

# What do panel data say on inequality and GDP? New evidence at US state-level

Mauro Costantini\*

*Department of Economics and Finance, Brunel University London, UK*

Antonio Paradiso

*Department of Economics of Ca' Foscari University, Venice, Italy*

---

## Abstract

This paper investigates the empirical relationship between inequality and per capita GDP for the US using panel data for 50 states over the period 1960-2015. The paper uses new recently released data for inequality and a novel panel estimator with interactive fixed effects. The empirical results seem to suggest the existence of a S-shaped relationship between inequality and per capita GDP.

*Keywords:* Inequality, GDP, panel data, interactive fixed effects.

*JEL Codes:* C33, O15, R10.

---

## 1. Introduction

The empirical literature concerning the effects of per capita GDP on inequality (Gini) is remarkable. The results are mixed, depending on the functional form and the methodology used, so that no unambiguous conclusions have been drawn upon the relationship. In particular, this seems to be true for those analyses related to the US, which have shown the existence of U-shaped, inverted U-shaped, S-shaped and N-shaped relationship between Gini and per capita GDP (for a brief survey, see Table A1 in online Appendix).

The aim of this paper is to revisit the empirical relationship between inequality and per capita GDP for the US using panel data for 50 states over the period 1960-2015. The advantages of using panel data at state level have been emphasized by Kim et al. (2011). This paper not only applies panel data at state-level, but it also conducts a preliminary inspection of the data using a nonparametric approach to determine the functional form of the relationship between inequality and GDP.

The paper attempts to contribute to the literature on the US in some respects. First, the paper uses new recently released data for inequality. Second, the relationship between Gini and per capita GDP is estimated using a recently developed penalized principal component (PPC) estimator by Li et al. (2016), which accounts for cross-sectional dependence and multiple structural breaks.

The empirical findings show the existence of a stable S-shaped relationship between Gini and per capita GDP when using the PPC estimator.

---

\*Corresponding author: Department of Economics and Finance, Brunel University London, Kingston Ln, London, Uxbridge UB8 3PH (UK). Telephone: +44 (0)1895 267958

*Email addresses:* [mauro.costantini@brunel.ac.uk](mailto:mauro.costantini@brunel.ac.uk) (Mauro Costantini), [antonio.paradiso@unive.it](mailto:antonio.paradiso@unive.it) (Antonio Paradiso)

The rest of the paper is organized as follows. Section 2 illustrates the estimator by Li et al. (2016). Section 3 describes data and discusses empirical results. Section 4 concludes.

## 2. Li et al. (2016)'s PPC estimation method

Li et al. (2016) consider the following panel data model with interactive fixed effects

$$Y_{it} = \beta_t' X_{it} + \lambda_i' f_t + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where  $X_{it}$  is a  $p \times 1$  ( $p \equiv p_{NT}$ ) vector of explanatory variables,  $\beta_t$  is a  $p \times 1$  vector of unknown coefficients that may vary over time,  $\lambda_t$  and  $f_t$  are  $R_0 \times 1$  vectors of unobservable factor loadings and common factors, respectively, which can be correlated with  $X_{it}$ .  $\varepsilon_{it}$  denotes the idiosyncratic error term.<sup>3</sup>

In order to estimate the regression coefficients and the factor loadings, Li et al. (2016) develop a penalized principal component (PPC) estimator. The following objective function is first defined

$$\tilde{Q}_{NT,\gamma}(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{F}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - X_{it}' \beta_t - \lambda_i' f_t)^2 + \frac{\gamma}{T} \sum_{i=2}^T \dot{\omega}_t \|\beta_t - \beta_{t-1}\|, \quad (2)$$

where  $\gamma \equiv \gamma_{NT} > 0$  is a tuning parameter and  $\dot{\omega}_t$  is a data driven weight.

As for the estimation,  $\mathbf{F}$  is first concentrated out

$$\hat{Q}_{NT,\gamma}(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\Lambda}) + \frac{\gamma}{T} \sum_{t=2}^T \dot{\omega}_t \|\beta_t - \beta_{t-1}\|, \quad (3)$$

where  $\hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = \frac{1}{NT} \sum_{t=1}^T (Y_t - X_t \beta_t)' \mathbf{M}_{\boldsymbol{\Lambda}} (Y_t - X_t \beta_t)$ . Then,  $\boldsymbol{\Lambda}$  is concentrated out

$$\bar{Q}_{NT,\gamma}(\boldsymbol{\beta}) = \bar{Q}_{NT}(\boldsymbol{\beta}) + \frac{\gamma}{T} \sum_{i=2}^T \dot{\omega}_t \|\beta_t - \beta_{t-1}\|, \quad (4)$$

where

$$\bar{Q}_{NT}(\boldsymbol{\beta}) = \frac{1}{N} \sum_{r=R_0+1}^N \mu_r \left[ \frac{1}{T} \sum_{t=1}^T (Y_t - X_t \beta_t)(Y_t - X_t \beta_t)' \right]. \quad (5)$$

The new penalized estimation procedure is called *adaptive group fused LASSO*. In order to estimate  $\boldsymbol{\beta}$ , Li et al. (2016) follow the method in Bai and Ng (2002)

