The Impact of Age Distribution on the Stock Price: A Nonparametric Approach

Cheolbeom Park*
National University of Singapore

December 2007

Abstract: This paper examines whether variations in demographic structure have influenced the stock price. It employs a nonparametric approach based on the Fourier Flexible Form (FFF) representation, which relates variations in the entire age distribution (instead of a single demographic measure focusing on one aspect of the demographic structure) to the normalized stock price, and which allows a flexible functional form to model the effects of the age distribution density. The main findings of this paper are that there is a significant impact from prime working age consumers on the stock price and that this impact is robust for all G5 countries (France, Germany, Japan, the UK, and the US). The fitted values from the nonparametric regression model closely track the normalized stock price. Finally, the prediction for the US normalized stock price is that it will decline for about a decade from 2009 or 2010 and then rise around 2020 when the relatively large cohorts born during the 1980s and early 1990s enter the prime working age stage.

JEL classification: G12, J11, C22

Key words: Age distribution, Fourier Flexible Form, Nonparametric regression, Stock price

---

* Address: Department of Economics, National University of Singapore, 1 Arts Link, Singapore 117570; Phone: +65-6516-6015; Fax: +65-6775-2646; E-mail: ecscp@nus.edu.sg.

1. Introduction

The variations in demographic structure have attracted many studies in various fields of economics because most developed economies are experiencing massive demographic changes and will be categorised as societies with an old population in the coming decades. Some recent examples of such studies are those on consumption and macroeconomic variables (Fair and Dominguez (1991) and Park, Shin, and Whang (2006)), economic growth (Bloom et al. (2007)), exchange rates (Rose and Supaat (2007)), and international capital flows (Domeij and Floden (2006)). Financial economics is no exception. In particular, studies on financial economics have examined whether the demand for financial assets, which is determined by variations in the demographic structure and different financial needs and decisions over the life-cycle, has affected the stock price. The basic theory is that consumers who are in the prime working age group have a strong demand for financial assets to prepare for their retirement period, while those who have retired usually dissave financial assets accumulated during their previous life-cycle stage. As a result, variations in the population age distribution result in different demand sizes for financial assets and thus variations in the demographic structure will be reflected in the stock price (see Abel (2001, 2003) and Geanakoplos, Magill, and Quinzii (2004) for theoretical discussions).

Although the theoretical prediction is clear, empirical results across studies have been mixed. Bakshi and Chen (1994), Geanakoplos, Magill, and Quinzii (2004), and DellaVigna and Pollet (2007)[1] find significant effects from demographic measures. However, Poterba (2001) argues, after exploiting various demographic measures, that no

[1] DellaVigna and Pollet (2007) show that demographics significantly affect cross-sectional stock returns at the industry level, and interpret this finding as evidence for the stock market's limited attention to relevant information.

significant relation between demographic structure and stock returns (or the stock price) has been found. Also, Ang, and Maddaloni (2005) claim that no significant relation has been found for US data, although the proportion of retired consumers has been found to have a significant negative impact on stock returns based on international data.

However, most previous studies make two implicit assumptions in their common statistical approach, a linear regression with one demographic measure. The first is that the selected demographic measure describes the variations of the entire age distribution fully, and the second is that the relation between the asset price and age distribution is linear. The findings of these studies could have problems resulting from these assumptions, which might be a reason for the mixed results and the incoherent interpretations across studies.

One demographic summary statistic usually describes one or just a few aspects of the entire population distribution. For example, the fraction of those in the prime working age group focuses on the effects from only those in that group, ignoring any possible effect from retired consumers, and vice versa. Furthermore, there is no clear line between the prime working age group and the retirement age group when selecting a demographic measure. It is well known that a substantial proportion of consumers retire at age 55 (the early retirement age for employer-provided pension plans), at age 62 (the early retirement age of the social security system), or at age 65 (the normal retirement age of the social security system) (see Gruber and Wise (1998)). As a result, the selection of only one or a few demographic measures from the various alternatives will be arbitrary and may be responsible for the conflicting results reported in the previous literature.

More importantly, the linear regression of the normalized stock price on a demographic summary has a potential misspecification problem, in addition to the possible spurious regression noted by Poterba (2001). It is well documented in the consumption literature that the optimal consumption rule is concave, and that the rules for optimal consumption and optimal asset allocation depend on the ages when consumers face uncertainty in their labor income. Also, studies in this literature find that both consumption and labor income show hump-shaped profiles against age, as a result of which total wealth grows nonlinearly over the working age period (see Carroll (1997), Gourinchas and Parker (2002), and Gomes and Michaelides (2005)). These findings imply that empirical studies examining the relation between demographic structure and the asset price should allow different impacts on the asset price determination from different ages; also that a relation more flexible than the linear one should be considered because it is unlikely that the asset market linearizes all the nonlinear patterns between these variables and age.

After presenting the statistical evidence for the misspecification of the linear regression, this paper relates variations in the normalized stock price to the variations of the probability density function of the age distribution (instead of a particular demographic measure) by employing a nonparametric regression model based on the Fourier Flexible Form (FFF). Since economic theories provide few guidelines for the functional form of the response function with respect to the population density function, the nonparametric regression based on the FFF is ideal because it allows a flexible functional form for the response.

Through an empirical analysis based on the nonparametric approach, this paper shows that the age response function is hump-shaped and significantly positive over the prime working ages, which is consistent with theoretical life-cycle models. Unlike the mixed results found in previous studies that used international data, the significant impact from prime working age consumers is robust in the four biggest non-US stock markets (France, Germany, Japan, and the UK) as well as in the US. Furthermore, the fitted values from the nonparametric regression model track the log price-dividend ratio or the log price-earnings ratio closely, even at times when the stock price rose too rapidly to be explained by fundamentals (for example, Japan in the 1980s and the US in the 1990s), or when the stock price suddenly fell (for example, Japan in the early 1990s). The effect from retired consumers is significantly negative, or at most insignificant, possibly due to the very slow decline of the wealth held by these people and the limitations of the age distribution data. Using the estimated coefficients from the regression model and US population projections from the US Census Bureau, the prediction for the normalized stock price is made to assess the impact of the retirement of the baby boomers. The normalized stock price is predicted to decline for about a decade from 2009 or 2010 and then to rise at around 2020 when the relatively large cohorts born during the 1980s and early 1990s enter the prime working age stage, in spite of the uncertainty regarding the magnitude of projected fluctuations.

The paper is organized as follows. Section 2 describes the US data and the evidence for the potential misspecification problem in the linear regression. Section 3 discusses the econometric methodology used in this study. Empirical results and the discussion of the results are presented in Section 4, and concluding remarks are offered in Section 5.

2. Data and the Misspecification Test of the Linear Regression

This section presents an explanation of the US data used in this study and the evidence of misspecification problems by exploiting a wide range of different measures of demographic structure used in existing studies.

When examining the relation between demographic structure and the aggregate asset price, this paper uses the stock price normalized by either dividends or earnings instead of using stock returns. There are two reasons for this: first, theoretical models such as in Abel (2001, 2003) and Geanakoplos, Magill, and Quinzii (2004) suggest that life-cycle consumption decisions under the overlapping generations set up affect the level of stock prices rather than stock returns; second, stock returns introduce noise into the low-frequency relation with demographic changes.

Data for the S&P 500 index, dividends, and earnings are taken from Robert Shiller's homepage[2] and are used to construct the log price-dividend ratio and log price-earnings ratio. The log price-dividend ratio is conceptually a better-normalized price but could have been affected by the change in firms' payout policies from dividend payment to the repurchase of shares (see Boudoukh et al. (2007) and Robertson and Wright (2006)). Therefore, I also consider the log price-earnings ratio to check whether the results are robust. The sample period is 1900 – 2006. Data for the age distributions of the US population are obtained from the homepage of the US Census Bureau.[3]

[2] The website address is http://www.econ.yale.edu/~shiller/.
[3] The website address is http://www.census.gov/.

Using the data described above, Table 1 shows linear regression results with a wide range of measures for demographic structure that have been used in previous studies. The results are from linear regressions with each of these measures as an independent variable in addition to a constant term. The dependent variable, either the log price-dividend ratio $(p_t - d_t)$ or the log price-earnings ratio $(p_t - e_t)$, is indicated in the first column of the table. The first two rows in each panel of the table show the estimates of the coefficient of each demographic summary statistic, associated standard errors estimated by the Newey and West (1987) procedure, and the goodness of fit ($\overline{R}^2$). Most demographic summary statistics are highly significant and can explain at least 15% of the variations in the log price-dividend ratio (5% for the log price-earnings ratio). The exceptions are the results for the fraction of the population between the ages of 20 and 64, which are insignificant for both dependent variables, and the ratio of the numbers of consumers between the ages of 40 and 64 to those over the age of 65, which is insignificant for the price-earnings ratio.

