Group Local Projections*

Jiaming Huang

Universitat Pompeu Fabra and Barcelona School of Economics

November 17, 2023

Abstract

This paper considers the estimation of heterogeneous impulse responses in large panels. I introduce an efficient data-driven clustering methodology for grouping heterogeneous responses within the local projection-IV framework. The proposed group local projection (GLP) estimator consistently recovers the latent group structure and the group-specific impulse responses when the panel dimensions increase. Simulation evidence illustrates the reliable finite sample performance of the estimator even under misspecification of the group structure. With the GLP estimator I revisit the debate on the effects of monetary policy shocks on house prices and document significant price appreciation after a contractionary shock in an economically large cluster of MSAs in the US. Importantly, this cluster is ignored by conventional grouping criteria.

Keywords: local projections, heterogeneous impulse responses, clustering, monetary policy, house prices

JEL: C32, C33, C38, E52, R31

*Contact: jiaming.huang@upf.edu. I would like to thank Geert Mesters, Barbara Rossi, Christian Brownlees, Davide Debortoli, Jordi Galí, Priit Jeenas, Regis Barnichon, Stéphane Bonhomme, Daniel Lewis, Òscar Jordà and seminar participants at UPF Econometrics seminar, CREI Macro Lunch and EABCN Conference on Advances in Local Projections and Empirical Methods for Central Banking for comments and suggestions. I gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation (FPI Grant PRE2020-092551). All errors are my own. `MATLAB` code for implementing the GLP estimator and the Supplemental Material are available on my webpage.

1 Introduction

Individual entities, being households, firms, or regions, typically respond differently to economic shocks. Empirical studies where heterogeneous responses—in terms of signs, magnitudes, and persistence—have been documented include the effects of fiscal shocks (Beetsma and Giuliodori, 2011; Cloyne and Surico, 2017), monetary shocks (Ottonello and Winberry, 2020; Cloyne, Ferreira and Surico, 2020), and income shocks (Arellano, Blundell and Bonhomme, 2017; Baker, 2018), among many others. The documented heterogeneity is crucial for understanding the transmission channels and distinguishing macroeconomic models (Kaplan and Violante, 2018). Consistently capturing the heterogeneity among the responses is thus of utmost importance.

There are two popular approaches to capturing the heterogeneity in panel data models. The first approach allows for unit-specific coefficients (e.g., Pesaran and Smith, 1995). However, this comes with a significant efficiency loss and wide confidence bands—notably in studies with limited time series observations—implying that policy recommendations are difficult to determine (e.g., Schmitt-Grohé and Uribe, 2018). The second approach defines a priori criteria to classify entities with grouped coefficients (Cloyne, Ferreira and Surico, 2020; Ando and Bai, 2015; Bester and Hansen, 2016), which faces the distinct downside that important heterogeneity may be missed or incorrectly attributed to the assigned variable in the case of endogenous assignment. Consequently, many studies simply concentrate on the average responses and ignore potential heterogeneity (e.g., Arezki, Ramey and Sheng, 2017; Acemoglu et al., 2019).

This paper proposes a data-driven methodology that strikes a balance between unit-specific and common impulse responses (hereafter IRs). Within the panel local projection instrumental variable (LP-IV) framework (Jordà, Schularick and Taylor, 2015), I develop a *group local projection* (GLP) estimator that recovers heterogeneous IRs by grouping individual responses using clustering methods. Specifically, the GLP iterates between estimating IRs and updating group membership to find the optimal group partitioning and IR estimates. Under the assumption of strong instruments, I show that the GLP consistently recovers the latent group structure and the group-specific IRs. Moreover, I show that it allows for standard inference on the IRs when the panel dimensions become large.

An important tuning parameter for all clustering methods is the choice of the number of groups. Within the GLP framework, I propose an information criterion in the spirit of Bai and Ng (2002) that consistently selects the number of groups. Besides, I find the distinctiveness of the IR estimates a practically useful qualitative indicator (Theodoridis and Koutroumbas, 2009, Chapter 16).

Simulating data from a dynamic panel data model, I show that the GLP estimator performs satisfactorily in finite samples in terms of the classification accuracy, the length of the confidence bands, and the root mean squared error. Moreover, the proposed information criterion correctly selects the group number as the sample size increases, and throughout, the estimation errors associated with the wrong group number remain moderate. Interestingly, even when the true IRs do not admit a group structure, the GLP estimates are often more precise than unit-specific and common IR estimates.

With the GLP estimator in hand, I revisit the literature on the responses of house prices to monetary policy shocks. The existing literature has predominantly focused on estimating aggregate effects at the country level (Jarociński and Smets, 2008; Iacoviello and Neri, 2010; Jordà, Schularick and Taylor, 2015). Some have considered imposing a pre-defined group structure on the disaggregated data (e.g., Del Negro and Otrok, 2007). The conventional finding is that house prices decrease following contractionary monetary shocks. However, I show that this result is largely driven by the inability of existing methods to identify the latent groups correctly.

Specifically, I consider a large panel of 382 MSAs from 1991:1 to 2007:12 in the US. The monetary shocks are identified by using as instruments the high frequency market surprises that are robust to the Fed's information effect (Miranda-Agrippino and Ricco, 2021). The GLP procedure detects three latent groups of housing responses, among which an economically large cluster—6.3% of MSAs accounting for 13.1% of total personal income in our sample—experiences housing booms after an unexpected interest rate hike. In stark contrast, 41.4% of MSAs, which account for only 29% of total personal income, suffer from housing busts of a similar scale. Hence, curbing house prices through monetary policy is problematic given the indeterminate IRs (Galí and Gambetti, 2015). Moreover, conventional grouping criteria fail to recover the documented patterns due to rich within-group heterogeneity, e.g., in terms of income and location.

The remainder of this paper is organized as follows. I continue this introduction by carefully relating the group local projection estimator to the existing literature. Section 2 describes the local projection model and establishes the notation for grouping. Section 3 introduces the GLP estimator for which the asymptotic properties are developed in Section 4. Section 5 discusses the criteria for determining the number of groups. Section 6 reports simulation results and Section 7 presents the empirical application. Section 8 concludes. All proofs as well as additional simulation and empirical results are relegated to the Supplemental Material.

Relation to the literature

The GLP builds on a large literature on clustering latent group structures in panel data. This paper is closest to Lin and Ng (2012) and Sarafidis and Weber (2015) who estimate grouped regression coefficients using K-means clustering. The GLP complements these papers by (a) extending their analysis to allow for endogenous variables and (b) developing an inference theory for the proposed GLP estimator.

Further, to address the incidental parameter problem, Bonhomme and Manresa (2015) and Bonhomme, Lamadon and Manresa (2022) propose to use K-means clustering to group fixed effects. They derive the asymptotic properties under strict exogeneity. The GLP extends their framework to allow for the clustering of heterogeneous coefficients of multiple endogenous variables in a GMM framework. This extension is crucial for macroeconomic studies that commonly use local projection-IV methods, as the variables of interest are, almost exclusively, endogenously determined in these settings (Stock and Watson, 2018). Moreover, the GMM objective function provides a flexible way to weigh information across different moment restrictions, which is particularly useful for grouping impulse responses that are known to be noisier at large horizons (Barnichon and Brownlees, 2019).[1]

Alternative methods for grouping heterogeneous IRs include the class of dynamic factor models and vector autoregressive models; see Stock and Watson (2016) for a recent review. Prominent examples are FAVAR (Boivin, Giannoni and Mihov, 2009), Panel VAR (Canova, Ciccarelli and Ortega, 2012), and Global VAR (Dees et al., 2007). However, these models can lead to misleading results because of the parametric restrictions required to reduce the VAR dimension. In particular, I show in the empirical application that FAVAR cannot recover the positive IRs or the group patterns revealed by the GLP since the panel of house prices has no strong factor structure. To capture enough variations in the data, the number of factors would be so large that the resulting VAR entails large estimation uncertainty due to the curse of dimensionality.

In this paper, I extend the K-means clustering approach to the local projection literature following Lin and Ng (2012) and Bonhomme and Manresa (2015). Alternatively, one could consider extending penalized regression methods such as the C-LASSO procedure proposed by Su, Shi and Phillips (2016) and the panel-CARDS approach of Wang, Phillips and Su (2018), to group impulse responses. The GLP procedure has several advantages. First, the GLP admits analytical solutions and can be implemented with a simple iterative algorithm, yielding computational efficiency. Second, the GLP requires only a single parameter input, i.e., the number of groups, for which both statistical and economic criteria are available.[2]

[1]More distantly related are studies that use predetermined criteria to group coefficients (e.g., Ando and Bai, 2015; Bester and Hansen, 2016; Cloyne, Ferreira and Surico, 2020).

[2]In contrast, LASSO-type estimators involve two hyperparameters to (i) shrink the IRs and (ii) determine the number of groups. A common practice is to choose the tuning parameters based on cross-validation. However, it is difficult to construct training/test samples for clustering tasks (Theodoridis and Koutroumbas, 2009).

2 Local projections and grouping

In this section, I introduce the panel local projection model and show how a group structure can be imposed on the impulse responses. To this extent, let y_t be the N-dimensional vector of outcome variables, with the ith element denoted by $y_{i,t}$, for $i = 1, \ldots, N$. The h-period ahead predictions for $y_{i,t}$ are modeled by local projections (e.g., Jordà, 2005). Specifically, I postulate that

$$y_{i,t+h} = w'_{i,t}\alpha_{i,h} + \epsilon_{i,t+h} \ , \qquad h = 0, \ldots, H \ , \tag{1}$$

where $w_{i,t}$ is the $K \times 1$ vector of possibly endogenous variables whose influence is measured by the impulse responses $\{\alpha_{i,h}\}_{h=0}^{H}$, and $\epsilon_{i,t+h}$ is the error term. The parameters $\alpha_{i,h}$ are allowed to vary across entities i and horizons h. To estimate $\alpha_{i,h}$ I assume that there exists an $L \times 1$ vector of instruments $z_{i,t}$ with $L \geq K$. Popular examples of model (1) include the estimation of the effects of monetary policy to firm investment (Ottonello and Winberry, 2020), where $y_{i,t+h}$ is the h-period ahead investment growth of firm i, and monetary shocks are identified by instrumenting the federal funds rate by high-frequency monetary surprises (e.g., Gürkaynak, Sack and Swanson, 2005; Miranda-Agrippino and Ricco, 2021).

In principle, we can estimate model (1) equation-by-equation (e.g., for each i and h) using instrumental variable methods (hereafter, individual LP-IV). However, when the time series dimension is modest, this approach typically leads to large confidence bounds for $\{\alpha_{i,h}\}_{h=0}^{H}$, which makes policy recommendations difficult. Specifically, in most macro studies the total number of periods T will be around a few hundred (e.g., Jordà, Schularick and Taylor, 2015).

In many applications, it is often reasonable to assume some commonality across the responses of individual entities. To this extent, I partition $w_{i,t} = (x'_{i,t}, c'_{i,t})'$ and assume that the impulse responses to the variables of interest $x_{i,t}$ can be summarized by G distinct groups, such that the individuals within the same group share the same impulse responses.[3]

[3] In principle, the model could be adjusted to allow group membership to vary across the horizons of the impulse responses. Doing so nevertheless generates irregular IR shapes that are difficult to be reconciled with existing macroeconomic theories (e.g., Woodford, 2003).

Note that both $x_{i,t}$ and $c_{i,t}$ may include exogenous and endogenous variables.

