A Hierarchical Factor Analysis of US Housing Market Dynamics

Serena Ng ∗ Emanuel Moench†

February 10, 2010

Abstract

This paper studies the linkages between housing and consumption in the United States taking into account regional variation. We estimate national and regional housing factors from a comprehensive set of US price and quantity data available at mixed frequencies and over different time spans. Our housing factors pick up the common components in the data and are less affected by the idiosyncratic noise in individual series. This allows us to get more reliable estimates of the consumption effects of housing market shocks. We find that shocks at the national level have large cumulative effects on retail sales in all regions. While the effects of regional shocks are smaller, they are also significant. We analyze the driving forces of housing market activity by means of Factor-Augmented VAR’s. Our results show that lowering mortgage rates has a larger effect than a similar reduction of the federal funds rate. Moreover, lower consumer confidence and stock prices can slow the recovery in the housing market.

Keywords: hierarchical factor models, FAVAR, housing crisis.

JEL classification: C10, C20, C30, E2, E3.

∗Columbia University, 420 W. 118 St., MC 3308, New York, NY 10025, serena.ng@columbia.edu.
†Federal Reserve Bank of New York, 33 Liberty St., New York, NY 10045, emanuel.moench@ny.frb.org. We would like to thank Chris Otrok, Dan Copper, seminar participants at Columbia University and the Bundesbank for helpful comments and discussions. Evan LeFlore provided valuable research assistance on this project. The first author would like to acknowledge financial support from the National Science Foundation under grant SES 0549978. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System.
1 Introduction

This paper provides a quantitative assessment of the dynamic effects of housing shocks on retail sales. The econometric exercise consists of estimating (national and regional) housing factors from large non-balanced panels of data. The economic analysis consists of studying the dynamic response of (national as well as regional) retail sales to housing market shocks, and assessing the sensitivity of the housing factor to economic conditions and stimulus. As a by-product, we obtain estimates of ‘house price factors’ that summarize the common information in the different published house price indicators.

Our approach has three features. First, we make a distinction between ‘national’ and ‘regional’ housing markets in recognition of the fact that not all variations in housing are pervasive. Second, we consider shocks to the housing ‘market’ as opposed to shocks to home ‘prices’ only. The analysis thus makes extensive use of housing data rather than relying on a single measure of house prices. Third, we use diverse measures of price and volume to determine the (latent) state of the housing market. The non-balanced panel of data covers series sampled at different frequencies and that are available over different time spans.

Our analysis consists of two steps. We first use a dynamic hierarchical (multi-level) factor model to disentangle information on the housing market into national, regional, and series-specific components. For each region, we embed the estimated national and regional housing factors along with other variables that control for the effects of regional business cycles into factor augmented vector-autoregressions (FAVAR). An analysis of the impulse responses then allows us to study the propagation mechanism of regional and national housing shocks.

Several considerations motivate our analysis. First, over the last few years, the US housing market has experienced an extended period of expansion followed by an abrupt and pronounced downturn. Newspaper articles and the media often suggest that a housing boom stimulates, while a housing bust slows non-housing activity, in particular consumption. For example, reporting on a speech made by Federal Reserve Chairman Greenspan, the Los Angeles Times (March 09, 1999) wrote ‘Capital generated by a booming housing market have probably spurred consumer spending and given a strong boost to the U.S. Economy’. In its October 16, 2006 issue, Newsweek magazine wrote that ‘if home prices drop too much, the damage to consumer confidence and spending won’t be easily offset.’

Although two-thirds of U.S. households are homeowners, evidence on the macroeconomic effects of housing shocks have been limited to point estimates from micro data for certain demographic groups or short panels, neither of which are ideal for studying the dynamic response to housing shocks at the aggregate level. There are few VARs estimated for con-
sumption and housing at the national level and even fewer at the regional level because it is
difficult to find housing and consumption data that are available for a long enough period to
make estimation of a VAR suitable. We circumvent this problem by not restricting ourselves
to house price data alone. We also do not restrict ourselves to data sampled at the same
frequency. Instead, we extract common housing factors at the national and regional level
from data on house prices and housing market quantities. We then use FAVARs to trace out
the aggregate effects of housing without fully specifying the structure of the housing market.

Second, the notion of a 'US housing market' disregards the fact that there is substantial
variation in housing activity across markets, a point also raised by Calomiris et al. (2008).
Indeed, of the four regions defined by the Census Bureau, the Northeast (including New
York and Massachusetts) and the West (including California, Arizona, and Nevada) have
historically had more active housing markets than the South (including Texas, Florida and
Virginia) and the Midwest regions (including Illinois, Ohio and Michigan). Consumers in
regions not used to large variations in the housing market might respond differently from
those that are accustomed to housing cycles. Consumption responses might also depend on
regional business cycle conditions. It is very much an empirical matter whether regional
differences in the consumption response to housing shocks exist.

Third, there exist numerous measures of house prices, including the well known Case-
Shiller price index, the indices published by the Federal Housing Finance Administration
(FHFA, formerly Office of Federal Housing Enterprise Oversight or OFHEO), and indices
published by the National Association of Realtors (NAR). Some series are available for time
spans longer than thirty years, while others are available for a little over a decade. Calomiris
et al. (2008) note that many issues surrounding empirical estimates of the wealth effect of
housing relate to the definition of the house price. We address this problem by estimating
a house price factor that extracts the common variations underlying all indicators of house
prices thereby filtering out idiosyncratic noise.

While price is a key indicator of the state of the housing market, data on the volume of
transactions are also available. Leamer (2007) argues that the 'volume cycle' rather than the
'price cycle' is what makes housing important in US business cycles. Some regional markets
may be more prone to high volatility in housing prices and to high volume of transactions
than others. In order to assess how the activities in the housing 'market' (as opposed to
house prices) affect consumption, we construct broad based measures of the state of the
housing markets at both the regional and the national level. In our analysis, this is handled
using a hierarchical factor model framework.
Our results can be summarized as follows. First, we find that the national and regional factors are of comparable order of importance in three of the four regions, but regional variations are much more important in the West. Second, there is a marked difference between the house price and the housing market factor in the Midwest and the South. Third, retail consumption in all regions responds positively to a national housing shock with the peak occurring about fifteen months after the shock. A two standard deviation shock in the housing factor can reduce aggregate consumption by as much as 8%, all else equal. However, the consumption responses are largely driven by responses to house price shocks. Shocks to volume have significantly smaller effects. A FAVAR in variables that might affect the housing factor indicates that interest rate cuts will stimulate housing activity, with reductions in mortgage rates potentially having a larger impact than similar cuts in the fed funds rate. Moreover, consumer confidence and stock prices both have a significant effect on housing market activity.

2 Data

We assemble a dataset that covers the four census regions Northeast, Midwest, West, and South. We also have aggregate measures of housing activity for the United States, which we will refer to as ‘national’ data. We combine information from various sources, including federal agencies and private institutions. Instead of focusing on house price indicators alone, our dataset consists of both price and volume information. This allows us to capture the different dimensions of the housing market. The data are further sampled at different frequencies: some series are monthly, and some are quarterly. Moreover, some series start as early as 1963, while some are not available until 1990. We thus face an unbalanced panel. We take January 1973 to be the beginning of our sample, and the last data entry is May 2008. The key series are listed in Tables 1 and 2. To motivate the analysis to follow, it helps to have a quick review of the data used.