There are counterintuitive results across demographic summary statistics such as the positive coefficient for the fraction of the population between ages 40 and 64 and the negative coefficient for the ratio between the share of this age group and the share of the population over the age of 65. Furthermore, this regression has two potential econometric problems. First, since demographic measures are slowly moving and both the log price-dividend ratio and the log price-earnings ratio are well known to be persistent, the findings in the first two rows in each panel could be spurious. In order to check this possibility, I conducted the augmented Dickey-Fuller test for regression residuals; the results are shown in the third row of each panel.[4] The results suggest that the underlying regression may be misspecified for at least three demographic measures, when the dependent variable is the log price-dividend ratio. These are the fraction of the population between ages 20 and 64, the ratio of the numbers of consumers between the ages of 40 and 64 to those over the age of 20, and the ratio of numbers of consumers between the ages of 40 and 64 to those over the age of 65. When the dependent variable is the log price-earnings ratio, none of the regression is misspecified, mainly because the log price-earnings ratio is less persistent than the log price-dividend ratio.

Secondly, the relation between demographic statistics and the normalized stock price may not be linear. In the fourth row of each panel are the results of another misspecification test, checking whether nonlinear transformations of the demographic summary statistics have any explanatory power for the normalized stock prices. The regression equation specification error test of Ramsey (1969) is employed for this purpose. The square of the fitted value of the dependent variable is added to the original regression, and a significant coefficient for this newly added variable indicates the potential misspecification problem of the original linear regression. The fourth row of each panel present the robust t-statistic for the square of the fitted value of the dependent variable. The Ramsey test statistics are highly significant, suggesting that all linear regressions, except the one with the ratio of numbers of consumers between the ages of 40 and 64 to those over the age of 65 as a regressor, have evidence for misspecification at the 10% or higher significance level, regardless of the dependent variable.

[4] The augmented Dickey-Fuller test without a constant under the Ng-Perron procedure is conducted. The 10% and 5% critical values are -1.95 and -2.24, respectively. Similar Dickey-Fuller tests are also employed in Poterba (2001) and Geanakoplos et al. (2004) to detect a possible misspecification.

Even the regression with the ratio of numbers of consumers between the ages of 40 and 64 to those over the age of 65 has problems. The augmented Dickey-Fuller test already indicates that it is spurious when the log price-dividend ratio is the dependent variable. Also, when the log price-earnings ratio is used as the dependent variable, the above variable has no explanatory power. Therefore, the results in Table 1 would seem to warrant a reconsideration of the regression results on the relation between age structure and the stock price. The linear regressions in existing studies seem to be spurious, misspecified, or both.

3. Econometric Methodology

Instead of choosing a linear regression and one demographic measure from many alternatives, the econometric model considered in this paper is

$$y_t = \int_T f_t(s)g(s)ds + u_t \,, \tag{1}$$

where y_t is the normalized stock price, which is either the log price-dividend ratio or the log price-earnings ratio, f_t denotes the density function of age distribution at time *t*, and *T* is a common compact support for f_t. Since f_t, the regressor, is the density function of age distribution, $\int_{T_j} f_t(s)ds$ (where T_j is a subinterval of *T*) is the fraction of individuals for age group T_j in the total population. $g(s)$ can be interpreted as an age response function which reflects the impact of population age distribution on the normalized stock price.[5]

[5] Similar econometric models are considered in Fair and Dominguez (1991) and Park, Shin, and Whang (2006) to study the effect of population age distribution on aggregate consumption. However, Fair and

The above econometric model relates the variation in the normalized stock price to the variation of the entire age distribution instead of focusing just on a specific age range or a particular demographic measure. Once one has obtained the estimate of $g(s)$, one is able to see which age group has a significant impact on the movement of the stock price instead of choosing an arbitrary age group a priori. One can also see whether the estimated $g(s)$ is consistent with the predictions from theoretical models, and whether the movement of the stock price (especially the stock market boom during the 1990s) can be explained by variations in population age distribution. Also, one can make a projection about whether the stock price will fall dramatically after the baby boomers retire.

While estimating $g(s)$, no functional form is imposed for $g(s)$. The only assumption required for $g(s)$ is that it is smooth enough to be approximated by a series of polynomials, trigonometric functions, or a mixture of both series. That is, suppose that $\|g_\kappa - g\| \to 0$ as $\kappa \to \infty$, where $g_\kappa(s)$ is an approximation of $g(s)$ given by a combination of a finite series of functions $\varphi_1,...,\varphi_\kappa$. Since $g_\kappa(s) = \sum_{i=1}^{\kappa} \alpha_i \varphi_i$, equation (1) can be written as

$$
\begin{aligned}
y_t &= \int_T f_t(s) g(s) ds + u_t \\
&= \int_T f_t(s) g_\kappa(s) ds + u_{\kappa t} \\
&= \int_T f_t(s) \sum_{i=1}^{\kappa} \alpha_i \varphi_i(s) ds + u_{\kappa t} \\
&= \int_T \sum_{i=1}^{\kappa} \alpha_i f_t(s) \varphi_i(s) ds + u_{\kappa t}
\end{aligned}
$$

Dominguez (1991) impose a quadratic parametric restriction on $g(s)$, and Park, Shin, and Whang (2006) add a linear conintegration regression part to equation (1).

$$= \sum_{i=1}^{\kappa} \alpha_i \int_T f_t(s) \varphi_i(s) ds + u_{\kappa t}$$

$$= z_t' a_\kappa + u_{\kappa t} \quad (2)$$

where $u_{\kappa t} = u_t + \int_T f_t(s)(g_\kappa - g)(s) ds$, $z_t = [\int_T f_t(s)\varphi_1(s)ds, ..., \int_T f_t(s)\varphi_\kappa(s)ds]'$, and $a_\kappa = [\alpha_1, ..., \alpha_\kappa]'$. $\int_T f_t(s)\varphi_i(s)ds$ can be approximated by $\sum_{h=1}^{d} f_t(s_h)\varphi_i(s_h)$, where s_h is a number on *T*. Letting $Y = [y_1, ..., y_N]'$ and $Z = [z_1', ..., z_N']'$, the LS estimator of a_κ can be written as

$$\hat{a}_\kappa = (Z'Z)^{-1} Z'Y . \quad (3)$$

As a result, the corresponding series estimator of $g(s)$ can be written as

$$\hat{g}(s) = \sum_{i=1}^{\kappa} \hat{\alpha}_i \varphi_i(s) = \Pi_\kappa' \hat{a}_\kappa , \quad (4)$$

where $\Pi_\kappa = [\varphi_1, ..., \varphi_\kappa]'$.

Once $Var(\hat{a}_\kappa)$ is obtained from the regression residuals with the use of the Newey and West (1987) procedure, $Var(\hat{g}(s))$ can be written as

$$Var(\hat{g}(s)) = \Pi_\kappa' Var(\hat{a}_\kappa) \Pi_\kappa . \quad (5)$$

In order to approximate $g(s)$, the FFF is employed. The FFF, which was introduced by Gallant (1981), extends the traditional Fourier theorem. The FFF expansion of $g(s)$ can be written as

$$g_\kappa(s) = \beta_0 + \beta_1 s + \beta_2 s^2 + \sum_{j=1}^{J} [\delta_j \cos(js) + \gamma_j \sin(js)], \quad (6)$$

where the vector of parameters is $\theta = [\beta_0, \beta_1, \beta_2, \delta_1, \gamma_1, ..., \delta_J, \gamma_J]'$ and $\kappa = 3 + 2J$. It is worthwhile to note the robustness of the FFF approach. Because economic theories provide few guidelines for $g(s)$, except the conjecture that $g(s)$ might be positive over a prime working age group and negative for young and retirement age groups, the FFF is ideal because it approximates $g(s)$ under a flexible representation. The trigonometric functions in the FFF representation are also well suited for characterizing the effect of the periodic pattern of the birth rates over the last century, as illustrated in Geanakoplos, Magill, and Quinzii (2004). Finally, the asymptotic normality of LS estimators for the FFF representation was developed by Andrews (1991).