Formally, let $g_i \in \{1, \ldots, G\}$ denote the group membership of individual i and $\beta_{g_i,h}$ the group impulse responses. The group partition is summarized by an $N \times 1$ vector $\gamma = (g_1, \ldots, g_N)' \in \mathcal{G}$, where $\mathcal{G}$ is the set of all possible partitions. The group local projection model is then given by

$$y_{i,t+h} = x'_{i,t}\beta_{g_i,h} + c'_{i,t}\phi_{i,h} + \epsilon_{i,t+h} \ , \qquad h = 0, \ldots, H \ , \tag{2}$$

which is identical to the standard local projection model (1) except that for the subset of variables $x_{i,t}$ the impulse responses of interest are grouped. When $G = N$, assuming groups are non-empty, the models are identical. Note that the setup is general since researchers are free to specify variables in $x_{i,t}$ whose responses are grouped, and place the remaining variables in $c_{i,t}$ whose responses may or may not follow a group structure.

The goal is to develop a method that consistently recovers the group partition γ and the group-specific impulse responses $\{\beta_{g_i,h}\}$.

3 The Group Local Projection estimator

This section introduces the *group local projection* (GLP) estimator. I start by defining some convenient notation. Let $\beta_{g_i,h} \in \Theta$ and $\phi_{i,h} \in \Phi$ be group- and individual-level impulse responses at horizon h. Denote the collection of parameters as $\beta_{g_i} = (\beta'_{g_i,0}, \ldots, \beta'_{g_i,H})'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_G)$, $\phi_i = (\phi_{i,0}, \ldots, \phi_{i,H})'$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$. The corresponding parameter space is written compactly as $\boldsymbol{\beta} \in \Theta_G$ and $\boldsymbol{\phi} \in \Phi_N$. I define the set of entities in group j as $\mathcal{S}_j = \{i \mid g_i = j\}$, with the group size given by the cardinality $N_j = |\mathcal{S}_j|$. Unless otherwise noted, $\|\cdot\|$ denotes the Euclidean norm, i.e., $\|X\|^2 = \text{tr}(XX')$ for a matrix or vector X. Finally, two collections of parameters $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_{G_1})$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{G_2})$ with possibly different G_1 and G_2, define the Hausdorff distance between them as

$$d_H(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}) = \max\left\{ \max_{g\in\{1,\dots,G_2\}} \min_{\tilde{g}\in\{1,\dots,G_1\}} \|\tilde{\beta}_{\tilde{g}} - \beta_g\|, \max_{\tilde{g}\in\{1,\dots,G_1\}} \min_{g\in\{1,\dots,G_2\}} \|\tilde{\beta}_{\tilde{g}} - \beta_g\| \right\} . \tag{3}$$

The GLP estimator aims to consistently recover the underlying group structure γ and the impulse response matrix $\boldsymbol{\beta}$. The individual-specific impulse responses $\boldsymbol{\phi}$ are not of explicit interest. Formally, I define the GLP estimator as

$$\{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}\} = \underset{\boldsymbol{\beta}\in\Theta_G, \boldsymbol{\phi}\in\Phi_N, \gamma\in\mathcal{G}}{\operatorname{argmin}} \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) \ , \tag{4}$$

where the objective function $\hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma)$ is

$$\hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) = \frac{1}{N} \sum_{i=1}^{N} \sum_{h=0}^{H} \hat{Q}_{iTh}(\beta_{g_i,h}, \phi_{i,h}), \quad \hat{Q}_{iTh}(\beta_{g_i,h}, \phi_{i,h}) = \hat{m}'_{i,h} \hat{\Omega}_{i,h} \hat{m}_{i,h}, \tag{5}$$

with $\hat{m}_{i,h} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t}(y_{i,t+h} - x'_{i,t}\beta_{g_i,h} - c'_{i,t}\phi_{i,h})$ and $\hat{\Omega}_{i,h}$ some $L \times L$ weighting matrix.

The GLP estimator is based on the minimization of the multiple-equation GMM objective function (5). Besides being able to handle endogenous variables, this also allows incorporating different weights for different units i and horizons h through $\hat{\Omega}_{i,h}$, which is particularly relevant in the local projection setup where the estimation errors tend to increase for longer horizons. In practice, we can set $\hat{\Omega}_{i,h} = \hat{V}_{i,h}^{-1}$ where $\hat{V}_{i,h}$ is the Newey and West (1987) variance estimate of $\hat{m}_{i,h}$ with $H+1$ lags, i.e.,

$$\hat{V}_{i,h} = \sum_{t=1}^{T} e^2_{i,t+h} z_{i,t} z'_{i,t} + \sum_{l=1}^{H+1} (1 - \frac{l}{H+2}) \sum_{t=l+1}^{T} e_{i,t+h} e_{i,t+h-l} (z_{i,t} z'_{i,t-l} + z_{i,t-l} z'_{i,t}) \tag{6}$$

with $e_{i,t+h} = y_{i,t+h} - w'_{i,t}\hat{\alpha}_{i,h}$ the residuals from individual LP-IV regressions. The performance of various weighting matrices can be found in Supplementary Material S2.3.

Note also that the estimator defined in (4) is by construction different from the conventional panel GMM estimator (e.g., Hayashi, 2000, Chapter 4). In particular, the resulting estimator is subject to the incidental parameter bias (due to the presence of $\phi_{i,h}$) even when the latent groups γ are known and $N/T \to \kappa \in [0, \infty)$; see Fernández-Val and Lee (2013) and Su, Shi and Phillips (2016) for detailed discussions on the bias. The advantage of formulation (5) is that the minimization problem can be solved by an iterative procedure in the spirit of the K-means algorithm, whereas alternative specifications can only be solved simultaneously with respect to $(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma)$; see e.g., Ramaswamy et al. (1993).

In particular, we can minimize (4) by iterating between two steps: (i) solving the group assignment problem and (ii) solving the IR estimation problem. The estimation procedure is summarized in the following algorithm.

Algorithm: Group Local Projection

1. *Initialization:* Obtain an initial guess for the IRs, $\hat{\boldsymbol{\beta}}^{(0)} = (\hat{\beta}_1^{(0)}, \ldots, \hat{\beta}_G^{(0)})$ and $\hat{\boldsymbol{\phi}}^{(0)} = (\hat{\phi}_1^{(0)}, \ldots, \hat{\phi}_N^{(0)})$;

2. *Assignment:* Given $\hat{\boldsymbol{\beta}}^{(r)}$ and $\hat{\boldsymbol{\phi}}^{(r)}$, solve

$$\hat{g}_i^{(r+1)}(\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\phi}}^{(r)}) = \underset{g_i \in \{1,\ldots,G\}}{\operatorname{argmin}} \sum_{h=0}^{H} \hat{Q}_{iTh}\left(\hat{\beta}_{g_i,h}^{(r)}, \hat{\phi}_{i,h}^{(r)}\right) \; ; \tag{7}$$

3. *Coefficient updating:* Given the group partition $\hat{\gamma}^{(r)} = (\hat{g}_1^{(r)}, \ldots, \hat{g}_N^{(r)})$, solve

$$\left\{\hat{\boldsymbol{\beta}}^{(r+1)}, \hat{\boldsymbol{\phi}}^{(r+1)}\right\} = \underset{\boldsymbol{\beta} \in \Theta_G, \boldsymbol{\phi} \in \Phi_N}{\operatorname{argmin}} \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \hat{\gamma}^{(r)}) \; ; \tag{8}$$

4. *Convergence:* If $\operatorname{dif}(r) \leq \epsilon_{tol}$ stop, where

$$\operatorname{dif}(r) = \max\left\{ \left| \hat{Q}_{NT}(\hat{\boldsymbol{\beta}}^{(r+1)}, \hat{\boldsymbol{\phi}}^{(r+1)}, \hat{\gamma}^{(r+1)}) - \hat{Q}_{NT}(\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\phi}}^{(r)}, \hat{\gamma}^{(r)}) \right|, \frac{\|\hat{\boldsymbol{\beta}}^{(r+1)} - \hat{\boldsymbol{\beta}}^{(r)}\|}{\|\hat{\boldsymbol{\beta}}^{(r)}\| + 0.001} \right\} ,$$

   else set $r = r + 1$ and go to Step 2.

Overall, the GLP algorithm generalizes the conditional K-means clustering algorithm of Lin and Ng (2012), where the main novelty is the GMM objective function.

Step 1 initializes the algorithm. In practice, it is recommended to use a number of different initial starting matrices $\hat{\boldsymbol{\beta}}^{(0)}$ by randomly drawing G entities (out of N) and computing IRs for each entity using individual LP-IV. Moreover, we fix $\hat{\boldsymbol{\phi}}^{(0)} = (\hat{\phi}_1, \ldots, \hat{\phi}_N)$ where $\hat{\phi}_i$ is the individual LP-IV estimates for unit i. In simulations, I find that using 50 random initializations works well.[4]

Given the initial guess $\hat{\boldsymbol{\beta}}^{(0)}$ and $\hat{\boldsymbol{\phi}}^{(0)}$, the GLP iterates between steps 2 and 3 until convergence. For the assignment step 2, I evaluate the individual objective function for $i = 1, \ldots, N$ at *each* group coefficient, i.e., obtaining $\hat{Q}_{iTh}\left(\hat{\beta}^{(r)}_{1,h}, \hat{\phi}^{(r)}_{i,h}\right), \ldots, \hat{Q}_{iTh}\left(\hat{\beta}^{(r)}_{G,h}, \hat{\phi}^{(r)}_{i,h}\right)$, and I assign unit i to group j such that $\sum_h \hat{Q}_{iTh}\left(\hat{\beta}^{(r)}_{j,h}, \hat{\phi}^{(r)}_{i,h}\right)$ is the minimum.

The coefficient updating step 3 can be carried out in closed form. A simple calculation shows that for all $j = 1, \ldots, G$ and $h = 0, \ldots, H$

$$\hat{\beta}^{(r+1)}_{j,h} = \left(\sum_{i \in \hat{S}^{(r)}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zx,i}\right)^{-1} \left(\sum_{i \in \hat{S}^{(r)}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zy,i,h}\right) , \tag{9}$$

and for all $i = 1, \ldots, N$

$$\hat{\phi}^{(r+1)}_{i,h} = \left(\overline{d}'_{zc,i} \hat{\Omega}_{i,h} \overline{d}_{zc,i}\right)^{-1} \overline{d}'_{zc,i} \hat{\Omega}_{i,h} \left(\overline{d}_{zy,i,h} - \overline{d}_{zx,i} \hat{\beta}^{(r+1)}_{\hat{g}_i,h}\right) , \tag{10}$$

where $\bar{d}_{zx,i} = \frac{1}{T}\sum_{t=1}^T z_{i,t}x'_{i,t}$, $\bar{d}_{zc,i} = \frac{1}{T}\sum_{t=1}^T z_{i,t}c'_{i,t}$, $\bar{d}_{zy,i,h} = \frac{1}{T}\sum_{t=1}^T z_{i,t}y_{i,t+h}$ and $\overline{M}_{zc,i,h} = \hat{\Omega}_{i,h} - \hat{\Omega}_{i,h}\bar{d}_{zc,i}(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i})^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}$. The closed-form solutions in (9) and (10) make the GLP attractive from a computational viewpoint.

Finally, step 4 assesses whether the algorithm has converged. The convergence criterion dif(r) is borrowed directly from Su, Shi and Phillips (2016) and I use $\epsilon_{tol} = 10^{-6}$ as a tolerance level in practice. Simulation shows little variation across different stopping rules.

