A large-dimensional structural factor analysis of Californian house prices*

VERY PRELIMINARY. PLEASE DO NOT QUOTE.

Lasse Bork[†]
Aalborg University

Stig Vinther Møller[‡]
CREATES, Aarhus University

September 2014

Abstract: In a large-dimensional dynamic factor analysis of Californian house prices, we decompose the metro-level house price variation into aggregate structural shocks and regional structural shocks. For the structural explanation of the house price variation, we estimate that aggregate supply and aggregate demand shocks play a bigger role than monetary policy shocks. However, the most important shock is the regional housing demand shock. We identify the factors by loading restrictions while the structural shocks are identified using a combination of zero and sign restrictions.

JEL Classification: C32; E32; E52; R11; R21; R31.

Key words: Housing; Monetary Policy; SVAR; housing supply and demand shocks; EM algorithm.

*Møller acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

[†]Department of Business and Management, Aalborg University, Fibigerstraede 2, 9220 Aalborg, Denmark. E-mail: bork@business.aau.dk.

[‡]CREATES, Department of Economics and Business, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Email: svm@asb.dk.

1 Introduction

This paper analyzes the role of standard economic shocks for the explanation of Californian house prices in a data-rich dynamic factor model. Specifically, we analyze to which extent standard aggregate shocks like monetary policy, aggregate demand, and aggregate supply as well as regional shocks like housing supply and demand, are able to account for the variation of Californian metro-level house prices. Any remaining unexplained variation is idiosyncratic so that we can decompose the sources of variation of Californian metro-level house prices into fundamental economic shocks and idiosyncratic shocks.

Decomposing the variation in Californian house prices into structural economic shocks and idiosyncratic shocks is interesting for a number of reasons. Compared to other states California is particularly challenging because house prices have been among the fastest growing during the last two decades with a big increase during the boom and a big decrease during the bust and identifying the main economic drivers is therefore interesting. Even within the state of California the metropolitan house price movements are quite different; see Figure 2 and 3. We attempt to identify potential economic drivers using a structural decomposition of the price variation which enables us to answer questions like: Are the Californian house prices mainly driven by local idiosyncratic shocks or are they mainly a result of the macroeconomic development? How much of the movements in the Californian house price can be attributed to regional housing demand and supply shocks? How sensitive are Californian house prices to US aggregate demand shocks and monetary policy shocks? Are there any differences among the metropolitan areas?

Our approach to the rather complex task of explaining Californian house prices builds on a large-dimensional dynamic factor model with almost 200 economic and financial time series that broadly cover the US aggregate economy and 75 Californian time series including metro-level house prices. As such, the Californian metropolitan house prices are potentially allowed to be a function of a lot of information; at the aggregate level as well at the state-level. The way this is made econometrically feasible, builds on the fact that many economic and financial time series comove, so that the panel of observed variables obey an approximate low-dimensional factor structure. So, essentially we model all the included aggregate and regional observed variables including house prices as a linear function of a few dynamic latent factors. Accordingly, a handful or more of dynamic factors drives the observed variables over time and any innovation in one of the factors can be traced back to the observed variables, for instance house prices in the San Diego metropolitan area, through the factor sensitivity (factor loading). Our approach to the identification of the structural shocks is through an economic identification of the factors so that the factors can be given an economic interpretation, for instance an inflation factor, an employment factor or a regional building permits factor.[1] This is done by identifying loading restrictions. Moreover, we also want to distinguish aggregate factors from regional factors and here loading restrictions accomplish this as detailed later. Finally, the structural identification of these economic interpretable factors follows recent standard practice in the structural VAR (SVAR) literature using a combination of zero restrictions and sign restrictions.

Now, in order to discuss structural shocks in a meaningful way we need to identify the factors so that they can be given an economic interpretation, for instance an inflation factor, an employment factor or a regional building permits factor. This is done by

[1] An alternative approach is through sign restrictions on the observed variables.

identifying loading restrictions. Moreover, we also want to distinguish aggregate factors from regional factors and also here loading restrictions accomplish this as detailed later. Finally, the structural identification of these economic interpretable factors follows recent standard practice in the SVAR literature using a combination of zero restrictions and sign restrictions.

1.1 Related literature

[consider to start with: What distinguishes this paper ...] Our paper is related to a number of papers that deals with the structural analysis of house prices; either of national house prices or regional house prices using either low-dimensional VARs or high-dimensional dynamic factor models. In a first step, Del Negro & Otrok (2007) estimate a common national house price factor together with a state-specific house price factor for each of the 48 contiguous states using a panel of state-level house prices that ends in 2005. In a second step, the national house price factor is added to a standard low-dimensional VAR with national economic variables and the response to a monetary policy shock can then be traced back to a particular state-specific house price via its loading on the national house price factor. They find that historical movements in house prices are mainly driven by local factors, although the national factor seems to play a bigger role during the 2001-2005 period. Furthermore, monetary policy shocks seem to play a small but nonnegligible role during the boom in house prices. Although we focus on the response of 25 Californian metro-level house prices, our set-up not only allows for the transmission of monetary policy shocks through multiple channels to the individual house prices but also for the transmission of other structural regional shocks as the house prices loads on both aggregate and regional factors. Jarocinski & Smets (2008) estimates the effect on aggregate US house prices of multiple structural shocks (housing demand shock, monetary policy shock and a term spread shock) using a nine-dimensional Bayesian VAR. Using a combination of zero and sign restrictions, they find that housing demand shocks and monetary policy explains a sizeable fraction of the house price boom. Our structural identification approach is quite similar to Jarocinski & Smets (2008) but because we also have measures of Californian housing construction activity (building permits) we can also identify a housing supply shock; in fact a regional housing supply shock. Moreover, we take into account an important shortcoming of the previous literature about the inference from the impulse responses analysis, when this is based on a combination of zero and sign restrictions; cf. Arias et al. (2014). Musso et al. (2011) use a standard VAR to analyse similarities and differences of the responses of euro area and US aggreate house prices to monetary policy shocks, housing demand shocks and also a credit shock using recursive identification. In sum, we note that the innovative feature of our paper is the way that multiple aggegate shocks as well as regional shocks in a data-rich setting can be traced back to the metro-level Californian house prices, and thus shed light on the role of structural shocks in explaining regional house prices.

The remaining part of the paper is organized as follows. In section 2 we specify a single dynamic factor model for all the US aggregate variables, Californian variables and metropolitan house prices. The economic identification of these dynamic factors as well as the identification of the structural shocks are then detailed. At the end of the section we briefly detail how the model is estimated using the EM algorithm. In section 3 the data are described and in section 4 we present the empirical results including the variance decomposition of key variables and Californian metropolitan house prices. Section 5 concludes.

2 A dynamic factor model for Californian house prices

In recent years, factor models have become a standard tool in applied macroeconomics and finance. Essentially, when the number of random sources of variation is less than the number of dependent variables, then a factor model enables the researcher to reduce the dimension of the number of explanatory variables to few latent factors. Since the first generation of (exact) factor models by Geweke (1977) and Sargent & Sims (1977), a considerable amount of research has been devoted to the econometric theory and empirical analysis of large dimensional approximate dynamic factor models.[2] Starting with the seminal paper of Chamberlain & Rothschild (1983), the large dimensional approximate dynamic factor model is introduced by notably Forni et al. (2000, 2004, 2005) (FHLR) in the frequency domain and Stock & Watson (2002*a*,*b*) (SW) in the time-domain. FHLR and SW estimate the large dimensional dynamic factor model non-parametrically by dynamic and static principal component methods, respectively, but recently these models have also been estimated by Bayesian methods (Otrok & Whiteman (1998), Kim & Nelson (1999)) as well as maximum likelihood methods (Doz et al. (2011*b*,*a*), Jungbacker & Koopman (2008)). Although the literature on applications of dynamic factor models is large, the literature on structural dynamic factor models is smaller with Bernanke et al. (2005) and

[2] By 'large' we mean large in the cross-section, i.e. large in the number of time series (N) and large in the number of observations (T) of the time series. By 'approximate' we refer to the relaxation of the iid error term assumption in the exact factor model such that the error terms are allowed to be weakly (locally) correlated; cf. Chamberlain & Rothschild (1983).

Forni & Gambetti (2010) as leading examples of monetary policy analysis.

In the following, we first present the dynamic factor model briefly, then we discuss identification and finally estimation.

2.1 Dynamic factor model (DFM)

The key implication of the dynamic factor model is that the variation of each of the N observed variables in the panel X can be decomposed into two orthogonal components; a component χ that is common to all variables and an idiosyncratic component ξ specifically related to the individual observed variable. Consequently, the ith variable in the panel X at time t can be written as:

$$X_{it} = \chi_{it} + \xi_{it} \tag{1}$$

for $i = 1, .., N$ and $t = 1, .., T$ with $E\left[\chi_{it}\xi_{js}\right] = 0 \; \forall \; i, j, t, s$ but with a potentially limited amount of correlation among the idiosyncratic components. Consider as in Forni et al. (2005), the specification of the $N \times 1$ vector of the common component at time t to be dynamically explained by a small number $q << N$ of common dynamic factors f_t such that $\chi_t = \lambda^{\top}(L) f_t$, where $\lambda(L)$ is a $q \times N$ matrix polynomial in the lag-operator L of finite order s and where the law of motion of the dynamic factors are commonly given by a standard reduced form $VAR(p)$ process. The dynamic factor model can now be written as:

$$\begin{aligned} X_t &= \lambda_0 f_t + ... + \lambda_s f_{t-s} + \xi_t \\ f_t &= \phi_1 f_{t-1} + ... + \phi_h f_{t-p} + u_t \end{aligned} \tag{2}$$

and assuming a limited response heterogeneity compared to the order of the VAR ($s < p$) allows us to write the dynamic factor model in (2) in a first-order state space representation:[3]

$$\begin{aligned} X_t &= \Lambda F_t + \xi_t \\ F_t &= \Phi F_{t-1} + U_t \end{aligned} \tag{3}$$

where X_t is $N \times 1$, $\Lambda = [\lambda_0, ..., \lambda_p]$ is a $N \times qp$ loading matrix, $F_t = \left[f_t^\top, ..., f_{t-p+1}^\top\right]^\top$ is a qp vector of dynamic factors and their lags, ξ_t is a $N \times 1$ vector with the idiosyncratic error terms, Φ is $qp \times qp$ matrix with autoregressive parameters, and the reduced form VAR residuals reside in $U_t = \left[u_t^\top, 0_{q(p-1)\times 1}^\top\right]^\top$. To fix ideas, we assume that $\xi_t \sim N(0, R)$ with R being a diagonal matrix corresponding to an exact factor model, but we will later relax this assumption. Throughout the paper we assume $u_t \sim N(0, W)$. All processes in (2) and (3) are assumed stationary.

Notice, that at this stage that our objective is to consider how the observed variables in X_t respond to a structural shock to one of the factors. However, the state space model in (3) is not yet econometrically identified; the purely latent factors lack economic interpretation; and we need to identify the structural shocks from the reduced form shocks U_t. Therefore, we now proceed with the identification of the model, then the economic identification of the factors and, finally, the identification of the structural shocks.

2.2 Econometric identification of the DFM

The state space model in (3) is not econometrically identified as it is possible to form observationally equivalent models by arbitrary rotations of the latent factors and the

[3] See Bai & Ng (2007) or Bai & Wang (2012) for more details.

loadings. In the non-parametric principal components approach the factors are usually restricted to be orthogonal in order to achieve identification. Under parametric estimation approaches different identifications schemes have been used, for instance the hierarchical loading matrix in Geweke & Zhou (1996) and Aguilar & West (2000) corresponding to Proposition 1 in Geweke & Singleton (1981), while a simple loading structure of a small subset of the loading matrix has been used in Bork (2008) corresponding to Proposition 2 in Geweke & Singleton (1981).[4] We prefer the latter identification scheme as this allows for correlated factors and thus resembles more closely the type of correlation that we would expect among observed economic variables. Furthermore, the variance-covariance matrix of the VAR innovations is left fully unrestricted, which is preferable in the structural analysis. To be precise, we ensure exact identification by restricting a small subset of the large dimensional loading matrix to be an identity matrix of size $q \times q$; without loss of generality the upper $q \times q$ block of λ_0 could be restricted to I_q.

