Economic Quarterly—Volume 95, Number 1—Winter 2009—Pages 1–24

Problems for a Fundamental Theory of House Prices

Andreas Hornstein

The recent turmoil in the U.S. residential housing market affects mainly the market for owner-occupied housing. In this market, most owners have less than a complete equity share in their home; rather, they obtain a mortgage and borrow against the value of their home. There is a presumption that over the last 30 years financial innovations have made it easier for households to borrow against the collateral value of their homes, thereby increasing the demand for housing and house prices. In this article we will argue that standard theories of the residential housing market do not predict that changes in collateral constraints significantly affect aggregate house prices. In fact, these standard theories find it difficult to account for the observed sustained house price increases. This suggests that we develop better theories of the underlying demand and supply for housing before we proceed to study the effects of financial frictions on the housing market.

There are two components of the market for single-family housing—the market for existing homes and the market for new homes. Changes in these two markets affect the aggregate economy in different ways, and over the last 30 years these two markets have behaved very differently. Almost by definition, the supply of existing homes in mature neighborhoods is less elastic than the supply of new homes in new neighborhoods. After all, the location of an existing home and the characteristics of its neighborhood cannot be easily replicated, whereas the supply of new land on the suburban fringe is relatively elastic, and the relative price of new homes is mainly determined by the price of residential structures. Thus, changes in the demand for housing should

I would like to thank Kartik Athreya, Ned Prescott, and Pierre Sarte for helpful comments and Nadezhda Malysheva for excellent research assistance. Any opinions expressed in this paper are my own and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: andreas.hornstein@rich.frb.org.

mainly show up in the relative price of existing homes and the construction of new homes. Indeed, existing house prices have increased substantially relative to new house prices. At the same time increased construction of new homes has directly contributed to gross domestic product (GDP) through its contribution to investment in residential structures. Higher relative prices of existing homes affect GDP only indirectly through wealth redistribution between current owners and potential future owners.

The ability to obtain credit is affected by household income and by the available credit arrangements. For example, if household income increases, not only is there likely to be a demand for more housing services, but also an increase in the rate at which households save. A higher savings rate should enable them to make a down payment for a house earlier in their life cycle; that is, they enter the housing market earlier in their life cycle and this increases the demand for owner-occupied housing. Similarly, allowing households to put down a smaller down payment on the home purchase is likely to increase the demand for housing.

The growth of the government sponsored enterprises (GSEs), Fannie Mae and Freddie Mac, has changed the market for owner-occupied housing. These GSEs purchase mortgages that satisfy certain criteria and they issue securities backed by these mortgages. They have thereby "commodified" mortgages and, in the process, have reduced the borrowing costs for homeowners. Similarly, the growth of the market for subprime mortgages after 2000 introduced new segments of the population to the market for owner-occupied housing. The increasing share of subprime mortgages in overall mortgages was, to a large extent, driven by the ability of mortgage issuers to securitize these mortgages. The subprime mortgage market collapsed in 2007 and it is uncertain if it will reemerge in the future and, if so, what form it might take. Overall, innovations in financial markets have affected the demand for housing in the past and they are likely to affect that demand in the future.

In Section 1, we review some of the data on house prices and the availability of mortgage credit to households. In Section 2, we describe a simple model of the housing market based on Davis and Heathcote (2005), where land is an essential input to the production of houses. This model attributes endogenous changes in the price of housing to changes in the relative scarcity of land. In order to understand long-run trends in house prices, we study the balanced growth path of this model and find that the model is reasonably successful at accounting for long-run changes in the price of new homes. In Section 3, we model the demand for mortgage-financed housing using the Campbell and Hercowitz (2006) representation of collateral constraints. We find that changes in collateral constraints hardly affect the balanced growth path of house prices. Like most aggregate models of the housing market, the baseline housing model treats new and existing homes as perfect substitutes even though we have seen a marked divergence in the relative price of both types of homes. Therefore, in Section 4 we argue that future research in housing should develop a theory that accounts for the differences between the market for new homes and the market for existing homes.

1. HOUSE PRICES AND FINANCIAL INNOVATIONS

The price of U.S. homes has increased significantly since the mid-1990s, and most of this price increase has shown up in the price of existing homes as opposed to the price of new homes. Over the last 30 years it has also become easier for owners to borrow against the collateral value of their home. The 2004–2006 boom of subprime mortgage lending was just another development that expanded the set of households that could enter the market for owner-occupied homes. One might, therefore, argue that house prices have increased because financial innovations that lowered the cost of owner-occupied housing have increased the demand for housing. In this section we summarize some of the developments in the U.S. housing market that pertain to house prices and the ability of homeowners to borrow against the value of their home. See the Appendix for a detailed description of the time series.

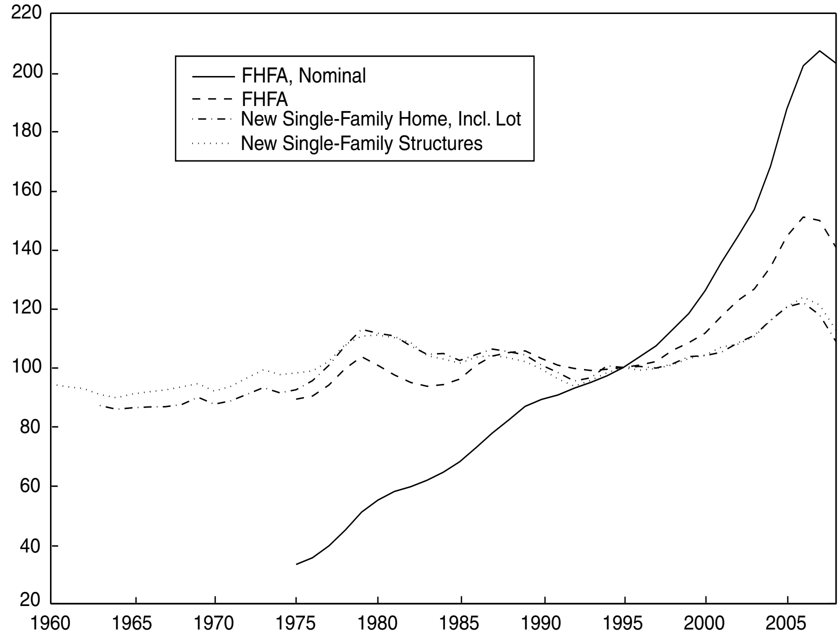
The nominal price of existing single-family homes in the United States has been steadily increasing since the 1970s and this process accelerated in the late 1990s (see Figure 1).[1] Even though the nominal price of existing homes increased nearly tenfold from 1970 to 2007, one has to keep in mind that the prices of other goods were also increasing, especially during the high inflation years of the 1970s. For reasons that will become clear later, we calculate the price of homes relative to the price of nondurable goods and services.[2] Relative prices of existing homes increased less than nominal prices, but even relative prices have almost doubled since 1970 and most of the price increase has taken place in the years since 1995. The relative price of homes peaked in 2006 after increasing by 50 percent in the 11 years since 1995. In contrast, this relative price increased by only 18 percent in the 25 years prior to 1995. One should note that even though the nominal price of existing homes never declined during this time period, the relative price of existing homes did decline in the early 1980s and 1990s.

The trend for the relative price of new single-family homes differs significantly from the relative price trend for existing homes. From 1970–2007, the relative price of new homes has increased by only one-third as much as the relative price of existing homes. Although new homes became relatively expensive in the late 1970s, their relative price then declined until the mid-1990s.

[1] All price indexes, for existing homes here and new homes below, are quality adjusted.

[2] The price index of nondurable goods and services is constructed using personal consumption expenditure data and excludes the service components related to housing. For a description of how the price index is constructed, see the Appendix.

Figure 1 Home Prices

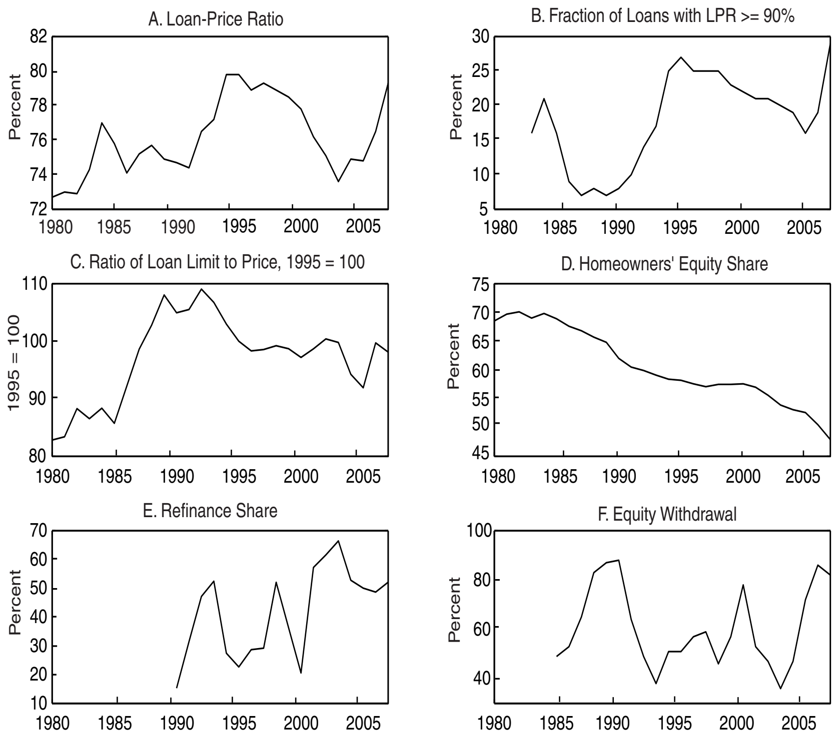


Notes: Except for the series "FHFA, Nominal," all prices are relative to a price index of nondurable consumption goods and services. All series are annual averages, except for 2008, which represents the average for the first three quarters. The series are described in the Appendix.

The price index of new single-family homes includes the value of the lot; thus, differences between the relative price of new and existing homes must be attributed to differences in the value of land embodied in the house price. In the National Income Account (NIA) measures of investment in residential structures, estimates of the value of land embodied in new single-family homes are removed from the new house price series. As we can see from Figure 1, the price index for single-family residential structures tracks the price index for new single-family homes quite closely. This suggests that the relative price of land used in the production of new homes has increased at about the same rate as has the price of residential structures. Finally, since there are persistent deviations of the price of new homes from the price of existing homes, we have to conclude that these two types of housing are imperfect substitutes.

The ability of owners to borrow against the collateral value of their house has increased over time. For example, there is some evidence that the average down payment on the purchase of a home declined significantly in the 1990s.

Figure 2 Mortgage Markets



Notes: Panels A and B graph the average loan-price ratio and the distribution of loan-price ratios for conventional loans used to purchase homes. Panel C graphs an index of the ratio of conforming loan limits to average house prices. Panel D graphs the homeowners' equity share from Flow of Funds data. Panel E graphs the share of refinance in total mortgage originations, and Panel F graphs the fraction of refinances that resulted in a 5 percent increase in the outstanding loan. Detailed descriptions of the series are provided in the Appendix.

The loan-price ratio for conventional mortgages used to purchase single-family homes increased from 75 percent to a peak of 80 percent in the mid-1990s (Figure 2, Panel A). Furthermore, the fraction of these conventional loans that had loan-price ratios in excess of 90 percent reached a peak of 25 percent in the mid-1990s (Figure 2, Panel B).

For the time period considered, the majority of mortgages originated are conforming; that is, they satisfy the underwriting guidelines of Fannie Mae and Freddie Mac and they do not exceed the loan limit imposed by either one. Fannie Mae and Freddie Mac purchase and securitize conforming mortgages. Up until September 2008, Fannie Mae and Freddie Mac were GSEs and mortgage market participants viewed them as being (implicitly) backed by the federal government.[3] Because of the implicit guarantee for GSE debt, the rates at which the two GSEs were able to borrow, and therefore the interest rates on conforming mortgages, tended to be low. Under these circumstances, homeowners can increase the loan share on which they pay relatively low interest rates; that is, they can lower the cost of a mortgage when the GSEs raise their loan limit relative to the average purchase price. Figure 2, Panel C plots the ratio of the loan limit imposed by Freddie Mac relative to the house price index for single-family homes purchased with conventional mortgages.[4] As we can see, the loan-limit to price ratio increased substantially in the late 1980s, and even today it is about 15 percent higher than in the 1980s.

A further sign that financial innovations made it easier for owners to borrow against the collateral value of their homes comes from the Flow of Funds data on homeowners' equity share in real estate. The homeowners' equity share declined from about 70 percent in 1980 to less than 50 percent in 2007 (Figure 2, Panel D). The fact that the decline in the homeowners' equity share is almost monotonic is a bit surprising since the evidence on down payment requirements for the purchase of homes suggests that these requirements started to increase again in the late 1990s. Yet, even though homeowners were apparently less able to borrow against the collateral of their house at the time of purchase, they were still able to extract some of the equity through refinancing their mortgages later on. With the exception of the mid-1990s and 2000, refinances constituted more than 40 percent of the total volume of mortgage originations (Figure 2, Panel E). In addition, more than 50 percent of all mortgage refinances resulted in a greater than 5 percent increase of the outstanding loan (Figure 2, Panel F).

Finally, the expansion of the market for subprime mortgages did introduce new population segments to the market for owner-occupied housing and made it possible for other homeowners to reduce their equity share substantially. It is, however, not straightforward to assess the quantitative importance of subprime mortgages since this market is less well-defined than the market for prime mortgages. Prime mortgages are essentially conforming mortgages and

[3] Weinberg and Walter (2002) discuss the possibility of implicit government guarantees on GSE debt. On September 7, 2008, Fannie Mae and Freddie Mac were taken over by the U.S. government, and it would appear that the guarantee on GSE debt was made explicit. The regulator of Fannie and Freddie, the Federal Housing Finance Agency (FHFA) has, however, stated that the guarantee is "effective," but not "explicit" (Natarajan 2008). Mortgage investors apparently also see a distinction between effective and explicit guarantees and, as of the end of November 2008, the interest rate spreads of GSE debt relative to comparable Treasury debt was 1.5 percentage points, about twice the spread before the takeover.

[4] Whereas the loan limit series is in current dollars, the home price series is an index normalized to 100 in 1987. Therefore, we renormalized the ratio to 100 in 1995.

jumbo mortgages, that is, mortgages that exceed the loan limit imposed by the two GSEs for borrowers with good credit histories. Subprime mortgages, according to most definitions, involve borrowers with impaired credit histories, which is reflected in low credit ratings. Subprime mortgages also tend to involve high loan-to-value ratios. Occasionally, subprime mortgages are grouped together with Alt-A mortgages. Unlike subprime mortgages, Alt-A mortgages are taken out by borrowers with good credit history, but the mortgage may involve a loan-to-value ratio that is too high or documentation that is insufficient for the mortgage to conform to the GSE standards.

Even though subprime mortgages lie at the heart of the financial market disruptions of the last year, they became a quantitatively important part of the mortgage market only after 2000, long after house prices started to increase. Mayer and Pence (2008) suggest that the share of subprime mortgages in the total number of all originated mortgages increased from less than 10 percent before 2000 to more than 20 percent after 2000.[5] Furthermore, Mayer and Pence (2008) argue that subprime originations were predominantly cash-out refinances. Gorton (2008) provides some statistics on the value of subprime and Alt-A mortgages, both in originations and in total outstanding volume. According to Gorton (2008, Table 3), the share of subprime mortgages in the total value of originations increased from 8 percent in 2000 to about 20 percent in 2004–2006. Consequently, the share of subprime mortgages in the total value of outstanding mortgages increased from 3 percent in 2000 to more than 10 percent in 2004–2006 (Gorton 2008, Table 2).

2. A SIMPLE MODEL OF HOUSING

We describe a simple general equilibrium model of the demand for housing where the price of housing is endogenous. A representative consumer has preferences over the consumption of nondurable goods and housing services. Housing services are proportional to the stock of housing. New housing is produced by combining new residential structures, structures for short, with land. New structures, together with nondurable consumption goods, are produced from aggregate output. The rate of transformation between nondurable consumption goods and structures is exogenous and determines the relative prices of structures. In this environment the relative price of housing depends on the supply of land and the relative price of structures.

We are interested in the model's ability to account for sustained house price increases such as those displayed in Figure 1. We will, therefore, study the model's balanced growth path, which reflects its long-run growth rates.

[5] Mayer and Pence (2008) discuss different definitions of subprime mortgages and their most preferred measure is based on the subprime lender list maintained by the U.S. Department of Housing and Urban Development.

The Environment

Time is continuous and the horizon is infinite. A representative agent derives utility from the consumption of a nondurable good, c_0, and the consumption of housing services, h_0. The agent's preferences are

$$\int_0^{\infty} e^{-\rho_0 t} \{\theta \ln h_0 (t) + (1-\theta) \ln c_0 (t)\} \, dt, \tag{1}$$

with time preference rate $\rho_0 > 0$ and $0 < \theta < 1$. The consumption of housing services is proportional to the stock of housing units owned by the agent. In this article, we will use the terms "housing services" and "housing stock" interchangeably.

The agent receives an exogenous endowment stream of an homogeneous good. The value of the endowment in terms of the nondurable consumption good is y_0. We express all prices in terms of the nondurable consumption good. The agent also receives l_0 units of new land and the price of new land is p_l. The agent can use his income for consumption, the purchase of new housing units, x_{h0}, at the relative price, p_h, or he can save it at an interest rate, r. The flow budget constraint of the household is

$$\dot{a}_0 (t) + c_0 (t) + p_h (t) \, x_{h0} (t) = y_0 (t) + p_l (t) \, l_0 (t) + r (t) \, a_0 (t) \,, \tag{2}$$

where a_0 is the agent's net financial wealth.[6] Housing depreciates at rate $\delta > 0$ and the stock of housing accumulates according to

$$\dot{h}_0 (t) = x_{h0} (t) - \delta h_0 (t) \,. \tag{3}$$

The homogenous good, y, can be used to produce the nondurable consumption good or it can be used to produce structures, x_s. The rate of transformation between nondurable consumption goods and structures is exogenous and the relative price of structures, p_s, is the inverse of the relative productivity of the structures sector. The aggregate resource constraint for nondurable consumption and structures is

$$c (t) + p_s (t) \, x_s (t) = y (t) \,. \tag{4}$$

Structures are combined with new land to produce new housing units using a Cobb-Douglas technology

$$x_h (t) = x_s (t)^{\beta} \, l (t)^{1-\beta} \,, \tag{5}$$

with $0 \leq \beta \leq 1$. The production of all goods is competitive.

The representative agent owns all of the endowment of land and the homogeneous output good. Market clearing for land, the output good, the nondurable consumption good, new housing structures, and the credit market

---

[6] The notation $\dot{z} (t) = \partial z (t) / \partial t$ denotes the time derivative of the variable, z, as a function of time, t.

imply (4), (5), and

$$l(t) = l_0(t), \; y(t) = y_0(t), \; c(t) = c_0(t), \; x_h(t) = x_{h0}(t), \; 0 = a_0(t). \tag{6}$$

We assume that the economy is growing over time. In particular, the endowments y and l and the relative price p_s all grow at constant rates γ_y, γ_l, and γ_s:

$$y(t) = \bar{y}e^{\gamma_y t}, \; l(t) = \bar{l}e^{\gamma_l t}, \text{ and } p_s(t) = \bar{p}_s e^{\gamma_s t}. \tag{7}$$

Before we proceed, some remarks on the properties of this environment are in order. First, new and existing housing are perfect substitutes in consumption. Therefore, new homes sell at the same price as do old homes, and this model cannot address the fact that the price of existing homes has been increasing at a faster rate than the price of new homes. Second, there is no meaningful distinction between renting housing or owning the housing stock. In other words, this model can be interpreted as one of owner-occupied housing, as is done here, or it can be interpreted as a model of rental housing.[7] Finally, this model entails some peculiar assumptions concerning the supply and use of land. The supply of new land used in the production of new homes is exogenous, and once land is embedded in new homes it depreciates at the same rate as do structures. In other words, once the structures of a house have depreciated, the plot cannot be reused for another house. The total stock of land then grows at the same rate as does the stock of new land.

Optimal Consumption and Production on the Balanced Growth Path

Hornstein (2008) provides a complete analysis of the optimization problem of the representative agent and the representative producer of new homes. We now summarize this analysis; we will drop the time index when not needed.

Optimal consumption of housing and nondurable consumption goods is such that the marginal rate of substitution between the two commodities is equated with their relative price,

$$\frac{\theta / h_0}{(1-\theta)/c_0} = \left(r + \delta - \hat{p}_h\right) p_h. \tag{8}$$

Here the price of nondurable goods is normalized at one and the price of housing services is equal to the user cost of housing, that is, the implicit rental rate paid for the use of the housing stock. This rental rate is the required return on the housing asset plus depreciation minus capital gains due to the

[7] Alternatively, one could assume that renting a home simply yields less utility than owning a home. Together with assumptions on financial frictions, this can generate a well-defined demand for rental and owner-occupied housing, e.g., Kiyotaki, Michaelides, and Nikolov (2007).

changes in the capital value of the housing stock.[8] The optimal allocation of consumption over time is determined by a standard Euler equation,

$$r = \rho_0 + \hat{c}_0. \tag{9}$$

Competitive production of new housing implies that for the two inputs, structures and new land, the value of an input's marginal product is equal to the price of the input,

$$p_h \beta \frac{x_h}{x_s} = p_s \text{ and} \tag{10}$$

$$p_h (1 - \beta) \frac{x_h}{l} = p_l. \tag{11}$$

On a balanced growth path (BGP), all variables grow at constant, but potentially different, rates. The resource constraint for the output good (4) implies that the BGP nondurable consumption and the value of structures grow at the same rate as does output,

$$\hat{c} = \gamma_s + \hat{x}_s = \gamma_y. \tag{12}$$

The production function for new housing, equation (5), implies that investment in new housing grows at a rate that is a weighted average of the growth rates of new structures and land,

$$\hat{x}_h = \beta \hat{x}_s + (1 - \beta) \gamma_l. \tag{13}$$

The market clearing conditions (6) imply that the representative household's choice variables grow at the same rates as the corresponding aggregate variables,

$$\hat{c}_0 = \hat{y}_0 = \gamma_y \text{ and } \hat{x}_{h0} = \hat{x}_h. \tag{14}$$

The accumulation equation for the housing stock, (3), implies that the stock of housing grows at the same rate as does investment in new housing,

$$\hat{h} = \hat{x}_h \text{ and } \frac{x_h}{h} = \hat{h} + \delta. \tag{15}$$

Finally, the growth rates for the price of new housing and land are determined by the first-order conditions for optimal input use in the production of new housing, equations (10) and (11),

$$\hat{p}_h + \hat{x}_h = \hat{p}_l + \gamma_l = \gamma_s + \hat{x}_s = \gamma_y. \tag{16}$$

We can now express the growth rates for the housing stock and the relative price of housing on the BGP as functions of the exogenous growth rates of output, the relative price of structures, and the supply of new land. The impact of a higher output growth rate on the rate at which relative house prices increase is immediate. Combining expressions (13) and (15) yields the rate at which

[8] The growth rate of a generic variable, z, is denoted $\hat{z}(t) = \dot{z}(t) / z(t)$.

the housing stock changes, and combining expressions (13) and (16) yields the rate at which the relative price of housing changes,

$$\hat{h} = \beta\left(\gamma_y - \gamma_s\right) + (1-\beta)\,\gamma_l \text{ and} \tag{17}$$

$$\hat{p}_h = \beta\gamma_s + (1-\beta)\left(\gamma_y - \gamma_l\right). \tag{18}$$

On the one hand, a higher growth rate of aggregate output increases both the rate of house price appreciation and the rate of housing stock accumulation. On the other hand, if the relative price of structures increases at a faster rate, or the rate at which new land becomes available declines, the relative price of housing increases at a faster rate but the housing stock is accumulated at a slower rate.

The impact of changes in the exogenous growth rates on the house price appreciation rate depends on the share of land in the production of homes. If land is not an input to the production of homes, that is, $\beta = 1$, then home production is proportional to the use of structures. Thus, house price appreciation is determined by the rate at which the relative price of structures changes and is independent of output growth and the availability of new land. Otherwise, if new homes are in fixed supply, that is, $\beta = 0$, then house price appreciation depends on the difference between the output growth rate and the land supply growth rate.

We normalize all variables such that they remain constant on the BGP. If the variable z grows at the rate $\hat{z}$ on the BGP, we define its normalized value as

$$\tilde{z}(t) = z(t)\,e^{-\hat{z}t}. \tag{19}$$

Essentially, the normalized value of a variable represents the level of its growth path. By construction the normalized variables do not change on the BGP, that is, $\dot{\tilde{z}} = 0$. In Hornstein (2008) we derive the solutions for the normalized levels of the BGP.

Quantitative Implications

What are the quantitative implications of our simple model for the rate at which house prices change over time? In particular, can the model account for the apparent increase of the house price appreciation rate after 1995? To answer this question we first calibrate the model by choosing parameter values to match certain statistics of the U.S. economy for the pre-1995 period. We then ask if changes in output growth rates or the rate at which the relative price of residential structures appreciate can account for the changes in house price appreciation rates.

We consider the U.S. economy from 1975 to 2007. For 20 years (1975–1995), average per capita GDP growth and average per household GDP growth were about 1 percent-per-year (see Table 1). Since the focus of analysis

Table 1 House Prices, Output, and Residential Investment: 1975–2007

	1976–1995	1996–2007
Prices		
OFHEO Home Price Index	0.6	3.5
New Single-Family Homes, Incl. Lot	0.4	1.4
Single-Family Residential Structures	0.1	1.4
Quantities, Per Capita		
Output	1.3	1.3
Single-Family Residential Structures	4.0	2.7
Quantities, Per Household		
Output	1.0	1.2
Single-Family Residential Structures	4.3	2.4

Notes: All prices are relative to the personal consumption expenditures (PCE) price index for nondurable goods and services (excluding housing services), and all quantities are nominal values deflated with the PCE price index for nondurable goods and services (excluding housing services). Detailed descriptions of the series are in the Appendix.

is residential housing, normalizing output per household seems to be more appropriate than the more standard per capita normalization and we choose $\gamma_y = 0.01$. For the same time period, the relative price of residential structures first increased and then declined such that the average annual appreciation rate from 1975 to 1995 was close to zero, $\gamma_s = 0$.

We calculate the housing accumulation rate based on the BGP equation (17) and, therefore, need a value for the share of land in the production of new homes and the rate at which new land becomes available. For new homes sold, the Bureau of Economic Analysis assumes a land value share of about 11 percent when it constructs the residential structures price index from the price index for new homes sold, where the latter includes the value of the lot (Davis and Heathcote 2007, 2602). This suggests $\beta = 0.9$. For the time period 1975–1995, this price index for new homes increased at an annual rate of about half a percentage point (Table 1). Davis and Heathcote (2007) also calculate an overall share of land in all home values, existing and new, that fluctuates between 30 and 45 percent. We, therefore, study the two extreme cases, $\beta = 0.9$ and $\beta = 0.5$.

The evidence on the rate at which new land becomes available is mixed at best. As part of their calculation of the value share of land in overall housing, Davis and Heathcote (2007) derive constant quality quantity indexes for residential land use. For the time period 1975–2006, their index of residential land use increases steadily at an average rate of 0.7 percent-per-year. At the same time, the number of households increased by 1.5 percent-per-year. Thus, according to Davis and Heathcote (2007), constant quality land use per household declined at an average annual rate of 0.8 percent-per-year. Overman, Puga, and Turner (2007) calculate the change in actual residential land use in the continental United States from 1976–1992 based on satellite survey data. They find that actual residential land use increased at an annual rate of 2.4 percent-per-year. Accounting for different population growth and land use patterns across states, they estimate that the average land use per household increased by about 0.7 percent-per-year during this period. Overman, Puga, and Turner (2007) do not account for the quality of the residential land used, but, for their estimate of land use to be consistent with Davis and Heathcote's (2007) estimate, one would have to assume that the average quality of land declined at a rate of 1.5 percent-per-year.[9] This seems unlikely. We do not take a stand on land use and simply set the growth rate to zero for the analysis, $\gamma_l = 0$, and assume that the rate at which land has become available has not changed over time.

