

## Supporting Information on the Modeling Framework

### Data:

Socio-economic data is drawn largely from Chinese statistical yearbooks (Guangdong Statistical Yearbooks, various years) but also based on the literature and interviews. While the systems model is initialized using the socio-economic data in tabular format, the statistical logit model reads the road networks, topography from a geographical database through GIS. The results of the previous remote sensing analyses of the region are also incorporated to track the location and growth of urban land over time.

A problem we encountered in data collection was obtaining reasonable population and labor force time series that has proven impossible due to the *hukou* system that continues to regulate location of official residence and frequent changes in official definitions. Official statistics show population levels lower by almost an order of magnitude than is generally believed to be the case and decreases in population during periods of rapid urbanization. Since an understanding of both the economy and the physical expansion of urban area are highly dependent on population levels, we estimate population time series of the counties based on the literature and interviews. This provides reasonable estimates of the population dynamics in tune with the actual situation in the region.

### System Dynamics Model:

Although, systems analysis has been utilized in spatially-explicit dynamic problem contexts before (Voinov, Costanza et al. 1999; Muetzelfeldt and Massheder 2003; Lee, Huang et al. 2008), such spatially-explicit system models are built to primarily study biophysical systems such as ecosystem or hydrologic dynamics (Ruth and Pieper 1994; Ahmad and Simonovic 2004; BenDor, Metcalf et al. 2006) the finer-scale representations of which can be relatively straightforward compared to those of socio-economic systems. The socio-economic components of such models, if present, are represented with either traditional econometric methods (Bockstael 1996) or at best, without clearly delineating the embedded feedback processes that drive the dynamics (Berry, Hazen et al. 1996; de Kok, Engelen et al. 2001).

The primary assumptions in the model are in Table S1. The population submodel simulates total population in a county and interacts with the three economic sectors through the labor force variable. The structures of the three economic sectors conform to a modified form of Cobb–Douglas production function, one which conforms to the basic tenets of an ecological economics perspective. While our formulation still retains the implicit assumption of substitutability between labor and capital, it differs from the traditional use of the function in three significant ways: 1) our framework places the economy in a larger system that includes the environment; 2) it acknowledges the nonlinear feedbacks and delays between the economy and the environment; 3) we challenge the assumption of an infinite resource base by incorporating land availability and potential shortages in natural resources. This variable reflects economic policies in China, which,

over the past decade, have put more emphasis on western and northeastern provinces than the south, thereby reducing the region's comparative advantage.

### ***Decision functions and variables***

The model incorporates functional forms that reflect the responsiveness of decision-makers to environmental and economic signals. Most of the functions representing decision-making processes in the model are related to economic decisions regarding investment flows, but some account for changes in population flow such as in-migration to the city. The general forms of these functions are developed from the literature and are also based on our knowledge of the study area. Nevertheless, due to the uncertainty surrounding their exact functional specifications, we run comprehensive sensitivity tests to account for uncertainty in the simulation results.

### ***Validation***

Overall, we judge a given model's validity by the degree of its usefulness in gaining insight about the real system it represents and by its utility in making decisions. The validation of the model components in this framework are carried out in a sequential manner starting from the calibration and validation of each sector of each submodel and progressing as the submodels are brought together. This was followed by further calibration and validation of the entire model consisting of structural and behavioral validity tests (Barlas 1996). Structural tests assess the logical validity of model equations by evaluating them one by one and by testing their behavior under extreme conditions. With behavior validation, the patterns of major variables generated by the model are evaluated against the available data on each county and the literature. The comparison between model-generated values and observed values allows us to evaluate if the model is an adequate representation of the real-world system. We emphasize that our model is expected to capture the broad dynamic patterns of the system (behavioral validation) for the correct reasons (structural validation). In addition to these formal tests, we also consulted experts on Pearl River Delta and China during the model building and validation processes.