As house price is the key indicator of housing wealth, most studies quite naturally use some measure of house prices to study the effect of housing market shocks on real economic activity. However, there exist several indicators of house prices published by various data providers, each employing different data sources and aggregation methods. While these house price series are correlated, each has some distinctive features. Ideally, one would therefore like to use a genuine measure of price movement. In this paper, this is taken to be variations that are common to all observed price measures.

In general terms, the literature distinguishes between three main ways to measure house
prices. As discussed in Rappaport (2007), the simplest method computes the average or median of house prices observed in a period. This ‘simple approach’ can be volatile due to changing composition of high and low priced units. The ‘repeat sales method’ focuses on houses that have been sold more than once. It is a price index and does not measure the price level itself. Furthermore, the number of repeat transactions can be small relative to total transactions, and it is subject to continual revisions. The third is the ‘hedonic approach’ which uses statistical methods to control for differences in quality. As a general matter, price measures based on repeat transactions are often thought to give more precise estimates of house price appreciation. Prices subject to compositional effects are believed to be better at measuring the amount required to purchase housing than at estimating the rate of house price changes.

The Federal Housing Finance Administration (FHFA) indices include only homes with mortgages that conform to Freddie Mac and Fannie Mae guidelines. Data are available at the national, regional, and state level, as well as for the major metropolitan areas. They are based on transactions and appraisals, and are then adjusted for appraisal bias. The FHFA also publishes a purchase-only index that excludes refinancing. These indices equally weight prices regardless of the value of the house. The coverage of the indices is broad because Freddie Mac and Fannie Mae provide loans throughout the country. However, the so-called jumbo loans over $417,000 are not included.

The S&P/Case-Shiller home price indices, published by Fiserv Inc., are based on information from county assessor and recorder offices. The index started with data from 10 cities in 1987 but was extended to cover 20 cities in 2000. The Case-Shiller indices do not use data from thirteen states and have incomplete coverage for 29 states. Compared to the FHFA, the Case-Shiller indices thus have a narrower geographical coverage. However, homes purchased with subprime and other unconventional loans are included in the indices. As they cover defaults, foreclosures, and forced sales, these indices show more volatility than the FHFA indices. Note also that the Case-Shiller indices are value weighted and hence give more weight to higher priced homes.

The FHFA and the Case-Shiller indices are both based on repeat sales. In contrast, the National Association of Realtors (NAR) report the mean/median purchase prices of homes directly. The NAR represents real estate professionals and has close to 2000 local associations and boards offering multiple listing services. The NAR surveys a fixed subset of its associations. Based on reported transactions from the sample, the NAR calculates a median price for each of the four Census Bureau regions. The national price is then taken
as a weighted average of the regional medians. The NAR price indices can be volatile due
to compositional changes. An increase in the difference between high priced relative to low
priced units will increase the regional and hence the national median. The NAR indices are,
however, available for each region on a monthly basis over a long time period.

The Bureau of Census publishes several house price series. A monthly national series
is available since 1963, but the regional data are available only quarterly. The Census also
provides an average price of new homes of constant quality from 1977 onwards on a quarterly
basis, both for the U.S. and for the four regions. The indices are based on a monthly survey
of residential construction activity for single-family homes. These indices are also subject to
compositional effects that might arise from the sales sample rather than any true changes
in price. The Census Bureau also publishes an index of one family homes sold based on the
hedonic approach.

The Conventional Mortgage Home Price Index (CMHPI) is provided by Freddie Mac. It
is calculated on a quarterly basis at both the national and regional level from 1975 onwards.
The index is based on conventional conforming mortgages for single unit residential houses
that were purchased or securitized by Freddie Mac or Fannie Mae. The CMHPI overlaps
with the FHFA series to some extent. We are primarily interested in the common variations
that underlie these series.

In addition to prices, volume data on transactions and turnovers are also informative
about the level of housing activity. Dieleman et al. (2000) found that demographic changes
are largely responsible for turnovers in the housing market, three-quarters of which are
generated by renters. In contrast, house prices are mainly affected by household income. In
a frictionless world, prices adjust and sales occur instantly after a shock. But frictions in
the housing market might prevent house prices from adjusting, which slows turnover. Stein
(1995) observed a positive contemporaneous correlation between changes in house prices and
sales and that there is more intense trading activity in rising markets than in falling markets.
He suggests that down payment and other borrowing constraints might be responsible for
market frictions. Case and Shiller (1989) argue that the rational response in a falling market
is for a homeowner to hold on to his/her investment in anticipation for higher future returns.
Berkovec and Goodman (1996) suggest that transactions might act as a forward indicator
of price changes. As Stein (1995) suggests, if an initial shock knocks prices down, the loss
on existing homes could undermine the ability of would-be movers to make down payments
on new homes. This lack of demand could further depress prices.

The transactions data used in our factor analysis include new and existing single family
homes under construction, sold, and for sale, as well as data on employment in the construc-
tion sector. Additionally, the Census bureau publishes data on homeowner vacancy rates, home-ownership rates, as well as the rental vacancy rate. These latter indicators are informative about the tightness of the prevailing housing markets.

It is also useful to discuss data that are available but are not used in our factor analysis. The BLS publishes data on housing starts, permits as well as rent. Housing starts and permits are informative about the future as opposed to the present market conditions. We do not use these variables in the factor analysis as our model only allows for variables to load on contemporaneous and lagged factor observations. The rent data are based on the Consumer Expenditure Survey of which two-thirds of the sample are homeowners. To the extent that rent captures the capitalized value of housing, it provides a measure of the fundamental instead of the market value of houses. During periods of speculative housing booms, rents and house prices can diverge quite substantially. However, rent is regulated in many areas. We also have data on prices of mobile homes. These data are not used in the base case, which focuses on single-unit and multi-family housing.

Finally, we note that all price variables are deflated by the (all items) CPI to control for increases in the overall price level. The NAR data are not seasonally adjusted. We use the X11 seasonal filter in Eviews to adjust these series. In all cases, we first annualize the data by taking year to year differences of the log level of the series. This means that for monthly data, we take the log difference of a series over a twelve month period. For a quarterly series, we take the log difference over four quarters. Since many series remain non-stationary, we transform the series into annual differences of the annual growth rates before estimating the factors. The effective sample is thus January 1975 to May 2008.

3 Econometric Methodology

Our econometric framework is set up with three issues in mind. First, shocks to the housing sector need not be the same as shocks to house prices. Second, there are substantial variations in housing market conditions across regions. Third, we have non-balanced panels of data.

To deal with the aforementioned issues, we use an extension of the ”Dynamic Hierarchical Factor Model” framework developed in Moench et al. (2009). In such a model, variations in an economic time series can be idiosyncratic, common to the series within a block, or common across blocks. We treat a block (identified as $b$) as one of the four major geographical regions, the NorthEast (NE), MidWest (M), South (S), and West (W). More precisely, we posit that

\footnote{The paper is available at \url{http://www.columbia.edu/~sn2294/papers/dhfm.pdf}.}
for each \( b = \text{NE}, \text{M}, \text{S}, \text{and} \ W \), we observe \( N_b \) housing indicators which have zero mean and unit variance. The data, stacked in the vector \( X_{bt} \), have a factor representation given by

\[
X_{bt} = \Lambda_{Gb}(L)G_{bt} + e_{Xbt}
\]

(1)

where \( \Lambda_{Gb}(L) \) is a \( N_b \times k_{Gb} \) matrix polynomial in \( L \) of order \( s_{Gb} \). According to the model, the housing indicators in a block are driven by a set of \( k_{Gb} \) regional factors denoted \( G_{bt} \), and idiosyncratic components, \( e_{Xbt} \). Stacking up \( G_{bt} \) across regions yields the \( K_G \times 1 \) vector \( G_t = (G_1t \ G_2t \ \ldots \ G_Bt)' \). Observed indicators of the national housing market are stacked into a \( K_Y \times 1 \) vector \( Y_t \). At the national level, we assume that

\[
\begin{pmatrix}
G_t \\
Y_t
\end{pmatrix} = \Lambda_F(L)F_t + \begin{pmatrix}
e_{Gt} \\
e_{Yt}
\end{pmatrix}
\]

(2)

where \( K_F \) factors, collected into the vector \( F_t \), capture the comovement common to all regional factors, and where \( \Lambda_F(L) \) is a \( (K_G + K_Y) \times K_F \) matrix polynomial of order \( s_F \).