Due to the characteristics of trigonometric functions, it is desirable to scale the data into the interval [0,1]. That is, with a given common support $T=[\lambda_1,\lambda_2]$ for f_t, f_t is transformed by $f_t^*(s)= f_t(\lambda_1+(\lambda_2-\lambda_1)s)$, so that f_t^* has the common support [0,1]. The original response function *g* with respect to f_t can be recovered from the response function g^* with respect to f_t^* by the transformation $g(s)=g^*((s-\lambda_1)/(\lambda_2-\lambda_1))$.

Since we cannot expand $g(s)$ with an infinite number of terms, it is important to decide κ (or equivalently *J* in equation (6)) so that the FFF representation of $g(s)$ gives a good approximation in the empirical analysis. Since the dependent variable (*Y*) and regressors (*Z*) are assumed to have no stochastic trend, the *h*-block cross-validation (*CV*) and the modified *h*-block *CV* criteria, which are suggested by Burman, Chow, and Nolan (1994) and Racine (1997), are used as selection criteria for κ. For a given block size (*h*), the *h*-block *CV* criterion can be written as

$$CV = N^{-1}\sum_{t=h}^{N-h}(y_t - z_t'\hat{a}_\kappa(t,h))^2, \tag{7}$$

where $\hat{a}_\kappa(t,h)$ is the estimator of a_κ obtained by removing the *t*-th observations in *Y* and *Z* and *h* observations preceding and following the *t*-th observations in both *Y* and *Z*. The modified *h*-block *CV* criterion, motivated by cases where κ / N is not negligible, can be written as

$$MCV = N^{-1}\sum_{t=h}^{N-h}(y_t - z_t'\hat{a}_\kappa(t,h))^2 + N^{-2}\sum_{t=h}^{N-h}\sum_{n=1}^{N}(y_n - z_n'\hat{a}_\kappa(t,h))^2$$

$$+ N^{-1}\sum_{n=1}^{N}(y_n - z_n'\hat{a}_\kappa)^2 \ . \qquad (8)$$

The κ that minimizes the above *CV* criteria is chosen.

The results for the *CV* criteria are presented in Table 2. In this table and the empirical analysis in the next section, I consider 50 age groups: ages 26, 27,…, 74, and ≥ 75, although the population estimates program at the US Census Bureau has published the US population estimates by single year age groups of 76 or more since 1900. I choose the first age group, those at age 26, because common stock holdings by consumers in their early 20s or younger are negligible compared with other age groups (see Poterba, 2001) and might be influenced by their parents, especially in the case of teenagers. The last age group, those at age 75 or above, is dictated by the availability of population estimates in 1900. The US Census Bureau published the US population estimates by 76 age groups (ages 0, 1, …, 74, ≥ 75) between 1900 and 1939. It then published the US population estimates by 86 age groups (ages 0, 1, …, 84, ≥ 85) between 1940 and 1979, and the number of age groups has been 101 (ages 0, 1, …, 99, ≥ 100) since 1980. Due to this limitation in population data, those at age 75 or above are put in the last age group to make the population data go back as far as possible. Hence, the total population is the population at least 26 years of age.

Each column in Table 2 shows the results for each alternative value for κ. Since *h* observations either side are removed to construct *CV* criteria, these are appropriate model selection criteria for observations from a general stationary series. Also, since κ in the FFF expansion is supposed to increase with *N*, both criteria are free from the problems noted by Shao (1993). The block size, *h*, is set as the nearest integer to $N/6$, which is 18, in accordance with the suggestion of Burman, Chow, and Nolan (1994). Since both statistics for *CV* criteria are minimized when $\kappa = 5$ (or equivalently $J = 3$) regardless of the dependent variable, I choose $\kappa = 5$ in the empirical analysis. In other words, $g(s)$ is estimated by the FFF expansion, which includes a constant, s, s^2, cos(s), and sin(s). Finally, when $\kappa = 5$, the augmented Dickey-Fuller test statistics for the regression residuals are -5.7513 (-5.3035) for the log price-dividend ratio (the log price-earnings ratio), which implies that there is no spurious relation for the nonparametric regression.

4. Empirical Analysis

4. 1. Evidence from the US Data

This sub-section reports the estimation results of the nonparametric regression model in equation (1) with the US data. Since the estimates of the individual parameters in the FFF expansion are void of economic interpretation,[6] the estimated age response function ($\hat{g}(s)$) and the fitted value from the nonparametric regression model are plotted in Figure 1 for both the log price-dividend ratio and the log price-earnings ratio.

First, note that the estimated age response functions for both normalized stock prices show that different age groups have different impacts on the determination of the aggregate stock price. More specifically, the estimated age response function is

---

[6] The estimates of the individual parameters in the FFF expansion and their robust standard errors are available from the author upon request.

significantly positive and hump-shaped over prime working ages between 37 and 56 (between 36 and 56) for the log price-dividend ratio (the log price-earnings ratio). Following the prime working ages, the estimated age response function becomes significantly negative and U-shaped over retirement ages between 59 and 70 (between 60 and 70) for the log price-dividend ratio (the log price-earnings ratio). These patterns of the age response function appear consistent with the distinct financial needs and asset allocation decisions at different life-cycle stages of consumers, as predicted by life-cycle models. It may seem strange that the peak of the age response function over the prime working ages appears in the late 40s, although the amount of wealth is predicted to reach its peak around the retirement age. However, life-cycle models such as that of Gomes and Michaelides (2005) also predict that the share of wealth invested in the stock market declines with age from the mid 30s for forward-looking consumers with labor income uncertainty. Hence, changing total wealth and the optimal share of wealth invested in the stock market together could explain the hump-shaped age response function over prime working ages. The significant age response functions imply that variations in age distribution play an important role in explaining the movements of the log price-dividend ratio or the log price-earnings ratio.

Second, consistent with these significant age response functions, the fitted values of the aggregate normalized stock price are impressive, as shown in the right-hand side panels of Figure 1. The fitted values from the nonparametric regression model track the actual movements of the normalized stock prices remarkably closely, even if a myriad of factors that affect the normalized stock price are ignored. This is particularly true for the boom of the stock market during the 1990s when the fitted values from the linear regressions with a demographic measure often lag behind the actual movements of the normalized stock prices. The $\overline{R}^2$ is 0.74 (0.46) when the log price-dividend ratio (the log price-earnings ratio) is the dependent variable, and is higher than any $\overline{R}^2$ from the linear regressions reported in Table 1. This suggests that the fact that baby boomers enter the prime working age group is a reason for the surge of the stock price during the 1990s.

Third, although the shape of the age response functions in Figure 1 is generally consistent with the predictions from the life-cycle consumption models or overlapping generations models, the behavior of the last age group is odd. The age response function for the log price-dividend ratio (the log price-earnings ratio) rises rapidly and becomes significantly positive from the age of 72 (the age of 73) at the 5% significance level. Even with the fact that financial assets held by older consumers decline very slowly (see Poterba (2001)), it is hard to justify the eccentric behavior of the age response function at the age of 72 or above. This part of the age response function implies that older consumers suddenly become aggressive buyers in the stock market. However, this odd behavior of the age response function could be an artifact of the traditional way of reporting the population age distribution. That is, everyone of age 75 or above is categorized into one single age group, the last one, and this group is treated equally in the empirical analysis. The proportion of consumers in this group is far higher than in any other age group in recent years, due to developments in medical science. For example, the fraction of consumers in the last age group in 2006 is 6.1%, while the fraction of consumers in the second to last age group (those at age 74) is a mere 0.5%. As a result, the last age group has a larger impact on the asset price determination. Hence, the behavior of the estimated age response function in the last age group may not be due to asset demand over the life-cycle, but a consequence of the fact that the last age group of the age distribution has a different nature.

In order to check this possibility, I conduct the same econometric analysis with the US data since 1940. From this year, the US Census Bureau began to publish the US population estimates by 86 age groups (ages 0, 1, …, 84, ≥ 85). This change in publishing the US population estimates would be expected to cause the eccentric behavior of the age response function to occur at later age groups than those reported in Figure 1. Figure 2 presents the results with the US data since 1940. Consistent with the conjecture, the age response function rises more slowly and its rapid rise can be observed from the ages of the mid-70s rather than the early 70s. Furthermore, with the data since 1940, the age response function is also hump-shaped over prime working ages and is significant at ages between 34 and 62 (between 33 and 56) for the log price-dividend ratio (the log price-earnings ratio). Consumers at age 30 or younger now have a significantly negative age response for each case. Finally, the fitted value from the nonparametric regression model is also remarkable and the $\overline{R}^2$ is 0.87 (0.65) when the log price-dividend ratio (the log price-earnings ratio) is the dependent variable.

