Identification and Estimation of Demand in Large Concentrated Markets

Saman Banafti [†] Tae-Hwy Lee [‡]

Abstract
Gabaix and Koijen (2020) introduces the Granular Instrumental Variables (GIV) methodology, which takes advantage of panel data to construct instruments to estimate structural time series regression models that involve endogenous regressors. The GIVs are constructed based on panel data models with factor structures, where the idiosyncratic error terms may have extraordinarily useful information. In this paper, we extend their GIV methodology by developing the GIV identification procedure to a large N and large T framework (current identification is for fixed N and large T) by establishing and restricting the asymptotic behavior of the Herfindahl index for large N markets as a function of the tail index of the size distribution of the cross-sectional units.

Key Words: Interactive effects, Factor error structure, Simultaneity, Power-law tails, Asymptotic Herfindahl index, Global crude oil market, Demand elasticity.

1. Introduction

In the absence of randomized control trials, finding valid and strong instruments to circumvent unobserved confounders is a very challenging task. The promising Granular Instrumental Variables (GIV hereater) methodology, Gabaix and Koijen (2020), proposes a systematic way to construct instruments from suitably weighted idiosyncratic shocks, from observational datasets and use them as instruments for aggregate endogenous variables.

How can idiosyncratic shocks be relevant for endogenous aggregate variables? Gabaix (2011) provided an initial theoretical solution to the debate by showing that when the firm size distribution is heavy tailed, the central limit theorem does not apply and idiosyncratic volatility decays much slower than $\frac{1}{\sqrt{N}}$. Gabaix (2011) coined this mechanism as the so-called "granular" hypothesis, in which the economy is composed of incompressible grains as opposed to infinitesimally small micro units. Acemoglu et al. (2012) formulated a network approach to demonstrate that sectoral idiosyncratic shocks generate non-negligible aggregate volatility when there exists sufficient asymmetry in the input-output relationships. Pesaran and Yang (2020) build off of the theoretical approach of Acemoglu et al. (2012) and develop econometric theory to measure the degree of network dominance and in their application they find some evidence of sector-specific shock propagation albeit not overwhelmingly strong for the US input-output accounts data over the period 1972-2002. More empirical evidence for such propagation mechanism is presented in Gatti et al. (2005), Canals et al. (2007), Koren and Tenreyro (2007), Blank et al. (2009), Malevergne et al. (2009), Yan (2011), Gabaix (2011), Carvalho and Gabaix (2013), Schiaffi (2013), Acemoglu et al. (2017), Jannati (2017) and Lera and Sornette (2017).

Gabaix and Koijen (2020), hereafter GK, illustrate that when the market under consideration is sufficiently concentrated, then one can use the the collection of idiosyncratic shocks to individual micro units, at each time period t, as an instrument for endogenous aggregate variables. The instrumental relevance follows heuristically from above. The exogeneity condition, as in any instrumental variables procedure, requires assumptions on unobserved random variables. However it should be noted that the exogeneity condition exploited in this framework is a relatively mild assumption that is often made in factor models (e.g. Bai

[†]Department of Economics, University of California, Riverside. Email: sbana001@ucr.edu.
[‡]Department of Economics, University of California, Riverside. Email: taelee@ucr.edu.

and Ng (2002)) for identification purposes. The insight and contribution of GK opens the doors to a wide possibility of ways in which one can continue building on the promising new GIV methodology. Our contribution to the GIV methodology is that we naturally extend their identification procedure to a large N and large T framework (GK formally introduced GIV for a fixed N and large T) by establishing and restricting the aymptotic behavior of the Herfindahl index for large N markets as a function of the tail index of the size distribution.

Notation. Throughout, we distinguish vectors and matrices from scalars by making an object bold. Let $\{X_{it}, i = 1, \ldots, N; t = 1, \ldots, T\}$ be a double index process of random variables where N denotes the number of cross-sectional units and T denotes the number of time periods. When we stack across i, we obtain $\underset{N\times 1}{\boldsymbol{X_{\cdot t}}} \equiv (X_{1t}, \ldots, X_{Nt})'$. Similarly, if we stack across t we obtain $\underset{T\times 1}{\boldsymbol{X_{i\cdot}}} \equiv (X_{i1}, \ldots, X_{iT})'$. Define X_{wt} as the cross-sectionally weighted average of X_{it}, that is $X_{wt} = \sum_{i=1}^{N} w_i X_{it}$. Let $\underset{N\times 1}{\boldsymbol{\iota}}$ denote a vector of ones. Let $\widetilde{X}_{it} = X_{it} - \bar{X}_t$, where $\bar{X}_t = \frac{1}{N}\sum_{i=1}^{N} X_{it}$, denote a cross-sectionally demeaned variable. Joint convergence of N and T will be denoted as $(N, T) \xrightarrow{j} \infty$ without any restriction on the relative rates.

