Modeling Systemic House Price Risk

By Francois-Michel Boire and Simon van Norden*

Economists and policy makers have become increasingly aware of the contribution of house price risk for overall financial fragility. We model the determinants of house price risk at the state level, including supply, demand and non-fundamental factors. Using quantile regressions, we find that some variables principally affect mean house price changes, while others skew the distribution of such changes towards relatively more or less downside risk. Interstate differences in these variables imply that some states faced much larger increases in downside risk during the recent Financial Crisis, consistent with the observed differences in housing price declines. *Keywords: Housing Bubbles, Systemic Risk, Quantile Regression, Value at Risk*

During the second half of the 20th century, the share of domestic banking credit allocated to the business sector in many industrial economies plummeted while that of mortgage loans swelled, making residential housing finance a cornerstone of developed economies banking system.[1] As a result, real estate repricing risk became an important source of systemic risk to the financial system, and many of the most financially-advanced economies (including the United States, the United Kingdom and Japan) experienced profound financial crises in which drops in real estate prices played an essential role.[2]

In such crises, the financial sector has tended to grossly underestimate the risk of a simultaneous housing crash.[3] Indeed, at the eve of the Great Recession, the joint distribution of regional house prices was commonly estimated with Gaussian copulas, effectively assuming independence of house price changes across regions. However, Zimmer (2012) found that even before the market crash, a simple model specification procedure could strongly reject the Gaussian copula in favour of highly dependent house prices for tail events. In this way, the Gaussian copula led credit rating agencies to underestimate house price dependence and over-price mortgage-backed securities.

In this paper, we present a new model of house price risk based on the use of a panel of state-level price indices and quantile regressions. Comovements in house prices across states are modeled as the response to national economic

* Boire: HEC Montreal and University of Western Ontario, fboire2@uwo.ca. van Norden: HEC Montreal, CIRANO and CIREQ, simon.van-norden@hec.ca. The authors would like to thank Philippe d'Astous and Lars Stentoft for their suggestions, and the UK Housing Observatory at LUMS for the use of their GSADF code. All errors are our own faults.

[1] For example, see Jorda, Schularick and Taylor (2016).

[2] Jorda, Schularick and Taylor (2017) explore the role of real estate risk in financial fragility.

[3] Glaeser (2013) provides a review of housing market speculation and crashes in US history.

fundamentals (such as interest rates) as well as state-level fundamentals that may themselves be highly correlated across states (such as unemployment rates.) The results are distributions of house price changes whose predicted movements are both highly persistent across time and correlated across states.

The framework also allows us to investigate the explanatory power of variables capturing changes in both the supply of and demand for housing, as well as the influence of non-fundamental or speculative factors. Our preferred model finds that both fundamental and non-fundamental factors significantly explain changes in house price risk. We also find that while some variables shift the whole distribution of house price changes up or down, others tend to change the shape of the distribution. The ability to capture both types of these effects is one of the advantages of the quantile regression framework.

In the next Section of the paper, we briefly review the literature on the post-WWII determination of house prices. We also roughly divide explanatory variables into supply-side, demand-side, financial and non-fundamental factors. Section II describes the data used in our analysis, including details on the measures of market "exuberance" proposed by Pavlidis et al. (2016). Section III explains the panel quantile regression methods used for estimation and inference while Section IV presents and analyses our results. Section V concludes.

I. Literature Review

The literature on modelling house prices has focused on a wide range of variables covering many different conceptual approaches. From a purely financial standpoint, one can think of a house as an asset whose price reflects the present value of the stream of housing services (or rent for non-occupying owners) it will provide. We refer to factors affecting that present value as "fundamental" factors. These include supply and demand factors, each of which we review below. Non-fundamental factors are those which come into play exclusively when one intends to sell the asset at a future date rather than simply hold it indefinitely. With some oversimplification, non-fundamental factors are most often associated with "irrational" behaviour and/or speculative bubbles.

A. Fundamentals

We organise our discussion of the fundamental determinants of house prices by dividing them into supply and demand factors. Supply factors affect the supply of new houses. Demand factors influence the demand for existing houses, mainly via changes in household wealth. However, housing demand may also be affected by the ease with which households can finance their purchase. The Global Financial Crisis greatly stimulated research on the linkages between housing prices and the financial sector, so much so that we discuss this section of the literature separately.

Supply Factors. — The supply-elasticity of land varies considerably and is an important influence on the evoluation of house prices. When constructors are capable of promptly responding to a surge in demand, they are effectively clearing the scarcity effect that would otherwise inflate prices, thereby impeding house price appreciation. During the 1980s and 2000s housing bubbles, states with highly elastic supply experienced markedly shorter house price booms than metropolitan areas with inelastic supply (Glaeser, Gyourko and Saiz (2008)). Most recently, Knoll, Schularick and Steger (2017) show that in a sample of 14 developed countries dating back to 1870, land prices play an important role in explaining long-run house price growth. They find that at the turn of the 20th century, housing growth was relatively stable and widely attributed to the tangible improvements made on a home, such as access to electricity and water. In the late 1900s however, land appreciation appears to be a primary factor in house price growth, implying that the supply elasticity of housing play a critical role in understanding long-term price trends. We allow for supply elasticity to vary across states in our analysis by including state-specific fixed effects in our models, as we explain below.

Supply factors may also affect house prices, among many other things, via the terms of trade. An improvement in the terms of trade (for example, a fall in oil prices for an oil importer) implies a decrease in the price of imports relative to that of domestic consumption. As housing is a non-traded good, this implies higher *real* house prices. Corrigan (2017) quantifies the cointegrating relationship between the relative price of housing and the terms of trade. With a panel VECM framework of 18 developed economies, Corrigan identifies a strong negative long-run correlation between import and house prices over the 1994-2015 period, where persistent declines in import prices explain a substantial portion of house price variance. He finds evidence that a decline in real import prices tends to drive down domestic non-housing consumption prices, resulting in decreased inflation expectations and pressuring short-term interest rates downward. In turn, as the price of non-housing services and interest rates fall, real house prices and household debt increase. The relationship is much weaker for net exporters, however. We will use the terms-of-trade index from the US Bureau of Economic Analysis to control for international-exchange driven housing supply shocks.

Demand Factors. — One might contend that demographics are inherently connected to housing demand, and prices should change accordingly. Although this claim holds true in the long run, total population is not a leading indicator of real estate investment. Monnet and Wolf (2017) recently found that in OECD countries, demographic trends of the 20-49 age group are better predictors for house price variations than any other macro-financial correlate. However, this analysis is undependable in instances of intense migratory flows because demographic cycles are hard to assess when a region experiences migration. Since migration patterns widely differ across the US, state level demographic trends are difficult to measure and compare, so we omit this predictor in our empirical work.

Household income is also thought to be an important determinant of house prices. In their influential study, Case and Shiller (2003) found that income per capita proved to be a powerful predictor of house prices in almost all states between 1985 and 2002. Indeed, markets with low house price-to-income ratios had stable house prices that were highly correlated with per capita income trends. Conversely the states with the largest ratios had the largest house price variability. In these states, Case and Shiller (2003) found that additional variables played important roles, such as unemployment, mortgage rates, housing starts, population, and income-to-mortgage payments ratios. For these states, unemployment had a negative impact on house prices between 1985 and 1999, although the relation was less clear over the 1985 to 2002 period.

Financial Factors. — It should be no surprise that the availability of credit is important factor in determining housing demand. However, given the importance of mortgage finance to modern banking that we noted at the outset, the health of the housing market may also have important effects on the state of the banking sector. The aftermath of the 2008 housing market collapse and financial crisis in the US has spurred empirical research into their links. Jorda, Schularick and Taylor (2016) point out that real estate downturns are liable to lead to deeper recessions and slower recoveries. Bauer (2014) shows that the implementation of restrictive credit policies often precedes house price corrections, and that total credit-to-GDP and bank credit-to-GDP ratios help forecast housing turning points. These are the same credit ratios that the BIS had found help to predict financial crisis.[4] Laeven and Valencia (2008) show that mortgage leverage stands as a critical element in detecting systemic banking crises and their financial cost.[5]

For our empirical analysis, we looked for variables that could capture both the relative scarcity of credit for housing investment as well as the potential riskiness of housing investment. We included long-term mortgage interest rates as an measure of the cost of mortgage credit, and its spread over the treasury rate as a measure of perceived riskiness.[6] We also included the mortgage debt-to-GDP as measure of mortgage leverage in the financial system as well as an indicator of housing affordability.

[4]See also Reinhart and Rogoff (2009) and Schularick and Taylor (2012).

[5]Work by Zimmer (2012) and Ho, Huynh and JachoChavez (2016) show that state-level US house price movements are much more highly correlated in the tails of their distributions, implying that the systemic risk to the financial system increases much more rapidly than standard linear models would predict.

[6]The Treasury spread could also be interpreted as measure of credit tightening. Bauer (2014) shows that the implementation of restrictive credit policies often precedes house price corrections.

B. Non-Fundamentals

As substantial body of work has documented the apparent role of non-fundamental factors in the determination of house prices.[7] Perhaps the most influential of these is Case and Shiller (2003), who examined two nationwide US housing market surveys carried out in 1988 and 2003. They found that a widespread optimism in the market may prompt panic buying, artificially inflating prices. Conversely, when signs of a slowdown become too obvious and the public becomes more pessimistic, fire sales may provoke a sharp price correction, inducing and maintaining high rates of mortgage delinquency. Case and Shiller (2003) conclude that public expectations about the market, despite their evident signs of bias, can sustain house price overvaluations. Since then, the fall of the US housing market in 2007-2009 as well as the improved availability and quality of housing price series greatly stimulated the growth of this literature. Among many other works, Glaeser (2013) chronicles US real estate history, documenting the publics serious flaws in predicting regional real estate price movements. Glaeser and Nathanson (2017) present a model in which excess volatility and realistic bias in household forecasts can be explained by buyers with modestly extrapolative expections.

Pavlidis et al. (2016) measure the contribution of non-fundamental factors using the Backward Supremum Augmented Dickey-Fuller (BSADF) unit-root test developed by Phillips, Shi and Yu (2015), which is designed to detect (possibly multiple) episodes of explosive dynamics in univariate time series. Periods during which the BSADF exceeds the critical value are said to show "exuberance". Using data from the International House Price Database of the Federal Reserve Bank of Dallas (FRB Dallas), the authors extract yearly country-specific BSADF test statistics for real house prices and price-to-income ratios. The latter is intended to capture more precisely periods during which house price movements do not reflect changes in market fundamentals. The authors found that episodes of exuberance for house prices and price-to-income ratios roughly coincide, but periods of exuberance are shorter for price-to-fundamentals. Negative exuberance levels are also observed when time-series display descending trajectories. To predict the likelihood of a housing bubble, Pavlidis et al. (2016) use probit models to predict an indicator of their housing exuberance index and show that the long-term interest rates, private credit growth, personal disposable income growth, unemployment, and GDP growth are significant predictors of price and price-to-income exuberance.

As shown in Figure I.B, we see that their exuberance measures for US house prices started in the late 1990s during the dot-com bubble, but the price-to-income ratio only started to show explosiveness in the early 2000s. One explanation Pavlidis et al. (2016) propose for this is that fundamental economic growth during

[7] The role of non-fundamental factors has frequently been contested. For example, compare Baker (2011) with Himmelberg, Mayer and Sinai (2005) on US house prices, or Roche (2001) and McQuinn and O'Reilly (2008) on Irish house prices.