2.3 Economic identification by overidentifying restrictions

The estimated factors from the exact identified model are, however, purely latent factors without a clear economic interpretation. We now impose over-identifying loading restrictions to enhance the economic interpretation of these factors but also to separate aggregate factors from regional (Californian) factors.

To fix ideas, we present a small stylistic dynamic factor model with q factors which are composed of q_Z aggregate factors $\left(f_t^Z\right)$, q_H Californian house price factors $\left(f_t^H\right)$,

[4] Recently, Bai & Wang (2012) and Bai & Wang (2014) has formulated propositions similar to these propositions targeted especially for dynamic factor models.

q_C Californian economic factors $\left(f_t^C\right)$ and, finally, a perfectly measured monetary policy factor i_t.[5] Hence, $q = q_Z+q_H+q_C+1$. The panel X_t consists of n observed time series which are composed of n_Z are US aggregate time series (Z_t), n_H Californian metro-level house price series (H_t), n_C Californian economic time series (C_t), and finally also the federal funds rate i_t. Therefore, $n = n_Z+n_H+n_C+1$. These observed variables are assumed to load only on contemporaneous factors, hence $s = 0$. In the corresponding state space model below two points should be noted. Firstly, assume that the exactly identifying restrictions have already been imposed on the upper $q_Z \times q_Z$ block of the $n_Z \times q_Z$ loading matrix λ_{ZZ}. The same identification principle applies to λ_{CC} and λ_{HH}. Secondly, note also the over-identifying exclusion restrictions imposed on f_t^C and f_t^H which facilitate the economic identification of separate Californian factors. In particular, US aggregate variables in Z_t are assumed to be contemporaneously unrelated to Californian economic factors, f_t^C and Californian housing factors f_t^H. In contrast the observed Californian economic variables in C_t and house prices in H_t are allowed to load on the overall aggregate variables. This means that we condition the estimate of e.g. the Californian house price factors f_t^H on a lot of information and that metro-level house prices are allowed to load on both aggregate and regional information. Finally, note that the dynamic interactions among the factors are fully unrestricted in the VAR with the implication that e.g. Z_t could depend on f_{t-1}^C

[5] Note, that i_t is not really a factor but a perfectly mesured variable that enters the factor dynamics; hence the notion of factor-augmented VAR of Bernanke et al. (2005)

through f_t^Z.

$$
\begin{bmatrix} Z_t \\ C_t \\ H_t \\ i_t \end{bmatrix} = \begin{bmatrix} \lambda_{ZZ} & 0 & 0 & \lambda_{Zi} \\ \lambda_{CZ} & \lambda_{CC} & \lambda_{CH} & \lambda_{Ci} \\ \lambda_{HZ} & \lambda_{HC} & \lambda_{HH} & \lambda_{Hi} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t^Z \\ f_t^C \\ f_t^H \\ i_t \end{bmatrix} + \begin{bmatrix} \xi_t^Z \\ \xi_t^C \\ \xi_t^H \\ 0 \end{bmatrix} \tag{4}
$$

$$
\begin{bmatrix} f_t^Z \\ f_t^C \\ f_t^H \\ i_t \end{bmatrix} = \Phi(L) \begin{bmatrix} f_{t-1}^Z \\ f_{t-1}^C \\ f_{t-1}^H \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} u_t^Z \\ u_t^C \\ u_t^H \\ u_t^i \end{bmatrix} \tag{5}
$$

The identified aggregate factors and Californian factors are, however, still latent without a clear economic interpretation. We now add one additional layer of identifying exclusion restrictions to achieve a clear economic interpretation.[6] Specifically, we identify an aggregate inflation factor $\left(f_{\pi,t}^Z\right)$ that loads exclusively on observed inflation series in $Z_{\pi,t}$ among the Z_t series, and we identify another aggregate economic activity factor $\left(f_{y,t}^Z\right)$ that loads exclusively on aggregate economic activity related series $Z_{y,t}$ among the Z_t series. Accordingly, λ_{ZZ} is block diagonal with $\lambda_{\pi,ZZ}$ and $\lambda_{y,ZZ}$ along the diagonal, while the exclusion restrictions are seen in the off-diagonal blocks of λ_{ZZ}. Although the loadings λ_{CZ} and λ_{HZ} are left free, we find in our empirical application, that the already imposed exclusion restrictions are sufficient to pin down the economic interpretation; otherwise more restrictions could be imposed. Similarly, we identify a regional (Californian) economic activity factor $\left(f_{y,t}^C\right)$, a specific Californian factor for building permits $\left(f_{b,t}^C\right)$ and a

[6] To conserve space, the fully specified state space system is presented in the appendix.

housing factor $\left(f_{h,t}^{H}\right)$. Thus, in this stylistic dynamic factor model we end up with $q = 6$ factors in $f_t = \text{vec}\left(f_{\pi,t}^{Z}, f_{y,t}^{Z}, f_{y,t}^{C}, f_{b,t}^{C}, f_{h,t}^{H}, i_t\right)$ that all have an economic interpretation and we can now proceed with a structural interpretation of shocks to these factors.

2.4 Structural identification by zero and sign restrictions

Based on the state space system in $(4)-(5)$ and the identified factors f_t we now identify five structural shocks; a positive aggregate demand shock $\left(\varepsilon_t^{AD}\right)$, a positive supply shock $\left(\varepsilon_t^{AS}\right)$, a positive regional (Californian) housing supply shock $\left(\varepsilon_t^{HS}\right)$, a positive regional (Californian) housing demand shock $\left(\varepsilon_t^{HD}\right)$ and a contractionary monetary policy shock $\left(\varepsilon_t^{MP}\right)$. As detailed below, these shocks are essentially given by a mapping from the structural shocks (ε) to the reduced form shocks (u). Hence, once these shocks are identified we can analyze how much of the variation in e.g. the house prices, H_t, that is due to aggregate demand shocks or due to Californian housing demand shocks; or alternatively, how much of the variation in Californian economic variables in C_t that is due to the identified Californian housing supply shock.

We rely on a combination of zero and sign restrictions to identify these structural shocks which has some advantages compared to the popular use of the recursive identification by the Cholesky decomposition and the pure sign restrictions approach. In particular, we do not have to assume that some variables respond with a lag to others, for instance that house prices respond with a delay to interest changes via monetary policy shocks. This implies, that we can remain agnostic about this particular impulse response. Instead, we identify the shocks as the set of randomly drawn candidate shocks that satisfies theoretically motivated sign restrictions on some of the impulse responses but remain agnostic about other impulse responses; see Faust (1998), Canova & Nicolo (2002), Uhlig (2005) and Fry & Pagan (2011) for important contributions to the literature. However, we also use zero restrictions to distinguish between aggregate shocks and regional (Californian) shocks, as this approach is more straightforward than the alternative sign restriction approach, where one would typically require that a particular aggregate variable grows relatively more/less than the regional variable. The details of the sign restrictions and zero restrictions are now discussed and then, finally, we provide a few technical details on the sign restriction methodology.

2.4.1 Identification of structural shocks by zero and sign restrictions

A positive *aggregate supply shock* (AS) is defined as shock that leads to a decrease in inflation, an increase in aggregate output, and an accommodating contractionary monetary policy, while the remaining factors are left unrestricted including the Californian variables; see Table 1 below. A similar definition is seen in Peersman (2005), Iacoviello (2005), Furlanetto et al. (2014). A positive *aggregate demand shock* (AD) moves inflation, output and the interest rate in the same direction which can be distinguished from an expansionary *monetary policy shock* (MP) that would have the opposite sign on the interest rate; see Peersman (2005), Peersman & Straub (2009), and Furlanetto et al. (2014). A Californian *housing supply shock* (HS) moves the house prices and the supply of housing units approximated by Californian building permits in opposite directions, while the Californian *housing demand shock* (HD) moves prices and building permits in the same direction; cf. Jarocinski & Smets (2008) and Abdallah & Lastrapes (2013) for a similar definition. To further distinguish these regional shocks from aggregate shocks, we assume that the aggregate variables are initially unaffected by these housing shocks. A summary of the shocks are given in Table 1, where $(-)$ or $(+)$ indicates the required sign, $(*)$ means unrestricted and (0) indicates a zero restriction. Furthermore, one may define the sign restriction to hold in each of J_i periods or cumulatively over the J_i periods, while the zero restrictions may only be required to hold for one period.

	Shock:	AS	AD	HS	HD	MP
Factor response at horizon $j = 1...J$						
Aggregate inflation	$f^Z_{\pi,j}$	$-$	$+$	0	0	$-$
Aggregate output	$f^Z_{y,j}$	$+$	$+$	0	0	$-$
Regional (Californian) output	$f^C_{y,j}$	$*$	$*$	$*$	$*$	$*$
Regional (Californian) building permits	$f^C_{b,j}$	$*$	$*$	$+$	$+$	$*$
Regional (Californian) house prices	$f^C_{h,j}$	$*$	$*$	$-$	$+$	$*$
Federal funds rate	i_t	$-$	$+$	0	0	$+$

Table 1: Sign restrictions on a small dynamic factor model

2.4.2 Generating shocks that satisfy the zero and sign restrictions

A useful starting point is to consider the popular recursive identification of a structural VAR which is computational convenient because of the simple mapping of the structural shocks, ε_t, to the reduced form shocks u_t via a Cholesky decomposition of the covariance matrix $\hat{\Omega}$ of the estimated residuals given by $\hat{\Omega} = A_0 A_0^\top$. It can then be shown that the uncorrelated structural shocks with unit variance is given by $\hat{\varepsilon}_t = A_0^{-1}\hat{u}_t$; see Christiano et al. (1999) and recently Fry & Pagan (2011). Consider a linear combination of this base set of uncorrelated shocks[7], ε_t, given by a orthogonal rotation matrix Q, resulting in a

[7] It is not important that the these shocks are the results of a Cholesky decomposition; it is merely convenienent.

new set of uncorrelated structural shocks ε_t^*

$$\begin{aligned}
\varepsilon_t^* &= Q\varepsilon_t \\
&= QA_0^{-1}\hat{u}_t \\
&= Q^*\hat{u}_t
\end{aligned} \tag{6}$$

where $Q^* = QA_0^{-1}$ and where the orthogonality implies that $\mathrm{Cov}\left(\varepsilon_t^*\right) = I_q = \mathrm{Cov}\left(\varepsilon_t\right)$. Moreover, the variance of the variables in the reduced form VAR is reproduced using ε_t^*, but the impact on $\hat{u}_t$ and hence the variables in the VAR will be different under the rotation Q; cf. Fry & Pagan (2011). Hence, one can think of sign restrictions as an algorithm that generates orthogonal rotation matrices Q and where only candidate shocks ε_t^* are kept if they satisfy all the restrictions on the impulse responses. At this point, we emphasize that one can trace the response of each of the N observed variables in the panel X_t to the candidate shock ε_t^* through the moving average representation of the dynamic factor model. In particular, rewriting the model in equations $(4) - (5)$, or the more general model in equation (3) in its moving average representation yields

$$X_t = B(L)Q^{*-1}\varepsilon_t^* + \xi_t, \tag{7}$$

where $B\left(L\right) = \Lambda\left[\left(I - \Phi\left(L\right)\right]^{-1} V$ and $V = \left[I_q, 0_{q(p-1)\times q}^{\top}\right]^{\top}$. Subsequently, the median and a lower and upper percentile of the set of candidate impulse responses are typically shown in a graph. However, as noted by Fry & Pagan (2011), the median is not represented by only one model but in fact a mixture of many different structural models, so they advocate to report the model closest to the median.

Implementing the sign restrictions approach to a structural identification of VARs or dynamic factor models requires an efficient algorithm for imposing zero and sign restrictions. Recently, Rubio-Ramirez et al. (2010) propose an algorithm for imposing zero restrictions and long-run restrictions in exactly identified models. Moreover, they propose an algorithm for sign restrictions using the QR decomposition. Arias et al. (2014) and Binning (2013) extend the work by Rubio-Ramirez et al. (2010) to allow for zero restrictions. As discussed by Arias et al. (2014), whenever sign restrictions are combined with zero restrictions, it becomes crucial to condition the draw of Q on the zero restrictions. We build our structural identification on the work by Arias et al. (2014) and Binning (2013) and the algorithm of the former is repeated in the appendix of this paper.