We assume an equilibrium real interest rate of 4 percent, which is standard in the literature. Equation (9) then implies the household's time discount factor, ρ_0. The Bureau of Economic Analysis (2004) reports depreciation rates between 1.1 and 3.6 percent for one- to four-unit residential structures and we chose a 1.5 percent depreciation rate. We determine the utility coefficient on housing services, θ, based on the share of nondurable consumption expenditures on the BGP. Since we do not model the consumption of durable goods services, it is not possible to construct a model-consistent measure of the share of nondurable goods. We, therefore, consider two alternative measures. First, we calculate the average share of nondurable consumption goods and services in total personal consumption expenditure plus expenditures on residential structures.[10] From 1975–1995, this expenditure share was about 80 percent and fluctuated between 76 and 82 percent. This measure probably understates the expenditure share of nondurable goods since its measure of residential structures includes multifamily units and we have included the purchase of durable consumption goods. Alternatively, we calculate the expenditure share of nondurable goods and services when total expenditures include only housing services next to the expenditures on nondurable goods and services. The latter share fluctuates between 82 and 84 percent between 1975 and 1995. Combining the two measures, we match an 80 percent expenditure share for nondurable goods and obtain the utility coefficient on housing services, $\theta = 0.556$. The parameter values are summarized in Table 2.

We find only one noticeable change in the driving forces of house price appreciation after 1995, namely a faster appreciation of the relative price of residential structures. As we can see from Table 1, whereas the appreciation rate of the relative price of residential investment increased by one percentage

[9] Davis and Heathcote's (2007) estimate for land use is quite smooth and very similar average growth rates apply for subsamples, in particular for the time period 1976–1992.

[10] We exclude housing-related services from the service component of personal consumption expenditures.

Table 2 Model Calibration

		Baseline Model			
$\rho_0 = 0.03$	$\theta = 0.556$	$\delta = 0.015$	$\beta = 0.5$	$\gamma_y = 0.01$	$\gamma_s = \gamma_l = 0$
		With Collateral Constraints			
$\rho_1 - \rho_0 = 0.02$	$\pi = 0.175$	$\phi = 0.0385$	$\alpha = 0.3$		

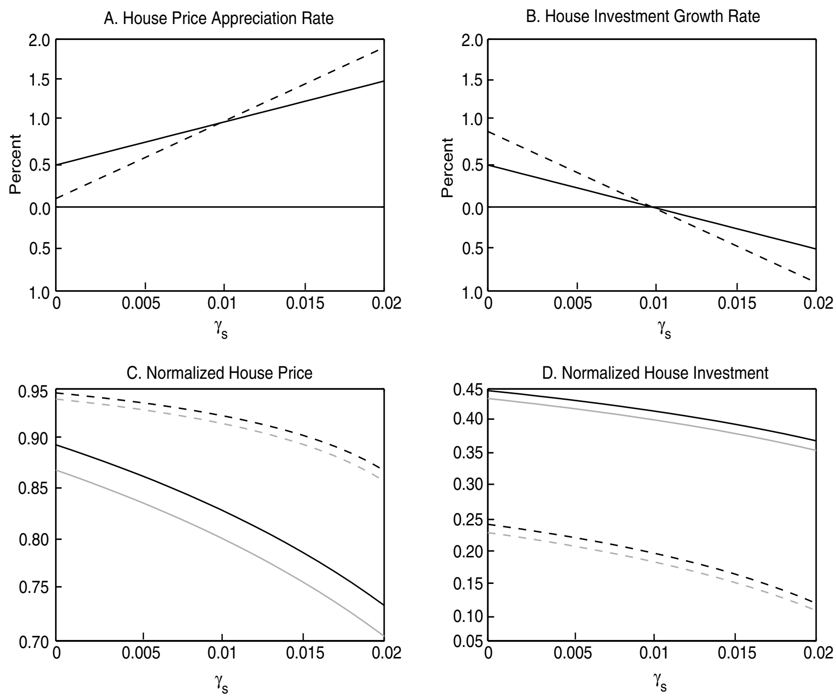
point after 1995, there was no corresponding significant increase of the output growth rate, either per capita or per household.

In Figure 3 we plot the growth rate and normalized level of the house price and investment as a function of the appreciation rate of the relative price of structures. The black lines denote this relation for the economy described in this section, and the gray lines denote the relation for the economy with collateral constraints, to be described in the next section. The solid (dashed) lines denote the economy with a large (small) new land value share in production, $\beta = 0.5$ ($\beta = 0.9$). We can see that the higher post-1995 new house price appreciation rate and the lower growth rate for residential structures is consistent with the higher price appreciation rate for residential structures (Table 1). While the change of the price appreciation rate for new homes fits qualitatively and quantitatively, the model does not capture the change of the growth rate for investment in residential structures quantitatively. Even for the period before 1995, the housing accumulation rate is predicted to be less than 1 percent, independent of the share parameter, β. This prediction is substantially below the observed 4 percent growth rate for residential investment (Table 1).

The comparison of BGP characteristics for different rates of relative price changes for residential investment probably overstates the model's ability to capture changes in the new house price appreciation rate. As we can see from Figure 3, Panel C, a higher appreciation rate of the price for structures not only increases the house price appreciation rate, but it also permanently lowers the price and investment path for homes. It is, thus, quite possible that for some time the transition to this new lower level of the BGP exerts a negative impact on the growth rates of prices and investment in new homes. This appears to be more of an issue when the contribution of new land to the production of new homes is large, since the normalized levels of new house prices and investment are more sensitive to parameter changes when the land value parameter is large.

Finally, note that the model does not make a distinction between new and existing homes. The model, therefore, does not capture the much faster price appreciation rate for existing homes after 1995. We now introduce financial frictions into the model and ask if innovations that eliminate some of these financial frictions can account for changes in house price appreciation rates.

Figure 3 The Impact of Price Appreciation for Residential Structures



Notes: Black (gray) lines refer to the model without (with) collateral constraints. Solid (dashed) lines refer to model calibrations with a large (small) share of land in the production of homes, $\beta = 0.5$ ($\beta = 0.1$).

3. FINANCIAL CONSTRAINTS AND THE DEMAND FOR HOUSING

We modify the simple general equilibrium model of the previous section and introduce a second consumer that is more impatient than the consumer studied above. At the equilibrium interest rate, the impatient agent will borrow from the patient agent. In fact, the impatient agent would like to borrow unlimited amounts. We, therefore, impose a borrowing constraint on the impatient agent that states that total borrowings are constrained by the collateral value of the agent's housing stock. We study how changes in the collateral constraint affect the equilibrium relative price of housing. Henceforth, we will distinguish between the lender, type 0 agent, and the borrower, type 1 agent.

Collateral Constraints for Housing

The borrower and lender have the same preferences with respect to the consumption of housing services and nondurable goods, (1), but the impatient borrower discounts future utility at a higher rate than the patient lender, $\rho_1 > \rho_0$. The amount of credit that the borrower can obtain is limited by the collateral value of the housing stock he owns. We assume that the required equity share of a borrower for a home of vintage τ is

$$\omega(\tau) = 1 - (1 - \pi) e^{-(\phi - \delta)\tau}, \tag{20}$$

with $\phi \geq \delta$. The down payment requirement for the purchase of new housing is $\omega(0) = \pi \in [0, 1]$. The required equity share remains constant if $\phi = \delta$, and increases with the age of the vintage to one if $\phi > \delta$.[11] The collateral constraint states that the household can borrow against the value of its undepreciated housing stock; that is, he can have negative financial net-wealth a_1, but the household has to retain a total equity position,

$$p_h(t) \int_0^\infty \omega(\tau) \left[e^{-\delta\tau} x_{h1}(t - \tau) \right] d\tau \leq p_h(t) h_1(t) + a_1(t). \tag{21}$$

Using the definition of the vintage-specific equity requirement, (20), the collateral constraint simplifies to

$$(1 - \pi) p_h(t) q_1(t) \geq -a_1(t), \tag{22}$$

where q_1 represents the part of the housing stock against which the household can borrow after a minimum down payment has been made. This collateralizable housing stock evolves according to

$$\dot{q}_1(t) = x_{h1}(t) - \phi q_1(t). \tag{23}$$

Thus, new purchases add to the collateralizable housing stock, but their use as collateral "depreciates" at rate ϕ rather than at rate δ, as does the physical housing stock. We refer to the collateralizable housing stock as the "collateral stock."

The borrower is assumed to maximize utility subject to a budget constraint and accumulation equation for the housing stock, analogous to equations (2) and (3). In addition, the borrower's choices have to satisfy the collateral constraint, (22), and the accumulation equation for the collateral stock, (23). Given these additional constraints, the capital value of a unit of housing stock for a borrower has to be adjusted for its contribution to the collateral stock. The marginal value of a unit of housing in terms of the nondurable consumption

[11] When the required equity share is increasing with the age of the housing vintage, a borrower would like to own only the newest vintage since he wants to borrow as much as possible against the collateral value of his housing stock. To prevent this outcome we assume that the borrower cannot continuously turn over his housing stock but has to hold on to vintages purchased in the past.

good becomes

$$\frac{\mu_1}{\lambda_1} = p_h - \frac{\varphi_1}{\lambda_1}, \tag{24}$$

where μ_1 is the marginal value of a unit of housing in utility terms, λ_1 is the marginal utility of income, and φ_1 is the marginal value of an additional unit of collateral. Analogous to the lender's consumption of housing and nondurable consumption goods, the borrower's optimal choice again equates the marginal rate of substitution between the two commodities with their relative price,

$$\frac{\theta / h_1}{(1-\theta)/c_1} = \left(\rho_1 + \delta - \hat{\mu}_1\right)\left(p_h - \frac{\varphi_1}{\lambda_1}\right). \tag{25}$$

Because the housing stock not only provides direct consumption services but also collateral services, the borrower's effective price of a unit of the housing stock is reduced and this lowers the user cost of housing.

We now assume that the representative borrower interacts with the representative lender from Section 2 in a competitive equilibrium. Production of nondurable consumption goods, structures, and new homes continues to be determined by equations (4) and (5). We assume that the lender receives a fraction, α, of the endowment of the output good and the remainder goes to the borrower,

$$y_0(t) = \alpha y(t) \text{ and } y_1(t) = (1-\alpha) y(t). \tag{26}$$

We also continue to assume that the lender receives all of the endowment of new land. Market clearing for the nondurable consumption goods, new housing, and the credit market now imply that

$$c(t) = c_0(t) + c_1(t),\ x_h(t) = x_{h0}(t) + x_{h1}(t),\ 0 = a_0(t) + a_1(t). \tag{27}$$

The growth rates of aggregate variables on the BGP are determined as before by equations (17) and (18) since the aggregate resource constraints have not changed. From the definition of market clearing, (27), it follows that, on the BGP, consumption of nondurable goods and housing, wealth, etc., for borrowers and lenders grows at the same rates

$$\hat{c}_i = \hat{a}_i = \gamma_y \text{ and } \hat{x}_{hi} = \hat{h}_i = \hat{q}_1 = \hat{h}, \text{ for } i = 0, 1, \tag{28}$$

and we normalize all variables as described by equation (19).

The interest rate on the BGP continues to be determined by the lender's time discount rate and the output growth rate (Equation [9]). One can show that on the BGP the collateral constraint is binding for the borrower since the borrower's marginal utility of wealth is positive and he is more impatient than the lender. Detailed derivations are in Hornstein (2008).

Quantitative Implications

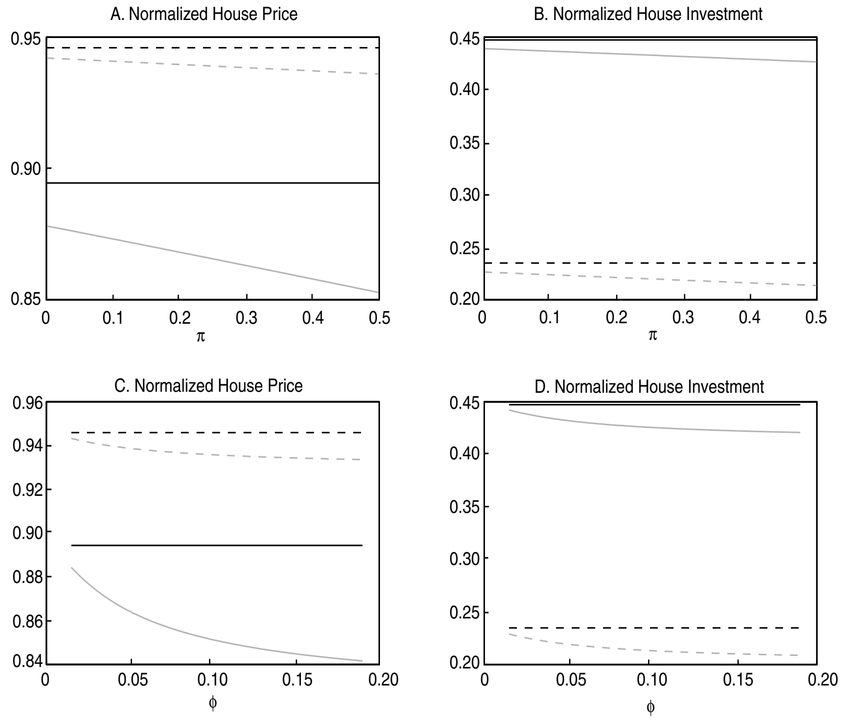
Collateral constraints have only a limited impact on the equilibrium allocations and prices of the economy's BGP. The first thing to note is that collateral constraints cannot affect the growth rates on the BGP since the growth rates are determined by the aggregate resource constraints that are not affected by the presence of collateral constrained agents. This means that collateral constraints can only affect the levels of the BGP. We now show that the impact of collateral constraints on these growth path levels is quantitatively limited. This also means that collateral constraints are unlikely to have a great impact on the transition to a new BGP.

Our model of collateral constraints is based on Campbell and Hercowitz (2006) and we follow their parameterization closely. The impatient borrower's time discount rate is set two percentage points higher than the lender's time discount rate, $\rho_1 = \rho_0 + 0.02$. In their analysis, Campbell and Hercowitz (2006) take a broad view of the role of collateral constraints and they model them as applying to the purchase not only of homes, but also of durable goods. Our view is more narrowly focused on the home mortgage market and we, therefore, only use their estimates of the down payment parameter and the equity accumulation rate as it applies to home mortgages. Hercowitz and Campbell (2006) argue that, for the time period before 1982, collateral constraints for homes are best represented by a down payment parameter, $\pi = 0.23$, and an equity accumulation rate, $\phi = 0.052$. The latter reflects an average term to maturity for mortgages of about 20 years.

Campbell and Hercowitz (2006) argue that post-1982 initial down payments declined by six percentage points and the average term to maturity increased by six years. Their collateral constraint parameters for the post-1982 period are $\pi = 0.175$ and $\phi = 0.0385$. Campbell and Hercowitz (2006) set the break point for changes in the collateral constraints in the mid-1980s because they want to argue that weaker collateral reduced the aggregate labor supply elasticity and thereby contributed to the "Great Moderation" in the mid-1980s. Our focus is on the housing market and we want to account for the increased rate of house price appreciation since the mid-1990s. In Section 1 we argued that financial innovations most likely loosened collateral constraints further during the post-1995 period. Therefore, we study the impact of even bigger reductions of the down payment requirement and bigger increases of the duration to maturity than considered by Campbell and Hercowitz (2006).

Campbell and Hercowitz (2006) allocate about one-third of the output endowment to lenders and two-thirds to borrowers, $\alpha = 0.3$. Underlying this distribution of the endowment are the assumptions that lenders own all the capital and borrowers own all the labor in the economy. If we were to assume that the output good is produced using capital and labor as inputs to a constant-returns-to-scale production function and we were to allow for capital accumulation, then the first assumption is an equilibrium outcome since only

Figure 4 The Impact of Collateral Constraints



Notes: Panels A and B display the response of normalized house prices and investment on the balanced growth path to changes in the down payment rate, π. Panels C and D display the response of normalized house prices and investment on the balanced growth path to changes in the equity accumulation rate, ϕ. Otherwise, see notes to Figure 3.

the patient lenders will own capital. If only borrowers supply labor, then their share of the output good is the labor income share. In the U.S. economy, the labor income share is about two thirds and the capital income share is one third. The calibration of the housing coefficient in the agents' utility functions is not affected by the collateral constraints.

In Figure 4 we plot how normalized house prices and investment, that is, the growth path levels, relate to the collateral constraint parameters. Gray lines denote the economy with collateral constraints and black lines denote the relation for the corresponding economy without collateral constraints.[12]

[12] Obviously, house prices and investment in the economy without collateral constraints do not respond to changes in the parameters, π and ϕ.

Lowering down payment requirements and the equity accumulation rate increases house prices and investment, but the effects are quantitatively small. We obtain the biggest effect on house prices and investment when the share of land in production is largest, $\beta = 0.5$. But even in this case, either completely eliminating down payment requirements or reducing equity accumulation rates to their lower bound does not increase house prices permanently by more than about 7 percent.

Returning to Figure 3, we see that the presence of collateral constraints does not affect much the impact of changes in the appreciation rate of the price of structures. With or without collateral constraints, normalized house price and investment levels decline with a faster rate of price appreciation. House prices and investments respond a bit more in the economy with collateral constraints, but the difference is marginal at best.

4. CONCLUSION

We have argued that models of the aggregate housing market, such as Davis and Heathcote (2005), may be able to account for the trend of new house prices, but these models cannot account for the differential price trends in the market for existing homes. Furthermore, including an explicit model of the mortgage market apparently does not improve the model's ability to match house price trends. One might argue that the model is too stylized for it to be able to account for sustained increases in house prices, but two more elaborate versions of the basic framework have not been more successful.

Iacoviello and Neri (2008) use the same basic model of housing but add a more elaborate production structure with capital accumulation, and they add other nominal and real rigidities to the model. They are mainly interested in the cyclical implications of collateral constraints and their simulation studies indicate that collateral constraints may play some limited role for the cyclical behavior of nondurable consumption. Even though their model's production structure is quite complicated, it shares with our baseline model the feature that growth rates on the BGP are independent of collateral constraints.

Kiyotaki, Michaelides, and Nikolov (2007) provide a more detailed representation of the life-cycle aspects of housing consumption in a heterogeneous agent economy with collateral constraints. They find that even though changes in collateral constraints have a significant distributional impact in the sense that they affect the choices between owning and renting homes, these changes have only a minor impact on house prices. Kiyotaki, Michaelides, and Nikolov (2007) do find that permanently higher labor productivity growth rates can significantly increase house prices, but this feature seems to be independent of the presence of collateral constraints.

Overall, it appears that the long-run growth properties of any model that is consistent with a balanced growth path, in particular the rates of house price appreciation, are likely to be determined by the basic supply and demand structure of the housing market and not by collateral constraints. Furthermore, given the persistent differences between the prices for new and existing homes, these two types of housing clearly represent imperfect substitutes. The first step toward improving our understanding of the housing market is then to develop a model that distinguishes between the market for new and existing homes. One possibility is to incorporate the recent externality-based theory of city structures, e.g., Lucas (2001), into models of the aggregate economy. This theory predicts land and house price gradients; that is, homes in different locations are imperfect substitutes. Conditional on a criterion that distinguishes between existing and new homes, one could work out the theory's implications for the determinants of the relative price of existing and new homes.

APPENDIX

The Office of Federal Housing Enterprise Oversight (OFHEO) publishes a house price index based on repeat sales transactions for single-family homes that are financed with mortgages that are conforming and conventional. The price index measures the average price change involved in the sale or refinancing of properties for which price data on previous transactions are available. The repeat sales feature of the price index is supposed to purge quality change from the measured price change. Mortgages are called *conforming* if they do not exceed a loan limit and they satisfy the underwriting guidelines of the two government sponsored agencies that purchase and securitize mortgages, the Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac). Mortgages are called *conventional* if they are neither insured nor guaranteed by the Federal Housing Administration, the Veterans Administration, or other federal government entities. Thus, conforming mortgages are prime mortgages while conventional mortgages can include both prime and subprime mortgages. OFHEO publishes a price index that involves actual transactions prices (purchases) and assessments (refinancing) since 1975. OFHEO also publishes a purchase-only price index since 1991. The Haver mnemonics for the comprehensive house price index is USHPI@USECON and for the purchase only price index it is USPHPI@USECON. This price index used to be known as the OFHEO house price index, but with the October 2008 merger of OFHEO into the new Federal Housing Finance Agency (FHFA), it is now referred to as the FHFA house price index.

We consider two other housing-related price series. First, the Census Bureau's price index for new single-family homes sold (HPDEX@USECON). Second, the price index for single-family structures from the national income accounts (JAFRSH1A@USNA). Whereas the first price index includes the value of the lot, the second price index applies only to new structures. Both series are constant quality price indexes.

We construct a price index for nondurable consumption goods and services, excluding housing services, from the NIA's data on PCEs. The growth rate of this price index is a Divisia index, that is, a weighted average of the components' quantity index growth rates, where the weights are the nominal expenditure shares of the components. The Haver mnemonics for the series involved are CNA@USNA, CSA@USNA, and CSRA@USNA for the nominal series, and CNHA@USNA, CSHA@USNA, and CSRHA@USNA for the chained 2000 dollar series.

The Federal Housing Finance Board publishes terms for conventional mortgages used to purchase single-family homes. For Figure 2, Panels A and B, we use the annual time series for loan-to-price ratios (FCMR@USECON) and the fraction of loans with loan-to-price ratios above 90 percent (FCMR4@USECON). These series represent national averages of major lenders and they include fixed rate and adjustable rate mortgages, but they exclude refinances. The alternative measure on down payment requirements for conventional mortgages in Figure 2, Panel C, is calculated as the ratio of the Fannie Mae conventional loan limit for a first mortgage on a single-family home (FCL1@USECON) to the average price of a single-family home financed with a conventional mortgage (USCMPHP1@USECON). The latter is also a repeat sales price index published by Freddie Mac.

From the Federal Reserve Board's Flow of Funds data, Balance Sheets of Households and Nonprofit Organizations, Table B.100, we obtain homeowners' equity as the market value of household real estate less the value of outstanding mortgages. The homeowners' equity share (PL15HOM5@FFUNDS) in Figure 2, Panel D, is then the share of homeowners' equity in the market value of household real estate.

The Mortgage Bankers Association provides data on the composition of mortgage originations, whether they are used for the purchase of homes (HMTOP@USECON) or to refinance an existing mortgage (HMTOR@USECON). Figure 2, Panel E plots the value share of refinance originations and Figure 2, Panel F plots the fraction of refinances that resulted in at least a five percentage point higher loan amount (HRFHA@USECON).

For Table 1, we use the nominal value of single-family residential structures investment (FRSH1A@USNA) and the nominal value of GDP (GDPA@USNA) for output. Both series are deflated by the above-described price index for nondurable consumption goods and services, excluding housing services. We then calculate per capita series using the U.S. resident population 16 years and older (POP16O@USECON) and per household series using the number of U.S. households (POPH@USECON).

REFERENCES

Bureau of Economic Analysis. 2004. "BEA Depreciation Estimates." http://www.bea.gov/national/FA2004/Tablecandtext.pdf.

Campbell, Jeffrey R., and Zvi Hercowitz. 2006. "The Role of Collateralized Household Debt in Macroeconomic Stabilization." Federal Reserve Bank of Chicago Working Paper 2004-24.

Davis, Morris A., and Jonathan Heathcote. 2005. "Housing and the Business Cycle." *International Economic Review* 46 (August): 751–84.

Davis, Morris A., and Jonathan Heathcote. 2007. "The Price and Quantity of

Residential Land in the United States." *Journal of Monetary Economics* 54 (November): 2,595–620.

Gorton, Gary B. 2008. "The Subprime Panic." Working Paper 14398. Cambridge, Mass.: National Bureau of Economic Research (September).

Hornstein, Andreas. 2008. "Notes on a Model of Housing with Collateral Constraints." Federal Reserve Bank of Richmond Working Paper.

Iacoviello, Matteo, and Stefano Neri. 2008. "Housing Market Spillovers: Evidence from an Estimated DSGE Model." Mimeo, Boston College.

Kiyotaki, Nobuhiro, Alexander Michaelides, and Kalin Nikolov. 2007. "Winners and Losers in Housing Markets." Mimeo, London School of Economics.

Lucas, Robert E. 2001. "Externalities and Cities." *Review of Economic Dynamics* 4 (April): 245–74.

Mayer, Christopher J., and Karen Pence. 2008. "Subprime Mortgages: What, Where, and to Whom." Federal Reserve Board, Finance and Economics Discussion Series 2008-29.

Natarajan, Prabha. "Words Vex Mortgage Investors." *Wall Street Journal*. 24 October, 2008.

Walter, John R., and John A. Weinberg. 2002. "How Large is the Federal Financial Safety Net?" *Cato Journal* 21 (Winter): 369–93.

Economic Quarterly—Volume 95, Number 1—Winter 2009—Pages 25–51

Indeterminacy from Inflation Forecast Targeting: Problem or Pseudo-Problem?

Bennett T. McCallum

Monetary economists have been rather proud about developments in their subject over the past two decades. There has been great progress in formal analysis and also in the actual conduct of monetary policy. Analytically, the profession has developed an approach to policy analysis that centers around a somewhat standardized dynamic model framework that is designed to be structural—respectful of both theory and evidence—and therefore usable in principle for policy analysis. This framework includes a policy instrument that agrees with the one typically used in practice and, in fact, models of this type are being used (in similar ways) by economists in both academia and in central banks, where several economic researchers have gained leading policymaking positions. Meanwhile, in terms of practice, most central banks have been much more successful than in previous decades in keeping inflation low while avoiding major recessions (with a few exceptions) prior to 2008. Furthermore, these improvements have been interrelated: The "inflation targeting" style of policy practice that has been adopted by numerous important central banks—and that arguably has been practiced unofficially by the Federal Reserve[1]—is strongly related in principle to the prevailing framework for analysis. For a recent exposition discussing

The author is affiliated with Carnegie Mellon University and NBER and is a visiting scholar at the Federal Reserve Bank of Richmond. For many helpful discussions, the author is indebted to Marvin Goodfriend and Edward Nelson. In addition, useful referee reports were provided by Thomas Lubik, John Walter, and Sam Henly. Opinions expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: bmccallum@cmu.edu.

[1] For an argument to this effect, see Goodfriend (2005).

this development, by an author who has participated both as researcher and policymaker, see Goodfriend (2007).

There are, nevertheless, reasons for concern about current analysis including ongoing disputes about the empirical performance of key relationships in the semi-standard model; about communication and commitment mechanisms in theory and especially in practice; about the relationship of monetary policy to credit, fiscal, and foreign exchange policies; and about a myriad of technical details.[2] Also, there is much uneasiness about current policy approaches in the face of major credit market difficulties and indications of rising inflation.

In this context, the present article will be devoted to one specific problematic feature of the recent analytical literature, namely, a lack of agreement concerning the importance of multiple-solution indeterminacies in the analysis of monetary policy rules. References to "indeterminacy," in the sense of more than one dynamically stable solution, or "determinacy" appear on about 75 different pages of the hugely influential treatise on monetary policy analysis by Woodford (2003a) and are ubiquitous in the literature,[3] with a substantial majority of references expressed from the point of view that takes indeterminacy per se to be a matter of serious concern, e.g., implying that policies leading to model equilibria with that property should be rigorously avoided. The motivation is that indeterminacy should be avoided because it implies both that the policymaker cannot know which candidate equilibrium will prevail and also the possibility that "sunspot" effects may be created so as to greatly increase the volatility of crucial variables.[4] Several writers, however, have expressed the view that indeterminacy per se is not necessarily a problem—that a more appropriate criterion would be based on the concept of learnability of potential equilibria. This latter position has been taken overtly by McCallum (2001, 2003), Bullard and Mitra (2002), and Bullard (2006), and is stated or indirectly implied in a large number of writings by Evans and Honkapohja, including their influential and authoritative treatise (2001).

