Empir Econ (2010) 38:23–45
DOI 10.1007/s00181-009-0254-1

ORIGINAL PAPER

Analysis of county employment and income growth in Appalachia: a spatial simultaneous-equations approach

Gebremeskel H. Gebremariam · Tesfa G. Gebremedhin · Peter V. Schaeffer

Received: 15 April 2006 / Accepted: 15 September 2008 / Published online: 31 January 2009

Abstract County median household income and employment growth rates tend to be characterized by spatial interaction. A spatial simultaneous-equations growth equilibrium model was estimated using GS2SLS and GS3SLS. The results indicate strong feedback simultaneity between employment and median household income growth rates. They also show spatial autoregressive lag simultaneity and spatial cross-regressive lag simultaneity with respect to employment and median household income growth rates, as well as spatial correlation in the error terms. Estimates of structural parameters show strong agglomerative effects and significant conditional convergence with respect to employment growth and median household income growth in Appalachia in the 1990 s.

Keywords Employment · Income · Spatial analysis · Appalachia

JEL Classification C3 · R1 · R5

1 Introduction

State policy makers and local leaders have long placed a high priority on local economic development (Isserman 1993; Pulver 1989; Ekstrom and Leistritz 1988). The changing structure of traditional industries and the impact of those changes on local communities have challenged the efficacy of established policies and strategies.

G. H. Gebremariam
Department of Economics, Virginia Polytechnic Institute, State University, Blacksburg, USA

T. G. Gebremedhin · P. V. Schaeffer (✉)
Division of Resource Management, West Virginia University, Morgantown, USA
e-mail: peter.schaeffer@mail.wvu.edu

A better understanding of factors that influence local employment, earning capacity, and quality of life issues has therefore become important for state, regional, and local agencies in charge of rural development policies. One of the policy challenges at the local and county level is spatial interdependencies. Outside of the far west, counties are small and in most instances cannot be thought of as even rough approximations of labor markets, and often as market areas for many consumer goods, either. This is expressed by the mean travel time to work in a rural state such as West Virginia, which in 2005 ranked fourteenth in the nation, with 24.9 min just below the national average of 25.1 min, and ahead of more urbanized states such as Ohio and Michigan (US Census Bureau 2005). We single out West Virginia because of its rural nature and because it is the only state completely contained within Appalachia. The twelve states with counties in Appalachia have commuting times (one way) of 22.4 min or higher. Dealing with spatial interdependencies is therefore one of the major objectives of this research.

Many of the forces responsible for past economic and social changes continue to have an impact. One of these changes was the emergence of computer-based technology in production, administration and information, which has reduced the role of economies of scale in many sectors. Studies by Loveman and Sengenberger (1991) and Acs and Audretsch (1993), for example, have shown a shift in industry structure toward decentralization and an increased role for small firms. This was mainly due to changes in production technology, consumer demand, labor supply, and the pursuit of flexibility and efficiency. These factors led to the restructuring and downsizing of large enterprises and the entry of new firms. Brock and Evans (1989) provide extensive documentation of the changing role of small businesses in the US economy, which are likely the result of responses to structural adjustments.

Parallel with technical changes leading to new industrial structures, new patterns of consumer expenditures and demand resulting from rising living standards contributed to the emergence of fragmented consumer markets, which also favored small consumer-oriented firms over high volume, production-oriented firms. Thus, new business opportunities in small and medium size enterprises resulted as large firms downsized in response to a changing environment. The emerging view among policy makers is that small business is a key element and driving force in generating employment and realizing economic development. This paradigm shift has brought about a revival in small businesses promotion and entrepreneurial initiatives at local, national and international levels.

Most new businesses start small and small businesses create the majority of new jobs (Acs and Audretsch 2001; Audretsch et al. 2000; Carree and Thurik 1998, 1999; Fritsch and Falck 2003; Reynolds 1994; Wennekers and Thurik 1999). A growing literature has explored the determinants of the regional variations in new business formation (Acs and Armington 2003; Audretsch and Fritsch 1994; Callejon and Segarra 2001; Davidson et al. 1994; Fotopoulos and Spencer 1999; Fritsch 1992; Garofoli 1994; Guesnier 1994; Hart and Gudgin 1994; Johnson and Parker 1996; Kangasharju 2000; Keeble and Walker 1994; Reynolds 1994).

The geographic impact of the change from large production-oriented plants to smaller consumer-oriented firms and plants is uncertain. While smaller units would tend to make rural production sites relatively more competitive, the consumer-orientation, which tends to favor locations close to markets, is more likely to have the opposite effect. Hence, it is not possible to predict the impact of the changes discussed above on the geographic distribution of economic activity a priori.

The literature on economic growth at the regional level has focused attention on the so-called convergence hypothesis of neoclassical growth theory which predicts that poorer regions tend to catch up with the richer regions in per capita income as time passes, through the process of factor mobility. Because of the spatial structure of our model, we can test for convergence. Previous studies by Barro and Sala-i-Martin (1992, 2004) for US states, Japanese prefectures and between European countries, and by Persson (1997) and Aronsson et al. (2001) across Swedish counties, found income evidence of convergence. Similar studies by Arbia et al. (2005) of 92 Italian provinces (1970–2000), Ertur et al. (2006) of 138 European regions (1980–1995), and Rappaport (1999) of US counties (1970–1990), also found income convergence. However, a study by Glaeser et al. (1995) did not discover significant evidence of income convergence between US cities. Of particular interest are two papers by Higgins et al. (2006) and Young et al. (2008) that looked at per capita income net of government transfers in US counties in all fifty states from 1970 to 1998. They found a speed of convergence of between 6 to 8%, considerable faster than the approximately 2% typically reported. Higgins et al. (2006) also found a much a faster speed of convergence in counties located in southern than in northeastern states.

The relationship between economic growth and its determinants has been studied extensively. One issue is whether population is driving employment changes or employment is driving population changes (do 'jobs follow people' or 'people follow jobs'?). Empirical studies on identification of the direction of causality have resulted in empirical models of regional development that often reflect the interdependence between household residential choices and firm location choices (Steinnes and Fisher 1974). To account for this causation and interdependency, Carlino and Mills (1987) constructed a simultaneous system model with two partial location equations as its components. They used data for counties in the contiguous United States. The empirical result from their study of greatest interest to us is the finding that in the 1970 s family income had a strong impact on the growth of population density as well as employment density. Recently, Deller et al. (2001) expanded the original Carlino-Mills model and presented a three-dimensional model (jobs-people-income) that explicitly traces job quality and the role of income in the regional growth process. They also used county data, but restricted themselves to non-metropolitan counties; the time period studied was 1985–1995. Their empirical results indicate that initial conditions co-determine the eventual outcome and that counties with higher initial population levels tended to have higher employment growth. However, counties that had higher levels of population, employment, and per capita income in 1985 tended to have lower rates of overall growth.

There have also been efforts to model the interactions between employment growth and human migration (Clark and Murphy 1996; MacDonald 1992), per capita personal income and public expenditures (Duffy-Deno and Eberts 1991), and net migration, employment growth, and average per capita income (Greenwood and Hunt 1984; Greenwood et al. 1986; Lewis et al. 2002) in simultaneous-equations models. Among these contributions, Clark and Murphy's (1996) findings have been particularly influential. Their empirical analysis covered the period 1981–1989 and was conducted at the county level. They expanded the Carlino-Mills model by including amenity measures beyond climate (temperature), neighborhood poverty, and fiscal variables. Their results are consistent with those of Carlino and Mills (1987) and, specifically, they find simultaneity between employment density and population density.

The focus of this empirical analysis is Appalachia, a region that is for many a symbol of poverty and underdevelopment in the midst of prosperity (Pollard 2003). It is a region of about 23 million people. Forty-two percent of the population is rural, compared to 20% for the nation as a whole. Many parts of the region can also be considered remote because of topography and a comparatively poor transportation infrastructure. Appalachia also constitutes a separate policy region, with programs administered by the Appalachian Regional Commission. The unit of analysis is the county, so that we can trace local economic development in terms of employment and income growth data, respectively. The time period considered is 1990–2000. This was a decade of economic growth and expansion in most of the United States. It is of interest to study if and/or how the boom of the 1990 s impacted Appalachian counties.

Like the studies mentioned above, this article examines the determinants of regional variations in employment and household income growth rates using county data. Its novel contribution lies in a methodological innovation. Specifically, the model introduces both spatial lag and spatial error dependence into a simultaneous equation model and obtains estimation results using Generalized Spatial 3 Stage Least Squares (GS3SLS). This has not been previously done and yields more efficient and consistent estimates. The estimation strategy is discussed in the estimation issue section.

