

National and Regional Housing Vacancy: Insights Using Markov-switching Models

Jeffrey P. Cohen
Professor of Real Estate and Finance
University of Connecticut
Jeffrey.Cohen@business.uconn.edu

Cletus C. Coughlin
Senior Vice President and Chief of Staff
Federal Reserve Bank of St. Louis
Cletus.C.Coughlin@stls.frb.org

Jonas Crews
Research Associate
Walton Family Foundation
jcrews@wffmail.com

7/8/2019

Abstract

We examine homeowner vacancy rates over time and space using Markov-switching models. Our theoretical analysis extends the Wheaton (1990) search and matching model for housing by incorporating regime-switching behavior and interregional spillovers. Our approach is strongly supported by our empirical results. Estimations, using constant-only as well as Vector Autoregressions, allow us to examine differences in vacancy rates as well as explore the possibility of asymmetries within and across housing markets, depending on the state/regime (e.g., low or high vacancy) of a given housing market. Estimated vacancy rates, conditional on the vacancy regime, which are found to be persistent, vary across regions in all Markov-Switching Vector Autoregression (MS-VAR) models. Models allowing for interregional effects via lagged vacancy rates and controls relating to migration tend to perform better than models lacking this feature. These models track vacancies well. Noteworthy is their performance during the Great Recession/Financial Crisis. The importance and diversity of interregional effects are demonstrated, and vacancies in a specific Census region are affected by vacancies in other regions. Moreover, the sizes of these effects depend on the vacancy state of the specific region.

JEL Codes: R31, C24, R11

Keywords: housing vacancy, Markov switching, search and matching, interregional spillovers, Vector Autoregressions

Our work has benefited from research assistance by Andrew E. Spewak and from comments at the 2017 Urban Economics Association meetings in Vancouver, the 2017 2nd Homes up International Conference at the Leibniz-Institute of Ecological Urban and Regional Development in Dresden, and the 2018 46th AREUEA National Conference in Washington, DC. Comments from Shulin Shen, Lara Loewenstein, and two anonymous referees were especially helpful. The views expressed are those of the authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Introduction

Vacancies in the housing market (i.e., unoccupied housing units) are similar to unemployment in the labor market in that some level is desirable and expected in a well-functioning market. For example, homeowners may experience changes in their family or employment situations such that their existing house no longer meets their needs in terms of location and/or house characteristics, and, after a search, more appropriate housing is purchased. Thus, until the first house is sold, a given homeowner may own two houses, one of which is vacant. Vacancies also arise for other reasons. For example, newly-constructed houses may be vacant for a period before occupancy.¹

In our analysis, we explore the possibility that vacancy rates depend on the state of the housing market. Specifically, we estimate separate vacancy rate models for what we term a low-vacancy state and for a high-vacancy state, with the high-vacancy state being relatively higher than the low-vacancy state. There is neither an absolute minimum difference between these states, nor is there an absolute maximum for a low-vacancy state or an absolute minimum for a high-vacancy state. In some cases three states are identified: low, intermediate, and high. Upon entering a specific regime, the regime is highly persistent. Numerous circumstances can cause vacancies to rise and lead to a high-vacancy state. National and local recessions as well as geographic shifts in demand can generate rising vacancies and lead to what we characterize as a high-vacancy state.

Persistently high vacancy rates can indicate housing market problems. For example, high and geographically-concentrated vacancy rates can indicate an inefficient allocation of resources and can breed vandalism and crime. In addition, local governments may confront substantial management and demolition costs in dealing with abandoned, run-down houses, while financial intermediaries may incur costs/losses in taking possession of and ultimately selling vacant houses associated with foreclosures.

Recent history provides a stark example of a recession that propelled a sharp upward movement in vacancy rates. As shown in Figure 1, during the housing crisis associated with the Great Recession, homeowner vacancy rates in the United States reached levels far greater than at any time since measurement began in the mid-1950s. Rates reached 2.9% during the housing crisis. As far back as records have been kept, rates had always been below 2.0%. As of Q1/2019, the national vacancy rate was 1.4%.

¹ Two other reasons might also be noted. An individual might own two houses, one of which is occupied most of the year and the other which is used for vacations. The vacation house will likely be unoccupied for large portions of a year. Such vacancy can be viewed as intentional and does not suggest any housing market problem. What is termed long-term vacancy (i.e., non-seasonal housing units that have been vacant for an unusually long period of time) suggests the possibility of some fundamental problem, such as a declining neighborhood. See Molloy (2016) for a recent analysis of long-term vacancies.

From a regional perspective, homeowner vacancy rates differ across space and, similar to the United States as a whole, change over time.² Various factors, such as the cost of holding vacant units, search costs and the matching process, expectations about future housing prices, demand for specific housing characteristics, the quantity and quality of intermediaries, the specific characteristics of the existing housing stock, transaction costs, land use regulations, and credit market imperfections can differ regionally and over time.³ In addition, regional vacancy rates can be affected differentially by national economic shocks and by regional shocks. These differences across space and changes over time create the possibility of different regional vacancy rates. For example, as of Q1/2019, the vacancy rate was 1.2% in the Northeast Census region, 1.2% in the Midwest, 1.7% in the South, and 1.0% in the West.

We attempt to explain the spatial variation theoretically by adapting and extending the Wheaton (1990) search and matching model for housing to include a direct connection between interregional housing market differences via migration primarily motivated by job opportunities. Changing jobs is a key factor in migration within the United States. For example, Ihrke (2014) found that with-in county moves are associated with housing-related issues, while between county moves are associated with job-related issues.⁴ Molloy, Smith, and Wozniak (2011) report that approximately 1.5% of the population moves between two of the four Census regions annually.

Empirically, we provide insights into vacancy rates for the United States and the four Census regions via Markov-switching models. In the context of housing vacancy, this is the first known research that uses Markov-switching models.⁵ These models allow us to deal with the large changes (or regime shifts) in vacancy rates during major economic shocks and recoveries. They also allow us to test the level of interdependence of the Census regions and how that interdependence varies across regimes, by incorporating other regions' vacancy rates into the empirical models. Our research appears to be the first attempt to empirically identify and explain interregional vacancy relationships. Of particular interest is how our models perform around the

² While our research examines owner-occupied housing, a similar analysis could be done for rental property. An early study is Gabriel and Nothaft (1988). From the mid-1950s to the present, the correlation between homeowner and rental vacancy rates is 0.77. These rates move in opposite directions prior to the Great Recession, a period worthy of a separate study, as well as for a much more complicated analysis that considers the interaction between rental and homeownership vacancies.

³ See Fritzsche and Vandrei (2014) for a discussion of the theoretical causes of vacancies and a summary of empirical findings. A related paper by Cheshire, Hilber, and Koster (2015) finds that regulations restricting new house construction increases rather than decreases vacancy rates. Regulations lead to higher prices providing incentives for occupying houses, but also impede the matching process. Empirically, for a sample of local housing markets in England, this latter effect, which tends to increase vacancies, dominates the former effect.

⁴ Changes in internal migration have garnered much research attention in recent years. Molloy, Smith, and Wozniak (2011) highlight that migration rates between Census regions, states, and MSAs have trended downward since the early 1980s. While migration rates have declined, the population changing residences is large. Molloy, Smith, and Wozniak (2017) argue that a decline in job changing has caused migration to decline; however, they are unable to find specific reasons for the decline in job changing.

⁵ For a related article not using Markov switching, see Zabel (2016). Zabel develops and estimates a dynamic model of the housing market in which vacancies are related to an error-correction process.

financial crisis and Great Recession, a period of substantial upheaval and distress in U.S. housing markets.⁶

The rest of the paper proceeds as follows. We review the literature focused on vacancies and develop an extension of the Wheaton (1990) model. This extension leads to the estimation of a Markov-Switching Vector Autoregression (MS-VAR) model in the empirical section. We then discuss the basics of Markov-switching models, followed by an examination of our empirical results. The last section concludes.

Housing Vacancy: Background Literature

The possibility of vacant housing is illustrated in Figure 2 via a simple supply and demand framework. As shown in the top panel, let the supply of housing units be fixed at S_0 . With a demand for occupied housing units, D_0^o , and a total demand for housing units, D_0^+ , then the equilibrium price would be P_0 , the equilibrium quantity of occupied housing units would be D_0 , and the equilibrium quantity of all housing units would be the same as the quantity supplied of all housing units (i.e., S_0). Associated with this equilibrium is a vacancy rate (VR) equal to $((S_0 - D_0)/S_0) \times 100$.

Next, suppose demand for housing declines to D_1^+ for all units and D_1^o for occupied units. Note that the shifts are parallel and equidistant.⁷ Such a decline puts downward pressure on housing prices. If price adjusts completely and instantaneously, then the new equilibrium price would be P_1 and the equilibrium quantities of occupied and total housing would be unchanged. Vacancies and the associated vacancy rate would also remain unchanged.

However, there are frictions that might prevent a complete and instantaneous adjustment. Molloy (2016) provides a number of arguments and references that would lead one to expect that price would not fall immediately to P_1 . Goodman and Ittner (1992) argue that owners tend to overestimate the value of their property and Genesove and Mayer (2001) find a reluctance to sell property for less than property owners judge as its worth.⁸ Especially when demand declines, property owners might not recognize the decline in the value of their property. Anenberg (2016) and Guren (2018) argue that owners set their asking prices based on the transactions prices of comparable properties sold recently. Thus, given a decrease in demand, some owners might

⁶ For a summary of housing price developments during the financial crisis, see Cohen, Coughlin, and Lopez (2012). For a review of recent literature focused on foreclosures and sales of distressed properties, see Cohen, Coughlin, and Yao (2016).

⁷ This is a strong assumption that causes vacancies and the associated vacancy rate to remain unchanged. If the demand for occupied housing shifts relatively more (less) than the demand for total housing, then vacancies and the vacancy rate increase (decrease) relative to the initial equilibrium.

⁸ The loss aversion argument of Genesove and Mayer (2001) illustrates how psychological concepts from behavioral economics can affect vacancy rates, a topic discussed in Fritzsche and Vandreli (2014). Stein (1995) offers another argument related to Genesove and Mayer (2001). In a declining house price environment, potential sellers are adversely affected by the resulting decline in wealth and liquidity. This produces a reluctance to sell because ever-lower prices decrease the potential seller's options for relocating as their capability of making a given downpayment is reduced.

have unrealistic expectations concerning the values of their property. Finally, if an owner is offered less than the mortgage amount, a common occurrence during the recent housing crisis, then sales become quite complicated. To complete the sale, either the lender must forgive the difference between the mortgage amount and the transaction price or the seller must make up the difference.

In light of the preceding frictions, assume price declines only to P_2 . Given this partial adjustment, the quantity of occupied units is D_1 . As a result, the vacancy rate, VR , increases to $((S_0 - D_1)/S_0) \times 100$, higher than its previous rate.⁹ Turning to the case of an increase in demand, one can also argue that price will not adjust completely and instantaneously. If so, then the vacancy rate will decline below its previous rate.

The next question is what happens when more time is allowed for adjustments in the housing market. Because a supply curve for housing for periods longer than the short run likely has a positive slope and can shift, the quantity of housing units can adjust upward via new construction and downward by depreciation/teardowns. On the demand side, whether the shock is temporary or permanent is of utmost importance. If the shock is temporary, then one should expect price and vacancy to return to their original values. On the other hand, if the shock is permanent, then price and quantity will adjust further and their effects on the vacancy rate are uncertain without more detailed information on various quantitative relationships as well as the cause of the shock.¹⁰

To illustrate the preceding point, assume a permanent decline in demand due to outmigration from a region. Such a decline is shown in the bottom panel of Figure 2 and is assumed to be parallel and equidistant. As a result, demand for all housing units falls to D_1^+ and for all occupied units falls to D_1^o . Similar to the case with a fixed housing supply, such a decline puts downward pressure on housing prices. If price adjusts completely and instantaneously, the new equilibrium price would be P_1 and the equilibrium quantities of total and occupied housing would decline to S_1 and D_1 . The number of vacancies would be unchanged and the vacancy rate would increase. However, as discussed previously, the actual shifts of occupied housing and total housing need not be parallel and equidistant. It is possible that in this new case the decline in occupied housing is relatively less (more) than the decline in total housing demand. This creates the possibility that the new vacancy rate is lower (higher) than the original vacancy rate. What remains certain is that the quantity of housing supplied is now less, the quantity of occupied housing is less, and the price of housing is less.

⁹ The effect of vacancies on housing prices has been a topic of increased attention due to the housing bubble. See Zabel (2016) and Whitaker and Fitzpatrick (2013).

¹⁰ Colwell (2002) identifies two components of the demand for vacancies: a transactions component, which is associated with the natural vacancy state, and a speculative component. This speculative component is tied to expectations of future prices relative to current prices. When future prices are in line with current prices (i.e., neither too high nor too low), then the speculative demand for vacancies is zero.

The foundations for the preceding supply and demand discussion are a search and matching model. Wheaton (1990) provided a basic search and matching model that yields a vacancy rate. Prior to extending his model, we use it to illustrate the basics of the search and matching process.¹¹

Assume that there are two types of households (e.g., families and singles) and two types of housing units (e.g., large and small). Households are viewed as “matched” when a family is in a large unit and a single is in a small unit and “mismatched” when a family is in a small unit and a single is in a large unit. A matched household becomes mismatched when a single becomes a family or a family becomes a single.¹² A household moves from mismatched to matched by finding and purchasing the other appropriate unit. Then the previously occupied house is put up for sale. An additional simplification in Wheaton’s model is that while households can change between types, the aggregate distribution of households by type is stable.