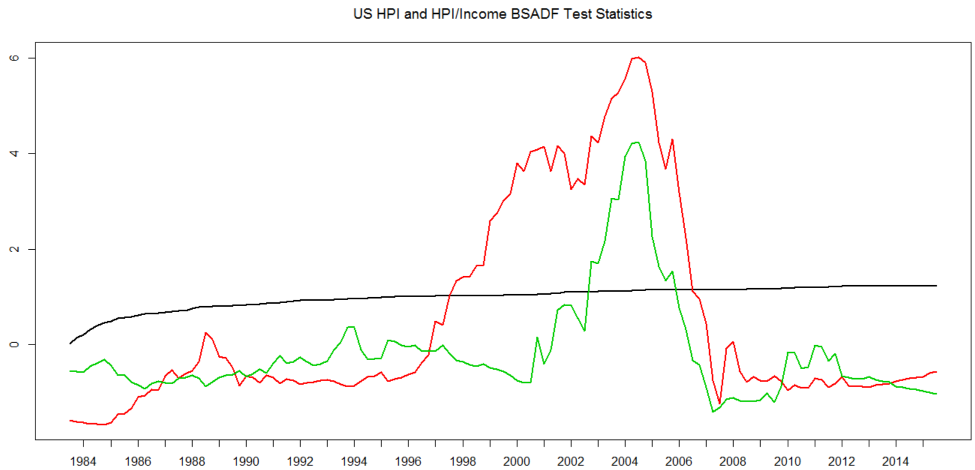


FIGURE 1. BSADF STATISTICS

Note: BSADF test statistics for US house prices (red) and price-to-income ratios (green) with 95 % critical values (black).
SOURCE: Federal Reserve Bank of Dallas: *https://www.dallasfed.org/institute/houseprice#tab1*

the tech boom caused the sequential price increases of the late 1990s and early 2000s, but later in 2003, the boom turned into a bubble. The price-to-income ratio then began to show exuberance, now indicating a rather unsustainable growth. In fact, after the dot-com bust, real economic growth came to a halt, yet real estate prices continued to grow. By the time the signs of a housing collapse became tangible in the financial and banking sectors, the run-up of house prices had already propagated around the world leading to a nearly simultaneous crash in 2006. They conclude that house price and price-to-income exuberance levels can give us clues about the underlying economic mechanism generating explosive prices.

II. Data

In this section we detail the data series we used as measures of the factors mentioned above. Unless specified otherwise, all series are quarterly and span the 30-year period from 1986Q1 to 2015Q4 (120 quarters, the longest for which all series were available.) Our balanced panel consists of the 50 states and the District of Columbia. Where state-level data were not available, national aggregates were used for all 51 “states”.

A. Dependent Variable

House Prices. — We measure fluctuations in house prices as quarterly changes in the natural logarithm of the Office of Federal Housing Enterprise Oversights (OFHEO) House Price Index (HPI). The state-level quarterly HPI is based on sales prices and appraisals of all repeat-sale detached single-family homes whose mortgages are purchased by Fannie May or Freddie Mac.[8] This therefore omits price trends related to vacant lands, multi-unit properties, and commercial real estate. Nevertheless, the share of total outstanding mortgages related to single-family homes historically dominated those of multifamily, commercial and farm mortgages. In fact, outstanding single-family home mortgages took about half of total outstanding mortgages in the mid-1970s. According to the Federal Reserves Flow of Funds Report, the share of single-family home mortgages increased to over 70% before the 2007 crisis and has since remained above that level.

B. Independent Variables: Supply Factors

Terms of Trade. — We measure the terms of trade using the Bureau of Economic Analysis' (BEA) quarterly national terms-of-trade index, which aims to reflect the purchasing power of the US in international markets. An increase in the index indicates a rise in the relative price of imports to exports, and conversely. Corrigan (2017) predicts a negative relationship between terms of trade and house prices.

Vacancy Rates. — In order to capture housing stock supply trends, we use yearly rental and home vacancy rate data from the US Bureau of the Census. The raw data are annual; we use the same annual value for all quarters in the calendar year. Vacant houses provide no stream of benefits; this should reduce their fundamental value. We therefore expect higher home vacancy rates to be associated with lower home price growth. Home vacancy rates can also be a proxy for regional real estate supply elasticity. Glaeser, Gyourko and Saiz (2008) note that supply elastic regions can respond to growth in housing demand with smaller price increases and more construction of new supply. Therefore, home vacancy rates may also change the scale of house price changes, where a high vacancy rate narrows the distribution of price changes.

Homes purchases and rentals are housing substitutes. Low rental vacancy rates could indicate a shift in demand from home purchases to rentals, suggesting a positive correlation between house prices and rental vacancy. Alternatively, rental vacancies could be low due to increased demand for all forms of housing and an inelastic supply, which in turn suggests a negative correlation. The expected sign of the coefficient on this variable is therefore ambiguous.

[8]For further details, see Calhoun (1996).

C. *Independent Variables: Demand Factors*

Personal Income. — We measure personal income as the state-level percent change in quarterly, current-dollar, seasonally-adjusted personal income series as reported by the BEA. We expect a positive correlation between changes in personal income and HPI changes across all quantiles. Kahn (2009) argues that expected productivity growth increases housing demand in anticipation of income growth, while Rünstler and Vlekke (2016) find that GDP and housing cycles are markedly synchronised in the US and the 5 largest European economies. Therefore, for good measure, we also include national GDP growth, as measured by the quarterly changes in the natural logarithm of current-dollar national GDP as reported by the BEA. We also expect this variable to have a positive correlation with changes in the HPI.

Unemployment. — After controlling for the income factors mentioned above, higher unemployment rates increase the perceived riskiness of household income, thereby reducing the investment demand for homes. We therefore include seasonally-adjusted state-level unemployment rates as published by the US Bureau of Labor Statistics (BLS). The monthly published series is converted to quarterly observations using end-of-quarter values. We also expect this variable to have a negative correlation with changes in the HPI.

Property Taxes. — Higher property taxes reduce the effective demand for housing by increasing housing ownership costs. They also differ substantially across states. Some states (e.g. Delaware, Hawaii, Oregon) have no personal property taxes. For most other states, personal property tax revenues vary between 0.05% and 0.2% of personal income. We control for this by constructing the property-taxes-to-personal-income ratio based on state-level Personal Income and Personal Property Taxes from the BEAs regional database. The raw data are annual; we use the same annual value for all quarters in the calendar year. We expect this variable to have the opposite effect of household income, and so be negatively correlated with the change in house prices.

D. *Independent Variables: Financial Factors*

Interest Rates. — We capture the impact of the primary mortgage market on housing demand with the national 30-year conventional mortgage rate. Based on Freddie Macs fixed rate mortgage commitments data, this monthly, non-seasonally adjusted series was released by the US Board of Governors of the Federal Reserve System . We convert this series to quarterly data using end-of-period observations. Since an increase in mortgage funding costs directly impacts housing affordability, higher mortgage rates are supposed to have a negative effect on house prices.[9] We add the 5- to 1-year yield spread of constant maturity treasury rates to our set of covariates to control for interest rate expectations. Indeed, interest rate spreads are commonly used as a forecasting tool of financial market returns. Positive yield spreads generally indicate expectations of stable economic growth, whereas trivial or negative yield spreads signal flat or inverted yield curves. In the event of a recession, house prices usually experience sharp corrections. We therefore predict that this covariate will be positively correlated with house prices. In other words, a flat or inverted medium-to-short-term yield curve can amplify the size of housing devaluation because it signals vulnerability in economic fundamentals. 5- and 1-year constant maturity treasury rate data are retrieved from the The Federal Reserve Bank of St. Louis FRED database and converted to quarterly series using end-of-period observations.

Mortgage Debt. — We also include the ratio of outstanding domestic mortgage debt to GDP as a real estate leverage measure. National outstanding mortgage debt data are taken from the Board of Governors of the Federal Reserve System and consist of quarterly measures of outstanding mortgages for all types of properties: one- to four-family residences, multifamily residences, farms, and non-farm/non-residential properties. Since real estate credit expansions often precede a downturn, high mortgage-to-GDP ratios are observed during a boom. For this reason, we expect an exacerbated negative effect of real estate leverage on the lower quantiles of house price changes. We also expect a credit expansion to stimulate house prices, possibly driving up the upper-quantiles. In other words, we predict that real estate leverage causes a scale shift that increases the dispersion of the distribution of house price changes.

E. Independent Variables: Non-Fundamental Factors

To measure the contribution of non-fundamental factors, we use the measures of housing market exuberance developed by Pavlidis et al. (2016) and maintained by the Federal Reserve Bank of Dallas as state-specific series of the exuberance levels of house prices and house-price-to-income ratios.[10] Instead of comparing the BSADF statistics to critical values, we directly include them as independent variables. This allows us to preserve information regarding the intensity of the explosive behaviour of regional prices and price-to-income ratios.

The explosive trends detected by the BSADF statistics are unsustainable by definition and reflect non-fundamental behaviours associated with models of rational speculative bubbles. In such models, the high capital gains possible should

[9] Interestingly, Case and Shiller (2003) found that mortgage rates had no apparent effect on house prices. They argued that simultaneity can cause the effects of mortgage rates to be ambiguous. For example, cheap financing may encourage housing demand, but banks may set low mortgage rates to restore the demand when real estate prices enter a downturn.

[10] Details on the construction of these series are provided in the Appendix.

the bubble continue to grow are counterbalanced by large losses possible should the bubble collapse. The larger the BSADF statistic, the larger the expected differences between the possible capital gains and losses. We should therefore expect increases in the BSADF statistic to be associated with a spread in the distribution of house price changes; it should increase the upper quantiles and reduce the lower quantiles.

Table 1 summarizes the variables and data series used in our model.

Table 1—Variable Summary

Category	Variable	Definition	Frequency (Raw Data)	Geography	Predicted Sign
	HPI	Change in log HPI	Quarterly	State	
Supply	**ToT**	Terms of trade	Quarterly	National	-
Supply	**Home Vacancy Rate**	Vacancy rate: single-family detached houses	Annual	State	-
Supply	**Rental Vacancy Rate**	Vacancy rate: multi-unit residential buildings	Annual	State	
Demand	**Income**	Percent change in real personal income	Quarterly	State	+
Demand	**GDP**	Change in log real GDP	Quarterly	National	+
Demand	**Unemployment**	Unemployment rate	Monthly	State	-
Demand	**Property Taxes**	Ratio of property taxes to income	Quarterly	State	-
Financial	**Mortgage Rate**	30-year conventional mortgage rate	Monthly	National	-
Financial	**Treasury Spread**	5- to 1-year Treasury Rate Spread	Monthly	National	+
Financial	**Mortgage Debt**	Total mortgage debt / GDP	Quarterly	National	-
Non-Fundamental	**Price Exuberance**	HPI BSADF statistic	Quarterly	State	+
Non-Fundamental	**Price/Income Exub.**	HPI/Income BSADF statistic	Quarterly	State	-

III. Methodology

In this section, we provide a brief introduction to quantile regression (QR) and define the empirical model we investigate. Next, we describe the inference issues that arise in a QR panel framework. Two robust inference procedures are implemented: bootstrapped confidence intervals and clustered covariance matrix (CCM) estimation. The former is adapted to a random effects model with clustered, serially-correlated error terms. The latter allows for fixed effects and heteroscedastic, serially-correlated error terms.

