Discussion Paper Series
No. 2005
March 2020

Demographic Structure and House Prices in the United States: A Reconciliation Using Metropolitan Area Data

Jihee Ann

Cheolbeom Park

The Institute of Economic Research – Korea University
Anam-dong, Sungbuk-ku, Seoul, 136-701, South Korea, Tel: (82-2) 3290-1632, Fax: (82-2) 928-4948

Demographic Structure and House Prices in the United States:
A Reconciliation Using Metropolitan Area Data[1]

Jihee Ann
Korea Real Estate Research Institute

and

Cheolbeom Park
Korea University

March 2020

Abstract: We apply a semiparametric approach to 19 metropolitan areas in the United States (US) to relate normalized house prices to entire age distributions in each area. We find that the shape of estimated age response functions differs across areas, although most areas show a negative impact of the elderly population on house prices. We further find that the age response function is more likely to be hump shaped when the population of an area becomes more aged, which also implies a negative aging impact. These results indicate that the impact of the elderly population will be negative as the population of an economy progresses toward becoming more aged.

JEL Classification: G12, J11, R30

Keywords: Demographic structure, Population aging, House prices, Semiparametric approach

[1] Jihee Ann: Associate Research Fellow, Research Division, Korea Real Estate Research Institute, 52 Bangbaero, Seochogu, Seoul, Korea 06705. Email: annji@kreri.re.kr.

Cheolbeom Park (corresponding author): Professor, Department of Economics, Korea University, 145 Anamro, Seongbukgu, Seoul, Korea 02841. Email: cbpark_kjs@korea.ac.kr. TEL: +82-2-3290-2203. Fax: +82-2-3290-2200.

I. Introduction

Many advanced economies are experiencing the phenomenon of rapid aging—a result of a drastic decline in fertility rates and death rates. Given this situation, policymakers, investors, households, and economists are keen to understand the impact of aging on house prices. Two opposing views can be found in the literature. Mankiw and Weil (1989), Takáts (2012), and Jäger and Schmidt (2017) show that an increase in the share of the elderly population will have a negative impact on house prices due to life cycle savings. That is, the prime working age generation would raise the size of a house whereas the elderly generation would reduce it. In contrast, Lisack et al. (2017) argue that the real interest rate will be lower in an aging economy, which will raise house prices. Hort (1998) also finds a positive impact of aging.

We provide empirical evidence on the relationship between demographic structure and house prices. We relate entire age distributions to house prices in 19 metropolitan areas in the US without imposing a functional form *a priori* for the relation. Consistent with the results of previous studies, we find both negative and positive impacts of aging. Although most areas show a negative impact, we reconcile these conflicting results by using the difference in the shapes of estimated age response functions. We show that the shape of age response functions relates with the demographic structure of the area. We demonstrate that as the share of the elderly population relative to the working population increases, the age response functions tends to be hump shaped, and the impact of the elderly population is more likely to be negative.

This paper is organized as follows: Section II presents the econometric methodology employed in this paper. Section III briefly discusses data, and Section IV provides the main empirical analysis. Concluding remarks are offered in Section V.

II. Econometric Model

To examine the relationship between the demographic structure and house prices, we consider the following econometric model:

$$y_t = \mu + \int_T f_t(m)\, g(m) dm + \beta' X_t + u_t \qquad (1)$$

where y_t is the log price–rent ratio, f_t denotes the density function of age distribution at time t, T is the common compact support for f_t, and X_t includes control variables other than the population age distribution. The log price–rent ratio (y_t) is a normalized house price and is examined in previous studies such as Campbell et al. (2009). As f_t, a regressor, is the density function of age distribution, $\int_{T_j} f_t(m)\, dm$ (where T_j is a subinterval of T) is the fraction of individuals for age group T_j in the total population. As we are relating the variation in the price–rent ratio to the variation in the entire age distribution, denoted as f_t, we can avoid any arbitrariness arising from the choice of a specific age range or a particular demographic measure.[2] $g(m)$ in regression equation (1) can be interpreted as the age response function that reflects the impact of population age distribution on the price–rent ratio. For the control variable, we include the growth rate of regional population and the growth rate of regional per capita income deflated by the corresponding regional consumer price index (CPI). Since migration between regions could be driven by current and expected income, which eventually affects the price-rent ratio, we have included the growth rate of per capita income in addition to the demographic structure.