Dynamics are introduced into the model by letting

\[
F_t = \Psi_F F_{t-1} + \epsilon_{Ft}
\]

(3)

\[
e_{Gb_{jt}} = \Psi_{Gb} e_{Gb_{jt-1}} + \epsilon_{Gb_{jt}}
\]

(4)

\[
e_{X_{bit}} = \Psi_{X_{bi}} e_{X_{bit-1}} + \epsilon_{X_{bit}}
\]

(5)

where \( \Psi_F \) is a diagonal \( K_F \times K_F \) matrix, \( \Psi_{Gb} \) is a diagonal \( k_{Gb} \times k_{Gb} \) matrix, and \( \Psi_{X_{bi}} \) is a diagonal \( N_b \times N_b \) matrix. Furthermore,

\[
\epsilon_{F_{kt}} \sim N(0, \sigma_{F_{kt}}^2), \quad k = 1, \ldots K_F
\]

\[
\epsilon_{Gb_{jt}} \sim N(0, \sigma_{Gb_{jt}}^2), \quad j = 1, \ldots k_{Gb}
\]

\[
\epsilon_{X_{bit}} \sim N(0, \sigma_{X_{bit}}^2), \quad i = 1, \ldots N_b.
\]

Equations (1) and (2) constitute what is known as a three level factor model:- the level one variations are due to \( \epsilon_{X_{bit}} \), the level two variations are due to \( \epsilon_{Gb_{jt}} \), and the level three variations are due to \( \epsilon_{F_{kt}} \). In our empirical application of this model, we will set \( K_F = k_{Gb} = 1 \quad \forall \quad b \). To identify the sign of the factors and loadings separately, we set the upper-left elements of \( \Lambda_F(0) \), and for \( b = 1, \ldots B \), we also set \( \Lambda_{Gb}(0) \) equal to one.\(^2\)

\(^2\)In the more general case with multiple factors at both levels, the \( k_{Gb} \) regional factors could, for example, be identified by requiring that for each \( b \), \( \Lambda_{Gb}(0) \) is a lower triangular matrix with diagonal elements of unity. Similarly, the \( K_F \) factors could be identified by requiring that \( \Lambda_F(0) \) is lower triangular, again with ones on the diagonal. In such a case of multiple factors, the ordering of the variables might potentially have an impact on the factor estimates.
Our model can be seen as having two sub-models, each with a state space representation. Specifically, if $G_{bt}$ was observed, (2) and (3) is a standard dynamic linear model where the latent factor is $F_t$. Then (1), (4) and (5) constitute the second dynamic linear model where the latent vector is $G_t$. It is in principle possible to use variables that are only available at the national level to estimate $F_t$. However, to the extent that $G_{bt}$ are correlated across regions, they also convey information about $F_t$ which we exploit in the estimation.

The three-level model implies that

$$X_{bt} = \Lambda_{Gb}(L)\Lambda_{Fb}(L)F_t + \Lambda_{Gb}(L)e_{Gbt} + e_{Xbt},$$

where $\Lambda_{Fb}$ is the sub-block of $\Lambda_F$ corresponding to block $b$. This is in contrast to a two-level factor model that consists of only a common and an idiosyncratic component. Omitting variations at the block level amounts to lumping $\Lambda_{Gb}(L)e_{Gbt}$ with $e_{Xbt}$ in the estimation of $F$. This can result in an imprecise estimation of the common factor space if variations in $e_{Gbt}$ are large. Moreover, explicitly specifying the block structure facilitates interpretation of the shocks.

In our setup, the regional factors contain information about the state of the housing sector at the national level. A different formulation of regional effects is a model specified as

$$X_{bit} = b_iF_t + c_{bi}e_{Gbt} + e_{bit}.$$ 

Fu (2007) uses such a model to decompose house prices in 62 U.S. metropolitan areas into national, regional, and metro-specific idiosyncratic factors using quarterly FHFA price data from 1980-2005. A similar model was also used by Del Negro and Otrok (2005) to estimate the common component of quarterly FHFA price data from 1986 to 2005. Stock and Watson (2008) use a variant of this model to analyze national and regional factors in housing construction. Kose et al. (2008) use it to study international business cycle comovements. Our model is more restrictive in that the responses of shocks to $F_t$ for all variables in block $b$ can only differ to the extent that their exposure to the block-level factors differs. However, the additional structure we impose makes the model more parsimonious, and it is easy to accommodate observed aggregate indices $Y_t$ in estimating $F_t$.

Numerous methods are available to estimate two level factor models. For models with a few number of series, maximum likelihood is widely used. For large dimensional factor models, the method of principal components is popular. The factor model used in the present study and introduced in Moench et al. (2009) is a multi-level extension of the simple two level factor model considered in various previous contributions. We use Markov Chain
Monte Carlo (MCMC) methods (specifically a Gibbs sampling algorithm) to estimate the posterior distribution of the parameters of interest and the latent factors. Unlike the method of principal components, estimation via Gibbs sampling requires parametric specification of the innovation processes. However, one practical advantage of MCMC is that credible regions can be conveniently computed. In contrast, there exists no inferential theory for multi-level models estimated by the method of principal components.

The problem considered here is non-standard because not every series on the housing market is available on a monthly basis. Aruoba et al. (2008) also consider estimation of latent factors when the data are sampled at mixed frequencies. However, we have the additional problem that not all our data series are available over the same time span. For example, house price data are available from NAR since 1970, from the FHFA at the regional level since 1975, and the Case-Shiller index is available since 1987.

The first problem that not all data are available on a monthly basis is easily handled in the Bayesian framework using data augmentation techniques. Suppose for now that we have data over the entire sample for $X_{bt}$, but it is only available on a quarterly instead of a monthly basis. The monthly value $X_{bt}$ when $t$ does not correspond to a month during which new data are released has conditional mean

$$X_{bt|t-1} = \Lambda_{Gb}(L)G_{bt} + e_{Xbt|t-1}$$

where $e_{Xbt|t-1} = \psi_{Xb,1}e_{Xbt-1}$. A monthly observation of $X_{bt}$ with conditional mean $X_{bt|t-1}$ and variance $\sigma^2_{Xb}$ is obtained by taking a draw from the normal distribution with this property. Similarly, if the $m$-th aggregate indicator is available on a quarterly basis, we make use of the fact that

$$Y_{t,m} = \Lambda_{Fm}'(L)F_t + e_{Yt,m}.$$ 

Conditional on $F_t$, $\Lambda_{Fm}$ and $\Psi_{Y,m}$, the monthly value of $Y_{mt}$ when data are not observed at time $t$ can be drawn from the normal distribution with mean $\Lambda_{Fm}'(L)F_t + e_{Yt,m|t-1}$ and variance $\sigma^2_{Y,m}$, where $e_{Yt,m}$ is an autoregressive process with parameters $\Psi_{Y,m}$.

