NBER WORKING PAPER SERIES

SPATIAL DEVELOPMENT

Klaus Desmet
Esteban Rossi-Hansberg

Working Paper 15349
http://www.nber.org/papers/w15349

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2009

We thank Elhanan Helpman, Gianmarco Ottaviano, Diego Puga, Steve Redding and various seminar participants for useful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Spatial Development
Klaus Desmet and Esteban Rossi-Hansberg
NBER Working Paper No. 15349
September 2009
JEL No. E32,O11,O18,O33,R12

ABSTRACT

We present a theory of spatial development. A continuum of locations in a geographic area choose each period how much to innovate (if at all) in manufacturing and services. Locations can trade subject to transport costs and technology diffuses spatially across locations. The result is an endogenous growth theory that can shed light on the link between the evolution of economic activity over time and space. We apply the model to study the evolution of the U.S. economy in the last few decades and find that the model can generate the reduction in the employment share in manufacturing, the increase in service productivity in the second part of the 1990s, the increase in land rents in the same period, as well as several other spatial and temporal patterns.

Klaus Desmet
Department of Economics
Universidad Carlos III
28903 Getafe (Madrid)
Spain
klaus.desmet@uc3m.es

Esteban Rossi-Hansberg
Princeton University
Department of Economics
Fisher Hall
Princeton, NJ 08544-1021
and NBER
erossi@princeton.edu

Spatial Development*

Klaus Desmet
Universidad Carlos III

Esteban Rossi-Hansberg
Princeton University

September 7, 2009

Abstract

We present a theory of spatial development. A continuum of locations in a geographic area choose each period how much to innovate (if at all) in manufacturing and services. Locations can trade subject to transport costs and technology diffuses spatially across locations. The result is an endogenous growth theory that can shed light on the link between the evolution of economic activity over time and space. We apply the model to study the evolution of the U.S. economy in the last few decades and find that the model can generate the reduction in the employment share in manufacturing, the increase in service productivity in the second part of the 1990s, the increase in land rents in the same period, as well as several other spatial and temporal patterns.

1. INTRODUCTION

Economic development varies widely across space. It is a common observation, as stated in the 2009 World Development Report, that the location of people is the best predictor of their income. This is clearly true when we move across countries, but there is also significant variation within countries. In the U.S., employment concentration and value added vary dramatically across space, and so does the rate of growth (see, e.g., Desmet and Rossi-Hansberg, 2009). Even though a casual look at the spatial landscape makes these observations seem almost trivial, there has been little work incorporating space, and

*We thank Elhanan Helpman, Gianmarco Ottaviano, Diego Puga, Steve Redding and various seminar participants for useful comments.

the economic structure implied by space, into modern endogenous growth theories. This paper addresses this shortcoming by presenting a dynamic theory of spatial development and contrasting its predictions with evidence on the spatial evolution of the U.S. in the last few decades.

The theory we present has four main components. First, it includes a continuum of locations that can produce in two industries: manufacturing and services. Production requires labor and land, with technologies being constant returns to scale in these two inputs. Since the amount of land at a given location is fixed, the actual technology experienced at a location exhibits decreasing returns to scale. This constitutes a congestion force. Second, locations can trade goods and services by incurring iceberg transport costs. Given these costs, national goods markets in both sectors clear in equilibrium. Labor is freely mobile and workers can relocate every period. As a result, in a given period all workers obtain a common utility level in equilibrium. Third, locations invest in innovation. Each location can decide to tax its firms and use the revenue to buy a probability of drawing a proportional shift in its technology from a given distribution. Hence, some locations may decide not to invest in technology, others may decide to invest but may be unlucky and not get a draw, and still others will get a draw and innovate. The benefits from innovation for a location last for only one period, since in subsequent periods land and labor arbitrage the gains away. The more labor works in a location before innovating, the more a potential innovation can be exploited next period, and thus the greater the incentives to invest. The model therefore exhibits a local scale effect in innovation, implying that more dense locations innovate more. Fourth, technology diffuses spatially. Locations close to others with a high technology get access to that technology through diffusion. Each location will produce using the best technology it has access to, whether through invention or diffusion.

We contrast the theory to U.S. macroeconomic and spatial data for the last two decades. A well known fact is that the employment share in manufacturing has declined over time and, correspondingly, the employment share in services has increased. This shift has been accompanied by a decline in the relative price of manufactured goods (see, e.g., Buera and Kaboski, 2007). Ngai and Pissarides (2007) show that a faster increase in manufacturing productivity, relative to service productivity, together with CES preferences and an elasticity of substitution less than one, can yield these effects. Our model starts off with a similar story. Initial conditions are such that in the beginning locations specializing in manufacturing innovate more. This implies a reduction in the manufacturing share and a drop in the relative price of manufactured goods, just as in Ngai and Pissarides (2007). Where we differ is that in our model this reallocation of employment toward services at some point accelerates innovation in some locations specializing in services. From then onward service productivity increases together with manufacturing productivity, ultimately leading to a fairly constant growth path in both industries and the economy. This is consistent with the evidence on manufacturing and service productivity in Triplett and Bosworth (2004), who document an acceleration in service productivity growth starting around 1995, while manufacturing productivity keeps growing at around 2% per year throughout.[1] Our model also generates a corresponding increase in land rents around that period, a prediction that is very clearly present in the data. Real wage growth exhibits a similar pattern, which is likewise corroborated by the data.

With respect to the spatial dimension, the theory predicts that, initially, when service productivity is about stagnant, manufacturing is more concentrated than services. Once the service sector starts innovating, concentration in the service sector increases in terms of both employment and productivity, implying a positive link between employment density, innovation and productivity growth. These theoretical predictions are borne out in the data: over the last decades the service sector has become more concentrated, in terms of both employment and productivity, making it look increasingly similar to manufacturing along this spatial dimension. This is consistent with the evidence in Desmet and Rossi-Hansberg (2009), who compare spatial growth in two different time periods, 1980-2000 and 1900-1920, and find that service growth at the end of the 20th century looked very similar

[1] Table A-4 in Triplett and Bosworth (2004) shows that growth in value added per worker in goods-producing sectors went from 2.11% between 1987 and 1995 to 1.94% between 1995 and 2001. In contrast, in service-producing sectors the growth rate went from 0.78% to 2.49%. If we focus only on the contribution of TFP, the difference is smaller but still there: growth in TFP went from 0.75% to 1.29% in goods-producing sectors and from 0.41% to 1.41% in service-producing sectors. (Note that since our model does not include capital, the value added per worker measure is more appropriate than the TFP measure.)

to manufacturing growth at the beginning of the 20th century. Both industries, in very different time periods, exhibited increasing concentration in medium-size locations.[2]

Since our theory incorporates both a time and a space dimension, it provides a link between the location decision of agents and their decision to innovate. Two parameters that govern this link are transport costs and the elasticity of substitution. Even though increases in transport cost lead to the standard static losses familiar from trade models, they also lead to dynamic gains by generating denser areas that, together with the scale effect in innovation, lead to faster growth.

Decreasing the elasticity of substitution between manufacturing and services implies that agents can substitute less and therefore prefer to be closer to areas specializing in the sector in which they do not work. This puts a break on the emergence of large clusters, as such clusters would increase the average distance to locations that specialize in the other sector. The result is more dispersion and therefore less innovation. However, lowering the elasticity of substitution also implies that agents in manufacturing areas, where the relative price of services is high, consume a greater share of their income on services. This increases the scale of service producers located close-by, thus leading to more innovation. The result is a non-monotonic pattern in the effects of the elasticity of substitution on location and growth. The first effect dominates for high values of the elasticity of substitution whereas the second effect dominates for lower values. To our knowledge, these spatial-dynamic effects are novel.

The existing literature on spatial dynamic models is fairly small. There is a successful literature in trade that has focused on dynamic models with two or more countries (see, among others, Grossman and Helpman, 1991, Eaton and Kortum, 1999, Young 1991, and Ventura, 1997).[3] The main difference with our work is that in these models there is no geography in the sense that locations are not ordered in space. In fact, most of these papers do not even introduce transport costs, let alone geography. In contrast, we introduce a continuum of locations on a line. Locations are therefore ordered geographically, and both

[2] However, Desmet and Rossi-Hansberg (2009) do not link their findings to the structural transformation and to other macroeconomic variables, which is the main focus of this paper.

[3] See also Baldwin and Martin (2004) for a survey of similar work within the 'New Economic geography' model.

transport costs and technology diffusion are affected by distance.

Incorporating a continuum of locations into a dynamic framework is a complicated task for two reasons: it increases the dimensionality of the problem by requiring agents to understand the distribution of economic activity over time and over space, and clearing goods and factor markets is complex because prices depend on trade and mobility patterns at all locations. These two difficulties make spatial dynamic models intractable, and the only way forward is to simplify the problem. A set of recent papers, such as Quah (2002), Boucekkine et al. (2009), and Brock and Xepapadeas (2008a,b), introduce a continuum of locations with geography and simplify the problem by assuming that each point in space is isolated, except for spatial spillovers or diffusion. By abstracting from transport costs, national goods markets, and factor mobility, they save the need to calculate price functions across locations over time. By imposing enough structure, they are able to mathematically characterize some aspects of social optima or equilibrium allocations, though they fall short of proposing a complete solution. In addition, they are unable to connect to the data, since the simplifying assumptions do not yield empirical predictions that are rich enough.

In contrast, our main goal is to propose a theoretical framework that can be used to study the spatial evolution of the U.S. economy over the last decades. To do so, it is crucial to have a model that is rich enough to capture a variety of spatial patterns of the economy. In order to deal with the complexity of forward-looking agents in a spatial context, the previous papers had to sacrifice many of the relevant spatial interactions. Another way around this problem, and the one we will follow, is to impose enough structure — through the diffusion of technology, the mobility of factors and the land ownership structure— that agents do not care to take their future allocations paths into account, given that they do not affect the returns of their current decisions. As for the problem of clearing factor and goods markets in a framework with a continuum of locations, we follow the method in Rossi-Hansberg (2005) that consists of clearing markets sequentially. These two simplifications are key to making a rich structure solvable and computable.

In Desmet and Rossi-Hansberg (2009) we use a similar methodology to study the dynamics of manufacturing and service growth across U.S. counties in the 20th century. Although that model also analyzes the link between innovation and spatial growth, our current paper is different in two ways. First, we explicitly model innovation as the outcome of a profit-maximizing problem and, in that sense, provide micro-foundations for why certain locations innovate more than others. Second, in Desmet and Rossi-Hansberg (2009) innovation in a given sector gets jump-started exogenously, thus making its timing *ad hoc* and independent of what is happening in the other sector. In our current paper innovation starts off endogenously as explained above.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 presents the data we use to empirically explore the theoretical predictions, carries out numerical simulations of the model, and discusses the link between our results and the data. In Section 3, we also analyze at length the novel spatial effects that result from changes in transport costs and the elasticity of substitution. Section 4 concludes.

2. THE MODEL

The economy consists of land and people located in the closed interval $[0,1]$. Throughout we refer to a location as a point in this interval, and we let the density of land at each location ℓ be equal to one. Hence, the total mass of land in the economy is equal to 1. We divide space into regions or 'counties' (connected intervals in $[0,1]$), each of which has a local government. For simplicity we make all counties of equal size, I. The total number of agents is given by $\overline{L}$, and each of them is endowed with one unit of time each period. Agents are infinitely lived.

2.1 Preferences

Agents live where they work and they derive utility from the consumption of two goods: manufactures and services. Labor is freely mobile across locations and sectors. Agents supply their unit of time inelastically in the labor market. They order consumption bundles according to an instantaneous utility function $U(c_M, c_S)$ with standard properties, where c_i denotes consumption of good $i \in \{M, S\}$. We also assume that $U(\cdot)$ is homogeneous of degree one. Agents are assumed to hold a diversified portfolio of land in all locations. Goods are non-storable, and there is no other savings technology apart from land. The problem of an agent at a particular location ℓ is given by[4]

$$\max_{\{c_i(\ell,t)\}_0^\infty} E \sum_{t=0}^{\infty} \beta^t U(c_M(\ell,t), c_S(\ell,t)) \tag{1}$$

$$s.t. \qquad w(\ell,t) + \frac{\bar{R}(t)}{\bar{L}} = p_M(\ell,t)\, c_M(\ell,t) + p_S(\ell,t)\, c_S(\ell,t)$$

for all t and ℓ.

where $p_i(\ell,t)$ denotes the price of good i, $w(\ell,t)$ denotes the wage at location ℓ and time t, and $\bar{R}(t)$ denotes total land rents per unit of land, so that $\bar{R}(t)/\bar{L}$ is the dividend from land ownership (since $\bar{L}$ is total population size) given that agents hold a diversified portfolio of land in all locations. The first-order conditions of this problem yield $U_i(c_M(\ell,t), c_S(\ell,t)) = \lambda(\ell,t)\, p_i(\ell,t)$, for all $i \in \{M,S\}$, where $U_i(\cdot)$ is the marginal utility of consuming good i and $\lambda(\ell,t)$ is a location- and time-specific Lagrange multiplier. Denote by $\bar{U}(p_M(\ell), p_S(\ell), w(\ell) + R(\ell)/L(\ell))$ the indirect utility function of an agent at location ℓ. Because of free mobility of labor, it must be the case that

$$\bar{U}\left(p_M(\ell), p_S(\ell), w(\ell) + \bar{R}/\bar{L}\right) = \bar{u}, \text{ for all } \ell \in [0,1], \tag{2}$$

where $\bar{u}$ is determined in equilibrium. In the numerical examples in the next section we will use a CES specification

$$U(c_M, c_S) = (h_M c_M^\alpha + h_S c_S^\alpha)^{1/\alpha} \tag{3}$$

with elasticity of substitution $1/(1-\alpha) < 1$.