Once the estimate of $g(m)$ is obtained, the age group that has a significant impact on the movement of the price–rent ratio can be determined, and the estimate of $g(m)$ can be compared with the implications of the theoretical models. Assuming a special form for $g(m)$, previous studies consider the variants of the econometric model in equation (1) to examine the

[2] Jäger and Schmidt (2017) instrument f_t based on past age distributions, and show that the results are robust, regardless of f_t being instrumented or not. Hence, we use f_t in the regression rather than instrument it.

relationship between the demographic structure and macroeconomic or financial variables. For example, the econometric model in Mankiw and Weil (1989) can be interpreted as a special case in which $g(m)$ is linear. Assuming a quadratic function for $g(m)$, Fair and Dominguez (1991) examine consumption and saving decisions. Higgins (1998) and Jäger and Schmidt (2017) assume a cubic function for $g(m)$ to estimate the impact of age distribution on international capital flows and house prices, respectively.

Unlike the above-mentioned studies, we do not impose *a priori* any special function for $g(m)$, which is consistent with Park (2010) in the sense that a flexible approach is employed. Instead, we assume that for the characteristic of $g(m)$, $g(m)$ must be sufficiently smooth to be approximated by a series of polynomials, trigonometric functions, or a mixture of both series. That is, we assume that $\|g_\kappa - g\| \to 0$ as $\kappa \to \infty$, where $g_\kappa(m)$ is an approximation of $g(m)$ given by a combination of a finite series of functions $\varphi_1, \dots, \varphi_\kappa$. When $g_\kappa(m) = \sum_{j=1}^{\kappa} \alpha_j \varphi_j(m)$, equation (1) can be expressed in a straightforward manner as:

$$y_t = \mu + \sum_{j=1}^{\kappa} \alpha_j \int_T f_t(m) \varphi_j(m) dm + \beta' X_t + u_{\kappa,t} = \mu + z_t a_\kappa + u_{\kappa,t}, \tag{2}$$

where $u_{\kappa,t} = u_t + \int_T f_t(m) (g - g_\kappa)(m) dm$, $a_\kappa = [\alpha_1, \dots, \alpha_\kappa, \beta']'$, and $z_t = [\int_T f_t(m) \varphi_1(m) dm, \dots, \int_T f_t(m) \varphi_\kappa(m) dm, X_t']$. Allowing Y and Z be vectors for y_t and z_t, respectively, the LS estimator of a_κ can be written as $\hat{a}_\kappa = (Z'Z)^{-1} Z'Y$. Then, the corresponding series estimator for the age response function can be expressed as

$$\hat{g}(m_h) = \sum_{j=1}^{\kappa} \hat{\alpha}_j \varphi_j(m_h), \tag{3}$$

where m_h is an interval on T.

In the empirical analysis, we test various series functions such as polynomials and mixtures of both polynomials and trigonometric functions, often referred to the Fourier Flexible Form (FFF) in Gallant (1981). The polynomial expansion of $g(m)$ can be written as $g_{\kappa}(m) = \alpha_1 m + \alpha_2 m^2 + \cdots + \alpha_\kappa m^\kappa$, and the FFF expansion of $g(m)$ can be expressed as $g_{\kappa}(m) = \alpha_1 m + \alpha_2 m^2 + \sum_{j=1}^{J}[\alpha_{3,j}\,cos(2\pi jm) + \alpha_{4,j} sin(2\pi jm)]$, where $\kappa = 2 + 2J$. The selection of κ (or equivalently the selection of J in the FFF expansion) is made based on the h-block cross-validation (CV) and the modified h-block CV criteria (MCV), as suggested by Burman et al. (1994) and Racine (1997), respectively. The CV can be expressed as:

$$CV = N^{-1} \sum_{t=h}^{N-h} (y_t - z_t' \hat{a}_\kappa(t,h))^2, \quad (4)$$

where $\hat{a}_\kappa(t,h)$ is estimators of the coefficients in equation (2) obtained by removing the *t*-th observation, and the h observations preceding and following the *t*-th observation in the dependent and independent variables in the regression. The modified h-block CV criterion, motivated by cases where $\frac{\kappa}{N}$ is not negligible, can be written as:

$$\begin{aligned} MCV = {} & N^{-1} \sum_{t=h}^{N-h} (y_t - z_t' \hat{a}_\kappa(t,h))^2 \\ & + N^{-2} \sum_{t=h}^{N-h} \sum_{i=1}^{N} (y_i - z_i' \hat{a}_\kappa(t,h))^2 \\ & + N^{-1} \sum_{i=1}^{N} (y_i - z_i' \hat{a}_\kappa)^2 \quad (5) \end{aligned}$$

The κ that minimizes CV or MCV is selected in the analysis.