A. *Quantile Regression*

Introduced by Koenker and Bassett (1978), linear quantile regression methods depart from Least Squares methods by directly estimating the entire quantile process instead of relying on error independence and normality assumptions to indirectly infer the quantiles from the conditional mean of the outcome variable[11]. The θ-th conditional quantile of y is defined as the value $q_y(\theta)$ such that $\theta = F\left(q_y(\theta)|X\right)$, where $F\left(.|X\right)$ is the CDF of y conditional on X.

Conditional quantiles can be estimated across a set $\theta \in (0, 1)$. By estimating QR processes at various quantiles, we can quantify the effect a given covariate has

[11]For an introduction to QR techniques, see Davino, Furno and Vistocco (2014) and Koenker and Bassett (1978).

on different regions of the conditional distribution, enabling the user to uncover location and scale effects of a covariate. Here, covariates whose effects are homogeneous across the quantiles are said to induce a "location shift" because their effect causes a parallel movement of the quantiles, resulting in a translation of the conditional distribution. Covariates that induce a "scale shift" have coefficients that vary across the quantiles. Such heterogeneous effects can in turn stretch or distort the conditional distribution of the outcome. In brief, QR methods make no assumptions regarding the distribution of the outcome variable, and therefore stand as a powerful tool to characterize asymmetrical responses in the distribution of the variable of interest. (A more technical introduction to the basic properties of quantile regressions may be found in the Appendix.)

One appealing feature is that quantile regressions are "distribution-free", referring to the absence of parametric distributional assumptions on the error terms. In fact, QR tolerates error terms with distributional asymmetries and skewness distortions, heteroscedasticity, and serial correlation. Coefficient estimates are asymptotically normally-distributed even in general cases of non-identically distributed or dependent errors.[12] In finite samples, slope estimates become skewed to the left for the lower quantiles and skewed to the right for higher quantiles as we move away from the median quantile estimate. Hypothesis testing performance with skewed coefficients can therefore lead to over-rejection rates at extreme quantiles, but simulation evidence shows that the estimators remain unbiased.[13]

Since its inception, QR has found a number of applications. With panel data applications becoming particularly popular in recent years, clustered data and fixed-effects are also growing topics of interest in QR modelling. To model housing price risk at the state-level, we consider a panel model with $N = 51$ states and $T = 120$ quarters of the form

$$q_{\Delta y}(\theta, i, t) = \alpha\left(\theta\right)_i + X_{i,t} \cdot \beta\left(\theta\right) + \varepsilon(\theta)_{i,t} \tag{1}$$

where $q_{\Delta y}(\theta, i, t)$ is the θ-quantile at time t of the conditional distribution of house price changes Δy in state i after conditioning on a vector of independent variables $X_{i,t}$. Unlike conventional linear regression, where the quantiles are jointly determined by fixed $\{\alpha_i, \beta\}$ coefficients and assumptions on the distribution of the regression errors, here the coefficients α_i and β are free to vary with θ. The alpha coefficients may also vary across states; these "fixed effects" capture the possible effects of omitted variables which vary across states but not time (for example, such as the supply elasticity of housing.) We also estimate the model without fixed effects (so that $\alpha_i = \alpha\ \forall i = 1, \ldots, N$.) Time variation in the quantiles enters solely through our vector of independent variables X_{it}. We do not allow for time-fixed effects.

[12]Davino, Furno and Vistocco (2014)
[13]*Ibid.*

Unlike conventional least-squares estimation, in which estimation of fixed-effect parameters may be separated from estimation of the remaining parameters, quantile regression requires that all coefficients be estimated simultaneously. This is because the first order condition of the QR problem requires that $F\left(\alpha\left(\theta\right) + X \cdot \beta\left(\theta\right)\right) = \theta$. Since the conditional quantile is not a linear operator, QR slopes cannot be estimated separately from fixed effects, which greatly increases the dimension of the optimization problem. (For example, see Powell (2016).) In our application, incorporating state-level fixed effects increases the dimension of our parameter vector from 13 to 64. The QR fixed-effects literature also cautions that in panel data, a large number of individuals relative to the number of observations may induce data multicollinearity (see Hagemann (2017) and Davino, Furno and Vistocco (2014)).[14]

We estimate the panel-data QR specified above with the 12 independent variables listed in Table 1, which cover a range of supply, demand, financial and non-fundamental factors. In addition, we also included a lagged dependent variable as there is substantial evidence of serial correlation in house price changes. (The median state-level first-order autocorrelation coefficient was 0.59. See Figure 2.)[15] We estimated regression coefficients for the model without fixed effects for $\theta =$ the quartiles and the deciles. We also estimated models with fixed-effects; results were generally similar across the two specifications.

B. *Robust Inference*

Despite including a lagged dependent variable in our model, Figure 3 shows that first-order serial correlations of the within-state residuals vary between -0.6 to 0.6. Simulation evidence implied that conventional QR standard errors in such cases led to tests with substantial size distortion. We therefore investigated two alternative methods for producing robust standard errors for panel QR regressions.

Bootstrapped Standard Errors. — The wild gradient bootstrap (Hagemann (2017)), extended from the wild bootstrap of Chen, Wei and Parzen (2004), enables us to compute valid standard errors with arbitrary forms of clustered autocorrelation. However, while Hagemann allows for heteroscedasticity, he does not consider the estimation of fixed effects. Standard errors are computed by resampling the QR estimators subgradient or first order conditions.

[14]Adding time-fixed-effects would have increased this to 183, which is substantially more than could be reliably estimated on a data set of our size. Furthermore, as some of our independent variables are national variables (such as the interest rate variables, which are the same across all states) the addition of time-fixed-effects would leave them unidentified.

[15]As we see below, this reduced, but did not elminate the serial correlation in the data. We take this into account in calculating appropriate standard errors for our estimated coefficients.

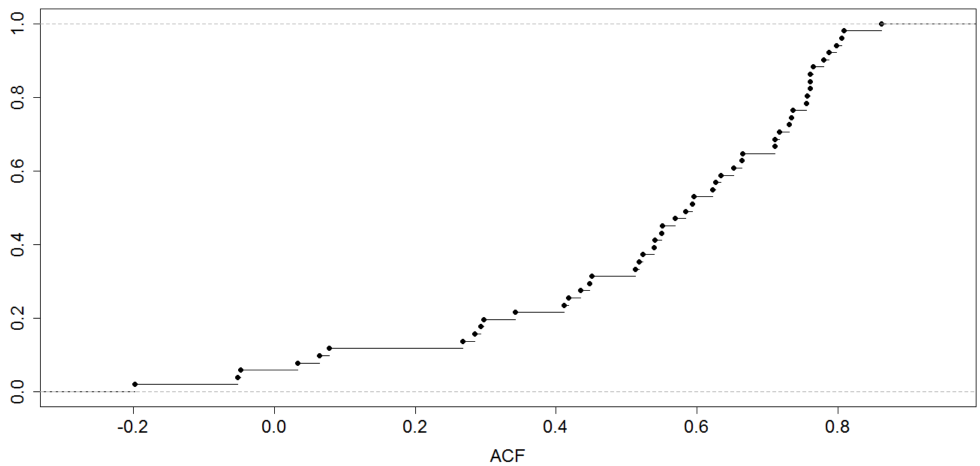


FIGURE 2. STATE-LEVEL AUTOCORRELATIONS

Note: The Figure shows is the cummulative distribution function (vertical axis) of the 51 estimated first-order autocorrelation coefficients (horizontal axis) for quarterly changes in state-level housing prices. Autocorrelation was substantial in many states, with most producing coefficients in the range of 0.4 to 0.8.

The QR estimates $(\alpha(\theta), \beta(\theta))$ solve the following first-order condition

$$(2) \qquad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j=1}^{T} \left[\theta - I\left(y_{i,j} - \alpha(\theta) - x'_{i,j} \cdot \beta(\theta) < 0\right)\right] \cdot x_{i,j} = 0$$

The wild gradient bootstrap process $W_N(\alpha(\theta), \beta(\theta), \theta)$ introduces state-level perturbations w_i such that
(3)

$$W_N(\alpha(\theta), \beta(\theta), \theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_i \cdot \sum_{j=1}^{T} \left[\theta - I\left(y_{i,j} - \alpha(\theta) - x'_{i,j} \cdot \beta(\theta) < 0\right)\right] \cdot x_{i,j},$$

where the w_i's are drawn from an i.i.d. random variable with $E[w_i] = 0$ and $E[|w_i|^q] < \infty$ for $q > 2$.[16] We now denote the QR objective function $M_N(\alpha(\theta), \beta(\theta), \theta)$, which can be expressed as:

$$(4) \qquad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j=1}^{T} \rho_\theta\left(y_{i,j} - \alpha(\theta) - x'_{i,j} \cdot \beta(\theta)\right)$$

[16]Hagemann (2017) suggests that w_i should be drawn from the Mammen (2012) 2-point distribution.

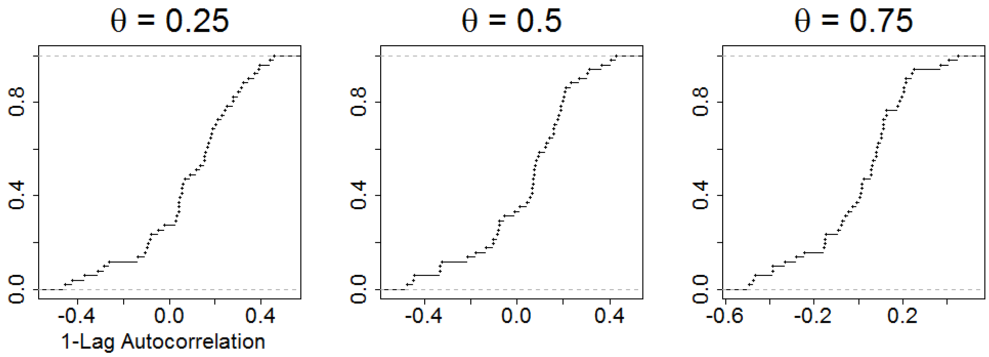


FIGURE 3. STATE-LEVEL RESIDUAL AUTOCORRELATIONS

Note: The figure shows is the cummulative distribution function (vertical axis) of the 51 estimated first-order autocorrelation coefficients (horizontal axis) for QR residuals with $\theta \in \{0.25, 0.50, 0.75\}$. Autocorrelation was substantial in many states, with most producing coefficients in the range of 0.4 to 0.8.

Each replication of the wild gradient bootstrap will then solve for $\left(\hat{\alpha}^*(\theta), \hat{\beta}^*(\theta)\right)$ that minimizes a new objective function M_N^*, expressed as
(5)

$$M_N^*(\alpha(\theta), \beta(\theta), \theta) = M_N(\alpha(\theta), \beta(\theta), \theta) + W_N(\alpha(\theta), \beta(\theta), \theta) \cdot \frac{(\alpha(\theta), \beta(\theta))}{\sqrt{N}}$$

The QR estimator's standard errors are subsequently estimated using the empirical distribution of the bootstrapped parameters. Hagemann (2017) presents Monte Carlo evidence that empirical critical values of standard errors offer a well-sized testing procedure.