[Does this section need a summary before moving on to estimation?]

2.5 Estimating the DFM by the EM algorithm

The linear Gaussian state space model in (3) with its latent factors f_t is well represented in a Kalman filter setting. However, the Kalman filter needs the parameters $\Theta = \{\Lambda, R, \Phi(L), W\}$ as input and therefore does not estimate these. Building on the seminal work by Dempster et al. (1977), Shumway & Stoffer (1982) introduce the Expectation Maximization (EM) algorithm to estimate the parameters in state space models as the model above. Essentially, the EM algorithm is an iterative maximum likelihood method that switches between an Expectation step and Maximization step. The maximization step results in the following closed form estimators at iteration j

$$
\begin{aligned}
\operatorname{vec}\left(\Lambda^{(j)}\right) &= \operatorname{vec}\left(DC^{-1}\right) && (8)\\
R^{(j)} &= \frac{1}{T}\left(E - DC^{-1}D^{\top}\right) && (9)\\
\operatorname{vec}\left(\Phi^{(j)}\right) &= \operatorname{vec}\left(BA^{-1}\right) && (10)\\
Q^{(j)} &= \frac{1}{T}\left[C - BA^{-1}B^{\top}\right] && (11)
\end{aligned}
$$

where the following moments are available from the Kalman smoother (indicated by subscript $t|T$):

$$
\begin{array}{ll}
A = \sum_{t=1}^{T}\left(\hat{F}_{t-1|T}\hat{F}_{t-1|T}^{\top} + \hat{P}_{t-1|T}\right) & B = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t-1|T}^{\top} + \hat{P}_{\{t,t-1\}|T}\right)\\
C = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t|T}^{\top} + \hat{P}_{t|T}\right) & D = \sum_{t=1}^{T} X_t\hat{F}_{t|T}^{\top}\\
E = \sum_{t=1}^{T} X_t X_t^{\top} &
\end{array}
\qquad (12)
$$

and where F_t is approximated by $\hat{F}_{t|T} = E\left[F_t | \mathcal{X}_T\right]$. $\mathcal{X}_T = \{X_1, .., X_T\}$ denotes the information set, $\hat{P}_{t|T} = var\left(F_t | \mathcal{X}_T\right)$ is the variance and $\hat{P}_{\{t,t-1\}|T} = cov\left(F_t, F_{t-1} | \mathcal{X}_T\right)$ is the lag-one covariance. The estimates $\Theta^{(j)}$ can then be used in the expectation step to compute a new set of moments from the Kalman smoother. Subsequently, the estimated moments are supplied to the maximization step above and the procedure continues until convergence of the likelihood.

In order to implement the identifying loading restrictions of the model in equations (4) – (5) we need a loading estimator subject to the imposed linear restrictions, denoted by Λ^*. Bork et al. (2009) show that the restricted Λ^* estimator subject to linear restrictions in the form $H_\Lambda \operatorname{vec} \Lambda = \kappa_\Lambda$ takes this form

$$\begin{aligned}\operatorname{vec}(\Lambda^*) &= \operatorname{vec}\left(DC^{-1}\right) \\ &\quad + \left(C^{-1} \otimes R\right) H_\Lambda^\top \left[H_\Lambda \left(C^{-1} \otimes R\right) H_\Lambda^\top\right]^{-1} \left\{\kappa_\Lambda - H_\Lambda \operatorname{vec}\left(DC^{-1}\right)\right\}\end{aligned} \tag{13}$$

- We are now ready to take our model to the data. Below we briefly present the data and subsequently the empirical results.

3 Aggregate and regional data

In our application, the data consist of a rich panel X of 183 US aggregate economic and financial time series as well as 75 Californian time series, of which 26 are metro-level house price indices. The sample spans 1986:Q3 - 2014:Q1 and the choice of the starting period is a comprimise between broad coverage and a sufficient length of the time series. Starting earlier would reduce the cross-sectional dimension, and the nominal house prices would be much more noisy as explained in Del Negro & Otrok (2007). A full description of the data is provided in the appendix. The US series represent the following categories of macroeconomic and financial time series: industrial production; capacity utilization; income; employment and hours worked; earnings, aggregate housing variables including housing starts; consumption; orders and inventories; money and credit; bond and exchange rates; consumer, producer and commodity prices; and stock prices. The Californian series represent: labor force, unemployment, metro-level employment, delinquency rates and charge-offs; consumer prices; consumer confidence; retail sales; building permits; and metro-level house prices indices from FHFA. All series are transformed into stationary variables with zero mean and unit variance.

Given our focus on the Californian metropolitan house prices, we depict the annualized mean nominal growth rates over the sample period in Figure 1. The mean annualized house price growth rate across all metropolitan areas is 4%, with the San Francisco area leading with 6%. The interior metropolitan areas like Bakersfield and Visalia-Porterville have experienced growth rates around 2%, so even within California there is significant differences in growth rates. The within state differences become even more significant when we focus on the recent boom period defined as 2000:Q1-2006:Q2 and the bust period from 2006:03-2009:Q3[8]. Although the house prices appreciated even more in some of the interior metropolitan areas during the boom compared to the San Francisco area, Los Angeles area and San Diego area, then nevertheless, these metropolitan areas also experienced a deep downturn during the bust; see Figure 2 and 3.

4 Empirical results

In this section, we present the results of taking the baseline model in baseline model in equations $(4)-(5)$ to the data described above. Specifically, we estimate a DFM with a similar number of factors ($q=6$) as in the baseline model. In fact, using the information criterion by Hallin & Liska (2007) applied to the combined US and Californian dataset, we estimate $\hat{q}=6$ dynamic factors. Moreover, we find $\hat{q}=5$ for the US dataset separtely,

[8]We define the boom and bust period similarly to Huang & Tang (2012)

while $\hat{q} = 3$ for the Californian dataset which indicates common sources of variation in the separate datasets. Later on, we will augment the baseline model with additional factors, based on the fact that our factors are more restricted than the unrestricted factors and to analyze whether there Californian house prices are in fact driven by a coastal house price factor and a interior house price factor. Figure 4 depicts the estimated factors and it can be seen that the first factor is very close related to US aggregate inflation $\left(f^{Z}_{\pi,t}\right)$. The second factor is closely related to aggregate US employment, so this is our approximation of US aggregate production $\left(f^{Z}_{y,t}\right)$, i.e. a measure of real activity. The third factor is defined as a Californian employment factor as a approximation of $\left(f^{C}_{y,t}\right)$. The fourth factor is a Californian building permits factor which is required in order to define a Californian housing supply and demand shock. The fifth factor is a common Californian house price factor and the sixth factor is the perfectly measured monetary policy rate.

Given the estimated dynamic factor model we can now conduct a structural analysis based on the estimated reduced form residuals, u_t, and the identification scheme in Tabel 1 above, where we impose cumulative sign restrictions over only $J = 2$ periods. The main message in the forecast error variance decomposition is conveyed in Figure 5. Specifically, in a forecast error variance decomposition, we calculate for a given forecast horizon what fraction of the total forecast error variance for a particular variable as a result of a specific shock, for instance the monetary policy shock. Hence, the forecast error variance decomposition is similar to the R^2 measure but for forecast errors at different horizons. Focusing on the Californian metro-level house prices it can be seen that the primary source of the variation in the house prices is regional housing supply shocks at the longer term while housing demand shocks play a bigger role at the medium term. Moreover, housing demand also plays a role in explaing e.g. Californian retail sales whereas housing supply plays a small role in explaining Californian unemployment. Notice also the almost non-existing role of moneary policy shocks in the explanation of Californian house prices.

The impulse responses in Figures 6, 7, 8, 9, and 10 show the response of a subset of observed variables to a one standard deviation shock to AS, AD, Housing supply, Housing demand, and monetary policy, respectively. Although, some of the confidence bands are wide, the impulse responses have the expected form. [[Add more comments]].

4.1 Augmenting the baseline model

Incomplete.

$f_t = \mathrm{vec}\left(f^Z_{\pi,t}, f^Z_{y,t}, f^C_{y,t}, f^C_{b,t}, f^H_{h,t}, i_t\right)$

- The idea is to augment the baseline model with an additional two factors, so we have $q = 8$ factors, $f_t = \mathrm{vec}\left(f^Z_{\pi,t}, f^Z_{y,t}, f^C_{y,t}, f^C_{b,t}, f^H_{h,t}, i_t\right)$, where $f^Z_{y,t}$ is defined as capacity utility factor and where an aggregate employment factor $f^Z_{e,t}$ is added as an additional factor. Moreover, $f^C_{y,t}$ is continues to be as a regional (Californian) employment factor $f^C_{e,t}$ to catch regional employment trends. Finally, the Californian house price factor is split into a coastal house price factor $f^C_{hc,t}$ and a Californian interior house price $f^Z_{hi,t}$ based on the insights from a principal component analysis of the panel of Californian house prices.

5 Conclusion

This paper advances a large-dimensional dynamic factor models to analyze the sources of variation in Californian metro-level house prices. Over-identifying loading restrictions are a key to identify two US aggregate dynamic factors, three regional (Californian) factors and a perfectly measured monetary policy rate. Specifically, we identify a US aggregate inflation factor; a US aggregate employment factor; a Californian employment factor; a Californian building permits factor and a common Californian housing price factor that allow us to identify housing shocks; and, finally, a monetary policy factor. The precise definition of the factors combined with zero restrictions help us in separating US shocks from regional (Californian) shocks.

Using a combination of zero restrictions and sign restrictions we do find that aggregate shocks play a role in explaining the variation of Californian metropolitan house prices, but Californian housing demand and Californian housing supply shocks are the most important.

A Appendix

A.1 Data decription

Data description of the expanded dataset of $N = 272$ series over the period $1986 : Q1 - 2014 : Q1$. The third column contains transformation codes. For the individual variable x, 1 means level, 2 means Δx, 4 means $\ln x_t$, and 5 means $\ln x_t - \ln x_{t-1}$.

Table 2: Data series

Variables	**Units/base/index**	**Code**
US indl prod - final products and nonindustrial supplies vola	index 2007=100	5
US indl prod - final products, total vola	index 2007=100	5
US industrial production - manufacturing (sic) vola	index 2007=100	5
US indl prod - electric and gas utilities vola	index 2007=100	5
US indl prod - fuels vola	index 2007=100	5
US indl prod - residential utilities vola	index 2007=100	5
US indl prod - mining naics=21 vola	index 2007=100	5
US indl prod - automotive products (consumer goods) vola	index 2007=100	5
US indl prod - materials, total vola	index 2007=100	5
US indl prod - consumer goods vola	index 2007=100	5
US indl prod - business equipment vola	index 2007=100	5
US indl prod - durable consumer goods vola	index 2007=100	5
US indl prod - nondurable consumer goods vola	index 2007=100	5
US indl prod - durable mfg (naics) vola	index 2007=100	5
US industrial production - nondurable manufactures vola	index 2002=100	5
US indl prod - nonenergy durable goods materials vola	index 2007=100	5
US indl prod - nondurb goods materials vola	index 2007=100	5
US real GDP pct. change at annual rates (ar) cona	pct.	1
US real GDP pct. change-gross priv dom. investment, fixed (ar) cona	pct.	1
US real GDP pct. change-gross priv fixed investment,nonres (ar) cona	pct.	1
US real GDP pct. change-gross priv fixed investment,resident(ar)	pct.	1
US real GDP pct. change - pce (ar) cona	pct.	1
US real GDP pct. change - pce, durables (ar) cona	pct.	1
US real GDP pct. change - pce, nondurables (ar) cona	pct.	1
US real GDP pct. change - pce, services (ar) cona	pct.	1
US indl utilization - manufacturing (sic) sadj	pct.	1
US indl utilization - durable mfg (naics) sadj	pct.	1
US indl utilization - nondurable mfg (naics) sadj	pct.	1
US indl utilization - selected high-technology industries sadj	pct.	1
US indl utilization - automobile and light duty motor vehicle sadj	pct.	1
US indl utilization - computer and electronic product sadj	pct.	1
US indl utilization - semiconductor and related equipment sadj	pct.	1
US indl utilization - food sadj	pct.	1
US ISM manufacturers survey: production index sadj	index	1
US ISM purchasing managers index (mfg survey) sadj	index	1
US ISM manufacturers svy results: production - net nadj	pct.	1
US personal income less transfer payments cona	usd bil 2009 chnd prc	5
US disposable personal income (monthly series) (ar) cona	usd bil 2009 chnd prc	5
US disposable personal income (ar) (chg p/p) cura	usd bil	1
US personal savings as a pct. of disposable personal income sadj	pct.	1
US personal dividend income (ar) cura	usd bil	5
US personal income receipts on assets cura	usd bil	5
US total civilian employment vola	ths. pers.	5
US employed, nonagriculture - (16 yrs+) (Househ. survey) vola	ths. pers.	5
US unemployment rate sadj	pct.	2
US unemployed (16 yrs and over) vola	ths. pers.	2
US unemployed for less than 5 weeks vola	ths. pers.	2
US unemployed for 5 to 14 weeks vola	ths. pers.	2
US unemployed for 15 to 26 weeks vola	ths. pers.	2
US unemployed for 15 weeks or more vola	ths. pers.	2