III. Data

We estimate the age response functions for 19 metropolitan areas in the US. All these areas along with the sample periods are listed in Table 1. The sample periods are 1975 – 2017 with the exceptions of Baltimore, Miami, Tampa, and Washington DC areas. House price indices are taken from the Federal Housing Finance Agency. Assuming that rents paid by renters are identical to rents accruing to owner-occupiers, tenant rent indices for the 19 areas are obtained from the Bureau of Labor Statistics. We adjust the rent scales using the average price–rent ratios for the corresponding areas in Campbell et al. (2009).[3] The population structure data for the sampled areas are extracted from the US Census Bureau, and we use population estimates by 14 five-year age groups (ages 20–24, 25–29, …, 80–84, and 85 or above) to make the annual age distribution. Regional population growth rates are also taken from the US Census Bureau. Finally, regional data for per capita income and CPI are extracted from the Bureau of Economic Analysis and the Bureau of Labor Statistics, respectively. All the variables have annual frequencies.

IV. Empirical Results

As we cannot expand $g(m)$ with an infinite number of functions, it is important to determine the number of series functions that should be included to achieve a good approximation. We use various series functions to compute CV and MCV. While computing CV and MCV, the block size, h, is set as the integer nearest to $N/6$ following the suggestion of Burman et al. (1994). The results are shown in Table 2, and CV and MCV are expressed in boldface when they are minimized among the alternatives for a given metropolitan area. The selected series functions differ across the 19 metropolitan areas, although the quadratic form is chosen most frequently.[4]

[3] The rent scales for Baltimore, Tampa, and Washington DC, which are not analyzed in Campbell et al. (2009), are adjusted using the national average price-rent ratio for the US.

[4] For a given metropolitan area, CV and MCV always indicate the same series functions among the alternative forms except Minneapolis area. For Minneapolis area, there is slight difference in CV between cubic polynomial

As different series functions are selected across the areas, we run the semiparametric regression in equation (1) separately for individual areas instead of running a pooled regression. Table 3 shows estimated coefficients for the growth rates of regional per capita income and regional population growth. The coefficients for the growth rate of regional per capita income are negative but insignificant in most cases. The coefficients for regional population growth are positive in 11 out of 19 cases, and most of them are insignificant.

Figures 1 and 2 present the estimated age response functions for all the areas in our study. As different series functions are selected across areas, the shapes of the estimated age response functions also differ across areas. However, the shapes of age response functions can be categorized into two types, as presented in Figures 1 and 2. Newey-West standard errors are used to construct the 95% confidence interval for the age response function. The estimated age response functions in Figure 1 are consistent with the life cycle hypothesis, implying a positive impact of the working age population and negative impact of the elderly population on house prices.[5] 14 metropolitan areas out of the 19 under study have this type of age response functions, and the negative impact of aging of this group is consistent with the results in Mankiw and Weil (1989), Takáts (2012), and Jäger and Schmidt (2017). In contrast, the estimated age response functions in type 2 of Figure 2 are compatible with Hort (1998) and Lisack et al. (2017), as they argue a positive impact of the elderly population on house prices. Five metropolitan areas are categorized as type 2. Figures 3 and 4 show that the fitted values from the semiparametric regression track the actual price–rent ratios quite closely, regardless

---

and FFF with $J = 1$, whereas FFF with $J = 1$ shows lower MCV score than cubic polynomial function. Since Racine (1997) demonstrates that MCV performs better with non-negligible $\frac{\kappa}{N}$, we use FFF with $J = 1$ in the estimation for Minneapolis area.

[5] Miami metropolitan area shows a somewhat unique shape of the age response function in which there are several peaks, but it is hump shaped for the range of ages above 70.

of the shapes of the estimated age response functions.