2 Method of analysis

Interdependence between employment and income exists because both households and firms are mobile and locate to maximize utility and profits, respectively. Households migrate if they can capture better income opportunities than those available at their current location and firms move to be near growing markets. The location decisions of firms are also expected to be influenced by factors such as local business climate, labor costs, tax rates, local public services and the supply of inputs. In addition, government-provided incentives may influence where firms locate. Such regional factors that affect households' and firms' decision making are also likely to exhibit spatial autocorrelation (Anselin 1988, 2003). These assumptions are expressed as three hypotheses to be tested: (1) Employment growth and median household income growth are interdependent and jointly determined by regional variables; (2) Employment growth and median household income growth in a county are conditional upon initial conditions of that county; and (3) Employment growth and median household income growth in a county are conditional upon business and median household income growth in neighboring counties. Emphasis is put on determining the linkages between employment growth and household median income, as well as on examining the elasticity of these variables with respect to each of the regional variables.

To test the three hypotheses, a spatial simultaneous equations model of business growth and household median income is used. Following Carlino and Mills (1987) and building on Boarnet (1994), a model that incorporates own-county and neighboring counties effects is specified as follows in matrix notation:

$$\boldsymbol{EMP}^*_{it} = f_1\left[\left(\boldsymbol{MHY}^*_t, \boldsymbol{WMHY}^*_t\right), \boldsymbol{WEMP}^*_t, \left|\boldsymbol{X}^{em}_{t-1}\right.\right] \quad (1a)$$

$$\boldsymbol{MHY}^*_{it} = f_2\left[\left(\boldsymbol{EMP}^*_t, \boldsymbol{WEMP}^*_t\right), \boldsymbol{WMHY}^*_t, \left|\boldsymbol{X}^{mh}_{t-1}\right.\right] \quad (1b)$$

$\boldsymbol{EMP}^*_t$ and $\boldsymbol{MHY}^*_t$ are the equilibrium levels of private non-farm employment and median household income, respectively, and t denotes time. $\boldsymbol{W}$ is a row standardized spatial weights matrix with typical element w_{ij}. Each element w_{ij} represents a measure of proximity between location i and location j. We define the adjacency criteria such that w_{ij} equals $1/n_i$; n_i is the number of nonzero elements in the ith row of $\boldsymbol{W}$. The row element is nonzero if location i and j are adjacent and 0 otherwise. $\boldsymbol{WEMP}^*_t$ and $\boldsymbol{WMHY}^*_t$ represent the equilibrium values of neighboring counties' effects for private non-farm employment and median household income, respectively. They are obtained by multiplying $\boldsymbol{EMP}^*_t$ and $\boldsymbol{MHY}^*_t$, respectively, with $\boldsymbol{W}$. The matrices of additional exogenous variables in the respective equations of the system of spatial simultaneous equations are given by $\boldsymbol{X}^{em}_{t-1}$ and $\boldsymbol{X}^{mh}_{t-1}$, respectively. The descriptions of these variables are given in the data section below. Note that equilibrium levels of private non-farm employment and median household income are assumed to be functions of the equilibrium values of the respective right-hand endogenous variables, their spatial lags and the vectors of the additional exogenous variables.

The system of equations in (1a, b) captures the simultaneous nature of the interactions between employment growth and median household income at equilibrium. The nature of interaction among the endogenous variables depends on the initial conditions in a county.

Based on the result of a generalized PE-test, a multiplicative log-linear form of the model was used. The model specification is discussed in greater detail in the section "Estimation Issues." The chosen functional form implies constant elasticity for the equilibrium conditions given in (1a,b). A log-linear (i.e., log-log) representation of the equilibrium conditions can thus be expressed as:

$$\boldsymbol{EMP}^*_t = \left(\boldsymbol{MHY}^*_t\right)^{a_1} \times \left(\boldsymbol{WEMP}^*_t\right)^{b_1} \times \left(\boldsymbol{WMHY}^*_t\right)^{c_1} \times \prod_{k=1}^{K_1} \left(\boldsymbol{X}^{em}_{kt-1}\right)^{x_{1k}} \quad (2a)$$

$$\boldsymbol{MHY}^*_t = \left(\boldsymbol{EMP}^*_t\right)^{a_2} \times \left(\boldsymbol{WMHY}^*_t\right)^{b_2} \times \left(\boldsymbol{WEMP}^*_t\right)^{c_1} \times \prod_{k=1}^{K_2} \left(\boldsymbol{X}^{em}_{kt-1}\right)^{x_{2k}} \quad (2b)$$

where a_i, b_i and $c_i i = 1, 2$ are the exponents on the endogenous variables and their spatial lags, x_{ik_q} for $i, q = 1, 2$ are vectors of exponents on the exogenous variables, $\prod$ is the product operator, and K_i for $i = 1, 2$ is the number of exogenous variables in the private non-farm employment and median household income equations, respectively. The log-linear specification has the advantage of yielding a log-linear reduced form for estimation, where the estimated coefficients represent elasticities. Duffy-Deno (1998) and Mackinnon et al. (1983) also show that, compared to a linear specification, a log-linear specification is more appropriate for models involving population and employment densities.

Previous empirical studies suggest that employment and median household income likely adjust to their equilibrium levels with a substantial lag (Aronsson et al. 2001; Barkley et al. 1998; Boarnet 1994; Carlino and Mills 1987; Deller et al. 2001; Duffy 1994; Duffy-Deno 1998; Edmiston 2004; Hamalainen and Bockerman 2004; Henry et al. 1999, 1997; Mills and Price 1984). Therefore, based on these studies, a distributed lag adjustment is introduced and the corresponding partial-adjustment process for Eqs. (1a,b) takes the form:

$$\frac{\boldsymbol{EMP}_t}{\boldsymbol{EMP}_{t-1}} = \left(\frac{\boldsymbol{EMP}_t^*}{\boldsymbol{EMP}_{t-1}}\right)^{\eta_{em}} \rightarrow \ln(\boldsymbol{EMP}_t) - \ln(\boldsymbol{EMP}_{t-1}) = \eta_{em} \ln\left(\boldsymbol{EMP}_t^*\right) - \eta_{em}\left(\boldsymbol{EMP}_{t-1}\right) \tag{3a}$$

$$\frac{\boldsymbol{MHY}_t}{\boldsymbol{MHY}_{t-1}} = \left(\frac{\boldsymbol{MHY}_t^*}{\boldsymbol{MHY}_{t-1}}\right)^{\eta_{mh}} \rightarrow \ln(\boldsymbol{MHY}_t) - \ln(\boldsymbol{MHY}_{t-1}) = \eta_{mh} \ln\left(\boldsymbol{MHY}_t^*\right) - \eta_{mh} \ln(\boldsymbol{MHY}_{t-1}) \tag{3b}$$

The subscript $t - 1$ refers to the variable lagged one period, one decade in this study, and η_{em} and η_{mh} are parameters representing the speed of adjustment of employment and median household income to their respective equilibrium levels. They are interpreted as the proportions of the respective equilibrium rate of growth that were realized in each period. If both η_{em} and η_{mh} are less than one, then the system is stable and guaranteed to converge.

The existence of spatial autocorrelation in the errors is tested by means of a Global Moran's I test statistic, as suggested by Anselin and Kelejian (1997) for models with endogenous regressors. A more general version of Moran's I test statistic and its asymptotic distribution is given by Kelejian and Prucha (2001). The results of the test (Table 2) indicate the existence of spatial autocorrelation in the errors of all equations in (3a, b). Therefore, we need a model that accounts for this spatial effect. We achieve this by substituting Eqs. (2a, b) into Eqs. (3a, b). Eliminating the unknown equilibrium values and simplifying the model yields the following system:

$$\begin{aligned} \boldsymbol{EMPR}_t = {} & \alpha_1 + \frac{\eta_{em} a_1}{\eta_{mh}} \boldsymbol{MHYR}_t + \frac{\eta_{em} b_1}{\eta_{em}} \boldsymbol{WEMPR}_t + \frac{\eta_{em} c_1}{\eta_{mh}} \boldsymbol{WMHYR}_t \\ & + \eta_{em} a_1 \ln(\boldsymbol{MHY}_{t-1}) + \eta_{em} b_1 \ln(\boldsymbol{WEMP}_{t-1}) \\ & + \eta_{em} c_1 \ln(\boldsymbol{WMHY}_{t-1}) \\ & + \sum_{k=1}^{K_1} \eta_{em} x_{1k} \ln\left(\boldsymbol{X}_{kt-1}^{em}\right) - \eta_{em} \ln(\boldsymbol{EMP}_{t-1}) + \boldsymbol{u}_t^{em} \end{aligned} \tag{4a}$$

$$\begin{aligned} \boldsymbol{MHYR}_t = {} & \alpha_2 + \frac{\eta_{mh} a_2}{\eta_{em}} \boldsymbol{EMPR}_t + \frac{\eta_{mh} b_2}{\eta_{mh}} \boldsymbol{WMHYR}_t + \frac{\eta_{mh} c_2}{\eta_{em}} \boldsymbol{WEMPR}_t \\ & + \eta_{mh} a_2 \ln(\boldsymbol{EMP}_{t-1}) + \eta_{mh} c_2 \ln(\boldsymbol{WEMP}_{t-1}) \end{aligned}$$

$$+ \eta_{mh} b_2 \ln(\boldsymbol{WMHY}_{t-1}) + \sum_{k=1}^{K_2} \eta_{mh} x_{2k} \ln\left(\boldsymbol{X}_{kt-1}^{ge}\right) - \eta_{mh} \ln(\boldsymbol{MHY}_{t-1}) + \boldsymbol{u}_t^{mh} \tag{4b}$$

