dataset.

3 The Estimation Process

This study aims at increasing the accuracy of the estimates of the long term trends of employment rate and labor force. In order to do this, we need to focus on which methology adopt to disentangle the cyclical component of a variable from its trend. There are two commont techniques used in the literature to achieve this goal. One is the adoption of a purely statiscal approach to estimate a reduced form relation. The other consists in estimating the equilibrium - or full-utilisation - level of the variables with a structural model, in which the extent of the gap between the equilibrium and the actual level of the variables strictly depends on the presence of frictions, or rationing, in the economic system.

However, both of the methods suffer of a crucial problem. When estimating the equilibrium level in the economy with a structural model, the frictions present in the model are considered exogenous, and hence cannot be a target measure for the policymakers (Borio et al. 2013). As a consequence, different assumptions on the presence and nature of these frictions lead to different results. If the equilibrium level of these variables is estimated using a reduced form univariate model (such as the Hodrick-Prescott filter, the Baxter and King

filter or other unobserved component methods), the analysis will suffer of the well-known end-point problem: the reliability of the end-of-the-sample estimates - which is usually the most important from the point of view of the policy maker - is limited, and this affects the usefulness of the results for real-time analysis or policy decisions. Moreover, in the most used technique,he Hodrick-Prescott filter, the amplitude of the frequency in which we are interested in is exogenousely set by the researcher.

For all these reasons we propose an alternative estimation process, using a refined version of the Kalman Filter. The higher accurancy of the Kalman filter with respect to the other procedures listed above is ensured by two different facts: the frequencies are not anymore set exogenously by the researcher, but computed by the filter itself with an estimation update algorithm that enlarges the convergence speed to the true signal-to-noise ratio. The second is the possibility to make the filter to better fit the data, by adding some additional explicative variables - that we will call conditioning variables - to the estimation process.

The Kalman filter can be represented by a state-space model of the form:

$$\Delta l f^* = \Delta l f_{t-1}^* + \epsilon_{1,t} \tag{1}$$

$$lf_{t} = lf_{t}^{*} + \gamma' x_{t} + \epsilon_{2,t} \tag{2}$$

where $lf_t = ln(LF_t)$, lf^* the labor force equilibrium level, and Δlf_t^* is the log of the labor force cyclical component. $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are assumed to be normally and indipendent distributed errors with zero mean and variance σ_1^2 and σ_2^2 . Equation (1) is the state equation, which estimates the variable in which we are interested - in this case the trend of labor force. Equation (2) is the measurement equation, which aims at disentangle the trend from the cycle including an error term and some proxies for the financial cycle in the vector x_t to help filtering out the cyclical component. Then, with this model we identify the state of the variable, which can be considered the "true" - or the equilibrium - level, getting rid of what can be considered a measurement error, which in our case is the cyclical component. The idea behind equation (2) is that the cyclical component, which fills the gap between the trend component of the variable and its actual value, is nothing more than an error term. As a consequence, in this state space model the measurement error has a well-defined behavior that might be better identified thanks to the conditioning variables. Of fundamental importance is the parameter $\lambda_1 = \frac{\sigma_1^2}{\sigma_2^2}$, the signal-to-noise ratio, which determines the relative variability of the estimated labor force equilibrium level. When λ_1 becomes very large, our equilibriul level will approximate a linear trend. If λ_1 gets close to zero, there will be no difference between the estimated trend and the actual measure. In our exercise, the value of λ_1 is settled equal to 100 for the HP filter, while we restrict the signal-to-noise $\lambda_1 = \frac{\sigma_1^2}{\sigma_2^2}$ to be:

$$\frac{y_t - y_{(hp),t}^*}{var(\Delta^2 y_{(hp),t}^*)} = \frac{y_t - y_{(kf),t}^*}{var(\Delta^2 y_{(kf),t}^*)}$$
(3)

in the complete model, where $y_{(2),t}^*$ and $y_{(3),t}^*$ are the potential output from the HP filtering and the Kalman Filter. In this way we preserve the business cycle implied in the standard HP filter also in the Kalman Filter.

The main advantage of such an approach (a là Borio et al., 2013) is that it is the model itself to suggest which variables (and which not) to incorporate in the model. The statistical theory tells that proxies of the financial cycle carry significant information regarding the output gap: those variables will be included in the state-space model and filtered out. In addition to these, we tested the significance of other variables which were considered helpful in detecting the cyclical component of the labor force, such as proxies for monetary and fiscal policy, the inflation rate, real gdp, and the potential output. While the relevance of the former variables is straightforward, the importance of potential gdp in detecting the cyclical component may be puzzling. It is worth to remind that, since the estimation procedure to obtain this variable may fail in detecting the cycle and filtering it out, it may still contain a cyclical component. Furthermore, we also enrich our specification testing the significance of lagged values of the labor force, since the presence of an integreted process in the level, or autoregressive order 1 or 2 in the differences, may lead to distorted estimates.

We propose to incorporate the structural breaks in the population by filtering the labor force cohorts by gender and age. Estimating the trend component of the labor force within the single cohorts, we account for all the structural breaks that can affect the labor market - such as aging, fertility, migration, schooling. For instance, the equilibrium measure of the male labor force in the 19-24 years old cohort will slowly embody the reduction of both population and people actively looking for a job due to schooling and fertility. Furthermore, also the business cycle and other structural dynamics should affect differently the probability to find a job of agents in different cohorts (Krusell, Mukoyama, Rogerson, and Sahin 2010; Elsby, Hobijn, and Sahin 2013). We conduct this exercise using OECD data on the labor force and population for the United States in the period 1960-2013. The data are annual and divided by gender and ages classes (the lenght of a class is 5 years, starting with the 19-24 and endig with the 65 and over).