Facing these different shapes of age response functions, we further check if these differences can be reconciled. For this, we conjecture that the different shapes of age response functions may be related with the different degrees of aging across the sampled areas. For example, the share of the population aged 60 or above in Dallas or Houston is approximately 16%, while the corresponding share in Tampa is around 26%. When the share of the elderly population is small, the house demand from the working age population can absorb the negative impact of the elderly population on house prices. As a result, the estimated age response function for the elderly can be estimated as positive. However, when the share of the elderly population exceeds a certain threshold, its negative impact on house prices could be more pronounced, and the estimated age response function will be negative for the elderly.

To investigate this possibility, we run the following regression:

$$ARF_i = a_0 + a_1 x_i + e_i, \tag{6}$$

where $ARF_i = \begin{Bmatrix} \text{peak} - \text{the terminal points in the age response function in type 1} \\ \text{bottom} - \text{the terminal points in the age response function in type 2} \end{Bmatrix}.$

ARF_i is designed to capture the shape of the age response functions. When the age response function is hump shaped, as shown in Figure 1, ARF_i will be positive. However, when the impact of aging is positive, as shown in type 2 of Figure 2, the terminal points of the age response functions are greater than the bottom points, which implies a negative value for ARF_i.[6] x_i is the ratio of the share of the population aged between 40 and 64 to the share of the population aged 65 and above in 2017. As the population of a metropolitan area becomes more

[6] ARF_i for Miami is the gap between the highest point and terminal point in the age response function because the age response function for the elderly population in Miami is hump shaped.

(less) aged, x_i will be lower (higher) and the shape of the age response function is more likely to be as shown in type 1 (2), which implies a more positive (negative) value for ARF_i. Hence, if our conjecture is correct, then a_1 in equation (6) will be negative. The results are presented in Table 4. Although we have a small number of observations, a_1 is significantly negative at the 5% level, as shown in the first column. When we switch x_i to the ratio of the share of the population aged between 20 and 64 to the share of the population aged 65 or above in 2017, we can also obtain supportive evidence to our conjecture. When we use the share of the population aged 60 or above in 2017 to measure the degree of aging, then x_i will increase with the degree of aging. As a result, a_1 will be positive under the conjecture, which is consistent with the result in the last column of Table 4. All these results suggest that the shape of the age response function depends on the degree of aging, and as the population of an economy becomes more aged, the shape of the age response function is more likely to be as shown in Figure 1, which is consistent with the life cycle hypothesis and negative aging impact on house prices.

V. Conclusion

We run semiparametric regressions for the price–rent ratios on population age distributions for 19 metropolitan areas in the US to understand the impact of aging on house prices. Although many areas have hump-shaped age response functions, different series functions are selected and different shapes of age response functions are estimated across areas, which is consistent with conflicting results reported in previous studies. Furthermore, we verify that the shape of the age response function is more likely to be hump shaped as the population of a metropolitan area becomes more aged, which implies that the shapes of age response function are related with the degree of aging. These results imply that the impact of aging on house prices will be negative when an area becomes more aged.

References

Burman, P., Chow, E., Nolan, D., 1994. A cross-validatory method for dependent data. Biometrika 81, 351-358.

Campbell, S.D., Davis, M.A., Gallin, J., Martin, R.F., 2009. What moves housing markets: A variance decomposition of the rent-price ratio. Journal of Urban Economics 66, 90-102.

Fair, R.C., Dominguez, K.M., 1991. Effects of the changing US age distribution on macroeconomic equations. American Economic Review 81, 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.

Higgins, M., 1998. Demography, national savings, and international capital flows. International Economic Review 39, 343-369.

Hort, K., 1998. The determinants of urban house price fluctuations in Sweden 1968-1994. Journal of Housing Economics 7, 93-120.

Jäger, P., Schmidt, T., 2017. Demographic change and house prices: Headwind or tailwind? Economics Letters 160, 82-85.

Lisack, N., Sajedi, R., Thwaites, G., 2017. Demographic trends and the real interest rate. Bank of England Working Papers No.701.

Mankiw, N.G., Weil, D.N., 1989. The baby boom, the baby bust, and the housing market. Regional Science and Urban Economics 19, 235-258.

Park, C., 2010. How does changing age distribution impact stock prices? A nonparametric approach. Journal of Applied Econometrics 25, 1155-1178.

Racine, J., 1997. Feasible cross-validatory model selection for general stationary processes. Journal of Applied Econometrics 12, 169-179.

Takáts, E., 2012. Aging and house prices. Journal of Housing Economics 21, 131-141.