At any time, a given household is in one of three occupancy states: 1) HM_i indicates a matched household (i.e., the household occupies an appropriate housing unit); 2) HD_i indicates a matched household with a house to sell (i.e., the household owns two units - one being appropriate and not for sale and the other inappropriate for them and for sale); and 3) HS_i indicates a mismatched household looking for an appropriate unit. With a fixed number of households and units of each type, households change states in the model according to the following differential equations:

$$(1) \dot{HS}_i = -m_i HS_i - \beta_i HS_i + \beta_j HM_j$$

$$(2) \dot{HD}_i = -q_i HD_i + m_i HS_i + \beta_j HD_j - \beta_i HD_i$$

$$(3) \dot{HM}_i = -\dot{HS}_i - \dot{HD}_i, i = 1, 2, j \neq i.$$

β_i is the transition rate from preferring type i to preferring type j . Wheaton (1990) defines a mismatched household as one that cannot find an appropriate house instantaneously, so this leads to a search process that produces matches. Wheaton (1990) assumes the arrival of matches in the entire market occurs with a Poisson process, with m_i the match rate. Sales of vacant houses also

¹¹ Not surprisingly, the literature on this topic has advanced substantially since 1990. For a recent literature review see Han and Strange (2015). See Williams (1995) for a continuous time version of Wheaton’s model. The role and impact of bargaining in the housing search process is examined by Ihlanfeldt and Mayock (2012) and by Merlo and Ortalo-Magné (2004). Piazzesi, Schneider and Stroebel (2015) extend the housing market matching literature by allowing for multiple market segments and heterogeneous searchers.

¹² Obviously, this is a major simplification. Households tend to move when a job change creates a large increase in commuting distance or the household experiences some other change in income or family size that makes a current house inadequate. Ihrke (2014) found that with-in county moves are associated with housing-related issues, while between-county moves are associated with job-related issues.

are assumed to occur with a Poisson process, q_i . The sales of vacant houses equals the flow of house purchases, so that:

$$(4) \quad q_i V_i = m_i H S_i$$

Equation (1) is the time rate of change of becoming mismatched. The first term is those who have become newly matched thus reducing the rate of becoming mismatched, the second term is those exiting from i (mismatched and looking) to j , and the third term captures the newly mismatched who are moving from j to i . Equation (2) is the time rate of change of those households being matched with a house to sell. The first term captures the sale of vacant houses of type i , thus reducing the rate of being matched with a house to sell. The second term captures those households who have become matched and now have a house to sell. The third term captures households who are newly matched with a type i house to sell. The fourth term captures those households who have changed from i into j . Equation (3) is time rate of change of those becoming matched, which is simply the difference between the negative of those households becoming mismatched and looking and those who have become matched with a house to sell.¹³

An Extended Search and Matching Model

¹³ Wheaton (1990) further simplifies the analysis by assuming that the two types of households are identical in number and behavior. With $\beta_1 = \beta_2$, $V_1 = V_2$, $H_1 = H_2$, $m_1 = m_2$, then $HS_1 = HS_2$, $HM_1 = HM_2$, and $HD_1 = HD_2$. The effect of this simplification is to reduce the system of six differential equations to the following two equations:

$$(i) \quad \dot{H}S_i = -HS(2\beta + m) + \beta H - \beta HD$$

$$(ii) \quad \dot{H}D_i = mHS(1 - HD/V)$$

Equations (9) and (10) allow the determination of HS and HD . The resulting steady state is characterized by:

$$(iii) \quad HS = (\beta(H - V))/(\beta + m)$$

and

$$(iv) \quad HD = V.$$

Equation (iii) captures mismatched households looking to buy a house, while equation (iv) captures matched households who have a (vacant) house to sell. In our adaptation, we retain the generalization that there can be two types, i and j , in our model and analysis.

We extend and adapt Wheaton's (1990) model to more explicitly consider interregional effects.¹⁴ One change is to incorporate an interactive aspect in which those attempting to buy and sell vacant homes of a given type (i.e., in one region) take the actions of all others (i.e., in other regions) as given. To accomplish this we modify the previous definitions to focus on two regions (or one region, i , and the rest of the country, j), rather than on households in one region. A second adaptation is allowing for multiple vacancy regimes, denoted by a subscripts on the model structural parameters. We also consider a different notion of "mismatch" here than in Wheaton (1990). Rather than mismatches due to owning a house that is inappropriately sized for a given household (i.e., a single individual owns a large house or a family owns a small house), in our context mismatches are associated with a desire to live in one region, but owning a home in the other region. Hence, $HS_{i,t}$, $HD_{i,t}$, and $HM_{i,t}$ are respectively those who live in region j and want to move to region i at a point in time; those who happily live in region i at a point in time, but still have a vacant home in region j ; and those who happily live in region i at a point in time. Also, $q_{s,i}$ is the probability of sale for a vacant region i house in regime s , $m_{s,i}$ is the probability of a mismatched household finding a region i home to purchase in regime s , and $\beta_{s,i}$ is the transition rate from preferring region i to preferring region j in regime s . Given the focus on "mismatched" households, together with the desire of these people to become matched, the mechanism by which vacancies in a particular region are determined is through-migration between regions. This is demonstrated in more detail below.

We have added time subscripts (denoted as t) to our variables and regime (or state) subscripts (denoted as s) to the parameters in equations (3) and (4). We additionally assume that all homes are owned by the market participants we explicitly consider in the model, which leads to $V_{i,t} = HD_{j,t}$ because the only vacant homes in region i would be owned by those who have moved to region j but still own a home in region i . The additions and assumption yield:

$$(3') \quad (HM_{i,t} - HM_{i,t-1}) = -(HS_{i,t} - HS_{i,t-1}) - (V_{i,t} - V_{i,t-1})$$

$$(4') \quad q_{s,i}V_{i,t} = m_{s,i}HS_{i,t}$$

Considering the activities of buyers and sellers of vacant homes in region i , we use equations (1) and (2) from above (i.e., Wheaton (1990)), slightly adjusted by adding time (t) and regime or state (s) subscripts and acknowledging that $m_{s,i}HS_{i,t} = q_{s,i}V_{i,t} = q_{s,i}HD_{j,t}$.¹⁵

¹⁴ For an alternative housing market model using a search-and-matching approach, see Lisi (2015). This model, similar to our model, highlights the existence of vacancies; however, in contrast to our model, it does not address interregional effects.

¹⁵ Wheaton (1990) drops the subscripts to indicate that the economy (i.e., one region) is in steady state. We are retaining the subscripts here to allow for multiple regions, each of which is taking the other regions' variables as exogenous. Without keeping these subscripts, it would be as if we were saying there was one large homogeneous region in the country, which is not consistent with the problem we are studying. In other words, these subscripts are a driver for our hypothesis that vacancies differ and are interdependent across regions.

$$(5) \quad \dot{HS}_i \sim HS_{i,t} - HS_{i,t-1} = -q_{s,i} HD_{j,t} - \beta_{s,i} HS_{i,t} + \beta_{s,j} HM_{j,t}$$

$$(6) \quad \dot{V}_i \sim V_{i,t} - V_{i,t-1} = -m_{s,i} HS_{i,t} + m_{s,j} HS_{j,t} + \beta_{s,i} HD_{i,t} - \beta_{s,j} HD_{j,t}.$$

These equations are different from equations (1) and (2) above (i.e., as in Wheaton (1990)) in the following ways. First, we assume a non-steady state scenario. Second, $-q_{s,i} HD_{j,t}$ (the number of homeowners who moved to region j but still own a home in a region, i, in a given regime, s, at a particular time, t) is on the right side of (5) instead of the scenario in equation (1) above where $-m_i HS_i$ is on the right side. Equation (6) is different from Wheaton (1990) (equation (2) above) since we are considering actions of housing market participants in region i, and $HD_{j,t}$ households (i.e., $V_{i,t}$) are the sellers in region i; we also have $-m_{s,i} HS_{i,t}$ on the right side of (6), as opposed to having $-q_{s,i} HD_{j,t}$ in equation (2) above (the analogous equation of Wheaton (1990)).

Solving (5) for $HS_{i,t-1}$, substituting this equation for $HS_{i,t-1}$ into (6), along with using the fact that $m_{s,i} HS_{i,t} = q_{s,i} V_{i,t} = q_{s,i} HD_{j,t}$, and then solving for $V_{i,t}$, yields:

$$(7) \quad V_{i,t} = \left[\frac{\beta_{s,i} - 1 + \beta_{s,j}(\beta_{s,i} - 1) + m_{s,i}q_{s,i}}{(\beta_{s,i} - q_{s,i} - 1)} \right] V_{i,t-1} + \left[\frac{m_{s,j}(\beta_{s,i} - 1)}{(\beta_{s,i} - q_{s,i} - 1)} \right] HS_{j,t-1} \\ + \left[\frac{\beta_{s,i}(\beta_{s,i} - 1)}{(\beta_{s,i} - q_{s,i} - 1)} \right] V_{j,t-1} + \left[\frac{-m_{s,i}\beta_{s,j}}{(\beta_{s,i} - q_{s,i} - 1)} \right] HM_{j,t-1}$$

Because the parameters in the brackets are transition rates, they are each expected to vary between 0 and 1. Given this, the anticipated signs of the terms in brackets implied by the theory are as follows. The term in front of $V_{i,t-1}$ is indeterminate. The terms preceding $V_{j,t-1}$, $HM_{j,t-1}$, and $HS_{j,t-1}$ are expected to be positive. We take this equation to the data, generalized to incorporate more than one j region, and estimate a Markov switching vector autoregression (MS-VAR) model, in order to examine how well our theory holds up empirically. The MS-VAR approach fits our theoretical model well by allowing us to simultaneously estimate equation (7) for multiple regions, with the designation of region i and regions j alternating across equations. To completely internalize the empirical dynamics, we add separate equations to our MS-VAR model for $HM_{j,t}$, $HS_{j,t}$, and $V_{j,t}$, each as a function of ($V_{i,t-1}$, $HS_{j,t-1}$, $V_{j,t-1}$, $HM_{j,t-1}$).

Underlying the parameters ($m_{s,i}$, $m_{s,j}$, $\beta_{s,j}$, $q_{s,i}$, $q_{s,j}$, $\beta_{s,i}$) are economic and other factors affecting the supply and demand of housing in each region. These parameters can vary across regimes. While there may be more than two regimes, we consider the situation that fits the data

the best. In general, in our empirical results below this implies that there are two possible regimes for 3 out of 4 of the U.S. regions – one associated with a low-vacancy situation, and another one with a relatively high-vacancy situation. One region – the West – has a preferred fit with a 3 state model: low-vacancy, intermediate-vacancy, and high-vacancy regimes.

The parameters $(m_{s,i}, m_{s,j}, \beta_{s,j}, q_{s,i}, q_{s,j}, \beta_{s,i})$ all have regime subscripts, s , where for the 2 state model, $s = \{1, 2\}$ (or equivalently, in a two-state model for our application, $s = \{\text{low vacancy, high vacancy}\}$). A key point is that there is a continuous equation constituting $V_{i,t}$ for the high vacancy regime, and a continuous equation constituting $V_{i,t}$ in a low vacancy regime. In other words, there are effectively jump discontinuities at the times of regime switches, with a binary regime indicator, but $V_{i,t}$ is continuous throughout a particular regime until the next switch. As a concrete example, suppose region i is in a low-vacancy regime (i.e., $s = 1$) from time t to time $t+10$, before a switch to a high vacancy regime (i.e., $s = 2$) at time $t+11$. From time t to time $t+10$, $V_{i,t}$ will be characterized by a continuous, linear equation between it and $V_{j,t}$. Then, from $t+11$ until the next regime switch, there will be a separate, regime 2-related continuous, linear equation describing the relationship between $V_{i,t}$ and $V_{j,t}$.

It is our hypothesis that major shifts in the U.S. economy, and thus the national housing market, also cause shifts in the movements of households across regional housing markets. For that reason, we attempt to fit switching models to the data to capture any shifts in the relationship between regional housing markets. Shocks are not restricted to the national economy as regional shocks are possible as well.

The relationships between $V_{i,t}$ and $HS_{j,t}$, $V_{j,t}$, and $HM_{j,t}$ can vary in different vacancy regimes, due to the regime subscripts underlying the parameters $m_{s,i}, m_{s,j}, \beta_{s,j}, q_{s,i}, q_{s,j}, \beta_{s,i}$. We aim to determine the sign and magnitude of these relationships, which can be different in each of the two (or three) vacancy states. We accomplish this by estimating a two- (or three-) state, reduced form Markov switching vector autoregression (MS-VAR) model.

A Markov-switching Model for Vacancy Rates

While we initially estimate constant-only and own-lag-only Markov-switching models to motivate the usage of regime switching, our model of interest is focused on the four U.S. Census regions. It is a reduced-form VAR model centered on a four-region generalization of equation (7). The preferred parametrization of this model would involve twelve equations, one for each region's V , HS , and HM variables as functions of the eleven other V , HS , and HM variables. This would allow a complete internalization of the interregional dynamics implied by equation (7). To illustrate, the equation for the South region's vacancy rate would be the following:

$$(8) V_{S,t} = C_{S,t} + a_{S,t} V_{S,t-k} + b_{S,t} HS_{S,t-k} + c_{S,t} HM_{S,t-k} + d_{S,t} V_{M,t-k} + e_{S,t} V_{N,t-k} + f_{S,t} V_{W,t-k} + g_{S,t} HS_{M,t-k} + h_{S,t} HS_{N,t-k} + i_{S,t} HS_{W,t-k} + j_{S,t} HM_{M,t-k} + k_{S,t} HM_{N,t-k} + l_{S,t} HM_{W,t-k} + \varepsilon_t,$$

where subscripts S, M, N, and W indicate South, Midwest, Northeast, and West, respectively.