Hence, the analysis is divided into consecutive steps. In the first one we filter each single cohort using the model described above, in which the equation (2) is rewritten as:

$$lf_t - lf_t^* = \gamma_1 cr_t + \gamma_2 m p_{t-1} + \epsilon_{4,t} \tag{4}$$

where the γ_1 is the coefficient for the credit rate share on GDP (cr_t) , while γ_2 is the coefficient of the lagged value of the monetary policy, mp_{t-1} , captured by the interest rate on 3-months treasury bills. $\epsilon_{4,t}$ is the usual indipendent and identical distributed error. In order to obtain the most parsimonious model, the choice of the variables to insert in equation (2) follows a general to particular procedure: we start by testing the significance of the variables (current GDP, inflation rate, monetary policy, and government primary balance) in each cohort, which were considered meaningfull to detect the cyclical component of the labor force, and their three lags. By eliminating the non-significative lags and variables, we obtained equation (2), which looks similar to the Borio et al (2013) one, with the crucial exception that we do not have an autoregressive term on the right hand side. Such difference may depend on a different behavior of the cyclical component. As Borio et al. (2013, 2014) discussed in their paper, the presence of an autoregressive term enhance the estimation robustness, while does not modify the punctual result. However, in our case the cyclical component of labor force appears as an erratic term.

Notably, the procedure above assures transparency in the estimation methodology and semplicity respect a structural model: the variable are excluded from the model because they lack significance in helping the filter ruling out the cyclical component, and this can be tested with a standart t-statistic. Furthermore, enriching the model with exogenous variables, improves the convergence to the true values narrowing the confidence intervals. Such a procedure allows us to avoid the problem of collinearity when evaluating the significance of the single variable, even if when we evaluate the complete model some of the variables which were significative in the take-alone test become irrelevant.

Once we have obtained the estimates of the labor force and population equilibrium levels for each cohort in every year, we aggregate them to obtain the yearly equilibrium levels of labor force and partecipation rate. The aggregation is done with a weighted average, were the weights are the shares of single cohorts in the total population - i.e. cohort population over whole population. The final result of these estimates is new series for labor force and parteciparion rates, with the cyclical component filtered out.

What is the rationale behind this? As we said, the labor force (share of population working or actively looking for a job) has a clear and strong cyclical component. It is in fact responsive to the fluctuations of the economy, in the sense that periods of strong economic growth induce people to enter the labor force, while prolonged recessions discourage unemployed to look for jobs. Hence,

crises contribute to deteriorate labor market conditions, and often result in significant drops in the labor force.

Less straigthforward is the other relationship suggested by our model: the one between the cycle and population. Population depends on three macro factors: life expectancy, fertility rate and net migration. At least in western countries, where acceptable standards of living, access to food and health care are generally available to everybody despite of the economic conditions, life expectancy does not seem to be strongly linked to economic conditions.

On the other hand, recent studies seem to associate pro-cyclical fluctuations to fertility rates: in contrast with a few decades ago, when fertility rates were considered counter-cyclical as they incresed the opportunity cost of childbearing (Butz and Ward, 1979), researchers have recently underlined the pro-cyclical behavior of fertility, which seems to respond positively to the ups and downs of the economy (Orsal and Goldstein, 2010, Sobotka et al., 2010). Fertility rates seem to be more responsive to econoimc fluctuations for younger cohorts, which have the possibility to postpone the decision to have kids, on the contrary to older couples closer to the biological limit of fertility (Goldstein et al., 2013).

Migration is the third fundamental component of the changes in population. In the last few decades in the US it has accounted for more than one third of the total population growth, and for up to a half of the labor force growth. Migration seems to have a pro-cyclical pattern, too: after six decades of continuous growth, the undocumented population in the US peaked in 2006 and declined during the recession, while also legal immigration levelled off (Massey, 2012). If we analyze the most usual case, that of net migration from Mexico to the US, we see that during the recession the number of Mexican emigrants dropped significantly, while the flows from the US to Mexico more than doubled between 1995-2000 and 2005-2010 (Passel et al., 2012).

4 Results

In this section we briefly report the main figures of our results. A more comprehensive presentation of all our estimations is contained in the Appendix.

Table (1) reports the statistical significance of our proxies, credit over gdp- cr_t - and monetary policy interest rate - $mp_t - 1$ -in estimating the labor force and population by gender and age cohort. As Borio et al. (2014) underline, the significance of a variable implies not only that this variable is correlated with the the labor force (or population), but also with the frequencies implicitly set by the scaling factor.

Moreover, the conditioning variables should have a stable mean. Since this characteristic is rarely present in economic time series, we demeaned the conditioning variablesvia Cesàro means¹, which increased the rate of convergence of our model.

As expected, the conditioning variables are always significant for the youngest cohorts of both population and the labor force. Indeed, the attachment to the labor force tends to be lower for the young workers, since they may decide either to go to college or back to high school instead of working, or just to wait for better job opportunities in the future. For the other cohorts, the significance of the two variables is more puzzling, and this might be explained both by a smoother path of the labor force and population, or a mismatch between the cycles of the two conditioning variables and the objective ones. The latter could be interpreted as a signal of an increasing attachements to the labor force shown by workers at an older age.

Turning to the population estimates, the significance of both the conditioning variables for the 15-19 cohort may depend on migration flows. In broad words, we expect migration inflows during booms, and outflows during busts, both correlated to credit expansion and a relative accommodant monetary policy. If this is the case, migration should be larger among young people, usually more likely to migrate (the 15-34 cohort represents over 30

One of the main problems we want to overcome with our estimates is that of the end-point problem, and the limited real-time reliability of the filtered series. This is a fundamental issue in understanding how well such a multivariate filtering method performs. From this point of view, our method seems to significantly increase the quality of the HP-filtered estimates.

If one observes - for instance - the graphs for the 15-19 cohort for male labor force reported in the Appendix, in 2008 the HP-filtered series still estimates the labor force to be above the equilibrium level, while our model suggests it to be much closer to the long-run level. Moreover, the HP series seems to suggest that as of 2012 (so close to the end of the sample) the labor force is once again above the equilibrium, differently with respect to our estimates which sets the equilibrium level above the actual series. Our model estimation seems to capture the dynamics of the financial cycle better than the HP filter.

Our model improves the current estimation procedures in two different ways. On the one hand, as depicted in the graphs reporting the comparison among the actual series, the Kalman Filter (KF), and the HP filtered (HP) one, we can see that the KF series displays a larger negative - or a tighter positive - cyclical labor force at the end of the period, claiming a better real-time estimation. On the other hand, the KF serier has a narrower confidence bands if compared with

¹Named after Ernesto Cesàro, who proved that if a sequence of numbers converge to a constant - the mean - the sequence of arithmetic means taken over the first n first elements also converge to the same costant.