Table 1. Metropolitan Areas and Sample Periods

Metropolitan Areas	**Sample Periods**
Atlanta-Sandy Springs-Roswell, GA	1975–2017
Baltimore-Columbia-Towson, MD	1998–2017
Boston-Cambridge-Newton, MA-NH	1975–2017
Chicago-Naperville-Elgin, IL-IN-WI	1975–2017
Dallas-Fort Worth-Arlington, TX	1975–2017
Denver-Aurora-Lakewood, CO	1975–2017
Detroit-Warren-Dearborn, MI	1975–2017
Houston-The Woodlands-Sugar Land, TX	1975–2017
Los Angeles-Long Beach-Anaheim, CA	1975–2017
Miami-Fort Lauderdale-West Palm Beach, FL	1977–2017
Minneapolis-St.Paul-Bloomington, MN-WI	1975–2017
New York-Newark-Jersey City, NY-NJ-PA	1975–2017
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1975–2017
San Diego-Carlsbad, CA	1975–2017
San Francisco-Oakland-Hayward, CA	1975–2017
Seattle-Tacoma-Bellevue, WA	1975–2017
St. Louis, MO-IL	1975–2017
Tampa-St. Petersburg-Clearwater, FL	1987–2017
Washington DC-Arlington-Alexandria, DC-VA-MD-WV	1998–2017

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

		Quadratic Polynomial	**Cubic Polynomial**	**FFF $J = 1$**	**FFF $J = 2$**	**FFF $J = 3$**	**FFF $J = 4$**
Atlanta	CV	0.017	**0.012**	0.019	0.035	0.030	0.077
	MCV	0.028	**0.020**	0.028	0.046	0.038	0.092
Baltimore	CV	**0.038**	0.052	0.060	0.068	0.144	3.846
	MCV	**0.055**	0.076	0.085	0.096	0.287	4.655
Boston	CV	**0.096**	0.125	0.178	0.163	0.230	0.256
	MCV	**0.145**	0.173	0.239	0.210	0.292	0.320
Chicago	CV	0.015	**0.015**	0.061	0.178	0.268	0.182
	MCV	0.028	**0.027**	0.090	0.229	0.330	0.240
Dallas	CV	0.012	0.019	0.046	**0.009**	0.014	0.024
	MCV	0.019	0.026	0.061	**0.012**	0.018	0.030
Denver	CV	**0.026**	0.063	0.132	0.041	0.094	0.154
	MCV	**0.044**	0.089	0.179	0.055	0.120	0.196
Detroit	CV	**0.041**	0.065	0.054	0.142	0.152	0.310
	MCV	**0.068**	0.098	0.089	0.189	0.200	0.390
Houston	CV	0.018	0.033	0.057	**0.017**	0.025	0.046
	MCV	0.027	0.047	0.079	**0.022**	0.033	0.060
Los Angeles	CV	0.058	0.094	**0.037**	0.158	0.475	0.635
	MCV	0.100	0.142	**0.069**	0.229	0.599	0.793
Miami	CV	0.583	0.680	1.083	**0.564**	1.143	0.616
	MCV	1.046	1.173	1.584	**0.863**	1.636	0.889
Minneapolis	CV	0.044	**0.034**	0.037	0.085	0.070	0.096
	MCV	0.072	0.055	**0.050**	0.108	0.087	0.121
New York	CV	0.198	0.174	**0.078**	0.099	0.088	0.083
	MCV	0.288	0.236	**0.104**	0.126	0.112	0.105
Philadelphia	CV	**0.018**	0.037	0.048	0.075	0.183	0.113
	MCV	**0.031**	0.054	0.064	0.092	0.229	0.142
San Diego	CV	**0.031**	0.040	0.046	0.167	0.522	0.489
	MCV	**0.061**	0.072	0.079	0.245	0.664	0.622
San Francisco	CV	**0.051**	0.063	0.185	0.692	1.281	0.900
	MCV	**0.096**	0.111	0.249	0.844	1.570	1.094
Seattle	CV	**0.029**	0.035	0.075	0.158	0.246	0.328
	MCV	**0.052**	0.059	0.110	0.207	0.303	0.396
St. Louis	CV	0.021	**0.014**	0.032	0.033	0.102	0.107
	MCV	0.033	**0.024**	0.058	0.053	0.156	0.165
Tampa	CV	**0.037**	0.067	0.068	0.223	0.213	0.151
	MCV	**0.057**	0.093	0.089	0.282	0.265	0.187
Washington DC	CV	**0.039**	0.733	0.251	0.593	0.382	0.090
	MCV	**0.067**	1.127	0.371	0.906	0.569	0.134