$V_{S,t-k}$ and $a_{S,t}$, as examples, represent the 1st through kth lags of $V_{S,t}$, and the corresponding parameter vector of length k. While equation (7) implies that only one lag be included, we allow for the possibility of $k > 1$, as dictated by estimation results, to deal with the possibility of housing market frictions resulting in additional lags being necessary.

Unfortunately, data limitations do not allow for the generation of individual region values for HS and HM , nor do they provide sufficiently many degrees of freedom to estimate all parameters. We are able to estimate models involving individual region vacancy measures, but with HS and HM aggregated to three regions at a time. We use the three-region-aggregated measures to estimate four different models, each producing an equation as close as possible to the preferred generalization of equation (7) for one of the regions. As an example, the model oriented toward the South is parametrized as follows:

$$(9a) V_{S,t} = C_{S,t} + a_{S,t} V_{S,t-k} + b_{S,t} HS_{MNW,t-k} + c_{S,t} V_{M,t-k} + d_{S,t} V_{N,t-k} + e_{S,t} V_{W,t-k} + f_{S,t} HM_{MNW,t-k} + \varepsilon_t$$

$$(9b) V_{M,t} = C_{S,t} + g_{S,t} V_{S,t-k} + h_{S,t} HS_{MNW,t-k} + i_{S,t} V_{M,t-k} + j_{S,t} V_{N,t-k} + k_{S,t} V_{W,t-k} + l_{S,t} HM_{MNW,t-k} + \varepsilon_t$$

$$(9c) V_{N,t} = C_{S,t} + m_{S,t} V_{S,t-k} + n_{S,t} HS_{MNW,t-k} + o_{S,t} V_{M,t-k} + p_{S,t} V_{N,t-k} + q_{S,t} V_{W,t-k} + r_{S,t} HM_{MNW,t-k} + \varepsilon_t$$

$$(9d) V_{W,t} = C_{S,t} + s_{S,t} V_{S,t-k} + t_{S,t} HS_{MNW,t-k} + u_{S,t} V_{M,t-k} + v_{S,t} V_{N,t-k} + w_{S,t} V_{W,t-k} + x_{S,t} HM_{MNW,t-k} + \varepsilon_t$$

$$(9e) HS_{MNW,t} = C_{S,t} + y_{S,t} V_{S,t-k} + z_{S,t} HS_{MNW,t-k} + \alpha_{S,t} V_{M,t-k} + \beta_{S,t} V_{N,t-k} + \gamma_{S,t} V_{W,t-k} + \delta_{S,t} HM_{MNW,t-k} + \varepsilon_t$$

$$(9f) HM_{MNW,t} = C_{S,t} + \vartheta_{S,t} V_{S,t-k} + \mu_{S,t} HS_{MNW,t-k} + \pi_{S,t} V_{M,t-k} + \rho_{S,t} V_{N,t-k} + \tau_{S,t} V_{W,t-k} + \varphi_{S,t} HM_{MNW,t-k} + \varepsilon_t,$$

where the subscript MNW represents the aggregated region encompassing the Midwest, Northeast, and West, or the portion of the U.S. excluding the South region. Variants of equations (9a)-(9f) oriented toward another region would thus incorporate HS and HM values for the portion of the U.S. exclusive of that region. Equation (9a) is as close as we can get to the fully generalized equation for $V_{S,t}$, (equation (8)). In correspondence with equation (7), with the South region as region i, HS_{MNW} and HM_{MNW} are the appropriate aggregated HS_j and HM_j . While each

multiple-equation model estimated will only have one equation of significant interest to us, it is important to note that central to the estimation of any switching model is the estimation of vectors of probabilities for being in a given state across the sample period. These state probability vectors would likely be very different if we attempted to estimate only each model's equation of interest in a single-equation model, given that key model dynamics would not be internal to the estimation.

In terms of our theoretical model in (7), the parameters in (9a), in a one-lag model, correspond to those in (7) as follows:

$$\begin{aligned}
a_{s,t} &\approx \left[\frac{\beta_{s,S} - 1 + \beta_{s,MNW}(\beta_{s,S} - 1) + m_{s,S}q_{s,S}}{(\beta_{s,S} - q_{s,S} - 1)} \right] \\
b_{s,t} &\approx \left[\frac{m_{s,MNW}(\beta_{s,S} - 1)}{(\beta_{s,S} - q_{s,S} - 1)} \right] \\
c_{s,t} &\approx \left[\frac{\beta_{s,S_M}(\beta_{s,S_M} - 1)}{(\beta_{s,S_M} - q_{s,S_M} - 1)} \right] \\
d_{s,t} &\approx \left[\frac{\beta_{s,S_N}(\beta_{s,S_N} - 1)}{(\beta_{s,S_N} - q_{s,S_N} - 1)} \right] \\
e_{s,t} &\approx \left[\frac{\beta_{s,S_W}(\beta_{s,S_W} - 1)}{(\beta_{s,S_W} - q_{s,S_W} - 1)} \right] \\
f_{s,t} &\approx \left[\frac{-m_{s,S}\beta_{s,MNW}}{(\beta_{s,S} - q_{s,S} - 1)} \right],
\end{aligned}$$

where, as examples of notation, $\beta_{s,S}$ is the transition rate from preferring the South to referring the rest of the country, and β_{s,S_M} is the transition rate from preferring the South to preferring the Midwest.

Data, Estimation Procedure and Results

We estimate regional variants of (9a-9f) using vacancy data for each of the four U.S. Census regions: Northeast, Midwest, South, and West, and three-region aggregates of $HS_{j,t}$ and $HM_{j,t}$ with the three regions varying to exclude the variant model's region of interest, per the discussion at the end of the last section. For generality, we include a constant (intercept) term in this MS-VAR reduced form estimation.

All data are quarterly. Vacancy data came directly from U.S. Census Bureau data. HS and HM vectors have been constructed from their definitions using regional owner-occupied housing stock, occupancy rate, population, and migration data from the U.S. Census Bureau. For the first several years of the sample, unavailable regional owner-occupied housing stock data were proxied by the regional population divided by the national number of individuals per household. Many of the underlying data vectors were yearly measures linearly extrapolated across intermediate quarters. $HS_{j,t}$ and $HM_{j,t}$ are proxied as follows: $HS_{j,t}$ is proxied as the number of homeowners in region i at time t times the migration rate out of region i at time $t+1$. $HM_{j,t}$ is proxied by the number of homeowners in region j at time t less [time- t outmigration from region i divided by the time- $t-1$ average number of region i families per household times the region- i homeownership rate at time $t-1$] minus [migration into region i divided by the weighted average number of individuals per household in region j times the weighted average homeownership rate in region j .]

All final data have been detrended, seasonally adjusted (using the Census' X-13 package,) and reshaped to simplify the information going into the empirical model and allow for accurate interpretations. Specifically, the seasonally adjusted vacancy rate is used in place of the number of vacancies, given the measure's tractability and common use as a housing market indicator. HS and HM have been seasonally adjusted, detrended, and normalized, with the latter bringing data values to ranges more comparable to the vacancy rates and hopefully bringing simplicity to a complex, nonlinear estimation procedure.

Simple National and Census Region Models

We begin with a special variant of equation (9a), for the U.S. as a whole (i.e., assuming there is only one region). This simple model essentially is a U.S. variant of (9a) that imposes $a_{s,t}$, $b_{s,t}$, $c_{s,t}$, $d_{s,t}$, $e_{s,t}$ and $f_{s,t}$ to equal vectors of zeros. Before examining the estimation results, let's re-examine the raw data for the vacancy rate for the United States from 1956:Q1 through 2019:Q1. Figure 1 shows that the U.S. vacancy rate has normally been less than 2.0 percent. Only the period associated with the housing crisis exhibited rates in excess of 2.0 percent. Prior to the Great Recession, no period, even recessions, exhibited a national vacancy rate greater than 2.0 percent.

The results of this simple Markov switching model for the United States are shown in Table 1. State 1 (2) is the low- (high-) vacancy state, p_{11} (multiplied by 100 to yield a percentage) is the probability of remaining in state 1, and p_{21} (similarly adjusted) is the probability of moving from state 2 to state 1.

For the United States the estimated low-vacancy rate is 1.16 percent, while the high-vacancy rate is 1.81 percent. As can be seen by the estimates of p_{11} and p_{21} , both states are highly persistent. The probability of starting in state 1, the low-vacancy state, and remaining in state 1 is 98.3 percent. Thus, once the nation is in a specific state/regime the state is likely to

persist. The probability of starting in state 2, the high-vacancy state, and moving to state 1 is 1.0 percent. These two probabilities also reveal the probability of starting in state 1 and moving to state 2, 1.7 percent, and the probability of starting in state 2 and remaining in state 2, 99.0 percent.

Figure 3 shows the actual vacancy rates compared with the estimated smoothed vacancy rate predictions and the estimated smoothed probabilities of the United States being in its high-vacancy state. With the exception of the Great Recession/Financial Crisis, the predicted vacancy rate tracks the actual vacancy rate quite closely. In turn, the estimated probability of the high vacancy state matches the actual and estimated vacancy rates. Generally speaking, the estimated probabilities suggest that the low-vacancy state prevailed for the majority of the period from the mid-1950s to the early 1980s, while subsequently, the high vacancy state has dominated.¹⁶

We next turn to the estimates for the Census regions, for a simple model assuming no inter-dependencies across regions at the regional level, where we estimate regional variants of (9a) using a Markov Switching model for each region, imposing $a_{s,t}$, $b_{s,t}$, $c_{s,t}$, $d_{s,t}$, $e_{s,t}$ and $f_{s,t}$ to equal vectors of zeros. In this region by region analysis, one observes much diversity across regions. The vacancy rate associated with the low-vacancy state varies across regions, ranging from 0.89 percent in the Northeast to 1.46 percent in the West. The rates in the Northeast and Midwest are below the national rate of 1.16 percent, while the rates in the South and West are above the national rate. Estimates for the high-vacancy state also reveal much diversity, ranging from 1.59 percent in the Northeast to 2.18 percent in the South. The rate for the high-vacancy state in the Northeast is below the national average, while each of the other regions is above the national average. It is also noteworthy that the national average is 16 basis points below the high-vacancy rate for the Midwest, 38 basis points for the South, and 36 basis points for the West. In fact, the national rate for the high-vacancy state is closer to the West's rate for the low vacancy state than the high-vacancy state. Finally, relative to its low-vacancy state, the Midwest shows the largest difference of 83 basis points and the Northeast shows the smallest difference of 70 basis points.

Concerning persistence, the results reveal that a given vacancy state in one quarter is likely to prevail in the next quarter. The lowest probability of starting in a low vacancy state and staying in the low vacancy state in the next period is 97.7 percent in the South. The lowest probability of starting in a high vacancy state and remaining in the high vacancy state in the next period is 94.3 percent in the West.

For estimated results for the Northeast, Figure 4 reveals some similarities and some dissimilarities with the estimates for the United States as a whole. Similar to the U.S. estimates, the probability of being in a low-vacancy state is high from the mid-1950s to the early 1980s. A

¹⁶ A close look at the data suggests a three-regime model; however, we found a simple three-regime model would not converge.

difference, however, is that these high probabilities continue for the Northeast until the late 1980s, while they stop for the nation. Similar to the national estimates, the estimates for the low-vacancy state are quite low for the late 1980s to the present. During this latter period, the actual vacancy rate shows much volatility relative to the predicted vacancy rate. This volatility becomes pronounced at the beginning of this century, years before the Great Recession. In terms of vacancy rates, the estimates provide no indication of an imminent return to a low-vacancy state.

For the Midwest, Figure 5 shows a pattern with similarities and dissimilarities to the United States. In contrast to the United States, the Midwest experienced a brief period of high vacancy in the early/mid 1980s and then returned to a low-vacancy state for more than a decade. The Midwest entry into its second period of high vacancy begins in roughly 2001/2002 and, similar to the United States, the region remains in the high-vacancy state. However, recent estimates suggest a slight movement toward the low-vacancy state. Also, similar to the U.S. results, with the exception of the Great Recession/Financial Crisis, the predicted vacancy rate tracks the actual vacancy rate quite closely.

For the South, Figure 6 shows more ups and downs than the United States and the Northeast and the Midwest. In contrast to the United States, estimates for the South indicate three periods of high vacancy – a brief period in the mid-1960s, a longer period in the mid-1980s to early 1990s, and finally a much lengthier period in late 1990s through the present. This last period begins long before the bursting of the housing bubble and even before the Midwest entered its most recent high vacancy state. Based on Figure 6, one sees that the South has generally been in its high-vacancy state since the mid-1980s. However, recent estimates suggest a slight movement toward the low-vacancy state. Also, similar to the U.S. results, with the exception of the Great Recession/Financial Crisis, the predicted vacancy rate tracks the actual vacancy rate quite closely.

For the West, Figure 7 shows a slightly larger upward movement in vacancy during the 1960s than the United States and other regions and a shorter duration of high vacancy during the recent housing crisis. With the exceptions of the late 1970s and the Great Recession/Financial Crisis, the predicted vacancy rate tracks the actual vacancy rate quite closely. What distinguishes the West from the nation and other regions recently, according to this simple model, is that the West has returned to its low vacancy state, while the United States and other regions have not. In fact, the West spent relatively little time in its high vacancy state in comparison to the United States and the other regions.