In the present article I wish to develop the position that indeterminacy is not necessarily problematic in the context of one particular application, namely, inflation forecast targeting in the sense of Taylor-style policy rules that respond, not to current (or past) inflation rates, but to currently expected values of inflation in future periods. That such indeterminacies might be brought about, and be undesirable, by strong responses of this type was first suggested

[2] In some quarters there is substantial concern over the neglect, in current mainstream analysis, of monetary aggregates. That topic is, however, distinct from those to be discussed in the present paper. For recent perspectives, see Woodford (2008) and McCallum (2008).

[3] Using the EBSCOhost search engine, one finds that, over the time span January 1995–June 2008, the number of papers and books with both "indeterminacy" and "monetary policy" appearing in their abstract is 78, while the number with these two terms appearing in their text is 166. There is some double-counting in these figures as both journal articles and working papers are included in the database.

[4] Some additional discussion is provided toward the end of Section 1.

by Woodford (1994)[5], and the argument was further developed by Bernanke and Woodford (1997), Clarida, Galí, and Gertler (2000), and (most thoroughly) Woodford (2003a, 256–61). Subsequently, many other authors have adopted this point of view, which is briefly mentioned in the textbooks of Walsh (2003, 247) and Galí (2008, 79–80). Indeed, it is apparently the prevailing point of view among analysts, despite the positive actual experience of the Bank of England over (say) 1996–2006.[6] I have briefly taken the opposing line of argument, that strong responses to expected future inflation rates will not be problematic, in McCallum (2001) and (2003), but those papers were primarily occupied with more general topics, which prevented a full development of this particular issue.[7]

In what follows, I will begin in Section 1 with an exposition of the nature of the indeterminacy problem in the context of inflation forecast targeting. Section 2 will then be devoted to the concept of learnability of a rational expectations (RE) solution. The position taken here is that the learnability of any particular RE solution should be considered a necessary condition for that solution to be plausible and, therefore, an equilibrium appropriate as a basis for thinking about real-world policy. In Section 3, numerical examples are developed to illustrate the points that have been made more generally, but also more abstractly, in Sections 1 and 2 and in previous writings. Section 4 then takes up the somewhat esoteric topic of "sunspot" solutions, i.e., solutions that include random components that are entirely unrelated to the specified model, including its exogenous variables. Finally, Section 5 provides a brief conclusion.

1. BASIC ANALYSIS

For concreteness, let us now adopt a simple model, representative of the recent literature, in which to discuss the issues at hand. It can be expressed in terms of a familiar three-equation structure as follows:

$$y_t = E_t y_{t+1} + b\,(R_t - E_t \pi_{t+1}) + v_t \qquad b < 0, \qquad (1)$$

[5] It has been, accordingly, referred to by Svensson (1997) as "the Woodford warning."

[6] It is well known that the Bank of England's policy during these years was to make adjustments in their policy rate in a Taylor-rule manner that responded to discrepancies between expected inflation two years ahead and the target rate.

[7] Woodford (2003b) criticized my 2003 paper primarily with regard to aspects concerning its subsidiary position with respect to a specific MSV (minimum state variable) solution. He did express an objection in the case of the application at hand, from a perspective that will be discussed below.

$$\pi_t = \beta E_t \pi_{t+1} + \kappa (y_t - \bar{y}_t) + u_t \qquad \kappa > 0; 1 > \beta > 0. \tag{2}$$

$$R_t = (1 - \mu_3) \left[(1 + \mu_1) \pi_t + \frac{\mu_2}{4} (y_t - \bar{y}_t) \right] + \mu_3 R_{t-1} + e_t \quad \mu_1 \geq 0. \tag{3}$$

Here, y_t and π_t are output and inflation expressed as fractional deviations from steady state and R_t is a one-period nominal interest rate that serves as the policy instrument. Thus, (1) is an IS-type relation consisting of a consumption Euler equation in combination with the overall resource constraint, (2) is a Calvo-style price adjustment equation, and (3) is the monetary policy rule.[8] Also, $\bar{y}_t$ is the flexible-price, natural rate of output, assumed to be generated exogenously by an AR(1) process[9] with AR coefficient ρ_a and innovation standard deviation SD(a). In the policy rule, the implicit target rate of inflation is zero. We will take the shock processes for u_t and e_t to be white noise (with standard deviations SD(u) and SD(e)) and the process for v_t to be AR(1) with AR coefficient ρ and standard deviation SD(v).

In the policy rule, the policy parameters μ_1 and μ_2 govern the strength of the central bank's policy response to deviations of inflation and output, respectively, from their target values, while μ_3 reflects the extent of interest rate smoothing. We begin with the central bank's policy responding to current observed inflation, π_t, and subsequently consider rules with a response to expected future inflation. In what follows, I will, for clarity, typically take μ_2 to be zero so that policy is responding only to inflation (usually with considerable smoothing). This practice (i.e., setting $\mu_2 = 0$) changes the numerical values at which effects such as indeterminacy occur but does not alter the arguments to be made in any essential manner.

Let us begin the analysis by also setting $\mu_3 = 0$ and $u_t = 0$ so that there is no smoothing and no price-setting shock. Then, substitution of equation (3) into (1) yields

$$y_t = E_t y_{t+1} + b \left[(1 + \mu_1) \pi_t + e_t - E_t \pi_{t+1} \right] + v_t, \tag{4}$$

and we can consider (2) and (4) as a two-equation system determining the evolution of π_t and y_t. The "fundamentals" or minimum-state-variable (MSV) solution will have these variables determined as linear functions of v_t, e_t, and $\bar{y}_t$.[10] As one final simplification we take the latter ($\bar{y}_t$) to be constant and

[8] Constant terms are omitted from (1) and (3) only for expositional simplicity. The analysis that follows implicitly presumes that nonzero constants are present in both of these relationships.

[9] That is, autoregressive of order one.

[10] Here I am using MSV as an alternative name for the fundamentals solution in the manner of Evans and Honkapohja (2001, 193–4), not in the manner proposed by McCallum (1983, 2003)

normalize it at zero. Then a fundamentals solution to the model (2)(4) will be of the form

$$y_t = \phi_{11} v_t + \phi_{12} e_t \text{ and} \tag{5}$$

$$\pi_t = \phi_{21} v_t + \phi_{22} e_t, \tag{6}$$

with constant terms again omitted only for simplicity.

In this case, the expected values one period ahead are $E_t y_{t+1} = \phi_{11} \rho v_t$ and $E_t \pi_{t+1} = \phi_{21} \rho v_t$. Then we can substitute these two expressions plus (5) and (6) into (2) and (4) to obtain undetermined-coefficient relations that express the ϕ_{ij} coefficients of the solution expressions in terms of the parameters of the structural equations (2) and (4). The results of that (tedious but straightforward) exercise are as follows:

$$\phi_{11} = \frac{1 - \beta\rho}{(1-\rho)(1-\beta\rho) - b\kappa(1 + \mu_1 - \rho)}, \tag{7}$$

$$\phi_{12} = \frac{b}{1 - \kappa b(1 + \mu_1)}, \tag{8}$$

$$\phi_{21} = \frac{\kappa}{(1-\rho)(1-\beta\rho) - b\kappa(1 + \mu_1 - \rho)}, \text{ and} \tag{9}$$

$$\phi_{22} = \frac{\kappa b}{1 - \kappa b(1 + \mu_1)}. \tag{10}$$

Here, the specified signs of the basic parameters imply that $\phi_{11} > 0$, $\phi_{12} < 0$, $\phi_{21} > 0$, and $\phi_{22} < 0$, so the solution equations show that a positive shock to demand (v_t) increases both inflation and output while a positive shock to the policy rule (e_t) decreases both inflation and output. From the expressions it can also be seen that (since $b < 0$) an increase in μ_1 increases the values of the positive denominators in all four expressions, implying that the variances of both inflation and output are decreased by a stronger positive policy response to observed inflation.

Are there other solutions, i.e., other expressions in addition to (5) and (6), that give values of the jointly dependent variables y_t and π_t in terms of exogenous and/or predetermined variables while satisfying the structural equations (1)–(3)? Without attempting an exhaustive search, let us (for comparison below) consider whether π_{t-1}, the lagged inflation rate, might be an additional

---

or Evans (1986). The latter, but not the former, involves a concept that is, in all cases, unique by construction.

variable that should appear in the solution equations. Thus, we add π_{t-1} to (5) and (6) with coefficients ϕ_{13} and ϕ_{23}, and then repeat the steps leading to (7)–(10). Upon doing so, we find that ϕ_{23} must satisfy the cubic equation

$$\left(\phi_{23} - \beta\phi_{23}^2\right)(1 - \phi_{23}) = \kappa b(1 + \mu_1)\phi_{23} - \kappa b\phi_{23}^2. \tag{11}$$

One root of the latter is clearly 0 and the other two must satisfy the quadratic $(1 - \beta\phi_{23})(1 - \phi_{23}) = \kappa b(1 + \mu_1) - \kappa b\phi_{23}$, which can be written as

$$\beta\phi_{23}^2 - [\beta + 1 - \kappa b]\phi_{23} + \left[1 - \kappa b(1 + \mu_1)\right] = 0. \tag{12}$$

For $\mu_1 = 0$, (12) can be written as $\left[(1 - \kappa b) - \beta\phi_{23}\right](1 - \phi_{23}) = 0$, from which we see that the two roots are 1 and $(1 - \kappa b)/\beta$. Since b < 0 and $\kappa > 0$, the latter is unambiguously greater than 1. With $\mu_1 > 0$ but very small, there are two positive real roots that both exceed 1, and for larger μ_1 there are conjugate complex roots with modulus greater than 1. Thus, with positive μ_1, there is no stable root to the quadratic (12) and therefore no stable root to the cubic except 0. With negative μ_1, by contrast, one root of (12) would equal 1 and the other would be positive and smaller than 1. Thus, there is an additional stable solution when μ_1 is negative but no additional stable solution—that is, additional to the fundamentals solution given by (5)–(10)—when it is positive. This finding, of course, represents the Taylor principle for the model at hand, in the case with $\mu_2 = 0$. This conclusion agrees exactly with that of Woodford (2003a, 254), which is obtained by an alternative procedure.[11]

With that background, we now turn to the case of inflation forecast targeting, in which the central bank's policy rule responds not to current inflation, but to $E_t\pi_{t+1}$, the current expectation of inflation one period into the future.[12] Going through steps to obtain the fundamentals solution as before, we find the equations comparable to (7)–(10) to be:

$$\phi_{11} = \frac{1 - \beta\rho}{(1 - \rho)(1 - \beta\rho) - b\kappa\rho\mu_1}, \tag{13}$$

$$\phi_{12} = b, \tag{14}$$

$$\phi_{21} = \frac{\kappa}{(1 - \rho)(1 - \beta\rho) - b\kappa\rho\mu_1}, \text{ and} \tag{15}$$

[11] Note that Woodford's condition (2.7, Ch. 4) is more general in that it permits responses to the output gap. His analytical procedure is more amenable to generalization than the one given here, but the latter is more elementary in terms of concepts utilized.

[12] This terminology does not agree with that of Svensson (1997) but is, I think, consistent with general usage.

$$\phi_{22} = \kappa b. \tag{16}$$

By inspection we can see that the signs are as before. We can also see that in this case a stronger policy response (larger μ_1) has no effect on the variability of inflation or output that occurs in response to a policy shock. Also, it is easy to determine that the denominators in (13) and (15) are smaller than in (7) and (9), so stabilization with respect to demand shocks is less effective than when policy responds to the current inflation rate.

Our present interest in the contrast, however, concerns the multiplicity of stable solutions that is possible under the policy rule that responds to expected inflation, $E_t\pi_{t+1}$. One way to demonstrate the existence of this multiplicity is to again go through the steps leading to solution expressions while including lagged inflation as an additional state variable in solutions such as (5) and (6), i.e., by using

$$y_t = \phi_{11}v_t + \phi_{12}e_t + \phi_{13}\pi_{t-1} \tag{5'}$$

$$\pi_t = \phi_{21}v_t + \phi_{22}e_t + \phi_{23}\pi_{t-1}. \tag{6'}$$

That change implies that $E_t y_{t+1} = \phi_{11}\rho v_t + \phi_{13}(\phi_{21}v_t + \phi_{22}e_t + \phi_{23}\pi_{t-1})$ and $E_t\pi_{t+1} = \phi_{21}\rho v_t + \phi_{23}(\phi_{21}v_t + \phi_{22}e_t + \phi_{23}\pi_{t-1})$. Then, undetermined coefficient reasoning implies that the values for the ϕ_{ij} are given by six relations analogous to those used in deriving (11) and (12), among which are

$$\phi_{13} = b\mu_1\phi_{23}^2 + \phi_{13}\phi_{23} \text{ and} \tag{17}$$

$$\phi_{23} = \beta\phi_{23}^2 + \kappa\phi_{13}. \tag{18}$$

From these, ϕ_{13} can be solved out, yielding the cubic equation

$$\phi_{23} = \beta\phi_{23}^2 + \kappa b\mu_1\phi_{23}^2/\left(1 - \phi_{23}\right). \tag{19}$$

Clearly, one solution to the latter is provided by $\phi_{23} = 0$, which then by (18) implies $\phi_{13} = 0$. This eliminates the π_{t-1} variable and leads back to the fundamentals solution obtained previously. But (19) is also satisfied by roots of the quadratic

$$\beta\phi_{23}^2 - \left[1 + \beta + \kappa b\mu_1\right]\phi_{23} + 1 = 0, \tag{20}$$

i.e., by

$$\phi_{23} = \frac{d \pm \left[d^2 - 4\beta\right]^{0.5}}{2\beta}, \tag{21}$$

where d is the term in square brackets in (20). Therefore, for some values of the parameters κ, β, b, and μ_1 there may be other real solutions, in addition to the fundamentals solution, that are stable.[13] (If such solutions are dynamically explosive, they do not create indeterminacy.) Tedious but simple algebra shows that with $\kappa > 0$, $b < 0$, and $0 < \beta < 1$, the region of determinacy includes all values of μ_1 between 0 and $\mu_1^c = -2(1+\beta)/\kappa b$, which is positive. For $\mu_1 < 0$ there are two positive real solutions with one of them stable (smaller than 1.0 in absolute value) so there is a second stable solution (indeterminacy), just as in the case with current inflation in the policy rule. But now it is the case that for $\mu_1 > \mu_1^c$ there are two real solutions to the quadratic, both negative, and one of them is stable. Thus, in this case we have a nonfundamentals solution for which there is no transversality condition to rule out the implied dynamic behavior as a rational expectations equilibrium. Instead, there is an infinite multiplicity of stable RE solutions indexed by the initial start-up value of π_{t-1}. In such cases, moreover, "sunspot" solutions are also possible in the sense of not being ruled out by the conditions of RE equilibria.[14] This is the problem suggested by the "Woodford warning" and presented in Woodford (2003a, 252–61, 2.11) where he generalizes our expression for μ_1^c.[15] Similar results are developed by King (2000, 78–82). The danger in question is made less likely, it should be mentioned, when values of μ_2 exceed zero.[16]

In what follows, it will be of considerable importance to take account of interest rate smoothing, that is, cases in which μ_3 in policy rule (3) is positive. For these cases, the analogous critical value will be denoted μ_1^{cc} and is given by

$$1 + \mu_1^{cc} = \left[1 + \frac{2\,(1+\beta)}{-b\kappa}\right]\left(\frac{1+\rho}{1-\rho}\right). \tag{22}$$

This expression is the special case, with no response to the output gap, of Woodford's expression (2003a, 258, 2.13) of Chapter 4.

---

[13] Analyses are provided by Bullard and Mitra (2002; 1,121–3) and Woodford (2003a, 256–60).

[14] A sunspot solution is one that includes random variables (of a martingale difference variety) that have no connection with other elements of the model. Such solutions will be considered in Section 5.

[15] Ours is a special case of Woodford's formula, as the latter formula admits the possibility of responses to the current output gap, as well as to expected inflation.

[16] See, e.g., Bullard and Mitra (2002) and Woodford (2003a, 257–8).

Thus far, I have discussed inflation forecast targeting only for the case in which the expected inflation rate, to which the central bank responds, is one period into the future, i.e., $E_t\pi_{t+1}$. But clearly it could instead be $E_t\pi_{t+j}$ with $j > 1$.[17] Since algebraic analysis of such cases is tedious, and since some additional concreteness to the discussion might in any case be useful, I will proceed by way of numerical calculations pertaining to a specific quantitative version of the model at hand. For our basic results, let us adopt the following calibrated parameter values, which have been chosen to be representative of semi-realistic specifications: $\beta = 0.99$, $b = -0.6$, $\kappa = 0.05$, $\rho = 0.5$, and $\rho_a = 0.95$.[18] Also, we have a policy rule that responds to expected inflation j periods into the future, with $j = 0, 1, 2, 3,$ or 4, that involves no response to the output gap and includes interest rate smoothing of a realistic magnitude. Thus, we use equation (3) with $\mu_2 = 0$ and $\mu_3 = 0.8$. The object now will be to look for critical values of μ_1 at which determinacy is lost, and multiple solutions begin, as $\mu_1$is increased. Results are shown in the next-to-last column of the following:

j	Inflation Variable in Policy Rule	Critical Value, μ_1^{cc} with $\mu_2 = 0$	Critical Value with $\mu_2 = 0.5$
0	Δp_t	None	None
1	$E_t\Delta p_{t+1}$	1,202–1,203	1,221–1,222
2	$E_t\Delta p_{t+2}$	105–106	117.8–117.9
3	$E_t\Delta p_{t+3}$	16–17	25.8–25.9
4	$E_t\Delta p_{t+4}$	3.5–3.6	10.4–10.5

Thus, with $j = 0$, we have the familiar result that the Taylor principle holds in the sense that, with current inflation in the Taylor-style rule, determinacy obtains for all values of $1 + \mu_1 > 1$. With $E_t\Delta p_{t+1}$ in the rule, determinacy holds for all μ_1 up through 1,202, but indeterminacy sets in before μ_1 reaches 1,203. And for expectations of inflation farther in the future, the determinacy region becomes progressively smaller until it disappears entirely at $j = 5$.[19] The importance of cases with j greater than 1 or 2 is underlined by the aforementioned example of the Bank of England, which in recent years conducted policy so as to bring the expected inflation rate average over $j = 5, 6, 7,$ and 8 into equality with its target value of 2.5 percent per annum.[20]

[17] Of course averages or other combinations would also be possible.

[18] The values of β and ρ_a are quite standard in the literature, while the value for b is representative of most analysts (though not Woodford [2003a]). For κ, 0.05 is a bit larger than Woodford's 0.024 and, finally, my choice of ρ_v is quite arbitrary but designed to reflect a moderate degree of positive autocorrelation.

[19] Complete disappearance might not occur, depending on specificational details, if the model's equations were enriched so as to imply more persistence. One reader has asked how to interpret the magnitudes of μ_1. One relevant point is that the original Taylor rule value is 0.5, but the literature seems to suggest that values up to 5.0, at least, are practical. More implausibly, but still of analytical relevance, some writers have implicitly suggested that no finite value is large enough.

[20] See, for example, Bean and Jenkinson (2001) and U.K. Treasury (2003).

As mentioned above, policy-rule responses to the output gap can also be helpful in creating determinacy. To illustrate this possibility, the final column in the table shows the critical value for μ_1 when the output response coefficient, μ_2, is set equal to 0.5. It will be seen that for each j, the critical value is increased, thereby reflecting a larger range of μ_1 values over which determinacy prevails. The quantitative magnitude of this improvement is not great, however.

What are the undesirable consequences of indeterminacy? Since I am doubtful that there are any, I will quote other writers. Woodford (2003a, 45) states that

> ...even if one restricts one's attention to bounded solutions ...there is an extremely large set of equally possible equilibria. These include equilibria in which endogenous variables such as inflation and output respond to random events that are completely unrelated to economic "fundamentals" (i.e., to the exogenous disturbances that affect the structural relations ...) and also equilibria in which "fundamental" disturbances cause fluctuations in equilibrium inflation and output that are arbitrarily large relative to the degree to which the structural relations are perturbed. Thus, in such a case, macroeconomic instability can occur owing purely to self-fulfilling expectations.

More compactly, Lubik and Schorfheide (2004, 190) state that "broadly speaking, indeterminacy has two consequences. First, the propagation of fundamental shocks, such as technology or monetary policy shocks, [through] the system is not uniquely determined. Second, sunspot shocks can influence equilibrium allocations and induce business-cycle fluctuations that would not be present under determinacy." From a more specific perspective, several writers have attributed poor performance of U.S. monetary policy during the 1970s to a policy rule that permitted indeterminacy.[21]

2. LEARNABILITY, NOT DETERMINACY

Having posed the indeterminacy problem, I now proceed to argue that it is in fact a pseudo-problem, i.e., one that should not be considered as relevant for policymaking in actual economies. The basic argument is that for any RE solution to be considered plausible, and therefore potentially relevant for policy analysis, it should be *learnable*, in the sense of Evans and Honkapohja (2001).[22] Then this requirement, which pertains to a specific least-squares

---

[21] For this interpretation, see, e.g., Clarida, Galí, and Gertler (2000) and Lubik and Schorfheide (2004).

[22] The Evans and Honkapohja (E&H) learning procedure has been described by various authors, including Bullard and Mitra (2002), Bullard (2006), and McCallum (2001, 2003, 2007), as well as a long list of papers by E&H.

learning procedure, eliminates the multiplicity of stable solutions described above, that is, for cases with $\mu_1 > \mu_1^{cc}$. Furthermore, from a broader and more general perspective, there is a strong tendency for the learnability requirement to rule out, as implausible, RE solutions other than a single, learnable, fundamentals solution.

The essential rationale for the learnability requirement is as follows: In any dynamic market economy, individual agents seeking optimal (or even desirable) outcomes for themselves will need to form expectations regarding future values of some endogenous variables. To do this in a manner consistent with the RE hypothesis, the agents must have a considerable amount of quantitative information regarding the time series properties of these variables, i.e., knowledge beyond a listing of relevant variables and functional forms. That quantitative knowledge cannot be gained by introspection or divine intervention or magic; instead, it must be obtained from information generated by the economy itself. A plausible model of this economy should then be one in which the agents can learn accurately about the quantitative properties of the model economy on the basis of data generated by that model. More specifically, any RE solution in the model must, to be plausible, be one that is *learnable* in the sense of there existing the possibility of the model's agents learning about the properties of that solution from data generated by the model economy. Of course, there are many conceivable learning schemes that an analyst could specify. For that reason, some economists have objected to results based on this particular learning mechanism, arguing that many others would be possible. I would argue, however, that the least-squares (LS) learning process is strongly "slanted" or "biased" toward a positive finding, i.e., toward generating a finding that the potential equilibrium in question is learnable. Specifically, the process is such that agents are depicted as forecasting on the basis of least-squares estimates of a vector-autoregression model that is correctly specified, i.e., includes the relevant lagged variables and the proper number of lags, and is re-estimated each period using data generated up through the previous period. Any particular RE solution is regarded as learnable if, as time passes with agents basing their expectations (forecasts) each period on these regressions, the system approaches this RE equilibrium. Thus, it is the case that "the LS learning process in question assumes that (i) agents are collecting an ever-increasing number of observations on all relevant variables while (ii) the structure is remaining unchanged. Furthermore, (iii) the agents are estimating the relevant unknown parameters (iv) with an appropriate estimator (v) in a properly specified model. Thus if a proposed RE solution is not learnable by the process in question—the one to which the Evans and Honkapohja (E&H) results pertain—then it would seem highly implausible that it could prevail in practice" (McCallum 2007; 1,378).

What is the relationship between a model's determinacy (or indeterminacy) and the learnability of its equilibria? McCallum (2007) demonstrates that, for a very broad class of linear models, determinacy implies learnability (of the single stable solution) under the assumption that the economy's agents have access to current-period values of endogenous variables in the learning process. If only lagged values are available in that process, however, determinacy does not imply learnability. More importantly for the topic at hand, it is shown that models with indeterminacy (more than one stable solution) may have one or more learnable equilibria.[23] If in fact they have one, and the others are not learnable, then there would seem to be only one equilibrium that is plausible and therefore relevant for policy analysis.

It may be useful to provide a brief summary of the formulation and results developed in McCallum (2007). Accordingly, we consider a model of the form

$$y_t = A\ E_t y_{t+1} + C\ y_{t-1} + D\ u_t, \tag{23}$$

where y_t is a m×1 vector of endogenous variables, A and C are m×m matrices of real numbers, D is m×n, and u_t is a n×1 vector of exogenous variables generated by a dynamically stable process,

$$u_t = P\ u_{t-1} + \varepsilon_t, \tag{24}$$

with ε_t a white noise vector and P a matrix with all eigenvalues less than 1.0 in modulus. It will not be assumed that A is invertible. This specification is useful in part because it is the one utilized in Section 10.3 of Evans and Honkapohja (2001), for which conditions relevant for learnability are reported on their p. 238.[24] Furthermore, the specification is very broad; in particular, any model satisfying the formulations of King and Watson (1998) or Klein (2000) can (with the use of auxiliary variables) be written in this form—which will accommodate any number of lags, expectational leads, and lags of expectational leads. In this setting, we consider solutions to model (1)–(2) of the form

$$y_t = \Omega\ y_{t-1} + \Gamma\ u_t, \tag{25}$$

in which Ω is required to be real. Then we have that $E_t y_{t+1} = \Omega(\Omega y_{t-1} + \Gamma u_t) + \Gamma P u_t$, and straightforward undetermined-coefficient reasoning shows that Ω and Γ must satisfy

[23] Cases with more than one learnable equilibrium seem to be quite rare but are possible in principle.

[24] Constant terms can be included in the equations of (1) by including an exogenous variable in u_t that is a random walk whose innovation has variance zero. In this case there is a borderline departure from process stability. The conditions on E&H (2001, 238) actually pertain to E-stability; see the discussion below.

$$A\Omega^2 - \Omega + C = 0, \text{ and} \tag{26}$$

$$\Gamma = A\Omega\Gamma + A\Gamma P + D. \tag{27}$$

For any given Ω, (27) yields a unique Γ generically,[25] but there are many m×m matrices that solve (26) for Ω. These result from different orderings of the generalized eigenvalues of the matrix pencil $B - \lambda A$. If more than one of the Ωs that satisfies (26) has all its eigenvalues less than 1 in modulus, there are multiple stable solutions, i.e., indeterminacy.

Let us then turn to conditions for learnability of specific solutions. First we review the main results outlined in McCallum (2007) with details of the argument relegated to Appendix A. We begin with the assumption that agents have full information on current values of endogenous variables during the learning process, and then we will mention a second assumption, namely, that only lagged values of endogenous variables are known during the learning process. The manner in which learning takes place in the E&H analysis is as follows. Agents are assumed to know the structure of the economy as specified in equations (1) and (2), in the sense that they know what variables are included, but do not know the numerical values of the parameters. What they need to know, to form expectations, are values of the parameters of the solution equations (25). In each period t they form forecasts on the basis of least-squares regression of the variables in y_{t-1} on previous values of y_{t-2} and any exogenous observables. Given those regression estimates, however, expectations of y_{t+1} may be calculated assuming knowledge of y_t or, alternatively, assuming that y_{t-1} is the most recent observation that is usable in the forecasting process. In the former case, the conditions reported by E&H (2001, 238) are that the following three matrices must have all eigenvalues with real parts less than 1.0:

$$F \equiv (I - A\Omega)^{-1} A, \tag{28a}$$

$$\left[(I - A\Omega)^{-1} C\right]' \otimes F, \text{ and} \tag{28b}$$

$$P' \otimes F. \tag{28c}$$

In the second case, however, the analogous conditions (E&H 2001, 245) are that the following matrices must have all eigenvalues with real parts less than 1.0:

[25] Generically, $I - P' \otimes [(I - A\Omega)^{-1} A]$ will be invertible, permitting solution for vec(Γ).

$$A\,(I+\Omega)\,, \tag{29a}$$

$$\Omega' \otimes A + I \otimes A\Omega, \text{ and} \tag{29b}$$

$$P' \otimes A + I \otimes A\Omega. \tag{29c}$$

Except in the case that $\Omega = 0$, which will obtain when $C = 0$, these conditions are not equivalent to those in (28).