To sum up, the GLP estimator can be easily implemented using existing tools. The procedure is computationally efficient because the assignment step amounts to simply eval-

[4]For more discussion on K-means initialization see Maitra, Peterson and Ghosh (2010) for an excellent overview.

uating GMM objective functions and the coefficient updating step can be solved in closed form. An implicit hyperparameter in the above algorithm is the number of groups G, which is discussed in greater detail in Section 5.

4 Asymptotic theory

This section develops the asymptotic properties of the GLP estimator when the cross-section and the time series dimensions become large. First, I discuss the conditions under which the latent group structure γ and the impulse responses $\boldsymbol{\beta}$ are consistently estimated. Second, I derive the limiting distribution of the impulse responses for large N and T. Throughout, the true group membership of individual i is denoted by g_i^0, the true number of groups by G^0, and the true impulse responses by $\boldsymbol{\beta}^0$.

4.1 Consistency

I begin by laying out the assumptions needed for consistency.

Assumption 1 (GMM)**.**

A. The parameter space Θ *and* Φ *are compact.*

B. For all i*,* $\{(y_{i,t}, w'_{i,t}, z'_{i,t})' \colon 1 \leqslant t \leqslant T\}$ *is a strictly stationary* α*-mixing sequence of random vectors with mixing coefficients* $a_i(\cdot)$ *such that* $a(\cdot) = \max_i a_i(\cdot)$ *satisfies* $a(\tau) \leqslant \exp\left(-c_1 \tau^{c_2}\right)$ *where* c_1, c_2 *are positive constants.*

C. For each i *and* h*, there exists some* $c_3, c_4 > 0$ *such that for each element* x *of* $\{(z_{i,t} w'_{i,t}, z_{it}\epsilon_{i,t+h})\}$*, we have* $\mathbb{P}(|x - \mathbb{E}[x]| > v) \leqslant \exp\left(1 - (v/c_3)^{c_4}\right)$*.*

D. For all i, t*,* $\mathbb{E}[z_{i,t} w'_{i,t}]$ *has full column rank and* $\mathbb{E}|z_{i,t,j} w_{i,t,k}|^{4(1+\delta)} < \infty$ *for* $j = 1, \ldots, L$ *and* $k = 1, \ldots, K$ *and some* $\delta > 0$*.*

E. For all i, t and h, $\mathbb{E}[z_{i,t}\epsilon_{i,t+h}] = 0$ and $\mathbb{E}|z_{i,t,j}\epsilon_{i,t+h}|^{4(1+\delta)} < \infty$ for $j = 1, \ldots, L$ and some $\delta > 0$.

F. For all h, there exists a deterministic sequence of symmetric positive definite matrices $\{\Omega_{i,h}\}_{i=1}^{N}$ such that for any $\nu > 0$, we have $\mathbb{P}(\|\hat{\Omega}_{i,h} - \Omega_{i,h}\| \geq \nu) = o(T^{-\delta})$ for all $\delta > 0$.

G. The maximal horizon H is finite.

Assumptions 1.A-1.F define a set of primitive conditions for the consistency of the GMM estimator (e.g., White, 2001, Chapter 3). Specifically, part 1.B assumes stationary strong mixing for each individual time series. Since we allow for individual-level parameters $\phi_{i,h}$, I follow Hahn and Kuersteiner (2011) and assume that $\phi_{i,h}$ are nonrandom, and the data are strong mixing conditional on their realization. Together with the tail probability condition 1.C, the exponential decay rate facilitates the use of Bernstein-type inequalities (e.g. Rio, 2017, Chapter 6). It is easy to extend the analysis with a slower decay rate at the cost of stronger moment bounds (e.g., Cheng, Schorfheide and Shao, 2019).

Parts 1.D and 1.E impose standard assumptions on the instrumental variables, relevance and exogeneity, and require the existence of $4(1+\delta)$ moments for the products $z_{i,t,j}w_{i,t,k}$ and $z_{i,t,j}\epsilon_{i,t+h}$. Assumption 1.F allows for flexible weighting across units and horizons, which is crucial in the current setup as IRs are uninformative and noisy for longer horizons. Combined with The rate condition can be relaxed to uniform convergence $\mathbb{P}(\sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\| \geq \nu) = o(1)$, at the cost of slower convergence of the misclassification error[5]. The final condition 1.G assumes that the maximal horizon of the impulse responses is finite, which is justified for most of the applications in macroeconomics where the responses converge to zero in the long run.

Intuitively, Assumption 1 ensures that the estimated group impulse responses converge in probability to a well-defined limit *given any group partition* γ. In the current setup,

[5]It is easy to verify that the uniform convergence of the objective function $\hat{Q}_{NT}$ to its population counterpart is determined by $\sup_i \|\overline{d}_{zw,i} - d_{zw,i}\|$, $\sup_i \|\overline{d}_{z\epsilon,i,h}\|$ and $\sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|$, which in turn determines the convergence of misclassification errors; see (124) in the Supplemental Material.

this assumption is insufficient as the latent group structure is unknown, and so additional assumptions are needed to infer the group membership. First we specify the latent group structure as follows.

Assumption 2 (Group Structure)**.**

A. For all $j = 1, \dots, G^0$, $\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{g_i^0 = j\} = \pi_j > 0$.

B. For all $j \neq k$, $\sum_{h=0}^{H} \|\beta_{j,h}^0 - \beta_{k,h}^0\| > 0$.

C. The true number of groups G^0 *is finite and known.*

Here part 2.A requires the group size to increase as N goes to infinity, which rules out the possibility of outliers that could have a nontrivial impact on the estimation. Part 2.B imposes that the true impulse responses of different groups are separable, and part 2.C requires that the number of groups is finite and known. Section 5 proposes an information criterion that consistently selects G^0. Extensions to the case where G^0 varies, e.g., across time and sample sizes, are left for future research.

Taken together, Assumptions 1-2 are comparable to Assumptions 1-2 in Bonhomme and Manresa (2015), but with the important difference that they pertain to a multiple-equation setting where the explanatory variables can be endogenous and valid instruments are available for inference. The following theorem summarizes the consistency of the GLP.

Theorem 1 (Consistency)**.** *Under Assumptions 1-2, as* $N, T \to \infty$*, we have for all* $\delta > 0$

$$\mathbb{P}\left(\sup_{i\in\{1,\dots,N\}} \mid \hat{g}_i - g_i^0 \mid > 0\right) = o(1) + o(NT^{-\delta}) \tag{11}$$

$$\hat{\beta}_j - \tilde{\beta}_j = o_p(NT^{-\delta}), \qquad j = 1, \dots, G^0\ . \tag{12}$$

where $\tilde{\boldsymbol{\beta}}$ is the infeasible estimator defined by $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}) = \underset{\beta,\phi}{\text{argmin}}\, \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma^0)$.

Theorem 1 contains two parts. The first part provides an upper bound of misclassification errors. The bound contains two parts. The first part is determined by $\mathbb{P}(\boldsymbol{\beta} \notin \mathcal{N}_\eta)$, the probability that $\boldsymbol{\beta}$ lies outside some neighborhood of the true parameters measured by Hausdorff distance; the second part is the misclassification probabilities given that $\boldsymbol{\beta} \in \mathcal{N}_\eta$.

The second part states that the GLP estimator is asymptotically equivalent to the infeasible estimator with known group structure. Our convergence rate is slower than the $o_p(T^{-\delta})$ in Bonhomme and Manresa (2015), because we allow for a large vector of nuisance parameter $\boldsymbol{\phi}$. Moreover, due to the presence of $\boldsymbol{\phi}$, even the infeasible estimator is asymptotically biased, as next section shows. To avoid the bias, Su, Shi and Phillips (2016) proposes a "post-lasso" estimator where conditional on estimated groups we can use the fully-pooled GMM objective function. However, doing so implicitly assumes that all parameters in $\boldsymbol{\phi}$ are grouped and with identical grouping structure, which is unlikely to hold in practice.

4.2 Asymptotic normality

Theorem 1 indicates that as T goes to infinity, the probability of misclassification converges to zero. To derive asymptotic distribution under this case, I impose the following assumption:

Assumption 3.

A. *For all i, $\{(y_{i,t}, w_{i,t}, z_{i,t})\}$ are independent and identically distributed across i.*

B. *For each $j = 1, \dots, G^0$, we have $N_j/T \to \kappa_j \in [0, \infty)$ as N and T go to infinity.*

Assumption 3 contains two parts. First, part 3.A imposes cross-sectional independence to facilitate the application of a central limit theorem (Fernández-Val and Lee, 2013; Su, Shi and Phillips, 2016; Wang, Phillips and Su, 2018). It is possible to relax this assumption by allowing for cross-sectional dependence; see for example Huang et al. (2021). Second, condition 3.B is a standard large N, T condition in the panel data literature (Hahn and Kuersteiner, 2011) where both the cross-section and time series dimension grow at the same rate. Notice that the GLP estimator suffers from the incidental parameter bias even with large T, which is formalized in the following theorem.

Theorem 2. *Under Assumptions 1-3, as N_j and T goes to infinity*

$$\sqrt{N_j T}(\hat{\beta}_{j,h} - \beta^0_{j,h}) \xrightarrow{\text{d}} \Sigma^{-1}_{j,h} N(\sqrt{\kappa_j}\mathcal{B}, \Psi_{j,h}), \quad j = 1, \dots, G^0 \tag{13}$$

where $\mathcal{B}$ is the asymptotic bias with analytical formula in Section S1, and

$$\Sigma_{j,h} = \underset{N,T\to\infty}{\text{plim}} \frac{1}{N_j} \sum_{i \in S^0_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zx,i}, \quad \Psi_{j,h} = \underset{N,T\to\infty}{\text{plim}} \, Var\left(\sqrt{\frac{T}{N_j}} \sum_{i \in S^0_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{z\epsilon,i,h}\right)$$

The above theorem states that we can carry out standard inference for each group IR under large T. In practice, we can estimate $\Sigma_{j,h}$ and $\Psi_{j,h}$ consistently by

$$\hat{\Sigma}_{j,h} = \frac{1}{N_j} \sum_{i \in \hat{\mathcal{S}}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zx,i}, \quad \hat{\Psi}_{j,h} = \frac{1}{N_j} \sum_{i \in \hat{\mathcal{S}}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \hat{V}_{i,h} \overline{M}_{zc,i,h} \overline{d}_{zx,i} \ , \tag{14}$$

where $\hat{V}_{i,h}$ is the variance estimate of $\overline{d}_{z\epsilon,i,h}$. We can use the HAC-robust estimator (6) with the GLP residuals $e_{i,t+h} = y_{i,t+h} - x_{i,t}\hat{\beta}_{\hat{g}_i,h} - c_{i,t}\hat{\phi}_{i,h}$. In the simulation section 6, I study the finite sample performance of the asymptotic approximation provided by Theorem 2 and find that for all $N_j/T \leq 1$ the approximation is very good.

5 Determining the number of groups

In this section, I propose an information criterion (hereafter, IC) that correctly selects the number of groups as the panel dimensions increase. In particular, I define

$$\text{IC}(G) = \hat{Q}_{NT,G} + \varrho_{N,T} \hat{Q}_{NT,G_{max}} G(H+1) \tag{15}$$

where G_{max} is the maximal number of groups, $\varrho_{N,T}$ a tuning parameter that controls the strength of penalty, and $\hat{Q}_{NT,G}$ the objective function (5) evaluated at the estimated parameters with G groups. The resulting estimator for the number of groups is given by

$$\hat{G}^{\mathrm{IC}} = \underset{G=1,\ldots,G_{max}}{\operatorname{argmin}} \ \mathrm{IC}(G). \tag{16}$$

Compared with conventional information criteria (e.g., Su, Shi and Phillips, 2016), the proposed IC is based on the GMM objective function $\hat{Q}_{NT,G}$ instead of the residual sum of squares because the GLP by construction does not aim to minimize the residual sum of squares, which can be an arbitrary function of the supplied number of groups. In contrast, $\hat{Q}_{NT,G}$ is monotonically non-increasing in the number of groups. In a similar vein, the penalty is scaled by $\hat{Q}_{NT,G_{max}}$. To show the consistency of $\hat{G}^{\mathrm{IC}}$ I make the following assumptions.