2. Model

In our exposition, we focus on the canonical setting of estimating the demand elasticity in the global crude oil market. Our stylized simultaneous equations model takes the simple form

$$d_t = \phi^d p_t + \varepsilon_t \tag{1}$$
$$y_{it} = \phi^s p_t + \boldsymbol{\lambda}_i' \boldsymbol{\eta}_t + u_{it}, \tag{2}$$

where d_t denotes the log change of aggregate crude oil consumption and y_{it} denotes the log change of country $i's$ crude oil production, p_t is the log change of real crude oil price (where we deflate the nominal oil price with the U.S. general price index), the coefficients ϕ^s and ϕ^d denote the crude oil supply and demand elasticities, respectively, and $\boldsymbol{\lambda_i}, \boldsymbol{\eta_t}$ are $r \times 1$ vectors of latent factor loadings and latent factors.[1,2] We assume no additional covariates for ease of exposition but they can be easily accomodated. The global market clearing condition is given by [3]

$$y_{St} \equiv \boldsymbol{S'y_{\cdot t}} = \sum_{i=1}^{N} S_i y_{it} = d_t,$$

where $\boldsymbol{S}$ is the $N \times 1$ vector of shares that are normalized such that $\sum_{i=1}^{N} S_i = 1$ and i and t take the values $i = 1, \ldots, N$ and $t = 1, \ldots, T$, respectively. Making use of the market clearing assumption we see that

$$p_t = \frac{1}{\phi^d - \phi^s}\left(u_{St} + \boldsymbol{\lambda_S'\eta_t} - \varepsilon_t\right), \tag{3}$$

[1] The main results extend relatively naturally to the case where both variables have a panel model.

[2] We estimate r using the ER and GR methods of Ahn and Horenstein (2013).

[3] As oil is a storable good, one could easily allow oil prices to adjust to the gap between supply and demand, e.g. as in Mohaddes and Pesaran (2016), who also allow prices to follow a general ARDL speciication. This introduces more complex notations without adding any substance to the main points for identification.

which makes the simultaneity clear, e.g., that prices are composed of size weighted idiosyncratic shocks and aggregate supply and demand shocks. The objective of the GIV methodology is to extract the idiosyncratic shocks and use them as instruments for price. For a large N extension, we require the tail of the size distribution to follow a power law, with tail index $\mu \in (0,1)$ or $\mu \to 0$. For example, if $\mu \in (1,2)$ it can be shown that the large N representation of prices satisfy $p_t = \frac{1}{\phi^d - \phi^s}\left(\boldsymbol{\lambda}'_{\boldsymbol{S}}\boldsymbol{\eta_t} - \varepsilon_t\right) + \mathcal{O}_p(1/(N^{1-\frac{1}{\mu}}))$; which renders a weak instrument based on size weighted idiosyncratic shocks for large N.

The case of uniform loadings. To momentarily fix ideas, it is helpful to consider a major simplification when constructing the instrument. Suppose that $\lambda_i = \lambda\,\forall i$ and define the $N \times 1$ vector $\boldsymbol{E} \equiv \boldsymbol{\iota}/N$. Then, the instrument can be formed as

$$\begin{aligned} z_t = y_{St} - y_{Et} &= (\phi^d p_t + \boldsymbol{\lambda}'\boldsymbol{\eta_t} + u_{St}) - (\phi^d p_t + \boldsymbol{\lambda}'\boldsymbol{\eta_t} + u_{Et}), \\ &= u_{St} - u_{Et} \equiv u_{\Gamma t}, \end{aligned} \tag{4}$$