2.2 Technology

Each location can produce in both sectors or specialize in one of them. The inputs of production are land and labor. Production per unit of land in the manufacturing sector is

[4] Since we assume labor mobility, utility levels equalize across space each period, so that we can study the optimization problem of an agent as if he stays in the same location forever.

given by

$$M\left(L_M\left(\ell,t\right)\right)=Z_M\left(\ell,t\right)^{\gamma}L_M\left(\ell,t\right)^{\mu},$$

and, similarly, in the service sector we have

$$S\left(L_S\left(\ell,t\right)\right)=Z_S\left(\ell,t\right)^{\gamma}L_S\left(\ell,t\right)^{\sigma},$$

where $Z_i\left(\ell,t\right)$ is the technology level[5] and $L_i\left(\ell,t\right)$ is the amount of labor per unit of land used at location ℓ and time t in sector i. We assume that a firm takes $Z_i\left(\ell,t\right)$ as given, so it does not take into account the effect of other producers on productivity. The problem of a firm in sector $i\in\{M,S\}$ at location ℓ is thus given by

$$\max_{L_i(\ell,t)}\left(1-\tau_i\left(\ell,t\right)\right)\left(p_i\left(\ell,t\right)Z_i\left(\ell,t\right)^{\gamma}L_i\left(\ell,t\right)^{\iota}-w\left(\ell,t\right)L_i\left(\ell,t\right)\right), \tag{4}$$

where $\iota\in\{\mu,\sigma\}$ and where $\tau_i\left(\ell,t\right)$ denotes taxes on profits charged by the local government to firms in industry i. The maximum per unit land rent that firms in sector i are willing to pay, the bid rent, is then given by

$$R_i\left(\ell,t\right)=\left(p_i\left(\ell,t\right)Z_i\left(\ell,t\right)^{\gamma}\hat{L}_i\left(\ell,t\right)^{\iota}-w\left(\ell,t\right)\hat{L}_i\left(\ell,t\right)\right)\left(1-\tau\left(\ell,t\right)\right). \tag{5}$$

We assume that $\tau_i\left(\ell,t\right)$ is the same across industries, and thus from now on, we drop the subscript on τ.

2.3 Diffusion and Timing

Technology diffuses between time periods. This diffusion is assumed to be local and to decline exponentially with distance. In particular, if $Z\left(r,t\right)$ was the technology used in location r in period t, in the next period $t+1$, location ℓ has access to (but does not necessarily need to use) technology

$$e^{-\delta|\ell-r|}Z_i\left(r,t\right).$$

[5] In the rest of the paper, we will use the terms 'technology level' and 'TFP' interchangeably when referring to $Z_i\left(\ell,t\right)$. Although strictly speaking TFP corresponds to $Z_i\left(\ell,t\right)^{\gamma}$, this difference will be irrelevant in the numerical section, where γ will be set to 1.

Hence, before the innovation decision in period $t+1$, location ℓ has access to

$$Z_i(\ell, t+1) = \max_{r \in [0,1]} e^{-\delta|\ell - r|} Z_i(r,t) \tag{6}$$

which of course includes its own technology of the previous period. This type of diffusion is the only *exogenous* source of agglomeration in the model.[6]

The timing of the problem is key. Figure 1 illustrates the assumed timing.

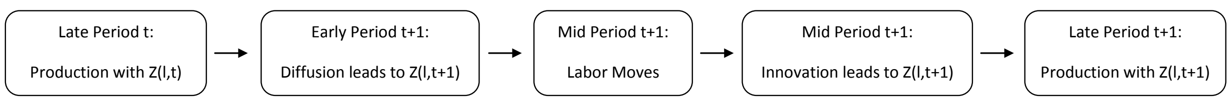


Figure 1: Timing

During the night, between periods t and $t+1$, technology diffuses locally as described above. This leads to a level of technology $Z_i(\ell, t+1)$, given by (6), in the morning. Labor then moves according to this technology and the wage determined by it. After labor moves, localities may decide to try to improve their technology by investing in innovation. This is done by taxing local firms as we describe in the next subsection. The level of technology a location uses in production in period $t+1$ is then either the one it woke up with or the improved technology provided it invested in innovation and was successful at doing so. Note that we are assuming that the number of people in location ℓ, denoted by $L(\ell, t)$, reacts to $Z_i(\ell, t+1)$ and adjusts before innovation is realized. That is, labor moves at the beginning of the period so that innovation has no contemporaneous effect on labor mobility. Given that agents hold a diversified portfolio of land in all locations and given that wages are determined before any possible innovation, agents do not need to build expectations about the future when deciding where to locate. The fact that labor cannot move to a location immediately as a result of a successful innovation is nonetheless important, since it implies that there are rents that can cover the costs of innovation.

[6] As we describe below, there is an endogenous source of agglomeration that results from trade. Locations that experience high relative prices of a given good are more likely to form clusters specialized in the production of that good.

2.4 Idea Generation

The government of a 'county' can decide to innovate by taxing firms to buy an opportunity to innovate. In particular, the county buys a probability $\phi \leq 1$ of innovating at cost $\psi(\phi)$ per unit of land in a particular industry i. Thus, given that all counties are of equal measure I, the total cost of innovation is $I\psi(\phi)$. This implies that with probability ϕ the county obtains an innovation and with probability $(1-\phi)$ its technology is not affected by the investment in innovation.[7]

If a county innovates, all firms (in all locations) in the county have access to the new technology. A county that obtains the chance to innovate draws a technology multiplier z_i from a Pareto distribution (with lower bound 1), leading to an improved technology level, $z_i Z_i(\ell, t)$, where

$$\Pr[z < z_i] = \left(\frac{1}{z}\right)^a$$

Thus, conditional on innovation, the average technology becomes

$$E\left(Z_i(\ell, t+1) \,|\, Z_i(\ell, t), Innovation\right) = \frac{a}{a-1} Z_i(\ell, t) \text{ for } a > 1. \tag{7}$$

Note that the average technology for a given ϕ, not conditional on innovating, is

$$\begin{aligned} E\left(Z_i(\ell, t+1) \,|\, Z_i(\ell, t)\right) &= \left(\frac{\phi a}{a-1} + (1-\phi)\right) Z_i(\ell, t) \\ &= \left(\frac{\phi + a - 1}{a-1}\right) Z_i(\ell, t). \end{aligned}$$

2.5 Innovation and Government Budget

The government of a county taxes profits of its firms to invest in innovation. We assume a balanced budget: if total investment in innovation in industry i in a county of measure I that includes location ℓ is $I\psi(\phi_i(\ell, t))$, the government taxes its firms exactly this amount. A county of size I that pays $I\psi(\phi_i(\ell, t))$ obtains in expectations a technology $\frac{\phi_i + a - 1}{a-1}$ times

[7] Instead we could assume that a county buys a realization of a Poisson distribution for a number of opportunities to innovate. In this case, we need to calculate the expectation of the maximum draw out of N realizations, which is distributed Fréchet, as discussed in Eaton and Kortum (2002).

greater than its current technology. Local governments maximize total output gains minus the investment cost of innovation. Hence, the local government maximizes

$$\max_{\phi_i(\ell,t)} \int_{C_I} \left(\left(\frac{\phi_i(\ell,t) + a - 1}{a-1} \right)^{\gamma} - 1 \right) Z_i(\ell,t)^{\gamma} p_i(\ell,t) L_i(\ell,t)^{\iota} d\ell - I\psi(\phi_i(\ell,t)) \quad (8)$$

where i denotes the industry a location specializes in, and C_I is the set of locations in the county.[8] The benefits of the extra production last only for one period. Since a county is by assumption small and innovation diffuses geographically in the next period, it has no power to affect its expected future level of technology. This explains why governments need not be forward-looking when deciding how much to invest in innovation. Furthermore, after a period new people move to the county and equalize utility across locations (people cannot be excluded after one period and everyone owns a diversified portfolio of land).

Note from (8) that the benefits of increasing ϕ are concave for $\gamma < 1$. Suppose the cost of a draw satisfies $\psi'(\phi) > 0$, and $\psi''(\phi) \geq 0$. A ready example would be

$$\psi(\phi) = \psi_1 + \psi_2 \phi \text{ for } \psi_2 > 0.$$

If $\psi_1 > 0$, there is a fixed cost to invest in innovation, so that we need

$$\int_{C_I} \left(\left(\frac{\phi_i(\ell,t) + a - 1}{a-1} \right)^{\gamma} - 1 \right) Z_i(\ell,t)^{\gamma} p_i(\ell,t) L_i(\ell,t)^{\iota} d\ell > I\psi(\phi_i(\ell,t))$$

for some $\phi_i(\ell,t) \in [0,1]$. We also need to satisfy the first-order condition (note that if $\psi(\cdot)$ is linear, the second-order condition is satisfied, since $\gamma < 1$) given by

$$I\psi_2 = \frac{(\phi_i^*(\ell,t) + a - 1)^{\gamma-1}}{(a-1)^{\gamma}} \int_{C_I} Z_i(\ell,t)^{\gamma} p_i(\ell,t) L_i(\ell,t)^{\iota} d\ell.$$

This implies

$$\phi_i^*(\ell,t) = \left(\frac{\frac{\gamma}{I} \int_{C_I} Z_i(\ell,t)^{\gamma} p_i(\ell,t) L_i(\ell,t)^{\iota} d\ell}{\psi_2 (a-1)^{\gamma}} \right)^{\frac{1}{1-\gamma}} - a + 1, \quad (9)$$

[8] As we will see, because of trade, each location will specialize in one industry. But given that counties consist of a set of locations, each county may very well produce in both sectors.

In the linear cost case we therefore have

$$\phi_i(\ell,t)=\begin{cases} 0 & \text{if } \psi\left(\phi_i^*(\ell,t)\right)\geq \frac{\left(\phi_i^*(\ell,t)+a-1\right)^\gamma-(a-1)^\gamma}{(a-1)^\gamma I}\int_{C_I} Z_i(\ell,t)^\gamma p_i(\ell,t)\, L_i(\ell,t)^\iota\, d\ell \\ & \text{and/or } \phi_i^*(\ell,t)\leq 0 \\ \\ \phi_i^*(\ell,t) & \text{if } \psi\left(\phi_i^*(\ell,t)\right)< \frac{\left(\phi_i^*(\ell,t)+a-1\right)^\gamma-(a-1)^\gamma}{(a-1)^\gamma I}\int_{C_I} Z_i(\ell,t)^\gamma p_i(\ell,t)\, L_i(\ell,t)^\iota\, d\ell \\ & \text{and } \phi_i^*(\ell,t)> 0 \\ \\ 1 & \text{if } \psi\left(\phi_i^*(\ell,t)\right)< \frac{\left(\phi_i^*(\ell,t)+a-1\right)^\gamma-(a-1)^\gamma}{(a-1)^\gamma I}\int_{C_I} Z_i(\ell,t)^\gamma p_i(\ell,t)\, L_i(\ell,t)^\iota\, d\ell \\ & \text{and } \phi_i^*(\ell,t)\geq 1 \end{cases}.$$

Note that a few results are immediate from these equations. Since $0<\gamma<1$, investment in innovation is weakly increasing in total industry output (or output per unit of land). This scale effect is consistent with the evidence presented by Carlino et al. (2007). They show that in the U.S. a doubling of employment density leads to a 20% increase in patents per capita.

Alternatively, we can also let

$$\psi(\phi)=\psi_1+\psi_2\frac{1}{1-\phi} \text{ for } \psi_2>0. \tag{10}$$

The advantage of this cost function is that it has an asymptote at 1. This prevents us from dealing with corner solutions at 1. In order to solve the FOC in closed form, let $\gamma=1$. The FOC is then given by

$$I\frac{\psi_2}{(1-\phi)^2}=\frac{1}{a-1}\int_{C_I} Z_i(\ell,t)\, p_i(\ell,t)\, L_i(\ell,t)^\iota\, d\ell,$$

which implies that

$$\phi_i^*(\ell,t)=1-\left(\frac{\psi_2(a-1)}{\frac{1}{I}\int_{C_I} Z_i(\ell,t)\, p_i(\ell,t)\, L_i(\ell,t)^\iota\, d\ell}\right)^{\frac{1}{2}}. \tag{11}$$

Then

$$\phi_i(\ell,t)=\begin{cases} 0 & \text{if } \psi\left(\phi_i^*(\ell,t)\right) \geq \frac{\phi_i^*(\ell,t)}{(a-1)I}\int_{C_I} Z_i(\ell,t)\, p_i(\ell,t)\, L_i(\ell,t)^{\iota}\, d\ell \\ & \text{and/or } \phi_i^*(\ell,t)\leq 0 \\ & \\ \phi_i^*(\ell,t) & \text{if } \psi\left(\phi_i^*(\ell,t)\right) < \frac{\phi_i^*(\ell,t)}{(a-1)I}\int_{C_I} Z_i(\ell,t)\, p_i(\ell,t)\, L_i(\ell,t)^{\iota}\, d\ell \\ & \text{and } \phi_i^*(\ell,t) > 0 \end{cases}. \tag{12}$$

As with the previous functional form, innovation exhibits once again a scale effect in total industry output, so $\phi_i(\ell,t)$ is increasing in $Z_i(\ell,t)\, p_i(\ell,t)\, L_i(\ell,t)^{\iota}$.