$\boldsymbol{EMPR}_t$ and $\boldsymbol{MHYR}_t$ are the log differences between the end and beginning period values of private non-farm employment and median household income, respectively, and denote the growth rates of the respective variables. α_r and ρ_r, for $r = 1, 2$, are unobserved parameters. $\boldsymbol{u}_t^{em}$ and $\boldsymbol{u}_t^{mh}$ are $n \times 1$ vectors of disturbances? Note that the disturbance vector in the rth equation is generated as:

$$\boldsymbol{u}_{t,r} = \rho_r \boldsymbol{W} \boldsymbol{u}_{t,r} + \boldsymbol{\varepsilon}_{t,r}, \quad r = 1, 2$$

This specification relates the disturbance vector in the rth equation to its own spatial lag. The vectors of innovations ($\boldsymbol{\varepsilon}_{it,r}$, $r = 1, 2$ or $\boldsymbol{\varepsilon}_t^{em}$ and $\boldsymbol{\varepsilon}_t^{mh}$) are distributed identically and independently with zero mean and variance-covariance σ_r^2, for $r = 1, 2$. Hence, they are not spatially correlated. The specification of the mode, however, allows for innovations that correspond to the same cross sectional unit to be correlated across equations. As a result, the vectors of disturbances are spatially correlated across units and across equations.

Equations (4a, b) constitute a system of simultaneous equations with feedback simultaneity, spatial autoregressive lag simultaneity, spatial cross-regressive lag simultaneity, and spatial autoregressive disturbances. The endogenous variables of the model are $\boldsymbol{EMPR}_t$ and $\boldsymbol{MHYR}_t$. If each equation is investigated separately, we notice that each of these variables is expressed in terms of the right hand endogenous variables and their spatial lags, the logs of the lagged endogenous variables and their spatial lags, and the logs of other exogenous variables. By structure, the spatial lags of the lagged endogenous variables are, however, included in the spatial lags of the respective endogenous variables. Hence, in order to avoid multicollinearity, the model is estimated by excluding all the spatial lags of the lagged endogenous variables.

3 Data types and sources

The data for the 417 Appalachian counties used for the empirical analysis were collected and compiled from County Business Patterns, Bureau of Economic Analysis, Bureau of Labor Statistics, Current Population Survey Reports, County and City Data Book, US Census of Population and Housing, US Small Business Administration, and Department of Employment Security. Data for county employment and county median household income were collected for 1990 and 2000.

3.1 Dependent variables

The dependent variables used in the empirical analysis include the growth rate of employment and the growth rate of median household income.

3.1.1 Growth rate of employment (EMPR)

The growth rate of employment is measured by the log-difference between the 2000 and the 1990 levels of private non-farm employment, exclusive of self-employment. Empirical research indicates that in the study period most new jobs were generated by new small businesses (Acs and Audretsch 2001; Audretsch et al. 2000; Carree and Thurik 1998, 1999; Wennekers and Thurik 1999; Fritsch and Falck 2003). Research by the US Small Business Administration also shows that job creation capacity in the US is inversely related to the size of the business. Between 1991 and 1995, for example, enterprises employing fewer than 500 people created new jobs as follows (size of enterprise in parenthesis): 3.843 million (1–4), 3.446 million (5–19), 2.546 million (20–99), and 1.011 million (100–499). During the same period, enterprises employing 500 or more people lost 3.182 million net jobs (US Small Business Administration (SBA) 1999).

3.1.2 Growth rate of median household income (MHYR)

The log-difference between the 2000 and 1990 levels of median household income in a given county is used to measure the growth rate of median household income. Median household income is used as an average overall measure of county-level income. Median household income is preferable to using the mean household income because unlike the mean, the median is not influenced by the presence of a few extreme values.

The spatial lags of the Growth Rate of Employment ($\boldsymbol{WEMPR}$) and Growth Rate of Median Household Income ($\boldsymbol{WMHYR}$) are included on the right hand side of each equation of (4a, b). These spatially lagged endogenous variables are created by multiplying each of the dependent variables by a row standardized queen-based contiguity spatial weights matrix $\boldsymbol{W}$.

3.2 Independent variables

The independent variables include demographic, human capital, labor market, housing, industry structure, and amenity and policy variables. In line with the literature, unless otherwise indicated, the initial values of the independent variable are used in the analysis. This type of formulation also reduces the problem of endogeneity. All the independent variables are in log form except those that can take negative or zero values. The descriptions of each of the independent variables of the models are given below.

Equation (4a) includes a vector of control variables ($\boldsymbol{X}_{kt-1}^{em}$) for $k = 1, \ldots, K_1$, which includes human capital, agglomeration effects, unemployment, and other regional socio-economic variables that are assumed to influence county employment growth (business growth) rate. Human capital is measured as the percentage of adults (over 25 years old) with college degrees and above (*POPCD*), and the percentage of adults (over 25 years old) with high school diploma (*POPHD*). It is expected that educational attainment is positively associated with employment growth. To control for agglomeration effects from both the supply and demand sides, county population size

(*POPs*) and the percentage of the population between 25 and 44 of age (*POP25-44*) are included and it is expected that agglomeration effects to have a positive impact on employment growth. The county unemployment rate (*UNEMP*) is included as a measure of local economic distress. Although a high county unemployment rate is normally associated with a poor economic environment, it may provide an incentive for individuals to form new businesses that can employ not only the owners, but also others. Thus, we do not know *a priori* whether the impact of *UNEMP* on employment growth is positive or negative. Establishment density (*ESBd*), which is the total number of private sector establishments in the county, divided by the county's population, is included to capture the degree of competition among firms and crowding of businesses relative to the population. The coefficient of *ESBd* is expected to be negative. Vector $\boldsymbol{X}^{em}_{kt-1}$ also includes *OWHU* (owner occupied housing) to capture the effects of the availability of resources to finance businesses and create jobs on employment growth in the county. The percentage of owner-occupied dwellings is expected to be positively associated with employment growth in the county. Also included in X^{em}_{kit} are property tax per capita (*PCPTAX*), percentage of private employment in manufacturing (*MANU*), percentage of private employment in wholesale and retail trade (*WHRT*), natural amenities index (*NAIX*), highway density (*HWD*), gross in-migration (*INM*), gross out-migration (*OTM*), median household income (*MHY*), and direct local government expenditures per capita (*GEX*). Since the percentage of the populations between 5 and 17 years of age (*POP5-17*) and above 65 years of age ($POP > 65$) do not constitute the prime working age of the population, they are not included in Eq. (4a). Direct federal expenditures and grants per capita (*DFEG*) in Appalachia have been mainly income support in the form of Food Stamps, Social Security Disability Insurance (SSDI), Temporary Assistance for Needy Families (TANF), and Supplemental Security Income (SSI) and hence not directly related to employment creation (Black and Sanders 2004). Homeownership (*OWHU*) and the social capital index (*SCIX*) are highly correlated. In order to avoid the problem of multicolinearity, *SCIX* is not included in Eq. (4a). *SCIX* is a county-level index that incorporates associational density of associations such as civic groups, religious organizations, sport clubs, labor unions, political and business organizations, percentage of voters who vote for presidential elections, county-level response rate to the Census Bureau's decennial census, and the number of tax-exempt non-profit organizations (Rupasingha et al. 2006).

We also use the natural amenity index created by McGranahan (1999) from standardized mean values of climate measures (January temperature, January days of sun, July temperature, and July humidity), topographic variation and water area as proportion of county area (see http://www.ers.usda.gov/Data/NaturalAmenities/natamenf.xls). Note that since both *SCIA* and *NAIX* are indices of many exogenous variables, they will constitute important parts of the instrument matrix that will be used to identify the endogenous variables of the system.

Equation (4b) contains a vector of exogenous variables ($\boldsymbol{X}^{mh}_{kt-1}, k = 1, \ldots, K_2$), which includes, among others, *POPs, POPd, FHHF, POPHD, UNEMP, MANU, WHRT*, and Social Capital Index (*SCIX*).

The initial levels of employment (EMP_{t-1}) and median household income (MHY_{t-1}) are also included in the respective equations of (4a, b). These variables are treated as predetermined variables because their values are given at the beginning of each period and hence are not affected by the endogenous variables. Table 1 provides the full list of the endogenous, and of the spatial lag and control variables, their descriptions and the sources of the data.