Table 1: Regression results: significance of explanatory variables

	Male Labor Force										
Variable	15-19	20 - 24	25 - 29	30 - 34	35 - 39	40-44	45-49	50 - 54	55 - 59	60 - 64	65+
$\overline{cr_t}$	Y	N	N	N	Y	N	N	N	Y	N	N
$mp_t - 1$	Y	Y	Y	N	Y	Y	N	N	N	N	Y
	Male Population										
Variable	15-19	20 - 24	25 - 29	30-34	35 - 39	40-44	45-49	50 - 54	55 - 59	60 - 64	65+
$\overline{cr_t}$	Y	N	N	N	N	N	N	N	Y	N	Y
$mp_t - 1$	Y	N	Y	N	Y	Y	Y	N	N	N	N
	Female Labor Force										
					Female	e Labor	Force				
Variable	15-19	20-24	25-29	30-34	Female 35-39	e Labor 40-44	Force 45-49	50-54	55-59	60-64	$\overline{65+}$
$\frac{\text{Variable}}{cr_t}$	15-19 Y	20-24 N	25-29 N	30-34 N				50-54 N	55-59 Y	60-64 N	$rac{65+}{ m N}$
					35-39	40-44	45-49				
cr_t	Y	N	N	N	35-39 N Y	40-44 N	45-49 N N	N	Y	N	N
cr_t	Y	N	N	N	35-39 N Y	40-44 N Y	45-49 N N	N	Y	N	N
$\frac{cr_t}{mp_t - 1}$	Y Y	N N	N N	N N	35-39 N Y Femal	40-44 N Y e Popula	45-49 N N ation	N N	Y N	N N	N Y

the usual HP ones, this implying more precise estimates.

Finally, we report the graphs of the total participation rates, as an average weighted on the share of each cohort in the total population. The figures compare the behavior of the OECD data (actual data) with our estimation of partecipation rate for male, female, and total population. Three different evidence emerge from the graphs: the first is that the male partecipation rate appears to be below its potential after the 1993. This gap, which seems to refill during the first half of 00s, starts to increase again with the recent crises. Second, the female participation rates are always above its potential and the spread between actual data and the estimated trend is increasing after the recent recession. This two evidences together suggest that the major contribution to the so-called jobless recovery puzzle - i.e. the empirical evidence that the recoveries from the recession in the last two decades did not imply an increase of the level of employment back to the pre-crises levels - was due to the behavior of the male labor force. Third, the partecipation rate for the whole population is perfectly in line with the US economic history: it becomes negative after the first oil shock, suggesting that the partecipation levels to the job market was to high, turning negative after the 2001 and again after 2008. This result validate the idea that the partecipation to the labor market is highly correlated with the business cycle.

Figure 4: Males, Smoothed and Estimated Participation Rates

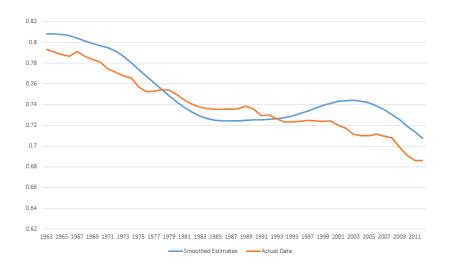
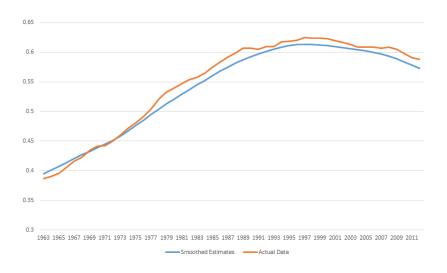


Figure 5: Females, Smoothed and Estimated Participation Rates



5 Conclusion

In this paper we have tried to derive an innovative method for filtering the cyclical component out of population and the labor force. The estimation procedure consists of a state-space model where the measurement equation is augmented

Figure 6: Total Participation Rates, Weighted Average

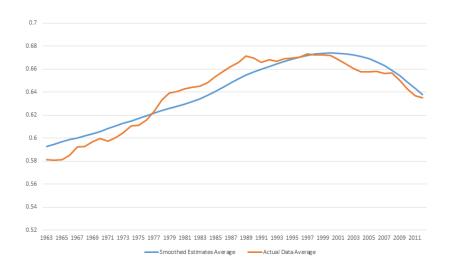
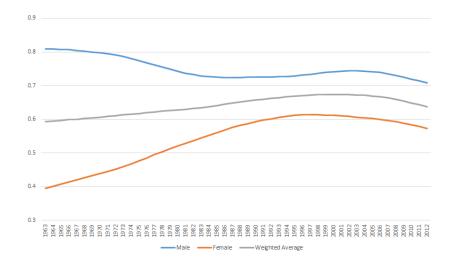


Figure 7: Participation Rates: Males, Females and Total (Weighted Average)



to include some proxies for the cycle.

The basic results obtained up to now are encouraging, and our model seems to perform better than traditional methods (especially with respect to the simple H-P filtering). The proxies we identified are significant in many of the cohorts analyzed, and the smoothed series seem to increase the reliability of the estimates with respect to other methods, in particular regarding the end of the sample.

However, many issues remain still open, both from the theoretical and from

the statistical point of view. Theoretically, the choice of the proxies and the lags is arbitrary, and there is no precise rule on how to choose them: they are simply the ones that maximize the statistical significance.

From the statistical point of vies, some of the results are not perfectly realistic, and the estimates need to be refined. Some proxies in fact intervene too much in modifying the original series, and for some cohorts the labor force reaches 100%.

However, the methodology seems promising and the results encouraging. Further developments of the model will focus on try to overcome these issues.