In order to finance $I\psi\left(\phi_i(\ell,t)\right)$ the government in location ℓ levies a tax $\tau(\ell,t)$ on firms in industry i such that

$$\psi\left(\phi_i(\ell,t)\right)=\frac{\tau(\ell,t)}{I}\int_{C_I}\left(Z_i(\ell,t)^{\gamma}\, p_i(\ell,t)\, L_i(\ell,t)^{\iota}-w(\ell,t)L_i(\ell,t)\right)d\ell \tag{13}$$

each period, where $\phi_i(\ell,t)$ is given by the expression above. This implies that each industry finances the investment in innovation in its own industry. Qualitatively the results would not change were we to allow for cross-subsidization. Note the timing. Taxes are set after the innovation is realized to cover its costs. So in (13), $\frac{1}{I}\int_{C_I} Z_i(\ell,t)^{\gamma}\, p_i(\ell,t)\, L_i(\ell,t)^{\iota}\, d\ell$ is actual average production per unit of land.

In all numerical exercises we make $\psi(\cdot)$ proportional to wages in each location. Hence, if an economy grows (and therefore wages increase), the cost of investment in innovation grows accordingly. Then, the model is such that –with enough locations so that the law of large numbers applies– the economy converges to a balanced growth path. Of course, for a finite number of locations, there will be fluctuations around this balanced growth path even in the long run. Of course, even if the law of large numbers holds, individual locations' employment, specialization, trade, etc. will keep changing over time.

2.6 Land, Goods, and Labor Markets

Trade allows locations to specialize in one industry.[9] Goods are costly to transport. For simplicity we assume iceberg transportation costs that are identical in manufacturing and services. This is without loss of generality given that the equilibrium depends only on the sum of transport costs in both industries. If one unit of any of the goods is transported from ℓ to r, only $e^{-\kappa|\ell-r|}$ units of the good arrive in r. Since the technology to transport goods is freely available, the price of good i produced in location ℓ and consumed in location r has to satisfy

$$p_i\left(r,t\right) = e^{\kappa|\ell-r|}p_i\left(\ell,t\right). \tag{14}$$

Land is assigned to the industry that values it the most. Hence, land rents are such that

$$R\left(\ell,t\right) = \max\left\{R_M\left(\ell,t\right), R_S\left(\ell,t\right)\right\}.$$

Denote by $\theta_i(\ell)$ the fraction of land at location ℓ used in the production of good i. If $R\left(\ell,t\right) = R_i\left(\ell,t\right)$, then $\theta_i\left(\ell,t\right) > 0$. Of course, with complete specialization this condition becomes $\theta_i\left(\ell,t\right) = 1$.

In order to guarantee equilibrium in product markets, we need to take into account that some of the goods are lost in transportation. To do this, let $H_i\left(\ell,t\right)$ denote the stock of excess supply of product i between locations 0 and ℓ. Define $H_i\left(\ell,t\right)$ by $H_i\left(0,t\right) = 0$ and by the differential equation

$$\frac{\partial H_i\left(\ell,t\right)}{\partial \ell} = \theta_i\left(\ell,t\right)x_i\left(\ell,t\right) - c_i\left(\ell,t\right)\left(\sum_i \theta_i\left(\ell,t\right)\hat{L}_i\left(\ell,t\right)\right) - \kappa\left|H_i\left(\ell,t\right)\right|, \tag{15}$$

where $x_M\left(\ell,t\right) = M\left(\hat{L}_M\left(\ell,t\right)\right)$ and $x_S\left(\ell,t\right) = S\left(\hat{L}_S\left(\ell,t\right)\right)$ denote the equilibrium production of good i at location r per unit of land. That is, at each location we add to the stock of excess supply the amount of good i produced and we subtract the consumption of good i by all residents of r. We then need to adjust for the fact that if $H_i\left(\ell,t\right)$ is positive

[9] Once again, counties are formed by many locations, so they do not need to specialize. Since counties are the ones that invest in innovation, we allow for the possibility of having one county invest in innovation in both industries.

and we increase r, we have to ship the stock of excess supply a longer distance. This implies a cost in terms of goods and services given by κ. The equilibrium conditions in the goods markets are then $H_i(1,t) = 0$ for all i.

We impose trade balance location by location. The value of the goods shipped to location ℓ must thus be identical to the value of the goods shipped from location ℓ, so that

$$p_M(\ell,t) H_M(\ell,t) + p_S(\ell,t) H_S(\ell,t) = 0 \text{ for all } \ell \text{ and } t. \tag{16}$$

The trade balance condition says that the value of goods produced and consumed at ℓ is equal, once transport costs in terms of goods are covered.

In equilibrium, labor markets clear. Given free mobility, we have to guarantee that the total amount of labor demanded in the economy is equal to the total supply $\overline{L}$. The labor market equilibrium condition is therefore

$$\int_0^1 \sum_i \theta_i(\ell,t) \hat{L}_i(\ell,t)\, d\ell = \overline{L}. \tag{17}$$

2.7 Definition of Equilibrium

An equilibrium in this economy is a set of real functions $(c_i, \hat{L}_i, \theta_i, H_i, p_i, R_i, w, Z_i, \phi_i, \tau)$ of locations $\ell \in [0,1]$ and time $t = 1, ...,$ for $i \in \{M, S\}$, such that:

- Agents choose consumption, c_i, by solving the problem in (1).
- Agents locate optimally, so w, p_i, R_i and L_i satisfy (2).
- Firms maximize profits by choosing the number of workers per unit of land, $\hat{L}_i$, that solves (4), and by choosing the land bid rent, R_i, that solves (5).
- Land is assigned to the industry that values it the most, so if $\max\{R_M(\ell,t), R_S(\ell,t)\} = R_i(\ell,t)$, then $\theta_i(\ell,t) = 1$.
- Goods markets clear, so H_i is given by (15) and $H_i(1) = 0$.
- Trade is balanced location by location, so (16) is satisfied.
- The labor market clears, so θ_i and $\hat{L}_i$ satisfy (17).

- The government budget is balanced each period, so ϕ_i and τ satisfy (13), and investment in innovation ϕ_i satisfies (12).
- Technology Z_i satisfies the innovation process that leads to (7) and the diffusion process given by (6).

3. EVIDENCE AND MODEL PREDICTIONS

We now proceed to solving the model numerically and to contrast the equilibrium allocation with the data. To do so, we need to propose values for all of the parameters in the model. These are based on the evolution of the U.S. economy over the period 1980-2005.

3.1 Evidence

Although many of the stylized facts will appear familiar from the literature on the structural transformation (see, e.g., Ngai and Pissarides, 2007, and Buera and Kaboski, 2007), we will also emphasize two less well-known aspects. First, in the last fifteen years, compared to the 1980s, many of those familiar stylized facts have undergone significant changes. Second, we will present evidence on the spatial dimension, an aspect generally ignored in this literature.

It is well known that employment has been moving out of goods and into services,[10] as can be seen in Figure 2.1 (where the extension .1 in the figure's name refers to the upper panel and the extension .2 to the lower panel).[11] The start of this shift dates back to the 1930s and has continued to the present day. However, since the mid 1990s this shift has clearly been slowing down. In fact, between 1980 and 1995 the share of service employment increased by about 10 percentage points but only by 4 percentage points between 1995 and 2005. This change in employment shares has been accompanied by a decrease in the price

[10] In the empirical section we distinguish between 'goods' and 'services' (where 'goods' is the aggregation of manufacturing, construction and mining) because this is the typical distinction in many of the data sources, such as the *Industry Economic Accounts* of the BEA. In the rest of the paper, we refer to the two sectors of interest as 'manufacturing' and 'services.'

[11] Figure 2 to 11 are included at the end of the paper.

of goods, relative to services. As shown in Figure 2.2., this decline was pronounced in the 1980s and early 1990s, but since then has been slowing down, with even a slight reversal in recent years.

The mid 1990s also marks a breakpoint for wages. Figure 3.1 shows how real hourly wages of production workers started to increase significantly around 1995, after two decades of decline.[12] This timing also coincides with the evolution of land and housing prices. Figure 3.2 shows sharp increases in the real values of land and housing post-1995, following a fairly stable pattern throughout the 1980s and the early 1990s. Of course, part of this dramatic increase is disappearing as a result of the current housing crisis, but it remains to be seen whether values will return to their pre-1995 levels in real terms.[13]

The dynamics in our theory are the result of innovations that translate into higher local productivity. We use value added per worker as the empirical counterpart to productivity. Figure 4.1 shows how in the 1980s services productivity growth, as measured by value added per worker, was falling behind that of goods, a phenomenon that goes back in time and was described by Baumol (1967), who argued that it was inherently more difficult to innovate in services than in goods. That same widening gap is also apparent in Figure 4.2, which reports the log of value added per worker in both goods and services. Since the mid 90s services productivity growth has clearly been catching up and, on some accounts, may even have surpassed productivity growth in the goods-producing sector (Triplett and Bosworth, 2004).

As for the spatial dimension, the goods sector has become more dispersed in terms of employment density, whereas the service sector has become more concentrated over time. Using U.S. county data, Table 1 shows the evolution of the standard deviation of log employment in both sectors between 1980 and 2005 (as well as the log difference between the 70th and 30th percentiles). For goods, the tightening distribution implies that counties are becoming more alike in terms of employment density. In contrast, for services the distri-

[12] For purposes of comparison with the numerical section, to obtain real wages we deflate by the services price index used in the *Industry Economic Accounts* of the BEA.

[13] Once again, we deflate by the services price index used in the *Industry Economic Accounts* of the BEA.

bution is widening, implying that service employment is becoming more concentrated in space. Note also that services started off being more dispersed than goods, and therefore the two distributions are becoming more similar. The increased spatial concentration in services also shows up when analyzing labor productivity, as measured by earnings per worker. Table 1 shows how the distribution of earnings per worker in the industrial sector did not change much over time. In contrast, in the service sector earnings per worker have become more unequal across counties, as reflected by the widening distribution. As with employment, sectoral differences have become mitigated over time.

	1980	1995	2005		1980	1995	2005
Log Employment Density				**Log Employment Density (net of local consumption)**			
Difference 70-30				*Difference 70-30*			
Goods	1.677	1.634	1.632	Goods	1.723	1.650	1.666
Services	1.340	1.462	1.510	Services	1.396	1.993	2.186
Standard deviation				*Standard deviation*			
Goods	1.764	1.740	1.695	Goods	1.731	1.676	1.636
Services	1.548	1.613	1.635	Services	1.548	1.613	1.635
Log Earnings per Worker				**Log Earnings per Worker (net of local consumption)**			
Difference 70-30				*Difference 70-30*			
Goods	0.296	0.294	0.325	Goods	0.328	0.261	0.274
Services	0.153	0.191	0.224	Services	0.164	0.261	0.314
Standard deviation				*Standard deviation*			
Goods	0.273	0.298	0.335	Goods	0.283	0.269	0.282
Services	0.159	0.207	0.252	Services	0.173	0.254	0.312
Source: REIS, Bureau of Economic Analysis							

Table 1: Distribution of Employment Density and Productivity

Given that in our theory locations fully specialize, whereas in the data they do not, we redo our exercise, after making the data more comparable to the theory. For each county we adjust the earnings in each sector to what they would be were the county fully specialized. Take a county that is a net exporter of goods. We compute the consumption of goods implied by the amount of services the county produces and subtract this amount from the total earnings of goods. To obtain the consumption of goods implied by the production of services, we use the aggregate ratio of spending on goods relative to spending on services from the U.S. economy and multiply this by services earnings. This gives us a measure of adjusted goods earnings were the county fully specialized in the production of goods. We do a parallel calculation for net service exporters. Obtaining a similarly adjusted measure for sectoral employment at the county level is straightforward: we just take the adjusted sectoral earnings and compute the implied sectoral employment. As can be seen in the right-hand side of Table 1, the results are essentially unchanged. Services start off spatially more dispersed than goods and, over time, become increasingly concentrated. Depending on the exact measure, services either converge or overtake goods in terms of the degree of spatial concentration.

3.2 An Equilibrium Outcome

The basic message we obtain from the evidence presented above is that between 1980 and 1995 productivity in goods, relative to services, was growing fast, relative prices of goods were declining, and employment in the goods-producing sectors was steadily falling. During that same period, service productivity growth was low, and real land rents and wages did not exhibit significant changes. Then, around 1995, land prices and wages started to increase in real terms and so did service productivity growth. Changes in employment shares and in the relative price of goods also slowed down or stopped altogether. This was accompanied by services becoming geographically more concentrated, making it more similar to goods in terms of its spatial distribution.

The model is rich enough to match all of these features of the evolution of the U.S. economy over the last 25 years, at least qualitatively and sometimes quantitatively. We now choose the parameters of the model and present a numerical exercise that can be compared to the data in the previous subsection.

To compute the model we need to specify initial productivity functions for both manufacturing and services. We let $Z_S(\cdot, 0) = 1$ and $Z_M(\cdot, 0) = 0.8 + 0.4\ell$. The key characteristic of the initial productivity functions is that service productivity is initially larger than that of manufacturing for locations close to the left border, whereas manufacturing productivity is larger than that of services close to the right border. Furthermore, the locations with the highest manufacturing productivity (namely, the right border) have a 20% larger productivity than the locations with the highest service productivity. These initial productivity functions imply that if all other parameters are identical between sectors, innovation always happens earlier in manufacturing than in services and always in the locations close to the right border.