4 Estimation issues

Equations (4a, b) constitute a model with feedback simultaneity, spatial autoregressive lag simultaneity, and spatial cross-regressive lag simultaneity with spatially autoregressive disturbances. This creates complications, of which the choice of the functional form of each equation, whether or not each equation is identified, and the choice of the estimator and instruments are the most important ones.

Concerning the functional form, a generalized PE test was performed (Kmenta 1986, pp. 521–522; Mackinnon et al. 1983) to determine whether a linear or log-linear specification is most appropriate. The test indicates that the log-linear specification is preferred to the linear form for all equations. Thus, the model is specified in log-linear form with two modifications involving the measurement of the explanatory variables. First, the natural log formulation is dropped for explanatory variables that can assume negative or zero values. Second, lagged 1990 values are used for all explanatory variables to avoid simultaneity bias.

Concerning identification, first, for each equation, the number of basic endogenous variables that appear on the right hand side is smaller than the number of control variables that appear in the model but not in that equation. Second, in those cases where there are more instruments than needed to identify an equation, a test statistic[1] was computed (Hausman 1983) to investigate whether the additional instruments are valid in the sense that they are uncorrelated with the error term. That is $E(\boldsymbol{Q}'\boldsymbol{u}_r) = \mathbf{0}$, where $\boldsymbol{Q}$ is an instrument matrix as defined below. Fulfillment of this condition ensures that the instrument $\boldsymbol{Q}$ allows us to identify the regression parameters $[\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}']$ of Eqs. (4a, b), where $\boldsymbol{\alpha}'$ is a vector of slope coefficients and $\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}'$ are vectors of coefficients of the right-hand dependent variables, the spatial lag variables, and the predetermined variables, respectively.

As to the choice of estimator, the Method of Moments is preferred over the Maximum Likelihood approach because the latter would involve significant additional computational complexity.[2] The conventional three-stage least squares estimation to

[1] This test statistic is nR_u^2, where n is the sample size and R_u^2 is the usual R-squared of the regression of residuals from the second-stage equation on all included and excluded instruments. In other words, estimate Eqs. (4a, b) by GS2SLS or any efficient limited-information estimator and obtain the resulting residuals, $\hat{u}_r$. Then, regress these on all instruments and calculate nR_u^2. The statistic has a limiting chi-squared distribution with degree of freedom equal to the number of over-identifying restrictions, under the assumed specification of the model.

[2] In the Maximum Likelihood approach, the probability of the joint distribution of all observations is maximized with respect to a number of parameters. This involves the calculation of the Jacobian that appears in the log-likelihood function, which is computationally challenging. The complexity becomes overwhelming if the sample size is large, which applies in our case, and if the spatial weights matrices are not symmetric, which also applies in our case, even if the sample size is moderate (Kelejian and Prucha 1999, 1998). We also do not expect the error terms in our model to be normally distributed, which is required for the Maximum Likelihood procedure.

Table 1 Descriptive statistics

Variable code	Variable description	Mean	SD	Minimum	Maximum
Constant		1.00	0.00	1.00	1.00
EMPR	Employment Growth Rate 1990–2000	0.17	0.25	−0.69	1.79
MHYR	Median Household Income Growth Rate 1990–2000	0.48	0.31	−0.49	1.40
WEMPR	Spatial Lag of *EMPR*	0.18	0.14	−0.18	0.81
WMHYR	Spatial Lag of *MHYR*	0.47	0.19	−0.11	1.02
POPs	Population, 1990	10.30	0.94	7.88	14.11
POPd	Population Density, 1990	4.28	0.90	1.85	7.75
POP5-17	Percent of Population between 5–17 Years, 1990	2.92	0.12	2.17	3.22
POP25-44	Percent of Population between 25–44 Years Old, 1990	3.38	0.08	2.79	3.74
$POP > 65$	Percent of Population above 65 Years Old, 1990	2.60	0.20	1.55	3.20
FHHF	Percent of Female Householder, Family Householder, 1990	2.32	0.20	1.81	3.19
POPHD	Persons 25 Years and over, % High School only, 1990	4.10	0.17	3.57	4.47
POPCD	Persons 25 Years and over, % Bachelor's Degree or above, 1990	2.27	0.41	1.31	3.73
OWHU	Owner-Occupied Housing Unit in Percent, 1990	4.33	0.08	3.87	4.47
MHV	Median Value of Owner Occupied Housing 1990	10.74	0.26	9.67	11.68
UNEMP	Unemployment Rate 1990	2.15	0.35	1.22	3.25
AGFF	% Employed in Agriculture, Forestry and Fisheries 1990	3.62	2.66	0.00	17.10
MANU	% Employed in Manufacturing 1990	3.14	0.57	0.79	3.98
WHRT	% Employed in Wholesale and Retail Trade 1990	2.92	0.19	2.16	3.32
FIRE	% Employed Finance, Insurance and Real Estate 1990	1.23	0.33	0.00	2.23
HLTH	% Employed Health Service 1990	1.95	0.34	0.74	3.44
NAIX	Natural Amenities Index 1990	0.14	1.16	−3.72	3.55
ESBd	Establishment Density 1990	2.93	0.34	1.87	4.09
EFIR	Earnings in Finance Insurance and Real Estate 1990	21075.08	96011.09	0.00	1638807.0
CSBD	Commercial and Saving Banks Deposits 1990	12.21	1.07	8.83	16.95
DFEG	Direct Federal Expenditure and Grants per Capita 1990	7.99	0.38	6.98	10.18
FGCE	Federal Government Civilian Employment per 10,000 Pop. 1990	60.48	101.03	0.00	1295.00
PCTAX	Per Capital Local Tax 1990	5.91	0.53	4.51	7.42
PCPTAX	Property Tax Per Capita 1990	5.52	0.62	3.91	7.36
SCIX	Social Capital Index 1987	−0.60	0.94	−2.53	5.64
HWD	Highway Density 1990	0.69	0.40	−0.34	2.63

Table 1 continued

Variable code	Variable description	Mean	SD	Minimum	Maximum
ESBs	Establishment Size 1990	2.53	0.30	1.49	3.60
AWSR	Average Annual Wage and Salary Rate 1990	9.75	0.19	9.31	10.35
EMP	Employment 1990	8.83	1.25	5.42	13.38
INM	In-Migration 1990	7.09	1.00	4.54	10.52
OTM	Out-Migration 1990	7.04	0.97	4.50	10.55
MHY	Median Household Income 1989	9.94	0.23	9.06	10.68
GEX	Direct General Expenditures per Capita 1992	7.23	0.28	6.49	8.11

All variables are expressed in logs except AGFF, EFIR, FGCE, SCIX, and NAIX

handle the feedback simultaneity is inappropriate, because of the spatial autoregressive lag and spatial cross-regressive lag simultaneities terms. The Spatial Generalized Methods of Moments approach used by Rey and Boarnet (2004) in a Monte Carlo analysis of alternative approaches to modeling spatial simultaneity is also inappropriate, because the model includes spatially autoregressive disturbances. Therefore, we use the Generalized Spatial Two-Stage Least Squares (GS2SLS) as suggested by Kelejian and Prucha (1998, 1999), and the Generalized Spatial Three-Stage Least Squares (GS3SLS) approach as outlined by Kelejian and Prucha (2004).

The GS2SLS and GS3SLS procedures are carried out in three and four step routines, respectively. The first three steps are common to both routines. In the first step, the parameter vector $\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}'$ is estimated by two stage least squares (2SLS), using an instrument matrix $\boldsymbol{Q}$ that consists of a subset of linearly independent columns $\boldsymbol{X}, \boldsymbol{WX}, \boldsymbol{W^2X}$, where $\boldsymbol{X}$ is the matrix that includes the control variables in the model. $\boldsymbol{W}$ is a weights matrix. The disturbances for each equation in the model are computed using the estimates of $\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}'$ from the first step. In the second step, the estimates of the disturbances are used to estimate the autoregressive parameter ρ for each equation, using Kelejian and Prucha (2004) generalized moments procedure. In the third step, a Cochran–Orcutt-type transformation is performed, using the estimates for ρ from the second step to account for the spatial autocorrelation in the disturbances. The GS2SLS estimates of $[\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}']$ are then obtained by estimating the transformed model using a subset of the linearly independent columns of $[\boldsymbol{X}, \boldsymbol{WX}, \boldsymbol{W^2X}]$ as the instrument matrix.