ROBUST SCORE AND WALD TESTS. — Yoon and Galvao (2016) derive a robust covariance estimator for QR models with fixed effects, leading them to propose new Score and Wald test statistics. For both tests, statistics are asymptotically distributed as chi-squared distributions with a number of degrees of freedom equal to the number of restrictions imposed under H_0. While their Wald test allows for both autocorrelated and heteroscedastic errors, their Score test permits serially-correlated errors but restricts errors to be homoscedastic. Their Wald test also requires estimation of the conditional density of the error terms, making it less reliable in smaller samples when the density is difficult to estimate. They find that their Score test has slightly better size properties than their Wald test. However, the Score test does not provide standard errors and is instead simply used to estimate the clustered covariance matrix (CCM) of the coefficients' subgradient.[17]

The CCM is of the following form[18]

$$\Sigma = \Lambda^{-1} V \Lambda^{-1} \tag{6}$$

In a panel with N entities, T periods, and p covariates, let X_t be a $N \times p$ matrix of observations at period $t \in \{1, \ldots, T\}$ and $c_i = \frac{E[f_{i,t}(0|X_{it})X_{it}]}{E[f_{i,t}(0|X_{it})]}$, where $f_{i,t}(.)$ is the density function of the residuals.[19] The Λ term captures heteroscedasticity and is defined as:

$$\Lambda \equiv \operatorname*{plim}_{N\to\infty} \frac{1}{N} \cdot \sum_{i=1}^{N} E\left[f_{i,t}\left(0|X_{it}\right) \cdot X_{it} \cdot \left(X_{it} - c_i\right)' \right]. \tag{7}$$

Additionally, let

$$\varrho_{its} \equiv E\left[\left(I\left(e_{it}\left(\theta\right) \le 0\right) - \theta\right) \cdot \left(I\left(e_{it}\left(\theta\right) > 0\right) - \theta\right) | X_{i1}, \ldots, \ X_{iT}, \alpha_i\right]. \tag{8}$$

V is then defined as:

$$V \equiv lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} V_{N,i}^{0} + V_{N,i}^{1} \tag{9}$$

where

$$V_{N,i}^{0} = \frac{1}{T} \cdot \sum_{t=1}^{T} \theta \cdot (1-\theta) \cdot E\left[(X_{it} - c_i) \cdot (X_{it} - c_i)'\right] \tag{10}$$

$$\begin{aligned} V_{N,i}^{1} = \frac{1}{T} \cdot \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} & E\left[\varrho_{it,t-s} \cdot (X_{it} - c_i) \cdot (X_{it-s} - c_i)'\right] \\ & + E\left[\varrho_{it-s,t} \cdot (X_{it-s} - c_i) \cdot (X_{it} - c_i)'\right] \end{aligned} \tag{11}$$

The V matrix captures serial correlation through $V_{N,i}^{1}$. In practice, the autocorrelation structure is not always computed up to $T-1$ lags. Rather, we consider a truncation parameter $r_T \le T-1$ that determines how many lags $V_{N,i}^{1}$ will account for. Theoretical support for the selection of r_T is a problem that has yet to be addressed in time-series literature and Yoon and Galvao (2016) find that their

[17]See Yoon and Galvao (2016), pp. 13-15.
[18]Yoon and Galvao (2016), p. 10
[19]Details on the estimation of $f_{i,t}(.)$ are provided in the Appendix.

results are sometimes sensitive to it. They found that $r_T = max\left(2, \lfloor 1.2 \cdot T^{\frac{1}{3}} \rfloor\right)$ worked well for panel data of dimension similar to ours. We follow their suggestion, which means that r_T is equal to 5 in our application. Hence, our $\hat{V}$ considers time-dependence in errors that can be detected within 5 consecutive quarters.

Yoon and Galvao (2016) also present the $\hat{V}^m$ estimator, a modified version of $\hat{V}$ that restricts serial dependence structures to be identical across entities ($\varrho_{its} = \varrho_{ts}$).[20] They claim that the modified variance estimator is robust to the truncation parameter choice. We follow them and set r_T equal to $T-1$ so as to cover the entire time dependence structure in our sample.

Simulation Results. — We investigated the reliability of the Yoon and Galvao's proposed test statistics with two different simulation experiments, both of which draw on Bertrand, Duflo and Mullainathan (2004).[21] The first follows Yoon and Galvao (2016) by generating a process with fixed effects and homogeneous, serially-correlated error terms

$$\begin{aligned} y_{i,t} &= \alpha_i + X_{i,t} \cdot \beta + z_{i,t} \cdot \gamma + e_{i,t} \\ &\text{where} \\ \alpha_i &\sim U(0,1) \\ X_{i,t} &= 0.3 \cdot \alpha_i + \varepsilon_{i,t} \\ \varepsilon_{i,t} &\sim i.i.d.\chi^2(3) \\ e_{i,t} &= \rho \cdot e_{i,t-1} + v_{i,t} \\ v_{i,t} &\sim i.i.d.N(0, 1-\rho^2) \end{aligned} \tag{12}$$

$z_{i,t}$ is generated using the approach of Bertrand, Duflo and Mullainathan (2004), which we detail below. Note that all states are assumed to have the same error dependence structure (given by ρ) so that we are under the null hypothesis of the Wald test with $\hat{V^m}$ as well as that with $\hat{V}$. We therefore use these simulations to assess the size of the Score and Wald tests. We report empirical test sizes with first-order serial correlation $\rho \in \{0, 0.4, 0.8\}$. In each case, we added a spurious random variable to the system (as $z_{i,t}$) and tested $H_0 : \gamma = 0$.

Our other approach simply added the spurious $z_{i,t}$ to our housing market data set and tested the same H_0.

In both cases, we used the two-step approach of Bertrand, Duflo and Mullainathan (2004) to generate our spurious treatment variable $z_{i,t}$.

1) We randomly pick half of the N states in the panel and select a random period $t \mid 0.1 \cdot T < t < 0.9 \cdot T$.

[20] See Yoon and Galvao (2016), p. 12.
[21] Hagemann's test was not included as its critical values are themselves determined by simulation.

2) We construct a dummy variable = 1 only for periods $\geq t$ in the selected states.

Results for the two simulation approaches are shown in Table 2 for estimation at the three quartiles $\theta \in \{0.25, 0.50, 0.75\}$.

Table 2—Empirical Test Size

		$\alpha = 0.05$			$\alpha = 0.10$		
Simulation	**CCM test**	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$
EQ (12) $\rho = 0$	W	0.193	0.09	0.026	0.272	0.146	0.052
	W^m	0.067	0.069	0.06	0.123	0.122	0.108
	S	0.078	0.08	0.064	0.13	0.121	0.109
	S^m	0.057	0.069	0.089	0.11	0.114	0.129
EQ (12) $\rho = 0.4$	W	0.264	0.097	0.014	0.349	0.147	0.034
	W^m	0.083	0.055	0.07	0.142	0.108	0.118
	S	0.073	0.066	0.064	0.12	0.121	0.103
	S^m	0.068	0.061	0.073	0.114	0.123	0.134
EQ (12) $\rho = 0.8$	W	0.326	0.081	0.004	0.405	0.125	0.017
	W^m	0.142	0.107	0.139	0.209	0.179	0.214
	S	0.078	0.066	0.072	0.129	0.128	0.124
	S^m	0.078	0.076	0.08	0.127	0.144	0.145
Actual Data	W	0.584	0.699	0.389	0.637	0.761	0.488
	W^m	0.54	0.634	0.421	0.602	0.69	0.507
	S	0.605	0.632	0.403	0.665	0.68	0.481
	S^m	0.495	0.487	0.358	0.58	0.561	0.442

Note: Figures shown are empirical test sizes based on 1000 replications.
α indicates the significance level of the test.
θ indicates the quantile at which the test is conducted.
ρ indicates the degree of persistence used in (12).
S indicates Score test, W indicates Wald test, and superscript m indicates use of the restricted estimator.

Results using the Yoon and Galvao (2016) simulation approach show that increasing ρ increases size distortion much more rapidly for the Wald tests than for the Score-based tests. Differences across quantiles and test size were modest. However, tests using the Bertrand, Duflo and Mullainathan (2004) approach with the housing market data produced entirely different results, with all tests showing severe size distortion.[22] For that reason, we focus our analysis on the results using the Hagemann (2017) tests. However, because the latter does not allow for the

[22]We conjecture that reflects inter-state dependence in the error structure for which the CCM estimator fails to adjust. However, we leave this question for future research.

estimation of fixed effects, we also used the simulated Wald and Score statistics used to construct the bottom panel of Table 2 to construct size-adjusted critical values for our data set and used them as a check on robustness of our results to the inclusion of fixed effects.

Table 3—Impact on Changes in Housing Prices

		θ		
Covariates	**Parameter**	**0.25**	**0.5**	**0.75**
Δlog **HPI**$_{t-1}$	$\hat{\beta}$	0.41823**	0.43022**	0.46840**
	$\hat{\sigma_{\beta}}$	0.03257	0.0331	0.04868
Terms of Trade	$\hat{\beta}$	0.0001	0.00020**	-0.00001
	$\hat{\sigma_{\beta}}$	0.00007	0.00006	0.00009
Housing Vacancies	$\hat{\beta}$	-0.00299**	-0.00190**	-0.00130*
	$\hat{\sigma_{\beta}}$	0.00048	0.00044	0.00053
Rental Vacancies	$\hat{\beta}$	0	-0.00015	-0.00027**
	$\hat{\sigma_{\beta}}$	0.00008	0.00008	0.0001
Δlog **Real Personal Income**	$\hat{\beta}$	0.00006	-0.00022	-0.00001
	$\hat{\sigma_{\beta}}$	0.00018	0.00017	0.00014
Δlog **Real GDP**	$\hat{\beta}$	0.05988**	0.02178*	0.02328*
	$\hat{\sigma_{\beta}}$	0.00733	0.00973	0.01125
Unemployment	$\hat{\beta}$	-0.00028	0.00006	0.00044
	$\hat{\sigma_{\beta}}$	0.00015	0.00011	0.00024
Property Taxes	$\hat{\beta}$	-0.07945	0.01333	0.00677
	$\hat{\sigma_{\beta}}$	0.36475	0.28108	0.56047
30-year Mortgage Rate	$\hat{\beta}$	0.00004	0.00034**	0.00059**
	$\hat{\sigma_{\beta}}$	0.00012	0.00011	0.00015
5- to 1-year Treasury Spread	$\hat{\beta}$	0.00099**	0.00044*	0.00081**
	$\hat{\sigma_{\beta}}$	0.00022	0.00019	0.00028
Mortgage Debt / GDP	$\hat{\beta}$	-0.02875**	-0.01589**	-0.00600*
	$\hat{\sigma_{\beta}}$	0.00285	0.00251	0.00305
HPI Exuberance	$\hat{\beta}$	0.00124**	0.00095**	0.00081**
	$\hat{\sigma_{\beta}}$	0.0001	0.00007	0.00013
HPI/Income Exuberance	$\hat{\beta}$	-0.00013	0.00018	0.00056*
	$\hat{\sigma_{\beta}}$	0.00017	0.00014	0.00024

Note: $\hat{\beta}$ is the estimated coefficient for the model without fixed-effects, $\hat{\sigma_{\beta}}$ is its standard error based on the Hagemann (2017) bootstrap with 300 replications. ** (*) indicates significance at the 1% (5%) significance level.

IV. Results

A. Model Estimates

The estimated QR coefficient at each quartile are shown in Table 3 for the model without fixed effects. The significance of the parameters indicated with asterisks is based on the Hagemann (2017) with standard errors and p-values inferred from the wild gradient bootstrap procedure. Model parameters were also estimated at each decile; Figures 4 through 9 illustrate how some of the regression coefficients vary across quantiles as well as provide estimated 95% confidence intervals. Those figures also show the OLS estimates and their confidence intervals as a set of solid and dashed horizontal lines for comparison.