Assumption 4 (IC)**.**

A. As N and T go to infinity, $\min_{1\leq G<G^0} \inf_{\gamma\in\mathcal{G}(G)} \hat{Q}_{NT,G} \xrightarrow{\text{p}} \underline{Q} > Q^0$ where

$$Q^0 = \underset{N,T\to\infty}{\text{plim}} \frac{1}{N}\sum_i \sum_h \hat{m}^{0\prime}_{i,h}\hat{\Omega}_{i,h}\hat{m}^0_{i,h} \ , \quad \hat{m}^0_{i,h} = \frac{1}{T}\sum_t z_{i,t}(y_{i,t+h} - x'_{i,t}\beta^0_{g^0_i,h} - c'_{i,t}\phi^0_{i,h}) \ .$$

B. $\varrho_{N,T} \to 0$ and there exists some constant $a > 0$ such that $\varrho_{N,T}T^a \to \infty$.

Overall, the assumption is similar to Assumptions B5-B6 in Su, Shi and Phillips (2016). First, part 4.A requires the objective function to be bounded below by the global minimum Q^0 when under-fitted, regardless of the group partition. Second, part 4.B disciplines the penalty strength $\varrho_{N,T}$. In simulations, I find that $\varrho_{N,T} = (NT)^{-1/4}$ works well. Given the assumption, I establish consistency in the following proposition.

Proposition 1. Under Assumptions 1-2 and 4, for $N, T \to \infty$ we have

$$\mathbb{P}\left(\min_{G\neq G^0} IC(G) > IC(G^0)\right) \xrightarrow{\text{p}} 1 \ . \tag{17}$$

The proposition establishes the consistency of the proposed information criterion. Nevertheless, when T is limited it tends to under-select the number of groups; see Section 6.3

and Supplementary Material S2.2.

For such cases the distinctiveness of the estimated IRs becomes a useful heuristic criterion. Intuitively, when the group IRs are close to each other, misclassification between groups could be large, as implied by the group separation condition 2.B. Hence, we may consider a smaller group number when IRs are inseparable (Theodoridis and Koutroumbas, 2009).

6 Simulation study

This section evaluates the finite sample performance of the GLP estimator. Specifically, I consider three settings where (i) the number of groups is known, (ii) the number of groups is unknown and determined by the information criterion (16), and (iii) there is no group structure and individual IRs are truly distinct. For each scenario, I compare the GLP estimator with the individual LP-IV (IND), the panel LP-IV (PAN), and the infeasible GLP (IGLP) where the group partition is known.[6]

6.1 Simulation design

I generate data from a dynamic panel data model given by

$$\begin{aligned} y_{i,t} &= \mu_i + \rho_{g_i} y_{i,t-1} + \delta_{g_i} x_{i,t} + \epsilon_{i,t} \\ x_{i,t} &= \mu_i + \pi \tilde{z}_{i,t} + u_{i,t} \end{aligned}, \tag{18}$$

where the individual fixed effects μ_i are drawn independently from the uniform distribution $U(0,1)$, the instrument $\tilde{z}_{i,t}$ is i.i.d. $N(0,1)$, and the error terms $\epsilon_{i,t}$ and $u_{i,t}$ are generated from a bivariate normal distribution

$$\begin{pmatrix} \epsilon_{i,t} \\ u_{i,t} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right).$$

[6]The IGLP is the (fully-pooled) panel LP-IV (Jordà, Schularick and Taylor, 2015) for each group, with standard errors clustered at the individual level (Cameron and Miller, 2015).

We are interested in estimating the effect of $x_{i,t}$ on $y_{i,t+h}$. To do so, let $\beta_{g_i,h}$ denote the horizon h impulse response for group g_i. We can recover this impulse response from the group local projection model

$$y_{i,t+h} = x'_{i,t}\beta_{g_i,h} + c'_{i,t}\phi_{i,h} + \epsilon_{i,t+h} \ , \tag{19}$$

where $c_{i,t} = (1, y_{i,t-1})'$ using $z_{i,t} = (\tilde{z}_{i,t}, 1, y_{i,t-1})'$ as instruments.[7] Overall, the DGP represents a large class of panel data models relevant for empirical studies (e.g., Acemoglu et al., 2019).

Table 1 reports the parameter values. The first two rows show different combinations of ρ_{g_i} and δ_{g_i} that control the persistence and the magnitude of the IRs respectively. For example, in Design 1 with two groups ($G^0 = 2$), the IRs share the same magnitude ($\delta_1 = \delta_2 = 3$) but have different levels of persistence ($\rho_1 = 0.2,\ \rho_2 = 0.6$). Row IV assigns 0.7 to π to ensure a strong IV setup, and row Fraction indicates the share of units in each group. To save space, I report $G^0 = 2$ and $G^0 = 3$ when the group number is known, and $G^0 = 3$ when the group number is estimated by the information criterion.

It is interesting to study the performance of the GLP estimator for the case where the impulse responses are heterogeneous, but there is no group structure. To this end, I modify DGP (18) by

$$y_{i,t} = \mu_i + \rho_i y_{i,t-1} + \delta_i x_{i,t} + \epsilon_{i,t} \ , \tag{20}$$

where the only change is that the coefficients that determine the impulse responses are individual-specific. I consider random coefficients $\rho_i \sim U(0.1, 0.9)$ and $\delta_i \sim U(1, 3)$.

[7] Alternatively, to avoid the dynamic panel bias in settings where T is small relative to N, we can take first differences and consider

$$\Delta y_{i,t+h} = \Delta x'_{i,t}\beta_{g_i,h} + \Delta y_{i,t-1}\phi_{i,h} + \epsilon_{i,t+h} \ ,$$

which can be estimated using $z_{i,t} = (\tilde{z}_{i,t-1}, y_{i,t-2})$ as instruments. I repeat the simulation exercise in Supplementary Material S2.5 with this estimator.

I evaluate the GLP using four criteria:[8] (1) the classification accuracy (AC), (2) the ratio of confidence lengths between the GLP and the IND (BR), (3) the root mean squared error (RMSE), and (4) the coverage rates. As a benchmark, the GLP confidence bands are computed according to Theorem 2, and the results of the fixed T inference are reported in Supplementary Material S2.4. Formally, these criteria are:

$$\begin{aligned}
\text{AC} &= \tfrac{1}{mN}\textstyle\sum_{r=1}^{m}\sum_{i=1}^{N}\mathbf{1}\{\hat{g}_i^{(r)} = g_i^0\} \\
\text{BR} &= \tfrac{1}{mNH}\textstyle\sum_{r=1}^{m}\sum_{i=1}^{N}\sum_{h=0}^{H}\frac{\text{GSE}^{(r)}_{\hat{g}_i,h}}{\text{SE}^{(r)}_{i,h}} \\
\text{RMSE} &= \sqrt{\tfrac{1}{mNH}\textstyle\sum_{r=1}^{m}\sum_{i=1}^{N}\sum_{h=0}^{H}(\hat{\beta}^{(r)}_{\hat{g}_i,h} - \beta^{0(r)}_{g_i^0,h})^2} \\
\text{Coverage}_h &= \tfrac{1}{mG}\textstyle\sum_{r=1}^{m}\sum_{j=1}^{G^0}\mathbf{1}\{\beta^{0(r)}_{j,h} \in CI_{j,h}\}
\end{aligned}, \tag{21}$$

where m is the number of replications.

With known group numbers, I run $m = 1000$ simulations for each parameter design and different combinations of $N = 100, 200, 300$, and $T = 100, 200, 300$. With unknown group numbers or truly distinct IRs, I run $m = 500$ simulations with eight guesses $G^{guess} = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and report the associated RMSE and the estimated group number.

6.2 Simulation results: known group number

Table 2 and Table 3 report the results. Five observations stand out. First, the classification accuracy improves dramatically as T increases, as Theorem 1 predicts. With a moderate sample size ($N = 100,\ T = 100,\ G^0 = 2$, Design 1), the GLP achieves 84.1% classification accuracy, which further rises to 96.2% with $T = 300$. To better illustrate the effectiveness, I report in Figure 1(a) a typical example of the individual IR estimates under ($N = 300,\ T = 100,\ G^0 = 2$, Design 1). The thin lines are individual IR estimates, and the thick lines indicate the maximal and the minimal estimates at each horizon. Although the individual

[8]The labels of estimated groups do not necessarily correspond to the true labels. To overcome this issue, I follow Bonhomme and Manresa (2015) and relabel the group estimates such that the accuracy is maximized, by going through all possible label permutations. The RMSE is then computed using the new labels.

IRs are highly overlapped, the GLP still achieves 84% accuracy.

Second, the GLP yields significant efficiency gains for all sample sizes. Column BR shows that the average lengths of the GLP confidence intervals closely track the infeasible counterparts, and are, in the worse case, only one-fifth of the individual LP-IV ones. Moreover, Table 3 shows that the coverage rates of the GLP bands are satisfactory when the time series dimension is large, but can be severely under-covered in short panels. This is because the GLP estimator is subject to two sources of bias: the incidental parameter bias, as 2 shows, and the misclassification bias. The former is well studied in the panel data literature and the magnitude of the bias is closely related to the N/T ratio. In contrast, Section S2.1 shows that the misclassification bias greatly alleviates as long as T gets large.

Third, the GLP leads to much lower RMSE than alternative methods for all sample sizes. Consider ($N = 300,\ T = 100,\ G^0 = 3$, Design 1) for example. For one thing, the average GLP RMSE is 36.2% (32.9%) lower than the PAN (IND) counterpart. For another, the GLP RMSE converges to the infeasible one at a rate faster than the individual LP-IV counterpart.

Fourth, the GLP performs well as we fix T and increase N. The reason is that, as noted in Section 4, T can grow polynomially slower than N. Supplementary Material S2.1 further corroborates this by the stable performance of the GLP under ($N = 1500, T = 300$). Therefore, the GLP is reliable in practically relevant sample sizes.

Finally, there is an overall deterioration in all evaluating metrics comparing $G^0 = 2$ with $G^0 = 3$, since the IRs are much less informative in the latter case; see for example Figure 1(b). Note that even in this case, the GLP still outperforms the alternatives.

Overall, the GLP estimator has excellent finite sample properties, and it outperforms both the panel LP-IV and the individual LP-IV.

6.3 Simulation results: unknown group number

Table 4 reports the RMSE of the GLP under different supplied group numbers, and Row IC shows the estimated number of groups. The table is revealing in several ways. First, the RMSE is U-shaped as the number of groups increases. Therefore, the GLP yields lower RMSE than PAN (biased) and IND (inefficient), even with the misspecified group number.

Second, the number of groups selected by the information criterion (15) converges to the true as the sample size increases. As we can see, although it may under-select when T is smaller than N, the estimated number of groups steadily converges. Supplementary Material S2.2 further confirms the result under alternative sample sizes.