### ***Population and Laborforce submodels***

The population submodel simulates total population in each county based on endogenous net birth rate and migration flows. The population submodel interacts with the three economy sectors through the labor force variable. The appeal of each county for migrant workers is influenced by job availability. These are determined in part by the comparative advantage of the PRD region relative to other Chinese regions and metropolitan areas (Yeung and Chu 1998). Job availability and comparative competitiveness of the region are proxies for the economic draw of a county (Fan 1996; Scharping and Huaiyang 1997). The primary determinants of migration rate are given in Equation 1. The subscript of the counties are omitted for clarity.

$$mr = mbr * e(j) * ra \quad (1)$$

where  $mr$ ,  $mbr$ ,  $e(j)$ ,  $j$ , and  $ra$  are migration rate, migration base rate, effect of job availability on migration, labor supply-demand ratio, and comparative advantage of PRD, respectively.  $e(j)$  and  $ra$  are functions of labor supply-demand ratio and time, respectively (Figure S2a-b). The labor submodel reallocates available workforce from the population submodel to different economic sectors based on their demand for labor.

### ***Economic Submodels***

The primary sector production is assumed to be composed of agricultural production (crops and orchards) (State Statistical Bureau various years). While the investment and capital stocks include all primary sector activities such as aquaculture, husbandry, and forestry in addition to agriculture, the production is dynamically computed only for the agricultural production. The production for husbandry, fishery and the production per unit forested land for forestry are assumed to be a function of cultivated land. This assumption does not significantly affect our results because the data shows declining GDP share of primary sector activities throughout the region over the past twenty years. The secondary sector includes all types of manufacturing activities, construction, mining and quarrying (State Statistical Bureau various years). The tertiary sector includes commercial, banking activities, tourism, and other service activities. The total factor productivity ( $tfp$ ) and labor technology factor ( $ltf$ ) variables for the economic sectors are estimated based on the data as well as the literature (Perkins 1996; IMF 2005).

Capital stock values are needed to calculate sectoral production, but are not reported in the statistical yearbooks. Consequently, we use data on the flows of new investment to estimate the capital stock (IMF 2005; State Statistical Bureau various years). New investments accumulate as installed capital, which depreciates with a fixed rate that is different for each sector (Bai, Hsieh et al. 2006). In the secondary and tertiary sectors, capital and labor employed (together with total factor productivity,  $tfp$  and labor technology variables,  $ltf$ ) drive sectoral production represented by sectoral GDP. Using sectoral GDP as a proxy for sectoral profitability does not fit well in an ecological economics framework (Table S1). However, sectoral GDP is still regarded as a common measure of the vitality of any particular sector. Parallel to this, its main function in the model is to inform potential investors of this aspect of the economy. In the primary sector, land is used in place of labor but a labor demand is still calculated based on estimated capital-labor ratio and the corresponding  $ltf$  variable. The change in the contribution of each sector to the total GDP in turn affects new investment flows in that sector.

Another factor that affects investment decisions in the secondary and tertiary sectors is the availability of land. In the primary sector, the capital to land ratio is used as a factor affecting investment in this sector. The function representing the effect of capital land ratio on new investment rate in the primary sector monotonically decreases as the ratio increases reflecting the

decreasing profitability from a unit cultivated land as more and more capital is installed. The tertiary sector structure differs from the secondary sector structure in that demand for tertiary sector serves by the increasingly affluent population is an additional factor driving investment decisions in this sector (Treyz 1993; Wu 1997). GDP per capita is used as a measure of the welfare of the population, which is a common practice in economic studies (Holtz-Eakin and Selden 1995). However, our use of GDP per capita should be interpreted as representing only the material well-being of the population. We assume a nonlinear relationship between increased material welfare and investments in the tertiary sector. The primary relationships in determining investment rates for the secondary and tertiary sectors are presented in Equation 2.

$$ir_s = ibr_s * g_s (f_u) * (h_s (GpCi) + n_s) * nc(rGDP) * ra \quad s=\text{secondary, tertiary} \quad (2)$$

where

$ir_s$      ≡ investment rate,

$ibr_s$     ≡ investment base rate,

$g_s$        ≡ effect of land availability,

$f_u$        ≡ fraction of urban land,

$h_s$        ≡ effect of material well-being of the population on investment,

$GpCi$     ≡ GDP per capita index (a measure of material well-being of the population relative to Shenzhen (the county with the highest GDP per capita in the base year 1988),

$nc$        ≡ the impact of economic development in neighboring counties, and

$rGDP = GDP / (\sum GDP_c + GDP)$  where  $GDP$  is gross domestic product of the target county and  $GDP_c$  is that of each of its neighbors.