where $\boldsymbol{\Gamma} \equiv \boldsymbol{S} - \boldsymbol{E} = \boldsymbol{S} - \boldsymbol{\iota}/N$ is a $N \times 1$ vector such that $\boldsymbol{\iota}'\boldsymbol{\Gamma} = \sum_i \Gamma_i = 0$, by construction. Identification by GIV requires that

$$\mathbb{E}(u_{it}\varepsilon_t) = 0\,\forall i,t, \tag{5}$$

$$\mathbb{E}(u_{it}\boldsymbol{\eta_t}) = 0\,\forall i,t. \tag{6}$$

Which gives $\mathbb{E}(p_t z_t) \neq 0$ (relevance) and $\mathbb{E}(z_t\varepsilon_t) = 0$ (exogeneity). Given relevance, exogeneity implies the following demand elasticity estimator $\widehat{\phi}^d = \frac{\sum_t d_t z_t}{\sum_t p_t z_t}$. Intuitively, z_t places larger weights on the idiosyncratic shocks to larger oil producers, these granular shocks will shift the supply curve while keeping the aggregate demand curve fixed since demand responds to these shocks only through their affects on prices. This allows for consistent estimation of the demand elasticity. The homogenous loadings assumption in this case tremendously facilitated the analysis. Uniform loadings allows one to construct the instrument, as in (4), from observables.

Although homogeneous loadings was only an abstraction to illustrate the instrument, GK advocate the use of $y_{\Gamma t} = y_{St} - y_{Et}$ in practice even when the loadings aren't uniform. In the general heterogenous loadings case, the instrument becomes

$$Z_t \equiv y_{\Gamma t} = u_{\Gamma t} + \boldsymbol{\lambda}_\Gamma \boldsymbol{\eta}_t \tag{7}$$

GK label this instrument with a capital case convention, to distinguish it because it is no longer solely composed of weighted idiosyncratic shocks, as the $\boldsymbol{\lambda}_\Gamma \boldsymbol{\eta}_t$ term is contaminating the instrument. However, this clever formulation is possible because they advocate estimation of the factors in practice, which they augment to their structural equations, thereby controlling for the second term which can potentially make their moment conditions different from zero (e.g., when $\mathbb{E}(S_i\boldsymbol{\lambda}_i) \neq 0$).

Homogeneous loadings are overly restrictive but can be easily accomodated in practice via PCA or iterative OLS-PCA methods e.g., Bai (2003), Bai (2009), Moon and Weidner (2017), Bai et al. (2015). Although in GK's asymptotic theory they assume homogenous loadings, which circumvents the need to estimate the factor structure, they indeed advocate augmenting their structural equations with estimated factors either via period-by-period cross sectional regressions when the loadings are known or via PCA in the case of non-parametric loadings. GK abstract away from the sampling error in suggesting the use of augmented factors, which only vanishes for both large N and T. Bai and Ng (2006) and

Greenaway-McGrevy et al. (2012) have developed the asymptotic distribution for structural parameters in factor augmented regressions in time series and panel models, respectively. In this paper, a variant of their corresponding results is established in showing the sampling error in the instrument is negligible when considering the asymptotic distribution of the structural parameter.

The general case when $\boldsymbol{\lambda_i}\neq\boldsymbol{\lambda}$. Here we formulate the estimation approach in the general case, which makes much heavier use of the cross-section. When we cross-sectionally demean the supply equation and stack across i we obtain

$$\widetilde{\boldsymbol{y}}_{\cdot\boldsymbol{t}} = \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\eta_t} + \widetilde{\boldsymbol{u}}_{\cdot\boldsymbol{t}}, \tag{8}$$

which is estimable with vanilla PCA when the factor structure is strong.[4] Letting $\boldsymbol{Q} = (\boldsymbol{I_N} - \widetilde{\boldsymbol{\Lambda}}(\widetilde{\boldsymbol{\Lambda}}'\widetilde{\boldsymbol{\Lambda}})^{-1}\widetilde{\boldsymbol{\Lambda}}')$, then $\boldsymbol{Q}\widetilde{\boldsymbol{y}}_{\cdot\boldsymbol{t}} = \boldsymbol{Q}\widetilde{\boldsymbol{u}}_{\cdot\boldsymbol{t}}$, completely purges the process of the common factors through the loading space. Premultiplying the share weights gives the instrument