Although the GS2SLS takes the potential spatial correlation into account, it does not utilize the information available across equations because it does not account for the potential cross equation correlation in the innovation vectors ($\varepsilon_{it}^{em}, \varepsilon_{it}^{mh}$). The correlation coefficient between the residuals of the GS2SLS (ε_{it}^{em} and ε_{it}^{mh}) is given in Table 2. The full system information is utilized by stacking the Cochran–Orcutt-type transformed equations (from the second step) in order to jointly estimate them. Thus, in the fourth step, the GS3SLS estimates of the betas, lambdas, and gammas $[\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}']$ are obtained by estimating this stacked model. The GS3SLS estimator is more efficient than the GS2SLS estimator. Further, consistent estimates of the covariance matrix are used to obtain the Feasible Generalized Three-Stage Least Squares (FGS3SLS) estimators of $\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}'$.

Table 2 Correlation matrix of the residuals from generalized spatial two-stage least squares (GS2SLS) estimation of the model

	Equation 1	Equation 2
Equation 1	1.0000	
Equation 2	−0.3974	1.0000

5 Discussion and analysis of results

The GS2SLS and GS3SLS parameter estimates of the system represented by Eqs. (4a, b) are reported in Table 3. These values are consistent with theoretical expectations and with the results of many other cross-sectional empirical studies (Boarnet 1994; Deller et al. 2001; Henry et al. 1997). The coefficients of the endogenous variables (EMPR and MHYR) are positive and statistically significant, indicating strong interdependence between employment and median household income growth rates. This interdependence is consistent with economic theory and empirical results. Increases in the demand for goods and services that result from increases in family median or per capita income are associated with increases in employment (Armington and Acs 2002), which create opportunities for even more people to work and earn income. However, the effect of median household income growth on employment growth is stronger than that of employment growth on median household income growth.

In the business employment (*EMPR*) equation, fifteen of the coefficient estimates are significantly different from zero at the 10% level or better. The results suggest a positive and significant parameter estimate for the spatial autoregressive lag variable (*WEMPR*). This indicates that employment growth tends to spill over to neighboring counties. The results also show a negative coefficient for (*WEMPR*) in the (*MHYR*) equation, indicating that employment growth rates in neighboring counties tend to unfavorably affect median household income growth rates (*MHYR*) in a given county. These estimates are important for policy because they indicate that employment growth in neighboring counties has positive and negative spillover effects on a given county's *EMPR* and *MHYR*, respectively. Furthermore, the significant spatial lag effects indicate that *EMPR* not only depends on characteristics within the county, but also on those of its neighbors. Hence, spatial effects should be tested empirically involving employment growth rates and household income growth rates. Our model specification incorporates a spatially autoregressive spatial process besides the spatial lag in the dependent variables. The negative estimate for ρ_1 (see Table 3) indicates that random shocks to *EMPR* do not only affect the county where the shocks originated and its neighbors, but also create negative shock waves across Appalachia.

To control for agglomeration effects, the model includes population statistics, such as the initial county population size (*POPs*) and the percentage of population between 25 and 44 years old (*POP25_44*). The result shows that both *POPs* and *POP25_44* have positive and significant effects on *EMPR*, even after accounting for potential spatial spillover effects. This result is consistent with the literature (Acs and Armington 2004) which indicates that a growing population increases the demand for consumer goods and services as well as the pool of potential entrepreneurs which encourage business formation. This result is important from a policy perspective. It indicates

Table 3 Generalized spatial 2SLS (GS2SLS) and full information generalized spatial 3SLS (GS3SLS) estimation results

Variables	GS2SLS				GS3SLS			
	EMPR Equation		MHYR Equation		EMPR Equation		MHYR Equation	
	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic
Constant	−7.5180***	−4.07	7.7602***	3.95	−8.53228***	−5.01698	8.6547***	4.714
EMPR			0.2825	1.66			0.6156***	4.0457
MHYR	0.1685	1.59			0.3735***	3.8956		
WEMPR	0.2492*	1.94	−0.1423	−0.98	0.2792**	2.2949	−0.2694*	−1.7
WMHYR	0.1657	1.44	−0.0559	−0.43	0.1147	1.1999	−0.1063	−0.8495
POPs	0.8367***	4.32	0.0877	0.78	0.7724***	4.3572	−0.0299	−0.2807
POPd	−0.0101	−0.3			−0.0123	−0.4054		
POP5-17			−0.1566	−0.9			−0.1072	−0.6642
POP25-44	0.2806	1.48			0.3093*	1.807		
$POP > 65$			0.1046	0.98			0.1576	1.6024
FHHF			−0.0031	−0.03			−0.0034	−0.3856
POPHD	−0.1589	−1.03	−0.2439	−1.15	−0.1487	−1.0167	−0.1556	−0.7667
POPCD	0.0561	1	−0.0989	−1.35	0.0789	1.4827	−0.1147	−1.6361
OWHU	−0.4079*	−1.77			−0.368*	−1.76		
MHV	−0.0309	−0.32	0.0955	0.76	−0.0483	−0.5198	0.0763	0.6308
UNEMP	−0.0825**	−2.05	0.0442	0.79	−0.079**	−2.0599	0.0706	1.3197
AGFF	−0.0055	−1.11	0.0025	0.38	−0.006	−1.2612	0.0032	0.5017
MANU	0.0856**	2.65	−0.0008	−0.02	0.0772**	2.5484	−0.0324	−0.8124
WHRT	0.3734***	4.5	−0.0727	−0.65	0.3719***	4.7178	−0.1916*	−1.8012
FIRE	0.0177	0.39	−0.0471	−0.86	0.0282	0.6542	−0.0616	−1.168
HLTH	−0.0079	−0.2	0.0297	0.56	−0.0157	−0.4067	0.0277	0.5475
NAIX	0.0072	0.72	−0.0063	−0.47	0.0062	0.645	−0.0064	−0.4944
ESBd	0.7049***	3.82	0.0242	0.27	0.6574***	3.9138	−0.0495	−0.5689
EFIR	−1.05216D-08	−0.09			−1.16242D-08	−0.1113		
CSBD	0.0406	1.14			0.0304	0.9565		
DFEG			0.0002	0.01			−0.0071	−0.1973
FGCE	0.0001	0.6			4.78*E*-05	0.5158		
PCTAX	−0.0706	−1.25			−0.062	−1.2314		
PCPTAX	0.0108	0.26			0.01095	0.2924		
SCIX			0.0439*	1.7			0.046*	1.974
HWD	−0.002	−0.04			−0.0062	−0.1303		
ESBs	0.5536**	2.87			0.5345***	3.0658		
AWSR	0.0912	0.94			0.0822	0.9521		
EMP	−0.8647***	−4.7	−0.0223	−0.28	−0.8151***	−4.8863	0.0941	1.2818
INM	0.1122	1.38	−0.1245	−1.25	0.1424*	1.8427	−0.1792*	−1.8725
OTM	−0.1382	−1.65	0.0693	0.65	−0.1401*	−1.7571	0.1248	1.215
MHY	0.2334	1.32	−0.7671***	−4.35	0.3636**	2.2161	−0.7976***	−4.7331
GEX	0.0608	1.33	0.0684	1.24	0.04105	0.9472	0.0477	0.8971
Rho (ρ)	−0.0428		0.1913		−0.0428		0.1913	

Table 3 continued

Variables	GS2SLS				GS3SLS			
	EMPR Equation		MHYR Equation		EMPR Equation		MHYR Equation	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
$nR^2 \sim \chi^2_{(30,36)}$ [a]	46.4608	0.02807[b]	39.1464	0.3305[b]	46.4608	0.02807[b]	39.1464	0.3305[b]
Moran I	−0.2058	−5.0284[c]	0.1336	3.0753[c]	−0.2058	−5.0284[c]	0.1336	3.0753[c]
Eta (η)	0.8647		0.7671		0.8151		0.7976	
Half-Life (years)	8.47		8.65		8.47		8.65	
PE test	log		log		log		log	
n	417		417		417		417	

*, **, and *** denote statistical significance level at the 10, 5, and 1%, respectively
[a] 30, 36 represent the degree of freedoms which are equal to the over-identifying restrictions in the EMPR, MHYR equations, respectively
[b] p-values
[c] Z-values for Moran I

that counties with high population concentration are benefiting from the resulting agglomerative and spillover effects that lead to localization of economic activities, in line with Krugman (1991a,b) argument on regional spillover effects.

The county unemployment rate (*UNEMP*) is included among the exogenous variables to measure local economic distress. The results suggest that a high unemployment rate is associated with low business growth. This indicates that the poor economic environment in Appalachia did not provide incentives for individuals to form new businesses that employ not only the owner, but others. Unemployed individuals may not have the capital to start a business. Furthermore, a high level of unemployment is indicative of a relatively low aggregate demand, which also discourages new firm formation. This result is consistent with the findings of Acs and Armington (2004). They found that unemployment is negatively associated with new firm formation during economic growth periods and positively during economic recession periods.