Third, even when the criterion under-selects, the GLP always leads to more precise estimates than the alternatives. Consider Design 1 with ($N = 300$, $T = 100$). Although the IC chooses two groups ($\hat{G} = 2$), the associated error (0.557) is much smaller than the alternatives (0.775 by the individual LP-IV and 0.814 by the panel LP-IV).

6.4 Simulation results: no group structure

Table 5 reports the results when the data have no group structure. Interestingly, even though the true number of groups is now N, the RMSE is still U-shaped because individual IR estimates can be noisy. Consequently, the GLP produces a reasonable approximation of the true responses.

Furthermore, the IC is rather conservative: it selects fewer groups than is needed to minimize the RMSE. For example, when $N = 100$ and $T = 100$, the IC leads to two or three groups, whereas the RMSE is minimized with six groups.

7 Heterogeneous house price responses in the US

There is abundant evidence that a tightening monetary shock dampens aggregate house prices (Jarociński and Smets, 2008; Iacoviello and Neri, 2010; Jordà, Schularick and Taylor, 2015). However, recent research argues that the responses of bubbly assets are indeterminate (Galí, 2014). Given that regional housing markets are highly fragmented with potentially different bubble components (Glaeser and Nathanson, 2015), it is natural to ask: Do regional house prices respond differently to monetary shocks? And if so, in what ways?

I answer the questions through the lens of MSA-level housing markets in the US. Specifically, I estimate the following group local projections model:

$$y_{i,t+h} = x_{i,t}\beta_{g_i,h} + c'_{i,t}\phi_{i,h} + \epsilon_{i,t+h}, \quad h = 1, \ldots, 24 \; . \tag{22}$$

Here, $y_{i,t+h}$ are the cumulative house price changes for 382 MSAs from 1991:1 to 2007:12.[9] $x_{i,t}$ is the Fed Funds rates, instrumented by $z_{i,t}$, the high frequency surprises in fed fund futures that are robust to the Fed information effect (Miranda-Agrippino and Ricco, 2021). The control variable $c_{i,t}$ contains the constant term, four lags of the independent and dependent variables, current and four lags of the 30-year fixed-rate for mortgage products, the industrial production growth, the PCE inflation, and the growth of real estate loans, making the model comparable to Del Negro and Otrok (2007).[10] Supplementary Material S3.1 provides details of the data construction.

7.1 Standard LP-IV results

As a benchmark, I present the results of standard LP-IV models in Figure 2. The Newey and West (1987) HAC robust F-statistic is 57.6, which eases the concern for weak instruments (Stock and Watson, 2018).

Consider first panel 2(a). House prices depreciate after a contractionary monetary shock, albeit the magnitude is small. The result suggests that monetary tightening can slow the pace of house price appreciation, which is consistent with conventional wisdom (e.g., Iacoviello and Neri, 2010). Interestingly, this policy recommendation reverses if we consider panel 2(b):

[9] I set the maximal horizon to two years so that IR estimates are more precise and the GLP procedure is more reliable. The results are robust to different choices of H and sample periods.

[10] The constant term in the baseline model raises the concern for the dynamic panel bias (Nickell, 1981). However, in our setup T is large, and the bias should be negligible. Moreover, Supplementary Material S3.4 shows that the results are robust to excluding the lagged dependent variables.

The individual IRs are highly dispersed with a large fraction exhibiting positive responses, suggesting that curbing soaring house prices using interest rate tools might instead cause instability.

While the message put forward by the individual LP-IV is alarming, drawing definitive conclusions based on these estimates is difficult due to the wide confidence bounds. As panels 2(c) and 2(d) show, the individual IRs are insufficient to support either of the claims above. The GLP is exactly designed to find a data-driven middle ground between panels 2(a) and 2(b), i.e., between unit-specific and common IR estimates.

7.2 Grouping housing dynamics

I now discuss the results of the GLP with three groups.[11] The estimated group IRs are in Figure 3. To better understand the group composition, I report in Table 6 three categories of group-specific economic indicators: economic development, housing market conditions, and household debt.

As is clear, the GLP recovers three distinct responses: (i) considerable housing appreciation (Group 1), (ii) muted responses (Group 2), and (iii) significant depreciation (Group 3). Specifically, Group 1, whose IRs are positive and sizable, is more economically developed: with only 6.3% of the MSAs, this group accounts for 13.1% of the total personal income. In stark contrast, Group 3 constitutes 41.4% of the MSAs, while the share of personal income is only 29%. Moreover, Group 1 has more regulated housing markets and less elastic house supply, implying more pronounced housing cycles (Saiz, 2010). Meanwhile, Groups 2 and 3 are similar in terms of housing conditions. Lastly, Groups 1 and 3 are inhabited by heavily indebted households, as is suggested by the high debt-to-income ratio. Overall, the revealed patterns of Group 1 are in accord with the narratives of housing bubbles (Glaeser and Nathanson, 2015).

[11] See Supplementary Material S3.2 for a detailed discussion on the selection of the number of groups. Importantly, the patterns revealed are robust to the choice of group number.

Interestingly, Table 6 also shows the large standard deviations of the indicators within each group, suggesting that external economic criteria may fail to recover the group structure. For instance, although income levels are positively correlated with housing responses, other latent factors are also at work. Therefore, classifying MSAs based *solely* on income levels will result in only negative IRs of different magnitudes. To see this, I report in Figure 4 the IRs using income as a classifying criterion. As panel 4(a) shows, the responses of the richest 10% of MSAs are, if anything, negative rather than positive. On the contrary, panel 4(b) shows that the poor MSAs in Group 1 have unambiguously positive IRs, even though they are generally poorer than MSAs with housing depreciation (Group 3).

Another widely used grouping criterion is geographical location (e.g., Ferreira and Gyourko, 2012). As is clear from Figure 5, the geographical distribution of MSAs across groups is far from random, with large spatial clusters such as California (Group 1) and Florida (Group 3) readily discernible. However, it is difficult to summarize the distribution through simple rules. For example, classifying MSAs into coastal, sunbelt, and interior areas would again average out the positive IRs.

In conclusion, the GLP estimator can capture well-established facts about housing cycles such as the impact of housing supply, credit conditions, and geography. More importantly, it outperforms those grouping criteria in that it takes into account hidden interactions among those factors. For example, perhaps location is more important than income level for poor MSAs in California, while for MSAs on the east coast, debt accumulation may be more decisive. Empirically, the GLP recovers positive housing responses to contractionary monetary shocks, which have not been documented in the literature.

7.3 Alternative: FAVAR

The previous section shows that the GLP procedure outperforms external grouping criteria in estimating heterogeneous IRs. A natural follow-up is whether alternative methods can also recover the latent group structure. To answer this question, I re-estimate the housing responses to monetary shocks using a FAVAR model, which provides a parsimonious way to characterize the heterogeneity (e.g., Del Negro and Otrok, 2007). Specifically, I estimate the following model:

$$\begin{bmatrix} i_t \\ w_t \\ f_t \end{bmatrix} = \sum_{j=1}^{q} A_j \begin{bmatrix} i_{t-j} \\ w_{t-j} \\ f_{t-j} \end{bmatrix} + u_t \ , \tag{23}$$

where i_t is the policy rate (FFR), w_t is a $P \times 1$ vector of controls, and f_t is an $r \times 1$ vector of factors extracted through

$$Y_t = \Lambda f_t + e_t. \tag{24}$$

Here the information variable Y_t is a vector of housing inflation in 382 MSAs, Λ is an $N \times r$ factor loading matrix, and e_t is the error term. The number of factors is determined by Onatski's (2010) criterion, leading to three factors ($r = 3$) that explain 47.7% of the variation in Y_t. To facilitate comparison, I estimate the above model with the same sampling period, control variables, number of lags ($q = 4$), and instrument as in the GLP before.

Figure 6 summarizes the results. As is clear, the confidence bands generated by FAVAR are exceedingly wide. Even for typical macro variables such as industrial production, the confidence intervals are so large that the sign of the responses can be ambiguous. Moreover, for some MSAs such as New York-Newark-Jersey City and Victoria, the confidence bands are larger than the individual LP-IV counterparts due to the additional estimation uncertainty in (23) caused by the factor estimation in (24).

Besides, there is fairly weak evidence of positive housing responses. Take New York-Newark-Jersey as an example. Compared with Figure 2(c), now the positive IRs are not only of much smaller magnitude but also associated with wider confidence bands. Considering all MSAs, Figure 6 shows that house prices decline in most areas after tightening monetary shocks.

Such observation is nonetheless misleading because the estimated factors only account for half of the variations in housing inflation, making the estimated IRs poor proxies for the true responses. As Figure 7 shows, increasing the number of factors leads to an opposite pattern: House prices rise after tightening monetary shocks, and with eight factors, there seem to be two different IR shapes. However, this pattern remains inconclusive since the eight factors capture only 65.38% of the variations. But to approximate the data well, the dimension of the VAR system would be too large, e.g., 29 factors are required to get R^2 over 90%, which introduces greater estimation uncertainty and makes the results unreliable.

In short, the FAVAR model fails to recover heterogeneous IRs when the variation explained by the factors is moderate, i.e., when the data do not have a strong factor structure, and the resulting IRs could even have the wrong sign. And even if the principal components explain the data well, the confidence bands can still be too large to be useful. In contrast, the GLP approach can capture the group structure well, as it directly groups the IRs without imposing an a priori dimension reduction technique.

8 Conclusion

Heterogeneous impulse responses are often ignored due to estimation difficulty. This paper proposes a solution to capture the heterogeneity that builds on clustering in panel data models. The proposed group local projection (GLP) estimator recovers the latent group structure of the impulse responses and is compatible with standard inference under large N and T. Compared with the individual LP-IV, the GLP facilitates policy recommendations by reducing confidence bands. Simulation studies demonstrate the reliable finite sample performance of the GLP.

Applying the GLP to the study of monetary policy and house prices shows that the GLP successfully recovers the latent groups and the group IRs. In particular, the paper finds that house prices climb up in economically important areas following an unexpected rise in the interest rate, which cautions against the use of monetary policy to control house prices. Notably, neither the external grouping criteria nor FAVAR can recover the group pattern.

The paper also raises several promising research avenues for the future. First, it is interesting to extend the GLP to alternative clustering algorithms that accommodate uncertain group assignments or outliers (e.g., Lewis, Melcangi and Pilossoph, 2019). Second, one may relax Assumption 2 to allow for richer group heterogeneity such as time-varying group membership. Yang, Yan and Huang (2019) serve as a promising first step in this direction.

References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2019. "Democracy Does Cause Growth." *Journal of Political Economy*, 127(1): 47–100.

Ando, Tomohiro, and Jushan Bai. 2015. "Asset Pricing with a General Multifactor Structure." *Journal of Financial Econometrics*, 13(3): 556–604.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica*, 85(3): 693–734.

Arezki, Rabah, Valerie A. Ramey, and Liugang Sheng. 2017. "News Shocks in Open Economies: Evidence from Giant Oil Discoveries." *Quarterly Journal of Economics*, 132(1): 103–155.

Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica*, 70(1): 191–221.

Baker, Scott R. 2018. "Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data." *Journal of Political Economy*, 126(4): 1504–1557.

Barnichon, Regis, and Christian Brownlees. 2019. "Impulse Response Estimation by Smooth Local Projections." *Review of Economics and Statistics*, 101(3): 522–530.

Beetsma, Roel, and Massimo Giuliodori. 2011. "The Effects of Government Purchases Shocks: Review and Estimates for the EU." *The Economic Journal*, 121(550): F4–F32.

Bester, C. Alan, and Christian B. Hansen. 2016. "Grouped Effects Estimators in Fixed Effects Models." *Journal of Econometrics*, 190(1): 197–208.