Figure 8: Females Labor Force Estimates

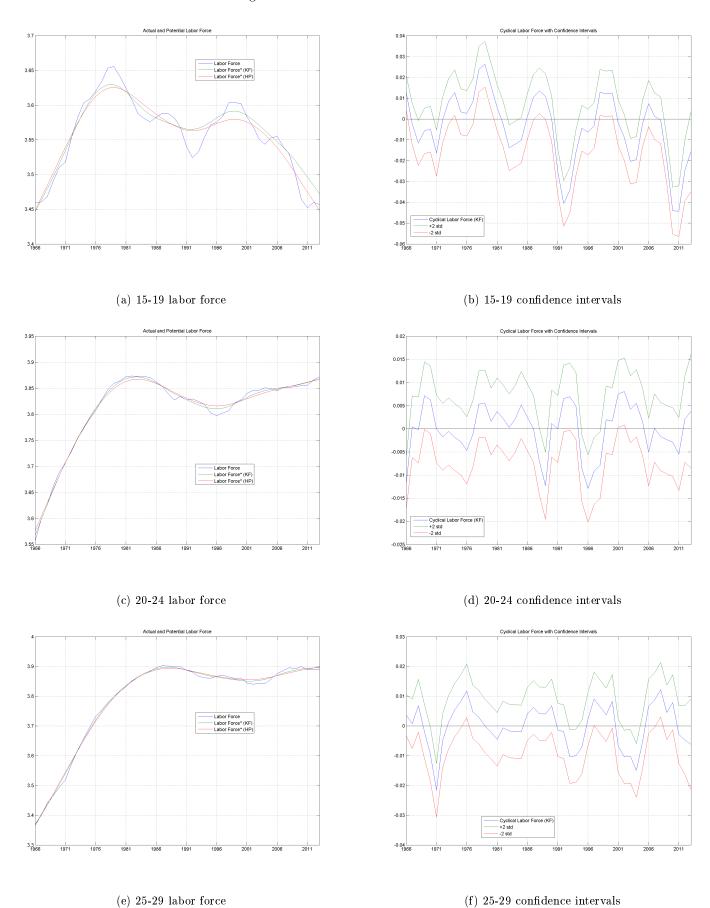


Figure 8: Females Labor Force Estimates

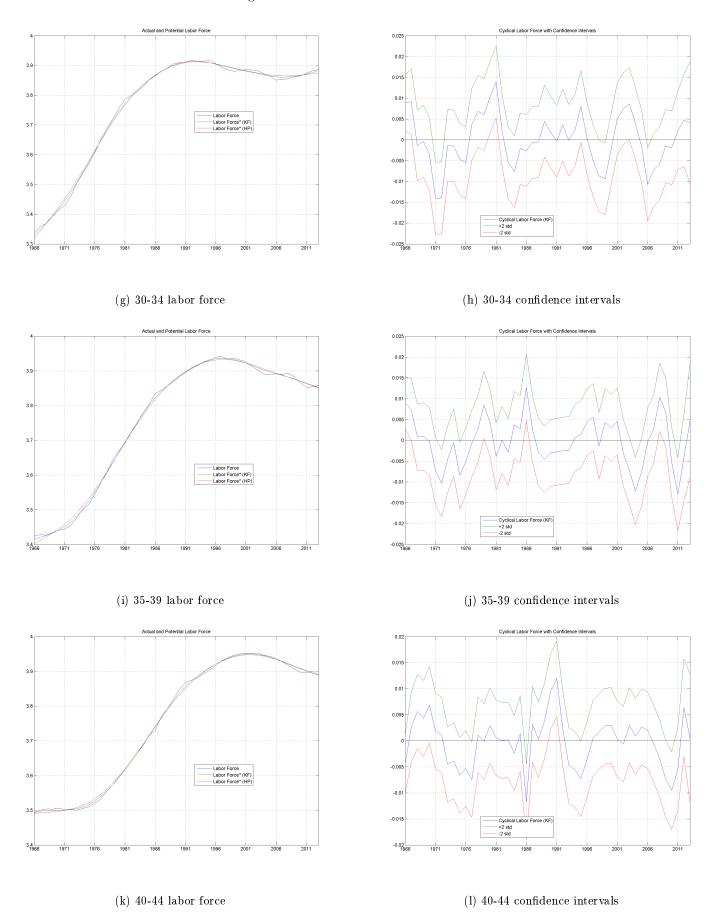


Figure 8: Females Labor Force Estimates

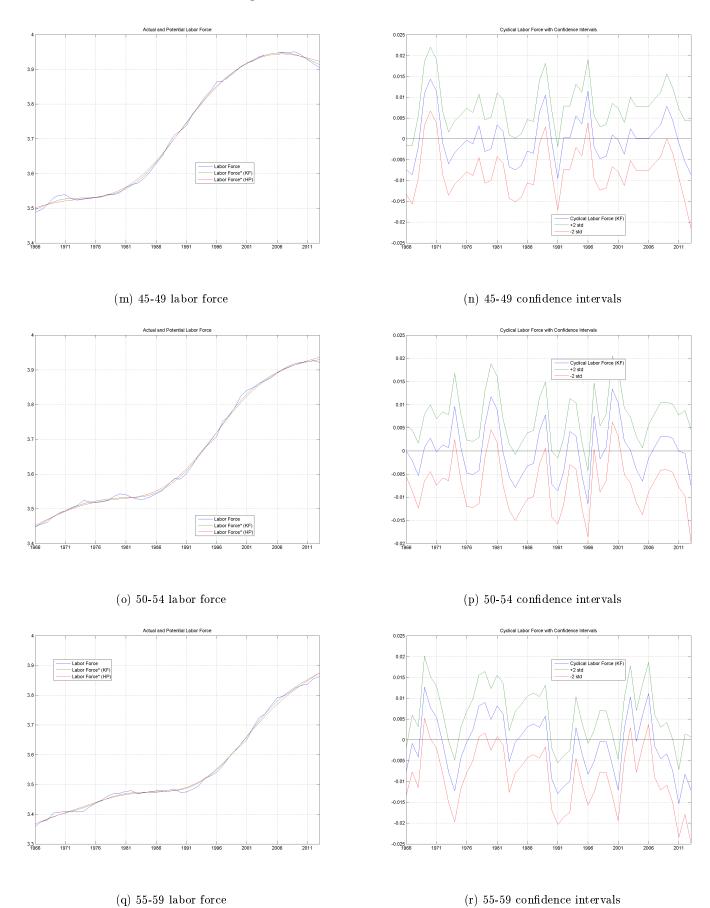
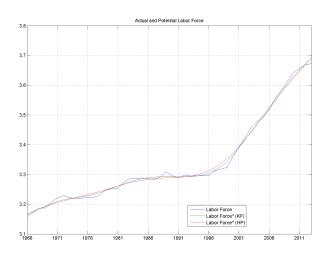
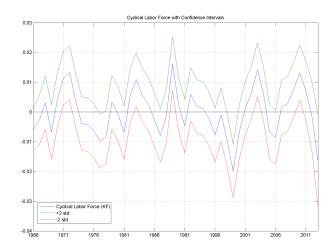
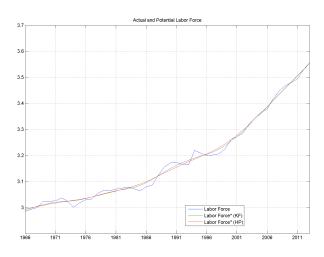