The elasticity of substitution between manufacturing and services, $1/(1-\alpha)$, is important for the results. A key mechanism in the model is that as productivity in one sector increases, relative to the other sector, the relative price of output in that sector decreases and so does its employment share. For this to happen, the elasticity of substitution between goods and services must be less than 1. This is consistent with empirical estimates. Stockman and Tesar (1995), for example, estimate it to be 0.44 for a set of 30 countries. Given this evidence, we set $\alpha = -1.5$, so the elasticity of substitution is $1/(1-\alpha) = 0.4$.

The elasticity of substitution is also important for the incentives to innovate in different sectors. With an elasticity below 1, when a sector's relative productivity increases and employment in that sector declines, the increase in employment in the other sector increases the incentives for innovation in that slow-growing sector. Eventually, enough people switch to the slow-growing sector for innovation to start there. In that sense, the economy self-regulates. Indeed, as more people move out of the fast-growing sector, thus tending to lower overall growth, the other sector starts innovating as well, thus tending to increase overall growth. As we show in the examples, the aggregate trend converges to a balanced growth path (apart from small random fluctuations). Given the importance of the elasticity of substitution, in later exercises we study the effect of changes in its value.

Using data from the BEA, Herrendorf and Valentinyi (2007) estimate labor shares in both sectors to be slightly above 0.6, so we set the share of labor in both sectors to $\mu = \sigma = 0.6$. Figure 2.1 shows that by 1980 the share of total employment in services was already substantially above that in manufacturing. To capture this, we set $1.4 = h_S > h_M = 0.6$, which generates an initial employment share in services of around 60%, roughly as in the data.

The timing is important. We let the model run for 50 periods and compare its predictions to the 25 years of data we have presented, so that a model period amounts to half a year. Throughout we let $\beta = 0.95$, although this parameter plays a limited roll in our results given that all decisions are essentially static. To make the simulations computationally feasible, we divide the unit interval into 500 locations, each of which we interpret as a county that makes autonomous decisions on innovation.

We simulate using the cost function in equation (10). We set $\psi_2 = -\psi_1 > 0$, so there are no fixed costs of investment in innovation. The intensity of innovation is then governed by two parameters: the cost parameter, ψ_2, and the shape parameter of the Pareto distribution, a, from which we obtain the productivity draws. Both parameters have a similar effect. Increasing ψ_2 leads to a higher cost, and increasing a gives a Pareto distribution with a thinner tail, so that both effects yield less innovation. We let $\psi_2 = 0.002054$ and $a = 43.4$, which results in aggregate productivity growth of around 3% per period in manufacturing and around 2% per period in services. This parameter configuration also implies an acceleration of services productivity growth around period 30, which in the data is interpreted as 1995 (15 years after 1980 and 2 periods per year). Once service innovation reaches full speed, its productivity growth rate is about the same as that of manufacturing, namely, 3% per period.

Aggregate productivity growth rates and therefore changes in employment shares are also determined by technological diffusion. We set the exponential decline of the diffusion of technology, δ, equal to 25. This results in employment shares in services that rise from 0.6 to 0.73, an increase of 13% for the 25-year period, as in the data.

We set the transport cost parameter $\kappa = 0.008$. This level of transport costs in general yields two main specialization areas: one for services and one for manufacturing. In particular, a cluster of service employment forms endogenously close to the manufacturing cluster. Other areas to the left of that cluster also specialize in services but produce much less. The formation of this cluster and its location, as well as the timing of the innovation in services, can vary significantly with the transport cost parameters. We study the effect of transport costs in more detail below. Of course, through the timing of the innovation in services, this parameter partially governs the magnitude of the decrease in the relative goods prices. In the model the manufacturing price falls by about 60%, more than the 40% observed in the data. Nevertheless we choose this parameterization because the timing is closer to the data, even if it yields a price decline that is somewhat too large.

The result of a numerical realization of the model is presented in Figure 5.[14] For all numerical simulations we present similar graphs, consisting of nine subplots. We denote subplots using 3 digits (e.g., 5.2.1) where the first digit denotes the number of the figure and the other two the corresponding row and column. In all figures services are plotted in red and manufacturing in blue.

Figure 5.1.1 presents the coefficient of variation of log employment (the dashed curve) and log value added (the solid curve) across counties in both industries.[15] The distribution of employment and value added vary in a parallel fashion. Of course, the coefficient of variation of value added is higher, since it includes employment, productivity and price dispersion. Initially manufacturing is innovating more, as reflected by the higher values for the coefficient of variation. Recall that a greater coefficient of variation points to a more disperse distribution, which means economic activity is spatially more concentrated.[16] Over time, as in the data, the service sector catches up with the manufacturing sector and surpasses it. Both sectors become more concentrated in space as in some of the measures in Table 1. The main feature of the data that the model is able to replicate is that the distribution of employment across counties is becoming more similar between manufacturing and services, with services becoming geographically more concentrated.

Figure 5.1.2 presents aggregate productivity calculated in two different ways. The solid

[14] We present examples of particular realizations of the innovation process. However, given the relatively large number of locations, results vary little for different realizations if we preserve the same parameter values.

[15] Throughout we exclude from this calculation all locations that have never innovated in the service sector as they make the coefficient of variation grow faster and the effects are harder to see. None of our conclusions are altered if we include all locations.

[16] Indeed, a tightly concentrated distribution implies that all counties are the same, so that economic activity is equally dispersed across space. In contrast, a widely disperse distribution means that counties are very different, with economic activity concentrating in some areas and by-passing others.

curves present aggregate productivity as

$$Agg_1 Z_i\left(t\right) = \frac{\int_0^1 x_i\left(\ell, t\right)\theta_i\left(\ell, t\right)d\ell}{\left(\int_0^1 \hat{L}_i\left(\ell, t\right)\theta_i\left(\ell, t\right)d\ell\right)^{\iota}}, \tag{18}$$

the dashed curves present an alternative statistic, namely,

$$Agg_2 Z_i\left(t\right) = \frac{\int_0^1 x_i\left(\ell, t\right)\theta_i\left(\ell, t\right)d\ell}{\int_0^1 \left(\hat{L}_i\left(\ell, t\right)\theta_i\left(\ell, t\right)\right)^{\iota} d\ell}, \tag{19}$$

where x_i denotes output in sector i. Given that there are decreasing returns to scale in labor at each location, it is not clear which one of them is preferred. $Agg_1 Z_i\left(t\right)$ is the equivalent of a Solow residual, but a shift in $Agg_2 Z_i\left(t\right)$ increases aggregate output by exactly the same amount. Note how, as time passes, we first observe the catching up of services in terms of aggregate productivity, but both manufacturing and services grow eventually at a roughly constant rate that is common to both sectors (up to the local random realizations that average out in space, but not fully since we have 500 locations). It is the process of shifting employment to the sector that innovates less that equates productivity growth in both sectors asymptotically, thus putting the economy on a balanced growth path.

Figure 5.1.3 presents the stock of excess supply, $H_M\left(\ell, t\right)$. Each curve represents excess supply in a different time period, $H_M\left(., t\right)$. In this simulation, lower curves coincide with later periods. A curve declines when locations specialize in services and it grows when locations specialize in manufacturing. It is a good way of tracking changes in specialization over space. A parameter that is key in determining the number of areas of specialization is the diffusion parameter δ. An increase in δ implies that diffusion dies out fast and so locations benefit little from it. To see this, Figure 6 presents a simulation with $\delta = 50$. Compared to Figure 5.1.3, we can see in Figure 6.1.3 that the slope of the stock of excess supply changes sign many times, indicating several switches in land-use specialization. The reason is clear: when diffusion is very local, being close (but not extremely close) to other regions producing the same good does not provide any advantage.

Figure 5.2.1 presents the value of land over time. It shows the value of the diversified portfolio of land held by all agents, as well as the value of land specialized in each sector.

Note from the figure how the value of manufacturing land decreases as technology in manufacturing improves faster than service technology. This happens because the decline in the value of manufactured goods more than compensates for the increase in productivity. The value of service land, on the other hand, increases throughout. Once innovation in the service sector accelerates, both the value of the portfolio of land and manufacturing land rents start increasing much faster, since both sectors are now competing for the same land close to each other. This is very clear in the U.S. data presented in Figure 3.2, and the timing coincides with the increase in service productivity shown in Figure 4.1. Note that both in the model and in the data we deflate by the price of service goods.

Figure 5.2.2 exhibits the price of manufactured goods relative to services. The initial increase in manufacturing productivity, together with an elasticity of substitution less than 1, implies that the relative price of manufactured goods declines over time. Once service productivity accelerates, the price stabilizes and declines much slower. The pattern is very close to the one we present in Figure 2.2 for the U.S. economy, although, as discussed, the magnitude of the decline is somewhat larger than the one observed.

We present the evolution of utility and wages in Figure 5.2.3. Note that wages do not increase significantly until service productivity starts growing. This is again consistent with the evidence in Figure 3.1, where wage growth in terms of service goods increases dramatically starting around 1995. Utility grows throughout, since productivity growth in any industry always increases welfare independently of the relative price and labor reallocation effects. There is also an acceleration in utility growth, but it is smaller than the one for wages.

Figure 5.3.1 shows employment shares in both sectors. Since there is no unemployment in this economy, one is the mirror image of the other. We chose parameter values to match the change in shares, so it is not surprising that this figure looks similar to Figure 2.1.

Finally, Figures 5.3.2 and 5.3.3 present the evolution of productivity over time and space. Since this is a three-dimensional object, we present colored contour plots. Dark blue areas represent low productivity, and lighter blue, followed by yellow and red areas, represent higher productivity levels. These figures are helpful in identifying the areas in which innovation is happening and how clusters of innovation are created and destroyed over time. As can be seen in the graphs, manufacturing productivity starts increasing immediately, and all innovation occurs in locations to the right (the top part in the graphs). In contrast, initially innovation in services happens only in very few locations, namely, the ones closest to the manufacturing cluster. Over time, as the employment share in services increases and diffusion takes hold, the set of regions that innovate grows, increasing the size of the service cluster. Regions to the left of the service cluster (the bottom part in the graphs) are not innovating but are specialized in services. Figure 5.1.3 is useful for assessing the relative magnitude of the production of all regions and therefore the pattern of specialization. This is consistent with the evidence presented in Table 1 that shows how the standard deviation of productivity across regions (in terms of earnings per worker) in both sectors increases throughout our time span.

Figures 6 and 7 present a comparative statics exercise when we vary the rate of decline of technology diffusion. A large value of δ implies that technological diffusion dies out faster in space. The benchmark parameterization uses a value of $\delta = 25$, and we present simulations with $\delta = 50$ and $\delta = 10$. With $\delta = 50$ we obtain less aggregate growth in both sectors. This is particularly evident for the service sector, as can be seen in Figure 6.3.3. It is also clear from Figure 6.1.3 that since technology is very local in this case, specialization switches many times in space. In contrast, when we make $\delta = 10$ in Figure 7, diffusion is widespread and there are only two clusters, with substantially more innovation and productivity growth over time. The parameter δ is related to our definition of a period. Letting the economy run for many more periods results in more diffusion even if δ is high because diffusion compounds over time.

3.3 The Effect of Transport Costs and the Elasticity of Substitution

A natural question to ask in the context of our theory is: what is the nature of the additional insights it provides relative to standard growth theories because it incorporates the distribution of economic activity in space? Adding space not only gives us more implications about, say, the dispersion of employment across sectors, but it also allows innovation to happen gradually as in Figure 5 in the service sector. Furthermore, modeling the location of economic activity in space adds economic effects that can overturn the standard reasoning behind the effect of particular parameters. This is the case for transport costs, τ, and the elasticity of substitution $1/(1-\alpha)$.

3.3.1 Transport Costs.—

In our theory transport costs have the standard negative effect on static welfare that is familiar from trade models. Higher transport costs imply that more goods are lost in transportation and agents obtain fewer gains from specialization. But here higher transport costs also imply that it is more important to produce in areas close to locations where the other sector is producing. So if transport costs are relatively high and one sector is already somewhat clustered (like the manufacturing sector in our benchmark case presented in Figure 5), economic activity in the other sector will cluster around it. In the example, the reason is that relative prices of manufactured goods will rise faster as we move away from manufacturing clusters (goods have to be transported and are therefore more expensive). Hence, the service-producing locations close to manufacturing areas have a larger scale, which results in more incentives to innovate. This is evident in Figure 5.3.3. Note also that once innovation starts in one location, it increases productivity in other close-by regions and therefore leads to even more innovation in the cluster. So diffusion, although not necessary to obtain this effect, reinforces it.

The next proposition proves this positive effect of higher transport costs on innovation for an initial condition in which the industry is stagnant.