Boivin, Jean, Marc P Giannoni, and Ilian Mihov. 2009. "Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data." *American Economic Review*, 99(1): 350–384.

Bonhomme, Stéphane, and Elena Manresa. 2015. "Grouped Patterns of Heterogeneity in Panel Data." *Econometrica*, 83(3): 1147–1184.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2022. "Discretizing Unobserved Heterogeneity." *Econometrica*, 90(2): 625–643.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, 50(2): 317–372.

Canova, Fabio, Matteo Ciccarelli, and Eva Ortega. 2012. "Do Institutional Changes Affect Business Cycles? Evidence from Europe." *Journal of Economic Dynamics and Control*, 36(10): 1520–1533.

Cheng, Xu, Frank Schorfheide, and Peng Shao. 2019. "Clustering for Multi-Dimensional Heterogeneity."

Cloyne, James, Clodomiro Ferreira, and Paolo Surico. 2020. "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism." *The Review of Economic Studies*, 87(1): 102–129.

Cloyne, James S, and Paolo Surico. 2017. "Household Debt and the Dynamic Effects of Income Tax Changes." *The Review of Economic Studies*, 84(1): 45–81.

Dees, Stephane, Filippo di Mauro, M. Hashem Pesaran, and L. Vanessa Smith. 2007. "Exploring the International Linkages of the Euro Area: A Global VAR Analysis." *Journal of Applied Econometrics*, 22(1): 1–38.

Del Negro, Marco, and Christopher Otrok. 2007. "99 Luftballons: Monetary Policy and the House Price Boom Across U.s. States." *Journal of Monetary Economics*, 54(7): 1962–1985.

Fernández-Val, Iván, and Joonhwah Lee. 2013. "Panel Data Models with Nonadditive Unobserved Heterogeneity: Estimation and Inference." *Quantitative Economics*, 4(3): 453–481.

Ferreira, Fernando, and Joseph Gyourko. 2012. "Heterogeneity in Neighborhood-Level Price Growth in the United States, 1993–2009." *American Economic Review*, 102(3): 134–140.

Galí, Jordi. 2014. "Monetary Policy and Rational Asset Price Bubbles." *American Economic Review*, 104(3): 721–752.

Galí, Jordi, and Luca Gambetti. 2015. "The Effects of Monetary Policy on Stock Market Bubbles: Some Evidence." *American Economic Journal: Macroeconomics*, 7(1): 233–257.

Glaeser, Edward L., and Charles G. Nathanson. 2015. "Housing Bubbles." In *Handbook of Regional and Urban Economics.* Vol. 5, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, 701–751. Amsterdam: North-Holland.

Gürkaynak, Refet S, Brian Sack, and Eric Swanson. 2005. "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *American Economic Review*, 95(1): 425–436.

Hahn, Jinyong, and Guido Kuersteiner. 2011. "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects." *Econometric Theory*, 27(6): 1152–1191.

Hayashi, Fumio. 2000. *Econometrics.* Princeton: Princeton University Press.

Huang, Wenxin, Sainan Jin, Peter C.B. Phillips, and Liangjun Su. 2021. "Nonstationary Panel Models with Latent Group Structures and Cross-Section Dependence." *Journal of Econometrics*, 221(1): 198–222.

Iacoviello, Matteo, and Stefano Neri. 2010. "Housing Market Spillovers: Evidence from an Estimated DSGE Model." *American Economic Journal: Macroeconomics*, 2(2): 125–164.

Jarociński, Marek, and Frank R. Smets. 2008. "House Prices and the Stance of Monetary Policy." *Federal Reserve Bank of St. Louis Review*, 90(4): 339–65.

Jentsch, Carsten, and Kurt G. Lunsford. 2022. "Asymptotically Valid Bootstrap Inference for Proxy SVARs." *Journal of Business and Economic Statistics*, 40(4): 1876–1891.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95(1): 161–182.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2015. "Betting the House." *Journal of International Economics*, 96: S2–S18.

Kaplan, Greg, and Giovanni L. Violante. 2018. "Microeconomic Heterogeneity and Macroeconomic Shocks." *Journal of Economic Perspectives*, 32(3): 167–194.

Lewis, Daniel J., Davide Melcangi, and Laura Pilossoph. 2019. "Latent Heterogeneity in the Marginal Propensity to Consume." Federal Reserve Bank of New York Staff Report 902.

Lin, Chang-Ching, and Serena Ng. 2012. "Estimation of Panel Data Models with Parameter Heterogeneity When Group Membership Is Unknown." *Journal of Econometric Methods*, 1(1): 42–55.

Maitra, Ranjan, Anna D. Peterson, and Arka P. Ghosh. 2010. "A Systematic Evaluation of Different Methods for Initializing the K-Means Clustering Algorithm."

Miranda-Agrippino, Silvia, and Giovanni Ricco. 2021. "The Transmission of Monetary Policy Shocks." *American Economic Journal: Macroeconomics*, 13(3): 74–107.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3): 703–708.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49(6): 1417–1426.

Onatski, Alexei. 2010. "Determining the Number of Factors from Empirical Distribution of Eigenvalues." *Review of Economics and Statistics*, 92(4): 1004–1016.

Pesaran, M.Hashem, and Ron Smith. 1995. "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics*, 68(1): 79–113.

Ramaswamy, Venkatram, Wayne S. Desarbo, David J. Reibstein, and William T. Robinson. 1993. "An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data." *Marketing Science*, 12(1): 103–124.

Rio, Emmanuel. 2017. *Asymptotic Theory of Weakly Dependent Random Processes.* New York, NY: Springer Berlin Heidelberg.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics*, 125(3): 1253–1296.

Sarafidis, Vasilis, and Neville Weber. 2015. "A Partially Heterogeneous Framework for Analyzing Panel Data." *Oxford Bulletin of Economics and Statistics*, 77(2): 274–296.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2018. "How Important Are Terms-of-Trade Shocks?" *International Economic Review*, 59(1): 85–111.

Stock, James H., and Mark W. Watson. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In *Handbook of Macroeconomics.* Vol. 2, , ed. John B. Taylor and Harald Uhlig, 415–525. Amsterdam: Elsevier.

Stock, James H., and Mark W. Watson. 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." *The Economic Journal*, 128(610): 917–948.

Su, Liangjun, Zhentao Shi, and Peter C. B. Phillips. 2016. "Identifying Latent Structures in Panel Data." *Econometrica*, 84(6): 2215–2264.

Theodoridis, Sergios, and Konstantinos Koutroumbas. 2009. *Pattern Recognition.* . 4th ed ed., Burlington, MA London: Academic Press.

Wang, Wuyi, Peter C. B. Phillips, and Liangjun Su. 2018. "Homogeneity Pursuit in Panel Data Models: Theory and Application." *Journal of Applied Econometrics*, 33(6): 797–815.

White, Halbert. 2001. *Asymptotic Theory for Econometricians.* . 2 ed., San Diego; London and Tokyo: Academic Press.

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

Yang, Xinfeng, Xiaodong Yan, and Jian Huang. 2019. "High-Dimensional Integrative Analysis with Homogeneity and Sparsity Recovery." *Journal of Multivariate Analysis*, 174.

Tables and figures

TABLE 1: SIMULATION DESIGN

		$G^0 = 2$	$G^0 = 3$
Design 1	δ	(3, 3)	(3, 3, 3)
	ρ	(0.2, 0.6)	(0.2, 0.6, 0.9)
Design 2	δ	(1, 2)	(1, 2, 3)
	ρ	(0.5, 0.5)	(0.5, 0.5, 0.5)
IV	π	0.7	0.7
Fraction		(0.5, 0.5)	(0.3, 0.3, 0.4)

Note: This table presents the true parameter values in the simulation study, corresponding to model (18). The row *Fraction* defines the fraction of the individuals in each group.

Table 2: GLP Performance (Known Group Number)

			$G^0 = 2$							$G^0 = 3$						
				BR (%)		RMSE (×100)					BR (%)		RMSE (×100)			
Design	N	T	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP
	100	100	84.1	15.8	15.7	29.6	33.5	63.8	10.0	78.3	20.3	19.3	52.7	81.7	77.5	15.3
	100	200	92.5	15.0	14.8	19.1	33.0	43.4	6.5	90.3	18.8	18.0	34.0	81.2	52.5	9.7
	100	300	96.2	14.7	14.5	13.6	32.8	35.0	5.1	95.0	18.2	17.7	24.7	81.1	42.3	7.5
	200	100	84.1	11.2	11.2	29.0	33.2	63.9	7.9	78.3	14.4	13.8	52.1	81.5	77.5	12.1
1	200	200	92.7	10.6	10.6	18.4	32.8	43.4	4.8	90.2	13.3	12.9	33.6	81.1	52.5	7.1
	200	300	96.3	10.4	10.3	13.1	32.7	35.0	3.8	95.1	12.9	12.6	24.1	81.0	42.3	5.5
	300	100	84.0	9.2	9.2	28.8	33.1	63.9	7.0	78.4	11.8	11.3	51.9	81.4	77.4	10.9
	300	200	92.8	8.7	8.6	18.2	32.8	43.4	4.2	90.2	10.9	10.6	33.8	81.0	52.6	6.3
	300	300	96.3	8.5	8.5	12.9	32.7	35.0	3.2	95.1	10.5	10.4	24.2	81.0	42.3	4.7
	100	100	99.7	15.9	15.7	6.4	22.3	38.5	5.9	99.6	19.5	19.1	9.7	36.8	50.7	9.0
	100	200	100.0	15.0	14.8	3.9	22.0	26.1	3.9	100.0	18.3	17.9	6.1	36.5	34.5	6.0
	100	300	100.0	14.7	14.5	3.1	21.9	21.1	3.1	100.0	18.0	17.6	4.8	36.4	27.8	4.8
	200	100	99.7	11.3	11.2	5.3	22.1	38.5	4.7	99.6	13.8	13.6	7.6	36.6	50.7	6.9
2	200	200	100.0	10.6	10.6	3.0	21.9	26.2	2.9	100.0	13.0	12.8	4.5	36.4	34.5	4.4
	200	300	100.0	10.4	10.3	2.3	21.9	21.1	2.2	100.0	12.7	12.6	3.5	36.3	27.7	3.5
	300	100	99.7	9.2	9.2	4.9	22.1	38.6	4.1	99.6	11.3	11.2	6.8	36.5	50.7	6.0
	300	200	100.0	8.7	8.6	2.5	21.9	26.2	2.4	100.0	10.6	10.5	3.9	36.4	34.4	3.8
	300	300	100.0	8.5	8.5	2.0	21.9	21.1	1.9	100.0	10.4	10.3	3.0	36.3	27.8	2.9

Note: This table reports the classification accuracy (AC), the confidence bands ratios between the GLP and the individual LP-IV (BR), and the RMSE of the GLP. GLP, PAN, IND and IGLP stand for the GLP, panel LP-IV, individual LP-IV and the infeasible GLP respectively. The infeasible GLP is group-by-group using standard panel LP-IV where we know the true group structure beforehand. Classification accuracy and band ratios are in percentage terms, and RMSE are multiplied by 100.