Figure 8: Females Labor Force Estimates

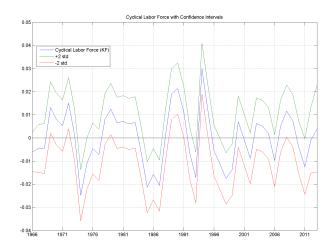




(s) 60-64 labor force



(t) 60-64 confidence intervals



(u) 65+ labor force

(v) 65+ confidence intervals

Figure 9: Females Population Estimates

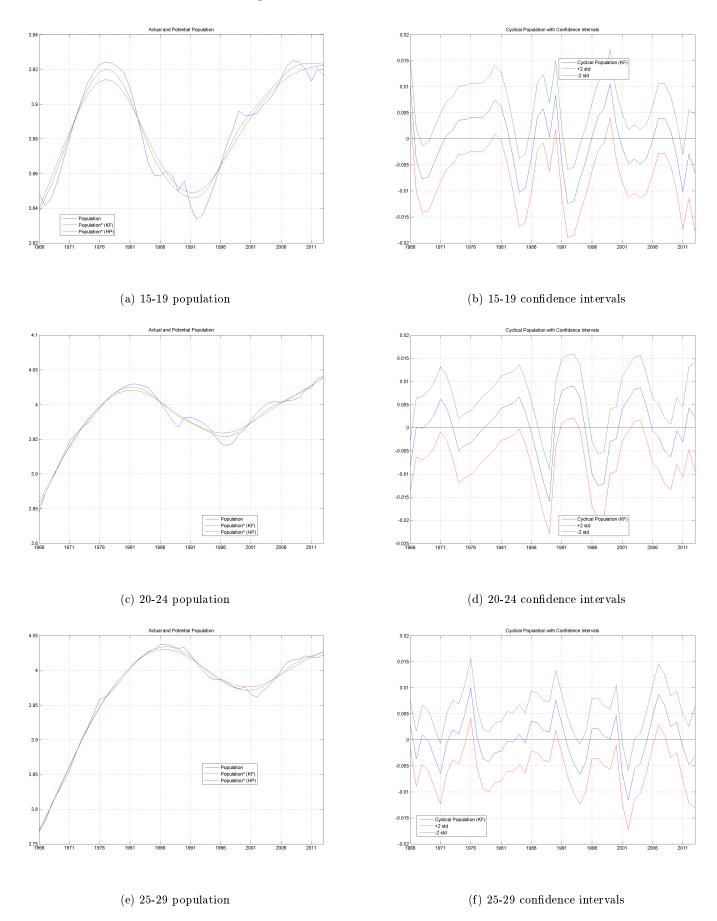


Figure 9: Females Population Estimates

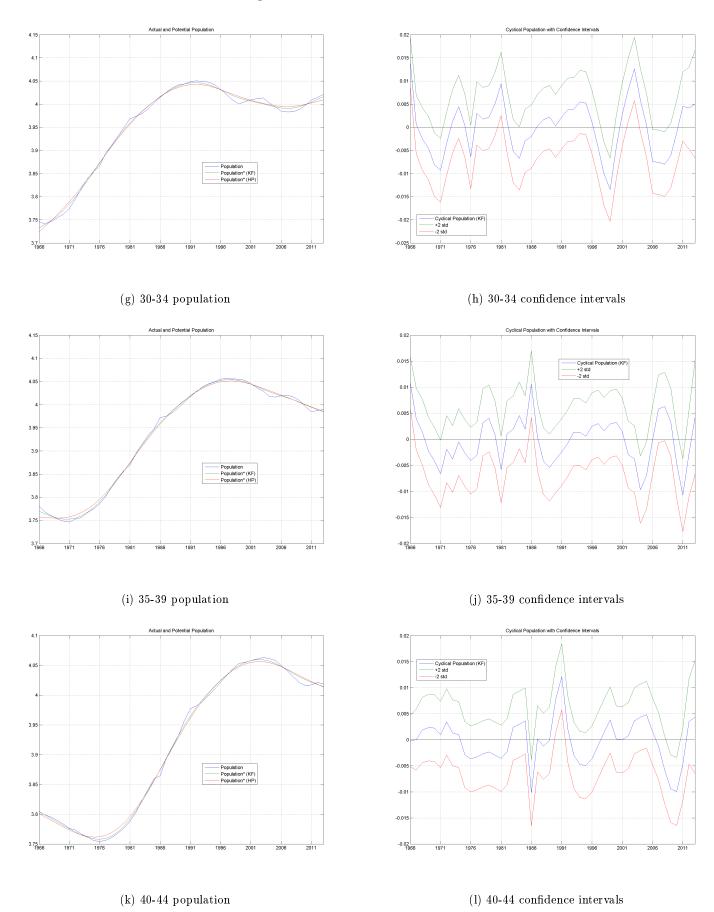
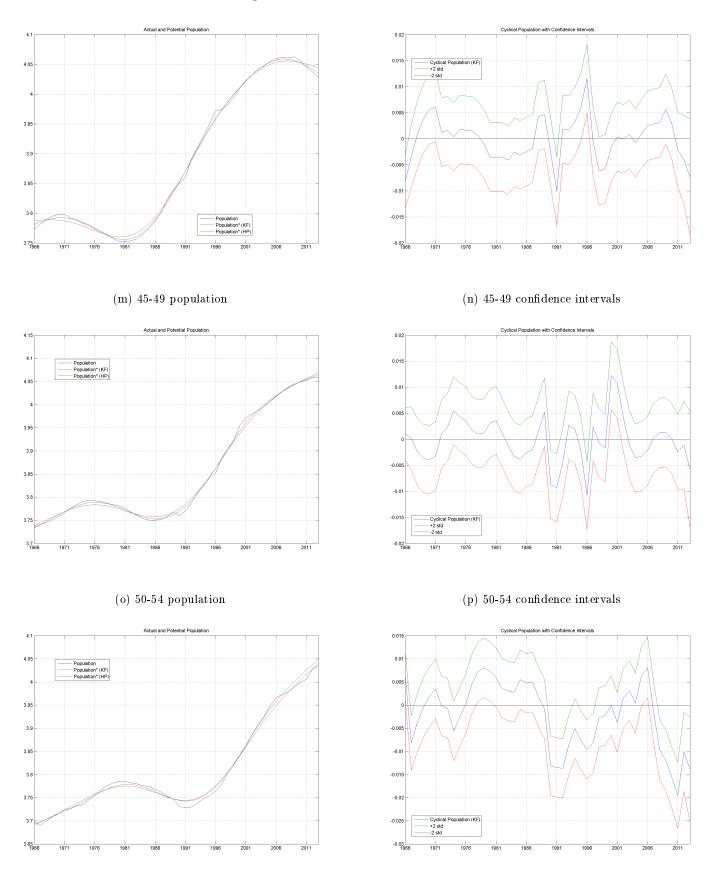