It is important to note that use of the first information assumption is not inconsistent with a model specification in which supply and demand decisions in period t are based on expectations formed in the past, such as $E_{t-1}y_{t+j}$ or $E_{t-2}y_{t+j}$. It might also be mentioned parenthetically that conditions (28) and (29) literally pertain to the *E-stability* of the model (23)(24)—see Evans (1986) and, for a heuristic introduction, McCallum (2003; 1,157–9)—under the two information assumptions, not its learnability. Under quite broad conditions, however, E-stability is necessary and sufficient for LS learnability. This near-equivalence is referred to by E&H as the "E-stability principle." Since E-stability is technically easier to verify, applied analysis typically focuses on it rather than on direct exploration of learnability.

Given the foregoing discussion, it is a simple matter to verify that if a model of form (23)(24) is determinate, then it satisfies conditions (28). First, determinacy requires that all eigenvalues of F must have moduli less than 1.0, so their real parts must all be less than 1.0, thereby satisfying (28a). Second, from equation (26) it can be seen that $(I - A\Omega)^{-1}C = \Omega$. Therefore, matrix (28b) can be written as $\Omega' \otimes F$. Furthermore, it is a standard result (E&H 2001, 116) that the eigenvalues of a Kronecker product are the products of the eigenvalues of the relevant matrices (e.g., the eigenvalues of $\Omega' \otimes F$ are the products $\lambda_\Omega \lambda_F$). Therefore, condition (28b) holds. Finally, since $|\ \lambda_F\ | < 1$, condition (28c) holds provided that all $|\ \lambda_P\ | \leq 1$, which we have assumed by specifying that (24) is dynamically stable.

3. NUMERICAL ANALYSIS

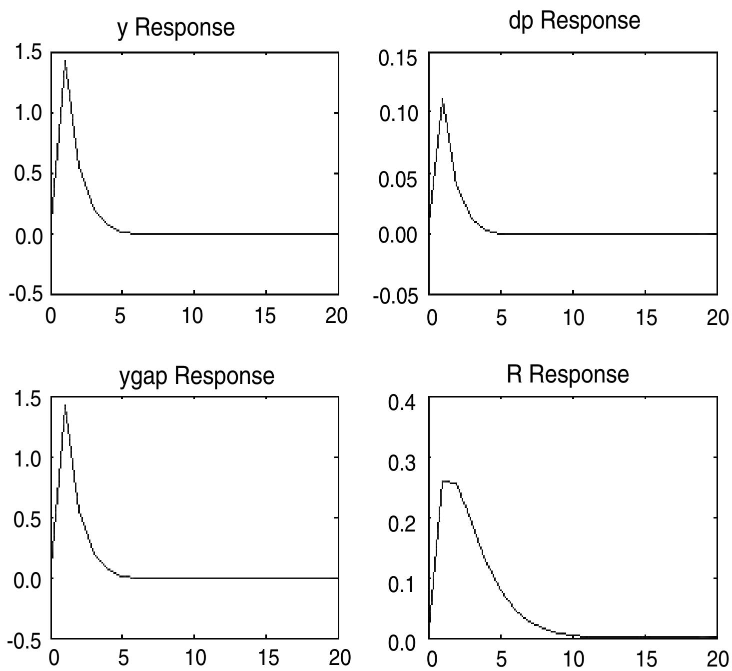
We now continue with the numerical example introduced in Section 1, where it was shown that the critical value, at which strong responses of monetary policy to expected inflation j periods into the future creates indeterminacy, decreases as j is increased—thereby making the problem (if it is one) more serious. We now look beyond that familiar finding, however, in a manner suggested by the discussion provided above. For specificity, let's focus on one particular case—that in which Δp_{t+2} is the inflation variable in the rule. Specifically, we inspect the system eigenvalues for a policy feedback value of $\mu_1 = 105$, just below the critical value. There the nonzero and finite eigenvalues are $0.6187 + 0.7857i$, $0.6187 - 0.7857i$, and 0.4920. Thus, the modulus of each of the complex values is 1.0000418, whereas the eigenvalue pertaining to the single relevant predetermined variable, R_{t-1}, is 0.4920. If we increase μ_1 to 106, however, the three nontrivial eigenvalues become $0.6159 + 0.7864i$, $0.6159 - 0.7864i$, and 0.4913. These are very little changed, but now the modulus of the two complex values is 0.998903. Thus, there are three stable eigenvalues but still only one predetermined variable, so there are three stable solutions. Among these, consider first the MOD solution, which has the same ordering as in the previous case. Then for E-stability we need, from (28a) and (A6) of Appendix A, the inverses of the two complex eigenvalues to have real parts less than 1. In fact, these inverses are $0.617236 - 0.78817i$ and $0.617236 + 0.78817i$, so both have real parts less than 1, meeting this criterion. The criteria (28b) and (28c) are both clearly met, as well. Thus, the MOD solution is E-stable and learnable.

What about the other stable solutions? One cannot exchange the place of the eigenvalue 0.4920 with either one of the complex numbers because that would give a solution expression that assigns complex values to reduced-form coefficients (thereby implying that numerical observations on interest rates, inflation rates, etc., are complex!). The only way to get a real solution is to reorder by shifting both complex eigenvalues to below the line, with both 0.4920 and 0 shifting to above the line.[26] But that implies that both of the latter have inverses—therefore eigenvalues of F—with real parts greater than 1. Accordingly, this solution violates the criterion (28a) necessary for E-stability and learnability. There is, therefore, only a single RE solution of form (25) for the calibration at hand. It features only a single RE solution that is plausible.

It is of some interest to continue with this example by examining impulse response functions (IRFs) pertaining to these alternative solutions. In Figure 1, we have impulse responses for the endogenous variables y_t, Δp_t, $\tilde{y}_t$, and R_t generated by a unit shock to the innovation in the policy rule when the μ_1 policy parameter is set at 104, just below the critical value at which indeterminacy arises. In Figure 2, we show the IRFs for the MOD solution when $\mu_1 = 106$, just above the critical value. It will be readily observed that the responses are very similar in these two cases—indeed, are visually identical. Next, consider the implications of the hypothesis that one of the alternative (non-MOD) stable solutions is relevant for values of $\mu_1 > 105$, even though the MOD solution is relevant at $\mu_1 = 105$. Thus, in Figure 3 we report the

[26] The line, that is, that separates eigenvalues associated with predetermined variables from the others.

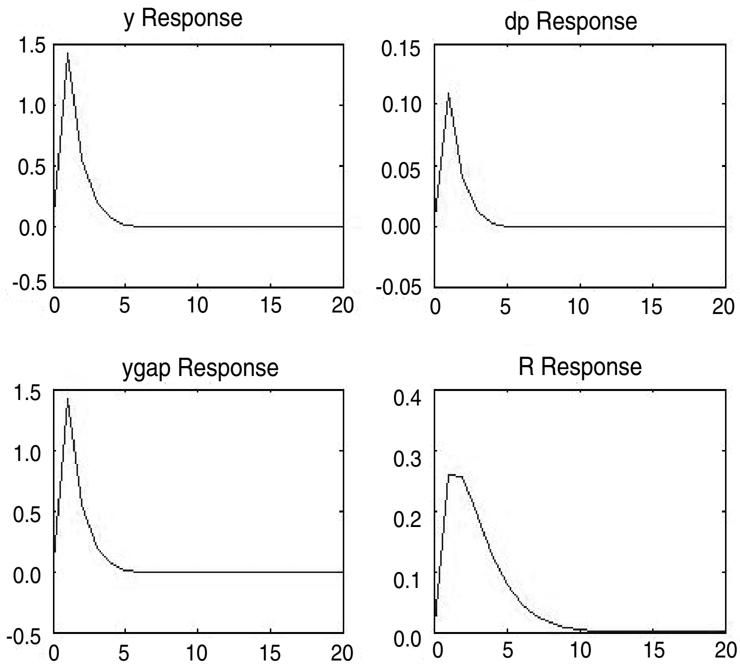
Figure 1 Responses to Unit Shock to IS, $\mu_1 = 104$



IRFs at $\mu_1 = 106$ for the alternative stable solution,[27] and in this case there is clearly no similarity to Figure 1. Indeed, the initial-period response is in the opposite direction from that in Figure 1 for both the inflation and interest rate variables. In sum, we find that there is no discontinuity pertaining to the impulse response functions around the critical value, according to the MOD solution. But with the alternative ordering (yielding a solution analogous to that with ϕ_{23} given by (21) in Section 1) there is a drastic change (relative to the IRFs for the only stable solution with $\mu_1 < 105$) resulting from a very small change in one parameter value (μ_1) (see Figure 3). This contrast seems to be strongly suggestive of the idea that the MOD solution is, in this case, much more plausible than the alternative solution, dynamic stability of the latter notwithstanding.

[27] The different arrangements of the complex eigenvalues yield identical impulse response functions.

Figure 2 Responses to Unit Shock to IS, $\mu_1 = 106$

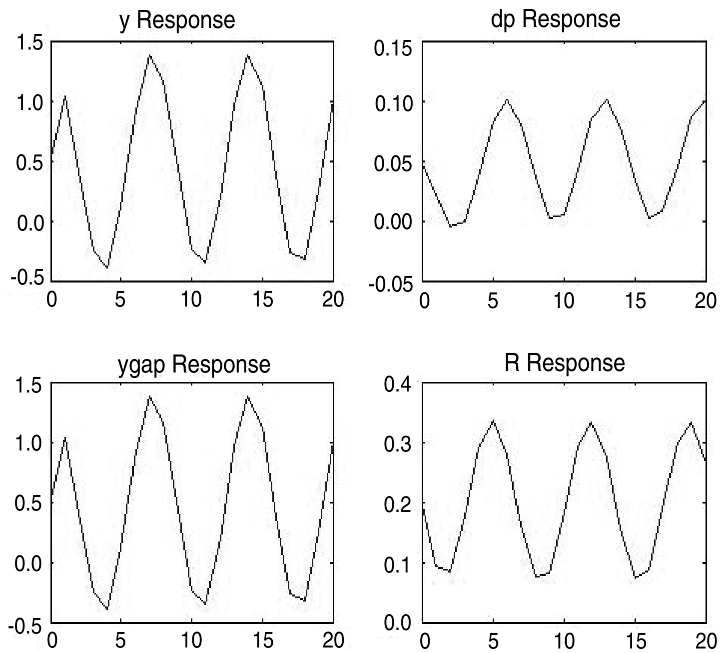


4. SUNSPOT EQUILIBRIA

Thus far, we have seen that nonfundamental equilibria of the form considered in (5')–(6') are not learnable, thereby lending support to the position that strenuous inflation forecast targeting is not dangerous. We have not, however, considered all possible forms of indeterminacy. In particular, in his critical review of the argument pertaining to inflation-forecast targeting in McCallum (2003), Woodford (2003b; 1,181–2) has shown that in cases with indeterminacy, solutions of a certain "sunspot" type—one that depends on an extraneous variable that evolves exogenously according to a finite-state Markov chain—can be learnable. This result accords with the analysis of Honkapohja and Mitra (2004; 1,753–4), who conduct a detailed analysis of Markov SSE (stationary sunspot equilibria) cases as well as non-Markov SSE cases. For some relevant discussion, see Appendix B.

It is nevertheless my belief that the indeterminacy under discussion pertains to RE solutions that are not plausible. This belief is based not on any refutation of the formal learnability analysis, but instead on a judgment

Figure 3 Responses to Unit Shock to IS, $\mu_1 = 106$ (alt. soln.)



concerning the way in which the formal analysis is used to gain insight into actual behavior in real-world economies. Specifically, the learning process described above in previous sections postulates individual agents who in effect base their expectations on forecasting rules implied by correctly specified (but unconstrained) vector-autoregression (VAR) models constructed using data from previous periods. To me this seems to be going as far as common sense allows in attributing sophistication to the expectation-formation processes of individuals in actual economies. But the forecasting rules needed for achievement of the Markov SSE solutions are not implied by VARs of this type; they require state-dependent intercept terms. That is, estimation of basic VARs over indefinitely long spans of time would not lead to forecast rules with the type of state-dependent parameters needed to support sunspot SSEs. Thus, I contend that such equilibria are simply implausible and should not be considered when discussing possible RE equilibria relevant for the design of monetary policy.[28]

Now, I can well imagine that some readers might be inclined to respond to this argument with the objection that rational expectations is itself implausible, so that there is an inconsistency in this argument. But of course, taken literally, RE itself *is* implausible—as early critics emphasized. Nevertheless, RE is rightly regarded by mainstream researchers as the appropriate assumption for the purpose of economic analysis, especially in the context of macroeconomic policy analysis. That is the case because RE is fundamentally the assumption that agents optimize with respect to their expectational behavior—just as they do (according to basic neoclassical economic analysis) with respect to other regular economic activities such as selection of consumption bundles, selection of quantities produced and inputs utilized, etc.—for a necessary condition for optimization is that individuals eliminate any systematically erroneous component of their expectational behavior. Moreover, RE is doubly attractive (to researchers) from a policy perspective, for it assures that a researcher does not propose policy rules of a type that is designed to exploit allegedly consistent patterns of suboptimal expectational behavior by individuals.

Accordingly, I contend that there is no inconsistency in using RE as one's expectational hypothesis while placing some limit on the scope of learning processes that can lead to RE equilibria. This is, as mentioned above, fundamentally a specification regarding information availability. In standard RE analysis it is assumed that agents have knowledge of the values of endogenous (and some exogenous) variables only in the present and past, not the future. Furthermore, in some influential papers, such as Lucas (1972), only partial information concerning current endogenous variables is available. There seems to be no difference in principle from our preceding argument in assuming that agents may not have observations on the current state of the system needed for the learning analysis of the Markov SSE variety.

There is also an alternative class of sunspot equilibria that will be dynamically stable when the determinacy condition of Section 1 is not satisfied. Specifically, in terms of our model (2) and (4), an arbitrary sunspot variable ξ_t with the property $E_{t-1}\xi_t = 0$ can be added to expressions like (5') and (6'), leading to stable solutions of that form. Honkapohja and Mitra (2004; 1,756) find, however, that "non-fundamental equilibria of the form (3) and (5) [their equation numbers] are never E-stable." Thus, this alternative class of

[28] An alternative way of presenting this position would be to *define* an RE equilibrium in a dynamic model for monetary policy analysis so as to require LS learnability (on the basis of *basic* VAR forecasting rules) as a necessary condition—one that represents *informational feasibility*.

sunspots does not, according to the learnability criterion promoted here, pose a problem for actual monetary policy.[29]

Some expository material pertaining to the two classes of sunspot solutions, and an apparent (but not actual) inconsistency in the results of Honkapohja and Mitra (2004) and E&H (2003), is provided in Appendix B.

5. CONCLUSIONS

In the contemporary mainstream literature on monetary policy analysis, it is typically contended that inflation forecast targeting—use of an interest rate policy rule that responds to currently expected inflation for some future period(s)—will, if applied too strongly, generate indeterminacy in the sense of a multiplicity of dynamically stable RE solutions. It is also concluded in this literature that this outcome represents a practical difficulty for monetary policymakers. By contrast, the present article argues that these findings of indeterminacy do not pose any actual problem for monetary policymakers. The reason is that in these analyses only one of the RE solutions possesses the property of least-squares learnability, a concept that is necessary for the plausibility of any rational expectations solution. Accordingly, other RE solutions could not plausibly prevail because there is no way for individual agents in the model (designed to depict reality) to obtain enough quantitative information about the economy's dynamics to form expectations in a way that would support the solution in question. Typically, however, this objection does not pertain to a single RE solution, which is the "natural" fundamentals solution. Thus, this article contends that indeterminacy of the type in question represents a pseudo-problem, not an actual problem for actual policymakers.

APPENDIX A

Here we provide additional development of results summarized in Section 2. Continuing from three lines below equation (27), the following analysis centers around (26). Since we do not assume that A is invertible, we write

[29] That finding is highly agreeable from the perspective of my argument. But even if it did not hold, I would continue to suggest that these arbitrary sunspot solutions are highly implausible. The crux of the matter is that the learning analysis treats the unspecified sunspot variables as observable by individual agents. But sunspot variables are, by definition, ones that represent no component of tastes, technology, government behavior, or institutional constraints—they represent merely the arbitrary beliefs of (individual and independent) market participants.

$$\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} \Omega^2 \\ \Omega \end{bmatrix} = \begin{bmatrix} I & -C \\ I & 0 \end{bmatrix}\begin{bmatrix} \Omega \\ I \end{bmatrix}, \tag{A1}$$

in which the first row reproduces the matrix quadratic (26). Let the 2m×2m matrices on the left and right sides of (A1) be denoted $\bar{A}$ and $\bar{C}$, respectively. Then instead of focusing on the eigenvalues of $\bar{A}^{-1}\bar{C}$, a matrix that does not exist when A is singular, we instead solve for the (generalized) eigenvalues of $\bar{C}$ with respect to $\bar{A}$. Thus, instead of diagonalizing $\bar{A}^{-1}\bar{C}$, as in Blanchard and Khan (1980), we use the Schur generalized decomposition, which establishes that there exist unitary matrices Q and Z such that $Q\bar{C}Z = T$ and $Q\bar{A}Z = S$ with T and S triangular.[30] Then eigenvalues of $\bar{C}$ with respect to $\bar{A}$ are defined as t_{ii}/s_{ii}. Some of these are "infinite," in the sense that some s_{ii} may equal zero. This will be the case, indeed, whenever A and therefore $\bar{A}$ are of less than full rank since then S is also singular. All of the foregoing is true for any ordering of the eigenvalues and associated columns of Z (and rows of Q). For the present, let us focus on the arrangement that places the t_{ii}/s_{ii} in order of decreasing modulus.[31]

To begin the analysis, premultiply equation (A1) by Q. Since $Q\bar{A} = SH$ and $Q\bar{C} = TH$, where $H \equiv Z^{-1}$, the resulting equation can be written as

$$\begin{bmatrix} S_{11} & 0 \\ S_{21} & S_{22} \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}\begin{bmatrix} \Omega^2 \\ \Omega \end{bmatrix} = \begin{bmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}\begin{bmatrix} \Omega \\ I \end{bmatrix}. \tag{A2}$$

The first row of (A2) reduces to

$$S_{11}\left(H_{11}\Omega + H_{12}\right)\Omega = T_{11}\left(H_{11}\Omega + H_{12}\right). \tag{A3}$$

Then, if H_{11} is invertible, the latter can be used to solve for Ω as

$$\Omega = -H_{11}^{-1}H_{12} = -H_{11}^{-1}\left(-H_{11}Z_{12}Z_{22}^{-1}\right) = Z_{12}Z_{22}^{-1}, \tag{A4}$$

where the second equality comes from the upper right-hand submatrix of the identity $HZ = I$, provided that H_{11} is invertible, which we assume without significant loss of generality.[32,33]

---

[30] Provided only that there exists some λ for which $\det[\bar{C} - \lambda\bar{A}] \neq 0$. See Golub and Van Loan (1996, 377) or Klein (2000). Note that in McCallum (2007) the matrices $\bar{A}$ and A are denoted A and A_{11}, respectively.

[31] The discussion proceeds as if none of the t_{ii}/s_{ii} equals 1.0 exactly. If one does, the model can be adjusted by multiplying some relevant coefficient by (e.g.) 0.9999.

[32] This invertibility condition, also required by King and Watson (1998) and Klein (2000), obtains except for degenerate special cases of (1) that can be solved by simpler methods than considered here. Note that the invertibility of H_{11} implies the invertibility of Z_{22}, given that Z and H are unitary.

[33] Note that it is not being claimed that all solutions are of the form (9).

As mentioned above, there are many solutions Ω to (26). These correspond to different arrangements of the eigenvalues, which result in different groupings of the columns of Z and therefore different compositions of Z_{12} and Z_{22}. Here, with the eigenvalues t_{ii}/s_{ii} arranged in order of decreasing modulus, the diagonal elements of S_{22} will all be nonzero provided that S has at least m nonzero eigenvalues, which we assume to be the case.[34] Clearly, for any solution under consideration to be dynamically stable, the eigenvalues of Ω must be smaller than 1.0 in modulus. In McCallum (2007) it is shown that

$$\Omega = Z_{22} S_{22}^{-1} T_{22} Z_{22}^{-1}, \tag{A5}$$

so Ω has the same eigenvalues as $S_{22}^{-1}T_{22}$. The latter is triangular, moreover, so the relevant eigenvalues are the m smallest of the $2m$ ratios t_{ii}/s_{ii} (given the decreasing-modulus ordering). For dynamic stability, the modulus of each of these ratios must then be less than 1. (In many cases, some of the m smallest moduli will equal zero.)

Let us refer to the solution under the decreasing-modulus ordering as the MOD solution. Now suppose that the MOD solution is stable. For it to be the only stable solution, there must be no other arrangement of the t_{ii}/s_{ii} that would result in a Ω matrix with all eigenvalues smaller in modulus than 1.0. Thus, each of the t_{ii}/s_{ii} for $i = 1, \ldots, m$ must have modulus greater than 1.0, some perhaps infinite. Is there some m×m matrix whose eigenvalues relate cleanly to these ratios? Yes, it is $F \equiv (I - A\Omega)^{-1}A$, which appears frequently in Binder and Pesaran (1995).[35] Regarding F, it is shown that, for any ordering such that H_{11} is invertible, including the MOD ordering, we have the equality

$$H_{11}\ F\ H_{11}^{-1} = T_{11}^{-1} S_{11}, \tag{A6}$$

which implies that F has the same eigenvalues as $T_{11}^{-1}S_{11}$. In other words, the eigenvalues of F are the same, for any given arrangement, as the *inverses* of the values of t_{ii}/s_{ii} for $i = 1, \ldots, m$. Under the MOD ordering these are the inverses of the first (largest) m of the eigenvalues of the system's matrix pencil. Accordingly, for solution (A4) to be the only stable solution, all the eigenvalues of the corresponding F must be smaller than 1.0 in modulus. This

[34] From its structure it is obvious that $\bar{A}$ has at least m nonzero eigenvalues so, since Q and Z are nonsingular, S must have rank of at least m. This necessary condition is not sufficient for S to have at least m nonzero eigenvalues, however; hence, the assumption.

[35] There is no general proof of invertibility of $[I - A\Omega]$, but if $A\Omega$ were by chance to have some eigenvalue exactly equal to 1.0, that condition could be eliminated by making some small adjustment to elements of A or C.

is a generalization of a result of Blanchard and Khan (1980) for a model with nonsingular A.

Thus, we have established notation and reported results showing that the existence of a unique stable solution requires that all eigenvalues of the defined Ω matrix and the corresponding F must be less than 1.0 in modulus. It will be convenient to express that condition as follows: all $|\lambda_\Omega| < 1$ and all $|\lambda_F| < 1$.

APPENDIX B

To illustrate some concepts pertaining to sunspot equilibria, consider the simple univariate model

$$x_t = aE_t x_{t+1} + u_t \qquad a \neq 0, a \neq 1, \tag{B1}$$

where u_t is generated by an AR(1) process with AR parameter ρ, assuming $|\rho| < 1$.[36] Then the fundamental RE solution will be of the form $x_t = \phi u_t$ so $E_t x_{t+1} = \rho\phi u_t$ and the undetermined-coefficient procedure relationship $\phi u_t = a\rho\phi u_t + u_t$ implies that $\phi = 1/(1 - a\rho)$. Thus, the fundamental solution is

$$x_t = \frac{1}{1 - a\rho} u_t. \tag{B2}$$

To introduce sunspot phenomena, consider solutions of the form

$$x_t = \phi_1 x_{t-1} + \phi_2 u_t + \phi_3 \xi_t, \tag{B3}$$

where ξ_t is a "sunspot" variable—i.e., an extraneous variable generated by any stochastic process such that $E_{t-1}\xi_t = 0$. Then we have

$$E_t x_{t+1} = \phi_1 \left(\phi_1 x_{t-1} + \phi_2 u_t + \phi_3 \xi_t\right) + \phi_2 \rho u_t + 0, \tag{B4}$$

and substitution of (B3) and (B4) into (B1) yields

$$\phi_1 x_{t-1} + \phi_2 u_t + \phi_3 \xi_t = a\left[\phi_1\left(\phi_1 x_{t-1} + \phi_2 u_t + \phi_3 \xi_t\right) + \phi_2 \rho u_t\right] + u_t. \tag{B5}$$

[36] Symbols are used here with meanings potentially different from those in the body of the article and in Appendix A.

For the latter to hold for all values of x_{t-1}, u_t, and ξ_t—i.e., to be a solution—it is necessary that

$$\phi_1 = a\phi_1^2, \tag{B6a}$$

$$\phi_2 = a\phi_1\phi_2 + \phi_2 a\rho + 1, \text{ and} \tag{B6b}$$

$$\phi_3 = a\phi_1\phi_3. \tag{B6c}$$

The first of these equations has two solutions, $\phi_1 = 1/a$ and $\phi_1 = 0$. The latter gives the fundamentals solution, but the former gives other solutions. Thus, with $\phi_1 = 1/a$, (B6b) becomes $\phi_2 = -1/a\rho$, while (B6c) reduces to $\phi_3 = \phi_3$ (i.e., is satisfied by any finite value). Accordingly, there are sunspot solutions satisfying

$$x_t = (1/a)\, x_{t-1} - (1/a\rho)\, u_t + \phi_3\xi_t, \tag{B7}$$

where ϕ_3 can be any real number. If $| \ a \ | < 1$, these solutions are explosive, but we have an infinity of stable sunspot solutions if $| \ a \ | > 1$.

The learnability of solutions (B2) and (B7) has been studied by E&H (2003), who show that the fundamentals solution is E-stable if $a < 1$ and is not E-stable if $a > 1$, whereas the sunspot solutions (B7) are not E-stable (or learnable) for any value of a.

E&H (2003) also consider, however, a special class of K-state Markov sunspots such that $x_t = \bar{x}\,(i)$ when the state variable s_t equals the constant $s(i)$, for $i = 1, 2, \ldots, K$, and evolves in accordance with a K-state Markov process with fixed transition probabilities. In this case, it transpires that non-fundamental solutions can be E-stable, even though such solutions can be written as special cases of the form (B3). That seemingly contradictory result is actually compatible with the result of the previous paragraph because adoption of the K-state Markov specification places additional structure on the system that is not implied by (B3), and consequently uses, in effect, a different "perceived law of motion" for the learnability analysis. E&H remark that "... one way to view our results is that the learning processes are attempting to learn different things for the different representations: for the AR(1) form of SSEs the learning rule corresponds to least squares estimation of the coefficients, while for the two-state Markov representation of 2-SSEs the learning scheme in effect estimates the support of the distribution" (2003, 179). My reasons for rejecting these solutions are developed in Section 5.

REFERENCES

Bean, Charles, and Nigel Jenkinson. 2001. "The Formulation of Monetary Policy at the Bank of England." *Bank of England Quarterly Bulletin* 41 (Winter): 434–41.

Bernanke, Ben S., and Michael Woodford. 1997. "Inflation Forecasts and Monetary Policy." *Journal of Money, Credit, and Banking* 24 (November): 653–84.

Binder, Michael, and M. H. Pesaran. 1995. "Multivariate Rational Expectations Models and Macroeconometric Modelling: A Review and Some New Results." In *Handbook of Applied Econometrics*, edited by M. H. Pesaran and M. Wickens. Oxford, U.K.: Basil Blackwell Publishers.

Blanchard, Olivier, and Charles M. Kahn. 1980. "The Solution of Linear Difference Models Under Rational Expectations." *Econometrica* 48 (July): 1,305–11.

Bullard, James. 2006. "The Learnability Criterion and Monetary Policy." Federal Reserve Bank of St. Louis *Review* 88 (May): 203–17.

Bullard, James, and Kaushik Mitra. 2002. "Learning About Monetary Policy Rules." *Journal of Monetary Economics* 49 (September): 1,105–29.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115 (February): 147–80.

Evans, George W. 1986. "Selection Criteria for Models with Non-Uniqueness." *Journal of Monetary Economics* 18 (September): 147–57.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics*. Princeton, N.J.: Princeton University Press.

Evans, George W., and Seppo Honkapohja. 2003. "Expectational Stability of Stationary Sunspot Equilibria in a Forward-Looking Linear Model." *Journal of Economic Dynamics and Control* 28 (October): 171–81.

Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton, N.J.: Princeton University Press.

Goodfriend, Marvin. 2005. "Inflation Targeting in the United States?" In *The Inflation Targeting Debate*, edited by B. S. Bernanke and M. Woodford. Chicago: University of Chicago Press.