Note: The polynomial expansion of $g(m)$ can be written as $g_\kappa(m) = \alpha_1 m + \alpha_2 m^2 + \cdots + \alpha_\kappa m^\kappa$, and the Fourier Flexible Form (FFF) expansion of $g(m)$ can be expressed as $g_\kappa(m) = \alpha_1 m + \alpha_2 m^2 + \sum_{j=1}^{J}[\alpha_{3,j}\, cos(2\pi jm) + \alpha_{4,j} sin(2\pi jm)]$.

Table 3. Coefficients for Regional Per Capita Income Growth and Regional Population Growth

	Regional Per Capita Income Growth	Regional Population Growth
Atlanta	-0.6** (0.24)	5.1** (2.29)
Baltimore	-2.5*** (0.64)	-37.5** (15.25)
Boston	0.4 (0.79)	2.4 (10.55)
Chicago	-0.6 (0.37)	-12.4** (5.69)
Dallas	-0.3** (0.13)	0.4 (1.05)
Denver	-0.6 (0.31)	0.5 (1.97)
Detroit	-0.8 (0.45)	2.3 (3.42)
Houston	-0.2 (0.13)	1.3 (0.68)
Los Angeles	-0.6 (0.63)	-19.5*** (6.61)
Miami	0.3 (0.59)	-0.2 (0.13)
Minneapolis	-0.4 (0.34)	5.8 (3.39)
New York	0.1 (0.45)	-26.2*** (4.11)
Philadelphia	-1.7*** (0.42)	14.8*** (3.24)
San Diego	-2.2*** (0.51)	-6.4 (3.18)
San Francisco	-0.6 (0.69)	-5.9 (5.96)
Seattle	-0.8 (0.43)	2.1 (2.29)
St. Louis	-0.4 (0.51)	12.8** (5.53)
Tampa	-0.6 (0.71)	13.8*** (3.56)
Washington DC	0.1 (0.73)	-18.7** (6.83)

Note: Numbers in parentheses are standard errors. **, and *** denote that each coefficient is significant at 5%, and 1% level, respectively.

Table 4. Shape of Age Response Functions and Aging Measure

$$ARF_i = \begin{Bmatrix} \text{peak} - \text{the terminal points in the age response function in Figure 1} \\ \text{bottom} - \text{the terminal points in the age response function in Figure 2} \end{Bmatrix}$$

x_i	**(40-64)/(65+)**	**(20-64)/(65+)**	**(60+)**
a_0	273.7 ** (127.1) [0.046]	242.9 ** (108.0) [0.038]	-341.6 ** (131.7) [0.019]
a_1	-124.0 ** (54.3) [0.036]	-59.2 ** (24.6) [0.028]	1616.6 ** (645.0) [0.023]
R^2	0.235	0.254	0.270

Note: Numbers in parentheses are standard errors and numbers in brackets are p-values. p-values are calculated from the t-distribution with 17 degrees of freedom. **, and *** denote that each coefficient is significant at 5% and 1% level, respectively.