As for the second problem that some data are missing for the early part of the sample, assume for the sake of discussion that the data (when they are available) come on a monthly basis. For the sub-sample over which data are not available, we fill the data with the value ‘NaN’. In a state space framework, these values contain no new information and contribute zero to the Kalman gain. To implement this, let $X_{bt}^+$ be the subset of $X_{bt}$ for which data are available at time $t$. If $X_{bt}$ is $N_b \times 1$, and $X_{bt}^+$ is $N_b^+ \times 1$, we work with the measurement equation:

$$X_{bt}^+ = \Lambda_{Gb}^+(L)G_{bt} + e_{Xbt}^+$$
where \( \text{var}(e_{X_{bt}}) \) is a \( N_{bt}^+ \times N_{bt}^+ \) matrix. Equivalently, let \( W_t \) be a \( N_{bt}^+ \times N_b \) selection matrix so that \( X_{bt}^+ = W_tX_{bt} \) is the \( N_{bt}^+ \times 1 \) vector of variables that contain new information at time \( t \). The measurement equation in terms of \( X_{bt} \) (which contains missing values) is

\[
W_tX_{bt} = W_t\Lambda_G(L)G_{bt} + W_te_{X_{bt}}. \tag{6}
\]

This is equivalent to using the entire \( T \times N_b \) matrix \( X_b \), which is padded with zeros when missing values are encountered, and then setting the Kalman gain to zero.

The third problem which makes our MCMC algorithm non-standard relates to the fact that \( G_{bt} \) conveys information about \( F_t \). More precisely,

\[
\Psi_G(L)G_{bt} = \alpha_{F, bt} + \epsilon_G \tag{7}
\]

where \( \alpha_{F, bt} = \Psi_G(L)\Lambda_F(L)F_t \) depends on \( t \). Given a draw of \( F_t \), this can be interpreted as a time-varying intercept that is known for all \( t \). By conditioning on \( F_t \), our updating and smoothing equations for \( G_t \) explicitly take into account the information carried by \( F_t \).

Summarizing, when a data series is unavailable for part of the sample, they are ‘zeroed out’ in the measurement equation. When a series is quarterly instead of monthly, then over the sample for which the data are available, we use data augmentation techniques to draw the monthly values. In Moench et al. (2009), we show that a simple extension of the algorithm in Carter and Kohn (1994) allows estimation of three level models that takes into account the dependence of \( G_b \) on \( F \). In this paper, we further modify the algorithm to accommodate the first two problems. Precisely, denote the observed national indicators of \( Y_t \) and the observed regional indicators \( X_{bt}, b = 1, \ldots, B \). Let

\[
\Sigma_{X_b} = \text{diag}(\sigma_{X_{b1}}^2, \ldots, \sigma_{X_{bN_b}}^2) \\
\Sigma_G = \text{diag}(\sigma_{G_{b1}}^2, \ldots, \sigma_{G_{bK_b}}^2) \\
\Sigma_F = \text{diag}(\sigma_{F_1}^2, \ldots, \sigma_{F_{K_F}}^2).
\]

These matrices are of dimension \( N_b \times N_b, k_G \times k_G \), and \( K_F \times K_F \), respectively. Collect \( \{\Lambda_G1, \ldots, \Lambda_GB\} \) and \( \Lambda_F \) into \( \Lambda \), \( \{\Sigma_G1, \ldots, \Sigma_GB, \Sigma_F\} \) into \( \Sigma \), and \( \{\Psi_G1, \ldots, \Psi_GB, \Psi_F, \Psi_X1, \ldots, \Psi_XB\} \) into \( \Psi \). We first use data available for the entire sample to construct principal components. These are used to initialize \( \{G_{bt}\} \) and \( \{F_t\} \). Based on these estimates of the factors, initial values for \( \Lambda_GB, \Psi_b, \Sigma_GB \) and \( \Sigma_F \) are obtained. Each iteration of the sampler then consists of the following steps:

1. Conditional on \( \Lambda, \Psi, \Sigma, \{G_t\} \) and \( \{Y_t\} \), draw \( \{F_t\} \).
2. Conditional on \( \{F_t\} \), draw \( \Psi_F, \Sigma_F \) and \( \Lambda_F \).

3. For each \( b \), conditional on \( \Lambda, \Psi, \Sigma \) and \( \{F_t\} \), draw \( \{G_{bt}\} \) taking into account time varying intercepts.

4. For each \( b \), conditional on \( \{G_{bt}\} \) and \( Y_t \), draw \( \Psi_{Gb} \) and \( \Sigma_{Gb} \).

5. For each \( b \), conditional on \( \{G_{bt}\} \), draw \( \Lambda_{Gbi} \). Also draw \( \Psi_{Xbi} \) and \( \sigma^2_{Xbi} \).

6. Data augmentation:

   i) For each \( b \) and conditional on \( \{G_{bt}\} \) and the parameters of the model, sample monthly values for elements of \( \{X_{bt}\} \) that are observed at lower frequencies.

   ii) Conditional on \( \{F_t\} \) and the parameters of the model, sample monthly values for those \( \{Y_t\} \) that are observed at lower frequencies.

   iii) Draw \( \Psi_Y \) using the augmented data vector for \( \{Y_t\} \).

We assume normal priors centered around zero and with precision equal to 1 for elements of \( \Lambda \) and \( \Psi \), and inverse gamma priors with parameters 4 and 0.01 for elements of \( \Sigma \). Given conjugacy, \( \Lambda_{Gb}, \Lambda_F, \Psi_{Xbi}, \Psi_{Gb}, \) and \( \Psi_F \) in steps 4 and 5 are simply draws from the normal distributions whose posterior means and variances are straightforward to compute. Similarly, \( \sigma^2_{Gb} \) and \( \sigma^2_{Xbi} \) are draws from the inverse chi-square distribution. Notice that the model for \( (G_{bt}, Y_t) \) is linear in \( F_t \) and it is Gaussian. Thus if there are no missing values in the data, we can run the Kalman filter forward to obtain the conditional mean of \( F_t \) at time \( T \) and the corresponding conditional variance. We would then draw \( F_T \) from its conditional distribution, which is normal. We then proceed backwards to generate draws \( F_{t|T} \) for \( t = T - 1, \ldots, 1 \) using the algorithm suggested by Carter and Kohn (1994) and detailed in Kim and Nelson (2000). Draws of \( \{G_{bt}\} \) can be obtained in a similar manner, as the model for \( X_{bt} \) is linear in \( G_{bt} \) and is Gaussian. This basic algorithm is modified to deal with a time varying intercept in the transition equation for \( G_{bt} \) and missing values as discussed above.

### 3.1 Estimates of \( F_t \) and \( G_{bt} \)

Our sample starts in 1975:1 and ends in 2008:5. The base case price factors are estimated from 7 series for each of the four regions, plus 9 national price series. The base case housing
market factors are estimated from 14 price and volume series for each of the four regions, plus 17 national series.