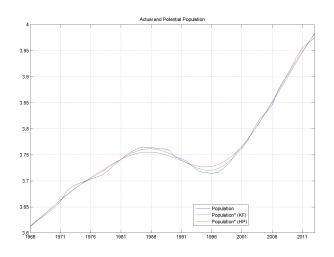
Figure 9: Females Population Estimates

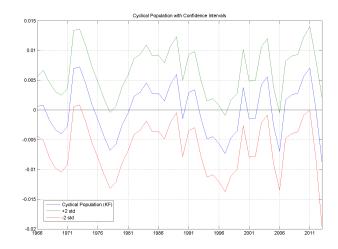


(r) 55-59 confidence intervals

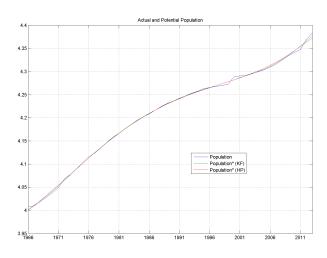
(q) 55-59 population

Figure 9: Females Population Estimates

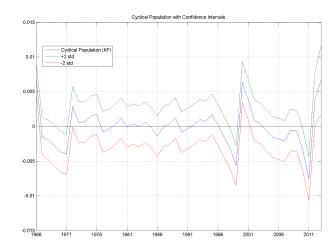




(s) 60-64 population



(t) 60-64 confidence intervals



(u) 65+ population

(v) 65+ confidence intervals