The coefficient of the variable representing the percentage of homes that are owned by their own occupants (*OWHU*) is negative and statistically significant at the 10% level. This result indicates that high home ownership is negatively associated with business formation in Appalachia. This is contrary to the expectation that high home ownership signals the availability of household assets and is therefore an indicator of the capacity to finance new businesses by potential entrepreneurs, either by using the house as collateral for loan or as indication of availability of other personal financial resources. The result, however, shows that in Appalachia during the study period home ownership was positively correlated with level of economic distress (Pollard 2003), and home ownership was higher in distressed counties (76%), and lowest in attainment counties (69%). Homeownership was also higher in central Appalachia (76%) than in the more developed northern or southern sub-regions; and Appalachia non-metro areas had higher ownership rates (76%) than its metro areas (72%). Thus, the result indicates that home ownership is not a good indicator of the availability of resources to start new business, at least in Appalachia.

The coefficients for *MANU* and *WHRT* are positive and significant at the 5 and 1% levels, respectively. These results indicate that counties with a higher initial percentage of their labor force employed in manufacturing and the wholesale and retail trade showed higher growth rates in business than other counties.

The percentage of people employed in manufacturing (*MANU*) and the percentage of people employed in wholesale and retail trade (*WHRT*) are included in the *EMPR* equation to control for the influence of sectoral employment concentration on the overall employment growth rate. The coefficient on *MANU* is positive and statistically significant at the 5% level, indicating a direct relationship between growths in overall employment and manufacturing employment at the beginning of the periods. The coefficient on *WHRT* is also positive and significant at the 1% level, indicating the positive role played by the service sector in expanding employment in Appalachia during the study period. Thus, these results tend to suggest that Appalachian counties that had a higher proportion of their labor force employed in manufacturing and whole sale and retail trade at the beginning the periods experienced higher growth rates in overall employment. This seems realistic since Appalachia has experienced a shift from resource-based economic activities to manufacturing and, particularly, to services. The coefficient on *WHRT* is higher and even more significant than the coefficient on *MANU* in the *EMPR* equation, indicating that the contribution of *WHRT* to overall employment growth was higher and more sustained than that of *MANU*.

Establishment density (*ESBd*), defined as the total number of private sector establishments in the county divided by the county's population, is included in the model to capture the degree of competition among firms and the concentration of businesses relative to the population density. The average size of establishment (*ESBs*), defined as total private sector employment divided by the number of private establishments in the county, is also included to capture the effects of barriers to entry of new small firms on employment growth. The coefficient for *ESBd* is positive and statistically significant at the 1% level, indicating that the Appalachian region is far below the threshold where competition among firms for consumer demands crowds businesses. According to the results, a high *ESBd* is associated with growth in employment (business growth), indicating that firms tend to locate near each other, possibly due to localization and agglomeration economies of scale. The coefficient for *ESBs* is also positive and significant indicating the existence of low barriers to new firm formation and employment generation in Appalachia during the study period.

The results indicate that the county employment level is dependent on gross in-migration, gross out-migration, and median household income. The coefficient for *INM*, for example, is positive and significant at the 5% level. The coefficient for *OTM* is negative and statistically significant at the 1% level. These are consistent with theoretical expectations and empirical findings (Borts and Stein 1964). In-migration tends to shift both the labor supply and labor demand curve right-wards, and out-migration tends to lead to leftward shift of the curves. Thus, in-migration leads to increases in employment, whereas out-migration leads to decreases in employment. A growing population increases the demand for consumer goods and services and is positively related to business formation (Acs and Armington 2004).

Consistent with theoretical expectations and empirical findings, the coefficient for *MHY* is positive and statistically significant at the 5% level. Increases in the demand for goods and services that result from increases in family median or per capita income are associated with increases in employment (Armington and Acs 2002).

An interesting observation from the empirical results pertains to the role of local government in employment growth. The model predicts that local governments, through their spending and taxation functions, play critical roles in creating and enabling economic environments for businesses to prosper. The empirical results, however, indicate that local governments have not played significant roles in employment growth in Appalachia. Given the economic hardship and high level of underdevelopment in Appalachia, these results are indications that local governments may need to reassess or step up their efforts to create incentives for employment growth in this region.

The elasticity of *EMPR* with respect to the initial employment level (*EMP*) is negative and statistically significant, indicating convergence in the sense that counties with low levels of employment at the beginning of the period (1990) tend to show a higher rate of business growth than counties with high initial levels of employment, conditional on the other explanatory variables. This result is consistent with prior studies on rural renaissance (Deller et al. 2001; Lundberg 2003). The speed of adjustment, η_{em}, is calculated at 0.8151, which indicates that just over 81% of the equilibrium rate of growth in the employment rate of growth was realized during the period 1990–2000. That is 8.151% annually, giving a half-life time of 8.47 years.

The parameter estimates for the *MHYR* equation also shows a positive estimate for ρ_2. This indicates that random shocks into the system with respect to *MHYR* not only affect the county where the shocks originate and its neighbors, but create positive spillover effects across Appalachia. The elasticity of *EMPR* with respect to the initial median household income (*MHY*) is negative and statistically significant, indicating convergence in the sense that counties with low median household incomes at the beginning of the period (1990) tend to show higher rates of growth of median household incomes than counties with high initial median household incomes, everything else being equal. The speed of adjustment, η_{mh}, is calculated at 0.7976, which indicates that about 80% of the equilibrium rate of growth in the median household income growth rate was realized during the period 1990–2000. That is 7.976% annually, giving a half-life time of 8.65 years. This result is comparable to the speed of convergence estimates obtained by Higgins et al. (2006) and Young et al. (2008).

The effect of out-migration on the growth rate of median household income is negative and statistically significant. If migrants' endowments of human capital in the form of education, accumulated skills, or entrepreneurial talents are higher compared to the sending population, then the loss of their skills, inventiveness and innovativeness would contribute to a decline in local productivity. Migrants may also own physical and financial capital that they may take with them leading to a loss in investment in the sending county. Moreover, out-migrants may contribute to a decline in the growth of markets and scale and agglomerations economies in the sending county. Such demand effects are the sources of loss in the growth of per capita personal incomes.

The coefficient for the index of social capital (*SCIX*) is positive and significant, suggesting that high levels of social capital increase the wellbeing of a county. The coefficients for the proportion of school age population (*POP5-17*), the proportion of the population above 65 years old ($POP > 65$), and the proportion of female headed households (*FHHF*) are negative, positive, and negative, respectively, as expected. Counties with high proportions of *POP5-17* and *FHHF* tend to have low levels of median household incomes, whereas counties with a high proportion of $POP > 65$ tend to have high levels of *MHY*. These results are consistent with empirical results of previous studies.

6 Conclusions

The main objective of this study was to test the hypotheses that (1) employment growth and median household income growth are interdependent and jointly determined by regional variables; (2) employment and median household income growth in a county are conditional upon initial conditions of the county; and (3) employment and median household income growth in a county are conditional upon employment and median household income growth in neighboring counties. To test these hypotheses, a spatial simultaneous equations model was developed. GS2SLS and GS3SLS coefficients of the parameters were obtained by estimating the model using data covering the 417 Appalachian counties for the 1990–2000 period. The empirical results of the study support the three hypotheses. In particular, the employment growth rate in one county is positively affected by the employment growth rate and the median household income growth rate in neighboring counties, and the median household income growth rate in one county is negatively affected by employment growth rate and median household income growth rate in neighboring counties.

A policy implication of the finding is that counties may be more successful in creating environments (business climate) to make themselves attractive to firms if several neighboring counties pool their resources. The results also indicate the presence of spatial correlation in the error terms, which implies that a random shock into the system spreads across the region. The results further indicate convergence across counties in Appalachia with respect to employment growth and median household income growth rates, conditional upon the initial conditions of the explanatory variables in the model. This information indicates that the divergence in the economic status among Appalachian counties is narrowing and could mean that the efforts of the Appalachian Regional Commission are showing results.

The empirical results indicate the presence of significant agglomerative effects: counties with higher population concentrations showed significant business growth. Combined with the findings of spillover effects, this might justify favoring focusing investments in areas capable of generating agglomeration effects.

The study also produces useful information concerning the creation of new or the expansion of existing businesses in Appalachia. Establishment density, which captures the degree of competition among firms and crowding of businesses relative to the population, indicates that Appalachia is below the threshold where competition among firms for consumer demands crowds businesses. In addition, the results indicate low barriers to new firm formation and employment generation during the study period.

While incorporating spatial interdependencies adds to the model's computational complexities, the returns are not only improved estimates, but the analysis also yields information about spatial relationships that would not otherwise be available. For the study period, this research suggests that a growth pole approach that spatially concentrates scarce policy investments could benefit the region. Such insight requires a spatially explicit model otherwise they are based on guesswork and intuition. Of course, given the short time period of our analysis, additional research is needed to determine if this result is stable over time or changes with the business cycle.