Fourth, although Figures 1 and 2 show a remarkable relation between the age distribution and the normalized stock price, one might still be skeptical of the relation because of movements in the interest rate. If the rapid rise in the aggregate stock price during the 1990s is a consequence of baby boomers' savings, these savings should also cause the interest rates to be historically low, which was not observed during the 1990s. Indeed, the upper panels in Figure 3 show that there is no tight joint dynamics of the interest rate[7] and the age distribution. The age response function is mostly insignificant and the nonparametric regression model explains very little of the variation in the short-term interest rate. The $\overline{R}^2$ is merely 0.01.

However, Barsky (1989) demonstrates that changes in risk or productivity growth can cause an independent movement of the interest rate and stock price in a general equilibrium model. Barsky (1989) shows that an increase (decrease) in risk lowers (raises) the interest rate due to the strengthened (weakened) precautionary saving, while the effect of the change in risk on the stock price is ambiguous because of two conflicting pressures. An increase (decrease) in risk lowers (raises) the demand for the risky stock, but the lower (higher) interest rate due to the change in risk raises (lowers) the stock price at the same time. As a result, the net effect of the change in risk on the stock price depends on the preference parameter values and/or the existence of other risky assets.[8]

Barsky (1989) characterizes the 1970s as a period with higher risk and concludes that it is possible to observe both a low real interest rate and a low stock price simultaneously. If the 1970s can be characterized as a period of high risk, then the 1990s can be characterized as a period of low risk.[9] The lower panels of Figure 3 show movements of the growth rate of real GDP per capita since 1952 in the US[10] and its standard deviations for non-overlapping five-year periods. These panels clearly show that the risk estimated as the standard deviation of real GDP per capita for non-overlapping five-year periods was high during the 1970s but extremely low during the 1990s. The

[7] The short-term real interest rate is also obtained from Robert Shiller's homepage.
[8] Barsky (1989) also shows that changes in the productivity growth have a qualitatively identical effect on the interest rate and stock price to that of changes in risk.
[9] The decline of US GDP volatility has been noted in macroeconomic literature such as Kim and Nelson (1999) and McConnell and Perez-Quiros (2000).
[10] The data for the US real GDP are taken from FRED of FRB of St. Louis. The website address is http://research.stlouisfed.org/fred2/.

decrease in risk during the 1990s could have raised the interest rate due to weakened precautionary saving and this might have offset the effect from the large savings of baby boomers. Furthermore, since the effect of the decrease in risk on the stock price is ambiguous, the effect of baby boomers' savings can be observed more clearly in movements of the stock price than the interest rate.

In summary, the age response function estimated from the nonparametric regression model is generally consistent with the predictions from theoretical models. Consumers in prime working age groups have a significantly positive impact on the stock price movements. However, one needs to be careful with the interpretation of the age response function for the last age group due to the limitations of the data, which seems impossible to resolve without a full description of the whole population up to the oldest person in the economy. Finally, the joint dynamics of the interest rate and age distribution may not be as clearly observable as that of the stock price and age distribution due to variations in the risk of the economy.

4.2. Evidence from International Data

Although there is a cyclical pattern in the birth rates of the US, as observed by Geanakoplos, Magill, and Quinzii (2004), the population age distribution is slowly moving, which implies that the remarkable relation between the stock price and age distribution in the US may suffer from a small sample problem, even with more than one hundred years of data. Because of this limitation, many studies, such as Poterba (2001), Geanakoplos, Magill, and Quinzii (2004), and Ang and Maddaloni (2005), examine data from other countries that have well-developed stock markets. The results are usually mixed, preventing a general conclusion, which means that a demographic summary statistic which has a strong explanatory power for one stock market often loses its power in another market. However, this problem might be a consequence of an arbitrary choice of demographic measure and the linear regression. Hence, this section also examines the relation between the stock price and age distribution with data from the four biggest non-US stock markets (France, Germany, Japan, and the UK) using the nonparametric regression model.

The log price-dividend ratio is constructed from the dividend yield data taken from Global Financial Data (GFD).[11] I choose the log price-dividend ratio for the analysis because the data series for the dividend yield are longer than the price-earnings ratio for all four countries in GFD, and the price-earnings ratio series are available only from 1971 in France and from 1969 in Germany.

The population estimates for France, Germany, and the UK are obtained from Eurostat.[12] Although the population estimates by 86 single year age groups (ages 0, 1, …, 84, ≥ 85) for France and Germany are available from 1950 and 1951 respectively, the population estimates by five year age groups (e.g., ages between 0 and 4, ages between 5 and 9, and so on) are available for the UK during the 1950s and 1960s. The population estimates by single year age groups of 86 for the UK are only available from 1972. As a result, the sample periods for France and Germany are 1950 to 2006 and 1951 to 2006 respectively, whereas the sample period for the UK is 1972 to 2006.

[11] The website address is http://www.globalfindata.com/index.php3.
[12] http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL is the website address.

The population estimates for Japan are taken from the homepage of the Japanese Statistics Bureau.[13] Although the population estimates by single year age groups begin to be available from 1920, the estimate series is broken from 1941 to 1943. Another data problem with Japanese population estimates is that figures for year 2005 population estimates are not found. Since the dividend yield series in GFD has a break from 1944 to 1948, the sample period of Japan is 1949 to 2004 and the population estimates have 80 age groups (ages 0, 1, …, 78, ≥ 79).[14]

Figure 4 presents the estimated age response functions for all four countries. Unlike previous studies' results that the explanatory power of a demographic measure differs greatly across countries, the age response function estimated from the nonparametric regressions is significantly positive over the prime working ages (from the mid-30s to the mid-50s) for all four countries. Also, the age response function is hump-shaped over the prime working ages, which is consistent with the results for the US. The age response function is negative over retirement ages for all four countries but the response function is significantly negative only for Japan and the UK. The rapid rise of the age response function in the last age group is found in three of the four countries (not the UK), but the standard errors are rising rapidly in the last age group for the UK, possibly due to the relatively small number of observations. The rapid rise of the age response function in the last age group is particularly pronounced for Japan where the fraction of old people is highest.

---

[13] The website address is http://www.stat.go.jp/english/.

[14] Although the number of single year age groups in Japanese population estimates is usually 85 or 90, the number of single year age groups in the 1952 population estimates is 79. Hence, the age distributions for other years are adjusted accordingly.

Figure 5 presents the fitted values of the log price-dividend ratio from the nonparametric regressions. The fitted values track the movements of the log price-dividend ratio particularly well in Japan and the UK, especially the rapid rise in the stock price during the 1980s and its sudden fall in the early 1990s in Japan. The $\bar{R}^2$s are 0.13, 0.20, 0.87, and 0.68 for France, Germany, Japan, and the UK respectively. Overall, the variations in the age distribution, especially consumers in prime working age groups, have played a crucial role in explaining the movements of stock prices in all international data considered in this sub-section.

4.3. Projection of the Aggregate Stock Price

One of the reasons to study the relation between the population age distribution and the aggregate stock price is a concern about whether the aggregate stock price will fall in the US, when the baby boomers enter the retirement stage, as dramatically as it rose during the 1990s. Using the estimates of the individual parameters in the FFF expansion and population projections provided by the US Census Bureau, this sub-section addresses this concern. The population projections used in this study are interim projections consistent with Census 2000, and these projections for the period from 2000 to 2050 can be found at the US Census Bureau website. Since realized stock market data are available up to 2006, the projections of the normalized stock price are for the period from 2007 to 2050.

Four sets of estimated coefficients in the FFF expansion are used for the projection, depending on the dependent variable of the estimation (the log price-dividend ratio or the log price-earnings ratio) and data for age distribution (the distribution for age groups between 26 and ≥ 75 and the distribution for age groups between 26 and ≥ 85).

As a result, four projections are formed, and are presented in Figure 6. The time (year 2007) when the projection starts is indicated as the vertical line in each panel.

The projected log price-dividend ratio and the log price-earnings ratio using the first age distribution (age groups between 26 and ≥ 75) are shown in the left-hand side panels of Figure 6. Both normalized stock prices are expected to begin to decline from the year 2009 until 2019 to the level of those in the early 1990s, and are expected to pick up after that point. When the other set of age distribution (age groups between 26 and ≥ 85) is used, projections of both normalized stock prices begin to decline from the year 2009 until 2019 to the level of those in the early 2000s instead of the 1990s. Again, both projections begin to pick up around 2020. Although there is a difference in the magnitude of the rise and fall of the projected normalized stock prices due to the difference in the estimated age response functions in Figures 1 and 2, the projected timing for the rise and fall of the stock market coincides almost perfectly. The rise in the stock price around 2020 appears to be related to the relatively high birth rates during the 1980s and early 1990s. This generation enters the prime working age stage at around that time.