Table 3 shows that supply, demand financial and non-fundamental factors all appear to have a significant effect on house price changes, although not every variable we included is individually significant. Among supply factors, state-level Housing Vacancies appear to have a significant impact across all quartiles, with both the Terms of Trade and Rental Vacancies appearing to play a role at some quantiles. Among the demand factors, only national real GDP growth appears to play a significant role; none of the state-level variables do. All three of the (national) financial factors seem to have a significant influence, while among the non-fundamental factors, HPI rather HPI/Income Exuberance appears to be the most significant. In addition, the autoregressive component was highly significant. The estimated coefficients also had the anticipated sign in all cases except that of the Mortgage Interest Rate, where higher rates were associated with higher, not lower, house price growth.[23]

B. Evidence of Asymmetries

The quantile regression framework also reveals that the many of the factors we identify as statistically significant have asymmetric effects on the distribution of house price changes, as we can see from Figures 4 through 9. For example, Figure 4 shows that the House Vacancy Rate has only a small and statistically insignificant effect on the upper tail of the distribution, but the coefficient increases steadily as we towards the lower tail, ending with an estimated coefficient roughly four times as large. This suggests that high levels of home vacancies do little to restrain house price increases but can nonetheless significantly add to downside risk in house prices. We also see that OLS overestimates the impact of home vacancies on the centre of the distribution.

[23]Results from the model with fixed effects and size-adjusted critical values for the Yoon and Galvao (2016) score tests tended to produce qualitatively similar coefficient estimates but found less evidence of statistical significance. In particular, none of the demand factors appeared to be significant, and Rental Vacancies were significant at some quantiles while House Vacancies were not. However, the Treasury Spread, the Mortgage Debt/GDP and the HPI Exuberance still appeared to be statistically significant at all three quartiles.

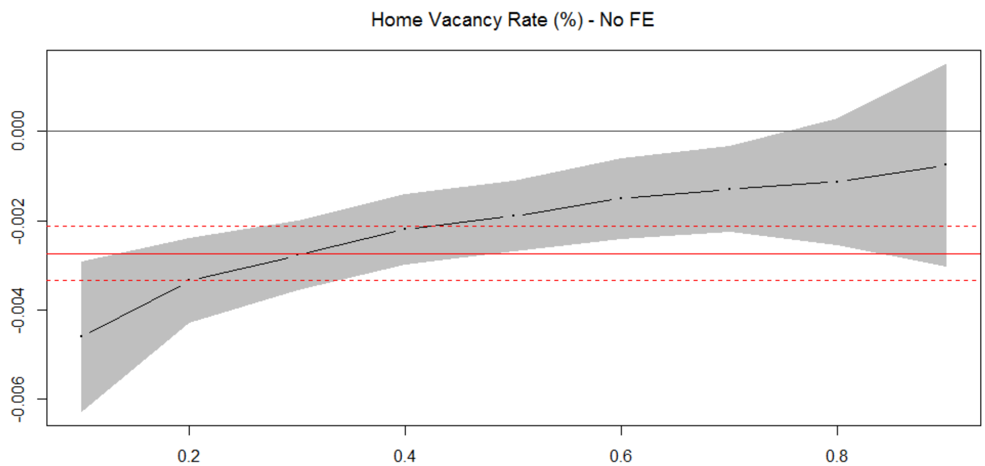


Figure 4. Estimated Coefficients by Quantile: Home Vacancy Rate

Note: The dashed-dotted line compares the estimated QR coefficients for the Home Vacancy rate, together with their 95% confidence intervals, to their OLS counterparts (shown as horizontal lines.) QR estimates are for $\theta \in \{0.10, 0.20, \ldots 0.90\}$.

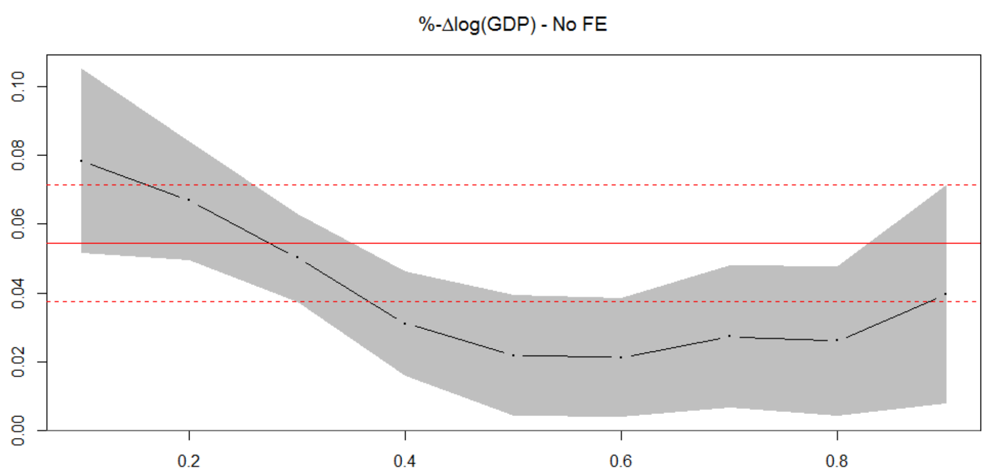


Figure 5. Estimated Coefficients by Quantile: Real GDP Growth

Note: The dashed-dotted line compares the estimated QR coefficients for real GDP growth, together with their 95% confidence intervals, to their OLS counterparts (shown as horizontal lines.) QR estimates are for $\theta \in \{0.10, 0.20, \ldots 0.90\}$.

Figure 5 shows that OLS also overestimates the impact of real GDP growth on median house price changes. While there is no significant variation in the QR coefficient estimates over the upper half of the distribution, GDP growth appears to have a much more important effect (roughly four times larger) on the lower tail. Results indicate that a 1% increase in GDP would drive the lower decile up by approximately 0.08, but upper quantiles would only increase by 0.02 to 0.04. Accordingly, when GDP increases, the conditional distribution of house prices moves slightly up and tightens as the lower quantiles are driven up further. This means that GDP growth reduces the magnitude of house price downturns in the next quarter by pushing up the lower quantiles. Conversely, a drop in GDP reduces most conditional quantiles and stretches the lower region of the distribution downward. Variations in GDP growth therefore cause a combination of location and scale shifts.

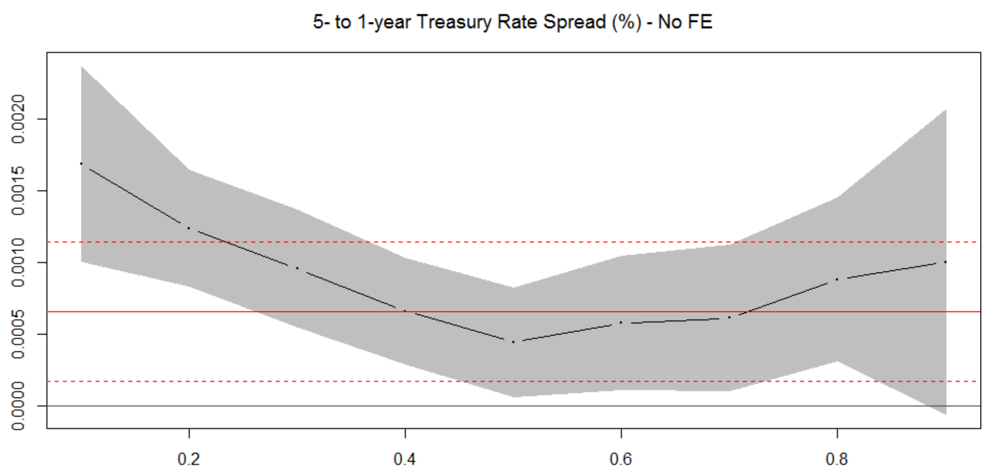


FIGURE 6. ESTIMATED COEFFICIENTS BY QUANTILE: TREASURY SPREAD

Note: The dashed-dotted line compares the estimated QR coefficients for the Treasury Spread, together with their 95% confidence intervals, to their OLS counterparts (shown as horizontal lines.) QR estimates are for $\theta \in \{0.10, 0.20, \ldots 0.90\}$.

All three of the significant financial factors also appear to have asymmetric effects on the distribution of house price changes. The effect of outstanding Mortgage Debt (8) attenuates steadily as we move from the lower to the upper tail of the distribution.[24] To give some idea of its economic importance, the fall in the mortgage debt ratio from just over 100% during the 2007-2009 financial crisis to its low of 69% in 2014 would have raised the first decile of quarterly house

[24]This is consistent with Jorda, Schularick and Taylor (2016), who find that increasing mortgage debt aggravates the risk of a financial crash in developed economies.

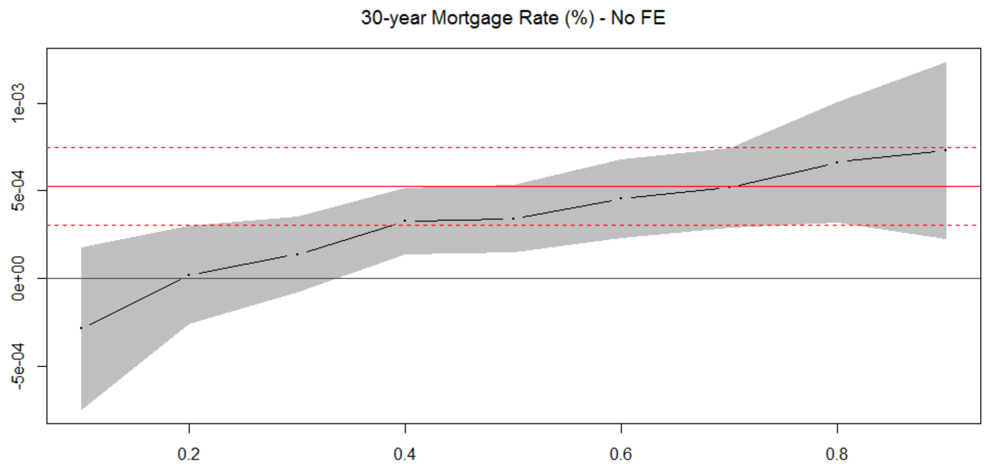


Figure 7. Estimated Coefficients by Quantile: Mortgage Rate

Note: The dashed-dotted line compares the estimated QR coefficients for the 30-Year Mortgage rate, together with their 95% confidence intervals, to their OLS counterparts (shown as horizontal lines.) QR estimates are for $\theta \in \{0.10, 0.20, \ldots 0.90\}$.