The term  $n_s$  represents the influence of nonlocal factors such as export activities and is assumed to be equal to one for both sectors. The functional forms of  $g_s$  for both sectors and  $h$  for tertiary sector are presented in Figure S2c-d and Figure S3a, respectively;  $h$  for the secondary sector is assumed to equal zero. The functional form for  $nc$  is given in Figure S3b. The equation for investment rate for primary sector is similar to Equation 2 with the exception that the impact of land availability is replaced by a capital-land ratio factor (Figure S3c).

**Spatial-logit model:**

Independent variables were built in a geographic information system from the classified Landsat images and other sources. The land cover type in the start year of a given period is assessed through the inclusion of three dummy variables drawn from classified Landsat imagery and representing the presence of agricultural lands, fishponds, and water. All three of these landcover classes are expected to decrease the probability of urbanization at a location because the first two represent profitable land uses competing with urbanization and the last implies major costs in urban development required for the filling in of major water bodies or the delta itself. However, with incredibly rapid economic and urban expansion, all of these may fail to hinder conversion to urban land cover. Similarly, the independent variable indicating percent slope at a site (derived from a DEM) is theoretically expected to increase the costs of development thus decreasing its probability. However, the exceptionally dynamic nature of growth in PRD and author observations of frequent flattening out of the region's topography leave the actual strength of the relationship in question (Field Visits [Enter earlier years of Karen's visits]], 2001, 2006, 2007, 2008).

Regional accessibility is represented with a number of distance gradients built in a GIS. Distance to the historical urban centers is captured by computing this distance to central Guangzhou and Dongguan. Distance from Shenzhen captures both access to the center of the Shenzhen Special Economic Zone and the Hong Kong border crossing. Access to the regional transport system is indicated by the distance of a given cell to the nearest major or medium roadway. All roads were digitized from Landsat imagery for a given start year with quality control and category assignment involving the consultation of historical highway maps and Google Earth. Major roads are multi-lane highways generally with controlled access and medium roads are any road that provides largely linear travel at a medium distance scale for through traffic or what would be called an "arterial" in the U.S. Local accessibility to the people and services present in existing urban land cover is proxied for by the inclusion of a variable built by passing a 1-kilometer mean filter of the urban landcover in the start year (i.e. the value indicates the proportion of the land within 1-kilometer of a site that is already urbanized in the start year). Greater accessibility of any sort is expected to increase the probability of urban land conversion though the significance and strength of the various types of accessibility are expected to vary greatly depending on the importance of particular spatial interconnections for the firms and people occupying the newly developed landscape. Additionally, variables indicating the simple geographic coordinates of each pixel are included to assess the spatial patterning of omitted drivers and lessen the impact of type-1 spatial autocorrelation.

Beyond these varied influences on the spatial configuration of development at the local level are three drivers of urbanization drawn from the upper level system dynamics model. The change in population and GDP during the time period (each controlling for the county's area) is included as a driver of land use conversion. The relationship between each of these changes and the probability of urban conversion on a given pixel is theoretically uncertain due to variation in the density of development. On a basic level, one might expect growth in population and GDP both to increase the probability of urbanization since more houses and firms take up more land. However, further consideration of these economic processes reveal that either variable could also have the opposite influence. For example, greater increases in population within a time period (i.e. more rapid increases) may drive up land prices (or pseudo-prices in the PRD case) leading to higher density residential development. More rapid increases in GDP may be associated with agglomerative effects resulting from a tighter spatial clustering of firms (Fujita and Thisse 2002). Either condition could result in a situation where more growth co-occurs with less landscape transformation so empirical work is necessary to understand the directionality of the relationship. Additionally, the relationship between economic growth and land consumption is likely tempered by the sectoral composition of the growth with services growth consuming less than industrial growth. This relationship is represented through the inclusion of a variable indicating the percentage of growth that was in the service industry. These three drivers connect our lower level spatially explicit model with an appropriately higher level representation of regional economic and population dynamics.