Switching Models for the U.S. Using Own Lagged Vacancy Rates

Next we consider a somewhat broader model, as in equation (9a), where there is only one region (i.e., the entire U.S., so that the dependent variable is for region i but there is no region j), but allowing lagged vacancies in the one region (i.e., the U.S.) to affect contemporaneous

vacancies. This is equivalent to assuming the parameters other than $a_{s,t}$ in (9a) are all set to zero. In Table 2, we provide the results for two Markov-switching models for the United States based on this restricted version of (9a). Both models use one (own) lag, with one model identifying three regimes and the other identifying two regimes. An examination of BIC values (i.e., -466.72 for the two-state model versus -468.27 for the three-state model) suggests the two-regime model is only slightly less preferred to the three-regime model; however, a comparison of Figures 8 and 9 suggests otherwise. The estimated probabilities shown in Figure 8 indicate that, virtually without exception, the United States has been in a high-vacancy state. Meanwhile, the estimated probabilities for the three-regime model are much more plausible. Figure 9 indicates a low-vacancy state for the first 40 plus years of the sample. Figure 9 then suggests an intermediate-vacancy regime beginning in the late 1990s leading up to the financial crisis. This regime was followed by a persistent high-vacancy regime throughout the financial crisis and slightly beyond. Subsequently, the United States has returned to a low-vacancy state. Thus, despite the close BIC values, we find the three-regime model far more convincing.

The estimates in Table 2 reveal that the own lag is statistically significant for the low- and intermediate-vacancy states, with the coefficient estimate slightly larger in the low-vacancy state. The disruptiveness of the financial crisis is suggested by the low estimate of the lagged coefficient for the high-vacancy state. All of the estimated constants are statistically significant, with the constant providing the explanatory power for the high-vacancy state. Finally, the probabilities of remaining in a specific state when beginning in that state are high and the probabilities of transitioning to another state are low. For example, the probability of remaining in the low-vacancy state is 99.4% and the probability of remaining in the high-vacancy state is 94.1%. Obviously, these probabilities leave only small probabilities for transitioning to another state in the subsequent quarter.

Switching Models for Census Regions: Impact of Interregional Effects

Now we allow for the possibility of multiple regions that could have interdependencies, and consider a MS-VAR model as in equations (9a), (9b), (9c), (9d), (9e) and (9f). We only report estimation results for the regional variants of equation (9a) due to it being our equation of interest and the other equations effectively being controls for overall model dynamics in estimating the state probabilities across time. The results for the Northeast are presented in Table 3 and Figure 10. We found that the best MS-VAR results are generated by a two-state, two-lag model. In Figure 10, we see that the estimated vacancy rates track the actual vacancy rates closely. Based on the estimates, we see persistent regimes. The low-vacancy state prevailed from the late 1960s to the mid-1980s, the late 1990s to the financial crisis in the mid-2000s, and currently from roughly 2011. The Northeast has had two major high-vacancy periods. It entered a high-vacancy state in the mid-1980s that persisted until the late 1990s. Not surprisingly, the Northeast entered another high-vacancy state during the financial crisis, one that persisted for a few years during the U.S. recovery.

Turning to the results in Table 3, for the low-vacancy state, we find the vacancy rates in the Northeast are related only to its own lagged vacancy rates. Both the one-lag and two-lag vacancy rates were positive, statistically significant determinants. In addition, the estimate for the own, one-lag model coefficient varies across regimes. For example, for the own, one-lag model, the estimate is 0.530 for the low-vacancy state and 0.731 for the high-vacancy state. We find a similar result for the South and West, as discussed later in this section. The probability of remaining in the low-vacancy state is 97.6% and the probability of remaining in the high-vacancy state is 97.4%. Vacancy rates in the Northeast were generally not affected in a statistically significant sense by vacancy rates in other regions. For the low-vacancy state, in addition to the own, one-lag vacancy rate, the effect of the one-lag vacancy rate in the Midwest was statistically significant and positively associated with the vacancy rate in the Northeast. For the high-vacancy state, in addition to the effect of the own, one-lag vacancy rate, the one-lag vacancy rate in the West was found to be statistically significant. A higher vacancy rate in the West was associated with a lower vacancy rate in the Northeast. Finally, none of the other control variables were statistically significant.

Turning to the results for the Midwest, which are provided in Table 4 and Figure 11, we found that a two-state, one-lag model worked best. In Figure 11, similar to the results for the Northeast, we see that the estimated vacancy rates track the actual vacancy rates closely. Generally speaking, the Midwest is in the low-vacancy state for the most of the period examined. Other than during the financial crisis, the other high-vacancy periods were brief.

In Table 4, for the low-vacancy state, we find that not only is the vacancy rate in the Midwest related to its own vacancy rate (lagged one period), but also to the lagged vacancy rates in the South and Northeast. In both cases, higher vacancy rates in these regions are related positively to the vacancy rate in the Midwest. In the high-vacancy state, we find that, once again, the vacancy rate in the Midwest is related to its own lagged vacancy rate. In addition, rather than the South or the Northeast, the current vacancy rate in the Midwest is related positively to the lagged vacancy rate in the West. Finally, the other two control variables were statistically significant in both states. In the low-vacancy state, both variables are related positively to the vacancy rate in the Midwest, while in the high-vacancy state, the variables are opposite in sign to each other. Finally, the probability of remaining in the low-vacancy state is 91.4% and the probability of remaining in the high-vacancy state is 98.3%.

Similar to the Midwest, a two-state, one-lag model performed the best for the South. The results for this region are presented in Table 5 and Figure 12. As shown in Figure 12, similar to the other regions, we see that the estimated vacancy rates track the actual vacancy rates closely. Excluding the financial crisis and a persistent aftermath, the South has resided in a low-vacancy state.

For the low-vacancy state, the results in Table 5 indicate that the South's own vacancy rate is related to its lagged own vacancy rate, but also to vacancy rates in Midwest and Northeast. In contrast, in the high-vacancy state, the vacancy rate in the South is related to only its own lagged vacancy rate. For the remaining control variables, there was only one case of statistical significance. Larger values for those living happily outside the South were associated with

higher vacancy rates in the South. Similar to the results for other regions, the states are persistent – nearly 100% for remaining in the low-vacancy state and 97% for remaining in the high-vacancy state.

Turning to the West, the best model is characterized by three states with one lag. See Table 6 and Figure 13 for these results. In Figure 13, once again, similar to the other regions, we see that the estimated vacancy rates track the actual vacancy rates closely. Until the early 1980s, the West was in a low-vacancy state and, until the financial crisis, an intermediate-state generally prevailed. The financial crisis generated a high-vacancy state that persisted until recently.

For each state, as shown in Table 6, the vacancy rate in the West is related positively to its own lagged vacancy rate. The vacancy rate in the West is also related in each state to the lagged vacancy rate in the Midwest. However, this relationship is negative in the low-vacancy state and positive in the other two states. A negative relationship in the low-vacancy state is also found for the impact of the lagged vacancy rates in the Northeast on the vacancy rate in the West. The remaining control variables tended to not be statistically significant. The only exception was in the high-vacancy state where a larger value for those desiring to relocate in the West as associated with a lower vacancy rate in the West. Finally, a given state tends to persist. In the West, the lowest percentage for remaining in a state was 96.4% for remaining in the high-vacancy state.

In summary, we find numerous differences as well as numerous similarities. While three of the regions had two distinguishable states, the West had three states. A given state tends to persist. We find that a region's own lag is always positive, with statistical significance in 10 of the 11 cases. Moreover, the responsiveness varies across a given state's regime. In terms of interregional linkages, each region was found to have, at least, two statistically significant lagged vacancy relationships. While the preferred model for three regions had one lag, the preferred model for the Northeast had two lags. Recall that our theoretical models stress the role of linkages, but do not suggest which regions are more likely to be linked to each other. Overall, of the 33 estimated relationships, 21 were signed positively, while 12 were signed negatively. Of the 21 positive estimates, 8 relationships were statistically significant, while of the 12 negative estimates, 3 relationships were statistically significant. With respect to our controls for those desiring to relocate and those not desiring to relocate from a specific region, we found that 6 (5) of 11 estimates to be positive (negative) for the former variable and 7 (4) of 11 estimates to be positive (negative) for the latter variable. For both variables, 3 of the 11 cases were found to be statistically significant. The estimated relationship was positive for 1 of the 3 cases for the former variable and all 3 of the cases for the latter variable. Thus, while our theory suggests that the coefficients for both $HS_{j,t}$ and $HM_{j,t}$ should be positive, that theoretical implication only seems valid for $HM_{j,t}$. However, it should be noted that our proxies for these variables are somewhat rough (although they are the best proxies that are available). The empirical values of $HS_{j,t}$ could also be highly influenced by foreign in-migration and outmigration, which are not internalized in the theoretical model. Only the Northeast did not have, at least, one statistically significant relationship for the control variables.

Conclusion

We extend the Wheaton (1990) model of search and matching for housing to allow for interactions across regions and multiple vacancy regimes. Our empirical analysis is motivated by the theory, and we examine homeowner vacancy rates using MS-VAR models. We estimate a national model (using a Markov Switching model) and models for regional housing markets (with MS-VAR). Most of these models identify two states (or regimes), one being a low-vacancy state and the other being a high-vacancy state; however, we do find cases where three states are identified. While there are numerous similarities between the national and regional results and the across-regions results, it is clear that no two regions are identical. Our estimations allow us to identify and examine differences in vacancy rates as well as explore the possibility of asymmetries within and across housing markets depending on the state/regime of a given housing market.

After demonstrating the usefulness of Markov-switching models using constants, based on a very restricted version of our extended Wheaton (1990) model, we examine slightly more general models using Vector Autoregressions. First, we examine a model for the United States. Our preferred model has three states. The estimations suggest that the U.S. has generally been in a low-vacancy state. The exception periods have been an intermediate-vacancy state in the years leading up to the financial crisis and then a high-vacancy state associated with the Great Recession/Financial Crisis. Subsequently, the United States returned to a low-vacancy state.

Next, using our full-blown extension of Wheaton's (1990) model, we explore vacancy rates in specific Census regions, especially their connections with other Census regions, with a Markov-switching Vector Autoregression model. We find numerous differences and similarities across regions. A given state tends to persist. Not surprisingly, the Great Recession/Financial Crisis and its aftermath affect the overall results in a number of identifiable cases. While three of the regions had two distinguishable states, the West had three states. We find that a region's own lag is always positive, with statistical significance in 10 of the 11 cases. Moreover, the responsiveness varies across a given state's regime. In terms of interregional linkages, each region was found to have, at least, two statistically significant lagged vacancy relationships. In addition, we identified other interregional linkages via our controls for those desiring to relocate and those not desiring to relocate from a specific region.

References

- Anenberg, Elliot. 2016. "Information Frictions and Housing Market Dynamics," *International Economic Review*, 57(4), pp. 1449-1479.
- Billio, Monica; Casarin, Roberto; Ravazzolo, Francesco; and van Dijk, Herman K. 2016. "Interconnections between Eurozone and US Booms and Busts Using a Bayesian Panel Markov-Switching VAR Model," *Journal of Applied Econometrics*, 31(7), pp. 1352-1370.
- Cermeño, Rodolfo. 2002. "Growth Convergence Clubs: Evidence from Markov-Switching Models Using Panel Data," 10th International Conference on Panel Data Discussion Paper D5-3, July.
- Cheshire, Paul; Hilber, Christian; and Koster, Hans. 2015. "Regulating Housing Vacancies Away? The Paradoxical Effects of Mismatch," Spatial Economics Research Centre Discussion Paper 181, July.
- Cohen, Jeffrey P.; Coughlin, Cletus C.; and Lopez, David A. 2012. "The Boom and Bust of U.S. Housing Prices from Various Geographic Perspectives," Federal Reserve Bank of St. Louis *Review*, 94(5), pp. 341-367.
- Cohen, Jeffrey P.; Coughlin, Cletus C.; and Yao, Vincent. 2016. "Sales of Distressed Residential Property: What Have We Learned from Recent Research?" Federal Reserve Bank of St. Louis *Review*, 98(3), pp. 159-188.
- Colwell, Peter F. 2002. "Tweaking the DiPasquale-Wheaton Model," *Journal of Housing Economics*, 11(1), pp. 24-39.
- Fritzsche, Carolin and Vandreij, Lars. 2014. "Theoretical Causes for Vacancies in the Housing Market," *Credit and Capital Markets*, 47(3), pp. 465-483.
- Gabriel, Stuart A. and Nothaft, Frank E. 1988. "Rental Housing Markets and the Natural Vacancy Rate," *Real Estate Economics*, 16(4), pp. 419-429.
- Genesove, David and Mayer, Christopher. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 116(4), pp. 1233-1260.
- Goodman, Jr., John L. and Ittner, John B. 1992. "The Accuracy of Home Owners' Estimates of House Value," *Journal of Housing Economics* 2(4), pp. 339-357.
- Guren, Adam. 2018. "House Price Momentum and Strategic Complementarity," *Journal of Political Economy*, 126(3), pp. 1172-1218.
- Hamilton, James D. 1989. "A New Approach to the Economic Analysis on Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), pp. 357-384.