Goodfriend, Marvin. 2007. "How the World Achieved Consensus on Monetary Policy." *Journal of Economic Perspectives* 21 (Fall): 47–68.

Honkapohja, Seppo, and Kaushik Mitra, 2004. "Are Non-Fundamental Equilibria Learnable in Models of Monetary Policy?" *Journal of Monetary Economics* 51 (November): 1,743–70.

King, Robert G. 2000. "The New IS-LM Model: Language, Logic, and Limits." Federal Reserve Bank of Richmond *Economic Quarterly* 86 (Summer): 45–103.

King, Robert G., and Mark W. Watson. 1998. "The Solution of Singular Linear Difference Systems Under Rational Expectations." *International Economic Review* 39 (November): 1,015–26.

Klein, Paul, 2000. "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model." *Journal of Economic Dynamics and Control* 24 (September): 1,405–23.

Lubik, Thomas A., and Frank Schorfheide. 2004. "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* 94 (March): 190–217.

Lucas, Robert E., Jr. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4 (April): 103–24.

McCallum, Bennett T. 1983. "On Non-Uniqueness in Rational Expectations Models: An Attempt at Perspective." *Journal of Monetary Economics* 11: 139–68.

McCallum, Bennett T. 2001. "Monetary Policy Analysis in Models Without Money." Federal Reserve Bank of St. Louis *Review* 83 (July): 145–60.

McCallum, Bennett T. 2003. "Multiple-Solution Indeterminacies in Monetary Policy Analysis." *Journal of Monetary Economics* 50 (July): 1,153–75.

McCallum, Bennett T. 2007. "E-Stability vis-a-vis Determinacy Results for a Broad Class of Linear Rational Expectations Models." *Journal of Economic Dynamics and Control* 31 (April): 1,376–91.

McCallum, Bennett T. 2008. "How Important is Money in the Conduct of Monetary Policy? A Comment." *Journal of Money, Credit and Banking* 40 (December): 1,783–90.

Svensson, Lars E. O. 1997. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." *European Economic Review* 41 (June): 1,111–46.

U.K. Treasury. 2003. *Policy Frameworks in the U.K. and EMU*. London: HM Treasury.

Walsh, Carl E. 2003. *Monetary Theory and Policy*, 2nd ed. Cambridge, Mass.: MIT Press.

Woodford, Michael. 1994. "Nonstandard Indicators for Monetary Policy: Can Their Usefulness Be Judged From Forecasting Regressions?" In *Monetary Policy*, edited by N. G. Mankiw. Chicago: University of Chicago Press.

Woodford, Michael. 2003a. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.

Woodford, Michael, 2003b. "Comment on: Multiple-Solution Indeterminacies in Monetary Policy Analysis." *Journal of Monetary Economics* 50 (July): 1,177–88.

Woodford, Michael. 2008. "How Important is Money in the Conduct of Monetary Policy?" *Journal of Money, Credit and Banking* 40 (December): 1,561–98.

Economic Quarterly—Volume 95, Number 1—Winter 2009—Pages 53–74

Semiparametric Estimation of Land Price Gradients Using Large Data Sets

Kevin A. Bryan and Pierre-Daniel G. Sarte

Traditional urban theory typically predicts land values that form a smooth and convex surface centered at a central business district (CBD) (Mills 1972 and Fujita 1989). The fact that land values are highest near the city center reflects a trade off between commuting costs and agglomeration externalities at the CBD. As distance from the city center increases, so do commuting costs for workers employed at the CBD. Agglomeration effects, however, such as knowledge sharing or decreased shipment costs from a common port, are highest near the CBD. In equilibrium, therefore, the price of residential land tends to be bid up most forcefully close to the city center where commuting costs are lowest. In empirical work, the shape of the land price surface is often estimated using a parametric regression that includes a measure of Euclidian distance from the CBD or a polynomial function of location data. The parameter associated with distance from the CBD, then, captures the rate at which land prices decline as one moves away from the city center and toward the rural outskirts. Though this is a straightforward method to obtain estimates of the rate of price decline, parametric methods can be misleading for two reasons.

First, as noted by Seyfried (1963) among others, cities are "not a featureless plain." Bodies of water, mountains, and geography more generally all distort the land price surface by influencing potential commuting patterns. Second, and more importantly, there is growing evidence, both theoretical and empirical, that the monocentric city of Mills (1972), for example, is

We wish to thank Nadezhda Malysheva, Yash Mehra, and Ray Owens for their comments. We also thank Sarah Watt for helpful research assistance. The views expressed in this article do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System. All errors are our own. E-mail: pierre.sarte@rich.frb.org.

being replaced by the polycentric city, where employment subcenters lead to land price gradients of a form that may be difficult to uncover parametrically. Anas, Arnott, and Small (1998) survey this literature, while Redfearn (2007) provides an example of the employment density surface in Los Angeles. A parametric model of such a city may smooth over important employment subcenters and high-price suburbs. As such, nonparametric estimates of the land price surface allow for a more robust description of the data.

Estimation of land gradients using nonparametric or semiparametric methods is somewhat involved relative to parametric regressions. In part, to economize on computations, early work in this area has tended to use only vacant lot sales (Colwell and Munneke 2003), but the sparseness of that data can lead to overly smooth price gradients. Furthermore, vacant lot sales are not as informative when considering land prices outside of dense urban cores, as there may be large areas without any nearby sale during the period studied. In contrast, the number of residential house sales in a given period can be substantial. This article, therefore, reviews a method for constructing land price gradients using a potentially large set of housing sales data. Drawing on work by Yatchew (1997) and Yatchew and No (2001), we estimate a semiparametric hedonic housing price equation where the contribution of housing attributes to home prices is obtained parametrically, but the component of home prices that varies with location is not assumed to lie in a given parametric family.

Using data from 2002–2006, we apply this method to the city of Richmond, Virginia, and three surrounding counties. The region under consideration covers approximately 1,218 square miles, comprises nearly one million people, and has boundaries that are agricultural in nature. Since our technique uses home transaction sales, and not simply vacant land, we are able to construct a land gradient from over 100,000 observations. Surprisingly, given the recent trend toward polycentric cities in the United States, we find that the price surface in Richmond is largely monocentric, with land prices falling from over \$100 per square foot (in 2006 constant dollars) around the CBD to less than \$1 per square foot in the rural outskirts. Though the CBD is the dominant feature on the price surface, larger suburbs, such as Mechanicsville, Ashland, Short Pump, and Midlothian, and corridors along Interstates 64 and 95, are easily identified. Furthermore, the presence of these subcenters distorts estimates of parametric surfaces even when they assume one dominant center.

An exponential function fitted to our estimates of land prices reveals that prices fall, on average, at the rate of 2.8 percent per mile as one moves away from Richmond's CBD. Put another way, land prices fall by $\frac{1}{2}$ every 25 miles. This rate is significantly higher than the rate of decline estimated with a least squares regression of home prices on housing characteristics and a measure of distance from the CBD, which finds prices falling at only 1 percent per mile. This difference arises because conventional parametric methods do not allow for local variations in housing prices and, consequently, achieve a considerably worse fit over a large area. In particular, the parametric regression is associated with a much poorer fit of the data relative to its semiparametric counterpart.

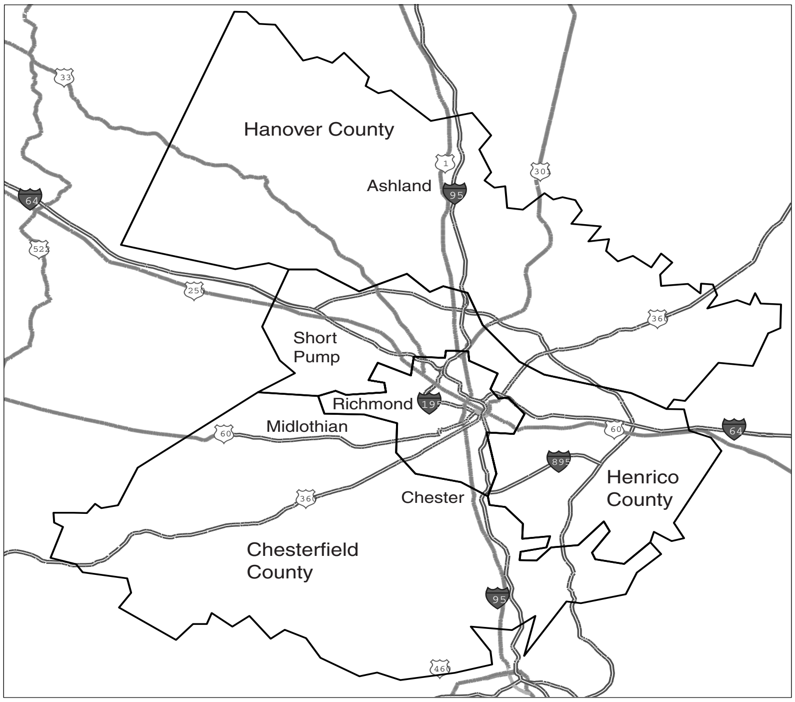
The technique for estimating land prices proposed in this article makes no assumptions about the geography of the Richmond area or the structure of the land price surface. In a parametric estimation of land prices, the inclusion of variables to represent location within a certain county or distance from an identified "employment subcenter" (such as Chicago O'Hare airport in Colwell and Munneke [1997]) is essential to achieving a reasonably accurate land price surface. However, identifying locations such as an employment subcenter can be a difficult and arbitrary task (Giuliano and Small [1991] and McMillen [2001]). Moreover, independent of commuting considerations, proximity to geographical features such as lakes may enhance the value of certain locations. By construction, this feature of land prices cannot be captured by parametric methods based on distance from a CBD. In contrast, because no arbitrary decisions about the functional form of the land surface need to be made with the method used in this article, it can be directly applied to any urban region of any shape and size.

The rest of the article proceeds as follows. In Section 1, we describe the data and features of the Richmond area. Section 2 describes the empirical model and discusses how semiparametric land price estimates are computed. In Section 3, we construct the land price surface and compare our estimates with those constructed using simpler polynomial and distance-from-CBD estimates. Section 4 concludes.

1. DATA DESCRIPTION

This article estimates the land price gradient from a full sample of residential sales in the city of Richmond and three nearby counties—Hanover, Henrico, and Chesterfield—from 2002–2006. Richmond is a mid-sized regional center with a population just over 200,000, lying 100 miles south of Washington, D.C., at the intersection of Interstates 95 and 64. The urban core was well developed by the late 19th century, when the city served as the Confederate capital during the Civil War. As such, the mean age of the housing stock in the city itself is more than 66 years. Hanover County lies due north of Richmond, with a largely rural population of less than 100,000, though significant suburbs do line the Interstate 95 corridor. Henrico County lies both to the west and to the east of Richmond, with a population just under 300,000, and is home to a number of quickly growing suburbs surrounding Interstate 64, notably the areas around Short Pump and Mechanicsville. Chesterfield County, with a population over 300,000, lies to the south of Richmond and is primarily made up of low density suburban areas, along with a few notable small towns such as Chester and Midlothian. A map of Richmond and surrounding counties is given in Figure 1. All told, the region includes almost one million people

Figure 1 Richmond and Surrounding Counties



residing in over 1,218.5 square miles. Aside from the far southern end of Chesterfield County, which abuts the cities of Colonial Heights and Petersburg, the edge of this region consists of rural farmland.

We acquired a full record of property sales, with matched housing characteristics, from the city and counties. These characteristics include the furnished square footage of a house, the number of years since the house was first built, its plot acreage, and the number of bathrooms available. We also include binary variables that indicate whether a house has air conditioning, whether its exterior is made of brick, vinyl, or wood, and whether it is heated using oil, hot water, or central air. Before carrying out the estimation, we filter the data along several dimensions. First, all nonresidential properties were removed, as the parametric portion of our estimation requires data on, for instance, livable square footage. Next, we remove houses that appear to have improperly entered data—this includes houses with construction dates before 1800, houses with sales prices of less than one dollar, houses that appear to have been sold in a lot where no breakdown of the sales price per house

Table 1 Data Summary

Variable	Mean	St. Dev.	Min.	Max.
Sales Price[a]	210,068.63	152,035.95	6.46	5,639,195.45
Age	33.14	28.61	1	207
No. of Bathrooms	2.13	0.90	0	72
Air Conditioning	0.66	0.47	0	1
Square Footage	1,876.0	903.4	319	63,233
Lot Acreage	0.70	2.44	0.02	98.77
Brick Exterior	0.20	0.40	0	1
Vinyl Exterior	0.31	0.46	0	1
Wood Exterior	0.18	0.38	0	1
Gas Heating	0.09	0.29	0	1
Oil Heating	0.17	0.38	0	1
Hot Water Heating	0.14	0.35	0	1
Central Air Heating	0.13	0.33	0	1

Notes: [a]Expressed in constant 2006 dollars.

is available, and houses with plot acreage lower than .02 acres (871 square feet).

We geocode data from street addresses in ArcGIS using SPCS NAD83 coordinates, which, unlike simple latitude and longitude, allow easy calculation of Euclidean distance in feet between any two points. In some cases, the geocoder was unable to positively identify a street address; unidentified houses are left out of the final sample. Descriptive statistics of our data are reported in Table 1.

2. THE EMPIRICAL MODEL

This section sets up the basic framework we use throughout the remainder of the article. We denote the city of Richmond (and its four surrounding counties) by C and a location in the city by $\ell = (x, y) \in C$, where x and y are Cartesian coordinates. We denote the (log) per-square-foot price of a home in Richmond by p. Our analysis begins with the following hedonic price equation,

$$p = \mathbf{X}\beta + f(\ell) + \varepsilon, \tag{1}$$

where $\mathbf{X}$ is a k-element vector of conditioning housing characteristics such that $cov(\mathbf{X}|\ell) = \Sigma_{\mathbf{X}|\ell}$, $f(\ell)$ is the component of a home price directly related to location, and ε is a random variable such that $E(\varepsilon|\ell, \mathbf{X}) = 0$ and $var(\varepsilon|\ell) = \sigma^2_\varepsilon$. The matrix $\mathbf{X}$ consists of all of the variables from Table 1 and quadratic terms for lot acreage and square footage, as Brownstone and De Vany (1991), among others, find that land price per acre is a concave function of parcel size. The coefficients, β, capture the reduced-form effects of particular housing attributes, such as the size of the living area of a house or whether air conditioning is available, on home prices. Moreover, since $p-\mathbf{X}\beta$ represents housing prices purged of the contribution from specific attributes, we think of $f(\ell)$ as capturing the value of land per-square-foot at a given location. While this general semi-log specification is standard in the analysis of real estate data, some differences exist regarding the functional form that describes the function $f(.)$. One option is to specify $f(.)$ as a polynomial function of location data, as in Galster, Tatian, and Accordino (2006),

$$f(x, y) = \alpha_0 x + \alpha_1 y + \alpha_3 x^2 + \alpha_4 y^2 + \alpha_5 xy. \tag{2}$$

Substituting equation (2) into equation (1), one can then consistently estimate the coefficients β and α_i, $i = 1, .., 5$, using least squares.

An alternative approach that uses least squares estimation is to parameterize $f(.)$ as a function of distance from the CBD, as in Zheng and Khan (2008),

$$f(x, y) = \alpha_d \sqrt{(x - x_c)^2 + (y - y_c)^2}, \tag{3}$$

where $\ell_c = (x_c, y_c)$ denotes the location of the CBD. Recalling that p is measured in log units, α_d then captures the exponential rate of change in land values as one moves away from the CBD.

In contrast to either of these approaches, this article does not assume that $f(\ell)$ lies in a given parametric family. The only restriction that we shall impose on $f(\ell)$ is that it satisfies a Lipschitz condition,

$$||f(\ell_a) - f(\ell_b)|| < L||\ell_a - \ell_b||, L \geq 0. \tag{4}$$

Semiparametric Regression

Estimating the nonparametric component of equation (1), $f(\ell)$, requires that we first address estimation of the parametric effects, β. One strategy would be to estimate equation (1) in two stages, first ignoring the nonparametric component, $f(\ell)$, in order to obtain estimates of β by regressing p on $\mathbf{X}$, and then applying nonparametric methods to purged home prices, $p - \mathbf{X}\hat{\beta}$, where $\hat{\beta}$ denotes the previously obtained estimates of β. However, since the reduced form model contains a component related to location that is being ignored, estimates of β obtained in this way will be biased when housing attributes, $\mathbf{X}$, are correlated with location, ℓ. Rather, a two-step estimation strategy must somehow "get rid" of the nonparametric component in the first step.

Let n denote the number of observations in our data set. A popular approach, pioneered by Robinson (1988), recognizes as a first step that equation (1) implies that

$$p - E(p|\ell) = [\mathbf{X} - E(\mathbf{X}|\ell)]\beta + \varepsilon. \tag{5}$$

In other words, the conditional differencing of equation (1) gets rid of the nonparametric component. Robinson (1988) then shows that by replacing $E(p|\ell)$ and $E(\mathbf{X}|\ell)$ with nonparametric kernel estimates (to be described below) $\widehat{E}(p|\ell)$ and $\widehat{E}(\mathbf{X}|\ell)$, respectively, and then regressing $p-\widehat{E}(p|\ell)$ on $[\mathbf{X}-\widehat{E}(\mathbf{X}|\ell)]$, yields estimates of β that are $\sqrt{n}$ consistent. Unfortunately, this method can be quite onerous since separate nonparametric regressions are required for each housing attribute in $\mathbf{X}$, where both the number of relevant housing attributes and the number of observations are large in our case. To avoid this problem, we summarize instead a differencing method developed more recently by Yatchew (1997) and Yatchew and No (2001), and adopted, for example, in Rossi-Hansberg, Sarte, and Owens (2008).

The basic idea behind differencing the data works as follows. We would like to re-order our data, $(p_1, \mathbf{X}_1, \ell_1), (p_2, \mathbf{X}_2, \ell_2), \ldots, (p_n, \mathbf{X}_n, \ell_n)$ so that the ℓ's are close, in which case differencing tends to remove the nonparametric effects. To get a sense of the implications of differencing, suppose that locations constitute a uniform grid on the unit square (the re-scaling is without loss of generality). Each point may then be thought of as taking up an area of $\frac{1}{n}$, and the distance between adjacent observations is therefore $\frac{1}{\sqrt{n}}$. Suppose further that the data have been re-ordered so that $||\ell_i - \ell_{i-1}|| = \frac{1}{\sqrt{n}}$. First-differencing of (1) then yields

$$p_i - p_{i-1} = (\mathbf{X}_i - \mathbf{X}_{i-1})\beta + f(\ell_i) - f(\ell_{i-1}) + \varepsilon_i - \varepsilon_{i-1}. \tag{6}$$

Assuming that equation (4) holds, the difference in nonparametric components in (6) vanishes asymptotically. Yatchew (1997) then shows that the ordinary least squares estimator of β using the differenced data (i.e., by regressing $p_i - p_{i-1}$ on $\mathbf{X}_i - \mathbf{X}_{i-1}$) is also $\sqrt{n}$ consistent. This estimator of β, however, achieves only $\frac{2}{3}$ efficiency relative to the one produced by Robinson's method. This can be improved dramatically by way of higher-order differencing. Specifically, define $\Delta\mathbf{p}$ to be the $(n-m)\times 1$ vector whose elements are $[\Delta\mathbf{p}]_i = \sum_{s=0}^{m} \omega_s p_{i-s}$, $\Delta\mathbf{X}$ to be the $(n-m)\times k$ matrix with entries $[\Delta\mathbf{X}]_{ij} = \sum_{s=0}^{m} \omega_s X_{i-s,j}$, and similarly for $\Delta\varepsilon$. The parameter m governs the order of differencing and the ω's denote constant differencing weights. Equation (6) can then be generalized as

$$\Delta\mathbf{p} = \Delta\mathbf{X}\beta + \sum_{s=0}^{m} \omega_s f(\ell_{i-s}) + \Delta\varepsilon,\ i = m+1, \ldots, n, \tag{7}$$

where the following two conditions are imposed on the differencing coefficients, $\omega_0, \ldots, \omega_m$:

$$\sum_{s=0}^{m} \omega_s = 0 \text{ and } \sum_{s=0}^{m} \omega_s^2 = 1. \tag{8}$$

The first condition ensures that differencing removes the nonparametric effect in (1) as the sample size increases and the re-ordered ℓ's get closer to each other. The second condition is a normalization restriction that implies that the variance of the transformed residuals in (7) is the same as the variance of the residuals in (1). When the differencing weights are chosen optimally, the difference estimator β_Δ obtained by regressing $\Delta \mathbf{p}$ on $\Delta \mathbf{X}$ approaches asymptotic efficiency by selecting m sufficiently large.[1] In particular, Yatchew (1997) shows that

$$\hat{\beta}_\Delta \sim^A N(\beta, (1 + \frac{1}{2m})\frac{\sigma_\varepsilon^2}{n}\Sigma_{\mathbf{X}|\ell}^{-1}),$$

$$s_\Delta^2 = \frac{1}{n}\sum_{i=1}^{n}(\Delta \mathbf{p}_i - \Delta \mathbf{X}_i \hat{\beta})^2 \to^P \sigma_\varepsilon^2, \text{ and} \tag{9}$$

$$\widehat{\Sigma}_{u,\Delta} = \frac{1}{n}\Delta \mathbf{X}' \Delta \mathbf{X} \to^P \Sigma_{\mathbf{X}|\ell}.$$

These results will allow us to do inference on $\hat{\beta}_\Delta$. By equation (9), the R^2 statistic associated with our original empirical specification (1) can be estimated as $R^2 = 1 - \frac{s_\Delta^2}{s_p^2}$. In our estimation exercise, we use $m = 10$, which produces coefficient estimates that are approximately 95 percent efficient when using optimal differencing weights.

Finally, note that because locations lie in R^2, the initial re-ordering of the ℓ's is not unambiguous in this case. A Hamiltonian cycle over distance is the ordering of housing sale coordinates such that the sum of differenced distances is minimized. However, computing a Hamiltonian cycle for over 100,000 points is not yet tractable on a personal computer.[2] In this article, we re-order locations using a nearest-neighbor algorithm that finds an approximate Hamiltonian cycle in the following way: First, we select an arbitrary starting location from which we then find the location of a sale in our data set nearest to it in Euclidean distance. From this second location, we then find the nearest third sale location among the set of remaining observations (i.e., those not already identified). This process is repeated until every sale location has been selected.[3] Figure 2 displays the path chosen by our algorithm. The median distance between points in our sample is 86 feet. As this figure is darker in areas where there are more residential sales, it also serves as a rough guide to residential density in the Richmond area. The origin represents the CBD of Richmond City.

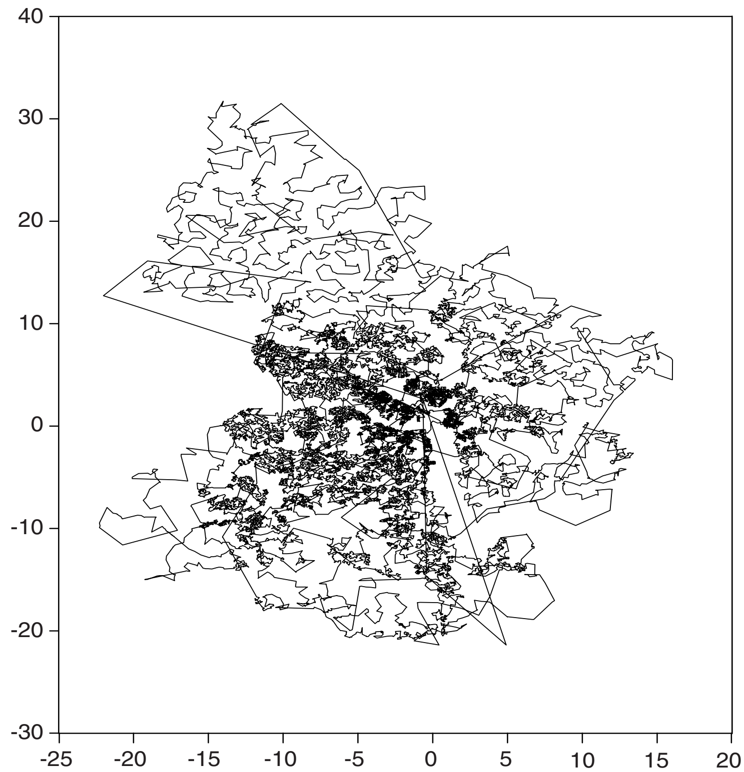
---

[1] Optimal differencing weights, $\omega_0, ..., \omega_m$, solve min $\delta = \sum_{k=1}^{m}(\sum_s \omega_s \omega_{s+k})^2$ subject to the constraints in (7). See Proposition 1 in Yatchew (1997).

[2] The Hamiltonian cycle is the solution to the famous Traveling Salesman Problem (TSP). As of 2007, the largest TSP ever solved on a supercomputer involved 85,900 points, which is smaller than our problem (Applegate et al. 2006).

[3] See Rosenkrantz, Stearns, and Lewis (1977). Although the starting point is arbitrary, it has little implications for our findings.

Figure 2 Path of Approximate Hamiltonian Cycle



Nonparametric Kernel Estimation of $f(\ell)$

Denote by z the price of a home "purged" of its contribution from housing characteristics, where z is obtained using first stage estimates, $z = p - \mathbf{X}\hat{\beta}_{\Delta}$, and construct the data (z_1, ℓ_1), (z_2, ℓ_2),...,(z_n, ℓ_n). Then, because $\hat{\beta}_{\Delta}$ is a consistent estimator of β, the consistency of $f(\ell)$ obtained using standard kernel estimation methods applied to purged home prices remains valid.

The *Nadaraya-Watson* kernel estimator of f at location ℓ_j is given by

$$f(\ell_j) = n^{-1} \sum_{i=1}^{n} W_{hi}(\ell_j) z_i. \tag{10}$$

In other words, the value of land at location ℓ_j is a weighted-average of the z's in our data sample. The weight, $W_{hi}(\ell_j)$, attached to each purged price,

z_i, is given by

$$W_{hi}(\ell_j) = \frac{K_h(\ell_j - \ell_i)}{n^{-1}\sum_{i=1}^{n} K_h(\ell_j - \ell_i)}, \tag{11}$$

where

$$K_h(u) = h^{-1}K(\frac{u}{h}),$$

and $K(\psi)$ is a symmetric real function such that $\int |K(\psi)|d\psi < \infty$ and $\int K(\psi)d\psi = 1$. Thus, we may choose to attach greater weight to observations on prices of homes located near ℓ_j rather than far away by suitable choice of the function K. In particular, as in much of the literature, our estimation is carried out using the Epanechnikov kernel,

$$K(\frac{u}{h}) = \frac{3}{4}\left(1 - \left(\frac{u}{h}\right)^2\right) I(\left|\frac{u}{h}\right| \leq 1), \tag{12}$$

where $I(.)$ is an indicator function that takes the value 1 if its argument is true and 0 otherwise. The distance between location ℓ_j and some other location ℓ_i in Richmond is simply measured as a Euclidean distance in feet. The kernel in (12) then implies that prices of homes located more than a distance of h feet from ℓ_j will receive a zero weight in the estimation of $f(\ell_j)$. In that sense, the bandwidth, h, has a very natural interpretation in this case. In practice, the estimation of $f(\ell)$ is affected to a greater degree by the choice of bandwidth rather than the choice of kernel.[4]

How then does one choose what bandwidth is appropriate? A seemingly natural method for choosing the bandwidth is to minimize the sum of squared residuals,

$$MSE = n^{-1}\sum_{i=1}^{n}[z_i - f(\ell_i)]^2.$$

However, because z_i is used when estimating $f(\ell_i)$, the mean squared error can be arbitrarily reduced by decreasing the bandwidth until all weight in $f(\ell_i)$ is effectively placed on z_i. To avoid this problem, the cross-validation method proposes that the bandwidth parameter be chosen by minimizing the sum of squared residuals from an alternative kernel regression in which z_i is dropped in the estimation of $f(\ell_i)$. Hence, we select h so that it solves

$$\min_h CV(h) = n^{-1}\sum_{i=1}^{n}[z_i - \widetilde{f}_h(\ell_i)]^2, \tag{13}$$

[4] See DiNardo and Tobias (2001).

where

$$\widetilde{f}_h(\ell_j) = n^{-1} \sum_{i \neq j}^{n} W_{hi}(\ell_j) z_i .$$

We estimate equation (7) using 103,543 observations over the period 2002–2006. All prices are deflated using the consumer price index and measured in 2006 constant dollars. We include among our conditioning variables, **X**, a set of time dummies associated with the sale date of a home that captures secular citywide increases in real home prices, where 2006 is set as the base year.