Continued on next page

Variables	Units/base/index	Code
US unemployed for 27 weeks and over vola	ths. pers.	2
US average duration of unemployment (weeks) vola	week	2
US employed - nonfarm industries total (payroll survey) vola	ths. pers.	5
US employed - total private vola	ths. pers.	5
US employed - goods-producing vola	ths. pers.	5
US employed - natural resources and mining vola	ths. pers.	5
US employed - construction vola	ths. pers.	5
US employed - manufacturing vola	ths. pers.	5
US employed - durable goods vola	ths. pers.	5
US employed - nondurable goods vola	ths. pers.	5
US employed - service-providing vola	ths. pers.	5
US employed - trade, transportation, and utilities vola	ths. pers.	5
US employed - wholesale trade vola	ths. pers.	5
US employed - retail trade vola	ths. pers.	5
US employed - financial activities vola	ths. pers.	5
US employed - government vola	ths. pers.	5
US employed - information vola	ths. pers.	5
US employed - private service-providing vola	ths. pers.	5
US employed - professional and business services vola	ths. pers.	5
US employed - education and health services vola	ths. pers.	5
US employed - leisure and hospitality vola	ths. pers.	5
US workers on involuntary parttime - all industries vola	ths. pers.	5
US initial claims for unemployment insurance (bci 5) vola	ths. pers.	2
US ISM manufacturers survey: employment index sadj	index	1
US ISM manufacturers svy results: employment - net nadj	pct.	1
US unemployed-job losers and completed temp jobs as a pct. labor forc	pct.	1
US nfib survey: pct. of firms with 1 or more hard to fill jobs sadj	pct.	1
US consumer confidence currently - jobs not so plentiful sadj	pct.	1
US consumer confidence currently - jobs plentiful sadj	pct.	1
US consumer confidence in 6 months - jobs fewer sadj	pct.	1
US consumer confidence in 6 months - jobs more sadj	pct.	1
US avg wkly hours - nondurable goods vola	hour	1
US hours worked of all persons - business sector (ar) vola	pct.	1
US hours worked of all persons - manufacturing sector (ar) vola	pct.	1
US hours worked of all persons-nonfarm business sector(pct. qoq,ar)	pct.	1
US avg overtime hours - manufacturing vola	hour	2
US avg overtime hours - durable goods vola	hour	2
US avg o/t hours prod wrkrs - industrial machinery vola	hour	2
US personal consumption expenditures (ar) cona	usd bil 2009 chnd prc	1
US chain-type quantity index for personal consmptn.expend.s sadj	index 2009=100	5
US chain-type quantity index for pce durables sadj	index 2009=100	5
US chain-type quantity index for pce nondurables sadj	index 2009=100	5
US chain-type quantity index for pce services sadj	index 2009=100	5
US chain-type quantity index for pce goods sadj	index 2009=100	5
US housing permits authorized - midwest (ar) vola	ths.	2
US housing permits authorized - northeast (ar) vola	ths.	2
US building permits to new private housing units vola	ths.	2
US housing permits authorized - south (ar) vola	ths.	2
US housing permits authorized - west (ar) vola	ths.	2
US new private housing units started vola	ths.	2
US housing started - midwest (ar) vola	ths.	2
US housing started - northeast (ar) vola	ths.	2
US housing started - south (ar) vola	ths.	2
US housing started - west (ar) vola	ths.	2
US new privately owned housing units completed (ar) vola	ths.	2
US monthly supply of new homes on market vola	month	1
US chg in priv invtry, autos, used (ar) cura	usd bil	1
US inventory change - autos (ar) cura	usd bil	1
US inventory change - new autos (ar) cura	usd bil	1
US nfib survey: pct. may add to inventories less pct. may reduce sadj	pct.	1
US ISM manufacturers survey: inventories index nadj	index	1
US manufacturers new orders, consumer goods and materials cona	usd mil 1982 prices	5
US ISM manufacturers survey: new orders index sadj	index	1
US ISM new orders sadj	pct.	1
US NFIB survey: pct. expecting higher sales less pct. expecting declin	pct.	1
US New York stock exchange composite share price index	index end 2002=5000	5
US Dow Jones industrials share price index (ep) nadj	index	5
US Standard and Poors' 500 composite - dividend yld	pct.	1
US Standard and Poors' composite index (ep)	index (1941-43=100)	5

Continued on next page

Variables	Units/base/index	Code
US Standard and Poors' share price index - industrials (ep)	index	5
US Standard and Poors' 500 composite - real p/e ratio	actual	1
CN Canadian dollars to 1 U.S. dollar (monthly average) voln	cad	5
JP Japanese yen to US $	jpy	5
SW Swiss Francs To USD	chf	5
UK US $ to 1	usd	5
US effective exchange rate narrow index - nominal nadj	index	5
US federal funds rate (avg.)	pct.	1
US T-bill spread 3 month - Fed Funds	pct.	1
US T-bill spread 6 month - Fed Funds	pct.	1
US corporate bond yield spread - Moody's AAA - Fed Funds	pct. period avrge.	1
US corporate bond yield spread - Noody's BAA - Fed Funds	pct. period avrge.	1
US US conventional ixed mortgages spread - 30 yr - Fed Funds	pct.	1
US money supply zero maturity cura	usd bil	5
US money supply m1 cura	usd bil	5
US money supply m2 cura	usd bil	5
US mny stock-institutional money funds(min investment over $50k)	usd bil	5
US broad money (m3) sadj	index 2010=100	5
US st. louis adjusted monetary base cura	usd bil	5
US money stock - currency in circulation cura	usd bil	5
US commercial and industrial loans outstanding (bci 101) cona	usd mil 2009 chnd prc	5
US ratio of consumer credit outstanding to personal income(bci 9	pct.	2
US nonrevolving consumer credit outstanding cura	usd bil	5
US consumer credit outstanding cura	usd bil	5
US commercial bank assets - commercial and industrial loans cura	usd bil	5
US commercial bank assets - consumer loans cura	usd bil	5
US commercial bank assets - real estate loans cura	usd bil	5
US resl mtg loans: all, 30 days delinquent sadj	pct.	2
US resl mtg loans: all, foreclosures started sadj	pct.	2
US ISM manufacturers survey: prices paid index sadj	index	1
Spot market price: All		5
Spot market price: Metals		5
Spot market price: Raw ind.		5
US PPI - finished consumer foods sadj	index 1982=100	5
US PPI - nonferrous metals nadj	index 1982=100	5
US PPI - crude materials sadj	index 1982=100	5
US PPI - finished consumer goods sadj	index 1982=100	5
US PPI - finished goods sadj	index 1982=100	5
US PPI - intermediate materials,supplies and components sadj	index 1982=100	5
US CPI - apparel sadj	index 1982-1984=100	5
US CPI - transportation sadj	index 1982-1984=100	5
US CPI - medical care sadj	index 1982-1984=100	5
US CPI - commodities sadj	index 1982-1984=100	5
US CPI - durables sadj	index 1982-1984=100	5
US CPI - all urban: all items sadj	index 1982-1984=100	5
US CPI - services sadj	index 1982-1984=100	5
US CPI - all items less food sadj	index 1982-1984=100	5
US CPI - all items less food and energy (core) sadj	index 1982-1984=100	5
US CPI - all items less shelter sadj	index 1982-1984=100	5
US CPI - all items less medical care sadj	index 1982-1984=100	5
US CPI - housing sadj	index 1982-1984=100	5
US CPI - nondurables sadj	index 1982-1984=100	5
US avg hourly real earnings - construction cona	usd 1982-84 prices	5
US ahe prod wrkrs-durable goods cona	usd 1982-84 prices	5
US ahe prod wrkrs-nondurable goods cona	usd 1982-84 prices	5
US avg hourly real earnings - goods-producing cona	usd 1982-84 prices	5
US avg hourly real earnings - manufacturing cona	usd 1982-84 prices	5
US avg hourly real earnings - financial activities cona	usd 1982-84 prices	5
US real hourly earnings - business sadj	pct.	1
US avg hourly real earn- trade, transportation, and utilities cona	usd 1982-84 prices	5
US avg hourly real earnings - private service-providing cona	usd 1982-84 prices	5
US consumer confidence - expectations sadj	index 1966m1=100	1
US CLI consumer sentiment sadj	index 2005=100	1
US TCB CEO confidence survey - conditions in 6 months nadj	pct.	1
CA labor force - California vola	ths. pers.	5
CA labor force-bal of california,state less LA-long bea	person	5
CA labor force - LA-long beach-glendale, ca md vola	person	5
CA unemployment - California vola	ths. pers.	2

Continued on next page

Variables	Units/base/index	Code
CA unemployment rate - California sadj	pct.	2
CA employment - California vola	ths. pers.	5
All Employees: Total Nonfarm in Modesto, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Oxnard-Thousand Oaks-Ventura, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Riverside-San Bernardino-Ontario, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Salinas, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in San Diego-Carlsbad-San Marcos, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Santa Barbara-Santa Maria-Goleta, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Santa Rosa-Petaluma, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Stockton, CA (MSA)	ths. of persons	5
All Employees: Total Nonfarm in Bakersfield, CA (MSA)	ths. of persons	5
FRM, 30yr, F.Mac W		1
US resl mtg loans: all, 30 days delinquent, California nadj	pct.	2
US resl mtg loans: all, 90+ days delinquent, California nadj	pct.	5
US resl mtg loans: all, total delinquent, California nadj	pct.	5
US resl mtg loans: all,foreclosure inventory, California nadj	pct.	5
US resl mtg loans: all,foreclosures started, California nadj	pct.	2
Total Nonperforming Loans for Commercial Banks in California	ths. of dollars	2
Total Net Charge-offs for Commercial Banks in California	ths. of dollars	2
CPI-U: Househ. energy in SF-Oakland-San Jose, CA (CMSA)	index 1982-84=100	5
CPI-U: Utility (piped) gas service in SF-Oakland-San Jose, CA (CMSA)	index 1982-84=100	5
CPI-U: Gasoline (all types) in SF-Oakland-San Jose, CA (CMSA)	index 1982-84=100	5
CPI-U: All items in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Commodities in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Food in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Housing in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Shelter in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Fuels and utilities in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Househ. energy in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Househ. furnishings and operations in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Medical care in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Nondurables less food in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Services less medical care services in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Owners' equiv. rent of residences in LA-Riverside-Orange County, CA (CMSA)	index december 1982=100	5
CPI-U: Utility (piped) gas service in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
CPI-U: Gasoline (all types) in LA-Riverside-Orange County, CA (CMSA)	index 1982-84=100	5
US consumer confidence index - pacific nadj	index 1985=100	1
Consumer Sentiment West		1
Real retail sales		5
Building Permits: total for California (SW)		5
Building Permits: 1-unit for California (SW)		5
Building Permits: 2-unit for California (SW)		5
Building Permits: 3-4 unit for California (SW)		5
Building Permits: 5-unit and more for California (SW)		5
House prices: Anaheim-Santa Ana-Irvine	FHFA metro index	5
House prices: Bakersfield	FHFA metro index	5
House prices: Chico	FHFA metro index	5
House prices: Fresno	FHFA metro index	5
House prices: LA-Long Beach-Glendale	FHFA metro index	5
House prices: Madera	FHFA metro index	5
House prices: Merced	FHFA metro index	5
House prices: Modesto	FHFA metro index	5
House prices: Napa	FHFA metro index	5
House prices: Oakland-Hayward-Berkeley	FHFA metro index	5
House prices: Oxnard-Thousand Oaks-Ventura	FHFA metro index	5
House prices: Redding	FHFA metro index	5
House prices: Riverside-San Bernardino-Ontario	FHFA metro index	5
House prices: Sacramento–Roseville–Arden-Arcade	FHFA metro index	5
House prices: Salinas	FHFA metro index	5
House prices: San Diego-Carlsbad	FHFA metro index	5
House prices: SF-Redwood City-South SF	FHFA metro index	5
House prices: San Jose-Sunnyvale-Santa Clara	FHFA metro index	5
House prices: San Luis Obispo-Paso Robles-Arroyo Grande	FHFA metro index	5
House prices: San Rafael	FHFA metro index	5
House prices: Santa Cruz-Watsonville	FHFA metro index	5
House prices: Santa Maria-Santa Barbara	FHFA metro index	5
House prices: Santa Rosa	FHFA metro index	5
House prices: Stockton-Lodi	FHFA metro index	5
House prices: Vallejo-Fairfield	FHFA metro index	5

Continued on next page

Variables	Units/base/index	Code
House prices: Visalia-Porterville	FHFA metro index	5

Proprietary data sources: Datastream. Public sources: Federal Reserve Economic Data (FRED). Federal Housing Finance Agency.