- Hamilton, James D. 1994. *Time-Series Analysis*, Princeton, NJ: Princeton University Press.
- Hamilton, James D. and Owyang, Michael T. 2012. "The Propagation of Regional Recessions," *Review of Economics and Statistics*, 94(4), pp. 935-947.
- Han, Lu and Strange, William C. 2015. "The Microstructure of Housing Markets: Search, Bargaining, and Brokerage," in *Handbook of Regional and Urban Economics, Volume 5B*, ed. by Duranton, Gilles; Henderson, Vernon; and Strange, William, North Holland, pp. 813-886.
- Ihlanfeldt, Keith and Mayock, Tom. 2012. "Information, Search, and House Prices: Revisited," *Journal of Real Estate Finance and Economics*, 44(1-2), pp. 90-115.
- Ihle, Rico; von Cramon-Taubadel, Stephan; and Zorya, Sergiy. 2009. "Markov-Switching Estimation of Spatial Maize Price Transmission Processes between Tanzania and Kenya," *American Journal of Agricultural Economics*, 91(5), pp. 1432-1439.
- Ihrke, David. 2014. "Reason for Moving: 2012 to 2013," *Current Population Reports, P20-574*, Washington, DC: US Census Bureau.
- Lisi, Gaetano. 2015. "Real Estate Macroeconomics and the Four-Quadrant Model: DiPasquale-Wheaton-Colwell Meet Mortensen-Pissarides," *Journal of Real Estate Practice and Education*, 18(1), pp. 87-106.
- Merlo, Antonio and Ortalo-Magné, François. 2004. "Bargaining over Residential Real Estate: Evidence from England," *Journal of Urban Economics*, 56(2), pp. 192-216.
- Molloy, Raven. 2016. "Long-Term Vacant Housing in the United States," *Regional Science and Urban Economics*, 59(C), pp. 118-129.
- Molloy, Raven; Smith, Christopher L; and Wozniak, Abigail, 2011. "Internal Migration in the United States," *Journal of Economic Perspectives*, 25(3), pp. 173-196.
- Molloy, Raven; Smith, Christopher L; and Wozniak, Abigail, 2017. "Job Changing and the Decline in Long Distance Migration in the US," *Demography*, 54(2), 631-653.
- Piazzesi, Monika; Schneider, Martin; and Stroebe, Johannes. 2015. "Segmented Housing Search," *NBER Working Paper* w20823.
- Stein, Jeremy C. 1995. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Constraints," *Quarterly Journal of Economics*, 110(2), pp. 379-406.
- Wheaton, William C. 1990. "Vacancy, Search, and Prices in a Housing Market Matching Model," *Journal of Political Economy*, 98(6), pp. 1270-1292.

Whitaker, Stephan and Fitzpatrick IV, Thomas J. 2013. "Deconstructing Distressed-Property Spillovers: The Effects of Vacant, Tax-Delinquent, and Foreclosed Properties in Housing Submarkets," *Journal of Housing Economics*, 22(2), pp. 79-91.

Williams, Joseph T. 1995. "Pricing Real Assets with Costly Search," *Review of Financial Studies*, 8(1), pp. 55-90.

Zabel, Jeffrey. 2016. "A Dynamic Model of the Housing Market: The Role of Vacancies," *Journal of Real Estate Finance and Economics*, 53(3), pp. 368-391.

Table 1

Markov-switching Models: Constant-only Models

	United States	Northeast	Midwest	South	West
State 1					
Constant	1.1644*** (0.0195)	0.8907*** (0.0150)	1.1346*** (0.0225)	1.4075*** (0.0271)	1.4595*** (0.0230)
Sigma	0.1546	0.1627	0.2044	0.2554	0.2616
State 2					
Constant	1.8055*** (0.0307)	1.5918*** (0.0273)	1.9611*** (0.0582)	2.1809*** (0.0341)	2.1652*** (0.0624)
Sigma	0.3724	0.2962	0.4716	0.3652	0.3623
p11	0.9834	0.9958	0.9892	0.9773	0.9831
p21	0.0103	0.0043	0.0162	0.0204	0.0575
Observations	248	248	248	248	248
Log-Likelihood	-42.50	19.10	-49.69	-78.48	-60.24
BIC	118.1	-5.124	132.5	190.0	153.6

Notes: *, **, *** denote 10%, 5%, and 1% statistical significance, respectively. Standard errors in parentheses.

Table 2

Markov-Switching Models for U.S.: Own 1-Lag Vacancy Models

						3 Regimes		2 Regimes	
Low State									
US ₋₁		0.943***		(0.015)		0.972***		(0.014)	
Constant		0.083***		(0.024)		0.049**		(0.023)	
Intermediate State									
US ₋₁		0.874***		(0.056)					
Constant		0.372**		(0.141)					
High State									
US ₋₁		0.136		(0.173)		0.243*		(0.134)	
Constant		1.520***		(0.301)		0.759***		(0.138)	
<hr/>									
p11		0.994		p12 0.000		p13 0.006		p11 0.857	
p21		0.025		p22 0.974		p23 0.001		p21 0.007	
p31		0.001		p32 0.058		p33 0.941			
<hr/>									
Observations		247				247			
Log-Likelihood		250.67				244.38			
BIC		-468.27				-466.72			

Notes: *, **, *** denote 10%, 5%, and 1% statistical significance, respectively. Standard errors in parentheses.

Table 3

Results for Northeast: 2 States with 2 Lags

	Low State	High State
South ₋₁ (V_j)	0.062 (0.099)	0.078 (0.116)
West ₋₁ (V_j)	-0.061 (0.083)	-0.163** (0.071)
Midwest ₋₁ (V_j)	0.182* (0.095)	0.138 (0.097)
Northeast ₋₁ (V_i)	0.530*** (0.095)	0.731*** (0.101)
HS ₋₁ (HS_j)	0.009 (0.020)	0.027 (0.026)
HM ₋₁ (HM_j)	-0.002 (0.081)	-0.001 (0.129)
South ₋₂ (V_j)	-0.088 (0.099)	-0.063 (0.106)
West ₋₂ (V_j)	-0.041 (0.075)	0.062 (0.070)
Midwest ₋₂ (V_j)	0.048 (0.098)	0.028 (0.103)
Northeast ₋₂ (V_i)	0.256*** (0.094)	0.105 (0.099)
HS ₋₂ (HS_j)	-0.012 (0.020)	-0.025 (0.026)
HM ₋₂ (HM_j)	-0.028 (0.082)	-0.024 (0.130)
constant	0.109 (0.078)	0.148 (0.144)
<hr/>		
p11	0.976	
p21	0.026	
<hr/>		
Observations	212	
Log-Likelihood	446.23	
BIC	-78.25	

Notes: *, **, *** denote 10%, 5%, and 1% statistical significance, respectively. Standard errors in parentheses.

Table 4

Results for Midwest: 2 States with 1 Lag

	Low State	High State
South ₋₁ (V_j)	0.153*** (0.046)	0.046 (0.175)
West ₋₁ (V_j)	-0.051 (0.035)	0.256* (0.133)
Midwest ₋₁ (V_i)	0.693*** (0.053)	0.675*** (0.135)
Northeast ₋₁ (V_j)	0.149*** (0.047)	-0.156 (0.157)
HS ₋₁ (HS_j)	0.052*** (0.013)	-0.101*** (0.035)
HM ₋₁ (HM_j)	0.071*** (0.015)	0.078* (0.044)
constant	0.070 (0.055)	0.197 (0.320)
<hr/>		
p11	0.914	
p21	0.017	
<hr/>		
Observations	213	
Log-Likelihood	379.32	
BIC	-329.73	

Notes: *, **, *** denote 10%, 5%, and 1% statistical significance, respectively. Standard errors in parentheses.

Table 5

Results for South: 2 States with 1 Lag

	Low State	High State
South ₋₁ (V_i)	0.811*** (0.044)	0.376** (0.161)
West ₋₁ (V_j)	0.037 (0.036)	-0.017 (0.101)
Midwest ₋₁ (V_j)	0.075* (0.044)	0.194 (0.144)
Northeast ₋₁ (V_j)	0.116*** (0.040)	0.034 (0.106)
HS ₋₁ (HS_j)	0.003 (0.011)	-0.007 (0.024)
HM ₋₁ (HM_j)	0.019 (0.014)	0.140*** (0.036)
constant	0.044 (0.062)	1.093*** (0.306)
<hr/>		
p11	0.995	
p21	0.030	
<hr/>		
Observations	213	
Log-Likelihood	376.24	
BIC	-323.58	

Notes: *, **, *** denote 10%, 5%, and 1% statistical significance, respectively. Standard errors in parentheses.

Table 6

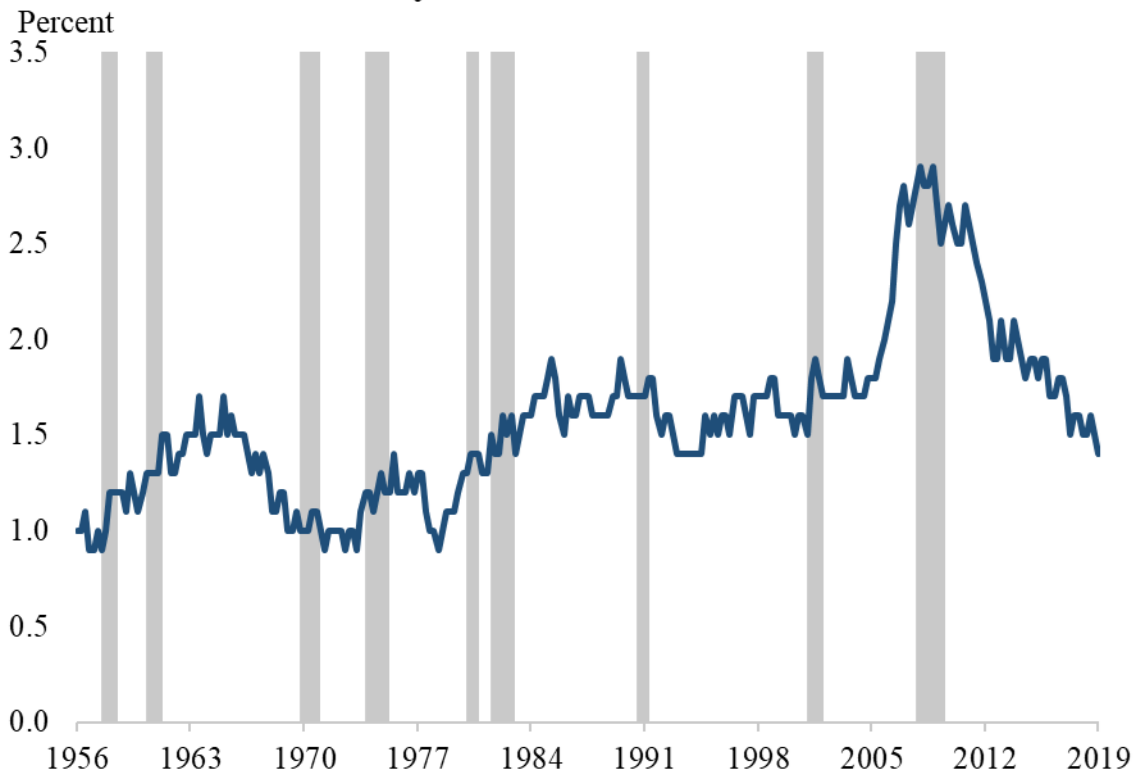
Results for West: 3 States with 1 Lag

	Low State	Moderate State	High State
South ₋₁ (V_j)	0.027 (0.086)	-0.179 (0.137)	0.198 (0.177)
West ₋₁ (V_i)	0.566*** (0.082)	0.344*** (0.115)	0.313** (0.136)
Midwest ₋₁ (V_j)	-0.132* (0.071)	0.363* (0.186)	0.575*** (0.198)
Northeast ₋₁ (V_j)	-0.146** (0.068)	0.189 (0.208)	-0.006 (0.137)
HS ₋₁ (HS_j)	0.004 (0.017)	0.041 (0.034)	-0.047* (0.027)
HM ₋₁ (HM_j)	0.014 (0.019)	0.022 (0.096)	0.022 (0.043)
constant	1.064*** (0.241)	0.564** (0.238)	-0.401 (0.339)
<hr/>			
	p11 0.987	p12 0.006	p13 0.007
	p21 0.025	p22 0.975	p23 0.000
	p31 0.035	p32 0.001	p33 0.964
<hr/>			
Observations	213		
Log-Likelihood	361.22		
BIC	-79.09		

Notes: *, **, *** denote 10%, 5%, and 1% statistical significance, respectively. Standard errors in parentheses.