Figures

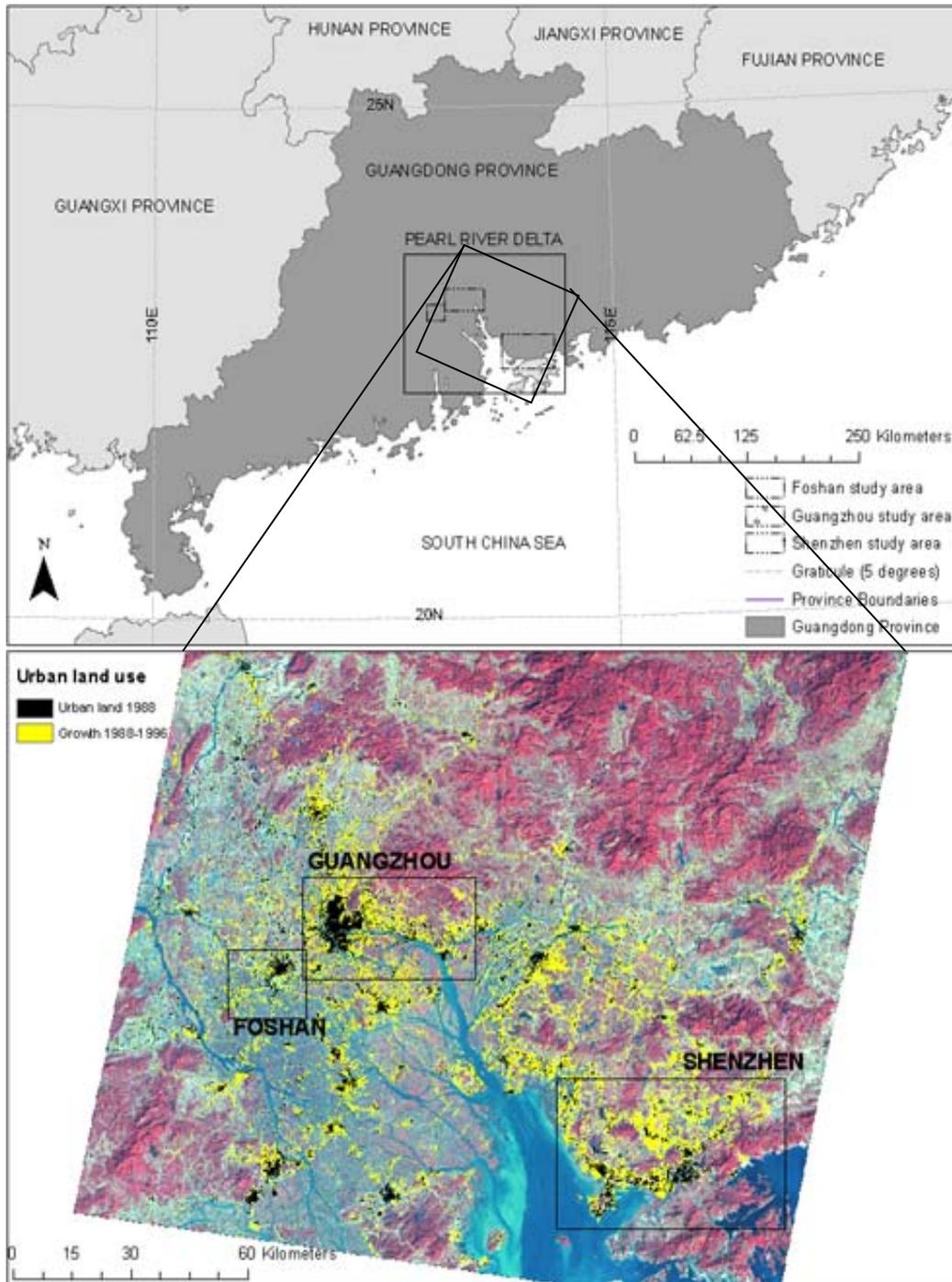


Figure S1. Location of Shenzhen in the Pearl River Delta.