Table 3: GLP Coverage Rates (%)

Design	N	T	$G^0=2$ h=0	h=1	h=2	h=3	h=4	h=5	h=6	$G^0=3$ h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	92.4	90.6	88.1	86.4	86.7	88.2	89.0	92.4	86.6	88.2	91.4	93.1	94.7	94.9
	100	200	94.1	94.0	94.9	94.4	92.0	93.3	92.3	94.3	91.3	91.9	93.3	94.0	94.9	94.7
	100	300	94.6	94.3	95.2	94.6	94.4	93.6	93.8	94.5	93.2	92.9	93.0	93.9	95.0	94.8
	200	100	90.2	86.2	78.7	77.2	80.2	83.9	84.8	91.6	75.7	77.1	83.0	87.9	91.5	93.3
	200	200	92.9	91.7	92.6	92.0	91.3	91.7	91.5	93.2	85.7	86.2	89.5	93.1	93.2	94.5
	200	300	93.2	93.6	95.7	95.6	93.9	93.4	92.4	93.7	90.6	92.1	93.4	93.9	94.3	94.6
	300	100	87.7	80.0	71.2	68.1	72.7	78.7	82.4	89.4	68.9	67.0	74.5	83.1	88.0	92.1
	300	200	92.1	89.9	91.7	89.0	88.9	89.3	88.1	92.4	81.1	81.6	84.9	90.2	93.0	94.1
	300	300	92.4	92.0	94.9	94.5	92.6	92.1	90.6	94.6	88.9	88.6	90.2	92.3	93.5	94.4
2	100	100	92.7	92.6	90.7	90.7	91.8	92.3	92.0	94.2	93.7	91.4	91.4	91.8	92.5	91.4
	100	200	94.1	93.8	92.9	93.3	92.4	92.4	93.9	94.3	93.7	92.5	91.9	93.0	92.6	93.8
	100	300	95.5	93.8	93.9	93.0	93.4	93.5	94.0	94.6	94.3	93.9	94.3	93.3	93.6	93.2
	200	100	90.9	91.5	89.4	87.5	86.4	85.7	86.4	92.3	92.7	90.1	89.3	89.2	88.9	89.0
	200	200	93.9	93.6	91.9	90.9	90.8	90.4	91.4	94.6	94.3	93.2	92.1	92.6	91.9	92.0
	200	300	94.5	94.5	93.4	93.0	92.3	93.4	92.8	94.4	94.2	93.4	93.4	93.1	92.8	92.0
	300	100	89.9	90.4	85.6	82.4	82.7	83.3	82.5	91.8	90.7	87.9	86.4	86.7	86.5	87.2
	300	200	92.6	92.4	91.2	90.2	90.1	89.2	89.3	93.2	92.4	90.8	90.6	90.4	89.6	90.3
	300	300	94.4	94.0	92.4	91.6	91.5	90.5	91.4	93.8	93.9	92.8	92.1	92.0	92.3	92.3

Note: This table reports the coverage probability of the large T inference as in Theorem 2. Alternative inference methods are compared in Supplementary Material S2.4.

Table 4: GLP with Unknown Group Number (RMSE ×100)

Design	N	100	100	100	200	200	200	300	300	300
	T	100	200	300	100	200	300	100	200	300
1	PAN	81.7	81.2	81.0	81.5	81.1	81.0	81.4	81.0	81.0
	$\hat{G} = 2$	**56.7**	40.2	33.0	**56.1**	39.8	32.6	**55.7**	39.6	32.4
	$\hat{G} = 3$	58.2	**39.2**	**27.1**	57.9	**39.0**	**26.7**	57.4	**38.5**	**26.3**
	$\hat{G} = 4$	59.2	39.7	30.3	58.4	39.3	29.7	57.7	39.2	29.4
	$\hat{G} = 5$	60.0	40.4	31.1	59.1	39.9	30.8	58.7	39.7	30.5
	$\hat{G} = 6$	61.0	41.1	31.9	60.0	40.6	31.2	59.4	40.2	31.1
	$\hat{G} = 7$	62.1	41.7	32.4	60.8	41.0	31.9	60.2	40.7	31.4
	$\hat{G} = 8$	62.5	42.1	32.9	61.4	41.4	32.2	60.7	41.1	31.8
	IND	77.5	52.5	42.2	77.7	52.6	42.2	77.5	52.6	42.3
	IC	2.0	2.7	3.0	2.0	2.9	3.0	2.0	3.0	3.0
2	PAN	36.8	36.5	36.4	36.6	36.4	36.3	36.5	36.4	36.3
	$\hat{G} = 2$	19.7	17.9	17.5	18.9	17.5	17.2	18.6	17.3	17.1
	$\hat{G} = 3$	**9.7**	**6.2**	**4.8**	**7.6**	**4.6**	**3.6**	**6.9**	**3.9**	**3.0**
	$\hat{G} = 4$	21.6	15.1	12.3	20.1	14.0	11.4	19.5	13.6	11.1
	$\hat{G} = 5$	25.1	17.5	14.3	23.4	16.4	13.3	22.8	15.9	12.9
	$\hat{G} = 6$	27.4	19.1	15.6	25.6	17.8	14.6	24.9	17.4	14.2
	$\hat{G} = 7$	29.1	20.3	16.5	27.1	19.0	15.4	26.3	18.4	15.0
	$\hat{G} = 8$	30.5	21.2	17.3	28.4	19.8	16.1	27.5	19.2	15.7
	IND	50.7	34.4	27.8	50.7	34.5	27.8	50.6	34.5	27.7
	IC	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0

Note: This table reports the RMSE (multiplied by 100) of the GLP with different supplied group number. Cells chosen by the information criterion are in bold.

Table 5: GLP with No Group Structure (RMSE ×100)

N	100	100	100	200	200	200	300	300	300
T	100	200	300	100	200	300	100	200	300
PAN	45.7	45.2	45.7	45.6	45.4	45.4	45.8	45.5	45.5
$\hat{G} = 2$	**39.9**	39.1	39.0	**39.7**	39.1	39.0	**39.8**	39.2	39.2
$\hat{G} = 3$	37.4	**34.0**	**33.6**	36.7	**34.2**	34.1	36.8	**34.6**	34.4
$\hat{G} = 4$	36.6	31.6	30.3	35.7	31.6	**30.3**	35.8	31.7	**30.6**
$\hat{G} = 5$	36.4	30.0	28.2	35.3	30.2	28.4	35.4	30.2	28.6
$\hat{G} = 6$	36.3	29.2	27.2	35.2	29.2	27.2	35.2	29.3	27.3
$\hat{G} = 7$	36.8	28.7	26.2	35.5	28.7	26.1	35.2	28.5	26.4
$\hat{G} = 8$	37.0	28.4	25.5	35.7	28.3	25.5	35.3	28.1	25.6
IND	50.5	34.2	27.8	50.5	34.4	27.6	50.6	34.3	27.8
IC	2.10	2.80	3.30	2.00	3.00	3.50	2.10	3.10	3.70

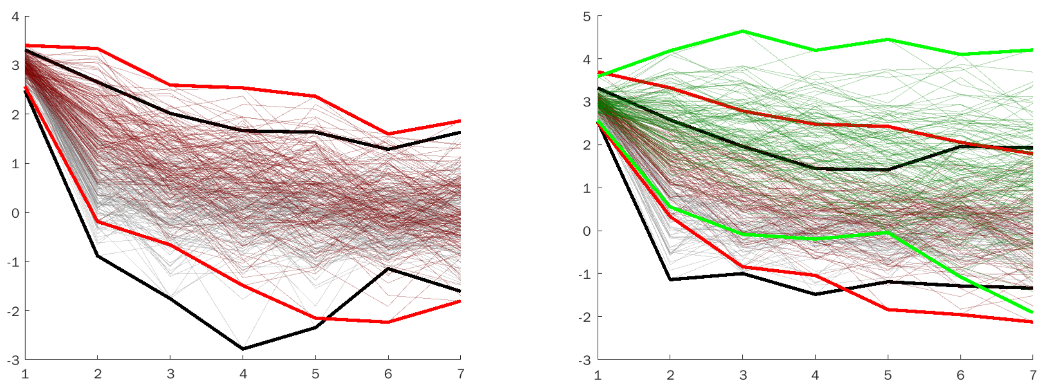
Note: This table reports the RMSE (multiplied by 100) of the GLP with different supplied group number. Cells chosen by the information criterion are in bold.

Table 6: MSA Group Economic Profile

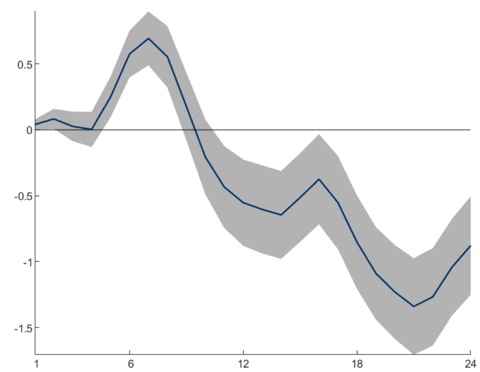
			GDP		Income		Population		Employment		Regulation		Elasticity		Debt (Low)		Debt (High)	
$\hat{G}$	g	count	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
2	1	59	52436	18145	48493	12688	1081	2061	669	1343	0.293	0.848	1.870	0.923	1.586	0.514	1.740	0.485
	2	323	46052	13190	45446	9399	666	1538	413	1001	-0.196	0.801	2.721	1.485	1.400	0.474	1.561	0.445
3	1	24	53686	22898	49063	14082	1486	2768	885	1782	0.513	0.396	1.530	0.775	1.841	0.449	1.954	0.435
	2	200	48452	14444	46352	9532	774	1861	490	1220	-0.172	0.871	2.685	1.465	1.359	0.440	1.529	0.418
	3	158	44238	11531	44888	9828	560	918	339	590	-0.153	0.780	2.631	1.443	1.453	0.511	1.614	0.481
4	1	24	53686	22898	49063	14082	1486	2768	885	1782	0.513	0.396	1.530	0.775	1.841	0.449	1.954	0.435
	2	87	49303	12894	47058	9575	842	2224	535	1462	0.135	0.985	2.131	0.870	1.413	0.465	1.579	0.446
	3	168	46887	14375	46086	10456	642	1311	404	857	-0.336	0.695	2.987	1.616	1.364	0.442	1.525	0.414
	4	103	43822	11592	43943	8158	603	1040	362	661	-0.154	0.815	2.616	1.466	1.450	0.530	1.622	0.499

Note: GDP and income are per capita measured in dollars. Population and employment are measured in thousands of person and thousands of jobs respectively. Regulation stands for the Wharton residential land use regulatory index, and Elasticity is the house supply elasticity. Debt (Low) and Debt (High) correspond to the lower bound and upper bound of the debt to income ratio.