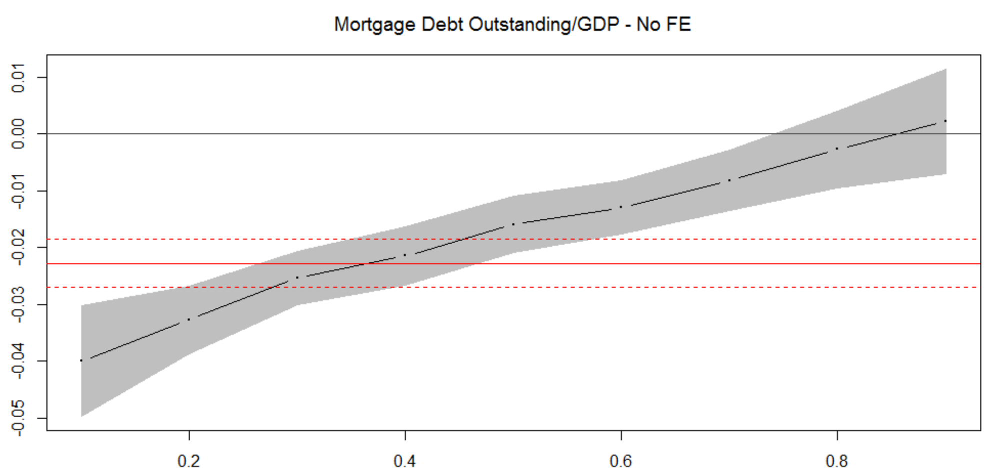


Figure 8. Estimated Coefficients by Quantile: Mortgage Debt

Note: The dashed-dotted line compares the estimated QR coefficients for the Mortgage Debt / GDP ratio, together with their 95% confidence intervals, to their OLS counterparts (shown as horizontal lines.) QR estimates are for $\theta \in \{0.10, 0.20, \ldots 0.90\}$.

price changes from by about 1.25 per cent while leaving the top decile effectively unchanged. Conversely, the Treasury Spread appears to have the largest effect at the lower tail of the distribution (see Figure 6.) Figure 7 shows that effect of changes in the Mortgage yield has approximately no effect on the lower tail of house price distribution, but that higher rates (counterintuitively) tend to raise the centre and upper tail of the distribution.

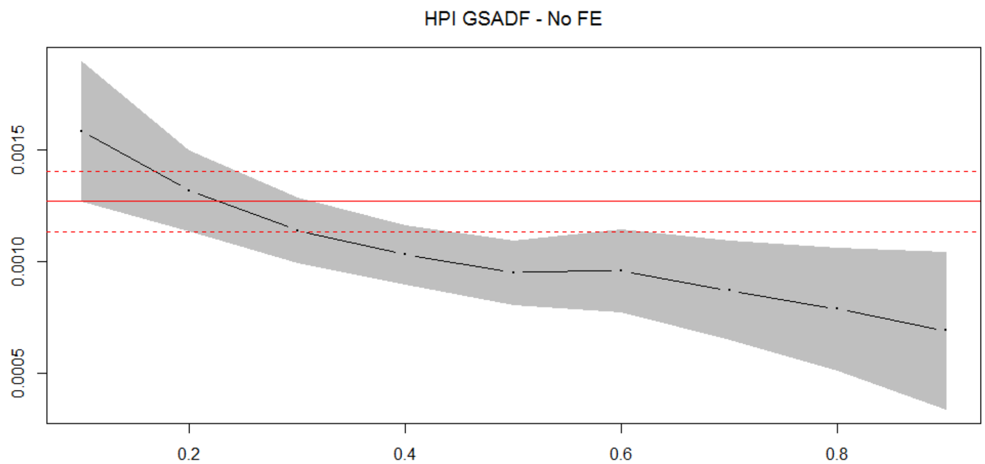


Figure 9. Estimated Coefficients by Quantile: House Price Exuberance

Note: The dashed-dotted line compares the estimated QR coefficients for the House Price Exuberance index, together with their 95% confidence intervals, to their OLS counterparts (shown as horizontal lines.) QR estimates are for $\theta \in \{0.10, 0.20, \ldots 0.90\}$.

Finally, Figure 9 shows that high market "exuberance", far from making markets more volatile, actually decrease downside risk by raising the lower quantiles of the house price change distribution.[25] This is particularly surprising as the exuberance assigns positive values to cases where prices are above or below fundamentals. This implies that increased exuberance tends to reduce risk (in the form of quarterly price changes) in the housing market. However, this is consistent with the interpretation of Pavlidis et al. (2016), who argue that unsustainable prices tend to be captured by the HPI/Income exuberance, while the exuberance statistics on the HPI itself capture market optimism.

C. *Implications for State-Level Risks*

To show the model's implications for the behaviour of house price risk, Figures 10 through 11 show how the fitted quantiles vary through time for selected states,

[25]When fixed effects are added, the coefficients appear more homogeneous across quantiles and closer to the OLS coefficient.

comparing them with the historical data as well as with quantiles implied from the same model when estimated by OLS and assuming normally distributed errors. Figure 10 shows results for California, with QR results for the 10, 25, 50, 75 and 90% quantiles in the upper panel, and the 10, 50 and 90% quantiles for OLS in the lower panel. In both cases, historical outcomes are shown by a dashed black line. The California market suffered sharp downturns in the late 1980s as well as from 2005 to 2009. While the QR model produces slightly tighter confidence bands than the OLS model (particularly near the peak of the market in the late 1980s), the differences are not especially striking, and QR quantiles are fairly symmetrically distributed around the median, suggesting that the model largely captures a location shifting over time.

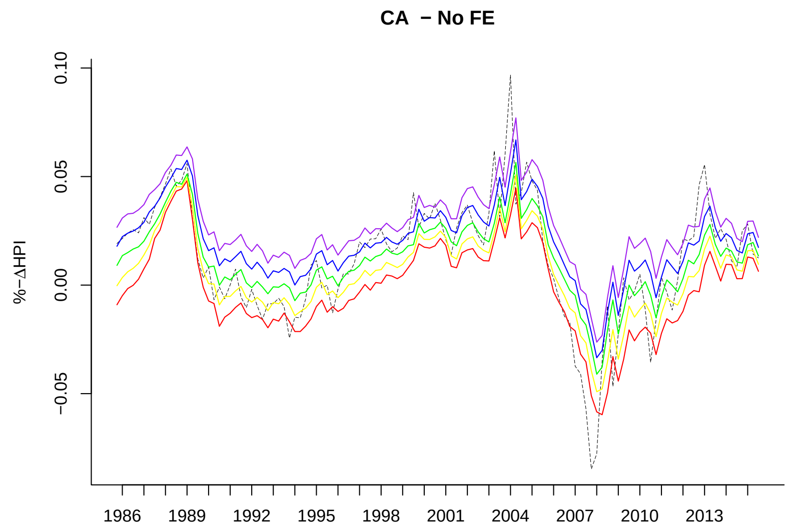
A different picture emerges for Ohio in Figure 11. The housing market in Ohio was generally stable until entering a decline shortly after 2000 which accelerated during the national market downtown and subsequently recovered. Here the QR model produces an upper 90% quantile that rarely dips below 2% annual growth while the lower 10% quantile ranges from almost 1% to -4%, implying wide variations in the riskiness of the housing market. Here the QR models captures both location and scale effects.

While Figures 10 and 11 show the degree of both location and scale shifts in the implied conditional distribution of house price changes, Figures 12 and 13 highlight the degree of asymmetry in these distribution and its variation over time. Each figure provides two complementary measures of asymmetry; the relative distance from the median to the quartile (green line) and that from the quartile to the outer decile (blue line.) A symmetric distribution would have a value of one; values less (greater) than one imply that upside (downside) risks are relatively larger. The green line captures asymmetries within the interquartile range while the blue line captures asymmetries outside this range.

Figure 12 shows that the relative degree of asymmetry shows substantial variation, briefly dipping well below 0.5 in the later part of the 1980s and briefly peaking around 2.0 following the 2008 financial crisis. Given that a value of 0.5 (2.0) for the green line implies that equally probable changes are half (twice) as large on the downside as on the upside, the implied asymmetries are substantial. We also see that the green and blue lines move closely together, so that changes in the asymmetries of small risks are largely mirrored by asymmetries in larger risks. Figure 13 shows that this is not always the case, however. Around the turn of the century we see a sharp increase in the green line while the blue line remains at low levels. This implies that while the odds of some small losses increased sharply relative to those of small gains, the odds of some large gains remained much greater than those of large losses.

V. Conclusions

We explored a new approach to modeling aggregate risk in housing markets using a panel quantile regression framework. The results imply that a mixture



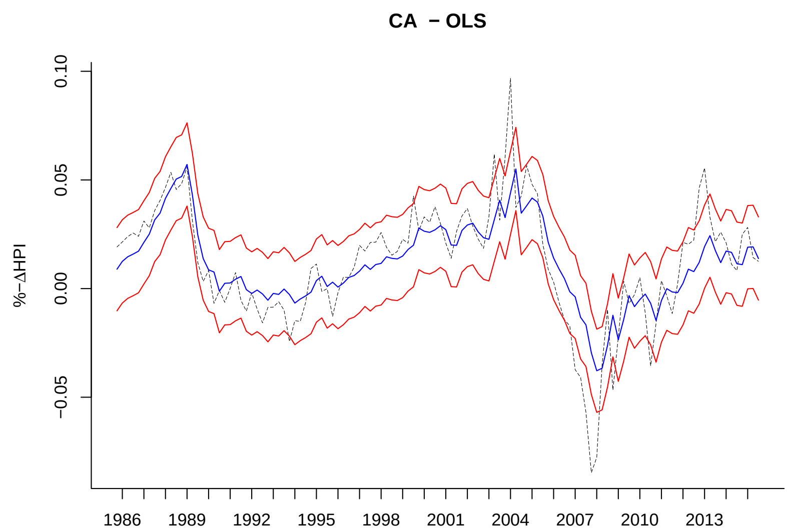


FIGURE 10. IMPLIED QUANTILES: CALIFORNIA

Note: The upper and lower panels compare the implied quantiles for the change in house prices from the QR model (upper panel) and OLS (lower panel) to historical data (black dashed line.) QR estimates are for $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ OLS estimates are for $\theta \in \{0.10, 0.50, 0.90\}$ under the assumption that errors are normally distributed.

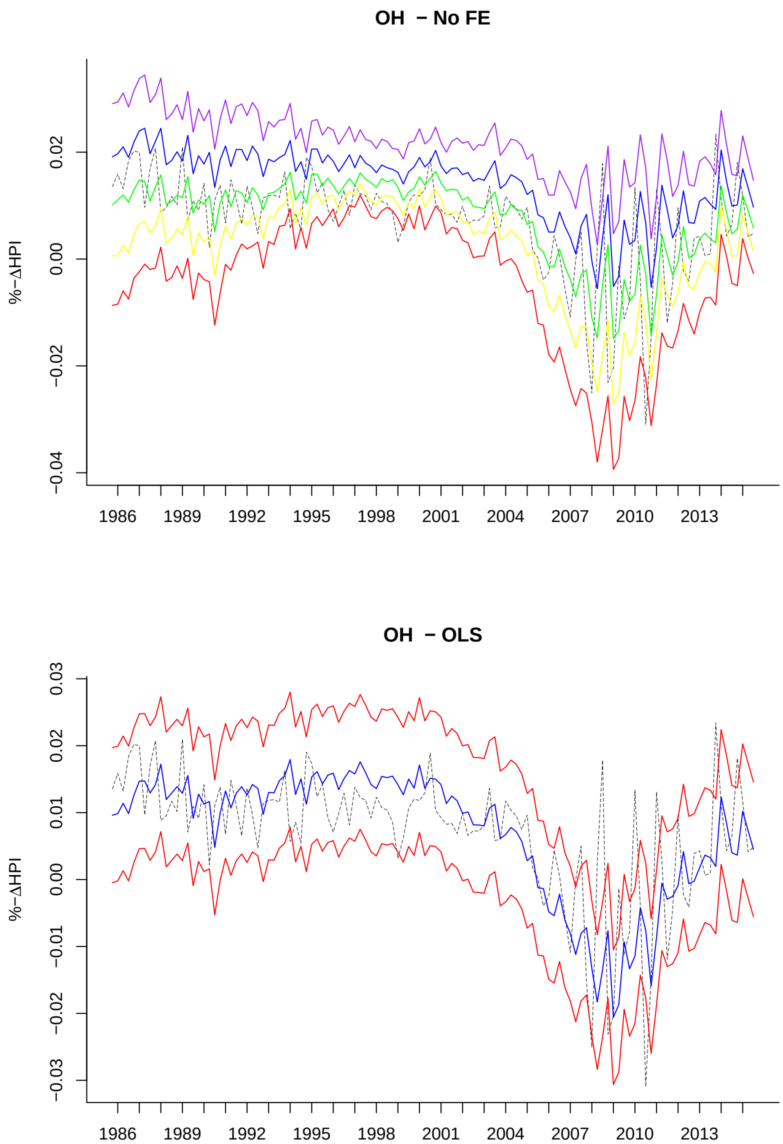


Figure 11. Implied Quantiles: Ohio

Note: The upper and lower panels compare the implied quantiles for the change in house prices from the QR model (upper panel) and OLS (lower panel) to historical data (black dashed line.) QR estimates are for $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ OLS estimates are for $\theta \in \{0.10, 0.50, 0.90\}$ under the assumption that errors are normally distributed.