In general, this study confirms the importance of spatial effects in regional development. The empirical results indicate the presence of spatial correlation in the error terms and of spatial autoregressive lag. Failure to account for spatial interaction effects results in less efficient and consistent estimates, as well as loss of insight.

Acknowledgments This research was partially funded by the West Virginia Agricultural and Forestry Experiment Station. We acknowledge helpful comments by Dale Colyer and two referees. We thank Anil Rupasingha, Stephan Goetz and David Freshwater for allowing the use of their Social Capital Index data set for US counties. The usual caveat applies.

Appendix A: Derivation of the reduced form of the model

Let the system given in (4a, b) be written as:

$$\boldsymbol{Y} = \boldsymbol{YB} + \boldsymbol{X\Gamma} + \boldsymbol{WY\Lambda} + \boldsymbol{U}. \qquad \text{(I)}$$

$$\boldsymbol{U} = \boldsymbol{WUC} + \boldsymbol{E} \quad \text{and}$$

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_G) \quad \boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K) \quad \boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_G)$$

$$\boldsymbol{WU} = (\boldsymbol{Wu}_1, \ldots, \boldsymbol{Wu}_G), \boldsymbol{C} = \text{diag}_{j=1}^{G}\left(\rho_j\right), \boldsymbol{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_G)$$

where $\boldsymbol{y}_j$ is the n by 1 vector of cross sectional observations on the dependent variable in the jth equation, $\boldsymbol{x}_l$ is an n by 1 vector of cross sectional observations on the jth exogenous variable, $\boldsymbol{u}_j$ is an n by 1 vector of error terms in the jth equation, and $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are correspondingly defined parameter matrices of dimension G by G and K by G, respectively. $\boldsymbol{B}$ is a diagonal matrix. $\boldsymbol{\Lambda}$ is G by G matrix of parameter estimates of the spatial lag variables. It not diagonal and hence each equation includes spatial cross-regressive lag variable in addition to its own spatial lag. Hence the model has the same structure as that in Kelejian and Prucha (2004).

Note that ρ_j denotes the spatial autoregressive parameter in the jth equation and since $\boldsymbol{C}$ is taken to be diagonal, the specification relates the disturbance vector in the jth equation only to its own spatial lag. Since it is assumed that $\boldsymbol{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $\boldsymbol{E}(\boldsymbol{\varepsilon\varepsilon}') = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, the disturbances, however, will be spatially correlated across units and across equations.

The system in Eq. (I) can be expressed in a form where its solution for the endogenous variables is clearly revealed. But, first consider the following vector transformations:

$$\begin{aligned}
vec(\boldsymbol{Y}) &= vec(\boldsymbol{YB}) + vec(\boldsymbol{X\Gamma}) + vec(\boldsymbol{WY\Lambda}) + vec(\boldsymbol{U})\\
vec(\boldsymbol{Y}) &= vec(\boldsymbol{YB}) + vec(\boldsymbol{X\Gamma}) + vec(\boldsymbol{WY\Lambda}) + vec(\boldsymbol{UWC} + \boldsymbol{E})\\
&= \left(\boldsymbol{B}' \otimes \boldsymbol{I}\right) vec(\boldsymbol{Y}) + \left(\boldsymbol{\Gamma}' \otimes \boldsymbol{I}\right) vec(\boldsymbol{X}) + \left(\boldsymbol{\Lambda}' \otimes \boldsymbol{W}\right) vec(\boldsymbol{Y})\\
&\quad + \left(\boldsymbol{C}' \otimes \boldsymbol{W}\right) \mathrm{vec}\boldsymbol{U} + vec\boldsymbol{E}
\end{aligned}$$

Letting $\boldsymbol{y} = vec(\boldsymbol{Y}), \boldsymbol{x} = \mathrm{vec}(\boldsymbol{X}), \boldsymbol{u} = \mathrm{vec}(\boldsymbol{U})$, and $\boldsymbol{\varepsilon} = \mathrm{vec}(\boldsymbol{E})$, it follows from Eq. (I) that:

$$\begin{aligned}
&\mathbf{y} = \left(\boldsymbol{B}' \otimes \boldsymbol{I}\right)\boldsymbol{y} + \left(\boldsymbol{\Gamma}' \otimes \boldsymbol{I}\right)\boldsymbol{x} + \left(\boldsymbol{C}' \otimes \boldsymbol{W}\right)\boldsymbol{u} + \varepsilon\\
&\quad \text{or}\\
&\boldsymbol{y} = \left(\boldsymbol{B}' \otimes \boldsymbol{I}\right)\boldsymbol{y} + \left(\boldsymbol{\Gamma}' \otimes \boldsymbol{I}\right)\boldsymbol{x} + \boldsymbol{u},\\
&\boldsymbol{u} = \left(\boldsymbol{C}' \otimes \boldsymbol{W}\right)\boldsymbol{u} + \boldsymbol{\varepsilon}
\end{aligned} \tag{II}$$

Let $\boldsymbol{B}^* = [(\boldsymbol{B}' \otimes \boldsymbol{I}) + (\boldsymbol{\Lambda}' \otimes \boldsymbol{W})]$, $\boldsymbol{\Gamma}^* = (\boldsymbol{\Gamma}' \otimes \boldsymbol{I})$ and $\boldsymbol{C}^* = \boldsymbol{C}' \otimes \boldsymbol{W} = \mathrm{diag}_{j=1}^{G}$ $(\rho_j \boldsymbol{W})$, then Eq. (II) can be written in more compact form as:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{B}^*\boldsymbol{y} + \boldsymbol{\Gamma}^*\boldsymbol{x} + \boldsymbol{u},\\
\boldsymbol{u} &= \boldsymbol{C}^*\boldsymbol{u} + \boldsymbol{\varepsilon}
\end{aligned} \tag{III}$$

Assuming that $\boldsymbol{I}_{nG} - \boldsymbol{B}^*$ and $\boldsymbol{I}_{nG} - \boldsymbol{C}^*$ are nonsingular matrices with $|\rho_j| < 1$, $j = 1, \ldots, G$, the system in Eq. (III) can be expressed in its reduced form as:

$$\begin{aligned}
\boldsymbol{y} &= \left(\boldsymbol{I}_{nG} - \boldsymbol{B}^*\right)^{-1}\left(\boldsymbol{\Gamma}^*\boldsymbol{x} + \boldsymbol{u}\right),\\
\boldsymbol{u} &= \left(\boldsymbol{I}_{nG} - \boldsymbol{C}^*\right)^{-1}\varepsilon
\end{aligned} \tag{IV}$$

Based on the results of our estimation, we found that $\boldsymbol{I}_{nG} - \boldsymbol{B}^*$ and $\boldsymbol{I}_{nG} - \boldsymbol{C}^*$ have full column ranks and $|\rho_j| < 1$, $j = 1, 2$. From this we can conclude that the reduced form of the system [Eq. (IV)] is properly defined and there also exists spatial multiplier working in the system.

References

Acs ZJ, Armington C (2003) Endogenous growth and entrepreneurial activity in cities. http://ideas.repec.org/p/cen/wpaper/03-02.html. Accessed 8 December 2008

Acs ZJ, Armington C (2004) The impact of geographic differences in human capital on service firm formation rates. J Urban Econ 56:244–278

Acs ZJ, Audretsch DB (1993) Introduction. In: Acs ZJ, Audretsch DB (eds) Small firms and entrepreneurship: an east-west perspective. Cambridge University Press, Cambridge

Acs ZJ, Audretsch DB (2001) The Emergence of the Entrepreneurial Society. Present. for the accept. of the Int. Award for Entrepr. and Small Bus. Res., Stockh, 3 May

Anselin L (2003) Spatial externalities, spatial multipliers and spatial econometrics. Int Reg Sci Rev 26(2):153–166

Anselin L (1988) Spatial econometrics: methods, and models. Kluwer, Dordrecht

Anselin L, Kelejian HH (1997) Testing for spatial error autocorrelation in the presence of endogenous regressors. Int Reg Sci Rev 20(1&2):153–182