Figure 1. Estimated Age Response Functions: Type 1

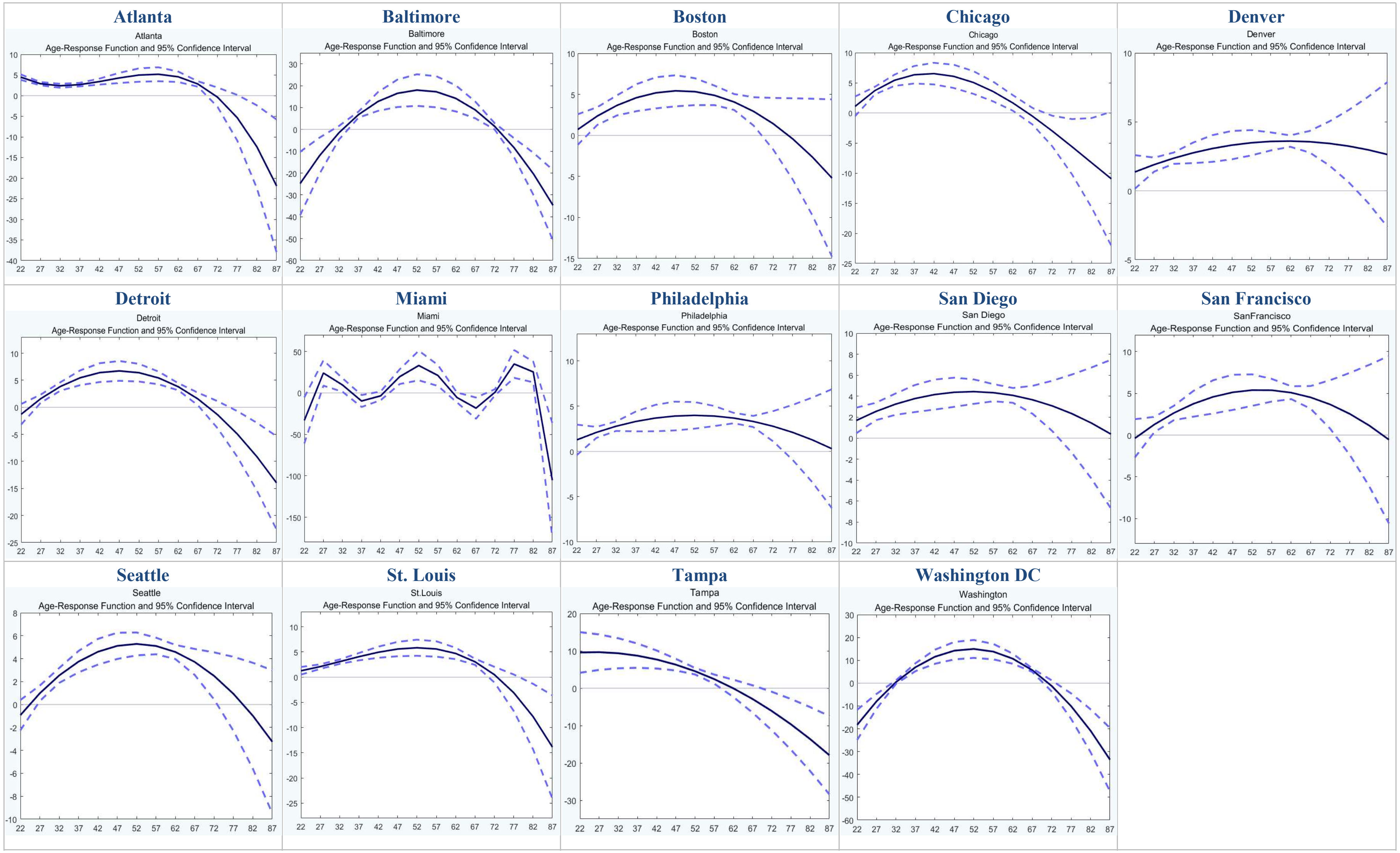

Figure 2. Estimated Age Response Functions: Type 2

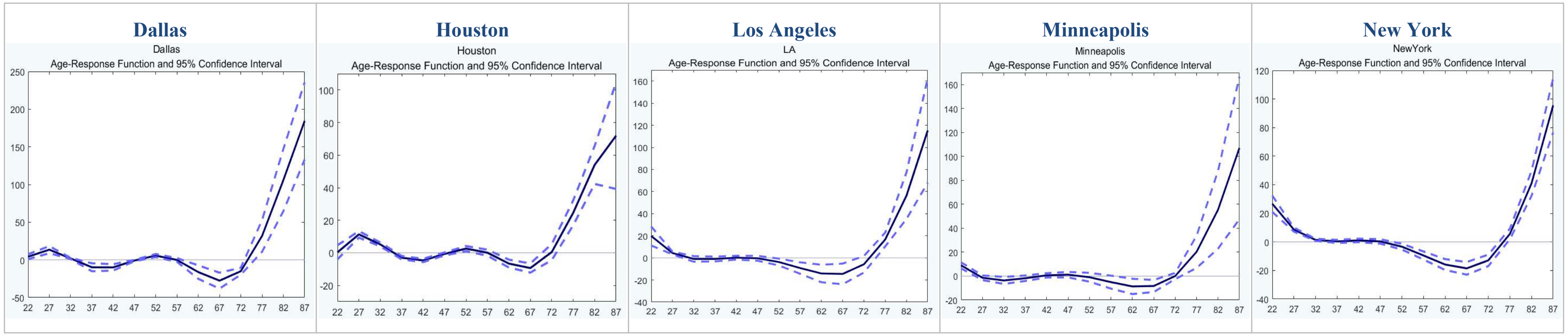