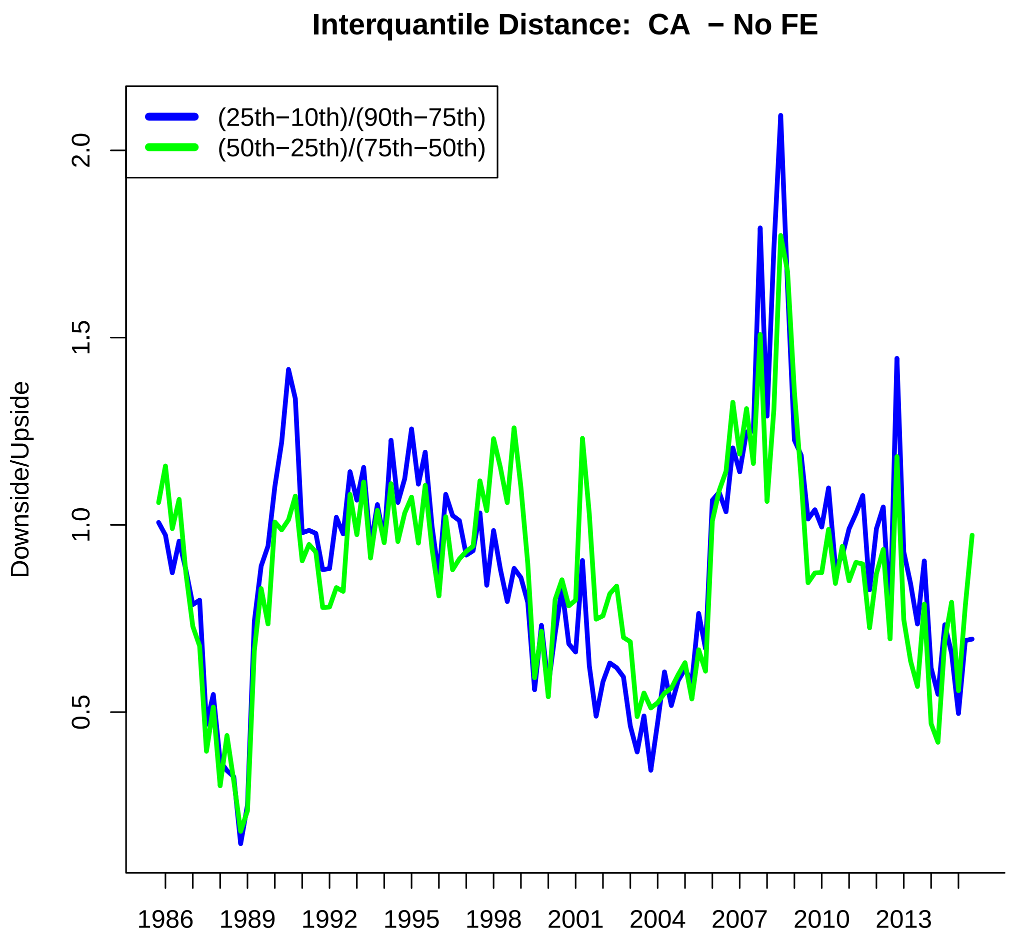


Figure 12. ITime-Varying Asymmetries: California

Note: The figure shows two simple measures of the degree of the asymmetry implied by the estimated quantiles. Each is the ratio of an inter-quantile distance in the lower half of the distribution to distance between the corresponding quantiles in the upper half. Values greater (less) than one imply that the corresponding portion of the conditional distribution is skewed towards lower (upper) tail.

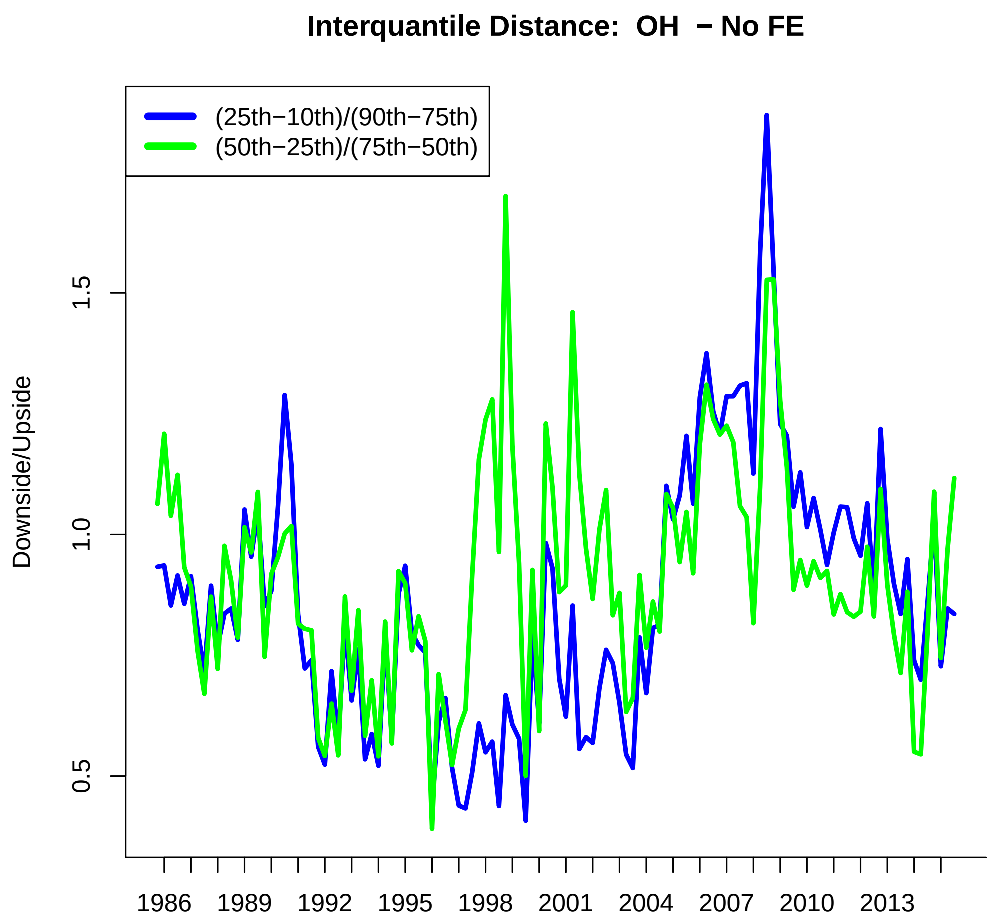


Figure 13. ITime-Varying Asymmetries: Ohio

Note: The figure shows two simple measures of the degree of the asymmetry implied by the estimated quantiles. Each is the ratio of an inter-quantile distance in the lower half of the distribution to distance between the corresponding quantiles in the upper half. Values greater (less) than one imply that the corresponding portion of the conditional distribution is skewed towards lower (upper) tail.

of supply, demand, financial and non-fundamental factors play significant roles in influencing the evolution of state-level US house prices. In several cases, they also show evidence of important asymmetries, with several factors (such as the Home Vacancy Rate or the level of Mortgage Debt) having a much greater impact on the lower quantiles, while others (notably the Mortgage Interest Rate) primarily affect the upper quantiles. We also show that implications of the model vary across states, with the quantile regression largely capturing location shifts for some states while for others it captures a mix of location and scale shifts.

REFERENCES

Baker, Dean. 2011. "The Menace of an Unchecked Housing Bubble." *The Economists' Voice: Top Economists Take On Today's Problems*, 288–295. Columbia University Press.

Bauer, Gregory H. 2014. "International House Price Cycles, Monetary Policy and Risk Premiums." Bank of Canada Staff Working Paper.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249–275.

Calhoun, Charles A. 1996. "OFHEO House Price Indexes: HPI Technical Description." Federal Housing Finance Agency.

Case, Karl E, and Robert J Shiller. 2003. "Is There a Bubble in the Housing Market?" *Brookings Papers on Economic Activity*, 2003(2): 299–342.

Chen, Li, Lee-Jen Wei, and Michael I Parzen. 2004. "Quantile regression for correlated observations." 51–69, Springer.

Corrigan, Paul. 2017. "Terms-of-Trade and House Price Fluctuations: A Cross-Country Study." Bank of Canada Staff Working Paper.

Dantzig, George B. 1963. "Linear Programming and Extensions. Princeton landmarks in mathematics and physics."

Davino, Cristina, Marilena Furno, and Domenico Vistocco. 2014. *Quantile Regression: Theory and Applications. Wiley Series in Probability and Statistics*, Oxford:John Wiley & Sons.

Glaeser, Edward L. 2013. "A Nation of Gamblers: Real Estate Speculation and American History." *American Economic Review*, 103(3): 1–42.

Glaeser, Edward L, and Charles G Nathanson. 2017. "An Extrapolative Model of House Price Dynamics." *Journal of Financial Economics.*

Glaeser, Edward L, Joseph Gyourko, and Albert Saiz. 2008. "Housing Supply and Housing Bubbles." *Journal of Urban Economics*, 64(2): 198–217.

Hagemann, Andreas. 2017. "Cluster-Robust Bootstrap Inference in Quantile Regression Models." *Journal of the American Statistical Association*, 112(517): 446–456.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. "Assessing high house prices: Bubbles, fundamentals and misperceptions." *The Journal of Economic Perspectives*, 19(4): 67–92.

Ho, Anson TY, Kim P Huynh, and David T JachoChavez. 2016. "Flexible Estimation of Copulas: An Application to the U.S. Housing Crisis." *Journal of Applied Econometrics*, 31(3): 603–610.

Jorda, Oscar, Moritz Schularick, and Alan M Taylor. 2016. "The Great Mortgaging: Housing Finance, Crises and Business Cycles." *Economic Policy*, 31(85): 107–152.

Jorda, Oscar, Moritz Schularick, and Alan M Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual*, 31(1): 213–263.

Kahn, James A. 2009. "Productivity Swings and Housing Prices." *Current Issues in Economics and Finance*, 15(July 3).