References

Abdallah, C. S. & Lastrapes, W. D. (2013), 'Evidence on the Relationship between Housing and Consumption in the United States: A State-Level Analysis', *Journal of Money, Credit and Banking, Blackwell Publishing* **45**(4), 559–590.

Aguilar, O. & West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business and Economic Statistics* **3**(18), 338–357.

Arias, J. E., Rubio-Ramirez, J. F. & Waggoner, D. F. (2014), Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications, International Finance Discussion Papers 1100, International Finance Discussion Paper (1100), Board of Governors of the Federal Reserve System.

Bai, J. & Ng, S. (2007), 'Determining the number of primitive shocks in factor models', *Journal of Business & Economic Statistics* **25**, 52–60.

Bai, J. & Wang, P. (2012), Identification and estimation of dynamic factor models, Mpra paper, University Library of Munich, Germany.

Bai, J. & Wang, P. (2014), 'Identification and bayesian estimation of dynamic factor models', *Journal of Business and Economic Statistics* **(forthcoming)**, 1–53.

Bernanke, B. S., Boivin, J. & Eliasz, P. (2005), 'Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach', *The Quarterly Journal of Economics* pp. 387–422.

Binning, A. (2013), Underidentified svar models: A framework for combining short and long-run restrictions with sign-restrictions, Working Paper 2013/14, Norges Bank.

Bork, L. (2008), Estimating US monetary policy shocks using a factor-augmented vector autoregression: An EM algorithm approach. CREATES Research Paper 2009-11.

Bork, L., Dewachter, H. & Houssa, R. (2009), Identification of Macroeconomic Factors in Large Panels, CREATES Research Papers 2009-43, School of Economics and Management, University of Aarhus.

Canova, F. & Nicolo, G. D. (2002), 'Monetary disturbances matter for business fluctuations in the G-7', *Journal of Monetary Economics, Elsevier* **49**(6), 1131–1159.

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica* **51**(5), 1281–1304.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (1999), Monetary policy shocks: What have we learned and to what end?, *in* J. B. Taylor & M. Woodford, eds, 'Handbook of Macroeconomics', Vol. 1A, Elsevier Science, North Holland, New York, pp. 65–148.

Del Negro, M. & Otrok, C. (2007), '99 Luftballons: Monetary policy and the house price boom across U.S. states', *Journal of Monetary Economics* **54**(7), 1962–1985.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society* **39**(1), 1–38.

Doz, C., Giannone, D. & Reichlin, L. (2011*a*), 'Slet a quasi maximum likelihood approach for large approximate dynamic factor models', *Review of Economics and Statistics (forthcoming)* (674). European Central Bank Working Paper no. 674.

Doz, C., Giannone, D. & Reichlin, L. (2011*b*), 'A two-step estimator for large approximate dynamic factor models based on Kalman filtering', *Journal of Econometrics (forthcoming)* (6043).

Faust, J. (1998), 'The robustness of identified var conclusions about money', *Carnegie-Rochester Conference Series on Public Policy* **49**, 207–244.

Forni, M. & Gambetti, L. (2010), 'The dynamic effects of monetary policy: A structural factor model approach', *Journal of Monetary Economics* **57**(2), 203–216.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000), 'The generalized dynamic-factor model: Identification and estimation', *The Review of Economics and Statistics* **82**(4), 540–554.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2004), 'The generalised dynamic factor model: consistency and rates', *Journal of Econometrics* **119**(2), 231–255.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2005), 'The generalized dynamic factor model: one sided estimation and forecasting', *Journal of the American Statistical Association* **100**(471), 830–840.

Fry, R. & Pagan, A. (2011), 'Sign restrictions in structural vector autoregressions: A critical review', *Journal of Economic Literature* **49**(4), 938–60.

Furlanetto, F., Ravazzolo, F. & Sarferaz, S. (2014), Identification of financial factors in economic fluctuations, Working Paper, Norges Bank 2014/09, Norges Bank.

Geweke, J. (1977), The dynamic factor analysis of economic time series, *in* D. J. Aigner & A. S. Golderger, eds, 'Latent Variables in Socioeconomic Models', North Holland.

Geweke, J. F. & Singleton, K. J. (1981), 'Maximum likelihood "confirmatory" factor analysis of economic time series', *International Economic Review* **22**(1), 37–54.

Geweke, J. & Zhou, G. (1996), 'Measuring the pricing error of the arbitrage pricing theory', *Review of Financial Studies* **9**(2), 557–587.

Hallin, M. & Liska, R. (2007), 'Determining the number of factors in the generalized factor model', *Journal of the American Statistical Association* **102**, 603–617.

Huang, H. & Tang, Y. (2012), 'Residential land use regulation and the US housing price cycle between 2000 and 2009', *Journal of Urban Economics* **71**(1), 93–99.

Iacoviello, M. (2005), 'House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle', *American Economic Review* **95**(3), 739–764.

Jarocinski, M. & Smets, F. R. (2008), 'House prices and the stance of monetary policy', *Federal Reserve Bank of St. Louis Review* (Jul), 339–366.

Jungbacker, B. & Koopman, S. J. (2008), Likelihood-based analysis for dynamic factor models, Tinbergen Institute Discussion Papers 08-007/4, Tinbergen Institute.

Kim, C.-J. & Nelson, C. R. (1999), *State-Space Models with Regime Switching*, The MIT Press.

Musso, A., Neri, S. & Stracca, L. (2011), 'Housing, consumption and monetary policy: How different are the US and the euro area?', *Journal of Banking & Finance* **35**(11), 3019–3041.

Otrok, C. & Whiteman, C. H. (1998), 'Bayesian leading indicators: Measuring and predicting economic conditions in Iowa', *International Economic Review* **39**(4), 997–1014.

Peersman, G. (2005), 'What caused the early millennium slowdown? Evidence based on vector autoregressions', *Journal of Applied Econometrics,* **20**(2), 185–207.

Peersman, G. & Straub, R. (2009), 'Technology shocks and robust sign restrictions in a euro area SVAR', *International Economic Review* **50**(3), 727–750.

Rubio-Ramirez, J. F., Waggoner, D. F. & Zha, T. (2010), 'Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference', *Review of Economic Studies* **77**(2), 665–696.

Sargent, T. J. & Sims, C. A. (1977), Business cycle modeling without pretending to have too much a priori economic theory, *in* C. A. Sims, ed., 'New Methods in Business Cycle Research', Federal Reserve Bank on Minneapolis, pp. 45–110.

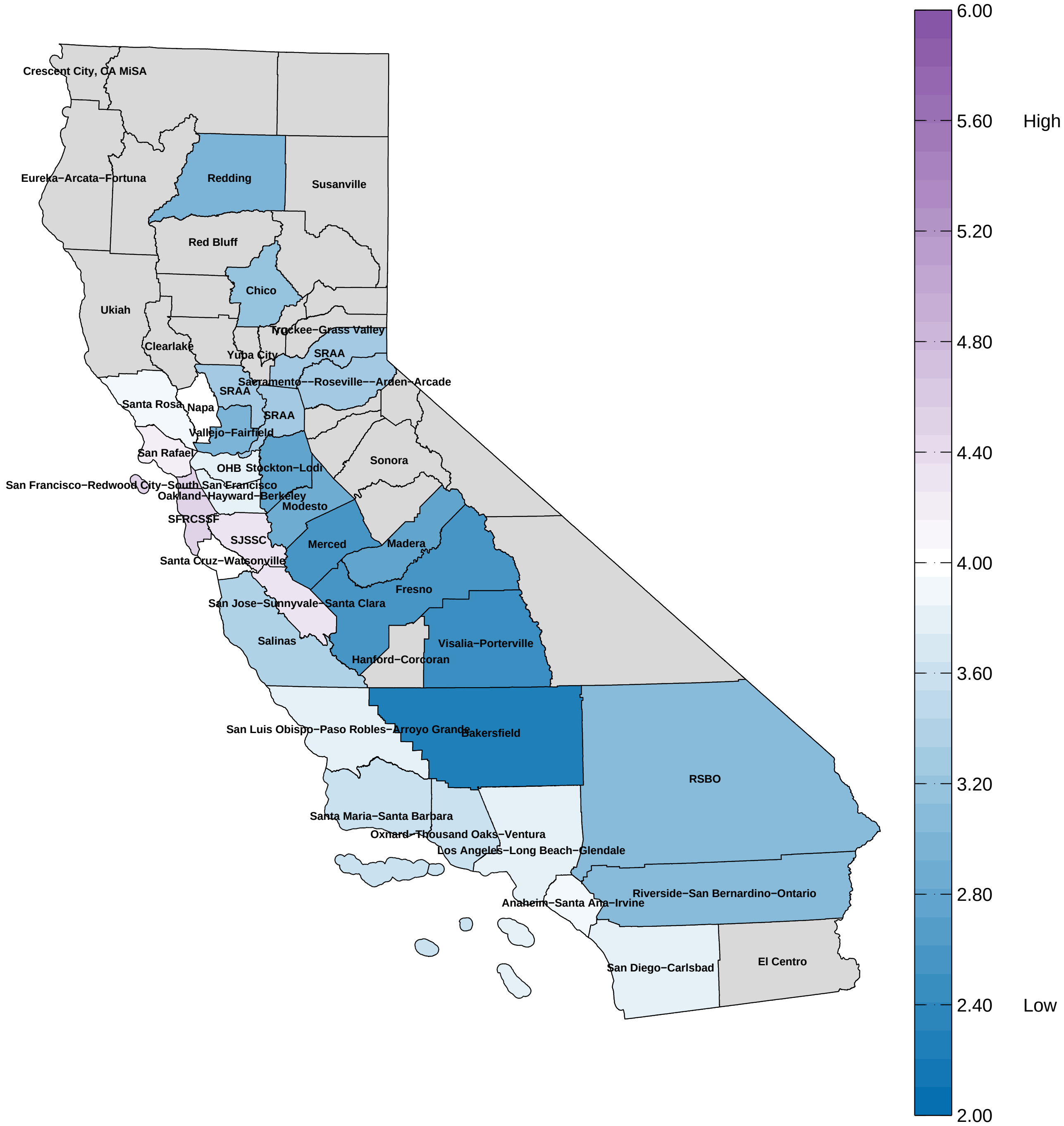
Shumway, R. H. & Stoffer, D. S. (1982), 'An approach to time series smoothing and forecasting using the EM algorithm', *Journal of Time Series Analysis* **3**, 253–226.

Stock, J. H. & Watson, M. W. (2002*a*), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J. H. & Watson, M. W. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business Economics and Statistics* **XX:II**, 147–162.