Figure 10: Males Labor Force Estimates

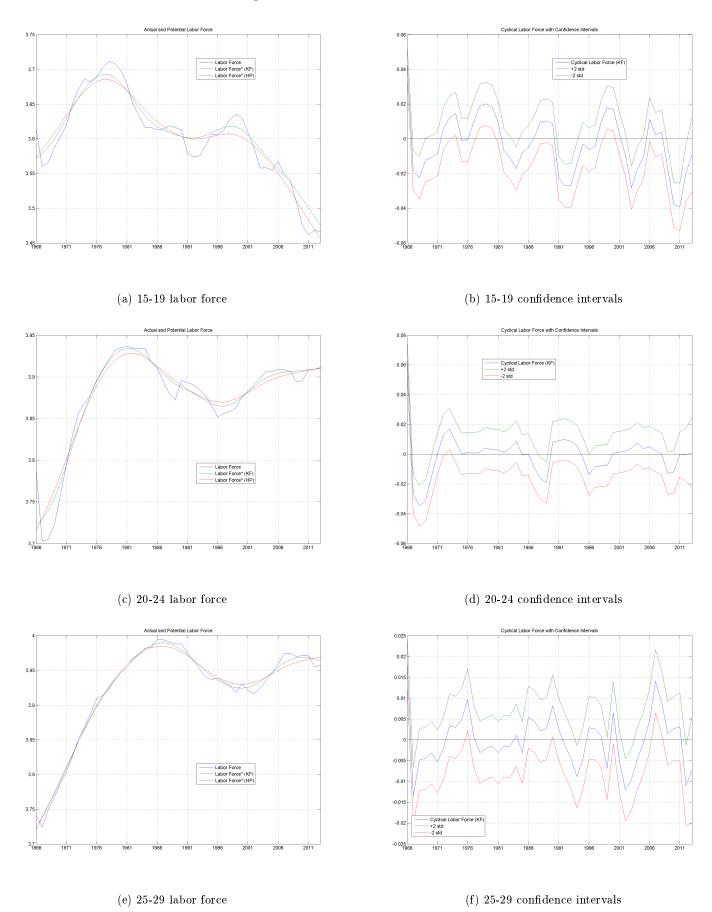


Figure 10: Males Labor Force Estimates

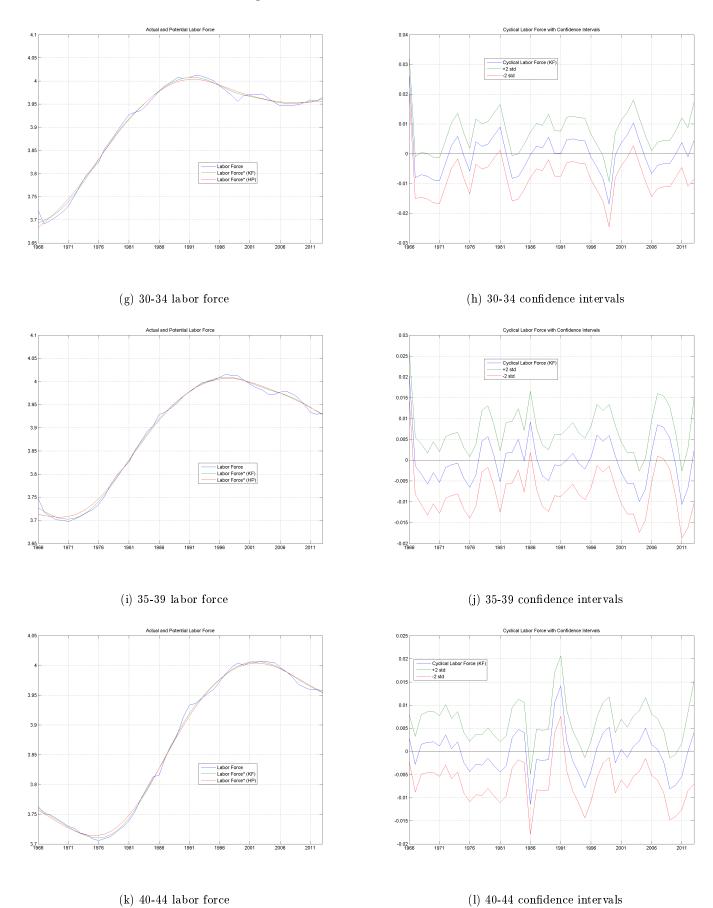


Figure 10: Males Labor Force Estimates

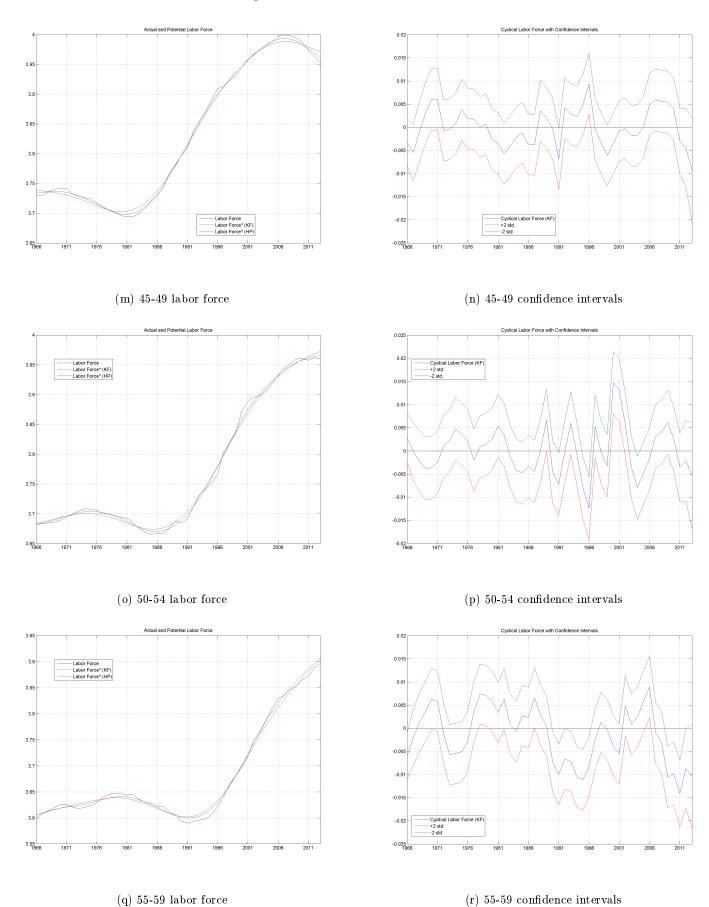
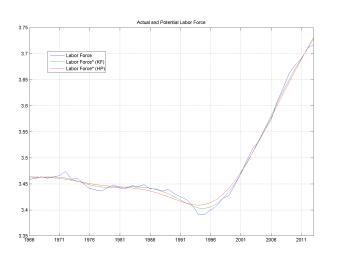
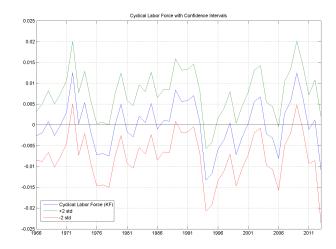
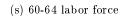
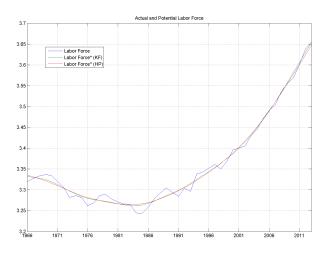