Knoll, Katharina, Moritz Schularick, and Thomas Steger. 2017. "No Price Like Home: Global House Prices, 18702012." *The American Economic Review*, 107(2): 331–353.

Koenker, Roger, and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica: Journal of the Econometric Society*, 46(1): 33–50.

Laeven, Luc, and Fabian Valencia. 2008. "Systemic banking crises: a new database." International Monetary Fund IMF Working Papers.

Mammen, Enno. 2012. *When Does Bootstrap Work?: Asymptotic Results and Simulations.* Vol. 77 of *Lecture Notes in Statistics*, New York:Springer Science & Business Media.

McQuinn, Kieran, and Gerard O'Reilly. 2008. "Assessing the role of income and interest rates in determining house prices." *Economic modelling*, 25(3): 377–390.

Monnet, Eric, and Clara Wolf. 2017. "Is demographics the housing cycle?" *Rue de la Banque*, , (41).

Pavlidis, Efthymios, Alisa Yusupova, Ivan Paya, David Peel, Enrique Martnez-Garca, Adrienne Mack, and Valerie Grossman. 2016. "Episodes of exuberance in housing markets: in search of the smoking gun." *The Journal of Real Estate Finance and Economics*, 53(4): 419–449.

Phillips, Peter CB, Shuping Shi, and Jun Yu. 2015. "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500." *International Economic Review*, 56(4): 1043–1078.

Portnoy, Stephen, and Roger Koenker. 1997. "The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators." *Statistical Science*, 12(4): 279–300.

Powell, David. 2016. "Quantile treatment effects in the presence of covariates." Available on BEPress.

Reinhart, Carmen M, and Kenneth S Rogoff. 2009. *This time is different: Eight centuries of financial folly.* Princeton, New Jersey:Princeton University Press.

Roche, Maurice J. 2001. "The rise in house prices in Dublin: bubble, fad or just fundamentals." *Economic Modelling*, 18(2): 281–295.

Rünstler, Gerhard, and Marente Vlekke. 2016. "Business, Housing and Credit Cycles." European Central Bank Working Paper.

Schularick, Moritz, and Alan M Taylor. 2012. "Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008." *The American Economic Review*, 102(2): 1029–1061.

Silverman, Bernard W. 1986. *Density estimation for statistics and data analysis.* London:Chapman & Hall.

Yoon, Jungmo, and Antonio F Galvao. 2016. "Robust Inference for Panel Quantile Regression Models with Individual Fixed Effects and Serial Correlation." November 30, 2016. Available at SSRN.

Zimmer, David M. 2012. "The role of copulas in the housing crisis." *Review of Economics and Statistics*, 94(2): 607–620.

The BSADF Test

The Backward Supremum ADF (BSADF) test is based on the Augmented Dickey-Fuller (ADF) unit-root test, which is in turn based on the regression:

$$\Delta y_t = a_{r_1,r_2} + \beta_{r_1,r_2} \cdot y_{t-1} + \sum_{j=1}^{k} \psi^j_{r_1,r_2} \Delta y_{t-j} + \varepsilon_t \tag{A1}$$

where y_t is a univariate time series, k denotes the number of auto-regressive lags in the model, and ε_t is an i.i.d. normally-distributed error term with standard deviation σ_{r_1,r_2}. The interval $[r_1, r_2]$ where $r_1, r_2 \in [0, 1]$ and $r_1 < r_2$ designates the portion of the sample used to calculate the ADF statistic, so with a sample with periods ranging from 0 to T, the $ADF^{r_2=m/T}_{r_1=n/T}$ statistic is based on a subset of periods ranging from n to m inclusively, where $n, m \in \{0, \ldots, T\}$ and $n < m$. The ADF test statistic is defined as:

$$ADF^{r_2}_{r_1} = \frac{\hat{\beta}_{r_1,r_2}}{\sigma_{\hat{\beta},r_1,r_2}} \tag{A2}$$

The Supremum ADF (SADF) (Phillips and Yu, 2011) is then defined as:

$$SADF\,(r_0) = \sup_{r_2 \in [r_0,1]} ADF^{r_2}_0 \tag{A3}$$

One can see that the SADF is calculated recursively with an expanding sample of periods with a minimum window size of r_0, while keeping the starting period fixed at $r_1 = 0$. The SADF is suited to detect a single period of unit-root behaviour in the sample (Phillips, Shi and Yu, 2015). In order to measure multiple episodes of explosiveness at a given period, we consider the BSADF test statistic :

$$BSADF_{r_2}\,(r_0) = sup_{r_1 \in [0,r_2-r_0]} ADF^{r_2}_{r_1}, \tag{A4}$$

where r_0 denotes the minimal window size. This procedure can produce real-time exuberance levels by setting r_2 as the present period (the t-th period corresponds to $r_2 = \frac{t}{T}$), and letting the start of our estimation period r_1 vary from the beginning of our sample (0) to $r_2 - r_0$. We can then recursively create a series of BSADF statistics for both HPI and HPI-to-income (See Figure 11 for HPI BSADF and Figure 12 for HPI-to-income BSADF in Annex 1).

Kernel Density Estimation

The CCM procedure requires state-specific error density estimates $f_i(.)$, which we estimate with the expression

$$\hat{f}_i(x) = \frac{1}{Th} \cdot \sum_{t=1}^{T} K\left(\frac{\hat{e}_{it} - x}{h}\right), \tag{B1}$$

where T is the number of periods in our sample. h and $K(.)$ are the user-chosen bandwidth and kernel. We will use the Gaussian kernel, such that $K(.)$ is the normal density function. Bandwidth validation is an issue that has not been resolved in the QR literature. The Silverman Rule of Thumb bandwidth, although *ad hoc*[26], is still widely used as a consistent bandwidth selection. For a given set of QR residuals across the panel e_{it} , we can compute the state-level Silverman bandwidths h_{1i} as

$$h_{1i} = \left(\frac{4\hat{\sigma}^5_{e,i}}{3n}\right) \tag{B2}$$

where $\sigma_{e,i}$ denotes the sample standard deviation of errors for state i. This bandwidth selection issue is of less concern for our purposes because the calculation of the Wald test does not require the whole density to be estimated. Rather, we estimate the density at 0 with $f_h(0)$. As suggested in Yoon and Galvao (2016), density estimates used for the calculation of the Λ term of the CCM estimator are conducted with h_2

$$h_{2i} = min\left(\hat{\sigma}_{e,i}\ ;\frac{IQR}{1.34}\right) \tag{B3}$$

where σ_e and IQR denote the QR residuals standard error and their inter-quartile range. This bandwidth is usually larger than the Silverman bandwidth, which leads to smoother densities.

Quantile Regression

Instead of minimizing the sum of squared residuals as in OLS, the QR linear programming problem for estimating the θ-th unconditional quantile q_θ exploits

[26]This method has been widely used since it was introduced by Silverman (1986). In fact, we will use the Silverman bandwidth in some initial steps of the computation of the Wald statistic. In cases where the error distribution is multimodal or asymmetric, the Silvermans rule of thumb is said to over-smooth the mode of the distribution causing a noisier estimate in the tails (Davino, Furno and Vistocco (2014)).

an asymmetric loss function $\rho_\theta(.)$ of residuals ε

$$\rho_\theta(\varepsilon) = [\theta - I(\varepsilon < 0)] \cdot \varepsilon \tag{C1}$$

where $I(.)$ is the indicator function. More generally, the loss function of the θ-th QR uses a similarly asymmetric weighting of absolute residuals (negative errors take weights of $1-\theta$ and positive errors take weights of θ). Let $\varepsilon = y - \alpha(\theta) - X \cdot \beta(\theta)$

$$\rho_\theta(\varepsilon) = [\theta - I(\varepsilon < 0)] \cdot \varepsilon = [(1-\theta) \cdot I(\varepsilon \leq 0) + \theta \cdot I(\varepsilon > 0)] \cdot |\varepsilon|. \tag{C2}$$

The conditional quantile function's coefficients $\{\alpha(\theta), \beta(\theta)\}$ are the solution to the following optimization problem:

(C3)

$$\begin{aligned}\{\alpha(\theta), \beta(\theta)\} &= argmin_{\alpha(\theta),\beta(\theta)}\ E\left[\rho_\theta(y - \alpha(\theta) - X \cdot \beta(\theta))\right] \\ &= argmin_{\alpha(\theta),\beta(\theta)} \int_{y \in R} \rho_\theta(y - \alpha(\theta) - X \cdot \beta(\theta))\ dF_Y(y) \\ &= argmin_{\alpha(\theta),\beta(\theta)} \Bigg((1-\theta) \cdot \int_{y<\alpha - X\cdot\beta} |y - \alpha(\theta) - X \cdot \beta(\theta)|\ dF_Y(y) \\ &\quad + \theta \cdot \int_{y>\alpha - X\cdot\beta} |y - \alpha(\theta) - X \cdot \beta(\theta)|\ dF_Y(y) \Bigg)\end{aligned}$$

To derive the first order condition of this minimization problem, let $\hat{q}_\theta = \alpha(\theta) + X\beta(\theta)$.

$$\begin{aligned} 0 &= \frac{d}{d\hat{q}_\theta}\ E\left[\rho_\theta(y - \hat{q}_\theta)\right] \\ &= \frac{d}{d\hat{q}_\theta}\left((1-\theta)\int_{y<\hat{q}_\theta} |y - \hat{q}_\theta| dF_Y(y) + \theta \int_{y>\hat{q}_\theta} |y - \hat{q}_\theta| dF_Y(y)\right) \\ &= \frac{d}{d\hat{q}_\theta}\left((1-\theta)\int_{-\infty}^{\hat{q}_\theta} (\hat{q}_\theta - y)\, dF_Y(y) + \theta \int_{\hat{q}_\theta}^{+\infty} (y - \hat{q}_\theta)\, dF_Y(y)\right) \\ &= (1-\theta)\int_{-\infty}^{\hat{q}_\theta} dF_Y(y) - \theta \int_{\hat{q}_\theta}^{+\infty} dF_Y(y) \\ &= (1-\theta)\, F(\hat{q}_\theta) - \theta\,(1 - F(\hat{q}_\theta)) \\ &\iff F(\hat{q}_\theta) = \theta \end{aligned} \tag{C4}$$

The first order condition leads to $F(\alpha(\theta) + X \cdot \beta(\theta)) = \theta$ where F is the CDF of the response variable y. It follows that the expected value of the θ-th quantile of the error ε is 0.

The coefficient estimates are the solution to the following optimization problem

(C5)

$$\left(\hat{\alpha}\left(\theta\right),\hat{\beta}\left(\theta\right)\right) = argmin_{\alpha(\theta),\beta(\theta)} \left\{ (1-\theta)\sum_{y\leq 0}\left|y_i - \alpha\left(\theta\right) - X_i\cdot\beta\left(\theta\right)\right|\cdot f\left(y\right) + \theta\sum_{y>0}\left|y_i - \alpha\left(\theta\right) - X_i\cdot\beta\left(\theta\right)\right|\cdot f\left(y\right)\right\}$$

which, for a given θ, is a convex minimization problem that can effectively be solved by the iterative simplex method proposed by Dantzig (1963). This well-known computing algorithm is easily implemented and works well with medium-sized samples.[27]

[27]The simplex method stands as a time-costly option for large problems of over 100,000 observations. The QR version of the Frisch-Newton interior-point algorithm introduced by Portnoy and Koenker (1997) is better suited to accelerate the procedure with samples of this size (Davino, Furno and Vistocco (2014)).