The model has a number of parameters that we need to specify. We assume \( k_{Gb} = 1 \) for all \( b \) and \( K_F = 1 \)^4. As discussed above, we set the upper-left element of the factor loading matrices \( \Lambda_F(0) \) and the four \( \Lambda_{Gb}(0) \) equal to one in order to separately identify the signs of the factors and factor loadings. We order the NAR’s ‘Median Sales Price of Single Family Existing Homes’ first both at the national and at the regional level. This implies that our estimated factors will have a positive correlation with the corresponding NAR price indexes, respectively. As stated above, we assume \( \epsilon_{Xbi}, \epsilon_{Gb}, \) and \( \epsilon_F \) to be AR(1) processes. Moreover, we let \( s_{Gb} = s_F = 2 \) so that the factors at both the regional and the national level are allowed to have a lagged impact of up to two periods on the respective observed variables. This allows us to accommodate lead-lag relationships between the housing cycles across the different regions and the nation as a whole. We begin with 20,000 burn in draws. We then save every 50th of the remaining 50,000 draws. These 1,000 draws are used to compute posterior means and standard deviations of the factors and parameters.

Table 3 reports estimates of the dynamic parameters and the variance of shocks to the common, regional, and series specific components. The unconditional variance of \( F_p \) and \( G_{p,b} \) are denoted \( \text{var}(F_p) \) and \( \text{var}(G_{p,b}) \), while the variances of \( \epsilon_{Fp} \) and \( \epsilon_{G,b} \) are denoted \( \sigma^2_{Fp} \) and \( \sigma^2_{G_{p,b}} \), respectively. The common price factor, \( F_p \) is more persistent but has a smaller variance than the regional factors, \( G_{p,b} \). Of the four regional factors, the West is the most persistent. Even though house price shocks in the Midwest are larger, the house price factor in the West has a larger unconditional variance once persistence is taken into account.

Figure 1 presents the national house price factor, denoted \( F_p \), along with the regional factors, denoted \( G_p \). The standard deviation of \( F_p \) is 0.636. The national factor (solid line) is notably smoother than the regional factors (dash-dotted line). The Northeast experienced housing busts in the early 1980s and the late 1980s that were much more pronounced than the national market. However, throughout the 1990s, the Northeast market is stronger than the national market. The West experienced a sharp decline in house prices in the mid 1970s and again in the early 1990s. These variations are larger than what was recorded for other periods, or in any of the other regions. Because of these two episodes, the \( G_p \) for the West has a standard deviation of 2.066, much larger than the 1.012 observed for the Northeast. The Midwest and the South have more tranquil housing markets. The standard deviation

---

^4 We considered \( k_{Gb} = K_F = 2 \), but the additional factors tend to have little variability and were subsequently dropped.
of $G_{p,b}$ are .515 and .078, respectively.

Table 3 also reports a decomposition of variance in the series used to estimate the factors. We find that shocks to the national house price factor, $\epsilon_{Fp}$ account for 34%, 24.7%, 34.1%, and 24.1% of house price variations in the four regions, respectively. Shocks to the regional factor $\epsilon_{Gp}$ have a share of 23.9% in the NorthEast, 27.1% in the West, 15.9% in the MidWest, and 8.3% in the South. Series specific shocks account for the remaining 42.1%, 48.2%, 50%, and 67.7% of the variation in the regional house prices series. Notably, the factor structure is strongest in the Northeast and is weakest in the South.

Figure 2 plots $F_p$ and four leading national house price indices. As expected, $F_p$ is somewhat smoother than the individual price series because $F_p$ is essentially a weighted average of all price indices. The Census monthly price index has a correlation of 0.67 with $F_p$, while the Conventional Mortgage Home (quarterly) price index has a correlation with $F_p$ of 0.93. Notably, both series are more volatile than $F_p$. The monthly FHFA series which is available since 1986 is also highly correlated with $F_p$: computed over the sample for which the two series overlap, the correlation coefficient is 0.99. The correlation between the Case-Shiller index and $F_p$ is 0.91. Notice, however, that there are significant differences between $F_p$ and the indicators in recent years. The four price indices seem to show sharper declines in house prices than the house price factor which incorporates information from various series.

One of our objectives is to investigate whether the house price cycle differs from the housing cycle, where the latter is defined based on data on prices as well as volume. Table 4 reports the posterior mean of the dynamic parameters and the variance of the shocks for the housing model. To distinguish them from the house price factors $F_p$, we denote the housing factors by $F_{pq}$ and the regional factors by $G_{pq}$. The common factor is still highly persistent. While a regional factor is not evident in house prices in the South, the data on volume help to isolate this factor.

A decomposition of variance of the housing market model reveals that national and regional shocks are equally important in the Northeast, the Midwest and the South, while regional shocks in the West are more important than the national shocks. However, the result that stands out is that idiosyncratic variation in the housing market data in all four regions are relatively more important than the common shocks and dominate the total variations in the data.

At the national level, the correlation between the house price factor $F_p$ and the housing market factor $F_{pq}$ is 0.85. In spite of this strong correlation and as seen from Figure 3, there is a notable difference between the two series during peaks and troughs. The discrepancy has
been especially pronounced since 2007. While $F_p$ is -2.006 at the end of our sample in May 2008, almost four standard deviations below the mean, $F_{pq}$ is -.517, roughly one standard deviation below the mean. The drop in housing activity as indicated by $F_{pq}$, estimated using both house price and quantity information, is thus less severe.

This section has focused on different measures of housing market activity, and several conclusions can be drawn. First, there is substantial regional variation in housing market activity with the regional component playing the largest role in the West. Second, our house price factor is highly correlated with each of the four widely used house price indices with the important difference that our $F_p$ is smoother. Third, $G_{pq}$ is generally similar to $G_p$ except in the South. At the national level, $F_p$ and $F_{pq}$ are well synchronized, but the decline of $F_{pq}$ since 2007 is much less pronounced than $F_p$. These observations suggest that there are important idiosyncratic movements in observed housing data. A particular house price series will not, in general, be representative of the true level of activity underlying the housing market. The more data we use to estimate the factors, the better we are able to ‘wash out’ the idiosyncratic noise. However, all indicators point to a sharp decline in housing market activity since 2007. This decline is pervasive and occurs at both the regional and national levels. We next investigate whether shocks to the housing market affect consumption.

4 Housing and Consumption

There exists little work on the regional aspect of housing variations. Using a dynamic Gordon model, [Ng and Schaller (1998)](NgAndSchaller1998) find that regional housing bubbles have predictive power for future consumption. [Campbell et al. (2008)](CampbellEtAl2008) find that housing premia are variable and forecastable and account for a significant fraction of the variation in the rent-price ratio at the national and regional levels. However, they do not assess the consumption effects of housing. One reason why there are so few estimates of the regional effects of housing is data limitation. Not only is it difficult to find regional housing data over a long time period, government statistical agencies do not publish consumption data at the regional level. We use regional retail sales data provided by the Census Bureau until 1997 and continued by the Bank of Tokyo-Mitsubishi (BTM) since then. These series are available monthly from 1970 onwards, both for each of the four Census regions, and also for the U.S. as a whole. This retail sales (which we will simply refer to as consumption) series is not seasonally adjusted. We run it through the X11 filter in Eviews, and deflate by the all items CPI. We then analyze the log annual difference of this seasonally-adjusted, real retail sales series.
4.1 Estimates from FAVAR

We are interested in quantifying the response of retail sales consumption to changes in regional and national housing market conditions. Retail sales, while representing only a subcategory of total consumption, have the advantage of being available for the main Census regions. We can therefore study the effects of housing market shocks on consumption both at the regional and the national level. The foregoing discussion suggests that a strong housing market will increase consumption of some but decrease the consumption of others. As those affected may have different propensities to consume, the aggregate effect of changes in housing market conditions on consumption is an empirical matter.