Proposition 1 *Given any level of transport costs κ, suppose aggregate productivity in industry i is stagnant in some period t. Then, an increase in the level of transport costs, κ, weakly increases aggregate productivity growth in industry i at time t.*

Proof. Let

$$m_\ell = \left\{ \ell : \lim_{\ell' \nearrow \ell} \theta_i(\ell, t) \neq \lim_{\ell' \searrow \ell} \theta_i(\ell, t) \right\}$$

denote the locations in which specialization changes from one industry to the other. Take $\ell \in m_\ell$ and ℓ' such that $\theta_i(\ell', t) = 1$ and $\ell = \arg\min\{|l - \ell'| \text{ for } l \in m_\ell\}$. Let $p(\ell', t) = p_i(\ell', t)/p_j(\ell', t)$ for $j \neq i$. Then either $\partial p(\ell', t)/\partial\kappa > 0$ for $|\ell - \ell'| < B$ for some $B > 0$ (since $p(\ell', t)$ decreases at a rate of 2κ with $|\ell - \ell'|$ by (14)), or $\partial p(\ell', t)/\partial\kappa < 0$ for all ℓ' such that $\theta_i(\ell', t) = 1$. In the latter case innovation in industry i is unaffected by κ. Thus, the rest of the proof assumes we are in the former case. Note that since $\bar{R}/\bar{L}$ is constant across locations, labor mobility (equation (2)) implies that $w(\ell', t)/p_j(\ell', t)$ increases less than $p(\ell', t)$ for $|\ell - \ell'| < B$ (since workers can substitute away from the expensive good). Thus, (4) implies that $L_i(\ell', t)$ increases for $|\ell - \ell'| < B$, since $Z_i(\ell', t)$ is predetermined as a function of $Z_i(\cdot, t-1)$ by (6). The result is that $Z_i(\ell', t)\, p_i(\ell', t)\, L_i(\ell', t)^\iota$ decreases with $|\ell - \ell'|$ and is increasing in κ for $|\ell - \ell'| < B$. Hence, $\max_{\ell'} Z_i(\ell', t)\, p_i(\ell', t)\, L_i(\ell', t)^\iota$ is increasing in κ. Equations (11) and (12), or alternatively, (9), then imply that $\max_\ell \phi_i(\ell, t)$ is increasing in κ. Note that $\min_{\ell'} Z_i(\ell', t)\, p_i(\ell', t)\, L_i(\ell', t)^\iota$ is decreasing in κ and that equation (11) is concave in $Z_i(\ell', t)\, p_i(\ell', t)\, L_i(\ell', t)^\iota$. However, since innovation is bounded below by zero, $\phi_i(\ell, t) \geq 0$, and we start from a situation where no region is innovating, namely $\phi_i(\ell, t) = 0$ all ℓ, proving that the $\max_\ell \phi_i(\ell, t)$ is weakly increasing in κ is sufficient to show that aggregate productivity in industry i is weakly increasing in κ. For a growing industry this is not necessarily the case, since (11) is concave in $Z_i(\ell', t)\, p_i(\ell', t)\, L_i(\ell', t)^\iota$ and so reductions in the price of good i in some locations may lead to declines in $\phi_i(\ell, t)$ that lead to aggregate declines in productivity. ∎

An immediate corollary of this proposition is that, if one of the industries is growing, productivity growth in the stagnant industry jump-starts earlier the higher are transport costs. Recall that innovation takes off when aggregate productivity growth in the other industry shifts enough labor to the stagnant sector. With higher transport costs, the increasing labor share of the stagnant sector will be more densely clustered, leading it to jump-start earlier.

Following the logic above, were we to increase κ from the benchmark value of 0.008, we would increase the density of the service cluster, leading to higher growth, wages and welfare. Qualitatively, the figures look similar to Figure 5 so we do not present them here.

It is easier to see the effect of trade costs when we make transport costs lower. We therefore present two additional simulations with lower transport costs, $\kappa = 0.001$ and $\kappa = 0.005$. Consistent with the argument above, we expect to see less innovation. In Figure 8, where $\kappa = 0.005$, innovation in the service sector is less concentrated in space, since being farther away from the areas specialized in manufacturing is less costly. More important, there is absolutely no innovation in services for the first 29 periods. Lowering κ even more to 0.001, as we do in Figure 9, spreads service employment even more as prices depend less on location. Now innovation happens only in period 40, and when it does, it happens in virtually all locations. As before, the lower transport costs imply lower wages and welfare.

In contrast to standard economic geography models, the static losses from higher transport costs are outweighed by the higher incentives to innovate in certain areas. The result is that growth and overall welfare are higher when transport costs are higher. Recall that the textbook two-region two-sector economic geography model with labor mobility concludes that higher transport costs lead to more dispersion (Krugman, 1991; Puga, 1999). The argument runs as follows: if transport costs are high enough and some factors are immobile, the cost of having to trade between the two regions ceases to compensate for the gains from agglomeration, so that it becomes beneficial for both regions to produce both goods. In as far as concentration of economic activity is related to economic growth, this implies a negative relation between transport costs and economic growth (Baldwin and Martin, 2004).

Whereas in those models higher transport costs lead to more dispersion, in our model they lead to more concentration. As argued by Helpman (1997), the key difference is that in our model, as in Helpman's, both goods face transport costs. This implies that larger transport costs induce services to locate closer to manufacturing. This leads to services becoming less dense in areas far away from manufacturing and more dense in areas closer to manufacturing. The increase in the scale of production then leads to more innovation in service regions that locate close to manufacturing. In contrast to standard economic geography models, the co-localisation of both sectors thus generates the emergence of a service cluster close to the manufacturing cluster. This co-localisation is facilitated in a world with many regions. Of course, in principle another possibility would be for manufacturing to disperse and locate closer to services, thus implying less concentration. This does not happen because the initial cluster of manufacturing gets reinforced over time through innovation and diffusion, a force absent in Helpman (1997). In other words, innovation and diffusion imply that there are more incentives for services to concentrate and form a cluster close to manufacturing than for manufacturing to disperse and locate close to services. The finding that higher transportation costs lead to more innovation, growth and welfare is an example in which having a rich spatial dimension leads to some novel economic effects.

The result that higher transport costs can lead to higher welfare can best be understood through a "second best" argument. In our model, the profits from innovation only last for one period. After that, profits get arbitraged away because workers can relocate and technology diffuses. This implies an externality, since firms do not get the full benefits from innovating. Higher transport costs bring the economy closer to its social optimum by increasing clustering and innovation, but come at the cost of losing resources. The optimal policy would be to introduce patents. However, because of the local scale effect in innovation, optimal patents would have to depend on time and location. Given its high information content, such a "first best" policy is probably infeasible.

3.3.2 Elasticity of Substitution

From standard aggregate logic we would expect a lower elasticity of substitution to lead to faster innovation in services. The reason is simple: as the elasticity of substitution drops, the initially higher productivity growth in manufacturing moves a larger share of the labor force into services, implying higher service density and faster growth.

However, the effect of changes in the elasticity of substitution has an important spatial component. The main logic is that changes in the elasticity of substitution change the willingness of agents to substitute services for manufactured goods and, therefore, their decision to locate in space. If the elasticity of substitution is low, agents are not willing to substitute consumption across sectors and so, given positive transport costs, care more about locating near areas that specialize in a different sector. This prevents the emergence of large service clusters, since those would increase the average distance to close-by manufacturing areas. Instead, many smaller service-producing areas locate across manufacturing areas. This lowers the scale of service-producing regions, implying less innovation in services.

There is another economic force that acts in the opposite direction. As we lower the elasticity of substitution, workers in manufacturing areas consume a higher share of their income in services the higher the price of services. Hence, locations specialized in services and close to areas that are specialized in manufacturing achieve a larger scale and therefore innovate more.

The result of these different effects leads to a non-monotonic relation between the elasticity of substitution and innovation. Starting from our benchmark value of 0.4, if we lower the elasticity of substitution to 0.33, innovation in services declines dramatically. We present these results in Figure 10. Innovation in services starts only in period 38 and is all close to manufacturing. Figure 10.1.3 shows how now we have several switches in specialization as we move across space.

If we lower the elasticity of substitution further to 0.25, we also obtain dispersed location of services close to manufacturing areas, but innovation starts in period 28 and is overall stronger. We present these results in Figure 11.[17] As with transport costs, this logic carries through for a wide range of parameterizations. In sum, the negative effect dominates for high values of the elasticity of substitution and the positive one for low values. This non-monotonicity is the result of the relocation motivated by the change in preferences, in combination with our innovation process. Once again, this result is unique to a spatial-dynamic setup.

4. CONCLUSION

In this paper we have presented a spatial dynamic growth model in which locations choose how much to invest in innovation, if at all, in each sector. To deal with the intractabil-

[17] Elasticities of substitution larger than the standard case lead to a larger, but similar-looking, cluster in manufacturing, so we omit the graphs.

ity of dynamic spatial frameworks, we have proposed a theory where agents solve static problems and markets clear sequentially. This allowed us to keep a rich structure that is able to capture many of the macroeconomic and spatial stylized facts of the U.S. economy. We find that employment relocation is crucial in balancing innovation across sectors. As innovation in one sector increases relative to the other sector, employment shifts from the more innovative to the less innovative sector. This increases incentives for innovation in the lagging sector that is gaining employment, especially in those locations close to the innovative sector's clusters. These effects balance the value of sectors in the economy and lead to a balanced growth path in which aggregate growth in the economy eventually stabilizes.

A fairly stable aggregate path hides important employment reallocation across space. As the economy grows, local clusters emerge and disappear. The pattern of clusters is related to the costs of innovation, the spatial scope of diffusion, transport costs and the elasticity of substitution, as we document numerically. For the latter two, incorporating the space and time dimensions overturns the standard logic of their effects familiar from trade and growth models. We argue that this process of innovation and employment reallocation helps rationalize many observed phenomena in the U.S. during the last few decades. Applying our theory to other time periods or countries could, perhaps, lead to a better understanding of both the theory and the evolution of economic activity in other contexts.

In calibrating the model, we have found that some of its quantitative implications are hard to reconcile with the data. In particular, in our benchmark calibration targeting the change in employment shares over time leads to a reduction in the relative price of goods that is somewhat too large. The model also generates too much innovation and therefore aggregate productivity grows faster than in the data. This is especially the case if we want to target the timing of innovation. Exploring other specifications of preferences (such as non-homotheticicities as in Buera and Kaboski, 2007) or innovation costs may yield a better fit, although we believe, it would obscure some of the spatial-dynamic economic forces we uncover. We therefore leave this exploration for future research.

REFERENCES

[1] Baldwin, R. E. and Martin, P., 2004. "Agglomeration and Regional Growth," in: J. V. Henderson & J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*, 4, 2671-2711.

[2] Baumol, W.J., 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *American Economic Review*, 57, 415-426.

[3] Boucekkine, R., Camacho, C., and Zou, B., 2009. "Bridging the Gap Between Growth Theory and the New Economic Geography: The Spatial Ramsey Model," *Macroeconomic Dynamics*, 13, 20-45.

[4] Brock, W. and Xepapadeas, A., 2008a. "Diffusion-Induced Instability and Pattern Formation in Infinite Horizon Recursive Optimal Control," *Journal of Economic Dynamics and Control*, 32, 2745-2787.

[5] Brock, W. and Xepapadeas, A., 2008b. "Pattern Formation, Spatial Externalities and Regulation in Coupled Economic-Ecological Systems," MPRA Paper 9105, University of Munich.

[6] Buera, F. and Kaboski, J.P., 2007. "The Rise of the Service Economy," working paper, UCLA.

[7] Carlino, G., Chatterjee, S. and Hunt, R., 2007. "Urban Density and the Rate of Invention," *Journal of Urban Economics*, 61, 389-419.

[8] Davis, M. and Heathcote, J., 2007. "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*, 54, 2595-2620.

[9] Desmet, K. and E. Rossi-Hansberg, 2009. "Spatial Growth and Industry Age," *Journal of Economic Theory*, forthcoming.

[10] Eaton, J. and Kortum, S., 1999. "International Technology Diffusion: Theory and Measurement," *International Economic Review*, 40, 537-570.

[11] Eaton, J. and Kortum, S., 2002. "Technology, Geography, and Trade," *Econometrica*, 70, 1741-1779.

[12] Grossman, G. and E. Helpman, 1991. *Innovation and Growth in the Global Economy*, Cambridge MA: The MIT Press.

[13] Helpman, E., 1997. "The Size of Regions," in: D. Pines, E. Sadka, and I. Zilcha, I. (eds.), *Topics in Public Economics: Theoretical and Applied Analysis,* Cambridge University Press: Cambridge.

[14] Herrendorf, B. and Valentinyi, A., 2007. "Measuring Factor Income Shares at the Sector Level - A Primer," CEPR Discussion Papers 6271, C.E.P.R.

[15] Krugman, P., 1991. "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483-499.

[16] Ngai, R.L. and Pissarides, C.A., 2007. "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97, 429-443.

[17] Puga, D., 1999. "The Rise and Fall of Regional Inequalities," *European Economic Review*, 43, 303-334.

[18] Quah, D., 2002. "Spatial Agglomeration Dynamics," *American Economic Review*, 92, 247-252.

[19] Rossi-Hansberg, E., 2005. "A Spatial Theory of Trade," *American Economic Review*, 95, 1464-1491.

[20] Stockman, A. and Tesar, L., 1995. "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements," *American Economic Review*, 85, 168-185.

[21] Triplett, J.E. and Bosworth, B.P., 2004. *Productivity in the U.S. Services Sector*, The Brookings Institution: Washington, D.C.

[22] Ventura, J., 1997. “Growth and Interdependence,” *Quarterly Journal of Economics*, 112, 57-84.

[23] World Bank, 2009, “Reshaping Economic geography,” *World Development Report,* The World Bank, Washington, D.C.

[24] Young, A., 1991. “Learning by Doing and the Dynamic Effects of International Trade,” *Quarterly Journal of Economics,* 106, 369-405.