3. FINDINGS

This section reviews our findings. Before describing the results from the semiparametric estimation, we first present estimates from the polynomial specification of Galster, Tatian, and Accordino (2006), equation (2), and the parameterized distance function of Zheng and Khan (2008), equation (3), as benchmarks.

Table 2 presents estimates from the specification where the value of land is modeled as a quadratic function of location data, as in equation (2), under the heading "Parametric Model 1." The estimation of the coefficients is carried out using least squares. Virtually all housing characteristics are statistically significant at the 5 percent critical level and most are significant at the 1 percent level. The coefficients associated with the sale date are significant over and above prices being measured in constant dollars. In particular, the findings suggest a general increase in real home prices over our sample period (recall that 2006 is set as the base year). In addition, the regression achieves a relatively good fit for cross-sectional data, with an R^2 of about .50.

Of central interest in the first two columns of Table 2 are the parameters that govern the value of land associated with the Cartesian location data. The coefficients of the polynomial in equation (2) are all highly significant, with the exception of the cross-term, which is statistically significant at the 10 percent critical level only. Figure 3 shows the land value surface associated with these parameters. The origin roughly represents the CBD of the city of Richmond, at the intersection of 7th Street and Canal Street. This location corresponds to coordinates within an area generally considered the employment center of Richmond, with high employment density and a preponderance of commercial and office buildings. The polynomial estimate of land prices, however, has a peak nearly 20 miles away from the CBD, roughly located in a far western suburb known as Short Pump. Since the polynomial in equation (2) permits, at most, one local interior maximum, this parametric regression imposes a land price surface typical of a monocentric city. Essentially, the polynomial will choose a maximum in an area for which there exist many house sales with a

Table 2 Modeling Land Prices as Functions of Local Data

	Parametric Model 1		Parametric Model 2		Semiparametric Model	
Variable	**Coeff.**	***t*-Statistics**	**Coeff.**	***t*-Statistics**	**Coeff.**	***t*-Statistics**
2002	−0.41	−52.37	−0.40	−50.10	−0.41	−74.64
2003	−0.32	−42.09	−0.31	−40.11	−0.33	−60.93
2004	−0.22	−29.32	−0.22	−27.81	−0.23	−43.07
2005	−0.09	−12.39	−0.09	−11.37	−0.10	−19.13
Age[a]	0.30	25.40	0.26	20.37	−0.40	−29.65
No. of Bathrooms	0.13	30.81	0.16	35.21	0.04	11.76
Air Conditioning	0.33	61.35	0.36	65.07	0.10	21.26
Sq. Ft.[b]	0.15	31.55	0.18	35.55	0.11	25.29
(Sq. Ft.)^{2}[c]	−0.38	−20.37	−0.45	−23.27	−0.25	−17.61
Acreage	−0.39	−222.58	−0.40	−225.31	−0.31	−161.19
(Acreage)^{2}[d]	0.54	123.50	0.56	124.35	0.38	106.39
Brick Exterior	0.04	6.02	0.04	5.52	−0.01	−1.79
Vinyl Exterior	0.06	10.05	−0.01	−1.48	−0.02	−4.37
Wood Exterior	−0.08	−11.01	−0.07	−9.21	−0.02	−3.99
Gas Heating	0.22	25.16	0.21	22.94	0.09	12.79
Oil Heating	0.04	4.56	0.18	25.03	0.06	6.96
Hot Water Heating	0.24	29.21	0.36	46.12	0.07	8.25
Central Air Heating	0.17	21.67	0.20	23.34	0.03	4.06
x	2.45	10.04				
y	−3.65	−31.88				
x^2	−0.09	−13.87				
y^2	−0.13	−37.07				
xy	−0.03	−3.88				
Distance from CBD			−0.01	−14.63		
R^2	0.50		0.47		0.77	

Notes: [a]Measured in 100 years; [b]measured in 1,000 sq. ft.; [c]measured in 100 million sq. ft. squared; [d]measured in 100 acres squared.

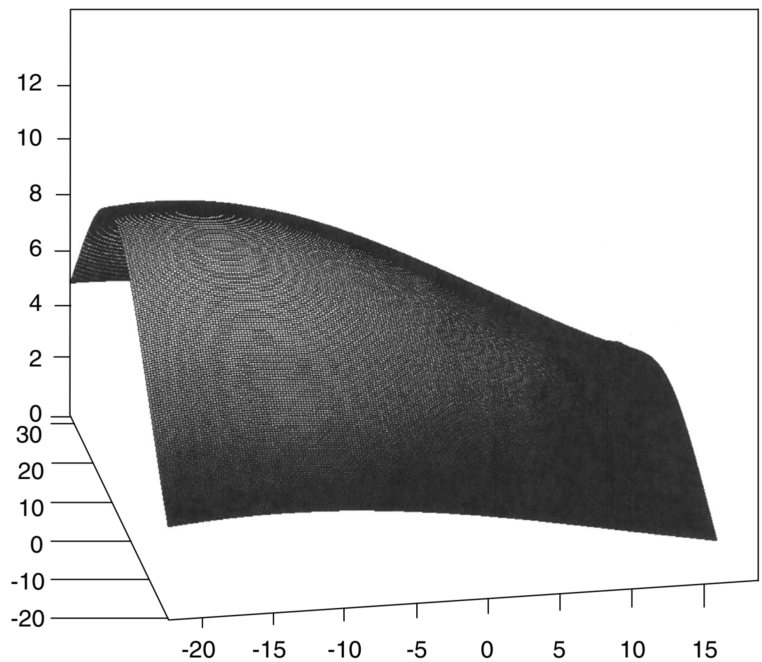
fairly high land value, even if there are other local maxima (near the CBD) on the true land price surface with higher prices but fewer sales.

Table 2 also presents estimates from the parametric specification where land prices (per square foot) are assumed to decline at an exponential rate with distance from the CBD, equation (3), under the heading "Parametric Model 2." The results generally mimic those of our first parametric model, both in terms of the coefficients associated with the different housing attributes and their statistical significance. For example, an additional bathroom adds approximately \$0.13 to the log per-square-foot price of a home under the first specification in Table 2 and \$0.16 under the second parameterization. This translates to an additional bathroom, adding \$2.39 and \$2.68, respectively, to the per-square-foot price of an average home. Note, however, that the parametric specification for Model 2 achieves a slightly worse fit than that for Model 1, with an R^2 statistic of 0.47.

Under the parametric specification including distance from the CBD, we find that the log price per square foot of land declines at a rate of .0103 per mile as one moves away from Richmond's CBD. This translates into a price per square foot of land that falls exponentially at a rate of 1.03 percent per mile as distance from the CBD increases. Alternatively, this exponential rate of decay implies that land prices fall by $\frac{1}{2}$ approximately every 67 miles. This rate of decline in land prices is significantly slower than those estimated by Zheng and Khan (2008) for Beijing, and Colwell and Munneke (1997) for Chicago. The difference occurs because our estimation exercise extends over an area much larger than that covered in the latter two papers, extending 50 miles in diameter in our case, using a similar specification. Because differences in geography are potentially much more pronounced over a larger area, the restriction embedded in the parametric specification related to location will be more stringent and its fit becomes poorer. Figure 4 shows the land price surface associated with our second parametric model. By construction, given the specification in equation (3), this surface reaches a peak at the location that we defined as the center of the CBD, $\ell_c = (x_c, y_c)$ in equation (3). As in the other parametric regression estimated above, this specification imposes a single peak on the land price surface as expected for a monocentric city. Compared to Figure 3, Figure 4 suggests considerably less variation in land prices throughout the city, with a peak price of \$7.60 per square foot at the CBD and approximately \$3.80 at the boundaries of greater Richmond. These figures translate to \$33,106 and \$16,553, respectively, for a 0.1 acre lot.

Given the variation in home prices in greater Richmond depicted in Table 1, this relatively narrow range in estimated land prices implied by Model 2 is potentially surprising. Moreover, the fact that these estimates stem from a parametric regression whose fit is slightly worse than that associated with the first parametric model in Table 2 should leave us somewhat skeptical. As we now discuss, the alternative specification where $f(\ell)$ in equation (1) is treated

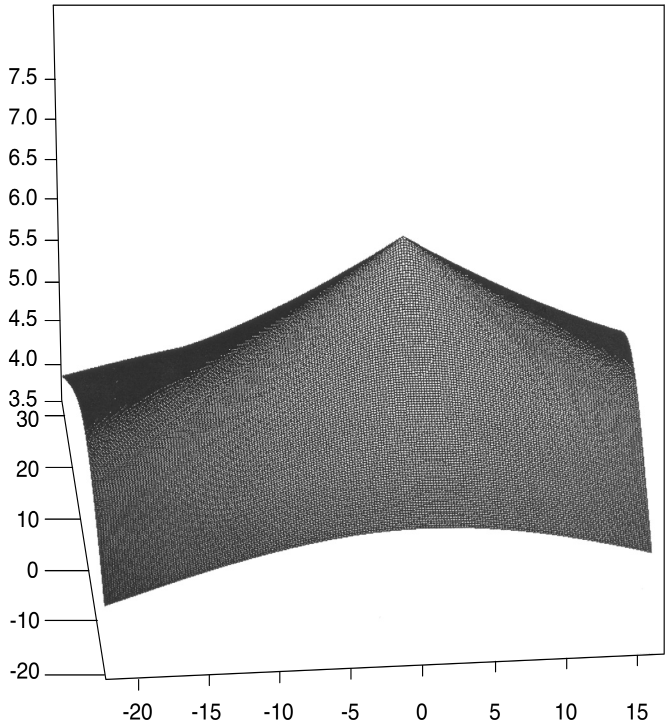
Figure 3 Land Price Surface Estimated from Parametric Model I



nonparametrically yields a significantly better fit, and implies a much more varied and greater range in land prices.

In contrast to the findings from the parametric approaches we have just reviewed, the last two columns of Table 2 present estimates from the semi-parametric method described in Section 2. As before, virtually all variables are statistically significant at the 1 percent critical level, but the magnitude of the coefficients differs somewhat from those of our first two parametric specifications. For example, an additional bathroom now contributes .038 to the log per-square-foot price of a home as opposed to .13 for Model 1 and .16 for Model 2. The difference stems from the fact that we now estimate the component of home prices associated with location nonparametrically. In particular, observe that this semiparametric specification achieves a noticeably better fit relative to the previous two parametric specifications with an R^2 statistic of 0.77 instead of 0.50. Put another way, the semiparametric method

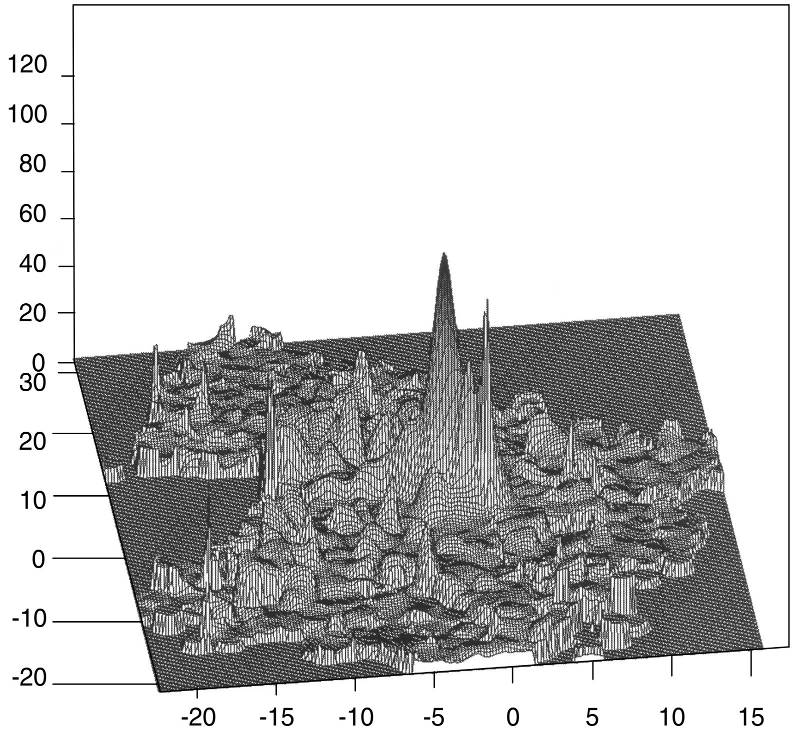
Figure 4 Land Price Surface Estimated from Parametric Model II



adopted here improves the fit of the parametric regressions carried out earlier by almost 60 percent.

Figure 5 illustrates the land price surface obtained using kernel estimation. Evidently, this surface differs considerably from those shown in Figures 3 and 4 along at least two important dimensions. First, this surface displays more variation in land prices across different areas of Richmond than could be obtained from parametric estimates. Observe, for instance, that the West End of Richmond is generally characterized by higher land prices than the area east of the city. The surface also displays multiple local peaks in prices associated with different parts of the city. Second, although the nonparametric estimation identifies one main peak, land prices where this peak is located are as high as \$130 per square foot, in contrast to \$10 per square foot using the polynomial specification in (2). A typical 0.1 acre lot in the most expensive neighborhoods of Richmond, therefore, is estimated at around \$566,280 as opposed to \$43,560 obtained earlier with the polynomial parameterization.

Figure 5 Land Price Surface Estimated from Semiparametric Model

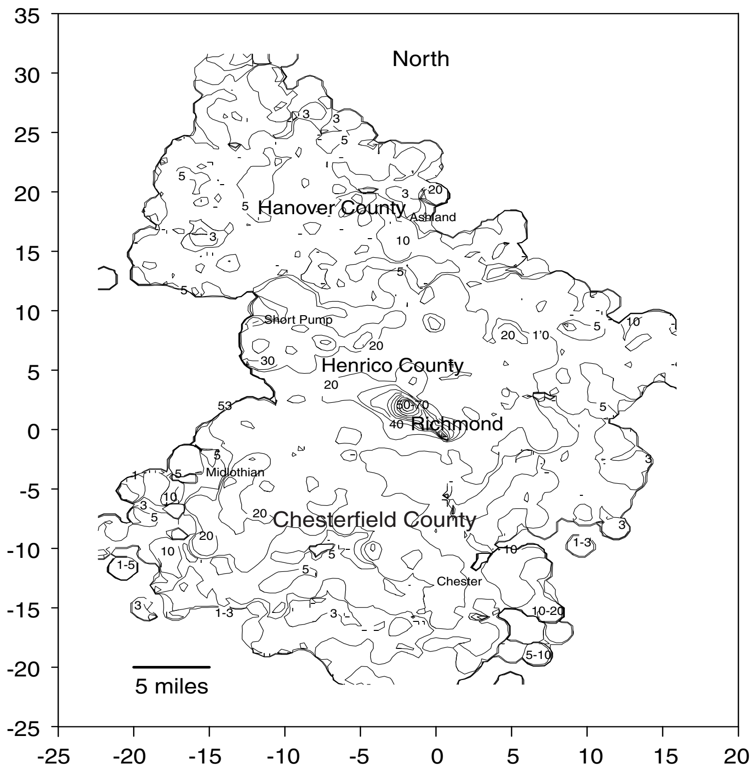


The main reason underlying these differences arises because nonparametric estimation relies on local averaging of the data—sharp peaks and valleys are much more easily discovered with a nonparametric estimation. More specifically, the bandwidth that minimizes the cross-validation criterion in equation (13) is around 5,000 feet in our case. In other words, in estimating land prices at any given location, our procedure uses data within 5,000 feet of that location, with weights that decay quadratically in (12) as one moves away from the point of estimation.

Figure 6 shows the contour map corresponding to the land price surface shown in Figure 5. A main peak is clearly visible just to the north and west of the CBD and corresponds to an area of expensive row houses known as the Fan District in Richmond. Prices in that neighborhood are as high as \$80 to \$130 per square foot.

In contrast, land prices near the boundaries of Richmond range from only \$1 to \$2 per square foot and capture the opportunity cost of land related to

Figure 6 Contour Map of Land Price Surface



agricultural activity at those locations. Local peaks in prices in Figure 6 are also visible five to 15 miles north and west of the city, around areas known as Short Pump and the West End more generally. These areas lie around Interstate 64 and consist of a number of newer suburban neighborhoods generally made up of single-family homes. The contour plot also shows evidence of local peaks extending north from the CBD around the Interstate 95 corridor and mid-sized towns such as Ashland and Mechanicsville. Finally, a series of local maxima can be found 12 miles west and 6 miles south of the CBD, in a region featuring a number of small lakes and golf courses.

Interestingly, despite the variations in land values shown in Figures 5 and 6, our findings suggest that Richmond remains largely a monocentric city. The more recent expansion in activity in the areas of Short Pump, west of the city, is associated with higher land prices on average compared to other areas located a similar distance away from Richmond's CBD. On the whole,

Table 3 Estimates versus Assessments

(Per Square Foot)	Assessments	Estimate
Mean	2.46	9.50
Median	2.02	9.15
Maximum	30.75	33.89
Standard Deviation	1.98	5.18

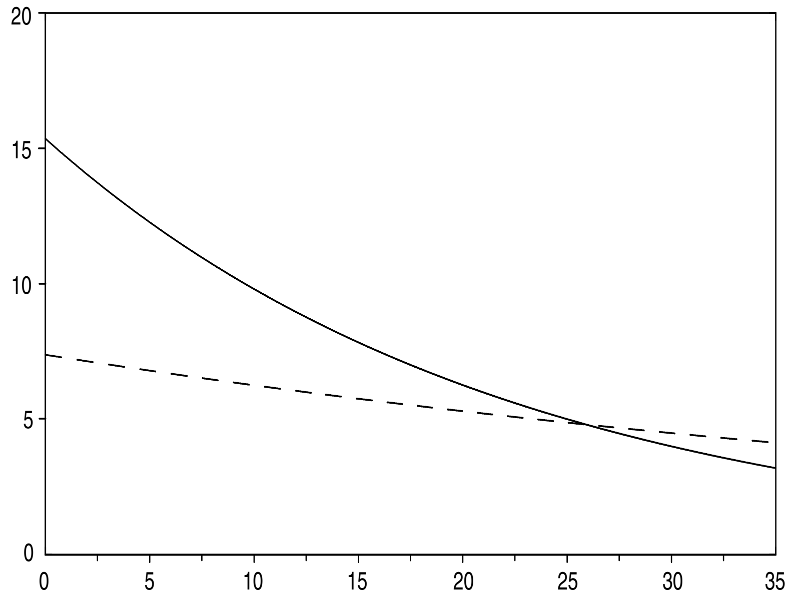
however, land values are highest in the older sections of Richmond near the center of the city.

To compare our estimated land prices with alternative estimates, we obtain land price assessments from Chesterfield County, the county that lies in the southern portion of our region.[5] For tax purposes, Chesterfield County computes assessments both for underlying land value and for improvements. The land value assessment is updated every year and is based on "comparable" vacant land sales from other regions in the county. Though this method allows for much less local variation than our semiparametric approach, it provides a rough estimate of land costs per square foot in the county. Over the period studied, land assessments average just under $3 per square foot, with a number of plots assessed at under $1 per square foot, and the most expensive county land assessed at $20 to $30 per square foot. Table 3 displays characteristics of each distribution. The most expensive plots were located near the golf courses and lakes southwest of Richmond's CBD that were identified as a local maximum in the semiparametric estimation. The assessed value of land tends to be $3 to $6 per square foot less than our semiparametric estimation. This difference reflects, in part, housing characteristics that we are unable to control for in the first step of the nonparametric estimation. For data availability reasons, we cannot remove every piece of a house—for instance, there is no dummy for the number of fireplaces, whether the lot is fenced, the particular layout of the house, the quality of interior materials, etc. The fact that a house exists on every land plot in our sample means that our definition of land is more precisely that of a plot with a zero square foot, zero bathroom house with some unidentified exterior and heating type. Practically, this means that houses in our sample have already installed utility hookups and already have the appropriate zoning for a residential house.

Unless the house is of very poor quality, a plot with those characteristics will be more valuable than empty land, and, therefore, our nonparametric land price estimates will be somewhat biased upwards. However, whatever bias

[5] In many counties, full records of assessment data are not free to obtain; the Chesterfield Assessment Office, however, was able to send us land assessments linked to our housing sales data.

Figure 7 Estimated Land Price Gradient



exists will be evident across all locations, so that differences in the price of land at one location relative to other locations should be similar in both the land assessments and our estimated land price; this indeed appears to be the case for Chesterfield County.

Finally, recall that the parametric specification, including distance from the CBD, delivers a small range of land values and a very shallow price gradient. Figure 7 shows an estimate of the land gradient obtained by projecting our estimates of (log) per-square-foot land prices at a given location onto distance from the CBD at that location. Put another way, we estimate the following equation,

$$\widehat{f}(\ell) = \alpha_0 + \alpha_d d(\ell) + e, \tag{14}$$

where $\widehat{f}(\ell)$ denotes our nonparametric estimate of (log) land value at location ℓ, $d(\ell)$ is the distance to the center of the CBD from ℓ, and α_d represents an exponential rate of change in prices. Our findings suggest a substantially faster rate of decay in land prices (the solid line) than estimated previously for Richmond using the parametric specification in equation (3) (the dashed line). The parametric specification estimated earlier gave a decline of 1.03

percent per mile. In contrast, we now find that α_d is approximately $-.0278$, as opposed to $-.0103$ in Table 2. This implies that land prices decline at a rate of 2.78 percent per mile on average as one moves away from Richmond's CBD. Alternatively, land prices fall by $\frac{1}{2}$ approximately every 25 miles as distance from the CBD increases.

4. CONCLUDING REMARKS

The complexity of the urban price surface means that the assumptions that prices decline monotonically from a CBD or reflect a simple polynomial function of location data are not innocuous. Transit corridors, bodies of water, parkland, golf courses, employment subcenters, and other topographical features can have significant effects on land prices around a city. While these features could in theory be controlled for, it is not straightforward to identify what features or employment centers might be worth identifying.

Drawing on recent work by Yatchew (1997) and Yatchew and No (2001), the semiparametric procedure outlined in this article allows for an approach that does not require a priori assumptions regarding what features of the landscape might affect land prices. It also allows a very large data set—that of all housing transactions in a region—to be used when estimating the land price gradient. Since this procedure does not, unlike earlier work on land prices, rely on local knowledge, it can be applied wholesale to any region or city.

Empirically, an application of this semiparametric approach to land price estimation in Richmond, Virginia, identifies local maxima in the land price surface principally along the Interstate 64 and 95 corridors, in the suburbs of Ashland and Short Pump, and around the lakes and golf courses south of Midlothian. The most expensive land in the region, by a large margin, lies in the historic district of the Fan located close to the CBD; prices in the Fan per acre are over 100 times more expensive than rural land in the surrounding counties.

REFERENCES

Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. "Urban Spatial Structure." *Journal of Economic Literature* 36 (September): 1,426–64.

Applegate, David L., Robert E. Bixby, Vasek Chvátal, and William J. Cook. 2006. *The Traveling Salesman Problem: A Computational Study*. Princeton, N.J.: Princeton University Press.

Brownstone, David F., and Arthur De Vany. 1991. "Zoning, Returns to Scale, and the Value of Undeveloped Land." *Review of Economics & Statistics* 73 (November): 699–704.

Colwell, Peter F., and Henry J. Munneke. 1997. "The Structure of Urban Land Prices." *Journal of Urban Economics* 41 (May): 321–36.

Colwell, Peter F., and Henry J. Munneke. 2003. "Estimating a Price Surface for Vacant Land in an Urban Area." *Land Economics* 79 (February): 15–28.

DiNardo, John, and Justin Tobias. 2001. "Nonparametric Density and Regression Estimation." *Journal of Economic Perspectives* 15 (Fall): 11–28.

Fujita, Masahisa. 1989. *Urban Economic Theory*. Cambridge, U.K.: Cambridge University Press.

Galster, George, Peter Tatian, and John Accordino. 2006. "Targeting Investments for Neighborhood Revitalization." *Journal of the American Planning Association* 72 (Autumn): 457–74.

Giuliano, Genevieve, and Kenneth A. Small. 1991. "Subcenters in the Los Angeles Region." *Regional Science and Urban Economics* 21 (July): 163–82.

McMillen, Daniel P. 2001. "Nonparametric Employment Subcenter Identification." *Journal of Urban Economics* 50 (November): 448–73.

Mills, Edwin S. 1972. *Studies in the Structure of the Urban Economy*. Baltimore: Johns Hopkins University Press.

Redfearn, Christian. 2007. "The Topography of Metropolitan Employment: Identifying Centers of Employment in a Polycentric Urban Area." *Journal of Urban Economics* 61 (May): 519–41.

Robinson, P.M. 1988. "Root-N-Consistent Semiparametric Estimation." *Econometrica* 56: 931–54.

Rosenkrantz, Daniel J., Richard E. Stearns, and Philip M. Lewis II. 1977. "An Analysis of Several Heuristics for the Traveling Salesman Problem." *SIAM Journal of Computing* 6: 563–81.

Rossi-Hansberg, Esteban, Pierre-Daniel G. Sarte, and Raymond E. Owens, III. 2008. "Housing Externalities: Evidence from Spatially Concentrated Urban Revitalization Programs." Federal Reserve Bank of Richmond Working Paper 08-03.

Seyfried, Warren R. 1963. "The Centrality of Urban Land Values." *Land Economics* 39 (August): 275–84.

Yatchew, Adonis. 1997. "An Elementary Estimator of the Partial Linear Model." *Economics Letters* 57 (December): 135–43.

Yatchew, Adonis, and Joungyeo Angelo No. 2001. "Household Gasoline Demand in Canada." *Econometrica* 69 (November): 1,697–709.

Zheng, Siqi, and Matthew E. Khan. 2008. "Land and Residential Property Markets in a Booming Economy: New Evidence from Beijing." *Journal of Urban Economics* 63 (March): 743–57.

Economic Quarterly—Volume 95, Number 1—Winter 2009—Pages 75–100

Consumption Smoothing and the Measured Regressivity of Consumption Taxes

Kartik B. Athreya and Devin Reilly

A maintained assumption of nearly all macroeconomic analysis is that households prefer their consumption to remain smooth across time and states of nature. Their ability to smooth consumption is affected by a variety of constraints, including fiscal policy, and, in particular, the choice of tax base. In practice, labor income and interest income on savings have constituted the bulk of taxed activities. However, the preceding forms of taxation create potentially important distortions. Prescott (2004) shows that labor income taxes may be important in depressing labor supply and average incomes to inefficient levels, while Atkeson, Chari, and Kehoe (1999) show that it can never be optimal to tax capital income in the steady state. In particular, capital income taxes hinder the household's ability to smooth consumption intertemporally by lowering the return on savings.

An alternative tax that avoids the hurdles to smoothing created by capital income taxes is a tax on consumption. In general, however, consumption taxes have been opposed on the basis that they are "regressive" in the sense that, at any point in time, the revenues may be disproportionately collected from households whose incomes are lower than average. Households in the U.S. economy also face substantial persistent and uninsurable idiosyncratic risks to their income (see, e.g., Storesletten, Telmer, and Yaron [2004]). As a result, many with currently low income will be those who have suffered an adverse shock in the past. From this perspective, a tax system that collects a

The views expressed in this article do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: kartik.athreya@rich.frb.org.

substantial portion of its revenues from those who find themselves with low income may seem undesirable.

Evaluating the burden of tax payments by income requires choosing a definition of income by which to order households. Two candidates are (i) income received in a given year and (ii) income realized over the lifetime. For each of these definitions, one can compute the regressivity of a given tax regime. The first measure of incidence, which we term *annual incidence,* will compare the cumulative contributions to tax collections of households collected at a point in time, and then ranked by current (annual) income. The second, which we refer to as *lifetime incidence*, will compare the cumulative contributions to tax collections of households ranked by their realized lifetime income.