Figure 1

U.S. Homeowner Vacancy Rate



Source: Census Bureau/Haver Analytics

Figure 2

The Housing Market and Vacancy Rates

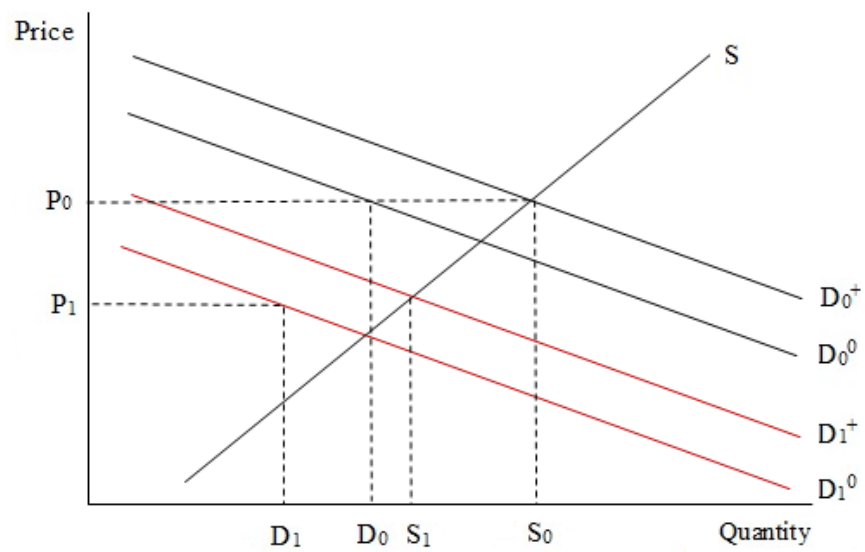
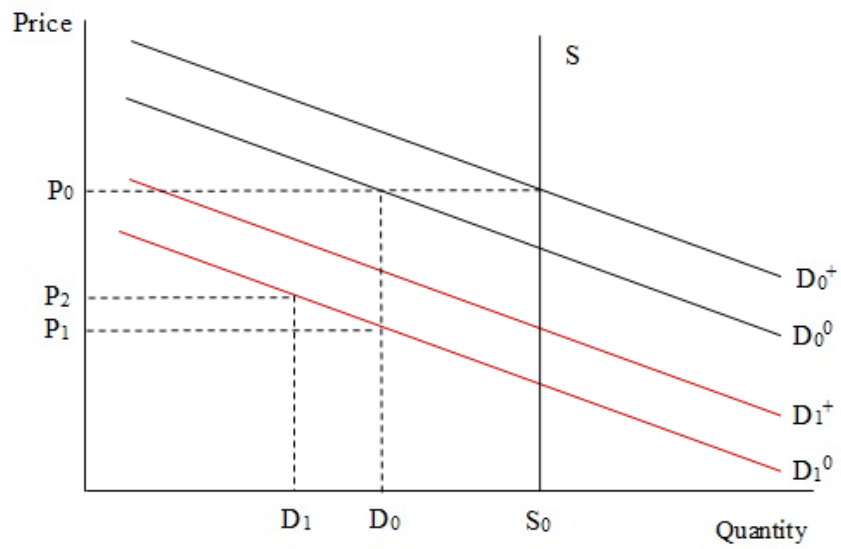
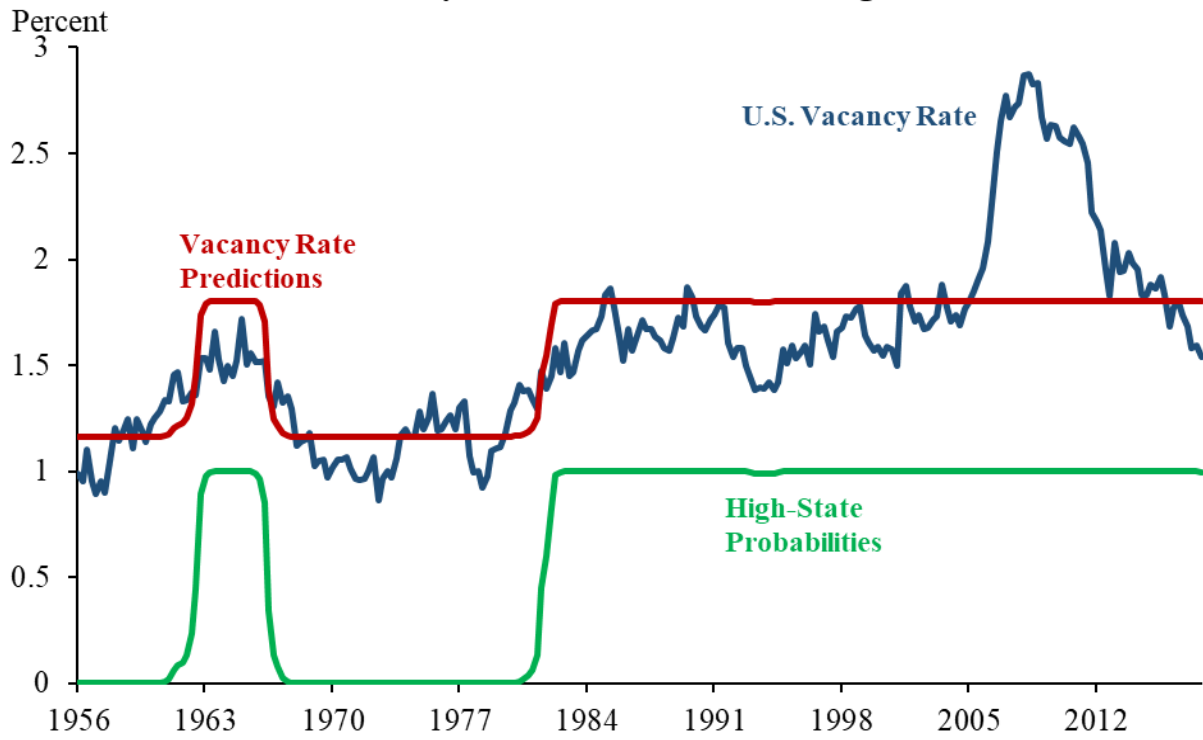


Figure 3

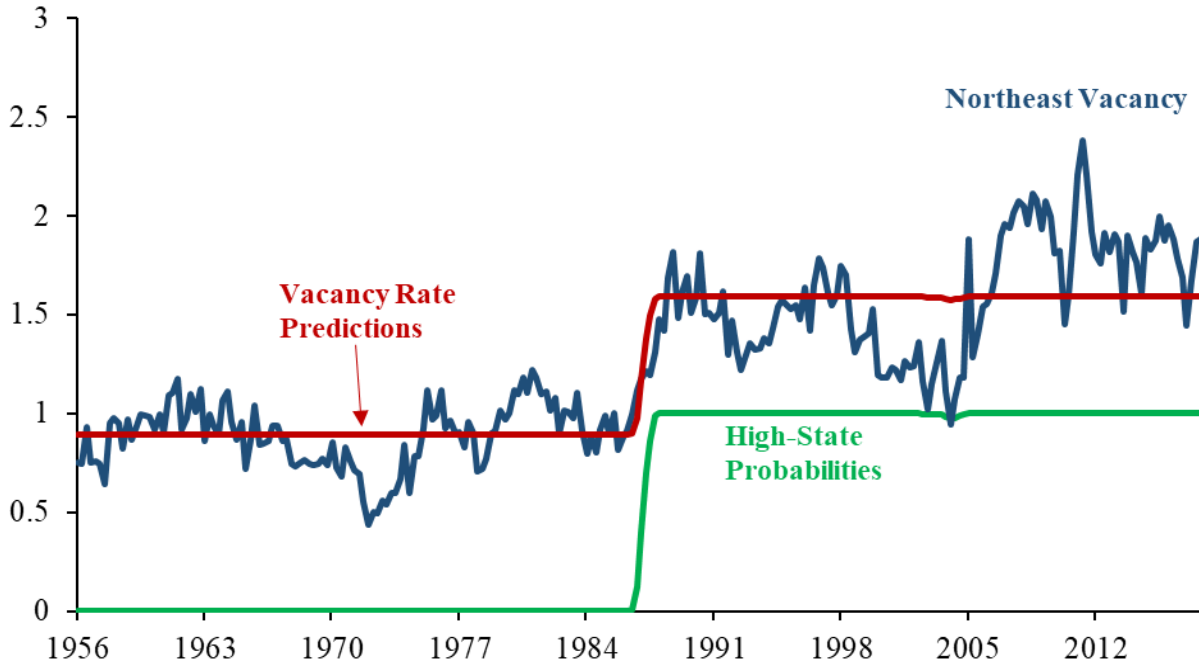
U.S.:Table 1-Based Vacancy Rate Predictions and High-State Probabilities



Sources: Census Bureau/Haver Analytics and Authors' Calculations

Figure 4

**Northeast:Table 1-Based Vacancy Rate Predictions
and High-State Probabilities**



Sources: Census Bureau/Haver Analytics and Authors' Calculations

Figure 5

**Midwest: Table 1-Based Vacancy Rate Predictions
and High-State Probabilities**

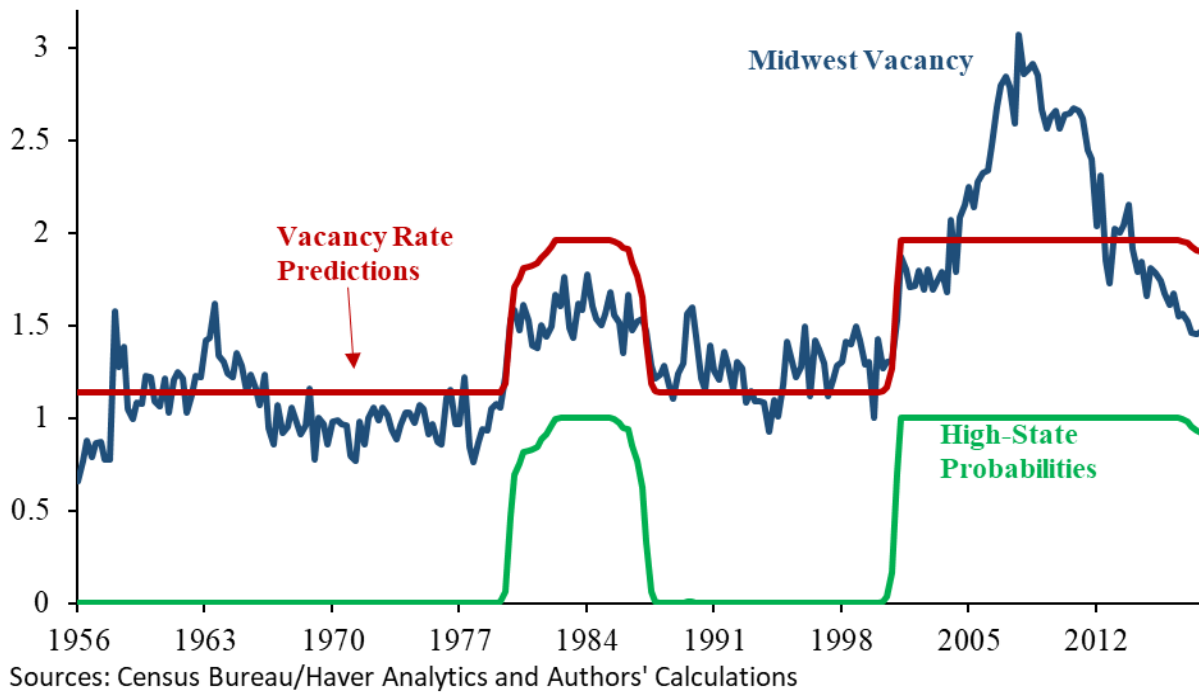


Figure 6

**South: Table 1-Based Vacancy Rate Predictions
and High-State Probabilities**

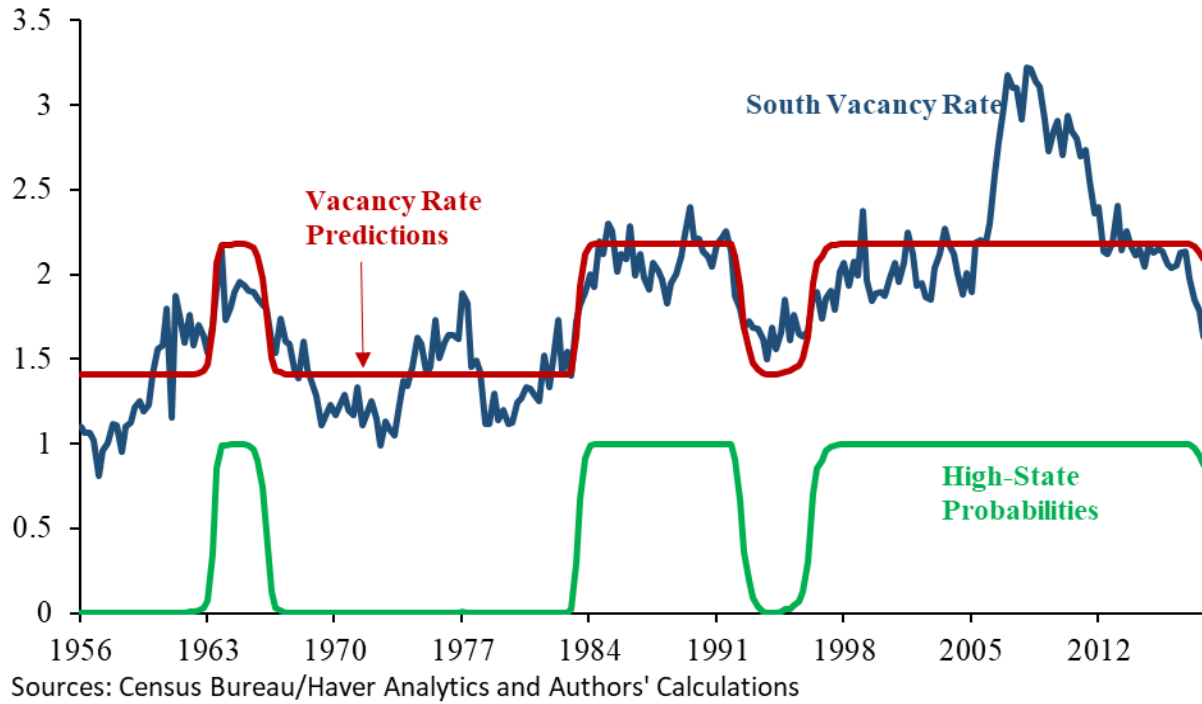
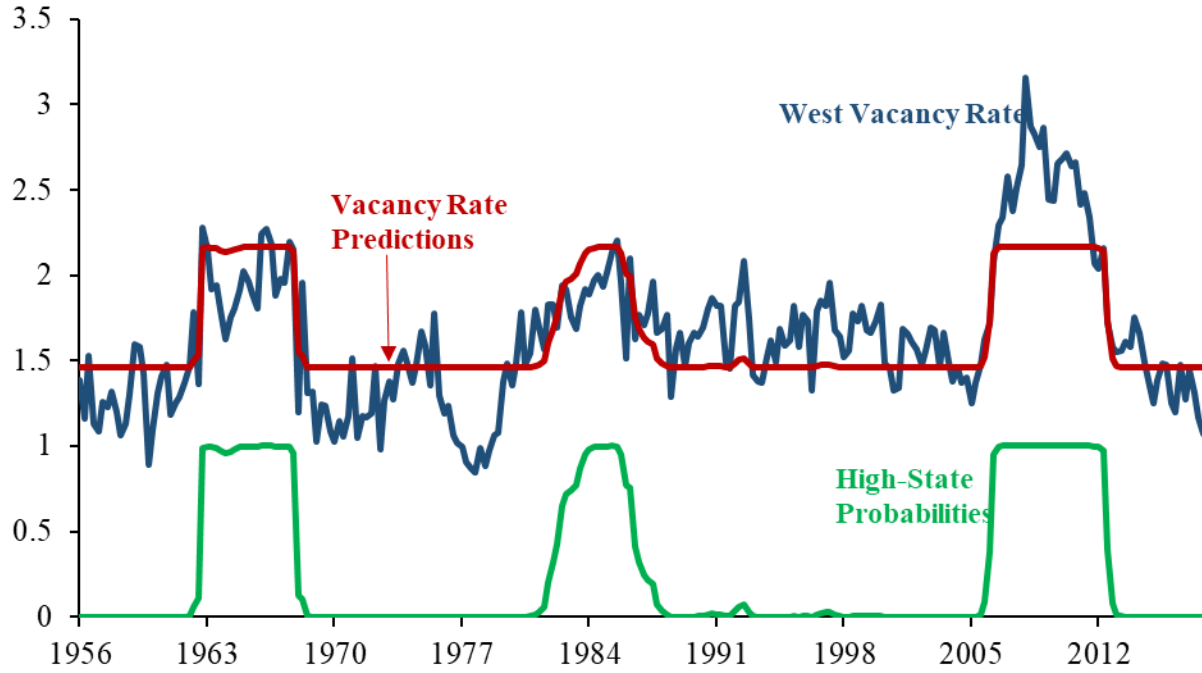