Figure 3. Fitted Values of Semiparametric Regression: Type 1

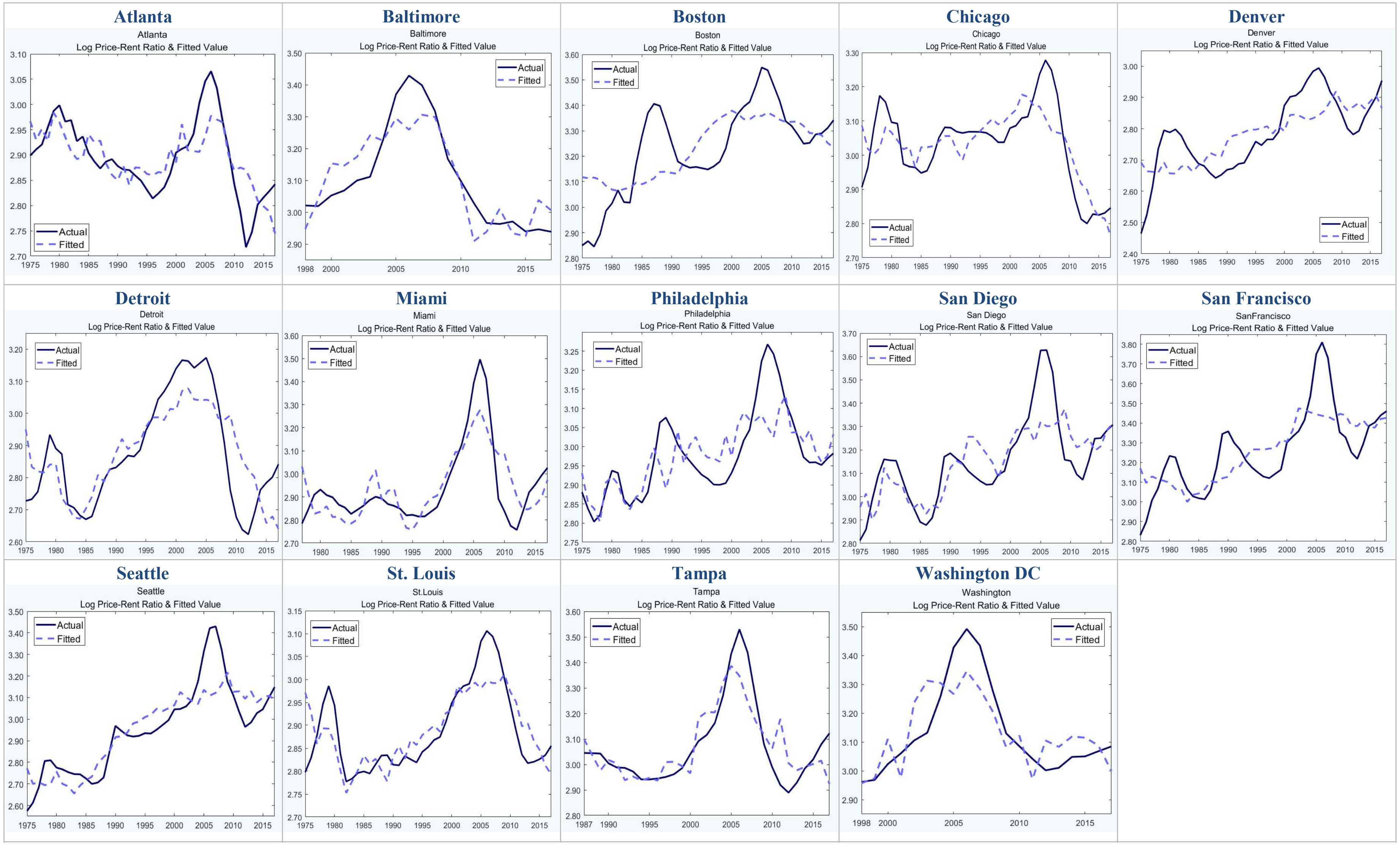

Figure 4. Fitted Values of Semiparametric Regression: Type 2