Arbia G, Basile R, Piras G (2005) Using panel data in modelling regional growth and convergence. Reg. Econ Appl Lab Work Pap No. 55. Univ. Ill, Urbana-Champaign
Armington C, Acs ZJ (2002) The determinants of regional variation in new firm formation. Reg Stud 36(1):33–45
Aronsson T, Lundberg J, Wikstrom M (2001) Regional income growth and net migration in Sweden 1970–1995. Reg Stud 35(9):823–830
Audretsch DB, Fritsch M (1994) The geography of firm births in Germany. Reg Stud 28(4):359–365
Audretsch DB, Carree MA, van Stel AJ, Thurik AR (2000) Impeded Industrial Restructuring: The Growth Penalty. Res Pap Cent for Adv Small Bus Econ Erasmus Univ., Rotterdam
Barkley DL, Henry MS, Bao S (1998) The role of local school quality and rural employment and population growth. Rev Reg Stud 28(1):81–102
Barro RJ, Sala-i-Martin X (1992) Convergence. J Polit Econ 100:223–251
Barro RJ, Sala-i-Martin X (2004) Economic growth, 2nd edn. MIT Press, Cambridge
Black DA, Sanders SG (2004) Labor market performance, poverty, and income inequality in Appalachia. http://www.arc.gov/images/reports/labormkt/labormkt.pdf. Accessed 8 December 2008
Boarnet MG (1994) An empirical model of intra-metropolitan population and employment growth. Pap Reg Sci 73(2):135–153
Borts GH, Stein JL (1964) Economic growth in a free market. Columbia University Press, New York
Brock WA, Evans DS (1989) Small business economics. Small Bus Econ 1(1):7–20
Callejon M, Segarra A (2001) Geographical determinants of the creation of manufacturing firms: the regions of Spain. http://www.ub.es/graap/pdfcallejon/RS01.pdf. Accessed 8 December 2008
Carlino OG, Mills ES (1987) The determinants of county growth. J Reg Sci 27(1):39–54
Carree MA, Thurik AR (1998) Small firms and economic growth in Europe. Atl Econ J 26(2):137–146
Carree MA, Thurik AR (1999) Industrial structure and economic growth. In: Audretsch DB, Thurik AR (eds) Innovation, industry evolution and employment. Cambridge University Press, Cambridge
Clark D, Murphy CA (1996) Countywide employment and population growth: an analysis of the 1980s. J Reg Sci 36(2):235–256
Davidson P, Lindmark L, Olofsson C (1994) New firm formation and regional development in Sweden. Reg Stud 28(4):395–410
Deller SC, Tsai TH, Marcouiller DW, English DBK (2001) The role of amenities and quality of life in rural economic growth. Am J Agric Econ 83(2):352–365
Duffy NE (1994) The determinants of state manufacturing growth rates: a two-digit-level analysis. J Reg Sci 34(2):137–162
Duffy-Deno KT (1998) The effect of federal wilderness on county growth in the inter-mountain western United States. J Reg Sci 38(1):109–136
Duffy-Deno KT, Eberts RW (1991) Public infrastructure and regional economic development: a simultaneous equations approach. J Urban Econ 30(3):329–343
Edmiston KD (2004) The net effect of large plant locations and expansions on county employment. J Reg Sci 44(2):289–319
Ekstrom B, Leistritz FL (1988) Rural community decline and revitalization: an annotated bibliography. Garland Publ, New York
Ertur C, Le Gallo J, Baumont C (2006) The European regional convergence process, 1980–1995: do spatial regimes and spatial dependence matter? Int Reg Sci Rev 29(1):3–34
Fotopoulos G, Spencer N (1999) Spatial variations in new manufacturing plant openings: some empirical evidence from greece. Reg Stud 33(3):219–229
Fritsch M (1992) Regional differences in new firm formation: evidence from West Germany. Reg Stud 26(3):233–244
Fritsch M, Falck O (2003) New firm formation by industry over space and time: a multilevel analysis. Discuss Pap German Inst for Econ Res Berlin
Garofoli G (1994) New firm formation and regional development: the Italian case. Reg Stud 28(4):381–393
Glaeser EL, Scheinkman JA, Shleifer A (1995) Economic growth in a cross-section of cities. J Monet Econ 36(1):117–143
Greenwood MJ, Hunt GL (1984) Migration and interregional employment redistribution in the United States. Am Econ Rev 74(5):957–969
Greenwood MJ, Hunt GL, McDowel JM (1986) Migration and employment change: empirical evidence on spatial and temporal dimensions of the linkage. J Reg Sci 26(2):223–234

Guesnier B (1994) Regional variation in new firm formation in France. Reg Stud 28(4):347–358
Hamalainen K, Bockerman P (2004) Regional labor market dynamics, housing, and migration. J Reg Sci 44(3):543–568
Hausman J (1983) Specification and estimation of simultaneous equations models. In: Griliches Z, Intriligator M (eds) Handbook of econometrics. North Holland, Amsterdam
Hart M, Gudgin G (1994) Spatial variations in new firm formation in the Republic of Ireland, 1980–1990. Reg Stud 28(4):367–380
Henry MS, Barkley DL, Bao S (1997) The hinterland's stake in metropolitan growth: evidence from selected southern regions. J Reg Sci 37(3):479–501
Henry MS, Schmitt B, Kristensen K, Barkley DL, Bao S (1999) Extending Carlino-Mills models to examine urban size and growth impacts on proximate rural areas. Growth Change 30(4):526–548
Higgins MJ, Levy D, Young AT (2006) Growth and convergence across the US: evidence from county-level data. Rev Econ Stat 88(4):671–681
Isserman AM (1993) State economic development policy and practice in the United States: a survey article. Int Reg Sci Rev 16(1–2):49–100
Johnson P, Parker S (1996) Spatial variations in the determinants and effects of firm births and deaths. Reg Stud 30(7):676–688
Kangasharju A (2000) Regional variations in firm formation: panel and cross-section data evidence from Finland. Reg Sci 79(4):355–373
Keeble D, Walker S (1994) New firms, small firms and dead Firms: spatial pattern and determinants in the United Kingdom. Reg Stud 28(4):411–427
Kelejian HH, Prucha IR (1998) A generalized two-stage least squares procedure for estimating a spatial autoregressive model with spatial autoregressive disturbances. J Real Estate Finance Econ 17(1):99–121
Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40(2):509–533
Kelejian HH, Prucha IR (2001) On the asymptotic distribution of the Moran I test statistic with applications. J Econ 104(2):219–257
Kelejian HH, Prucha IR (2004) Estimation of simultaneous systems of spatially interrelated cross sectional equations. J Econ 118(1):27–50
Kmenta J (1986) Elements of econometrics. Macmillan, New York
Krugman P (1991a) Increasing returns and economic geography. J Polit Econ 99(3):483–499
Krugman P (1991b) Geography and trade. MIT Press, Cambridge
Lewis DJ, Hunt GL, Plantinga AJ (2002) Does public land policy affect local wage growth. Growth Change 34(1):64–86
Loveman G, Sengenberger W (1991) The re-emergence of small-scale production: an international comparison. Small Bus Econ 3(1):1–37
Lundberg J (2003) On the determinants of average income growth and net migration at the municipal level in Sweden. Rev Reg Stud 32(2):229–253
MacDonald JF (1992) Assessing the development status of metropolitan areas. In: Mills ES, MacDonald JF (eds) Sources of metropolitan growth. Cent. for Urban Policy Res, New Brunswick
Mackinnon JG, White H, Davidson R (1983) Tests for model specification in the presence of alternative hypotheses: Some further results. J Econ 21(1):53–70
McGranahan DA (1999) Natural amenities drive rural population change. http://www.ers.usda.gov/publications/aer781/aer781i.pdf. Accessed 9 December 2008
Mills ES, Price R (1984) Metropolitan suburbanization and central city problems. J Urban Econ 15(1):1–17
Persson J (1997) Convergence across the Swedish counties, 1911–1993. Eur Econ Rev 41(9):1834–1852
Pollard KM (2003) Appalachia at the millennium: an overview of the results from census 2000. Popul. Ref. Bur., Washington
Pulver GC (1989) Developing a community perspective on rural economic development policy. J Community Dev Soc 20(2):1–4
Rappaport J (1999) Local growth empirics. Cent for Int Dev Work Pap No. 23. Harvard University Press, Cambridge
Rey SJ, Boarnet MG (2004) A taxonomy of spatial econometric models for simultaneous equations systems. In: Anselin L, Florax RJGM, Rey SJ (eds) Advances in spatial econometrics: methodology, tools and applications. Springer, Berlin

Reynolds PD (1994) Autonomous firm dynamics and economic growth in the United States, 1986–1990. Reg Stud 28(4):429–442

Rupasingha A, Goetz SJ, Freshwater D (2006) The production of social capital in US counties. J Socio-Econ 35(1):83–101

Steinnes DN, Fisher WD (1974) An econometric model of intra-urban location. J Reg Sci 14(1):65–80

US Census Bureau (2005) Mean travel time to work for workers 16 years and over who did not work at home (Minutes): 2005. (2005 American Community Survey). http://factfinder.census.gov/servlet/DatasetMainPageServlet?_ds_name=ACS_2005_EST_G00_&_lang=en&_ts=199031476495. Accessed 4 June 2007

US Small Business Administration (SBA) (1999) The state of small business: a report of the President. US Gov Print Press, Washington

Wennekers S, Thurik AR (1999) Linking entrepreneurship and economic growth. Small Bus Econ 13(1):27–55

Young AT, Higgins MJ, Levy D (2008) Sigma convergence versus beta convergence: evidence from US county-level data. J Money Credit Bank 40(5):1083–1093