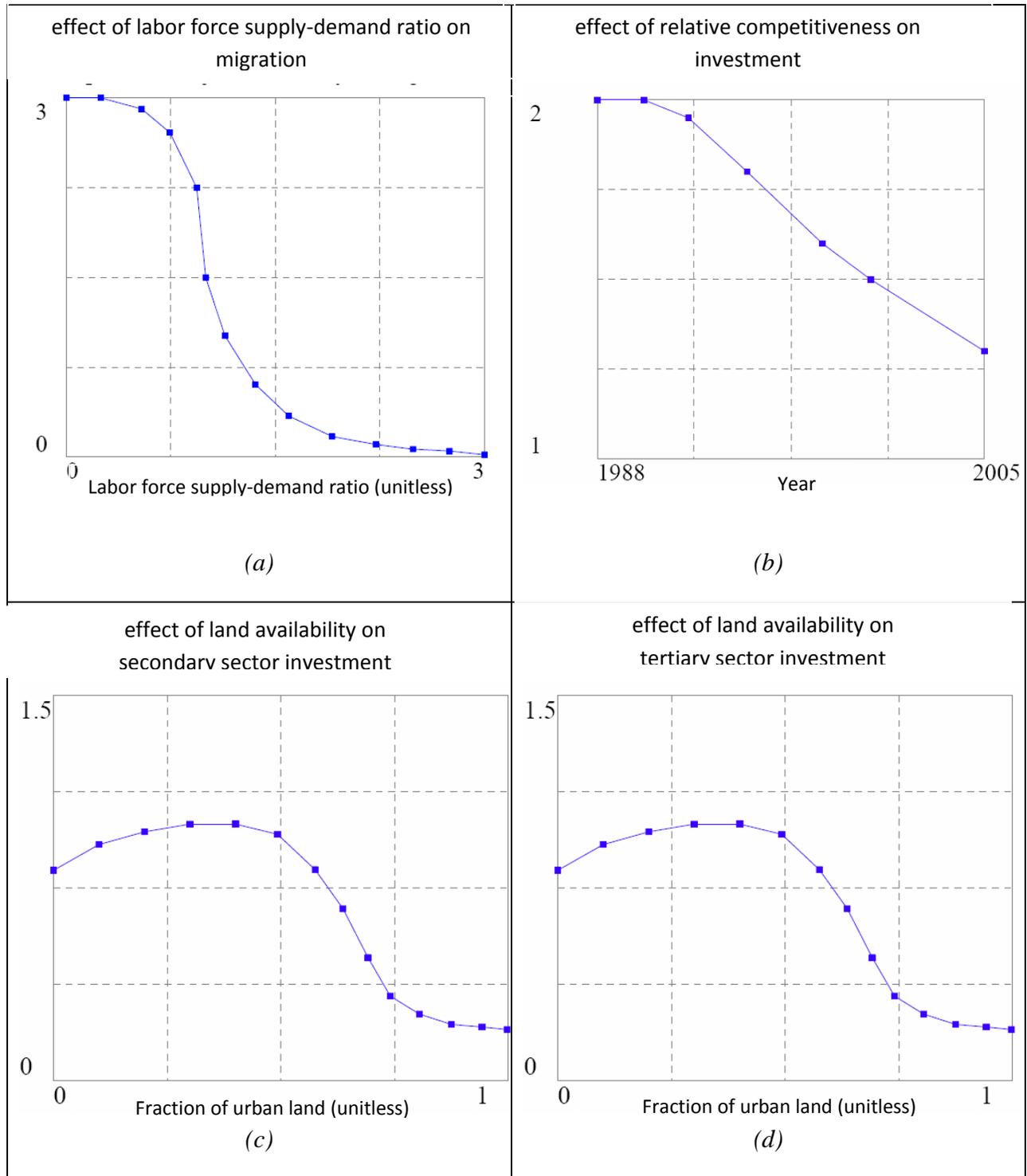


Figure S2. Most important functional relationships in the model regarding the effect of (a) job availability on migration, (b) comparative advantage of the PRD, and (c-d) fraction of developed land on, respectively, secondary and tertiary sector investment.