Some have noted that the measured incidence of consumption taxes depends on the notion of income being used. Notably, Metcalf (1997) shows that while the annual incidence of consumption taxes appears regressive, the lifetime incidence is roughly proportional. In particular, when income is deterministic and has a "hump" at middle age, relatively young households can expect income to grow, while relatively old households can expect income to fall. Under the presumption that households prefer smooth consumption, the young will generally borrow, if allowed, while the old will run down assets to finance consumption in retirement. This behavior implies that households will consume large amounts relative to their income when young while the reverse will hold when old. As a result, any cross-sectional assessment of "who pays" a consumption tax will conclude that it is paid disproportionately by the currently relatively poor. However, this apparent regressivity is merely an artifact of households successfully achieving smooth consumption.

The preceding intuition was derived in a purely deterministic setting. However, the logic extends to the more general case where income has both deterministic and stochastic components. In stochastic settings, people in any cross-sectional data will differ even if they share many characteristics such as age, gender, and education. However, the nature of the shocks that lead a priori similar households to differ matter for the assessment of the effects of tax policy. In particular, given the variance of innovations to income faced by a household, its ability to smooth consumption in the absence of complete insurance markets depends crucially on the *persistence* of shocks. Loosely speaking, the more that annual income "looks like" long-run average income, the more informative annual incidence of consumption taxes will be. When shocks are transitory, a current shock to productivity will have less influence on the lifetime resources that a household can expect over its remaining lifetime. As a result, consumption levels will not need to be adjusted by much in order for the lifetime budget constraint to be satisfied. In turn, unless the household is near a constraint on borrowing, its consumption will not respond to such shocks. By contrast, in an economy with highly persistent labor income risk, a household who has just received a bad shock may expect more of the same in the near and intermediate-term future; thus, expected lifetime resources have to be revised downward, possibly significantly (see, e.g., Deaton [1992]). Therefore, satisfaction of the household's lifetime budget constraint will require a commensurate reduction in current and future consumption. Conversely, if a household receives a good realization of a persistent shock, consumption is likely to jump up as lifetime expected resources are revised upwards.

From a policymaker's perspective, the issue is this: the less closely that consumption tracks income, the more effective we can say that consumption smoothing is. However, annual incidence measures will suggest regressivity. As a result, consumption taxes may appear undesirable in precisely those instances in which households are successful in managing the impact of income fluctuations on their standard of living. The preceding is a relevant consideration: A relatively large body of work has shown that households engage in significant consumption smoothing over their lifetimes (see, e.g., Attanasio et al. [1999] and Gourinchas and Parker [2002]). In contrast to annual incidence, lifetime incidence will not be distorted by the effectiveness of household consumption smoothing.

Unfortunately for policymakers, recent work has debated the persistence of income shocks (which are taken to represent productivity shocks), with estimates that lie substantially away from each other. At one end of the spectrum are the estimates of Storesletten, Telmer, and Yaron (2004) who argue that aggregate cross-sectional evidence suggests a unit-root component for shocks. At the other end of the spectrum are the more recent estimates of Guvenen (2007), who has argued that shock persistence is in fact far lower, and in an AR(1) setting, better approximated by a persistence parameter of 0.8. Given this large range, we present the implications of a switch to consumption taxes under a variety of values for shock persistence and show that measured regressivity does depend on the persistence of shocks.

In this article, we address two questions. First, how will a move to pure consumption taxation matter for aggregate outcomes, and how do the results depend on the persistence of shocks to productivity? Specifically, under varying shock persistence, how do the levels and variability of consumption, wealth, and labor supply respond to this tax reform? Second, how regressive are consumption taxes? Do annual and lifetime incidence measures of consumption taxes differ, and how do the results depend on the persistence of productivity shocks? Specifically, we utilize the Suits Index (Suits 1977), a standard measure of the incidence of taxes, to determine how regressivity depends on (i) the frequency at which income is measured and (ii) the stochastic structure of idiosyncratic household productivity. We will then describe the relationship of the Suits Index to direct cross-sectional measures of inequality, in particular, the Gini Index and the coefficient of variation.

Given our objectives, it is essential that we face the household with a stochastic productivity process that accurately captures both the true nature of household risk *and* the tools with which households smooth consumption. Therefore, our model features a stochastic process for productivity that contains a transitory component, a persistent component, and a well-defined "hump-shaped" life-cycle profile for average productivity. We equip households with the two tools thought to be empirically most relevant for consumption smoothing: self-insurance through asset accumulation, and flexible labor supply. Our work is most closely related to Fullerton and Rogers (1991) and Metcalf (1997), who study the dependence of measured regressivity on the frequency of income measurement, though in stylized models that abstract from uncertainty. Given the potential for uncertainty to alter consumption smoothing, our article contributes to the literature by allowing for stochastic shocks of varying persistence and flexible labor supply. It is also related to Ventura (1999), Nishiyama and Smetters (2005), Athreya and Waddle (2007), and Fuster, Imrohoroglu, and Imrohoroglu (2008). Our work differs from prior work as it derives the implications for tax incidence as a function of the stochastic properties of income.

Our main findings are as follows. In terms of aggregates, we find that a move to a consumption tax will increase savings taken into retirement but will not alter either labor supply or consumption variability substantially. The *level* of inequality does vary with the persistence of productivity shocks, especially when using lifetime measures of the relevant variables. With respect to regressivity, our results show that the findings of Metcalf (1997) carry over to a substantially richer setting: We show that regressivity is a measure that is quantitatively sensitive to the frequency of income being used. Our results obtain in spite of the fact that borrowing constraints bind for most households early in life. While annual incidence shows substantial regressivity, the lifetime incidence of a consumption tax is proportional, *irrespective* of the persistence of income shocks. Perhaps the central lesson of our article is that standard measures of the incidence of consumption taxes can be rather misleading as a guide to its implications for household consumption smoothing.

In what follows, Section 1 lays out some intuition for the role played by consumption taxes. Section 2 presents the model and equilibrium, Sections 3 and 4 present the parameterization and results, and Section 5 concludes.

1. WHY MIGHT A SWITCH TO CONSUMPTION TAXES MATTER?

First, we provide some intuition for why a switch to consumption taxation may indeed alter the optimization problem faced by agents. Notice that in some very simple settings, tax systems that tax both labor income and capital income are actually equivalent to systems in which there is a pure consumption tax.

As a result, the move to a consumption tax from a regime of labor and capital income taxes is not inherently a meaningful change, as it may not change the household's underlying optimization problem. Following Nishiyama and Smetters (2005), it is instructive to consider a simple two-period model in which households enter with zero wealth ($a_1 = 0$), work only in the first period of life whereby they earn a deterministic wage, w_1, pay a flat tax on labor income, τ_l, and save an amount, a_2. In the second period, households are taxed on their capital income at a flat rate, τ_k, and live off gross-of-interest (and net-of-capital income tax) savings $a_2(1+r(1-\tau_k))$ in the second period. The per-period budget constraints are as follows. In the first period, we have

$$c_1 + a_2 = w_1(1 - \tau_l),$$

and in the second period we have

$$c_2 = (1 + r(1 - \tau_k))a_2.$$

In the absence of borrowing constraints, the relevant constraint on households is the single lifetime budget constraint:

$$\frac{c_1}{(1 - \tau_l)} + \frac{c_2}{(1 - \tau_l)(1 + r(1 - \tau_k))} = w_1.$$

Next, consider the same environment, but where labor and capital income taxes have been replaced by consumption taxes alone. In this case, the lifetime budget constraint is

$$c_1(1 + \tau_{c_1}) + \frac{c_2(1 + \tau_{c_2})}{(1 + r)} = w_1. \tag{1}$$

Inspecting (1) reveals that a regime in which consumption taxes in period 1 are set at $\tau_{c_1} \equiv \frac{1}{1-\tau_l} - 1$ and $\tau_{c_2} \equiv \frac{1+r}{[1+r(1-\tau_k)](1-\tau_l)} - 1$ generates identical incentives and constraints for the household. In this case, a system of flat capital and labor income taxes is equivalent to a system of consumption taxes that vary with age. The age-dependency of the equivalent consumption tax regime is a direct result of nonzero capital income taxation: $\tau_{c_1} = \tau_{c_2}$ if $\tau_k = 0$. Thus, whenever $\tau_k \neq 0$, it is as if future consumption is being taxed at a rate different from current consumption. In particular, a positive capital income tax implies that $\tau_{c_1} < \tau_{c_2}$: Future consumption is more expensive than current consumption.[1]

[1] It is for this reason that Erosa and Gervais (2001) emphasize that wherever a consumption tax system would be optimally age-dependent, but is unavailable for exogenous reasons, positive *flat* capital income taxes can be used along with labor income taxes to acheive the same outcome.

More generally, in a longer (but still deterministic and finite-lived) model, flat capital and labor income taxes are equivalent to a regime in which there are (i) an age-dependent sequence of consumption taxes, $\{\tau_{c_j}\}_{j=1}^{J}$, and (ii) a lump-sum transfer, Υ_0, to all households to offset the difference in present values of labor income created by the presence of capital income taxes. That is, the equivalent consumption tax at any age j=1, 2,..., J, is given by

$$\tau_{c_j} = \frac{(1+r)^j}{(1+r(1-\tau_k))^j(1-\tau_l)} - 1,$$

where we see again that if $\tau_k = 0$, $\tau_{c_j} = \frac{1}{1-\tau_l}\ \forall\ j = 1, \ldots, J$. Letting $\widetilde{w}_j$ denote income/productivity while the lump-sum *transfer* to households under a consumption tax is given by

$$\Upsilon_0 = \sum_{j=0} \frac{\widetilde{w}_j}{(1+r(1-\tau_k))^j} - \sum_{j=0} \frac{\widetilde{w}_j}{(1+r)^j}.$$

Notice again that when $\tau_k = 0$, there is no age-dependence in the sequence of consumption taxes, nor is there any transfer (i.e., $\Upsilon_0 = 0$).

Given these cases, we now turn to the aspects of our preferred model that break the equivalence between consumption tax regimes and those regimes that tax labor and capital income. First of all, like both recent tax reform proposals and analyses, we will consider a move to a regime of a flat consumption tax, implemented here as a flat sales tax on all household purchases.[2] These are among the most practical forms of consumption taxes under consideration in policy discussions. Notably, the inherent difficulties in implementing age-dependent taxes perhaps account for the fact that they are not a feature of any major economy. The absence of age-dependence then immediately rules out any equivalence with income taxes. Second, the interest rate on savings in the model is strictly positive. As a result, regardless of the size of capital income, as long as it is positive, an age-dependent consumption tax will be required to obtain equivalence. Third, we do not augment household income with lump-sum transfers or taxes. Fourth, we do not allow households to hold negative asset positions. As a result, young households may find themselves unable to consume as much as they would like. To the extent that consumption tracks household income, consumption taxes will not look as regressive—even though the observed fall in regressivity is an artifact of a binding constraint! Given all these departures, it is likely that a switch to a flat consumption tax regime generates meaningful changes in the economic environment within which households operate.

[2] Alternative regimes to implement consumption taxes include making all savings fully tax-deductible, or imposing a value-added tax.

2. MODEL

The economy is closely related to that in Ventura (1999), in that it features a well-defined life-cycle path for labor productivity, stochastic shocks, taxes, and elastic labor supply. There is a large number of agents who consume and work for J periods and then retire. We will focus on stationary settings where there is a time-invariant measure of agents of each age j, and, moreover, that the age-distribution is uniform.

During working life, households' productivity has a deterministically evolving component, but is also subject to stochastic shocks. In each period, households must choose labor effort, consumption, and savings. After working life, households then enter "retirement," which lasts for K periods. Households in retirement are assumed to face no further labor market risk and, therefore, solve a simple deterministic consumption-savings problem. They face only the constraint that the optimal consumption path have a present value equal to the present value of resources brought into retirement, inclusive of transfers.

Preferences

Households value consumption and leisure. All households have identical time-separable constant relative risk aversion (CRRA) utility functions over an composite good defined by a Cobb-Douglas aggregate of consumption and leisure, c_j and l_j, respectively, at each age-j during working life, and a "retirement felicity function," ϕ, that is defined on wealth, x_R, taken into retirement.

Households discount future consumption of the composite good exponentially using a time-invariant discount factor, β, and weight total consumption expenditures, c_j, in each period by an adjustment for the age-specific average household size, ES_j (a mnemonic for "equivalence scale"). Effective consumption is then defined to be $\frac{c_j}{ES_j}$. The problem for the household is to choose a vector sequence, $\{c_j, l_j\}_{j=1}^{J}$, and retirement wealth, x_R, to maximize lifetime utility. The absence of labor income in retirement implies that the value to a household of entering retirement with a given level of wealth, x_R, is the solution to the following problem. Let the maximal leisure available to households be denoted by $\bar{l}$, and let $\Pi(x_R)$ be the feasible set of consumption sequences given that a household enters retirement with resources x_R:

$$\phi(x_R) = \max_{\{c_k\}\in\Pi(x_R)} \sum_{k=1}^{K} \beta^k \frac{\left[c_k^{\theta}\bar{l}\right]^{1-\alpha}}{(1-\alpha)}. \tag{2}$$

The overall objective of the household can now be expressed as the sum of the optimization problem applicable to working life and a "continuation" value given by resources brought into retirement. Let $\Pi(\Psi_0)$ denote the space of all feasible combinations $(\{c_j, l_j\}, x_R)$ given initial state Ψ_0. The household optimization problem is

$$\max_{(\{c_j,l_j\},x_R)\in\Pi(\Psi_0)} E_0 \sum_{j=1}^{J} \beta^j \frac{\left[\left(\frac{c_j}{ES_j}\right)^{\theta} l_j^{1-\theta}\right]^{1-\alpha}}{1-\alpha} + \phi(x_R). \tag{3}$$

Endowments

Households are endowed with one unit of time, which they can divide between labor and leisure. Household income is determined as the product of labor effort and labor productivity. Productivity in any given period is the outcome of a process that has both deterministic and stochastic components. We follow Storesletten, Telmer, and Yaron (2004) to represent the logarithm of productivity (wages per effective unit of labor), $\ln w_j$, of households as the sum of three components: an age-specific mean of log productivity, μ_j, persistent shocks, z_j, and transitory shocks, η_j. Therefore, we have

$$\ln w_j = \mu_j + z_j + u_j \tag{4}$$

with

$$z_j = \rho z_{j-1} + \eta_j,\ \ \rho \leq 1,\ \ j \geq 2 \tag{5}$$

$$u_j \sim i.i.d\ N(0, \sigma_u^2),\ \eta_j \sim i.i.d.\ N(0, \sigma_\eta^2),\ \ u_j, \eta_j \text{ independent.} \tag{6}$$

Households draw their first realization of the persistent shock from a distribution with a conditional mean of zero, i.e., z_0. The innovation to the persistent shock, η_1, is also mean-zero, but has variance, $\sigma_{\eta_1}^2$, that is drawn to help match the inequality of log labor income among those entering the labor force. The variance of persistent shocks drawn at all other ages differs from $\sigma_{\eta_1}^2$ and is denoted σ_η^2.

Market Arrangement

As is standard in models of exogenous uninsurable risks, households of age-j can save and dissave by choosing a position in only a single noncontingent bond, denoted x_{j+1}. The economy is a small open-economy setting, whereby savings earn an exogenous gross rate of return (net of taxes). The household can also vary its labor supply, both to respond to changes in labor productivity and to smooth consumption of the composite good. For example, if financial resources are low in the current period, a household with a given labor productivity may choose to supply more labor than they would if they had more financial wealth. This is because they would otherwise be forced into a current period allocation that had low consumption and high leisure, while their intertemporal smoothing motives dictate preventing a fall in consumption. According to the experiment under study, labor income, capital income from savings, and consumption may each be taxed at (time-invariant) flat rates denoted by τ_l, τ_k, and τ_c, respectively.[3] Notice that in this model, given the abstraction from multiple layers of production of the final consumption good, the consumption tax will also be identical to a value-added tax. Because we treat the economy as one that is open to world trade and, furthermore, one in which the households under study do not affect the total demand or supply of assets worldwide, the interest rate on risk-free savings is assumed to be unaffected across tax policies. Given elastic labor supply and the three taxes, the generalized household budget constraint in each period is

$$c_j(1+\tau_c) + x_{j+1} = \widetilde{w}_j(1-l_j)(1-\tau_l) + x_j(1+r(1-\tau_k)). \qquad (7)$$

Optimal Household Decisions

Retirement

Age-J households value retirement savings via $\phi(x_R)$. The consumption flow arising from a given level of savings is specified as follows. Households aged $J+1$ are guaranteed to have a minimal standard of living given by a threshold, $\underline{\tau}^R$, representing Social Security and Medicare. Transfers during retirement are therefore *not* means-tested and are given instead by a single lump-sum transfer, $x_{\underline{\tau}}$, to all retiring households. Our approach follows Huggett (1996). A household's wealth level at retirement is then the sum of the household's personal savings, x_{J+1}, and the baseline retirement benefit, $x_{\underline{\tau}^R}$,

$$x_R = x_{J+1}\widetilde{R} + x_{\underline{\tau}^R}. \qquad (8)$$

The amount $x_{\underline{\tau}^R}$ is the wealth level that, when annuitized at the gross after-tax interest rate $\widetilde{R} \equiv (1+r(1-\tau_k))$, yields a flow of income each period equal to the societal minimum retirement consumption floor, $\underline{\tau}^R$. That is, minimal retirement wealth, $x_{\underline{\tau}}$, solves

[3] An interesting extension for future work would be to allow for more general, possibly progressive, tax schemes.

$$\sum_{k=1}^{K} \frac{\underline{\tau}^R}{(\widetilde{R})^k} = x_{\underline{\tau}^R}. \tag{9}$$

To solve for indirect utility at retirement, define the budget constraint for a retiree in period-k of retirement as follows:

$$(1 + \tau_c)c_k + x_{k+1} = x_k(1 + r(1 - \tau_k)) + \underline{\tau}^R. \tag{10}$$

Given the objective function during retirement (equation 2), the optimal intertemporal allocation of consumption must satisfy the following Euler equation:

$$\frac{c_{t+1}}{c_t} = \left(\frac{1}{\beta \widetilde{R}}\right)^{\frac{1}{\theta(1-\alpha)-1}}. \tag{11}$$

Defining $\gamma = (\frac{1}{\beta \widetilde{R}})^{\frac{1}{\theta(1-\alpha)-1}}$, we then see that (11) implies that consumption at any date-k of retirement can be defined as:

$$c_k = \gamma^k c_0. \tag{12}$$

Given the preceding requirement on optimal consumption growth, we use the budget constraint to pin down the level of the sequence of retirement consumptions. First, we iterate on the per-period budget constraint (equation 10) to obtain a single present value budget:

$$\sum_{k=0}^{K} \frac{c_k(1 + \tau_c)}{\widetilde{R}^k} = x_R,$$

where x_R is defined in (8).

As a result, we obtain

$$c_0 = \frac{x_R}{\sum_{k=0}^{K} \gamma^k \frac{(1+\tau_c)}{\widetilde{R}^k}}.$$

The remaining sequence is given by (12), which we denote as $\{c^*_{R_k}\}_{k=0}^{K}$, which then yields the indirect utility of retirement:

$$\phi^*(x_R) = \sum_{k=0}^{K} \beta^k \frac{\left[c^*_{R_k} \bar{l}\right]^{1-\alpha}}{(1-\alpha)}. \tag{13}$$

Working life

The solution of the household's problem during working life is simplified by our use of Cobb-Douglas preferences. It is instructive to display the manner in which the various taxes alter the optimal allocation of consumption over time and the optimal mix of consumption and leisure. First, within any given period, it is useful to think of a household as first working full time and then "buying back" consumption and leisure. Therefore, if a household works full time (normalized to unity), has entered a period with savings x_j, and plans to save x_{j+1}, its resources available to purchasing consumption and leisure are pinned down. That is, consumption and leisure purchases must satisfy

$$c_j(1+\tau_c) + \widetilde{w}_j l_j(1-\tau_l) = \widetilde{w}_j(1-\tau_l) + x_j(1+r(1-\tau_k)) - x_{j+1}. \quad (14)$$

Letting $\Lambda_j \equiv \widetilde{w}_j(1-\tau_l) + x_j(1+r(1-\tau_k)) - x_{j+1}$ denote the total "resources" available for consumption and leisure, we have from the intratemporal first-order conditions of the household's problem that the optimal mix of expenditures on leisure and consumption satisfies

$$\frac{l_j}{c_j} = \frac{(1-\theta)}{\theta}\frac{(1+\tau_c)}{(1-\tau_l)\widetilde{w}_j}. \quad (15)$$

Notice that for any given realization of current productivity, $\widetilde{w}_j$, and elasticity of substitution, θ, the optimal mix of leisure and consumption depends only on the ratio $\frac{(1+\tau_c)}{(1-\tau_l)}$. That is, the levels of either tax alone do not determine how households divide their resources between leisure or consumption. The preceding expression, when substituted into the household budget constraint, gives the optimal levels of consumption and leisure, respectively, as a function of resources Λ_j:

$$c_j = \frac{\Lambda_j \theta}{1+\tau_c}, \text{ and} \quad (16)$$

$$l_j = \frac{\Lambda_j(1-\theta)}{(1-\tau_l)\widetilde{w}_j}. \quad (17)$$

Given these rules for optimal consumption and leisure for any given resources, the only remaining decision for the household is to choose what resources to keep in the current period; this is simply done by choosing the savings level x_{j+1} optimally. Let $U(.)$ denote the within-period utility function. During working life, the intertemporal first-order condition is given by

$$U'_{c_j}(c_j, l_j) = \beta(1+r(1-\tau_k))U'_{c_{j+1}}(c_{j+1}, l_{j+1}). \quad (18)$$

Notice here that consumption and labor income taxes do not appear, while the capital income tax does. This is the crux of the distortion to private savings decisions induced by capital income taxes. Moreover, as shown above, an equivalent system of consumption exists that implies systematically increasing tax rates on consumption in the increasingly distant future. This implies that capital income taxes lower the return to saving and thereby encourage current consumption; when consumption and leisure are complements, there is resultant reduction in work effort.

Recursive Formulation

The household's problem can be represented recursively as follows. At the beginning of each period, the household's options are completely determined by its age-j, its wealth, x_j, its current realized value of the persistent shock, z_j, and the current realization of the transitory shock, η_j. These items are sufficient to determine the budget constraint faced by households in the current period, and also to obtain the best forecast of next period's realization of the persistent shock.[4]

Optimal household behavior requires that in each period, given their state vector, the household chooses consumption, c_j, and savings, x_{j+1}, to satisfy the following recursion:

$$V(j, x_j, z_j, \eta_j) = \max_{c_j, x_{j+1}} U(c_j) + \beta E_{(z_{j+1}, \eta_{j+1}|z_j)} V(j+1, x_{j+1}, z_{j+1}, \eta_{j+1}), \tag{19}$$

subject to the budget constraint described in equation (7), and where $\beta E_{(z_{j+1}, \eta_{j+1}|z_j)}$ denotes the expectation of the value of carrying savings, x_{j+1}, into the following period when the shocks tomorrow (z_{j+1}, η_{j+1}) are drawn from the conditional joint distribution that reflects the current realization of the persistent shock, z_j. We focus on a stationary equilibrium: Households optimize given prices, and the distribution of the households over values of the state is stationary (time-invariant).

3. PARAMETERIZATION

The model period is one calendar year. Households work for $J = 44$ periods, where $j = 1$ represents real-life age 21, and $j = 44$ is retirement at age 65. Retirement lasts for $K = 25$ periods, so all agents die at real-life age 90. Risk aversion and discounting are set at $\alpha = 3$ and $\beta = 0.96$, respectively. The (gross) risk-free interest rate on savings is $R^f = 1.01$. Households are born

[4] Given that the tax rates are assumed to remain constant throughout time, they do not need to be included in the "state vector."

Table 1 Parameter Values

Variable	Value
β (Discount Factor)	0.96
J (Working Life)	44
K (Retirement Length)	25
R^f (Risk-Free Rate)	1.01
x_1 (Beginning of Life Assets)	0
$\bar{l}$ (Maximum Leisure)	1
θ (Elasticity of Labor Supply)	0.5

with zero financial wealth: $x_1 = 0$. Maximum leisure time, $\bar{l}$, is normalized to unity and the elasticity of labor supply, θ, is set to 0.5 to reflect that, on average, half of a household's discretionary hours (16 hours per day) are spent working. Our benchmark model features taxes on consumption, labor income, and capital income, and we follow Fuster, Imrohoroglu, and Imrohoroglu (2008) to assign the following values for these taxes: $\tau_c = 0.055$, $\tau_k = 0.35$, and $\tau_l = 0.173$. Under a switch to a pure consumption tax, we ensure revenue-neutrality relative to our benchmark economy.

A brief summary of the stochastic process for productivity is the following. We set $\rho = 0.99$, $\sigma_u^2 = 0.063$, $\sigma_{\eta_1}^2 = 0.22$, and $\sigma_\eta^2 = 0.0275$, as these values generate reasonable income variability (given optimal labor supply) among the youngest working-age households in the data, as well as the nearly linear life-cycle growth of cross-sectional variance in log income documented in Storesletten, Telmer, and Yaron (2004) and the total increase in cross-sectional (log) income variance over the life cycle. The parameters governing the income process also generate reasonable wealth-to-income ratios over the life cycle (see, e.g., Athreya [2008]).

The mean of log productivity is given by the profile $\{\mu_j\}_{j=1}^J$ and is based on the estimates of Hansen (1993). We approximate the continuous state-space stochastic process for income via a discrete state-space Markov chain using the method of Tauchen (1986). Specifically, we use a 32-point discretization of the persistent shock and a three-point discretization for the transitory shock. We employ standard discrete-state space dynamic programming and Monte Carlo simulation to solve for decisions and generate aggregate outcomes, respectively.[5] The values of all policy-invariant parameters are reported in Table 1.

[5] All code is available from the authors on request.

4. CONSUMPTION TAX REFORM

We first report the model's implications for the aggregate consequences of a move to pure consumption taxation for several specifications of income persistence and risk aversion. Specifically, we set $\tau_l = \tau_k = 0$ and set τ_c such that the change is revenue-neutral. We provide measurements of consumption, labor supply, and wealth distributions across tax regimes in each case. We then turn to the issue of the measurement of the progressivity of consumption taxes.

Consumption, Asset Accumulation, and Leisure

The means and coefficients of variation for the variables mentioned above appear in Table 2, while Table 3 contains the Gini Coefficients for annual income, lifetime income, annual consumption, lifetime consumption, and wealth. The model does fairly well under relatively high productivity shock persistence in reproducing estimates of observed labor income and wealth inequality. Rodríguez et al. (2002), using the 1998 Survey of Consumer Finances, report a wealth Gini of 0.8 and an annual income Gini of 0.55, very close to the model's predictions under our benchmark model, which features high persistence. The model also preserves the observed ordering of inequality seen in the data (e.g., Rodríguez et al. 2002), where wealth is more unequal than income, which in turn is more unequal than consumption. Therefore, the limited insurance that households provide through saving and dissaving is partially effective but nonetheless results in large wealth inequality.

Our first result is that the largest effects of a move to a consumption tax occur in savings. This is due to the removal of the intertemporal distortion created by the taxation of capital income, as well as the need for additional savings in retirement to offset the heavier tax burden faced by retirees who no longer escape taxation. The magnitude of the increase in average savings is similar to other recent work (see, e.g., Fuster, Imrohoroglu, and Imrohoroglu [2008], Table 3). In the cases with lower persistence, when assets are more useful for self-insurance, the increase in savings when switching to a pure consumption tax is even larger. That is, a move to a consumption tax regime under low income shock persistence induces a larger response in aggregate savings than with higher persistence. This makes clear that the size of the distortion created by a capital income tax depends on the shock process faced by households.[6] Of course, the increased savings means increased resources

[6] Changes in the persistence of the shock alter the unconditional variance of the shock. However, given that productivity is a log-normal random variable, changes in the varriance affect the mean of the level of productivity. When we lower the persistence of shocks, we therefore increase the variance of the transitory component such that the mean level of income always remains constant. Part of the effect on savings seen is due to the higher variance of transitory shocks under lower persistence.