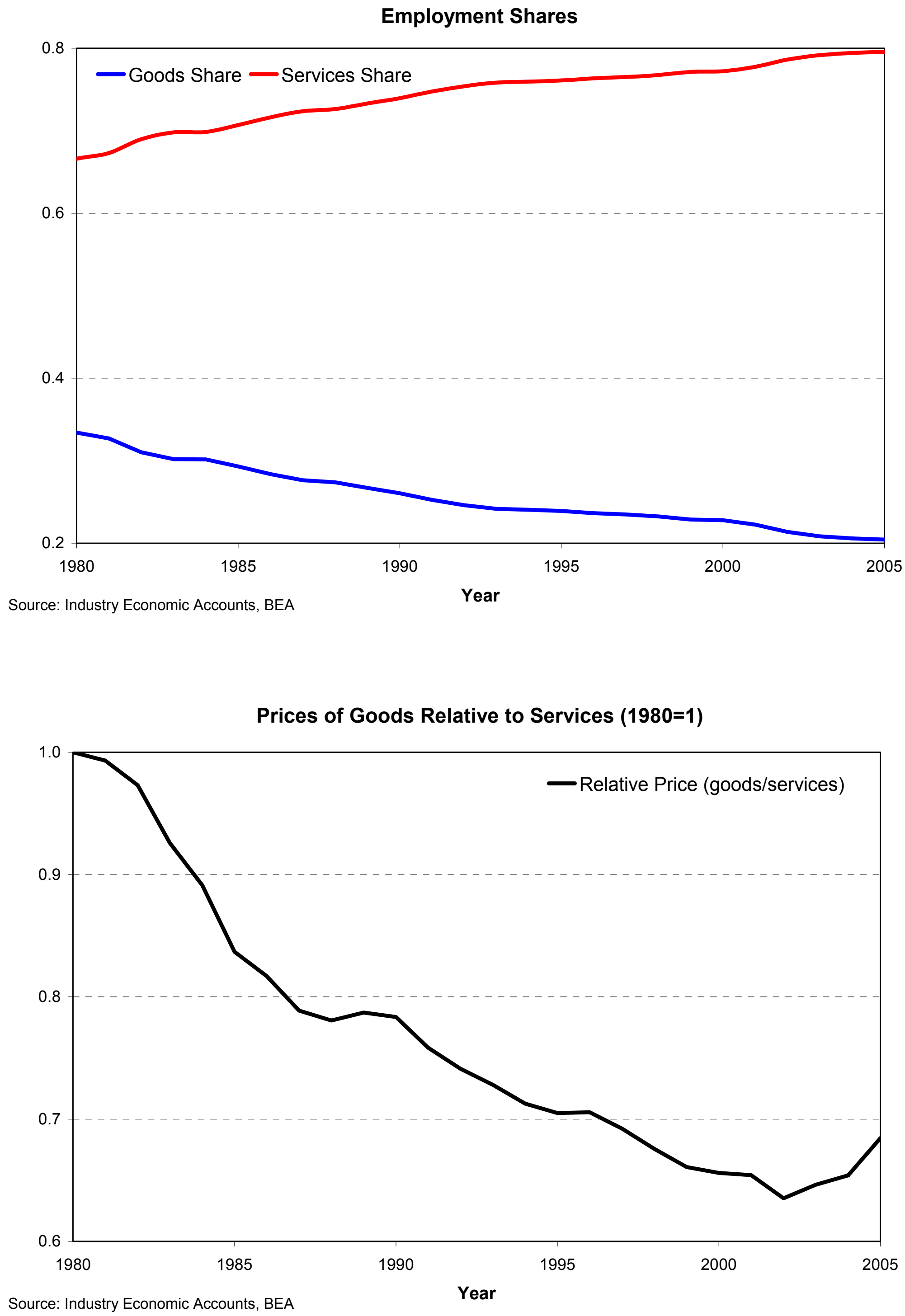
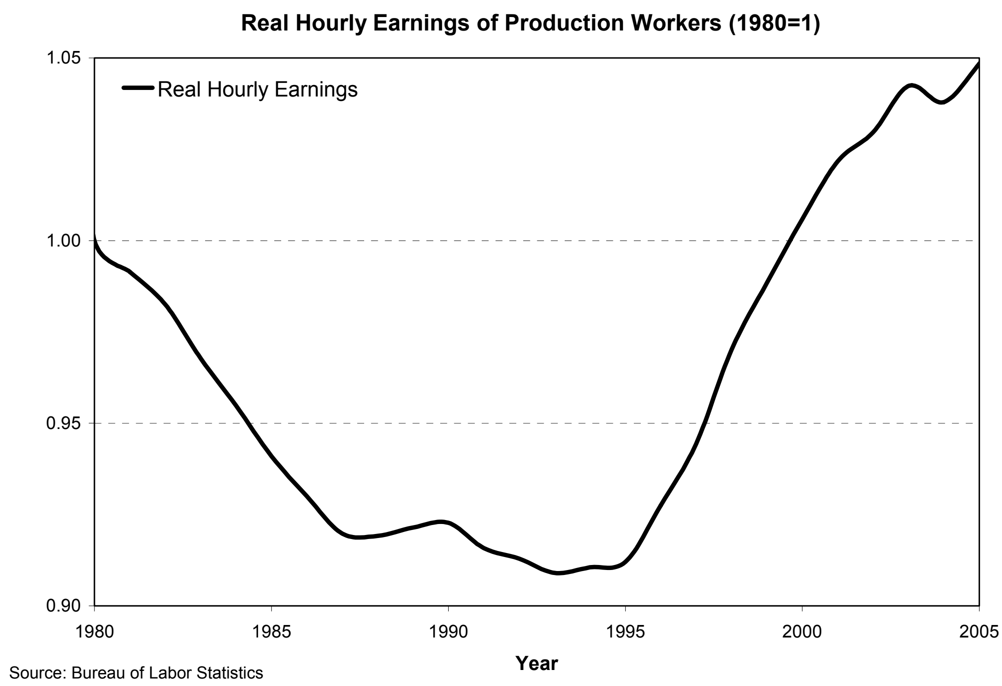


Figure 2: Employment Shares and Relative Prices



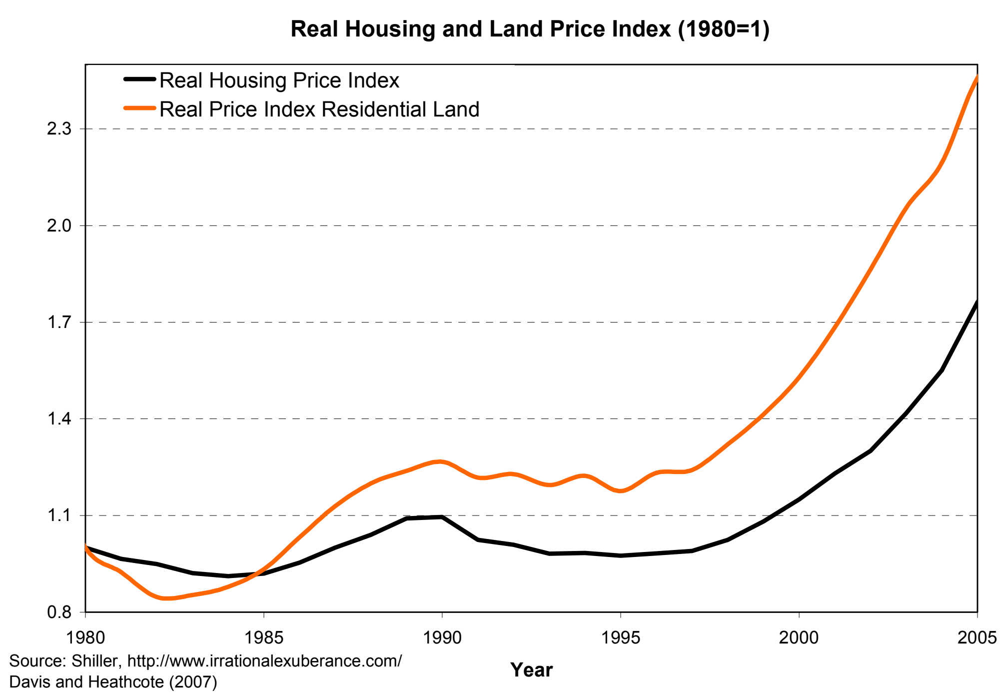
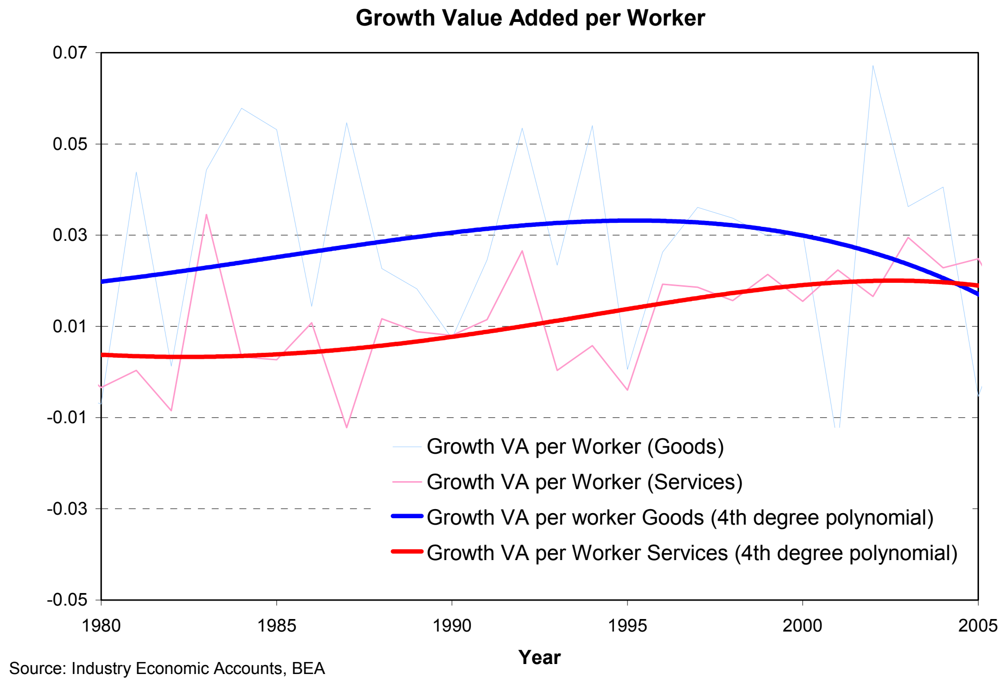