Our analysis is based on factor-augmented vector-autoregressions (FAVAR), a tool for analyzing macroeconomic data popularized by Bernanke et al. (2005). While a conventional VAR is an autoregressive model for a vector of observed time series, a FAVAR augments the observed vector of variables by a small set of latent factors often estimated by the method of principal components. Bai and Ng (2006) showed that if $\sqrt{T}/N \rightarrow 0$ as $N, T \rightarrow \infty$, the estimated factors that enter the FAVAR can be treated as though they are observed. The method of principal components is not, however, well suited for the present analysis for two reasons. First, the number of series available for analysis is much smaller than the typical large dimensional analysis in which principal components is applied. Second, we have a non-balanced panel with data sampled at mixed frequencies. Both problems are more easily handled by Bayesian estimation. Accordingly, our FAVAR is based on Bayesian estimates of the factors.

Our first set of FAVARs consist of five variables, respectively: $F_{pq}, G_{pq,b}, U, U_b$, and $RS_b$ where $U$ is the national unemployment rate, $U_b$ is the unemployment rate for region $b$, and $RS_b$ is the linearly detrended logarithm of real retail sales for region $b$. The variables $U_b$ and $U$ allow us to control for regional and aggregate business cycle conditions. We use the housing factors instead of the house price factors as these provide more comprehensive measures of the housing market. The regional unemployment rates are available only from 1976 onwards. Thus, for this exercise, the sample is 1976:1-2008:5. The standard deviations of the variables used in the FAVAR are given in Table 3.

We identify shocks to the housing market factor using a simple recursive identification scheme where the variables are ordered as they appear above. This identification implies that the national housing factor does not react to regional housing market shocks, regional and national unemployment shocks as well as regional consumption shocks within the same month. This assumption appears reasonable given that it usually takes at least a few weeks
from the time a decision is made to purchase, sell or construct a home before an actual transaction is being made. Our identification also implies that regional retail sales can respond within the same month to both national and regional housing shocks as well as national and regional labor market shocks, as would be the case if households can adjust their consumption decisions rather quickly in response to various kinds of economic shocks. As for the particular ordering between the national and regional housing market factors, it also appears intuitive to suppose that national housing market shocks may have an immediate impact on regional housing market dynamics whereas the reverse does not hold.

The impulse response functions are obtained as follows. First recall that we saved 1,000 of the 50,000 draws of $G_{pq,b}$ and $F_{pq}$ from the Gibbs sampler. For each draw of the regional and national housing factors, we estimate a five variable FAVAR with two lags for each of the four regions. The estimated FAVARs are then used to obtain impulse response of regional retail sales to shocks in $F_{pq}$ and $G_{pq,b}$. We also estimate a three variable VAR in $F_{pq}$, $U$, and $RS$ to study the impulse response of aggregate retail sales to housing market factor shocks. Repeating this for each of the 1,000 draws of $F_{pq}$ and $G_{pq,b}$ gives a set of impulse responses from which we can compute the posterior means and percentiles of the posterior distribution.

Figure 4 reports the posterior mean of regional consumption responses to a one standard deviation shock in $F$ along with the 90% probability intervals. The effect of a one standard deviation shock to $F$ is positive in all four regions. The response of retail sales is hump shaped. The shock triggers a permanent increase in the level of retail sales. The effects, similar across regions, peak about ten months after the shock. The cumulative effects in the four regions over two and a half years are .037, .057, .043, and .037, respectively. Hence, according to our estimates, a shock to the national housing market may result in a long-run increase of regional consumption between 3.7% and 5.7% above its trend level, all else being equal. While the positive consumption effect would be consistent with the idea that homeowners take advantage of a hot housing market, trading down their homes to enjoy realized capital gains, a more likely interpretation is that the positive response is due to the collateral effect brought about by increased home equity.

Figure 5 graphs the impulse response to a standard deviation shock in $G_{pq}$. Because $F_{pq}$ is also in the FAVAR and is ordered first, shocks to $F_{pq}$ can trigger a contemporaneous response in the regional housing market factor, while the reverse is not true. The results can thus be interpreted as shocks to regional housing market activity that are orthogonal to national housing market shocks. As indicated by wider probability bands, the effects of shocks to $G_{pq}$ are statistically less well determined than those to $F_{pq}$. While the effects at the
peak are about the same as the response to a national housing factor shock, the cumulative
effects of regional housing shocks are smaller and differ substantially across regions. The
long run effects are .028, .064, .026, and .043, respectively, implying an increase of regional
consumption due to regional housing market shocks between 2.6% and 6.4% above its trend
level, all else equal. Notably, the effects are largest in the West where variations in $G_{pq}$ are
also relatively more important. Even in this region, we find the effects of regional housing
shocks on consumption to be less pronounced than the national shocks.

In unreported results, we find that shocks to volume tend to reduce retail sales for a
few months after the shock, and have no substantial long-run effects. Furthermore, the
consumption response to $G_{pq}$ also seems to be largely due to the response to $G_p$. Thus, the
consumption responses we observe in Figures 4 and 5 are largely a consequence of shocks to
house prices rather than housing volume.

Given the heterogeneity in response across regions, what is the consumption response at
the national level? To assess this question, we estimate a FAVAR in three variables: aggregate
retail sales, $RS$, aggregate unemployment, $U$, and the national housing market factor, $F_{pq}$.
We again identify shocks to the housing market factor using a recursive identification scheme
where the ordering is as the variables appear above. The economic reasoning behind this
approach follows the discussion above.

The top panel in Figure 6 shows that the response of aggregate retail sales to a one
standard deviation shock in $F_{pq}$ is positive with a maximum effect occurring about 15 months
after the shock, and a cumulative effect of 0.04 after 30 months. This implies that a one
standard deviation shock may push aggregate retail sales 4% above their trend level. At the
same time, unemployment falls by over ten basis points as housing market activity increases.
Figure 6 also shows how $F_{pq}$ responds to its own shock. The response is gradual, and the
half-life of the shock is about 10 months.

We have presented results for counter-factual increases in housing market activity. A
policy question of interest is the quantitative consumption effect as the national housing
market contracts. Figure 4 implies that consumption is expected to fall immediately in all
regions as a result of a housing market shock, all else equal. As noted earlier, the housing
factor in May 2008 was -.571, about one standard deviation below the mean. A one standard
deviation $F_{pq}$ shock in the FAVAR is about .25. A two standard deviation shock to $F_{pq}$ can
thus have a cumulative consumption effect on the West of $2 \times .057$, or about 11 percent, and
about 8 percent in the other three regions.

The consumption effect is much larger if we look at the house price factor, recalling that
$F_p$ is estimated to be -2.006 in May 2008, almost four standard deviations below the mean. Our results then suggest that at its worst (about fifteen months after the shock), consumption can fall by 1.2% with an even higher cumulative effect than a shock to housing activity. The consumption effect thus depends on whether we think house prices alone reflect the state of the housing market, or whether volume information should be taken into account.

### 4.2 What Affects the Housing Factor?

The slumping U.S. housing market has been a deep concern for private citizens and policy makers alike. While as of May 2008, the last data point in our sample, our housing factor was only one standard deviation below average, housing market activities have further decelerated since. This raises the question of what might stimulate housing activity.