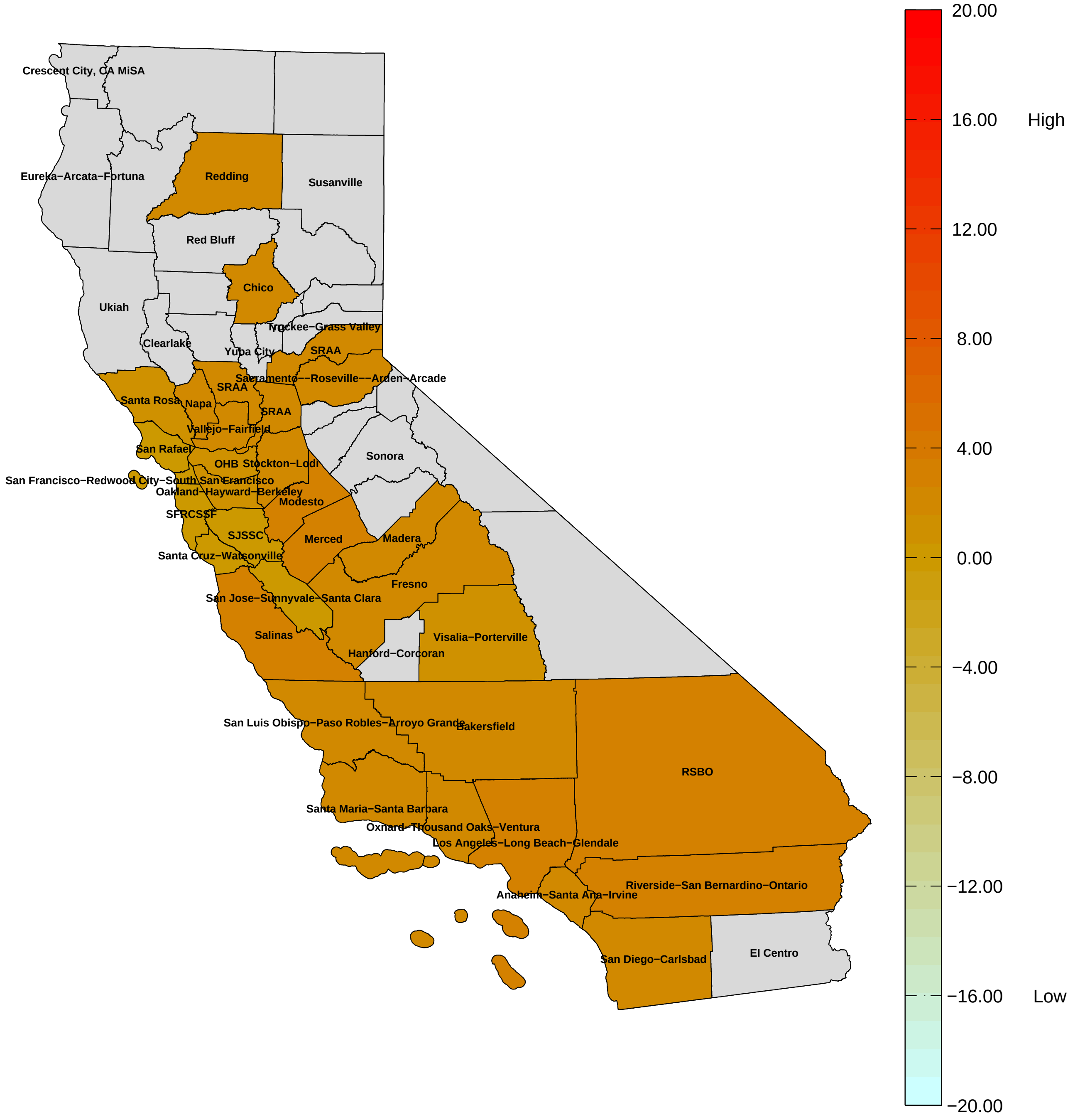
Uhlig, H. (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics* **52**(2), 381–419.

Figure 1: Mean growth rates of Californian metropolian house prices during the sample period.



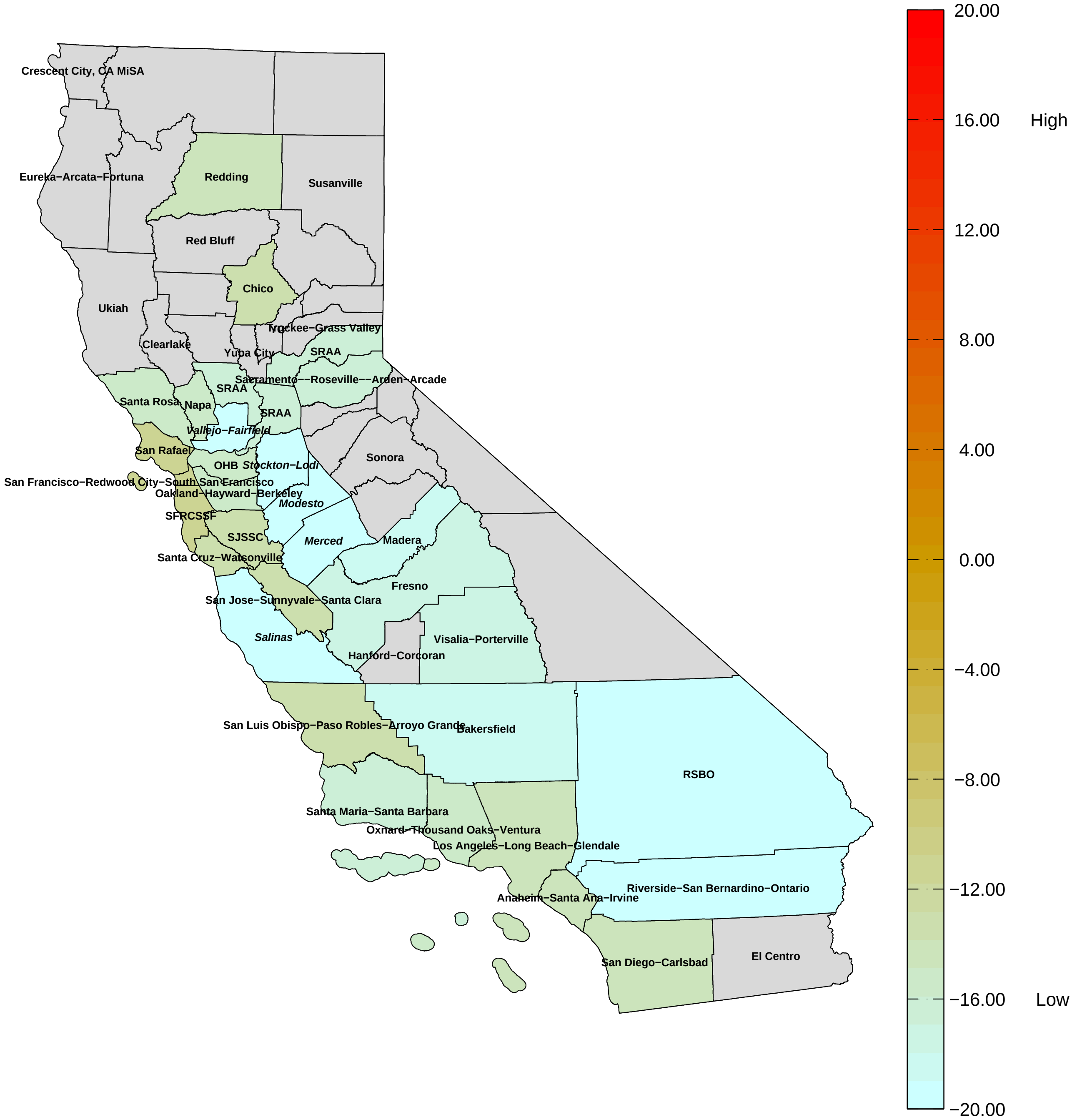
The figure displays the annualized mean nominal growth rates of metro-level house prices in California over the period 1983:Q3 - 2014:Q3. The grey counties are excluded because of insufficient sample length.

Figure 2: Mean growth rates of Californian metropolian house prices during the boom period.



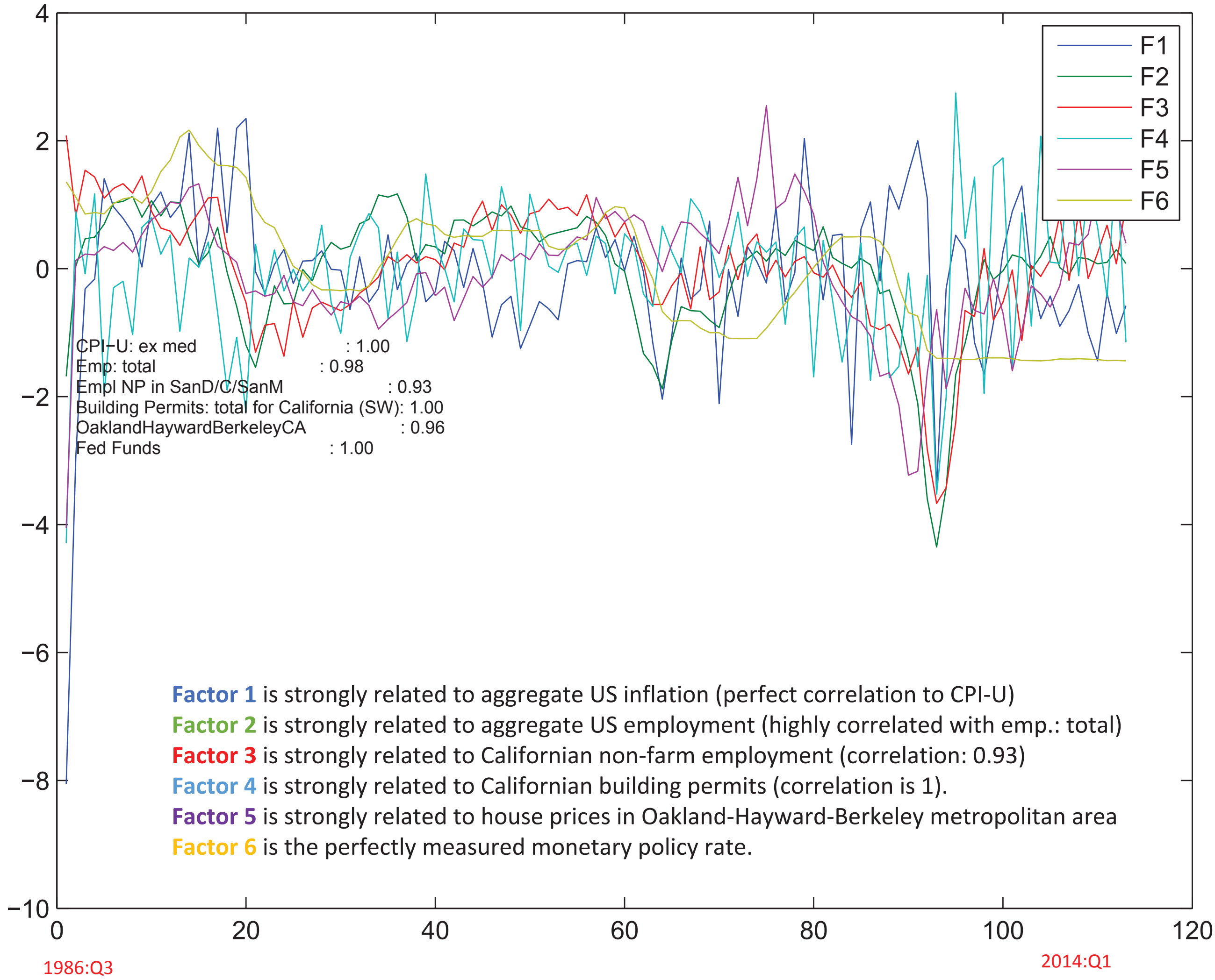
The figure displays the annualized mean nominal growth rates of metro-level house prices in California during the *boom* period defined as 2000:Q3 - 2006:Q2, similarly to Huang and Tang (2012). The grey counties are excluded because of insufficient sample length. A few metropolitan areas are written in italics which indicates that their growth rates are outside the bounds given by the right-hand-side bar. The use of capital letters only, are an abbreviation of the metropolitan name given in the adjacent county.