However, this projection based on the age distribution should not be used for forecasting short-term fluctuations of the stock price because a myriad of other factors (economic growth, risk, etc.) are not considered. Moreover, the projections should be interpreted with caution because they implicitly assume a closed economy for the US. The ongoing globalization phenomena could break the tight historical relation between the population age distribution and the stock price, and create a new relation between the joint demography of countries and the stock price.

5. Conclusion

In this paper, a nonparametric regression model based on the FFF expansion is proposed to examine whether variations in demographic structure have influenced the stock price. The proposed method provides a number of improvements over existing studies. For example, the method relates variations in the entire age distribution (instead of one demographic measure focusing on a single aspect of the demographic structure) to the normalized stock price. Also, the nonparametric regression based on the FFF allows a flexible functional form for the effects of the age distribution density function.

The main findings of this paper are that the estimated age response function is hump-shaped and significantly positive over the prime working ages. Unlike the mixed results with international data in previous studies, the significant impact from prime working age consumers is robust for all G5 countries (France, Germany, Japan, the UK, and the US). Furthermore, the fitted values from the nonparametric regression model track the log price-dividend ratio or the log price-earnings ratio remarkably well, although factors other than variations in the age distribution function are not considered. The effect from consumers in the last age group is not clearly estimated due to the limitation of the age distribution data. Finally, the prediction for the normalized stock price made by the estimated coefficients from the nonparametric regression model and the US population projections by the US Census Bureau states that the normalized stock price will decline for about a decade from 2009 or 2010 and then rise around 2020 when the relatively large cohorts born during the 1980s and early 1990s enter the prime working age stage.

References

Abel, A. B. (2001) "Will Bequests Attenuate the Predicted Meltdown in Stock Prices When Baby Boomers Retire?" *Review of Economics and Statistics*, 83:4, 589-595.

Abel, A. B. (2003) "The Effects of a Baby Boom on Stock Prices and Capital Accumulation in the Presence of Social Security" *Econometrica*, 71:2, 551-578.

Andrews, D. K. (1991) "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models" *Econometrica*, 59:2, 307-345.

Ang, A. and A. Maddaloni (2005) "Do Demographic Changes Affect Risk Premiums? Evidence from International Data" *Journal of Business*, 78:1, 341-379.

Bakshi, G. and Z. Chen (1994) "Baby Boom, Population Aging, and Capital Markets" *Journal of Business*, 67:2, 165-202.

Barsky, R. B. (1989) "Why Don't the Prices of Stocks and Bonds Move Together?" *American Economic Review*, 79:5, 1132-1145.

Bloom, D. E., D. Canning, G. Fink, and J. E. Finlay (2007) "Does Age Structure Forecast Economic Growth?" NBER working paper 13221.

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts (2007) "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing," *Journal of Finance*, forthcoming.

Burman P., E. Chow, and D. Nolan (1994) "A Cross-Validatory Method for Dependent Data," *Biometrika*, 81:2, 351-358.

Carroll, C. D. (1997) "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, CXII:1, 111-148.

DellaVigna, S., and J. M. Pollet (2007) "Demographics and Industry Returns," *American Economic Review*, forthcoming.

Domeij, D., and M. Floden (2006) "Population Aging and International Capital Flows," *International Economic Review*, 47:3, 1013-1032.

Fair, R. C., and K. M. Dominguez (1991) "Effects of the Changing US Age Distribution on Macroeconomic Equations," *American Economic Review*, 81:5, 1276-1294.

Gallant, A. R., (1981) "On the Basis in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211-245.

Geanakoplos, J., M. Magill, and M. Quinzii (2004) “Demography and the Long-Run Predictability of the Stock Market,” *Brookings Papers on Economic Activity*, 1:2004, 241-325.

Gomes, F., and A. Michaelides (2005) “Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence,” *Journal of Finance*, 60:2, 869-904.

Gourinchas, P.-O. and J. A. Parker (2002) “Consumption over the Life Cycle,” *Econometrica*, 70:1, 47-89.

Gruber, J., and D. Wise (1998) “Social Security Programs and Retirement Around the World,” NBER Working Paper 6134.

Kim, C.-J., and C. R. Nelson (1999) “Has the US Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle,” *Review of Economics and Statistics*, 81:4, 608-616.

McConnell, M. M., and G. Perez-Quiros (2000) “Output Fluctuations in the United States: What Has Changed since the Early 1980’s?” *American Economic Review*, 90:5, 1464-1476.

Newey, W., and K. West (1987) “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55:3, 703-708.

Park, J. Y., K. Shin, and Y.-J. Whang (2006) “A Semiparametric Cointegrating Regression: Investigating the Effects of Age Distribution on Consumption and Saving,” working paper, Seoul National University.

Poterba, J., (2001) “Demographic Structure and Asset Returns,” *Review of Economics and Statistics*, 83: 4, 565-584.

Racine, J., (1997) “Feasible Cross-Validatory Model Selection for General Stationary Processes,” *Journal of Applied Econometrics*, 12:2, 169-179.

Ramsey, J. B. (1969) “Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis,” *Journal of the Royal Statistical Society* (Series B), 31:2, 350-371.

Robertson, D., and S. Wright (2006) “Dividends, Total Cashflow to Shareholders and Predictive Return Regressions,” *Review of Economics and Statistics*, 88:1, 91-99.

Rose, A. K., and S. Supaat (2007) “Fertility and the Real Exchange Rate,” NBER working paper 13263.

Shao, J. (1993) “Linear Model Selection by Cross-Validation,” *Journal of the American Statistical Association*, 88: 422, 486-495.

Table 1. Misspecification Test of the Linear OLS Regression

		Median Age	Average Age of those 20+	Fraction of Population 20-64	Fraction of Population 40-64	Fraction of Population 65+	(Population 40-64) / (Population 20+)	(Population 40-64) / (Population 65+)	(Population 40-49) / (Population 20-29)
$p_t - d_t$	β	0.0750 (0.0201)	0.1172 (0.0311)	2.6942 (3.4349)	8.1565 (2.7371)	8.7944 (2.2808)	5.5654 (2.3794)	-0.2272 (0.0634)	1.5854 (0.3760)
	$\bar{R}^2$	0.3811	0.3833	0.0196	0.3161	0.3894	0.1397	0.2880	0.4207
	ADF test	-2.5641	-2.3411	-1.3667	-2.3565	-2.3208	-1.5806	-1.8604	-2.2925
	Ramsey test	4.6860	7.3078	4.4282	4.1019	2.3633	3.6978	0.4514	5.4684
$p_t - e_t$	β	0.0406 (0.0169)	0.0596 (0.0255)	1.3259 (2.5734)	5.1151 (2.1204)	4.0049 (2.0918)	4.6215 (1.8581)	-0.0883 (0.0618)	1.2148 (0.2760)
	$\bar{R}^2$	0.1334	0.1175	-0.0007	0.1504	0.0938	0.1190	0.0466	0.3059
	ADF test	-3.9013	-3.8292	-3.4826	-3.9476	-3.7509	-3.7964	-3.6064	-4.4746
	Ramsey test	3.1350	3.4872	3.8897	2.5628	1.6565	2.5744	-0.2190	1.6847