To address this question, we consider a FAVAR with six variables and four lags at the national level. These variables in the order they enter the FAVAR are the unemployment rate (U), the fed funds rate (FF), a 30 year effective mortgage rate (MR), our housing factor ($F_pq$), the University of Michigan's survey of consumer sentiment (MICH), and the log of the S&P 500 index (SP). The unemployment rate captures aggregate business cycle dynamics. The fed funds rate and the effective mortgage rate measure tightness in the money and loans market. The Michigan survey measures confidence for the economy, and the S&P 500 index is a proxy for changes in financial wealth. Arguably, each of these variables can be thought of having an effect on the level of housing activity.

We identify shocks in the FAVAR using a recursive ordering of the variables as they appear above. This ordering implies that the unemployment rate does not respond within the month to any of the shocks but its own. The Fed Funds rate is ordered second and hence responds on impact only to unemployment shocks and monetary policy shocks. The 30 year effective mortgage rate is ordered third which implies that it is assumed to respond on impact to unemployment and monetary policy shocks, but with a one-month lag to shocks to the housing market as measured by our estimated housing factor, consumer confidence, and the S&P 500 index. We put the housing factor in fourth position which implements the assumption that housing market activity cannot respond within the month to shocks to consumer confidence and stock prices. Finally, the fact that consumer confidence and the S&P 500 are ordered last implies that these two variables can respond within the month to unemployment, interest rate, and housing market shocks.

Figure 7 shows the response of $F_pq$ to shocks to each of the six variables.\footnote{Unreported results show that changes to the ordering of the six variables do not qualitatively affect our} We normalize
the shocks to the federal funds rate and the effective mortgage rate to have a contemporaneous -25 basis point impact on itself, respectively. All other shocks are one standard deviation shocks. Reductions in both interest rates lead to an increase in housing market activity. The maximum effect of a 25 basis points cut in the fed funds rate on the housing factor is 0.026 and the cumulative effect is 0.35. By contrast, the maximum effect of a 25 basis points reduction in the effective mortgage rate is 0.13 and the cumulative effect after 30 months equals 0.46. Hence, reducing mortgage rates has a larger maximum and cumulative effect on the housing factor than an equivalent cut of the fed funds rate. This result suggests that direct policy interventions in the mortgage market may represent an effective way to revive the housing market.

Interestingly, we find that a positive one standard-deviation shock to the unemployment rate boosts housing activity. This effect is due to a strong reduction of the Fed Funds rate following a negative shock to real activity. We further find that a one standard deviation increase in stock prices has a positive effect on housing activity both in the short-run and in the long-run. The maximum effect on $F_{pq}$ is 0.033, recorded three periods after the shock, and the cumulative effect is 0.51 which is about equal to two standard deviations of the housing factor. A one standard deviation increase in consumer confidence boosts the housing factor in the short term, but interestingly has a negative cumulative effect.

An overview of the results suggests that all else equal, housing market activity can be stimulated by a cut in the federal funds rate and a reduction of mortgage rates, the latter potentially having larger effects. Increases in stock prices will positively affect housing market activity in the short-run and in the long-run. Increased consumer confidence is found to stimulate housing market activity in the short, but may have a negative long-run effect on housing.

We also re-estimate the previous FAVAR using three different national house price series available for our full data sample as well as our estimated national house price factor, each standardized to have the same unconditional variance. We do not report these results here, but restrict ourselves to noting that the various indicators imply quite different reactions of house prices to the six shocks. The particular choice of house price measure thus potentially has a large impact on the conclusions one may reach from a quantitative analysis such as the one carried out above. conclusions.
5 Conclusion

This paper provides three new perspectives on the effects of housing shocks on consumption. First, we distinguish between house price shocks and shocks to general activity in the housing market. Second, we analyze regional as well as national level data. Third, our housing shock is not tied to a specific house price series. Instead, we extract house price and housing market factors from a large number of housing indicators. Our results indicate that in spite of large idiosyncratic variations, there is a national and a regional housing component in each of the regions, though the regional component is more important than the national component in the West. The aggregate response of consumption to national housing shocks is hump shaped. According to our estimates, the drop in housing market activity that began in 2006 can lead to a significant decline of consumption, all else being equal. Interest rate cuts can stimulate housing activity, and directly targeting lower mortgage rates may be an effective way to revive the housing market. However, without a boost in consumer confidence and the stock market, the housing market can remain depressed for a prolonged period of time. Our econometric framework permits a block structure and can handle data of mixed frequencies. Latent factors can also co-exist with observed factors. The methodology can be useful in other applications.
### Table 1: Regional Data

<table>
<thead>
<tr>
<th>Series</th>
<th>Source</th>
<th>Frequency</th>
<th>First Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Sales Price of Single Family Existing Homes</td>
<td>NAR</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Single Family Median Home Sales Price</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>Average Existing Home Prices</td>
<td>NAR</td>
<td>qly</td>
<td>Q1-1970</td>
</tr>
<tr>
<td>Average New Home Prices</td>
<td>NAR</td>
<td>qly</td>
<td>Q1-1970</td>
</tr>
<tr>
<td>Conventional Mortgage Home Price Index</td>
<td>FHLMC</td>
<td>qly</td>
<td>Q1-1970</td>
</tr>
<tr>
<td>FHFA Purchase-only Index</td>
<td>FHFA</td>
<td>mly</td>
<td>Jan1986</td>
</tr>
<tr>
<td>FHFA Home Prices</td>
<td>FHFA</td>
<td>qly</td>
<td>Q1-1970</td>
</tr>
<tr>
<td><strong>Volume data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New 1-Family Houses Sold</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>New 1-Family Houses For Sale</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Single-Family Housing Units Under Construction</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1980</td>
</tr>
<tr>
<td>Multifamily Units Under Construction</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1980</td>
</tr>
<tr>
<td>Homeownership Rate</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>Homeowner Vacancy Rate</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>Rental Vacancy Rate</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
</tbody>
</table>

### Table 2: National Data

<table>
<thead>
<tr>
<th>Series</th>
<th>Source</th>
<th>Frequency</th>
<th>First Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Sales Price of Single Family Existing Homes</td>
<td>NAR</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Median Sales Price of Single Family New Homes</td>
<td>Census</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Single Family Median Home Sales Price</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>Average Existing Home Prices</td>
<td>NAR</td>
<td>mly</td>
<td>Jan1994</td>
</tr>
<tr>
<td>Average New Home Prices</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1970</td>
</tr>
<tr>
<td>S&amp;P/Case-Shiller Home Price Index</td>
<td>S&amp;P</td>
<td>qly</td>
<td>Q1-1982</td>
</tr>
<tr>
<td>Conventional Mortgage Home Price Index</td>
<td>FHLMC</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>FHFA Purchase-only Index</td>
<td>FHFA</td>
<td>mly</td>
<td>Jan1986</td>
</tr>
<tr>
<td>FHFA Home Prices</td>
<td>FHFA</td>
<td>qly</td>
<td>Q1-1970</td>
</tr>
<tr>
<td><strong>Volume data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New 1-Family Houses For Sale</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Housing Units Authorized by Permit: 1-Unit</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Multifamily Units Under Construction</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Multifamily Permits US</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Multifamily Starts US</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Multifamily Completions</td>
<td>CENSUS</td>
<td>mly</td>
<td>Jan1968</td>
</tr>
<tr>
<td>Homeowner Vacancy Rate</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>Homeownership Rate</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
<tr>
<td>Rental Vacancy Rate</td>
<td>CENSUS</td>
<td>qly</td>
<td>Q1-1968</td>
</tr>
</tbody>
</table>
### Table 3: Estimates of $\Psi$ and $\sigma_F$: House Price Model