Figure 3: Wages, Land Prices, and Housing Prices



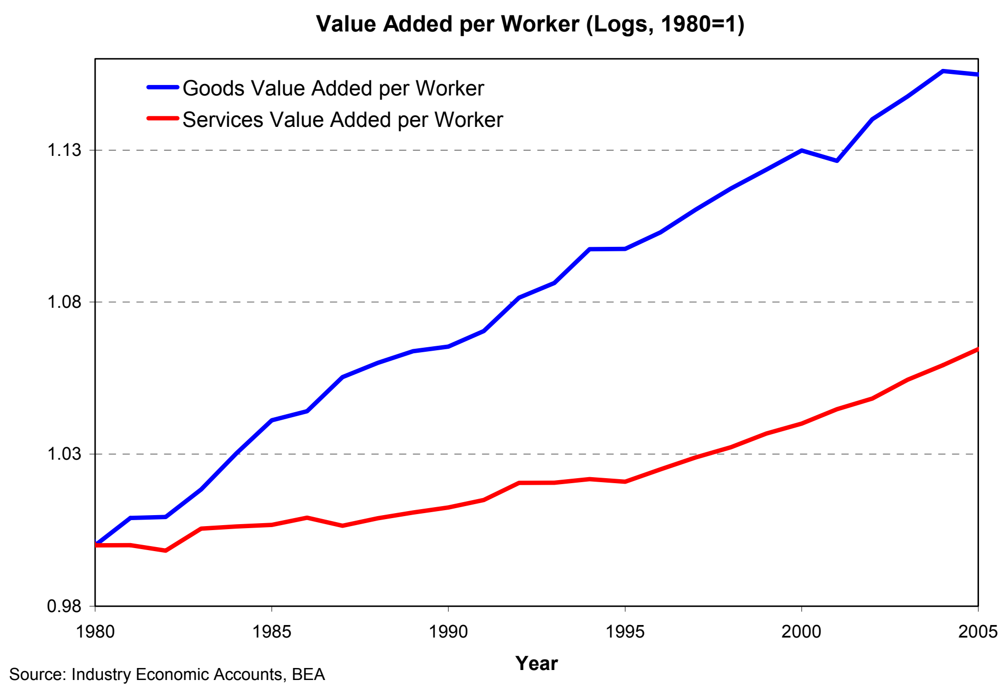


Figure 4: Value Added per Worker

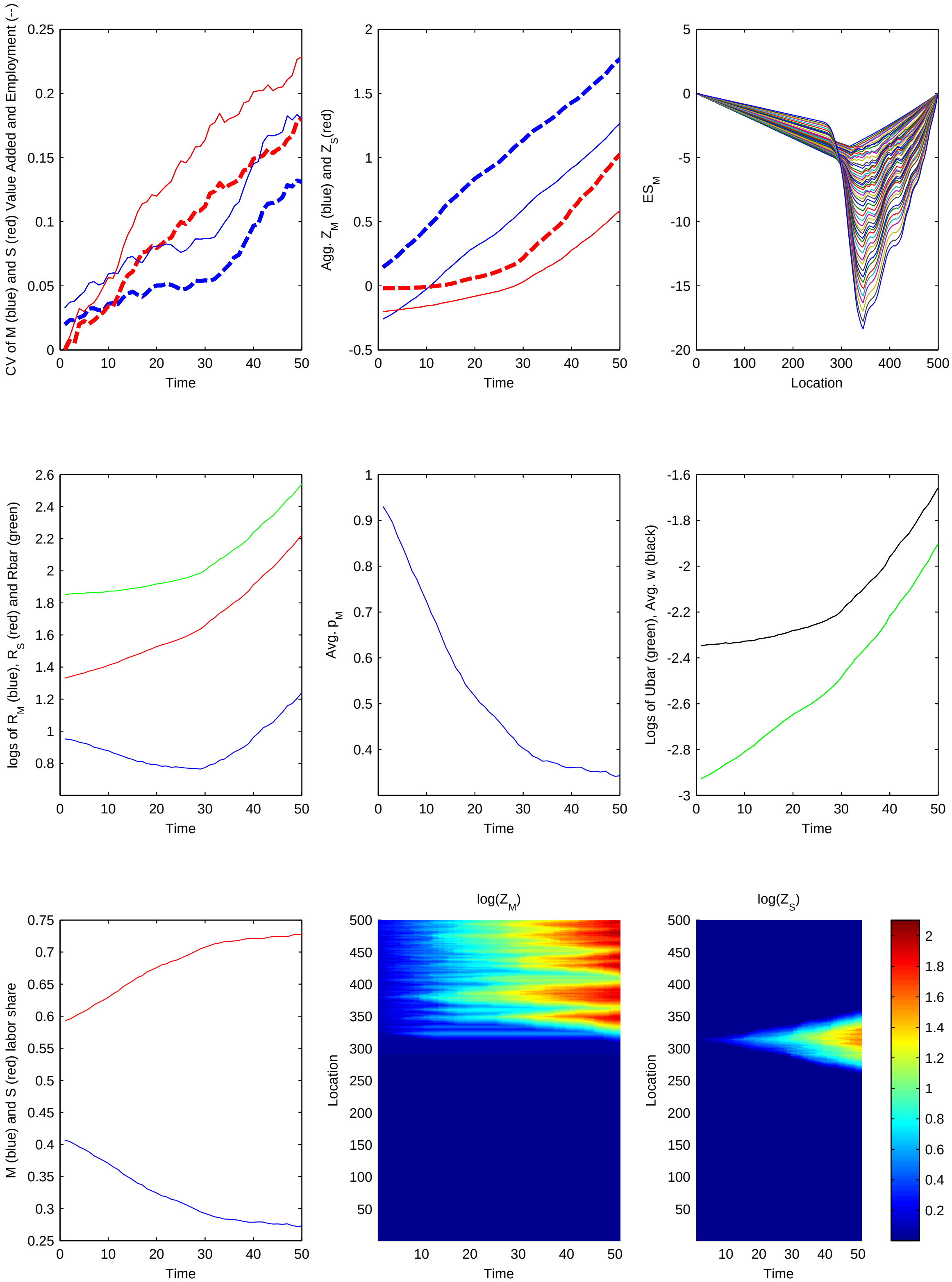


Figure 5: Simulation Results for Benchmark Parameterization

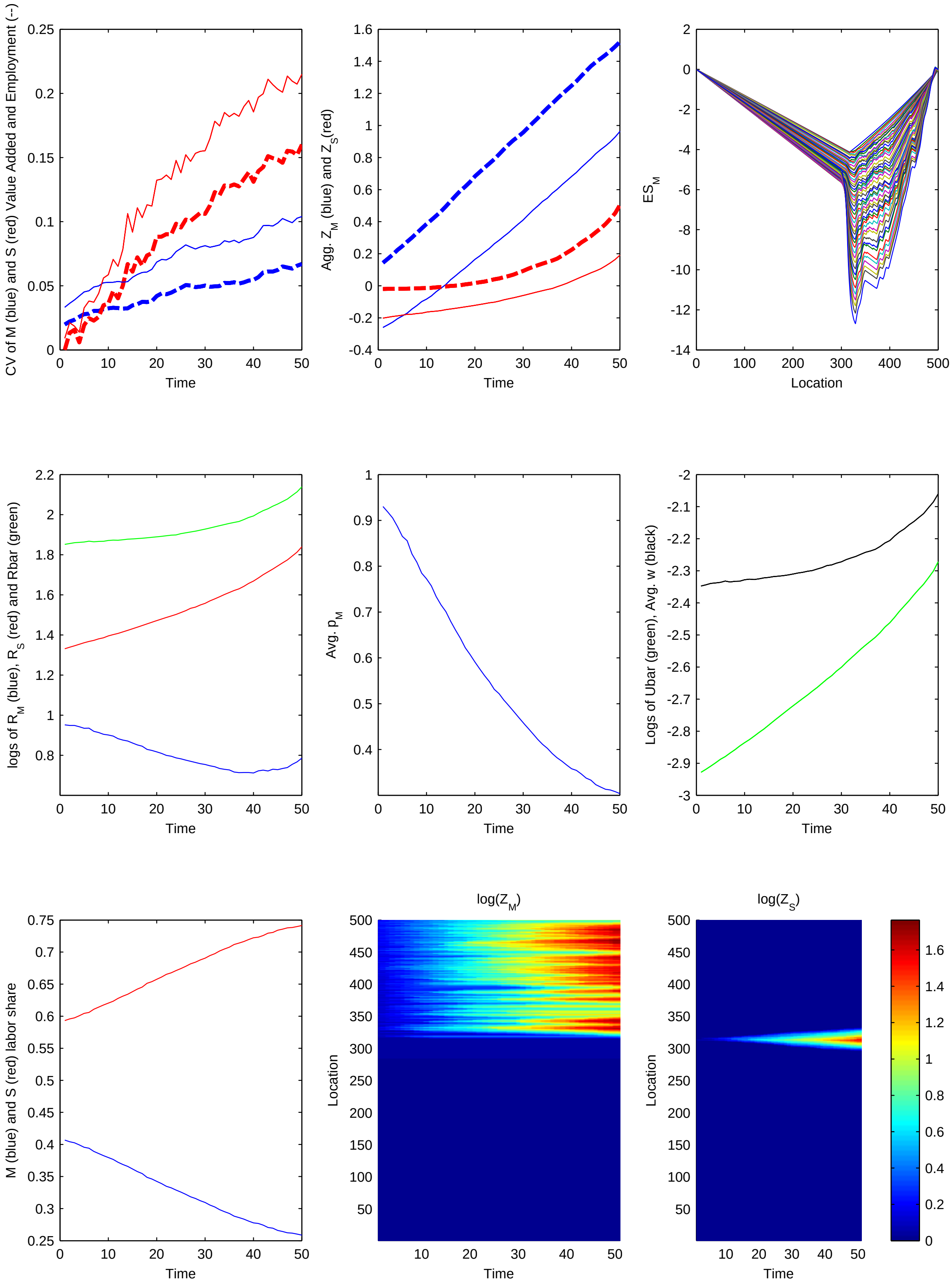


Figure 6: Fast Declining Diffusion ($\delta = 50$)

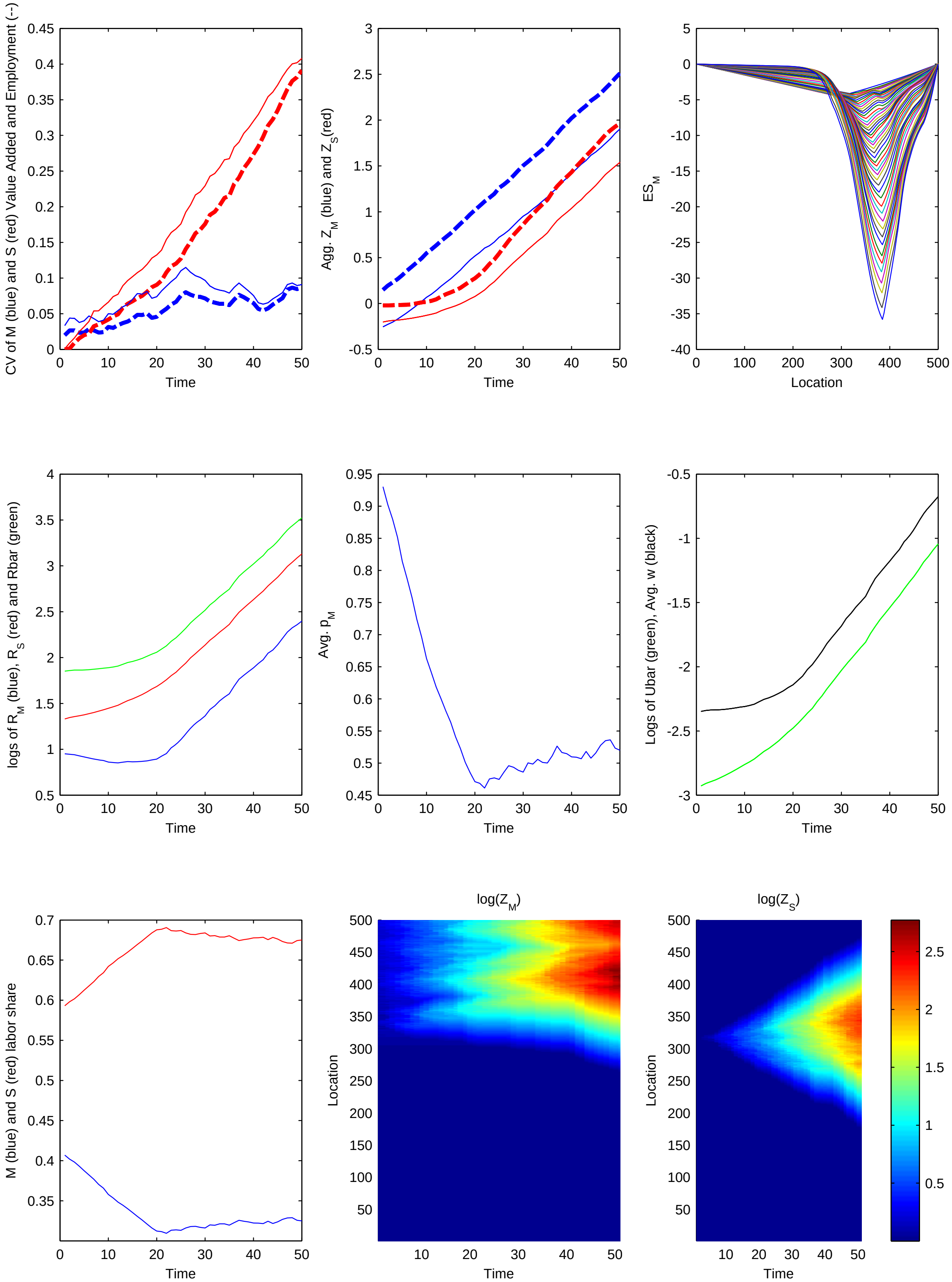


Figure 7: Slow Declining Diffusion ($\delta = 10$)

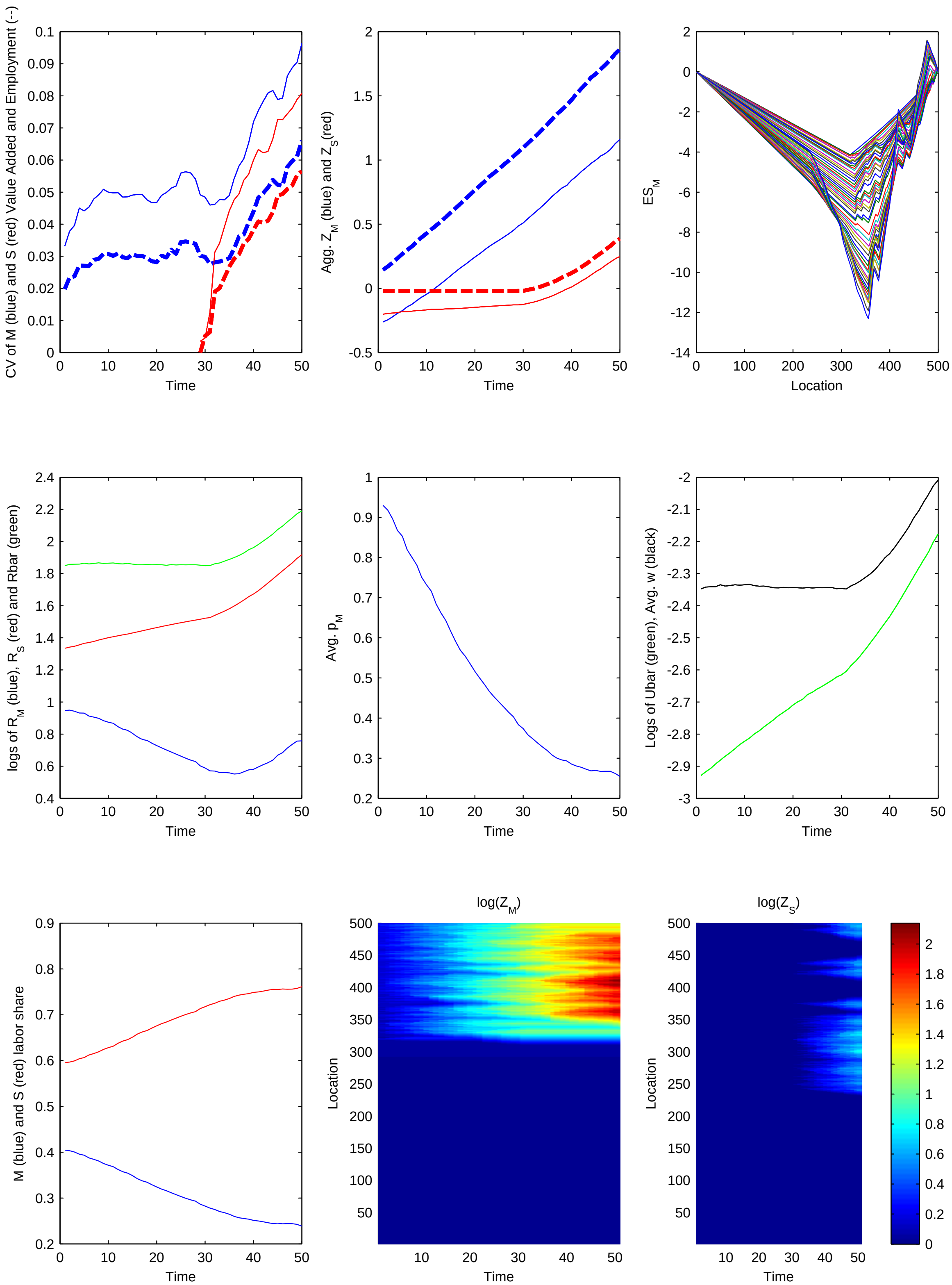


Figure 8: Low Transport Costs ($\kappa = 0.005$)