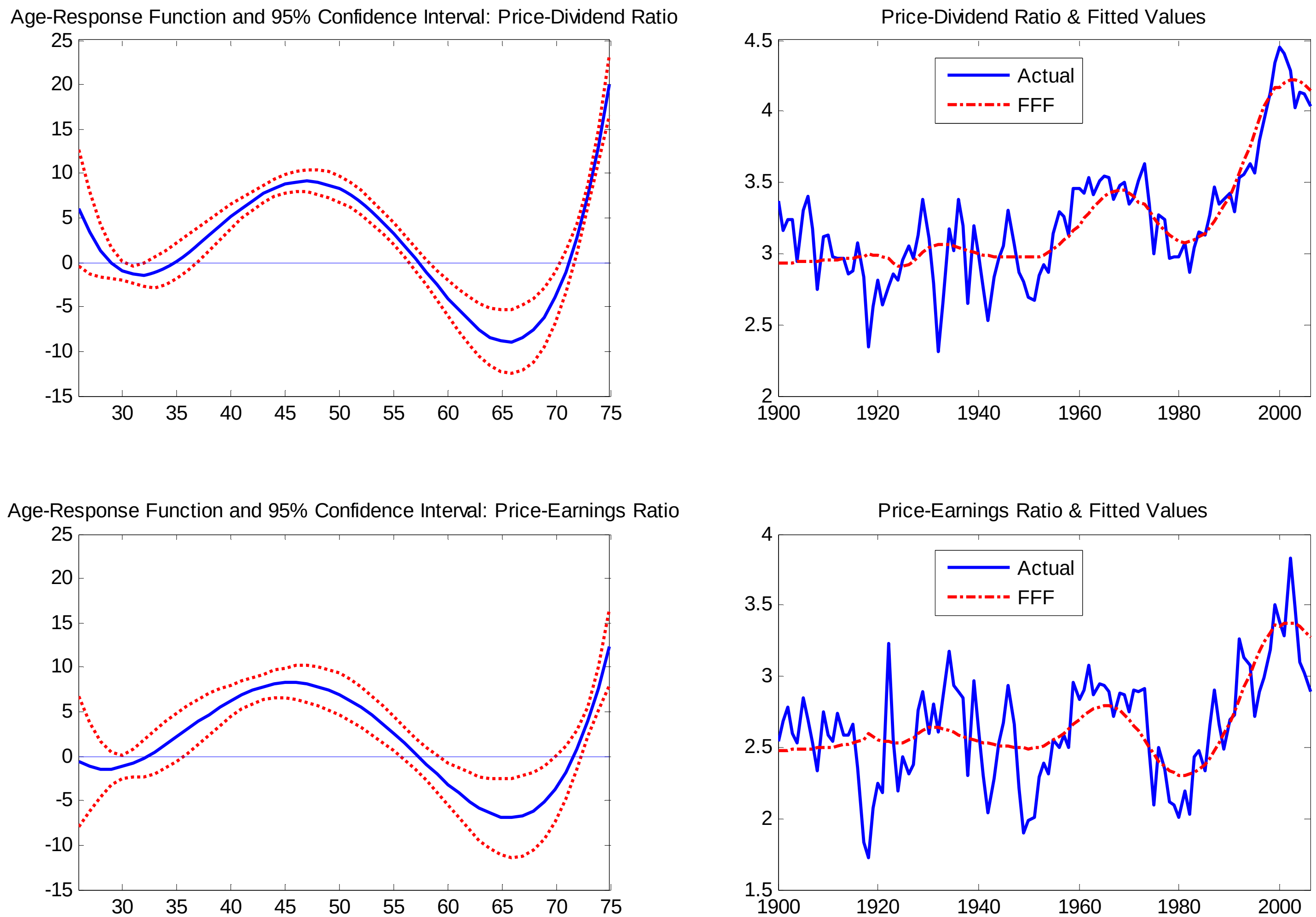
Note: The first two rows in each panel show the results from the linear regression, $y_t = \alpha + \beta x_{it} + \varepsilon_{it}$. Variables for x_{it} are listed in each column and standard errors for β are estimated by the Newey-West procedure with four lags and shown in parentheses. Results are not sensitive to the choice of the number of lags in the Newey-West procedure. The augmented Dickey-Fuller (ADF) test with the Ng-Perron procedure is conducted for regression residuals. The Ramsey test shows the robust t-statistic for γ in $y_t = \alpha + \beta x_{it} + \gamma \hat{y}_t^2 + \varepsilon_{it}$, where $\hat{y}_t^2$ is the square of the fitted value of y_t from the above linear regression. A significant coefficient for γ indicates the potential misspecification problem of the original linear regression.

Table 2. *h*-block and Modified *h*-block Cross-Validation Criteria

		$\kappa = 3$ ($J = 0$)	$\kappa = 5$ ($J = 1$)	$\kappa = 7$ ($J = 2$)	$\kappa = 9$ ($J = 3$)	$\kappa = 11$ ($J = 4$)	$\kappa = 13$ ($J = 5$)
$p_t - d_t$	*h*-block *CV*	0.4339	0.1157	41.2870	2.6900	336.0529	13123
	Modified *h*-block *CV*	0.6937	0.2088	80.4123	5.0676	614.8085	26100
$p_t - e_t$	*h*-block *CV*	0.2098	0.0586	28.1318	1.7800	236.7252	824.7
	Modified *h*-block *CV*	0.4436	0.1870	54.1044	3.5427	431.8795	16456

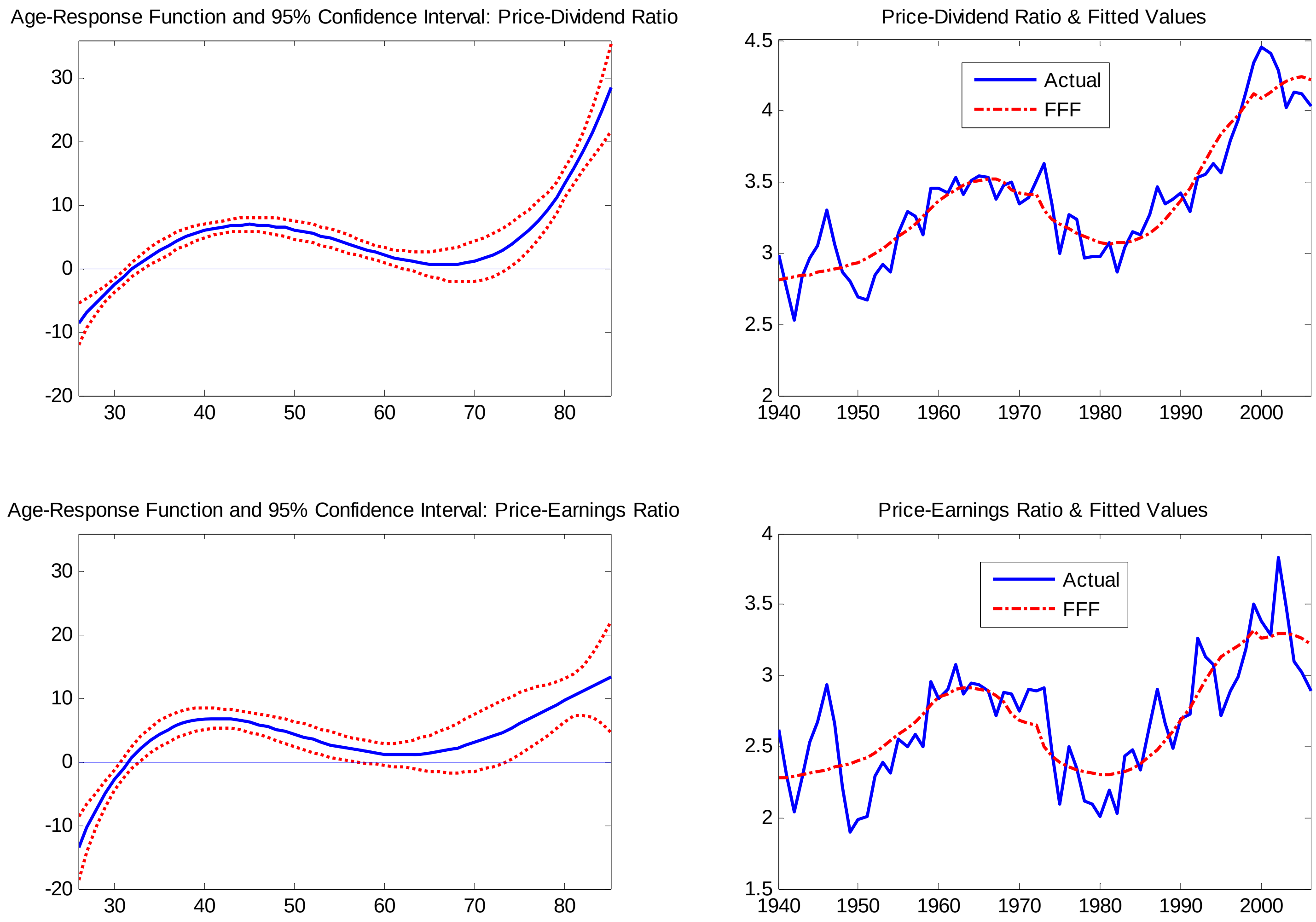
h-block cross-validation and modified *h*-block cross-validation are data-dependent criteria to select the optimal order of κ (or equivalently J in equation (6)) of the FFF representation. Statistics for *h*-block cross-validation and modified *h*-block cross-validation are computed from equations (7) and (8), respectively. The κ that minimizes the above *CV* criteria is chosen. See Burman, Chow, and Nolan (1994), and Racine (1997) for further discussion.