Figure 7

**West: Table 1-Based Vacancy Rate Predictions
and High-State Probabilities**



Sources: Census Bureau/Haver Analytics and Authors' Calculations

Figure 8

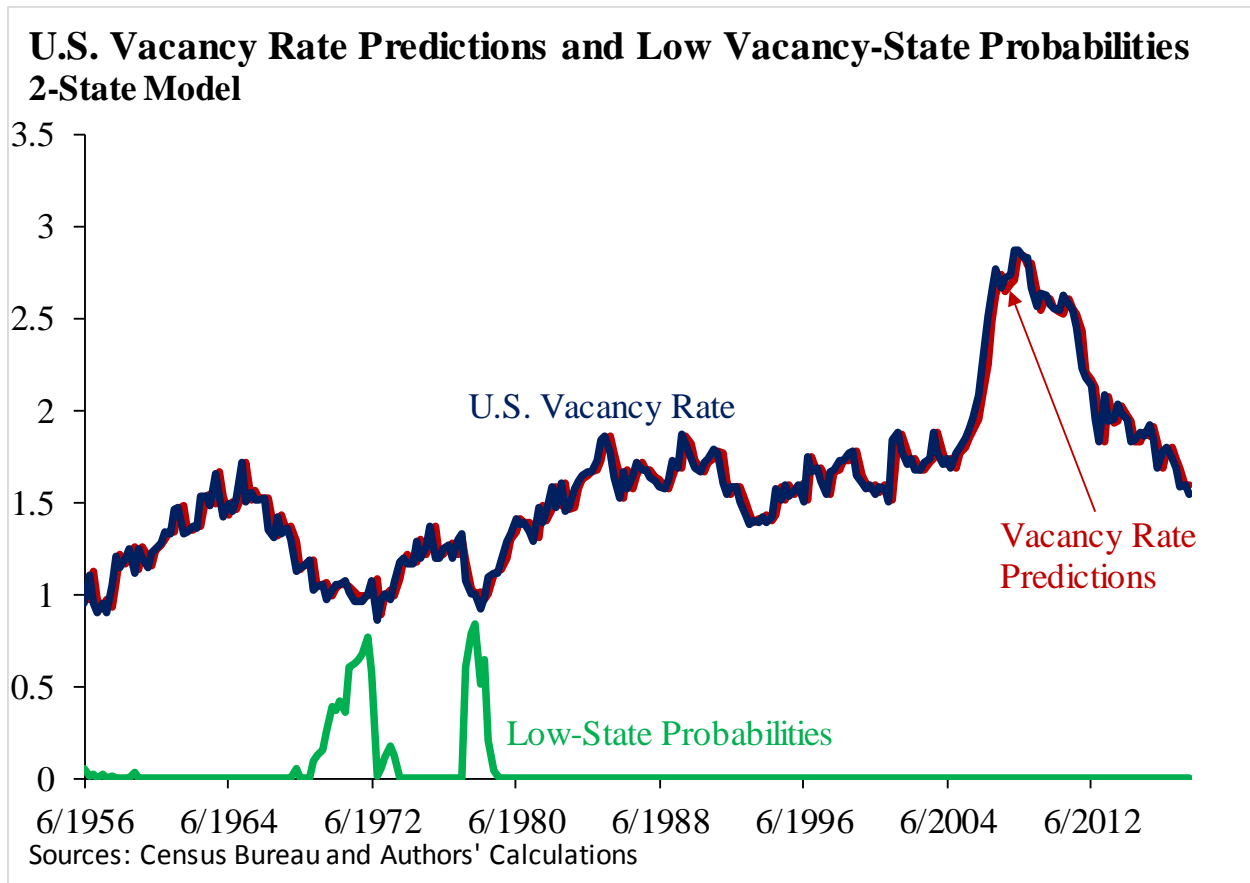


Figure 9

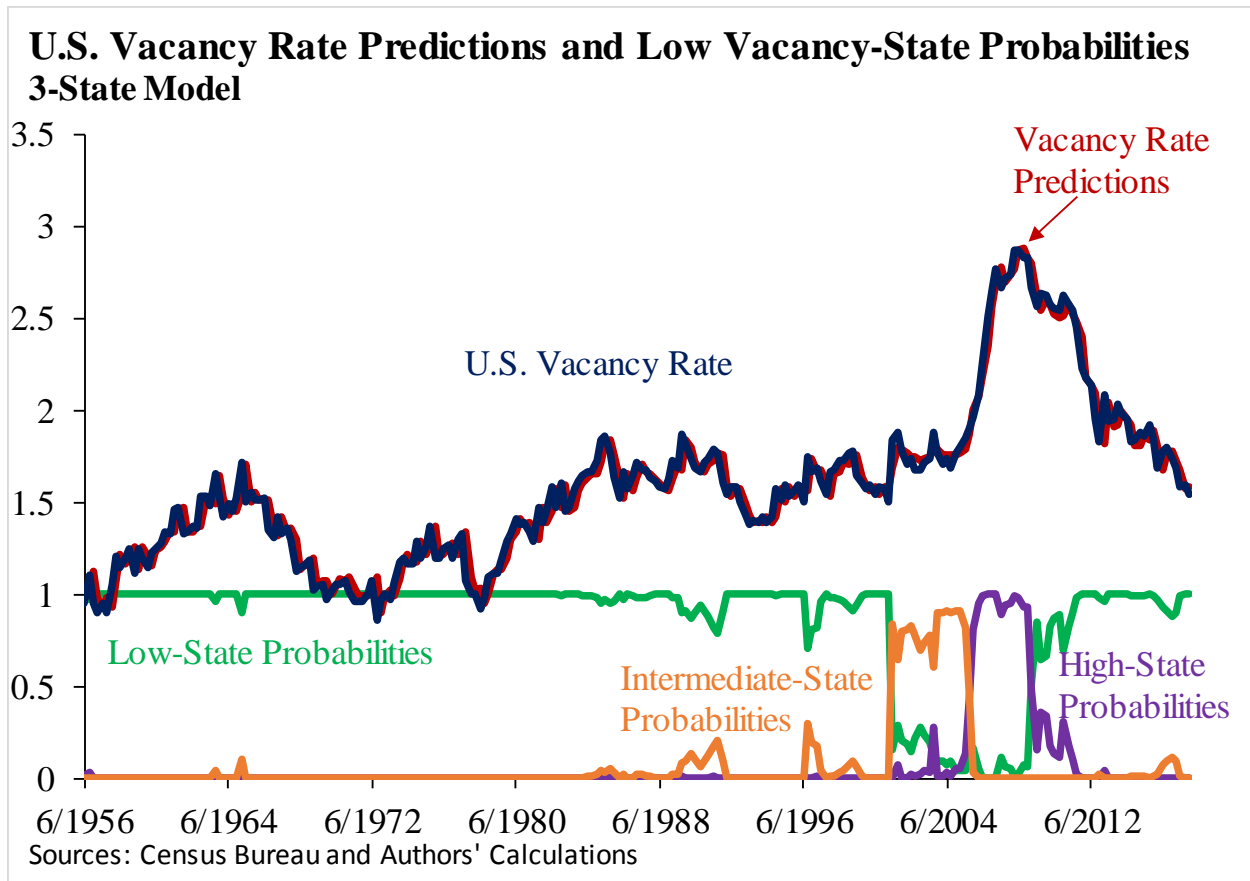


Figure 10

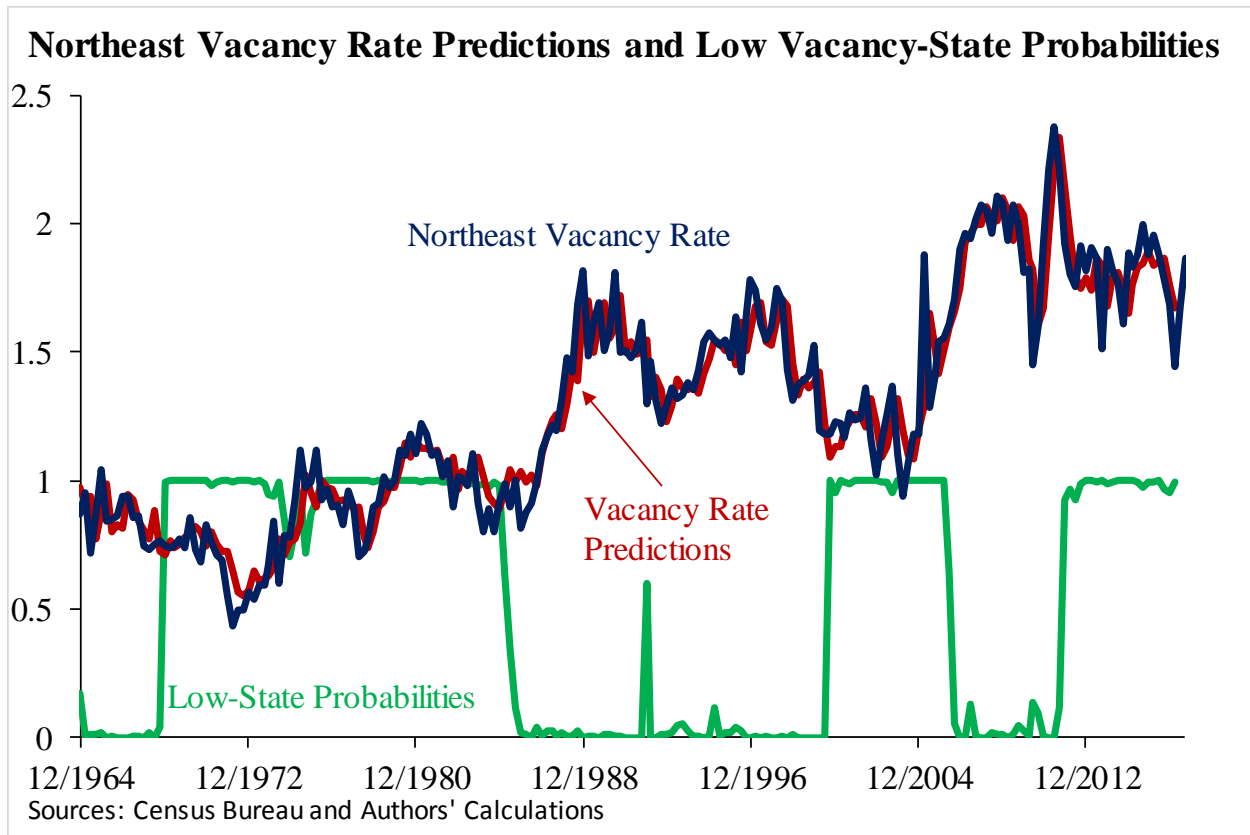


Figure 11

Midwest Vacancy Rate Predictions and Low Vacancy-State Probabilities

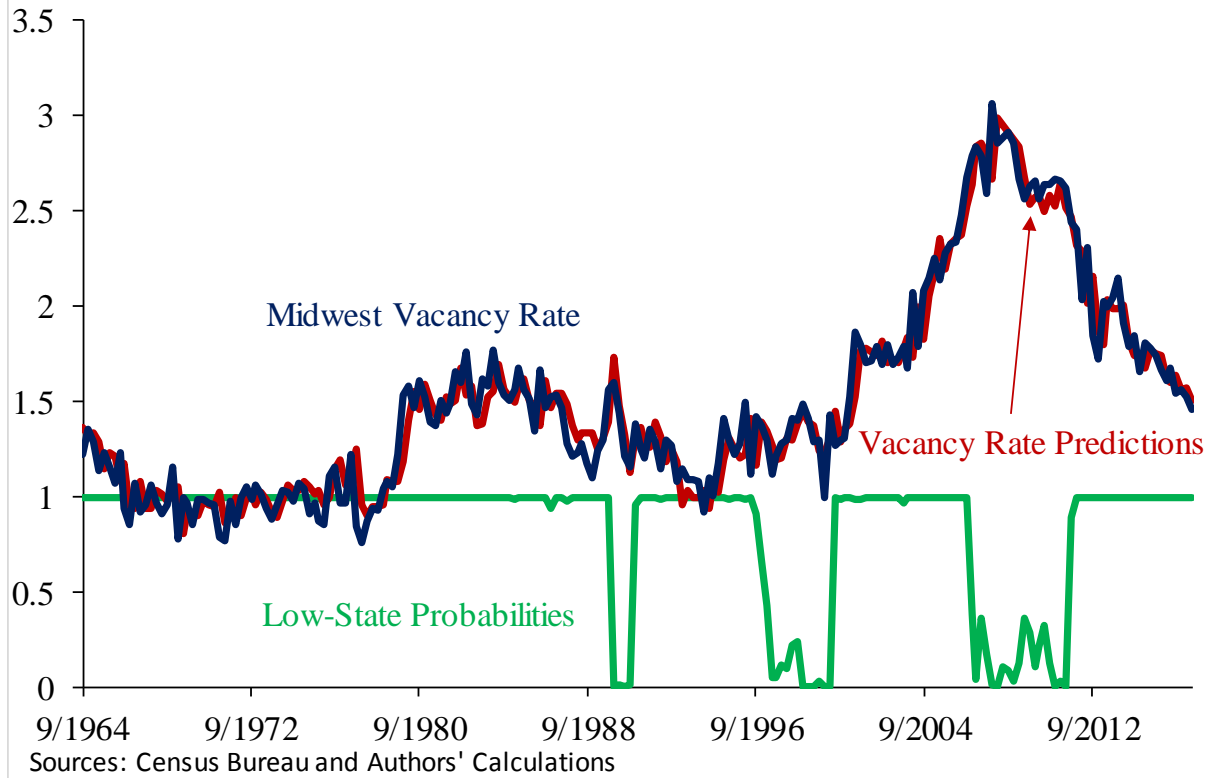


Figure 12

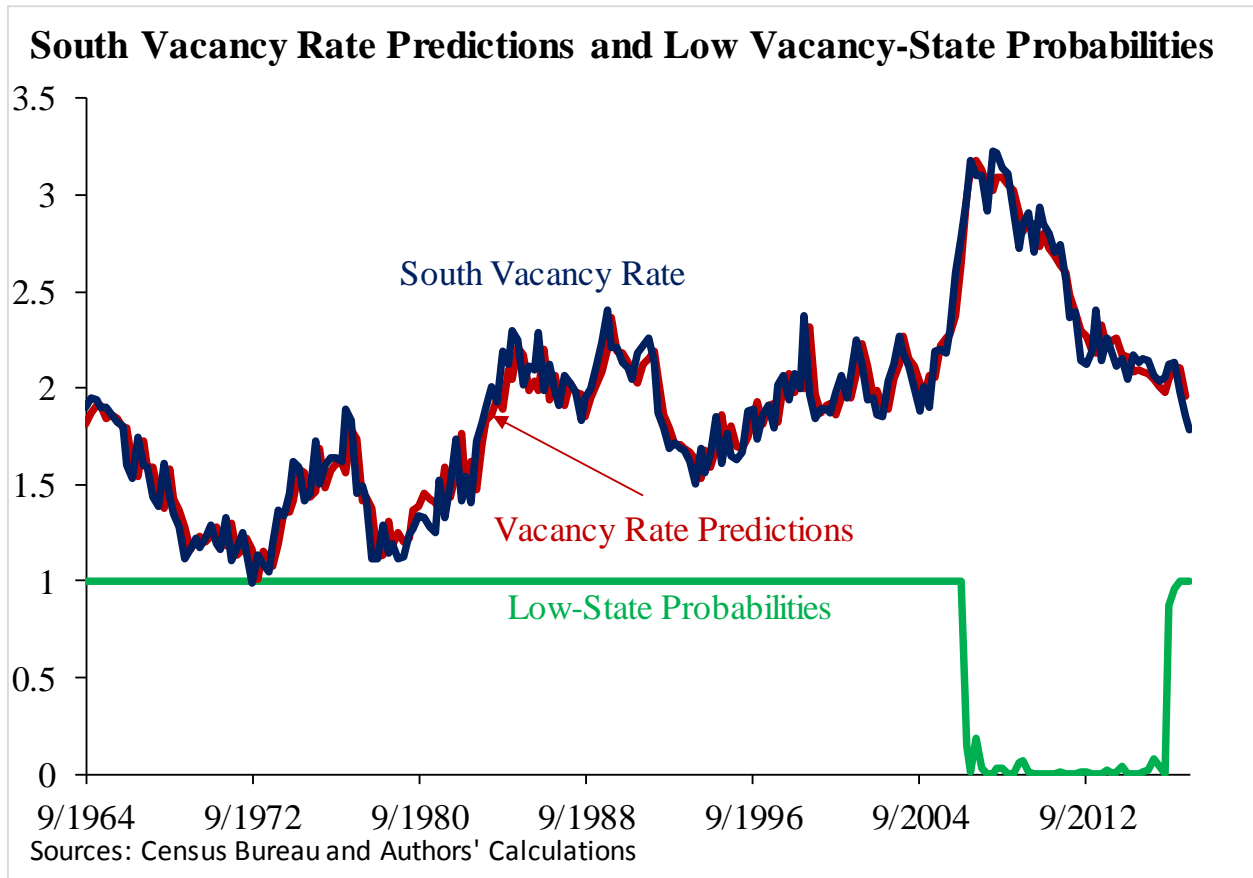
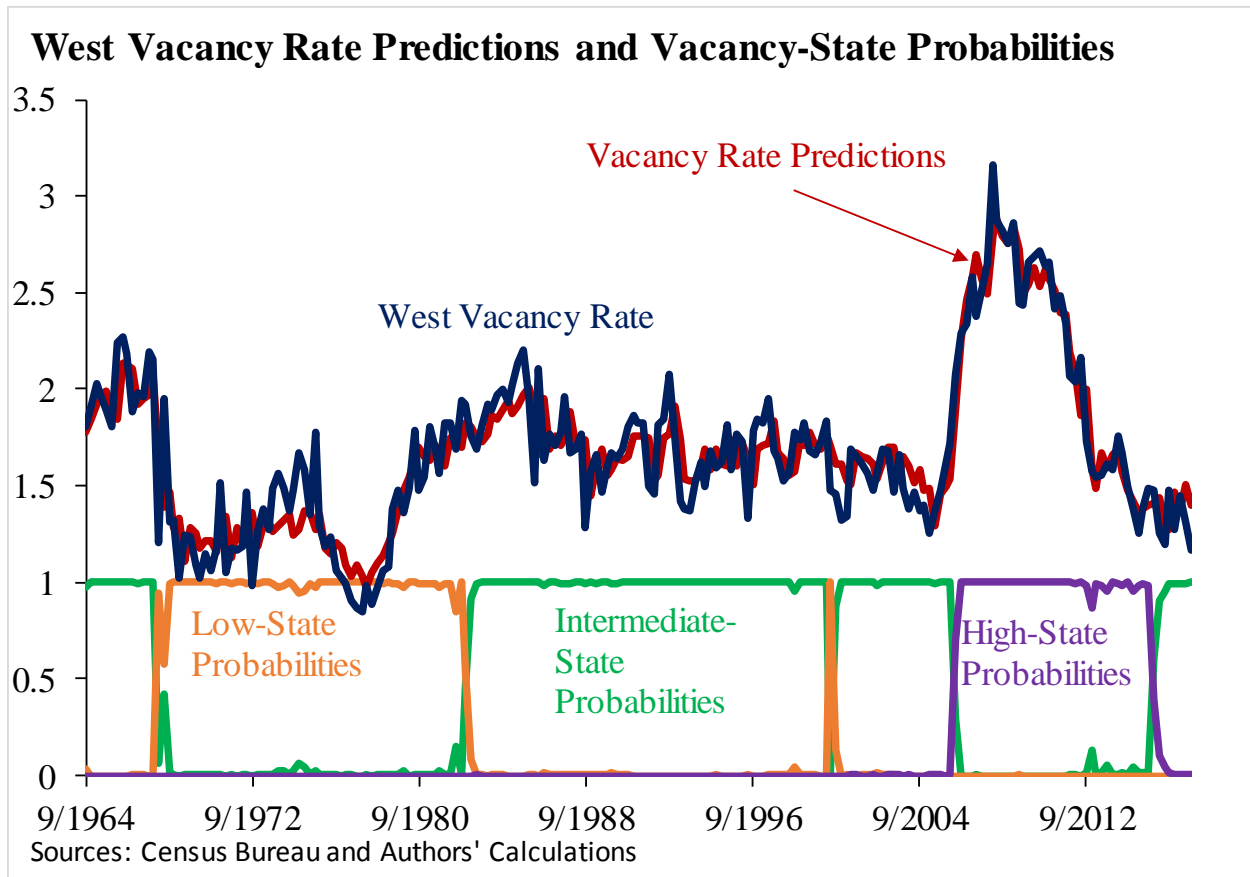


Figure 13



Appendix:

Discussion of Basic Markov-Switching Models

Some time series variables exhibit changes in behavior from one stretch of time to the next. For an example of such a variable, consider a city that has persistent periods of two types of easily distinguished weather - comfortable temperatures and heat waves. Daily construction activity, C_t , in a non-heat wave period at time t is described as $C_t = \beta_0 + \alpha_0' (Z_t)$, where β_0 and α_0 are respectively a constant and a coefficient column vector, both being specific to comfortable days. Z_t is a column vector of explanatory variables at time t . Meanwhile, construction activity during heat waves could be described as $C_t = \beta_1 + \alpha_1' (Z_t)$. It is reasonable to believe that, *ceteris paribus*, the former equation will predict much higher construction activity than the latter. Further, assuming we have all the data for independent variables and daily construction activity, estimation of these equations is relatively straightforward because we can distinguish whether past days were in a heat wave, and then estimate each equation using appropriate data.

Unfortunately, it is frequently difficult to distinguish between variables' different behaviors, which can be thought of as belonging to different states or regimes. Consider U.S. labor productivity growth, which likely goes through periods of low growth and high growth. We do not have any binary indicators that tell us when some shock has, for example, shifted labor productivity to a low-growth state. Such shocks are random and difficult to spot even after they have occurred. Markov-switching models, first discussed in Hamilton (1989) and further detailed in Hamilton (1994), attempt to deal with this problem of having a variable characterized by different states that have no simple way of being identified. That is, Markov-switching models attempt to identify unobservable states of the world, and describe how variables behave in each state. "Markov" refers to the fact that the models use Markov chains to characterize the unobserved states.