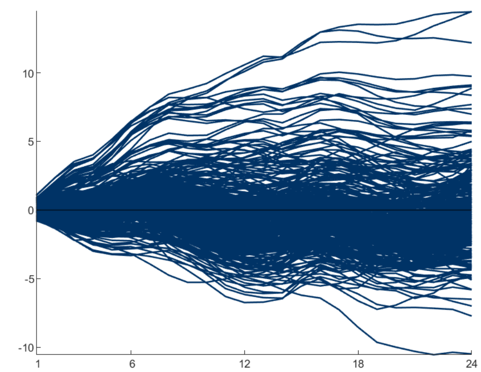
Figure 1: Simulation Example



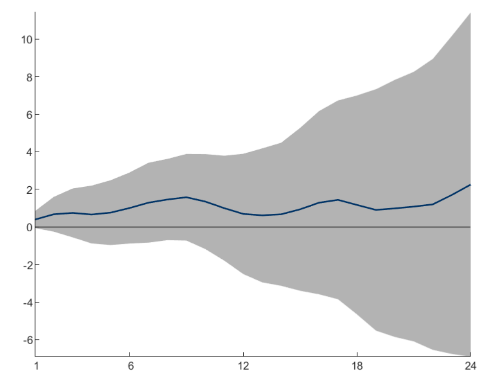
(a) Design 1 ($G^0 = 2, N = 300, T = 100$) (b) Design 1 ($G^0 = 3, N = 300, T = 100$)



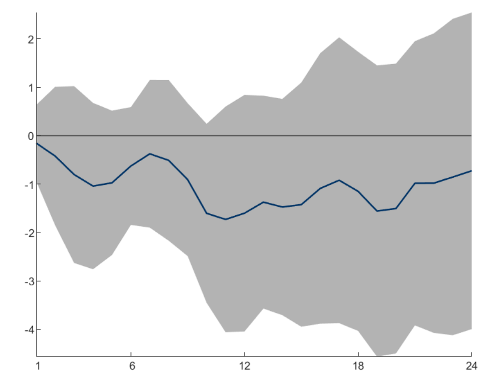
(A) Panel LP-IV



(B) Individual LP-IV



(C) New York-Newark-Jersey City



(D) Victoria, TX

Figure 2: House Prices Impulse Responses (LP-IV)

Note: All IRs are measured in percentage and correspond to a one percentage increase in the FFR. The shaded areas indicate 95% confidence bands. For panel LP-IV, standard errors are clustered at MSA level (Cameron and Miller, 2015), and for the individual LP-IV, NeweyWest HAC standard errors are used with $H + 1$ lags.

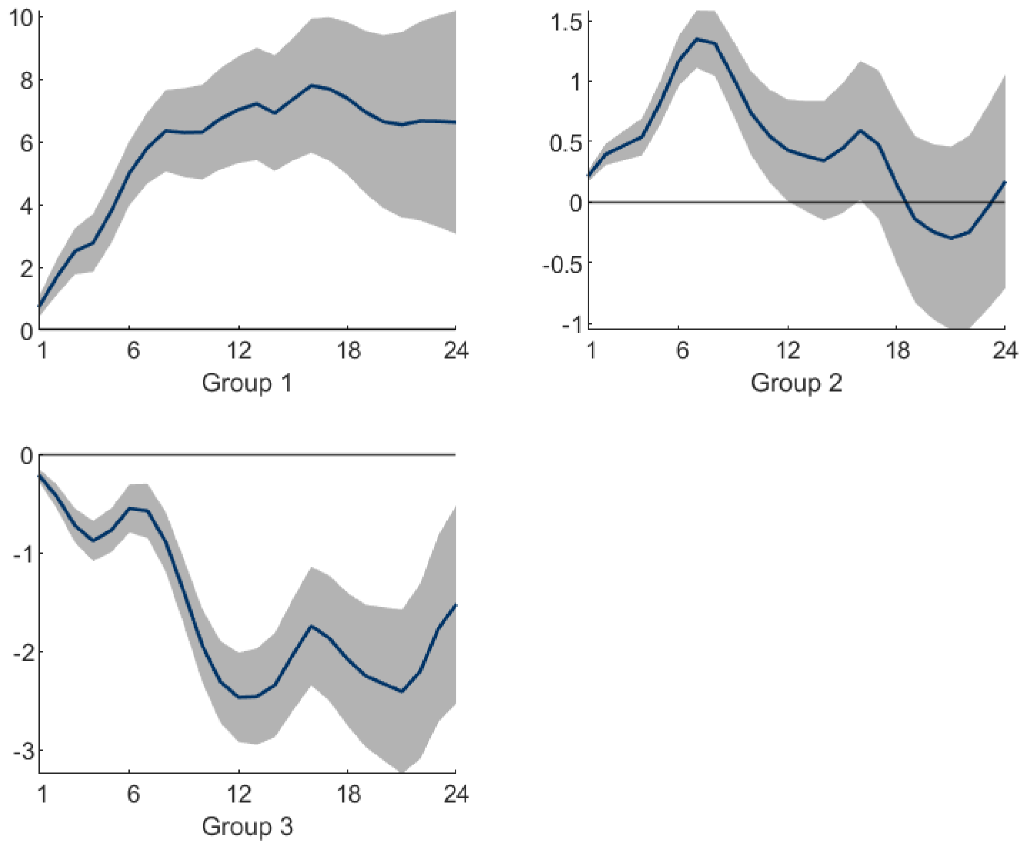
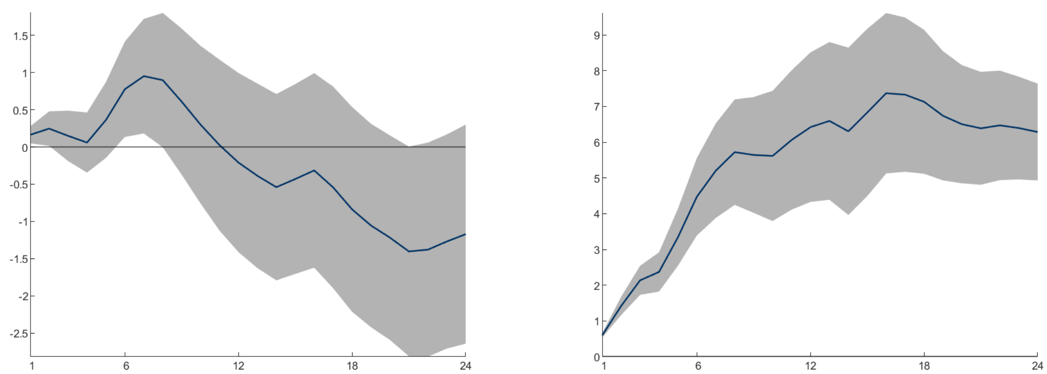


Figure 3: Group Impulse Responses ($\hat{G} = 3$)

Note: For the GLP estimates, the standard errors are computed with large T inference as in Theorem 2.



(a) The Richest 10% MSAs (b) Poorest 10% MSAs in Group 1

Figure 4: External Grouping Criterion (Income Level)

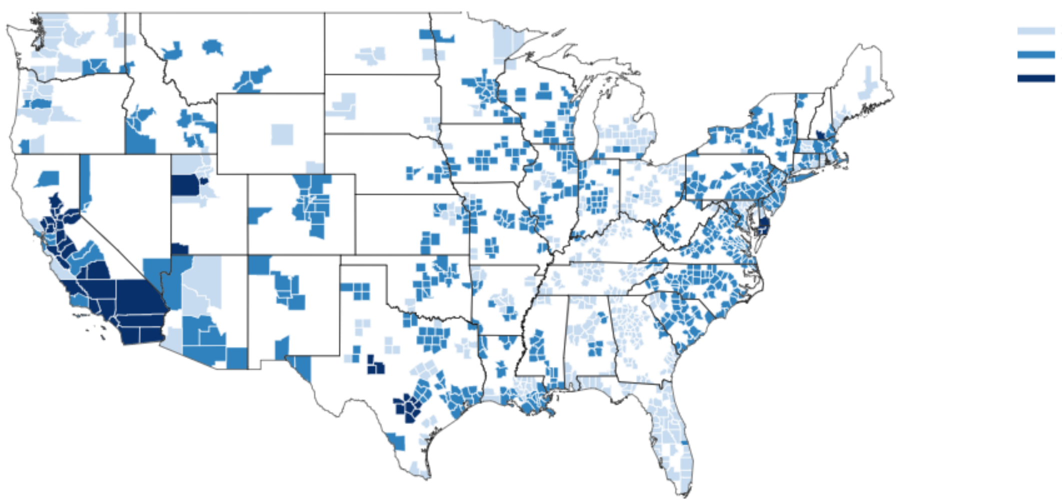


Figure 5: Geographical Distribution of MSA Groups

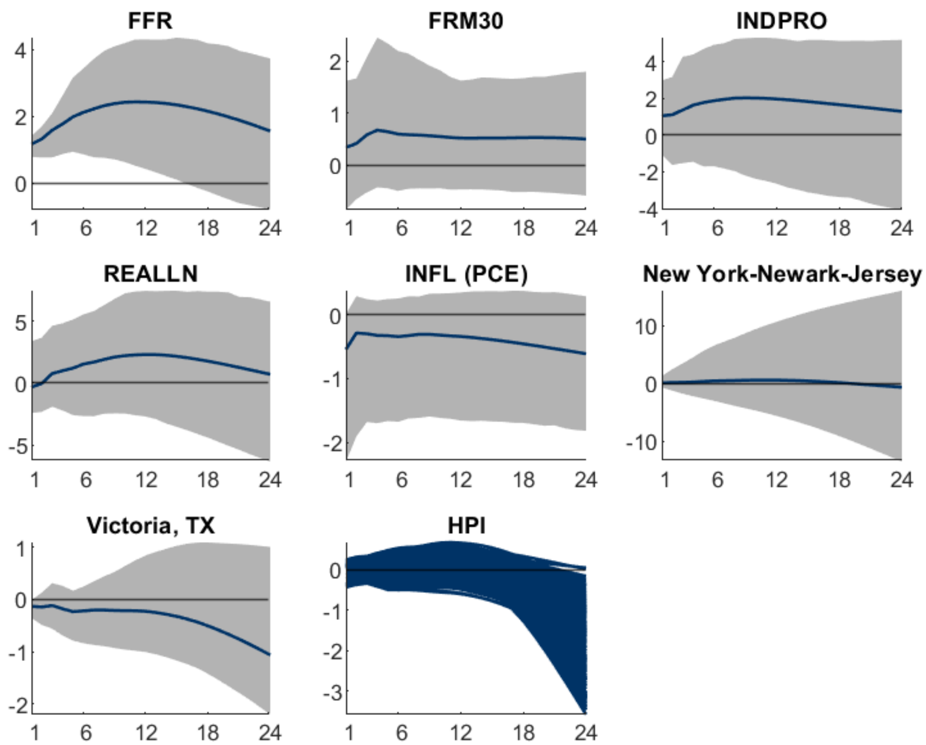
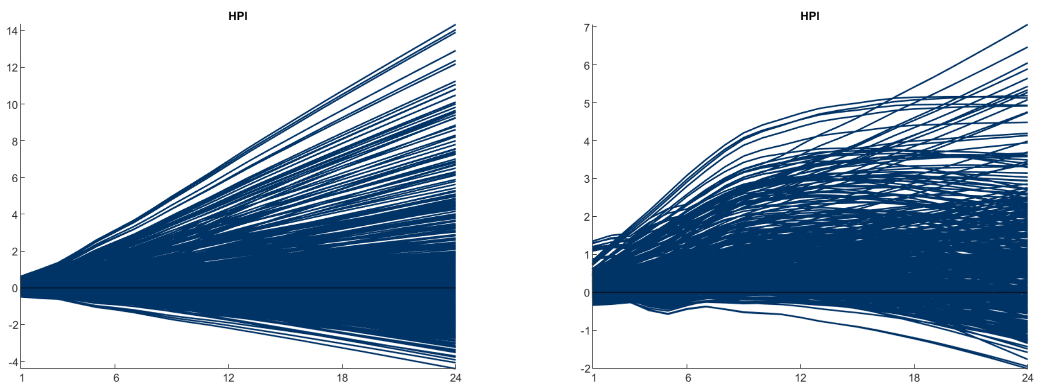


FIGURE 6: IMPULSE RESPONSES: FAVAR

Note: The shaded areas indicate 95% confidence bands using moving block bootstrap as in Jentsch and Lunsford (2022). Results are robust to different inferential methods.



(a) Five Factor (b) Eight Factors

Figure 7: Housing Responses: FAVAR with Different Factors