Figure 3: Mean growth rates of Californian metropolian house prices during the bust period.



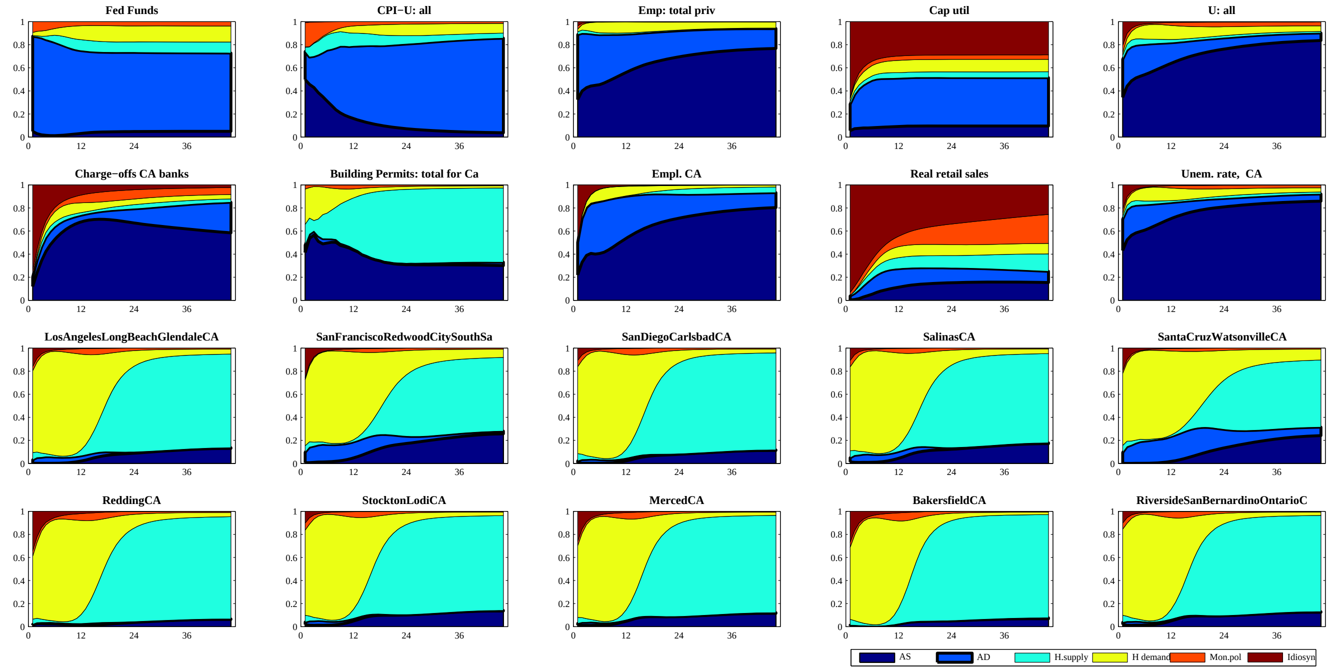
The figure displays the annualized mean nominal growth rates of metro-level house prices in California during the *bust* period defined as 2006:Q3 - 2009:Q3, similarly to Huang and Tang (2012). The grey counties are excluded because of insufficient sample length. A few metropolitan areas are written in italics which indicates that their growth rates are outside the bounds given by the right-hand-side bar. The use of capital letters only, are an abbreviation of the metropolitan name given in the adjacent county.

Figure 4: Estimated factors from the baseline model.



The figure displays the estimated dynamic factors from the baseline model using the EM algorithm.

Figure 5: Forecast error variance decomposition based on the baseline model.



The figure displays the variance decomposition of the baseline model. Specifically, we decompose the variation in e.g. house prices in Los Angeles Long Beach Glendale metropolitan area into a small idiosyncratic component, an even smaller role for monetary policy, a larger role of US aggregate demand and US aggregate supply, but with the predominant source of variation from housing supply (longer term) and housing demand (medium term). The horizontal axis display the horizon in quarters.

The figure displays the responses of a subset of observed variables from the panel due to a one standard deviation shock. Horizontal axis is measured in quarters and the vertical axis in standard deviation.

Figure 6: Impulse responses of the observed variables following a favourable AS shock (baseline model).

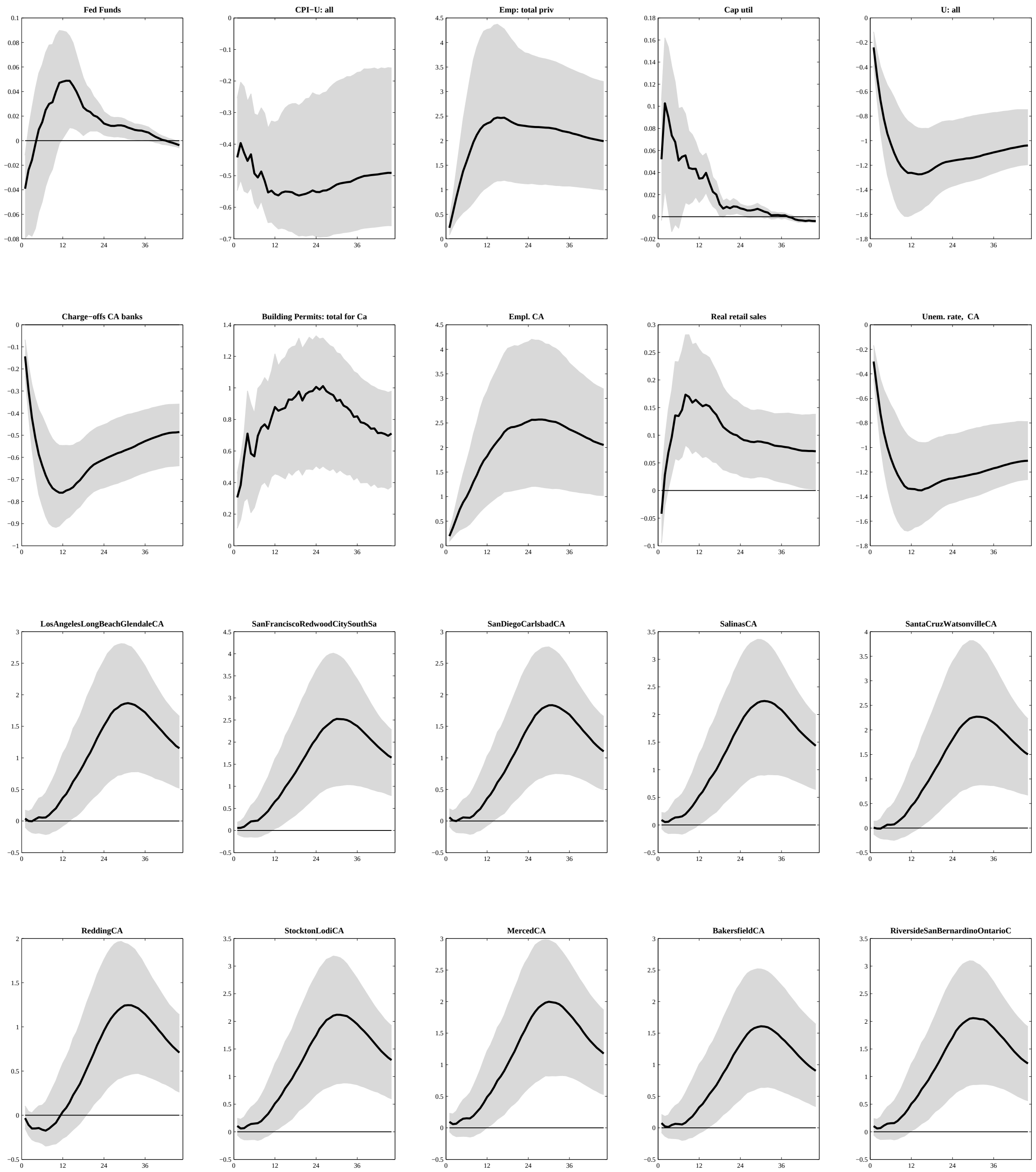
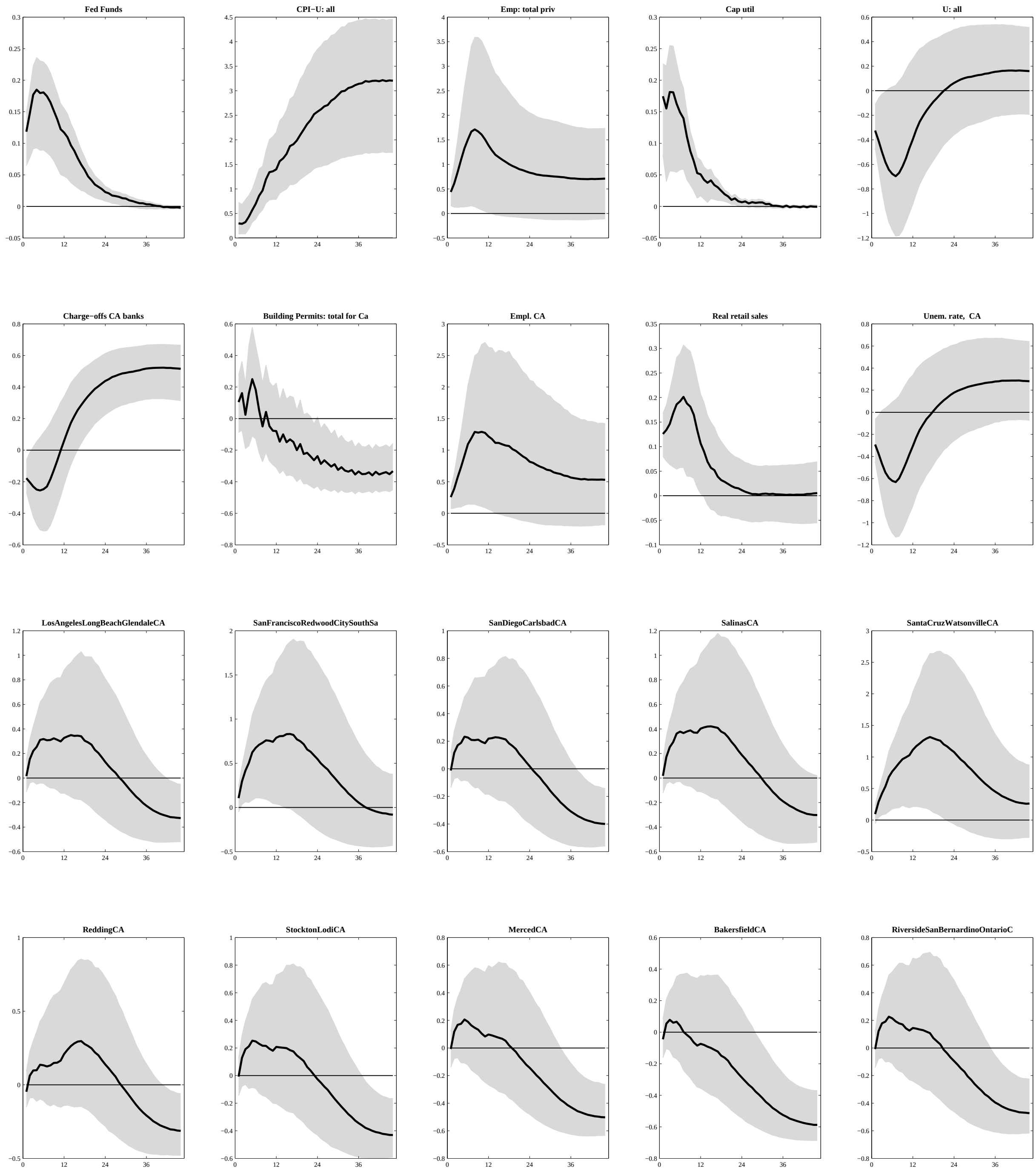
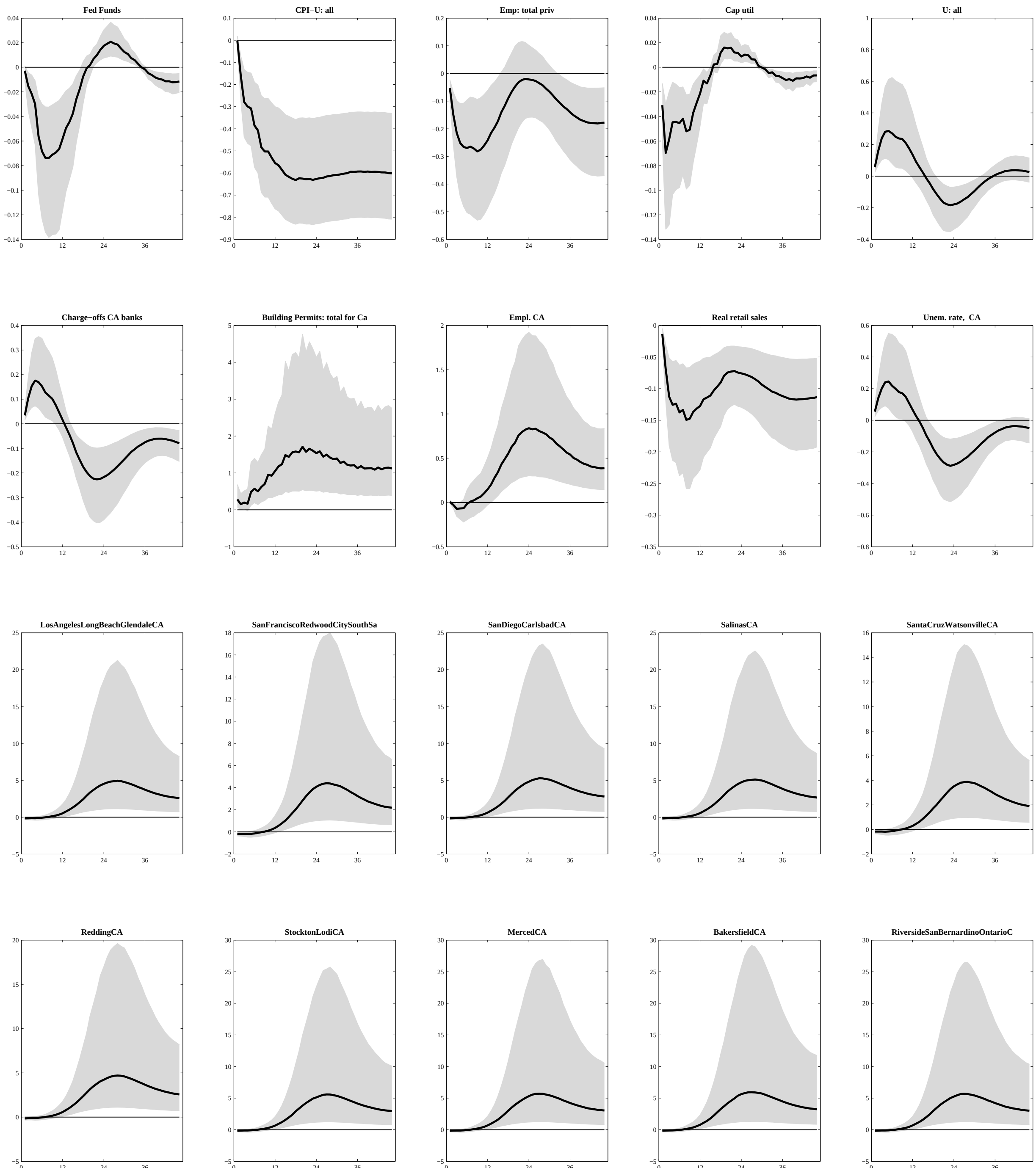