Figure 1. Age Response Functions and Fitted Values: Nonparametric Regression and Age Distribution 26-75



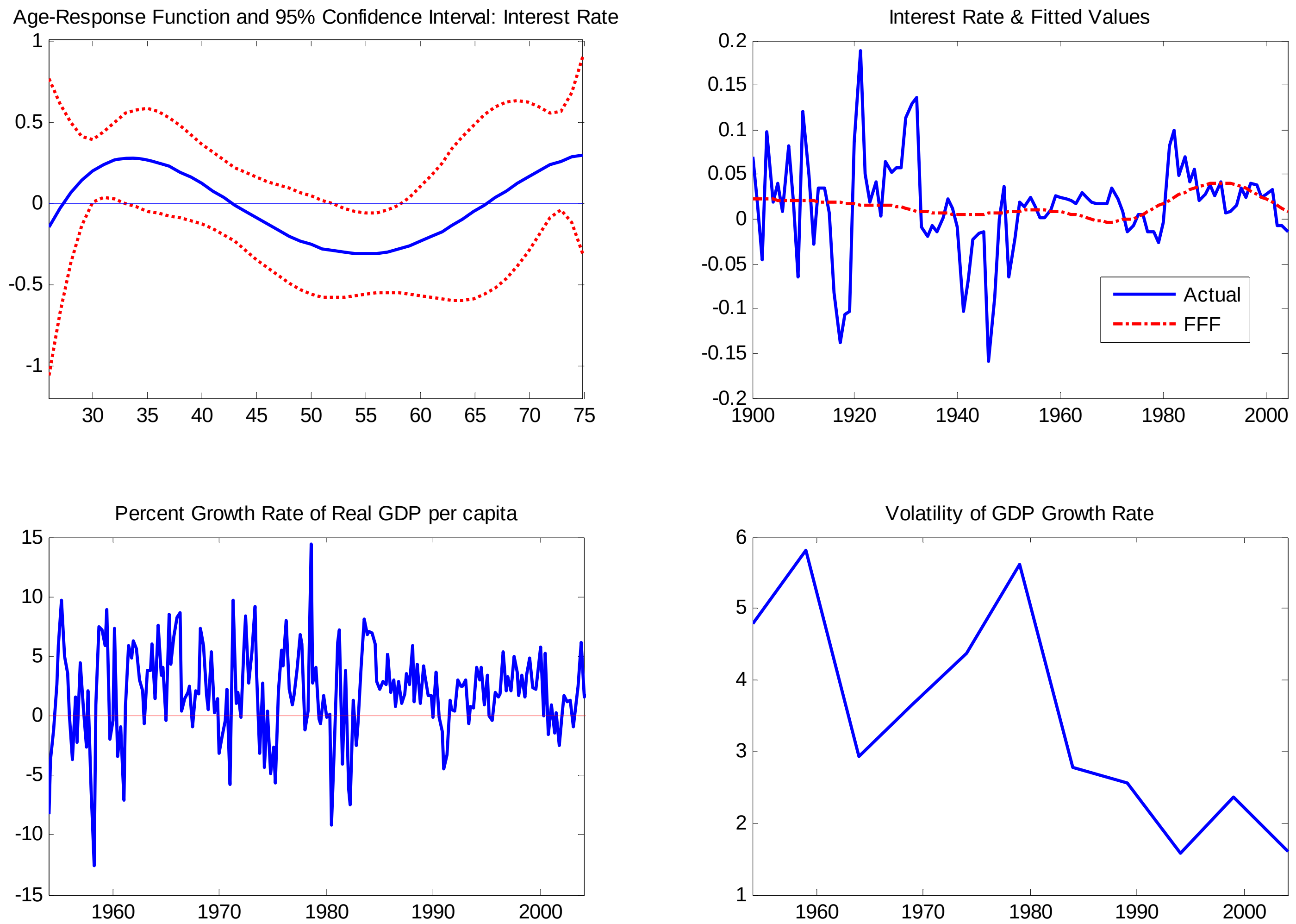
Note: This figure shows the estimates of $g(s)$ and the fitted values of the normalized stock price from the nonparametric regression in equation (1). US population estimates by single year age groups (ages 26, 27, …,74, and ≥ 75) are used.

Figure 2. Age Response Functions and Fitted Values: Nonparametric Regression and Age Distribution 26-85



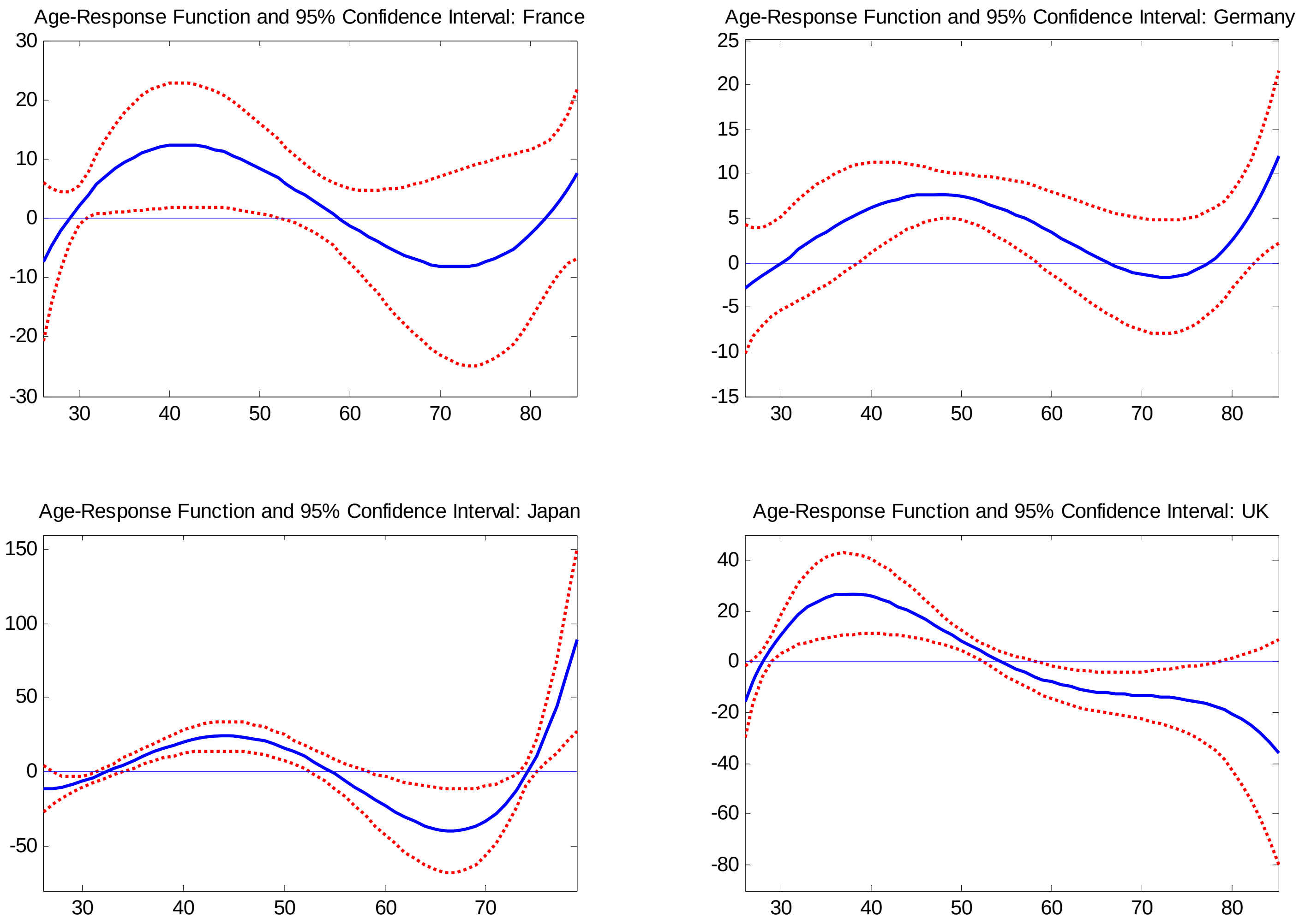
Note: This figure shows the estimates of $g(s)$ and the fitted values of the normalized stock price from the nonparametric regression in equation (1). US population estimates by single year age groups (ages 26, 27, …, 84, and ≥ 85) are used.