<table>
<thead>
<tr>
<th></th>
<th>$F_p$</th>
<th>$\text{var}(F_p)$</th>
<th>$\Psi_F$</th>
<th>s.e.</th>
<th>$\sigma^2_F$</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>0.639</td>
<td>0.942</td>
<td>0.054</td>
<td>0.020</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>2.066</td>
<td>0.744</td>
<td>1.058</td>
<td>0.126</td>
<td>1.300</td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>0.524</td>
<td>0.187</td>
<td>0.107</td>
<td>0.280</td>
<td>0.229</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>0.074</td>
<td>-0.018</td>
<td>0.002</td>
<td>0.079</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Decomposition of variance:

<table>
<thead>
<tr>
<th></th>
<th>$\epsilon_F$</th>
<th>s.e.</th>
<th>$\epsilon_{Gb}$</th>
<th>s.e.</th>
<th>$\epsilon_{Xb}$</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>0.340</td>
<td>0.087</td>
<td>0.239</td>
<td>0.053</td>
<td>0.421</td>
<td>0.051</td>
</tr>
<tr>
<td>W</td>
<td>0.247</td>
<td>0.083</td>
<td>0.271</td>
<td>0.060</td>
<td>0.482</td>
<td>0.056</td>
</tr>
<tr>
<td>MW</td>
<td>0.341</td>
<td>0.080</td>
<td>0.159</td>
<td>0.058</td>
<td>0.500</td>
<td>0.075</td>
</tr>
<tr>
<td>South</td>
<td>0.241</td>
<td>0.070</td>
<td>0.083</td>
<td>0.021</td>
<td>0.677</td>
<td>0.061</td>
</tr>
</tbody>
</table>

### Table 4: Estimates of $\Psi$ and $\sigma_F$: Housing Market Model

<table>
<thead>
<tr>
<th></th>
<th>$F_{pq}$</th>
<th>$\text{var}(F_{pq})$</th>
<th>$\Psi_F$</th>
<th>s.e.</th>
<th>$\sigma^2_F$</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>0.554</td>
<td>0.896</td>
<td>0.069</td>
<td>0.028</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>1.025</td>
<td>0.530</td>
<td>0.300</td>
<td>0.145</td>
<td>0.109</td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>2.019</td>
<td>0.766</td>
<td>1.636</td>
<td>0.139</td>
<td>2.341</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>0.127</td>
<td>0.206</td>
<td>0.009</td>
<td>0.101</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

Decomposition of variance:

<table>
<thead>
<tr>
<th></th>
<th>$\epsilon_F$</th>
<th>s.e.</th>
<th>$\epsilon_{Gb}$</th>
<th>s.e.</th>
<th>$\epsilon_{Xb}$</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>0.164</td>
<td>0.036</td>
<td>0.147</td>
<td>0.028</td>
<td>0.689</td>
<td>0.028</td>
</tr>
<tr>
<td>W</td>
<td>0.055</td>
<td>0.023</td>
<td>0.247</td>
<td>0.048</td>
<td>0.698</td>
<td>0.044</td>
</tr>
<tr>
<td>MW</td>
<td>0.114</td>
<td>0.029</td>
<td>0.128</td>
<td>0.019</td>
<td>0.758</td>
<td>0.016</td>
</tr>
<tr>
<td>South</td>
<td>0.130</td>
<td>0.032</td>
<td>0.154</td>
<td>0.027</td>
<td>0.717</td>
<td>0.026</td>
</tr>
</tbody>
</table>

22
References


Ng, S. and Schaller, H. 1998, Do Housing Bubbles Affect Consumption, Boston College, mimeo.


Figure 1: National and Regional House Price Factors

This figure shows the estimated posterior mean of the national ($F_p$) versus the regional house price factors ($G_p$) for each of the four Census regions. The sample period is 1975:01 - 2008:05.
Figure 2: National House Price Factor and Leading House Price Indices

This figure shows the estimated posterior mean of the national house price factor ($F_p$) versus four leading national house price indices. “NAR” is the Median Sales Price of Single Family Existing Homes from the National Association of Realtors (NAR); “CMHPF” is the Conventional Mortgage Home Price Index from Freddie Mac; “FHFA” is the Purchase-only House Price Index from the Federal Housing Finance Administration (formerly OFHEO); “Case-Shiller” is the S&P/Case-Shiller Home Price Index published by Fiserv Inc. While the NAR series is available at the monthly frequency, the latter three indices are only available quarterly. Our estimated house price factor is a monthly time series. The sample period is 1975:01 - 2008:05.
Figure 3: National Housing Factor and National House Price Factor

This figure shows the estimated posterior mean of the national housing factor, $F_{pq}$, and the estimated posterior mean of the national house price factor, $F_p$. The former is estimated using both price and volume data whereas estimation of the latter is exclusively based on home price data. The sample period is 1975:01 - 2008:05.
Figure 4: Impulse Responses of Regional Retail Sales to National Housing Shocks

This figure shows the estimated posterior mean and 90% probability interval of impulse responses from the FAVARs discussed in Section 4.1. For each region $b$, these contain the following five variables: $F_{pq}, G_{pq,b}, U, U_{b}$, and $R_{b}$ where $F_{pq}$ and $G_{pq,b}$ denote the national and regional housing factor, $U$ and $U_{b}$ the national and regional unemployment rate, and $R_{b}$ the linearly detrended logarithm of real retail sales for region $b$. We identify shocks using a recursive identification scheme of the five variables ordered in the way they appear above. The FAVAR has two lags. The sample period is 1976:01 - 2008:05.
Figure 5: Impulse Responses of Regional Retail Sales to Regional Housing Shocks

This figure shows the estimated posterior mean and 90% probability interval of impulse responses from the FAVARs discussed in Section 4.1. For each region $b$, these contain the following five variables: $F_{pq}, G_{pq,b}, U, U_b,$ and $RS_b$ where $F_{pq}$ and $G_{pq,b}$ denote the national and regional housing factor, $U$ and $U_b$ the national and regional unemployment rate, and $RS_b$ the linearly detrended logarithm of real retail sales for region $b$. We identify shocks using a recursive identification scheme of the five variables ordered in the way they appear above. The FAVAR has two lags. The sample period is 1976:01 - 2008:05.
Figure 6: **Impulse Responses of National Retail Sales to National Housing Shocks**

This figure shows the estimated posterior mean and 90% probability interval of impulse responses from the three-variable FAVAR discussed in Section 4.1. This contains national retail sales, $RS$, the national unemployment rate, $U$, and the national housing market factor, $F_{pq}$. We identify shocks using a recursive identification scheme of the three variables in the order they appear above. The FAVAR has two lags. The sample period is 1976:01 - 2008:05.
Figure 7: **Impulse Responses of National Housing Factor to Different Shocks**

This figure shows the estimated posterior mean and 90% probability interval of impulse responses from the FAVAR discussed in Section 4.2. This contains the national unemployment rate, $U$, the Fed Funds rate, $FF$, a 30 year effective mortgage rate from Freddie Mac, $MR$, our national housing factor $F_{pq}$, the University of Michigan’s survey of consumer sentiment, $MICH$, and the log of the S&P 500 index, $SP$. We identify shocks using a recursive identification scheme of the variables in the order they appear above. The FAVAR has four lags. The sample period is 1976:01 - 2008:05.