Figure 7: Impulse responses of the observed variables following an AD shock (baseline model).



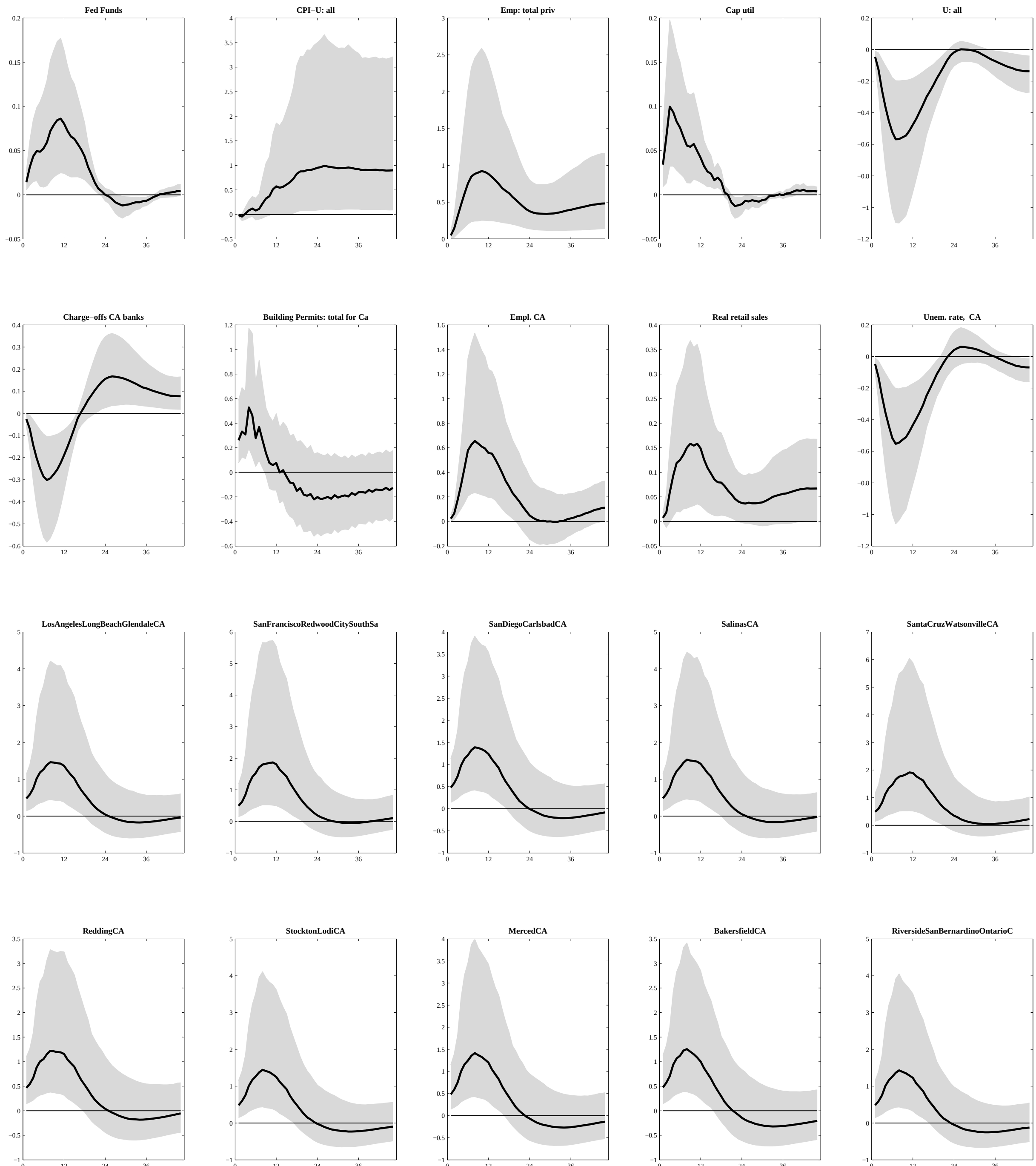
The figure displays the responses of a subset of observed variables from the panel due to a one standard deviation shock. Horizontal axis is measured in quarters and the vertical axis in standard deviation.

Figure 8: Impulse responses of the observed variables following a positive Californian housing supply shock (baseline model).



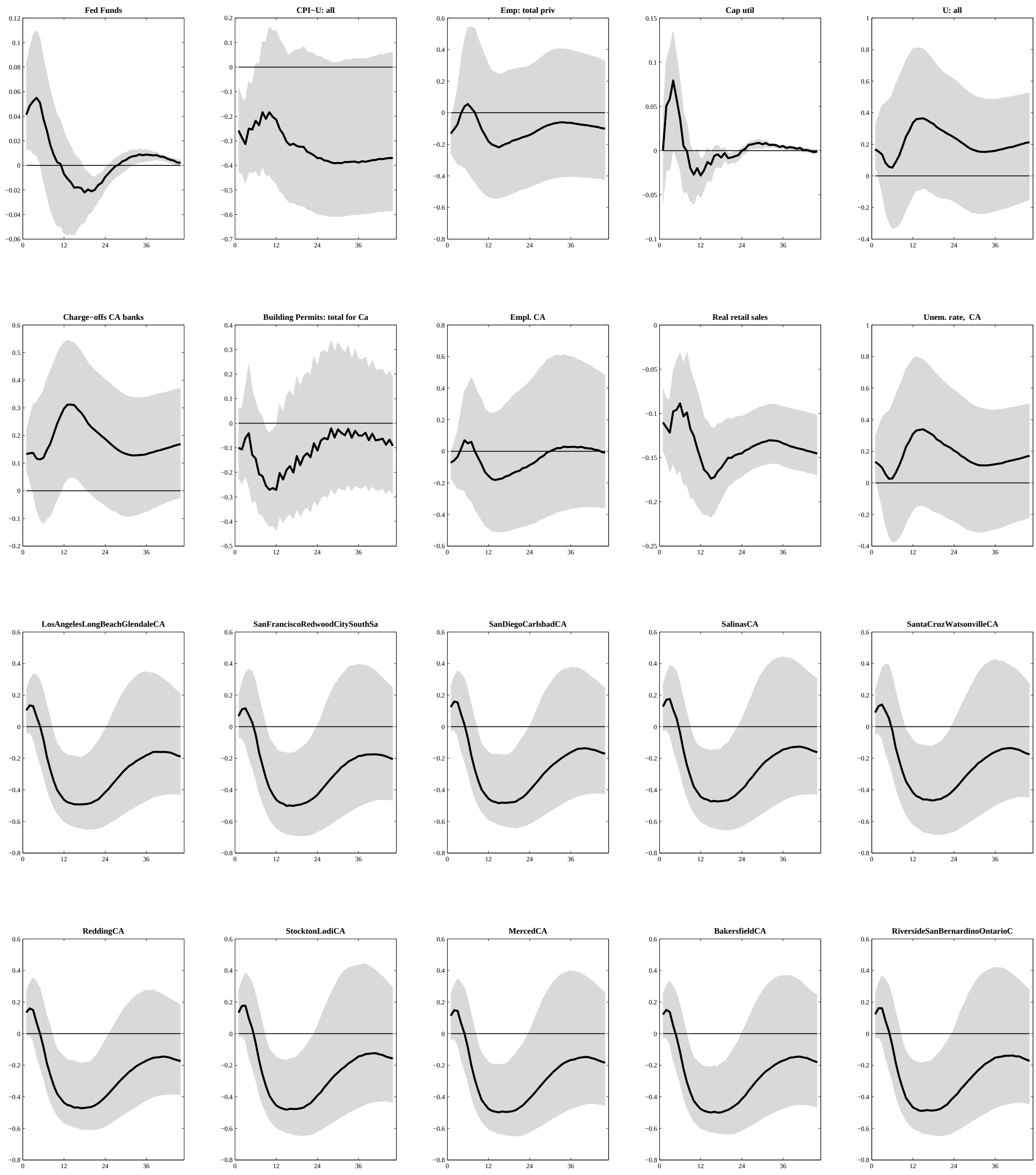
The figure displays the responses of a subset of observed variables from the panel due to a one standard deviation shock. Horizontal axis is measured in quarters and the vertical axis in standard deviation.

Figure 9: Impulse responses of the observed variables following a positive Californian housing demand shock (baseline model).



The figure displays the responses of a subset of observed variables from the panel due to a one standard deviation shock. Horizontal axis is measured in quarters and the vertical axis in standard deviation.

Figure 10: Impulse responses of the observed variables following a contractionary monetary policy shock (baseline model).



The figure displays the responses of a subset of observed variables from the panel due to a one standard deviation shock. Horizontal axis is measured in quarters and the vertical axis in standard deviation.