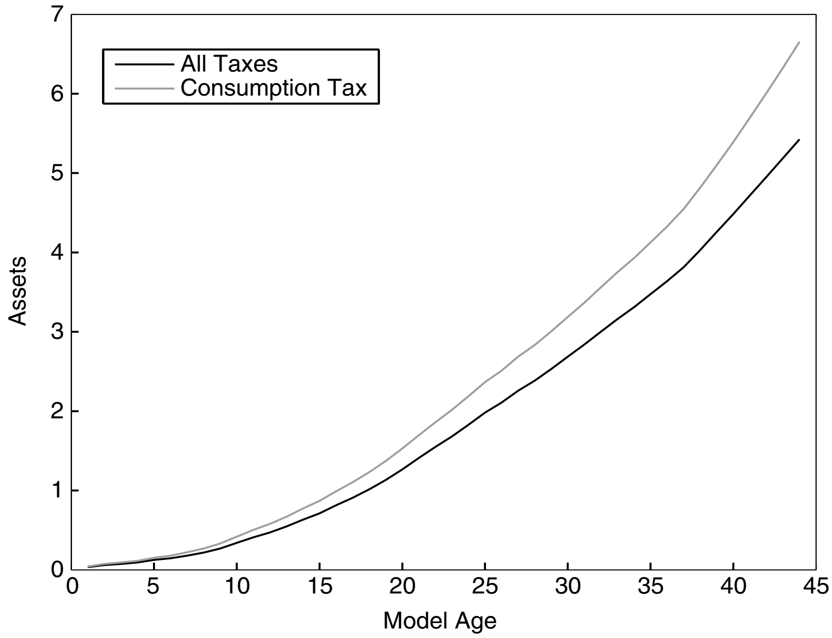
Table 2 Aggregate Results

	ρ	α	τ_l	τ_k	τ_c	E(l)	CV, l	E(x)	CV, x	E(c)	CV, c	E$\left(\frac{c}{ES}\right)$	CV, $\frac{c}{ES}$.	E (Lab. Inc)	CV Lab. Inc
Case 1	0.99	3	0.173	0.35	0.055	.503	0.188	1.958	2.114	0.606	1.088	0.530	1.062	0.908	1.396
Case 2	0.99	3	—	—	0.338	.503	0.188	2.351	2.012	0.580	1.108	0.507	1.082	0.905	1.379
Case 3	0.99	2	0.173	0.35	0.055	.503	0.194	1.650	2.344	0.605	1.082	0.530	1.057	0.910	1.410
Case 4	0.99	2	—	—	0.338	.502	0.195	1.948	2.238	0.579	1.102	0.506	1.077	0.907	1.394
Case 5	0.8	3	0.173	0.35	0.055	.528	0.265	1.984	1.449	0.579	0.890	0.507	0.864	0.885	1.395
Case 6	0.8	3	—	—	0.35	.527	0.268	2.616	1.363	0.547	0.904	0.479	0.878	0.885	1.383
Case 7	0.8	2	0.173	0.35	0.055	.525	0.265	1.659	1.653	0.578	0.899	0.506	0.874	0.885	1.400
Case 8	0.8	2	—	—	0.35	.524	0.269	2.219	1.543	0.545	0.912	0.477	0.887	0.887	1.388
Case 9	0.5	3	0.173	0.35	0.055	.538	0.282	1.994	1.254	0.578	0.864	0.506	0.834	0.889	1.448
Case 10	0.5	3	—	—	0.35	.538	0.284	2.635	1.191	0.545	0.876	0.478	0.846	0.890	1.437
Case 11	0.5	2	0.173	0.35	0.055	.535	0.283	1.670	1.448	0.577	0.875	0.505	0.846	0.890	1.453
Case 12	0.5	2	—	—	0.35	.535	0.286	2.238	1.366	0.544	0.886	0.476	0.856	0.892	1.442

Table 3 Gini Coefficients

	Annual Lab Inc.	Annual Cons.	Wealth	Lifetime Lab Inc.	Lifetime Cons.
$\rho = 0.99,\ \alpha = 3,\ \tau_c = .055,$ $\tau_l = .173,\ \tau_k = .35$	0.5342	0.4462	0.7574	0.4129	0.38
$\rho = 0.99,\ \alpha = 3,\ \tau_c = .338,$ $\tau_l = 0,\ \tau_k = 0$	0.5319	0.4494	0.7483	0.4101	0.3834
$\rho = 0.99,\ \alpha = 2,\ \tau_c = .055,$ $\tau_l = .173,\ \tau_k = .35$	0.5356	0.4464	0.7895	0.4129	0.3802
$\rho = 0.99,\ \alpha = 2,\ \tau_c = .338,$ $\tau_l = 0,\ \tau_k = 0$	0.5332	0.4499	0.7806	0.41	0.3836
$\rho = 0.8,\ \alpha = 3,\ \tau_c = .055,$ $\tau_l = .173,\ \tau_k = .35$	0.5487	0.3992	0.6599	0.219	0.1975
$\rho = 0.8,\ \alpha = 3,\ \tau_c = .35,$ $\tau_l = 0,\ \tau_k = 0$	0.5483	0.3999	0.6438	0.217	0.1997
$\rho = 0.8,\ \alpha = 2,\ \tau_c = .055,$ $\tau_l = .173,\ \tau_k = .35$	0.5477	0.4032	0.7073	0.2188	0.1982
$\rho = 0.8,\ \alpha = 2,\ \tau_c = .35,$ $\tau_l = 0,\ \tau_k = 0$	0.5476	0.4032	0.6904	0.2169	0.2004
$\rho = 0.5,\ \alpha = 3,\ \tau_c = .055,$ $\tau_l = .173,\ \tau_k = .35$	0.5591	0.3851	0.6145	0.1454	0.1271
$\rho = 0.5,\ \alpha = 3,\ \tau_c = .35,$ $\tau_l = 0,\ \tau_k = 0$	0.5589	0.3854	0.6004	0.144	0.1287
$\rho = 0.5,\ \alpha = 2,\ \tau_c = .055,$ $\tau_l = .173,\ \tau_k = .35$	0.5585	0.389	0.6683	0.1455	0.1276
$\rho = 0.5,\ \alpha = 2,\ \tau_c = .35,$ $\tau_l = 0,\ \tau_k = 0$	0.5585	0.3886	0.6528	0.1441	0.1292

Figure 1 Average Asset Position, by Age



taken into retirement. However, under a pure consumption tax regime, the ability of households to use these resources to finance consumption will be altered. Figure 1 shows that the removal of the intertemporal distortion in savings ultimately aids substantially the ability of households to transfer resources into retirement.

In contrast to outcomes under pure consumption taxes, the persistence of shocks to productivity does not play an important role in aggregate savings when income is taxed. The intuition here is that, with capital income taxes in particular, arranging for consumption in the distant future (e.g., at retirement) is more expensive than without a capital income tax. As a result, even though lower persistence makes self-insurance more effective, the distortion created by capital income taxation makes future consumption expensive enough to make the net increase in aggregate savings small.

With respect to wealth inequality, the coefficients of variation and Gini Coefficients for wealth show that a move to a consumption tax *lowers* wealth inequality and variability, irrespective of persistence and risk aversion. This is an important observation for those concerned with the long-run equity implications of consumption taxation. We also see that, for a given tax regime, low persistence leads to low wealth inequality. This occurs as lower persistence makes lengthy runs of good or bad luck less likely. Conversely, lower risk aversion, by making households more willing to allow for variation in their consumption, creates a wealth distribution with a lower mean and higher coefficient of variation for any given tax regime.

Turning next to effective consumption, we see that a move to a consumption tax has a significant effect. In all economies under study, our model predicts that a move to a pure consumption tax leads to about a 6 percent drop in average effective consumption, while leaving the coefficient of variation largely unchanged. Persistence matters for mean effective consumption only at the highest value, $\rho = 0.99$. However, a move from the benchmark tax regime to pure consumption taxation does not substantially affect the variability of consumption, as seen from the coefficient of variations (cv) shown in Table 2.

As a symptom of the effectiveness of self-insurance under transitory income risk, we see that consumption inequality falls substantially when the persistence drops below 0.99. This result does not depend on tax regime or risk aversion. Looking at Table 3, we see that the Gini Coefficients for consumption show a similar pattern of inequality as the coefficients of variation. When measured at an annual frequency, the consumption Gini decreases as persistence falls, indicating that inequality is higher in states with more persistent shocks regardless of tax regime or risk aversion. This result is accentuated when consumption is reported as a lifetime measure.

Unlike its effects on wealth accumulation and effective consumption, a move to consumption taxes has little impact on labor supply. Moreover, this is robust as it occurs for all levels of risk aversion and shock persistence that we consider. This is important, as the elimination of labor income taxation might have been thought to induce greater labor supply. However, recall equation (15), which shows that the optimal mix of consumption and labor depends on the ratio $\frac{1+\tau_c}{1-\tau_l}$. A move to a pure consumption tax increases both the numerator and the denominator, potentially undoing much of the change created by a jump in the consumption tax. This happens in the model on average. We see in Table 2 that, although consumption falls, leisure remains more or less constant. We also see that, regardless of tax regime and risk aversion, the mean and coefficient of variation of leisure rise with lower persistence. With higher persistence, each shock changes potential future earnings by more than if shocks were transitory. This means that the only way to keep consumption smooth over the life cycle is to work hard in both bad times and good times, which makes leisure less volatile. The result that labor supply does not move much with a consumption tax reform is somewhat telling. Recent work has made clear that household labor supply can be an important smoothing device (e.g., Pijoan-Mas [2006] and Blundell, Pistaferri, and Preston [2008]). Yet, in our experiments, labor hours and earnings respond very little in response to the elimination of income taxes in favor of consumption taxes. The behavior of labor supply, therefore, provides an additional source of evidence that consumption taxes do not expose households to increased risk.

Given the relative invariance of labor supply across economies, the induced stochastic process for labor income is also similar across economies. For example, we see only a slight increase in inequality as persistence decreases, which is reflected in the annual labor income Gini Coefficient, as well as small changes in response to risk aversion. With respect to persistence, our finding stems from the increased volatility in labor supply in low persistence states, which leads to more volatile income for agents. However, labor income inequality depends heavily on whether it is measured annually or over the lifetime. Given any combination of risk aversion and persistence, we see that annual income inequality is substantially higher than lifetime income inequality, as realized lifetime productivity will be much less volatile than its annual counterpart. Moreover, as the persistence of income grows, the level of annual income inequality decreases slightly, while lifetime income inequality increases dramatically. This is because the variance of realized productivity over the lifetime will be much larger when shocks are persistent.

Measured Progressivity and its Relation to Consumption Smoothing

Having laid out the aggregate implications of a move to a consumption tax, we now turn to the central questions of our article regarding the measurement of the incidence of consumption taxes and the relationship of these statistics to direct measures of consumption smoothing. To measure the progressivity of a given tax regime for a given economy, we use the Suits Index (Suits 1977). Let S_x represent the Suits index for a given tax regime and state, and $T_x(y)$ represent the cumulative tax burden for a given level of accumulated household income, y, then:

$$S_x = 1 - \int T_x(y)dy.$$

A Suits index can therefore range between -1 and 1. A positive index implies a progressive tax, while a negative index implies regressivity in the tax regime. An index of 0 is proportional. Table 4 reports the value of the Suits index across experiments and for three reference variables: realized annual income, realized lifetime income, and wealth. Our preferred "direct" measures of consumption smoothing are the coefficient of variation of consumption and the Gini Coefficient for consumption.

The basic input to the Suits index is a function mapping the relative contribution of households ranked by a given reference variable to tax revenues. However, instead of plotting tax contributions by accumulated percentages of households, as is the case with the Gini index, which is based on a Lorenz

Table 4 Suits Indexes

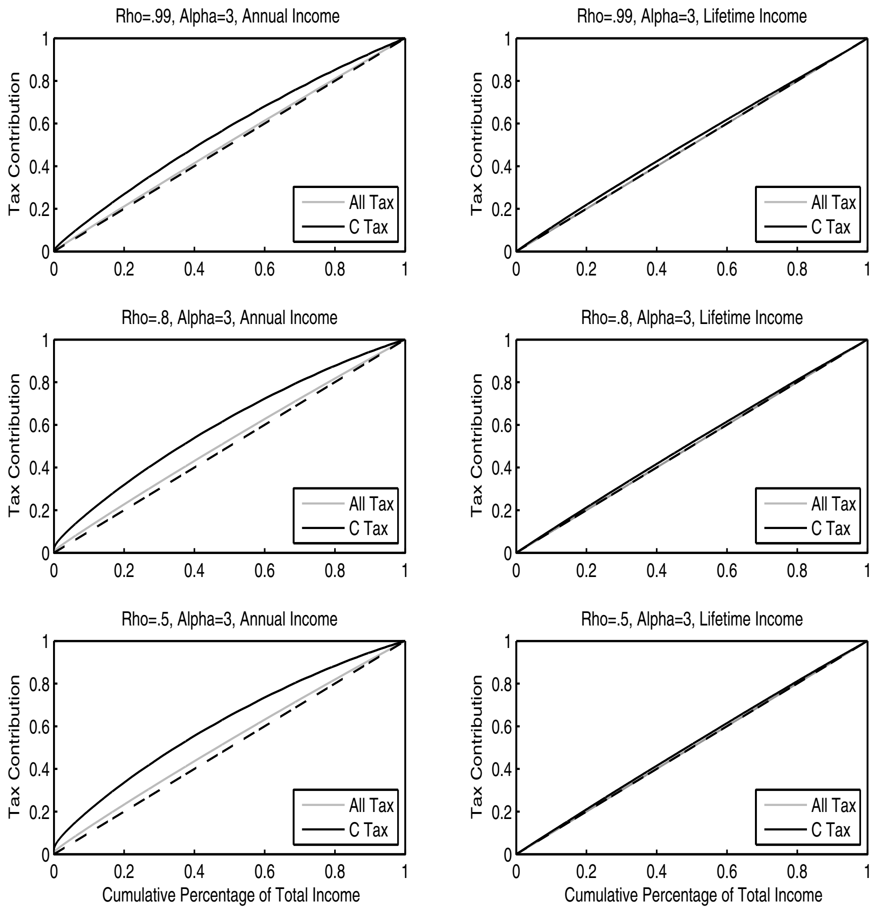
	By Annual Income	By Lifetime Income	By Wealth
$\rho = 0.99,\ \alpha = 3,\ \tau_c = .055,\ \tau_l = .173,\ \tau_k = .35$	−0.02	0.00	−0.41
$\rho = 0.99,\ \alpha = 3,\ \tau_c = .338,\ \tau_l = 0,\ \tau_k = 0$	−0.12	−0.03	−0.42
$\rho = 0.99,\ \alpha = 2,\ \tau_c = .055,\ \tau_l = .173,\ \tau_k = .35$	−0.02	0.00	−0.47
$\rho = 0.99,\ \alpha = 2,\ \tau_c = .338,\ \tau_l = 0,\ \tau_k = 0$	−0.12	−0.03	−0.49
$\rho = 0.8,\ \alpha = 3,\ \tau_c = .055,\ \tau_l = .173,\ \tau_k = .35$	−0.04	0.00	−0.44
$\rho = 0.8,\ \alpha = 3,\ \tau_c = .35,\ \tau_l = 0,\ \tau_k = 0$	−0.19	−0.02	−0.44
$\rho = 0.8,\ \alpha = 2,\ \tau_c = .055,\ \tau_l = .173,\ \tau_k = .35$	−0.04	0.00	−0.50
$\rho = 0.8,\ \alpha = 2,\ \tau_c = .35,\ \tau_l = 0,\ \tau_k = 0$	−0.19	−0.02	−0.51
$\rho = 0.5,\ \alpha = 3,\ \tau_c = .055,\ \tau_l = .173,\ \tau_k = .35$	−0.05	0.00	−0.46
$\rho = 0.5,\ \alpha = 3,\ \tau_c = .35,\ \tau_l = 0,\ \tau_k = 0$	−0.22	−0.02	−0.46
$\rho = 0.5,\ \alpha = 2,\ \tau_c = .055,\ \tau_l = .173,\ \tau_k = .35$	−0.05	0.00	−0.53
$\rho = 0.5,\ \alpha = 2,\ \tau_c = .35,\ \tau_l = 0,\ \tau_k = 0$	−0.22	−0.02	−0.53

curve, the Suits index relies on a curve constructed by plotting the cumulative percentage of the reference variable against the cumulative percent of total tax burden on the vertical axis. A given point on the x-axis of the more familiar Lorenz curve refers to a household whose realization of the reference variable (e.g., income) lies above a given fraction of households. By contrast, a given point on the x-axis of the Suits index function gives the accumulated percentage of the reference variable. For example, a value on the x-axis of 0.3 for a Lorenz curve of tax contributions by income refers to a household whose income is above 30 percent of *households*. A value on the x-axis of 0.3 under the Suits index refers to the entire set of households whose *collective contribution to total income is 30 percent*. In particular, unless the reference variable is distributed uniformly, these two measures will not coincide.

Our main finding is that the measured frequency of income is important for the assessment of the progressivity of consumption taxes, both in absolute terms and relative to that obtaining under income taxes. By contrast, measured income frequency matters very little in the assessment of progressivity under income taxes. This result is robust as it survives across varying levels of income shock persistence as well as risk aversion. Each row in Figure 2 displays the Suits function under annual and lifetime measures of income for a given level of income shock persistence. Risk aversion is held fixed at $\alpha = 3$. Since productivity is risky, income is a random variable. Therefore, we measure income ex-post. In the case of lifetime income, we compute, using our simulated income histories, realized lifetime labor incomes for a large sample of households. As seen in Figure 2, the Suits function for annual incidence lies significantly above the 45° line for the consumption tax regime. However, the Suits function for lifetime incidence is essentially proportional. This finding echoes the earlier finding of Metcalf (1997) and suggests that the presence of uninsurable productivity risk does not alter the implications for regressivity of a consumption tax. The measurement of income also affects the *relative* regressivity of consumption taxes versus income taxes. Figure 2 shows that, under annual measures of income, the consumption tax appears much more regressive when compared to a regime with income taxes.

As mentioned at the outset, the more transitory is productivity risk, the more labor earnings are likely to respond to a change in productivity. This is because the ability of a household to generate earnings over its remaining lifetime is relatively less affected when shocks to its productivity are transitory. As a result, the household smooths both consumption and its complement, leisure, quite effectively. In turn, households often consume amounts that are large in relation to their earned income when young and small relative to their earned income when old. As seen in Figure 2, the lower is persistence, the more annual incidence suggests that consumption taxes are regressive. Table 4 presents the numerical values of the Suits indexes.

Figure 2 Suits Functions by Labor Income



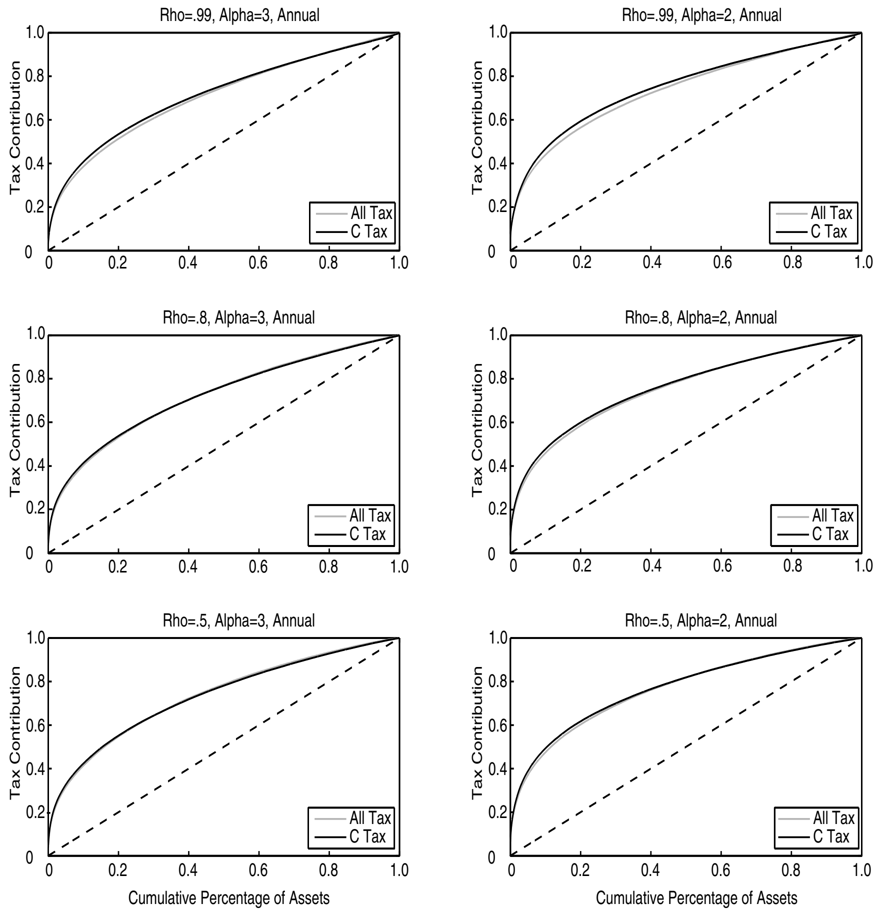
When measured by annual income, a move to a consumption tax from the benchmark tax system leads to more regressivity, and the measure of regressivity increases as income persistence falls. This is because transitory shocks are effectively smoothed via both asset accumulation or decumulation and changes in labor supply. However, as the persistence of productivity shocks rises, such smoothing becomes more difficult. Table 2 shows that the variability of both consumption and effective consumption rise systematically with persistence. Similarly, Table 3 shows that when computed either using lifetime or annual consumption, the Gini Coefficient remains remarkably stable across tax regimes. As with the coefficient of variation, the consumption Gini falls substantially as persistence falls. The preceding makes clear that measures of regressivity that are based on annual income may be misleading for household well-being because they *rise*, while two independent and direct measures of consumption smoothing indicate an improvement in insurance. In sharp contrast, lifetime incidence measures show little variation with shock persistence. It is also important to note that we disallow borrowing in the model; more ability to issue debt would further exaggerate the measured regressivity of consumption taxes. Lastly, while not shown here for brevity, the results in Figure 2 are nearly replicated when risk aversion is lowered below $\alpha = 3$. While we have focused on consumption, notice that in Table 1 the mean and coefficient of variation in labor effort are very similar across tax regimes for all the values of income persistence and risk aversion we consider. Therefore, consumption taxes are unlikely to damage household well-being along this dimension.

Taxes that fall disproportionately on households with low wealth may also be seen as regressive. Therefore, we turn now to measures of regressivity based on rankings of households by wealth. Wealth is a "stock" variable and so there is no "frequency" dimension to its measurement, but the issue of progressivity remains: Do the relatively wealthy pay disproportionately more than those who have fewer assets? As seen in Figure 3, the answer is no under either tax regime. In fact, the Suits index for tax progressivity shown in Table 4 indicates that when measured by wealth, both income and consumption taxes are quite regressive. The measured regressivity of taxes when the Suits function is constructed using wealth does respond to changes in risk aversion. As seen in both Figure 3 and Table 4, higher risk aversion implies lower measured regressivity. This is an implication of the increased precautionary savings motive under higher risk aversion, which leads low-wealth households to increase their savings disproportionately more than their high-wealth counterparts. Therefore, any given quantile of wealth represents a larger number of households under low risk aversion than under high risk aversion. As seen earlier in Table 2, neither mean consumption nor mean income changes substantially with risk aversion. Therefore, the contribution to total tax revenues of the lower wealth quantiles will be greater under low risk aversion. In addition to the preceding, comparing the columns within each row of Figure 3 shows that risk aversion has essentially no effect on the relative progressivity of income and consumption tax regimes. In terms of the overall regressivity of consumption taxes relative to current tax policy, the preceding results make clear that consumption is not inherently more regressive, especially when a lifetime perspective is taken.

5. CONCLUDING REMARKS

The smoother is consumption for a household, the more its tax burden remains invariant to its income. Ironically, this implies that when insurance and credit

Figure 3 Suits Functions by Assets



markets are most successful in delivering intertemporally and intratemporally smooth consumption, tax incidence using high frequency income measures (such as annual income) will, all else equal, imply the greatest regressivity. In this article, we have constructed and simulated a rich model of consumption, savings, and work effort over the life cycle. We have argued that while annual incidence suggests that consumption taxes are regressive, lifetime incidence suggests proportionality. Moreover, for a given level of income shock persistence, consumption taxes do not matter substantially for the variability of consumption. Lastly, we show that lifetime incidence is similar across tax regimes, labor productivity persistence, and risk aversion levels.

REFERENCES

Athreya, Kartik. 2008. "Default, Insurance, and Debt Over the Life-Cycle." *Journal of Monetary Economics* 55 (May): 752–74.

Athreya, Kartik, and Andrea L. Waddle. 2007. "Implications of Some Alternatives to Capital Income Taxation." Federal Reserve Bank of Richmond *Economic Quarterly 93* (Winter): 31–55.

Atkeson, Andrew, V. V. Chari, Patrick J. Kehoe. 1999. "Taxing Capital Income: A Bad Idea." Federal Reserve Bank of Minneapolis *Quarterly Review* 23 (Summer): 3–17.

Attanasio, Orazio, James Banks, Costas Meghir, and Guglielmo Weber. 1999. "Humps and Bumps in Lifetime Consumption." *Journal of Business and Economic Statistics* 17: 22–35.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." Mimeo, University College London.

Deaton, Angus. 1992. *Understanding Consumption*. New York: Oxford University Press.

Erosa, Andres, and Martin Gervais. 2001. "Optimal Taxation in Infinitely-Lived Agent and Overlapping Generations Models: A Review." Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Spring): 23–44.

Fullerton, Don, and Diane Lim Rogers. 1991. "Lifetime Versus Annual Perspectives on Tax Incidence." *National Tax Journal* 44 (September): 277–87.

Fuster, Luisa, Ayse Imrohoroglu, and Selahattin Imrohoroglu. 2008. "Altruism, Incomplete Markets, and Tax Reform." *Journal of Monetary Economics* 55 (January): 65–90.

Gourinchas, Pierre-Olivier, and Jonathan A. Parker. 2002. "Consumption Over the Life Cycle." *Econometrica* 70 (January): 47–89.

Guvenen, Fatih. 2007. "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review* 97 (June): 687–712.

Hansen, Gary D. 1993. "The Cyclical and Secular Behavior of the Labour Input: Comparing Efficiency Units and Hours Worked." *Journal of Applied Econometrics* 8: 71–80.

Huggett, Mark. 1996. "Wealth Distribution in Life-Cycle Economies." *Journal of Monetary Economics* 38: 469–94.

Metcalf, Gilbert E. 1997. "The National Sales Tax: Who Bears the Burden?" CATO Policy Analysis No. 289.

Nishiyama, Shinichi, and Kent Smetters. 2005. "Consumption Taxes and Economic Efficiency with Idiosyncratic Wage Shocks." *Journal of Political Economy* 113: 1,088–115.

Pijoan-Mas, Josep. 2006. "Precautionary Savings or Working Longer Hours?" *Review of Economic Dynamics* 9 (April): 326–52.

Prescott, Edward C. 2004. "Why Do Americans Work So Much More than Europeans?" Federal Reserve Bank of Minneapolis *Quarterly Review* 28 (July): 2–13.

Rodríguez, Santiago Budría, Javier Díaz-Giménez, Vincenzo Quadrini, and José-Víctor Ríos-Rull. 2002. "Updated Facts on the U.S. Distributions of Earnings, Income, and Wealth." Federal Reserve Bank of Minneapolis *Quarterly Review* 26 (Summer): 2–35.

Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron. 2004. "Consumption and Risk Sharing over the Life Cycle." *Journal of Monetary Economics* 51 (April): 609–33.

Suits, Daniel B. 1977. "Measurement of Tax Progressivity." *American Economic Review* 67 (September): 747–52.

Tauchen, George. 1986. "Finite State Markov Chain Approximations to Univariate and Vector Autoregressions." *Economic Letters* 20: 177–81.

Ventura, Gustavo. 1999. "Flat Tax Reform: A Quantitative Exploration." *Journal of Economic Dynamics and Control* 23 (September): 1,425–58.