Markov-switching models have been used in a variety of applications. Billio et al. (2016) and Hamilton and Owyang (2012) use the models to compare business cycles across countries and states, respectively. Cermeño (2002) uses a simple Markov-switching model to characterize low-growth and high-growth regimes for per capita output of U.S. states and several countries. Ihle, Cramon-Taubadel, and Zorya (2009) use a Markov-switching model to characterize the transmission of maize prices between two African countries.

A general Markov-switching model, with dependent variable at time t , y_t , and column vector of explanatory variables, possibly including lags of y , at time t , X_t , is:

$$(7) \quad y_t = \beta_{s_t} + \alpha_{s_t}' (X_t) + \varepsilon_t,$$

where s_t , derived through estimation of a Markov chain, indicates the state of y at time t , and $\varepsilon_t \sim \text{i.i.d. } N(0, \sigma_{s_t}^2)$.¹⁷ Thus, the variance of the error term, the value of the constant, β , and the coefficient column vector, α , can depend on s_t . It is not required that all parameters depend on the regime, but at least one must or (13) would become a simple regression. Note that if X contains no variables, (13) is reduced to a constant-only model where the constant for each regime will simply be the regime's average value for y .

There is also an explicit autoregressive Markov-switching model, thoroughly described in Hamilton (1994):

(8) $y_t = \beta_{s_t} + \alpha'_{s_t}(X_t) + \gamma_{s_t}^1 (y_{t-1} - \beta_{s_{t-1}} - \alpha'_{s_{t-1}}(X_{t-1})) + \gamma_{s_t}^2 (y_{t-2} - \beta_{s_{t-2}} - \alpha'_{s_{t-2}}(X_{t-2})) + \dots + \varepsilon_t$,
where $\gamma_{s_t}^1, \gamma_{s_t}^2, \dots$ are the coefficients for the differences between the first, second, ... lag terms and their estimates less autoregressive terms, and all other parameters and variables are as described above. Similar to the other parameters, the γ 's can, but do not have to, depend on the regime.

¹⁷ The inclusion of autoregressive variables tends to produce smoother regime changes. Models without autoregressive variables tends to generate abrupt switches.