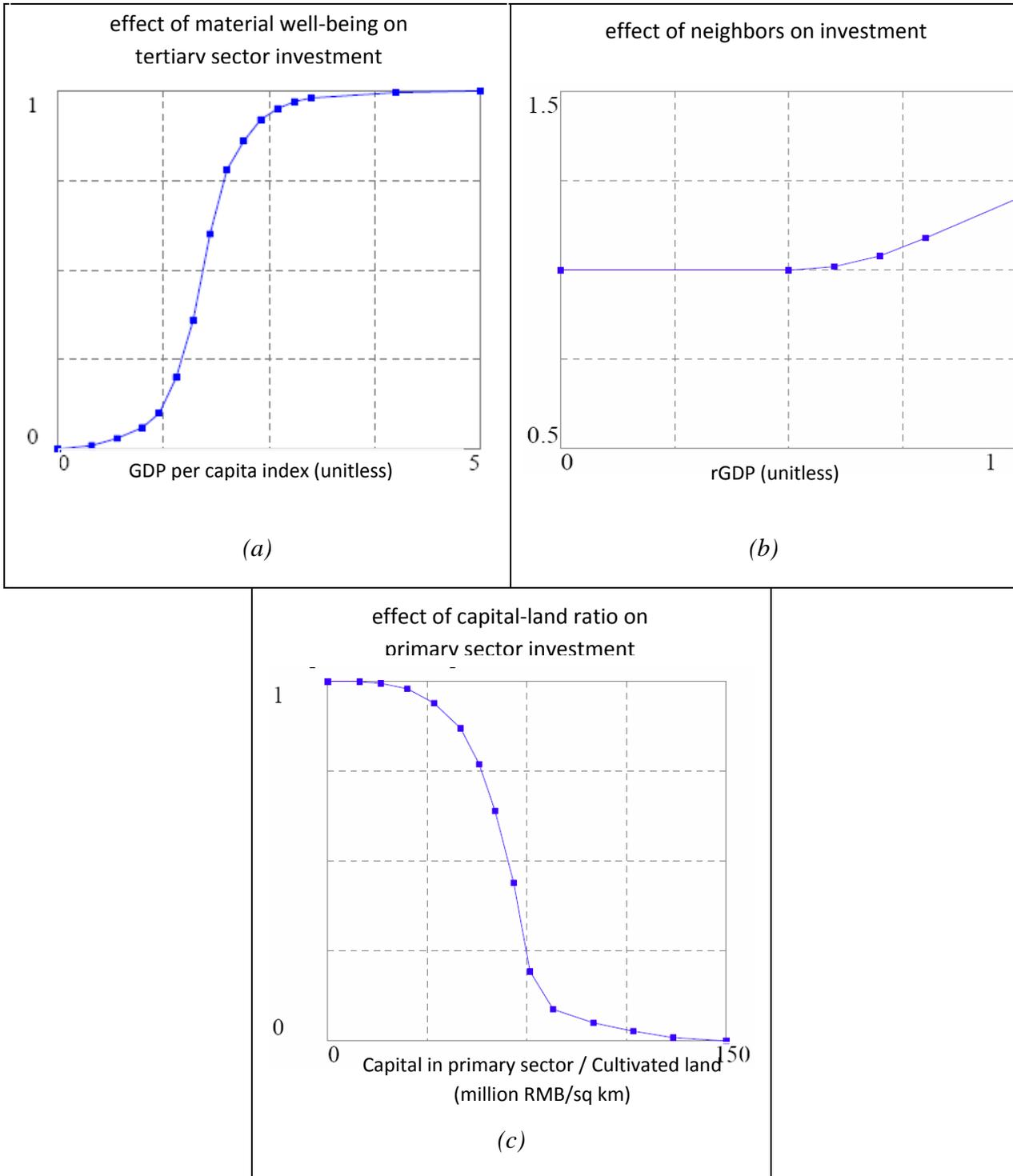


Figure S3. Most important functional relationships in the model regarding impacts of (a) material well-being on tertiary sector investment, (b) neighboring counties on investment, and (c) capital-land ratio on primary sector investment.