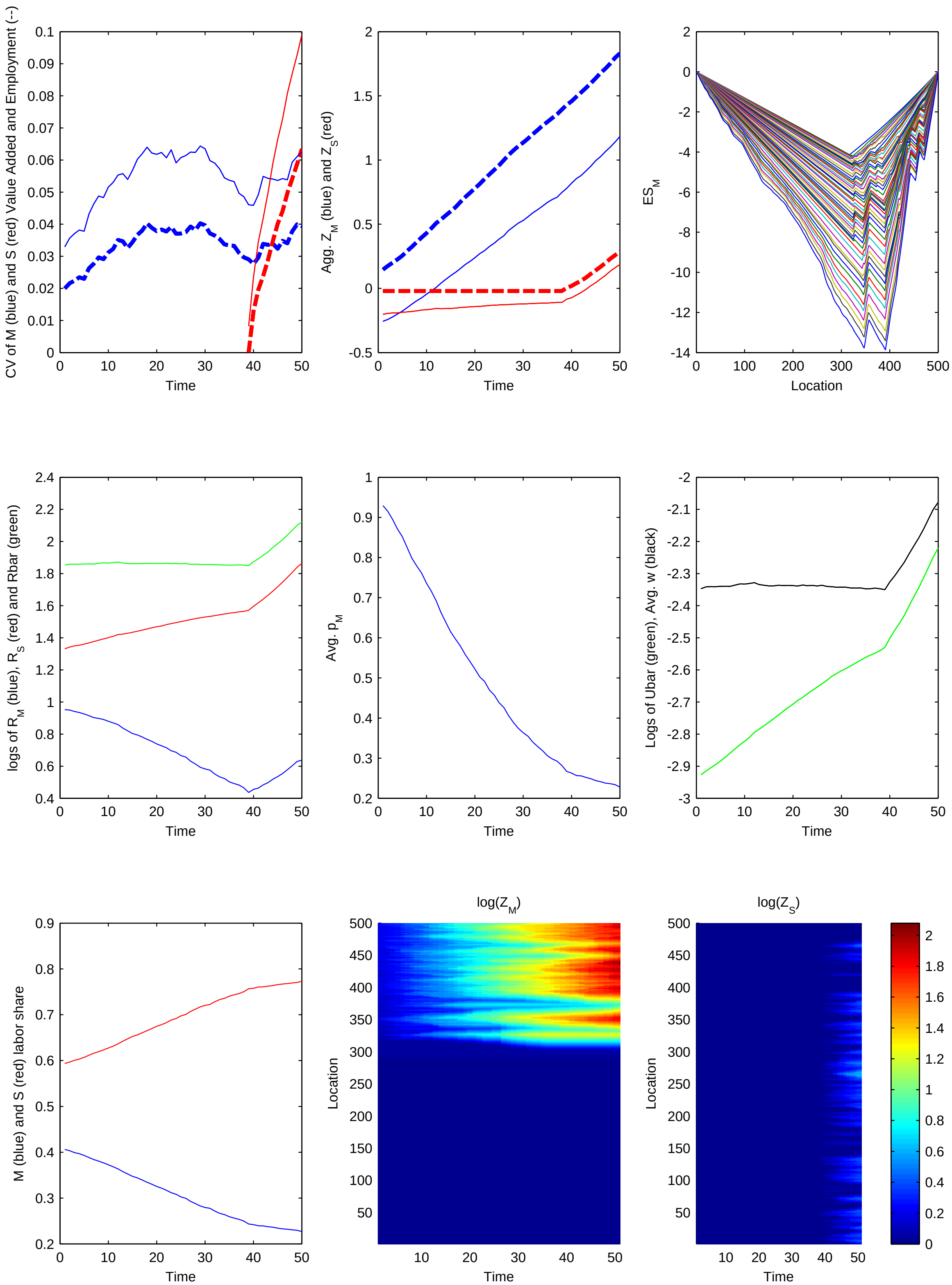


Figure 9: Very Low Transport Costs ($\kappa = 0.001$)

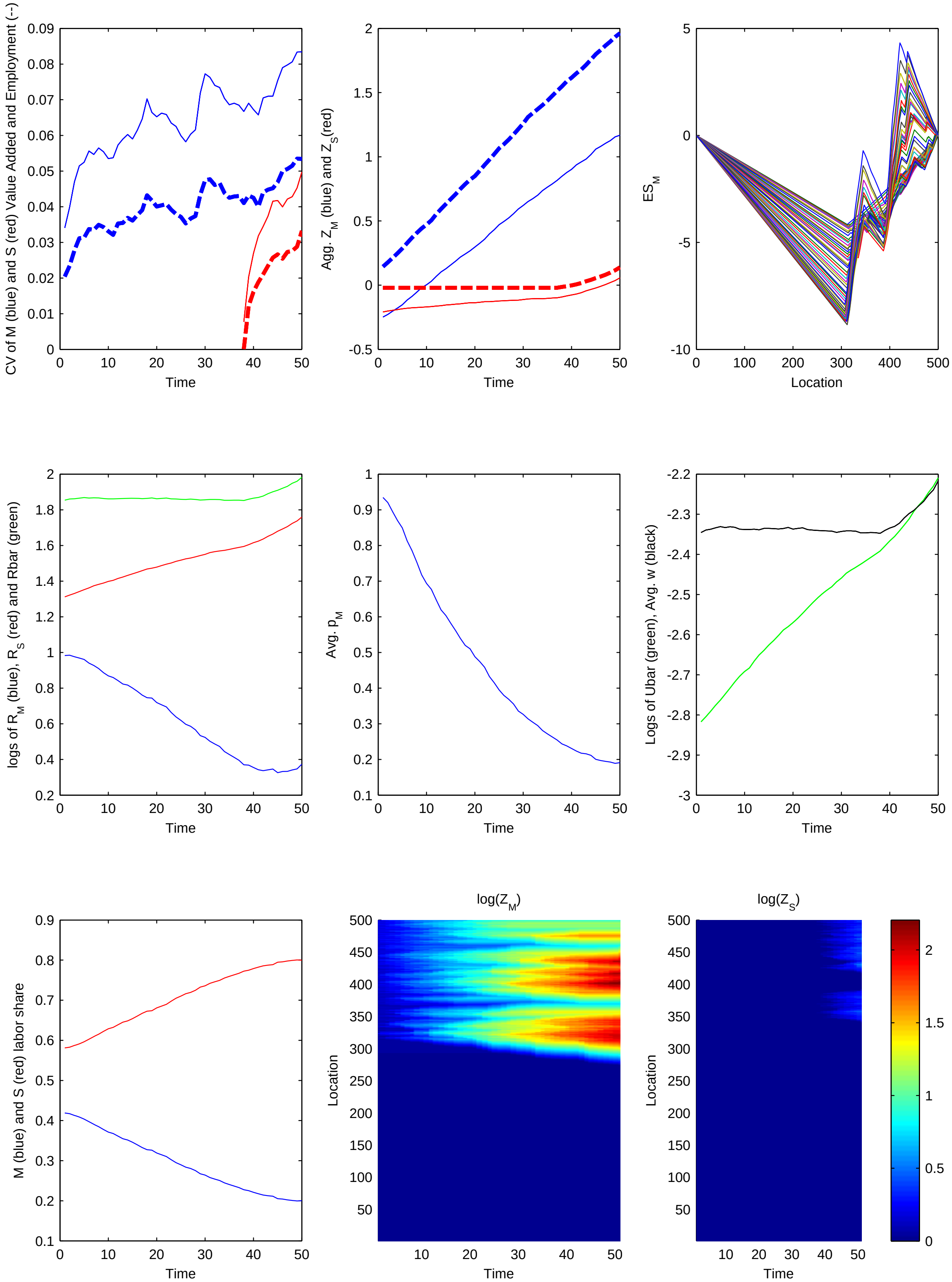


Figure 10: Low Elasticity of Substitution ($\alpha = -2$)

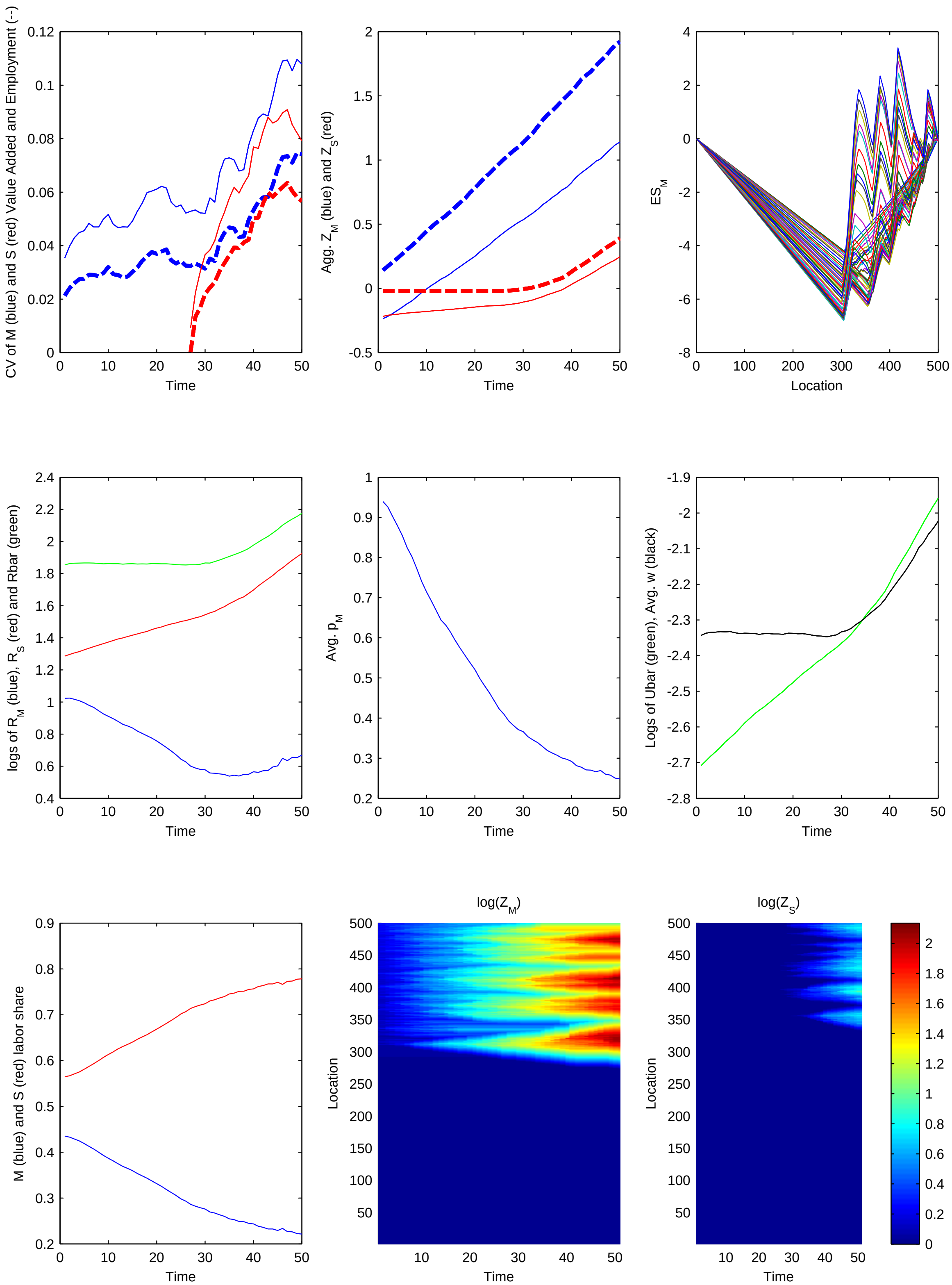


Figure 11: Very Low Elasticity of Substitution ($\alpha = -3$)