$$\begin{aligned} z_t &= \boldsymbol{S}'\boldsymbol{Q}\widetilde{\boldsymbol{y}}_{\cdot\boldsymbol{t}}, && (9)\\ &= \boldsymbol{S}'\boldsymbol{Q}\widetilde{\boldsymbol{u}}_{\cdot\boldsymbol{t}} \equiv \boldsymbol{\Gamma}'\widetilde{\boldsymbol{u}}_{\cdot\boldsymbol{t}}, && (10) \end{aligned}$$

where $\boldsymbol{\Gamma} \equiv \boldsymbol{QS}$ is unknown because $\boldsymbol{Q}$ is unknown, but $\boldsymbol{Q}$ is easily estimated from data. Once we have $\widehat{\boldsymbol{Q}}$, we form $\widehat{z}_t = \boldsymbol{S}'\widehat{\boldsymbol{Q}}\widetilde{\boldsymbol{y}}_{\cdot\boldsymbol{t}}$ from observables. Importantly, when $\boldsymbol{\lambda_i} = \boldsymbol{\lambda}\,\forall i$, then $\boldsymbol{\Gamma} = (\boldsymbol{I_N} - \widetilde{\boldsymbol{\Lambda}}(\widetilde{\boldsymbol{\Lambda}}'\widetilde{\boldsymbol{\Lambda}})^{-1}\widetilde{\boldsymbol{\Lambda}}')\boldsymbol{S} = \boldsymbol{S} - \boldsymbol{\iota}/N$ from the previous example with homogenous loadings. This gives rise to a more general demand elasticity estimator

$$\widehat{\phi}^d = \widehat{\phi}^d(\widehat{\boldsymbol{z}}) = \frac{\sum_t d_t \widehat{z}_t}{\sum_t p_t \widehat{z}_t}. \tag{11}$$

It follows that

$$\begin{aligned} \widehat{\phi}^d - \phi^d &= \left(\sum_t \widehat{z}_t p_t\right)^{-1}\left(\sum_t \widehat{z}_t \varepsilon_t\right), \\ &= \left(\sum_t z_t p_t + \sum_t (\widehat{z}_t - z_t) p_t\right)^{-1}\left(\sum_t z_t \varepsilon_t + \sum_t (\widehat{z}_t - z_t)\varepsilon_t\right). \end{aligned}$$

From above, it is apparent we need to show $\frac{1}{T}\sum_{t=1}^T(\widehat{z}_t - z_t)\varepsilon_t = \frac{1}{T}\sum_{t=1}^T \boldsymbol{S}'(\widehat{\boldsymbol{Q}} - \boldsymbol{Q})\widetilde{\boldsymbol{y}}_{\cdot\boldsymbol{t}}\varepsilon_t = o_p(1)$ and $\frac{1}{T}\sum_{t=1}^T(\widehat{z}_t - z_t)p_t = \frac{1}{T}\sum_{t=1}^T \boldsymbol{S}'(\widehat{\boldsymbol{Q}} - \boldsymbol{Q})\widetilde{\boldsymbol{y}}_{\cdot\boldsymbol{t}}p_t = o_p(1)$. The order of the sampling error generally relies, in part, on μ, the tail index of the size distribution. Results on the order of the Herfindahl as a function of the tail index parameter μ entails a total of six possible cases. However, for inference, we require $\mu \in (0,1)$ (regularly varying tails) or $\mu \to 0$ (slowly varying tails). Then, under regularity assumptions it can be shown that as

[4] Strong factors in the sense that $\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}'/N \xrightarrow{p} \mathbb{E}(\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}') > 0$; which is equivalently stated as $\gamma_{max}(\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}') = \Theta_p(N)$ where γ_{max} denotes the maximum eigenvalue and $a = \Theta(b)$ states that a and b rise jointly proportionally.