Table S1. Primary assumptions of the model.

1. The populations exhibit the same fertility and mortality behavior throughout the region.
2. The labor force is interchangeable among the three sectors.
3. Potential migrants are assumed to be attracted exclusively by job availability and comparative attractiveness of the region.
4. The comparative attractiveness of the region is assumed to decrease over time.
5. In driving investment in all three economic sectors, growth in sectoral GDP share is used as a proxy for the profitability of that sector.
6. 1980 real prices are used.

## References

- Ahmad, S. and S. P. Simonovic (2004). "Spatial system dynamics: New approach for simulation of water resources systems." Journal of Computing in Civil Engineering **18**(4): 331-340.
- Bai, C. E., C. T. Hsieh, et al. (2006). The return to capital in China. National Bureau of Economic Research Working Paper 12755. Cambridge, MA.
- Barlas, Y. (1996). "Formal aspects of model validity and validation in system dynamics." System Dynamics Review **12**(3): 183-210.
- BenDor, T. K., S. S. Metcalf, et al. (2006). "Modeling the spread of the Emerald Ash Borer." Ecological Modelling **197**(1-2): 221-236.
- Berry, M. W., B. C. Hazen, et al. (1996). "Lucas: a system for modeling land-use change." Computational Science & Engineering, IEEE **3**(1): 24-35.
- Bockstael, N. E. (1996). "Modeling economics and ecology: The importance of a spatial." American Journal of Agricultural Economics **78**(5): 1168.
- de Kok, J., G. Engelen, et al. (2001). "Modeling land-use change in a decision-support system for coastal-zone management." Environmental Modeling & Assessment **6**(2): 123.
- Fan, C. C. (1996). "Economic opportunities and internal migration: a case study of Guangdong province, China." Professional Geographer **48**: 28-45.
- Fujita, M. and J.-F. Thisse (2002). Economics of Agglomeration: Cities, Industrial Location, and Regional Growth. Cambridge UK, Cambridge University Press.
- Holtz-Eakin, D. and T. M. Selden (1995). "Stoking the fires? CO2 emissions and economic growth." Journal of Public Economics **57**: 85-101.
- IMF (2005). People's Republic of China: 2005 Article IV Consultation— Country Report No. 05/411. Washington, DC., International Monetary Fund.
- Lee, C. L., S. L. Huang, et al. (2008). "Biophysical and system approaches for simulating land-use change." Landscape and Urban Planning **86**(2): 187-203.
- Muetzelfeldt, R. and J. Massheder (2003). "The Simile visual modelling environment." European Journal of Agronomy **18**(3-4): 345-358.
- Perkins, D. H. (1996). "Productivity performance and priorities for the reform of China's state-owned enterprises." Journal of Development Studies **32**: 414-444.
- Ruth, M. and F. Pieper (1994). "Modeling spatial dynamics of sea-level rise in a coastal area." System Dynamics Review (Wiley) **10**(4): 375-389.
- Scharping, T. and S. E. Huaiyang (1997). Migration in China's Guangdong province: major results of a 1993 sample survey on migrants and floating population in Shenzhen and Foshan. Hamburg, Germany, Institut für Asienkunde.
- State Statistical Bureau (various years). China Statistical Yearbook New York, Praeger.
- Treyz, G. I. (1993). Regional Economic Modeling: A Systematic Approach to Economic Forecasting and Policy Analysis. Norwell, Kluwer Academic Publishers.
- Voinov, A., R. Costanza, et al. (1999). "Patuxent landscape model: integrated ecological economic modeling of a watershed." Environmental Modelling & Software **14**: 473-491.
- Wu, W. (1997). "Proximity and complementarity in Hong Kong-Shenzhen industrialization." Asian Survey **37**: 771-793.
- Yeung, Y. M. and D. K. Y. E. Chu (1998). Guangdong: Survey of a Province undergoing Rapid Change. Hong Kong, The Chinese University Press.