Figure 3. Nonparametric Regression with the Interest Rate and GDP Volatility



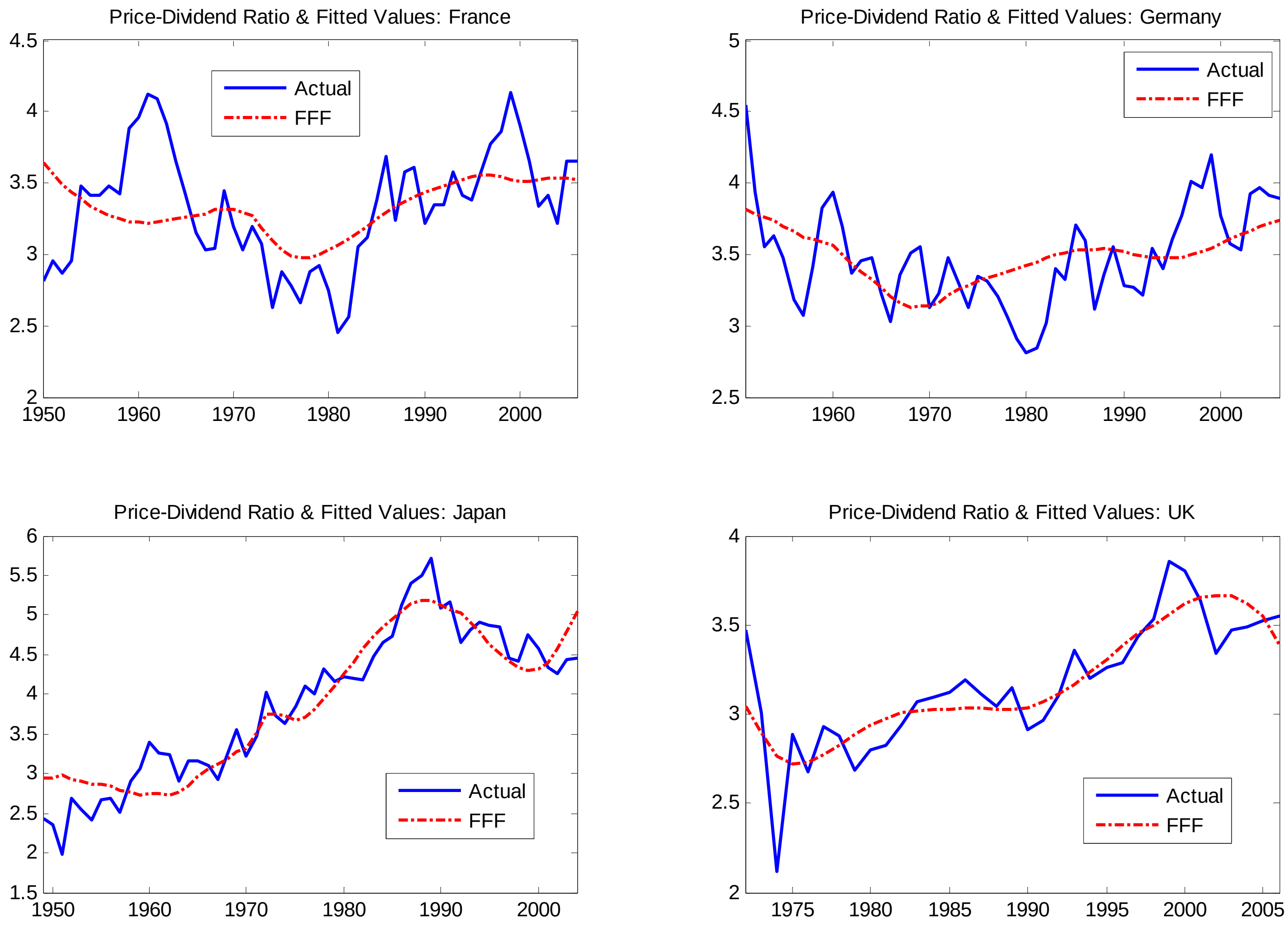
Note: The upper panels show the results of the nonparametric regression with the short-term interest rate. The lower panels show the movements of the annualized GDP growth rate and its volatility measured by non-overlapping five-year period standard deviations.

Figure 4. International Data and Age Response Functions



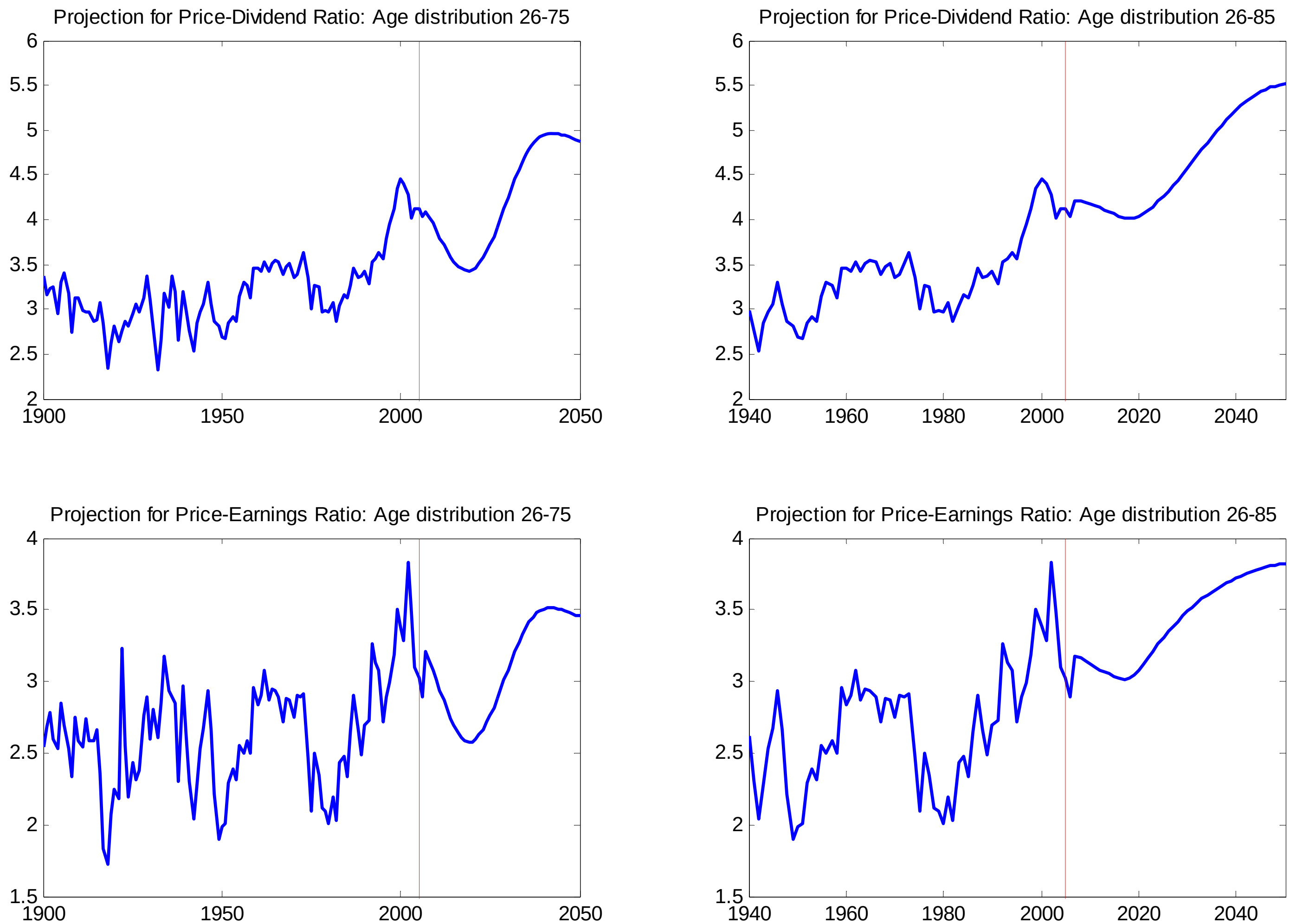
Note: This figure shows the estimates of $g(s)$ from the nonparametric regression for France, Germany, Japan, and the UK stock markets.

Figure 5. International Data and Fitted Values from Flexible Fourier Form



Note: This figure shows the fitted values of the log price-dividend ratio from the nonparametric regression for France, Germany, Japan, and the UK stock markets.

Figure 6. Projections with the US Data



Note: This figure shows the projections of the US normalized stock price with the use of the estimates of the individual parameters in the FFF expansion and population projections obtained from the US Census Bureau website. The vertical line indicates the time when the projections start.