$(N,T) \xrightarrow{j} \infty$, consistency and asymptotic normality are achieved

$$\widehat{\phi}^d - \phi^d = \left(\sum_t \widehat{z}_t p_t\right)^{-1}\left(\sum_t \widehat{z}_t \varepsilon_t\right) = \left(\sum_t z_t p_t\right)^{-1}\sum_t z_t \varepsilon_t + o_p(1), \quad (12)$$

$$\sqrt{T}(\widehat{\phi}^d - \phi^d) \xrightarrow{d} \mathcal{N}(0, \mathbb{v}_d). \quad (13)$$

3. Concluding Remarks

In this paper, we have further developed the GIV methodology introduced by Gabaix and Koijen (2020), which takes advantage of panel data to construct instruments for estimation of structural time series regression models that involve endogenous regressors. We focus on the underlying econometric issues involved in extending GIV to a large N and large T framework where the loadings are treated as unknown parameters to be estimated before constructing the instrument. We further establish that the sampling error arising from estimating the instrument does not affect the limiting distribution for the structural parameter of interest.

REFERENCES

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). "The network origins of aggregate fluctuations". *Econometrica* 80.5, pp. 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2017). "Microeconomic origins of macroeconomic tail risks". *American Economic Review* 107.1, pp. 54–108.

Ahn, S. C. and A. R. Horenstein (2013). "Eigenvalue ratio test for the number of factors". *Econometrica* 81.3, pp. 1203–1227.

Bai, J. (2003). "Inferential theory for factor models of large dimensions". *Econometrica* 71.1, pp. 135–171.

— (2009). "Panel data models with interactive fixed effects". *Econometrica* 77.4, pp. 1229–1279.

Bai, J. and S. Ng (2002). "Determining the number of factors in approximate factor models". *Econometrica* 70.1, pp. 191–221.

— (2006). "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions". *Econometrica* 74.4, pp. 1133–1150.

Bai, J., Y. Liao, and J. Yang (2015). "Unbalanced panel data models with interactive effects". *The Oxford Handbook of Panel Data*.

Blank, S., C. M. Buch, and K. Neugebauer (2009). "Shocks at large banks and banking sector distress: The Banking Granular Residual". *Journal of Financial Stability* 5.4, pp. 353–373.

Canals, C., X. Gabaix, J. Vilarrubia, and D. Weinstein (2007). "Trade Shocks, Trade Balances, and Idiosyncratic Shocks". *Banco de España, Documento de Trabajo* 721.

Carvalho, V. and X. Gabaix (2013). "The great diversification and its undoing". *American Economic Review* 103.5, pp. 1697–1727.

Gabaix, X. (2011). "The granular origins of aggregate fluctuations". *Econometrica* 79.3, pp. 733–772.

Gabaix, X. and R. S. Koijen (2020). "Granular instrumental variables". *Available at SSRN 3368612*.

Gatti, D. D., C. Di Guilmi, E. Gaffeo, G. Giulioni, M. Gallegati, and A. Palestrini (2005). “A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility”. *Journal of Economic Behavior & Organization* 56.4, pp. 489–512.

Greenaway-McGrevy, R., C. Han, and D. Sul (2012). “Asymptotic distribution of factor augmented estimators for panel regression”. *Journal of Econometrics* 169.1, pp. 48–53.

Jannati, S. (2017). “Geographic Spillover of Dominant Firms’ Shocks”. *8th Miami Behavioral Finance Conference*. Vol. 2019.

Koren, M. and S. Tenreyro (2007). “Volatility and development”. *The Quarterly Journal of Economics* 122.1, pp. 243–287.

Lera, S. C. and D. Sornette (2017). “Quantification of the evolution of firm size distributions due to mergers and acquisitions”. *PloS one* 12.8.

Malevergne, Y., P. Santa-Clara, and D. Sornette (2009). *Professor zipf goes to wall street*. Tech. rep. National Bureau of Economic Research.

Mohaddes, K. and M. H. Pesaran (2016). “Country-specific oil supply shocks and the global economy: A counterfactual analysis”. *Energy Economics* 59, pp. 382–399.

Moon, H. R. and M. Weidner (2017). “Dynamic linear panel regression models with interactive fixed effects”. *Econometric Theory* 33.1, pp. 158–195.

Pesaran, M. H. and C. F. Yang (2020). “Econometric analysis of production networks with dominant units”. *Journal of Econometrics* 219.2, pp. 507–541.

Schiaffi, S. et al. (2013). “The Granularity of the Stock Market: Forecasting Aggregate Returns Using Firm-Level Data”. *Rivista di Politica Economica* 4, pp. 141–169.

Yan, W. (2011). “Role of diversification risk in financial bubbles”. *Swiss Finance Institute Research Paper* 11-26.