Figure 10: Males Labor Force Estimates

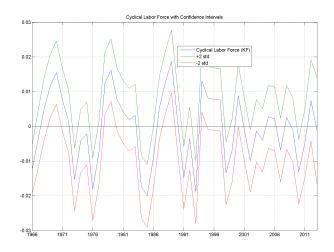








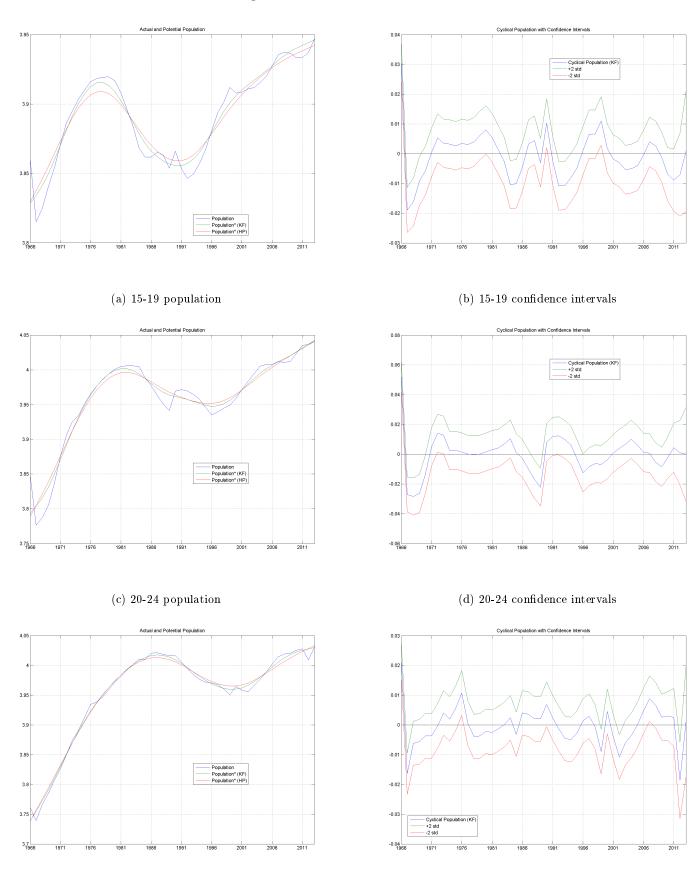
(t) 60-64 confidence intervals



(u) 65+ labor force

(v)~65+~confidence~intervals

Figure 11: Males Population Estimates



(f) 25-29 confidence intervals

(e) 25-29 population

Figure 11: Males Population Estimates

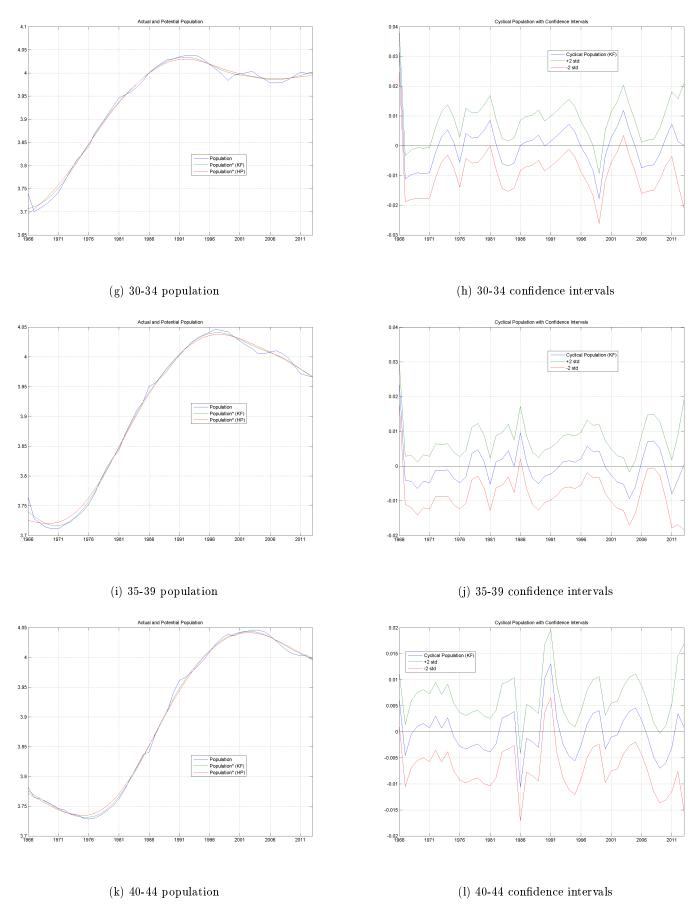
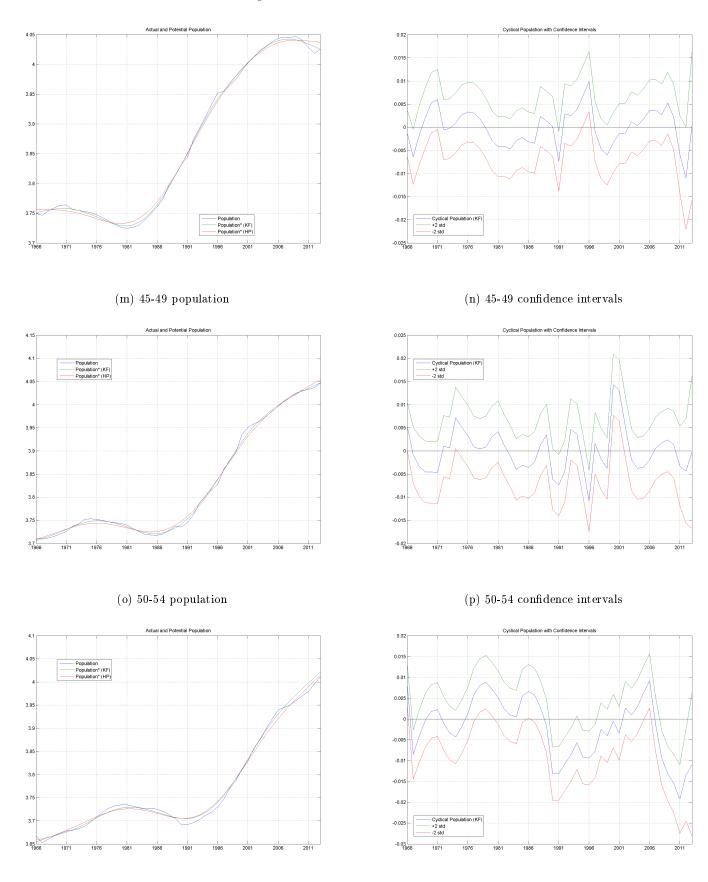


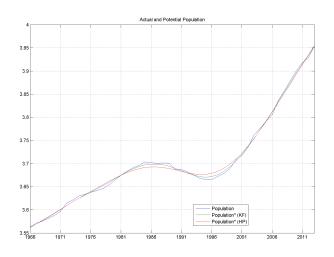
Figure 11: Males Population Estimates

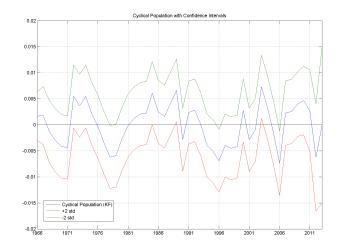


(r) 55-59 confidence intervals

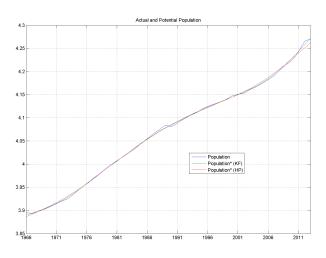
(q) 55-59 population

Figure 11: Maales Population Estimates

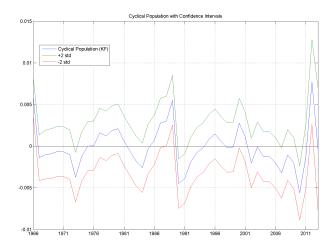




(s) 60-64 population



(t) 60-64 confidence intervals



(u) 65+ population

(v) 65+ confidence intervals