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \hat{Q}_{NT,\gamma}(\boldsymbol{\beta}, \hat{\boldsymbol{\Lambda}}) \quad (6)$$

and

---

<sup>3</sup>In Li et al. (2016)'s paper,  $\beta_t^0 = \alpha_j^0$  for  $t = T_{j-1}^0, \dots, T_j^0 - 1$  with  $j = 1, \dots, m^0 + 1$ ,  $T_0^0 = 1$ , and  $T_{m^0+1}^0 = T + 1$ , where the superscript 0 indicates the true value. Further,  $T_j^0$ ,  $j = 1, \dots, m^0$ , and  $m^0 + 1$  indicate the existence of  $m^0$  unobserved breaks points/dates and the number of regimes, respectively.  $\boldsymbol{\beta} = (\beta_1', \dots, \beta_T')'$ ,  $\boldsymbol{\alpha}_m = (\alpha_1', \dots, \alpha_{m+1}')'$ ,  $\boldsymbol{\Lambda} = (\lambda_1, \lambda_2, \dots, \lambda_N)'$ ,  $\mathbf{F} = (f_1, f_2, \dots, f_T)'$ ,  $\mathcal{T}_m = (T_1, \dots, T_m)$ ,  $\boldsymbol{\alpha}_{m^0}^0 = (\alpha_1^0, \dots, \alpha_{m^0+1}^0)'$ , and  $\mathcal{T}_{m^0}^0 = (T_1^0, \dots, T_{m^0}^0)$ .

$$\left[ \frac{1}{NT} \sum_{t=1}^T (Y_t - X_t \hat{\beta}_t)(Y_t - X_t \hat{\beta}_t)' \right] \hat{\Lambda} = \hat{\Lambda} \mathbf{V}_{NT}, \quad (7)$$

where  $\mathbf{V}_{NT}$  represents a diagonal matrix that consists of the  $R_0$  largest eigenvalues of the matrix in the square brackets in (7). The common factor  $\mathbf{F}^0$  is estimated as follows

$$\hat{\mathbf{F}} = (\hat{f}_1, \hat{f}_2, \dots, \hat{f}_T)', \quad (8)$$

with  $\hat{f}_t = N^{-1} \hat{\Lambda}' (Y_t - X_t \hat{\beta}_t)$ . In order to estimate  $\beta^0$  and  $\Lambda^0$ , an iterative algorithm based on (6) and (7) should be used (see Li et al., 2016).

To determine the number of factors, Li et al. (2016) propose the following information criterion

$$BIC(R) = \ln V(R, \hat{\beta}_R) + \rho_1 R, \quad (9)$$

where  $V(R, \hat{\beta}_R) = \frac{1}{N} \sum_{r=R+1}^N \mu_r \left[ \frac{1}{T} \sum_{t=1}^T (Y_t - X_t \hat{\beta}_{t,R})(Y_t - X_t \hat{\beta}_{t,R})' \right]$ ,  $\hat{\beta}_R = (\hat{\beta}'_{1,R}, \dots, \hat{\beta}'_{T,R})'$ ,  $\hat{R} = \arg \min_{0 \leq R \leq \max}$ .<sup>4</sup> To select  $\gamma$ , the following information criterion is considered

$$IC(\gamma) = \ln[\tilde{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_\gamma})] + \rho_2 p(\hat{m} + 1), \quad (10)$$

where  $\tilde{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_\gamma}) = Q_{NT}(\tilde{\alpha}_{\hat{m}_\gamma}, \tilde{\Lambda}, \tilde{\mathbf{F}}, \hat{\mathcal{T}}_{\hat{m}_\gamma})$ ,  $\tilde{\mathbf{F}}$  is defined as in (8), with  $\hat{\Lambda}$  and  $\hat{\beta}_t$  replaced by  $\tilde{\Lambda}$  and  $\tilde{\alpha}_j(\hat{\mathcal{T}}_{\hat{m}_\gamma})$  when  $\hat{T}_{j-1} \leq t \leq \hat{T}_j - 1$ , and  $\hat{\gamma} = \operatorname{argmin}_\gamma IC(\gamma)$  (see Li et al., 2016).<sup>5</sup>

### 3. Data and empirical results

Data on real per capita GDP (*Ypc*), expenditure on health care (*HCpc*) and welfare (*Wpc*) are taken from [http://www.usgovernmentspending.com/download\\_multi\\_year](http://www.usgovernmentspending.com/download_multi_year). Gini index data (Gini) are from Frank (2009) (updated from the author up to 2015) available at [http://www.shsu.edu/eco\\_mwf/inequality.html](http://www.shsu.edu/eco_mwf/inequality.html). In the spirit of Mushinski (2001) and Duncan (2016), we use a nonparametric estimation technique to determine the functional form of the empirical relationship between inequality (Gini) and real per capita GDP.<sup>6</sup> The advantage of using this approach is that it imposes no functional restrictions on the form. In Figure B1 (see online Appendix), we report the scatter-plots of nonparametric fit for all the states. Overall, it is observed a well-defined S-shaped relationship, with the exception of Alaska, Montana, North Dakota, and Wyoming. In addition, we use a semiparametric fixed effects estimator by Baltagi and Li (2002) for a further check concerning the existence of a S-shaped relationship between Gini index and GDP. Figure C1 (see online Appendix) offers a clear-cut evidence of a S-shaped relationship.

Therefore, we estimate the following equation by Li et al. (2016)'s PPC estimator

$$Gini_{it} = \beta_{0t} + \beta_{1t} \ln(Ypc_{it}) + \beta_{2t} \ln(Ypc_{it})^2 + \beta_{3t} \ln(Ypc_{it})^3 + \lambda'_i f_t + u_{it}. \quad (11)$$

<sup>4</sup> $\hat{\beta}_{t,R}$  denotes the PCA estimator of  $\beta_t$  without penalization. In the analysis,  $\rho_1 = \frac{(N+T)p}{NT} \ln\left(\frac{NT}{N+T}\right)$ .

<sup>5</sup> $\tilde{\alpha}_{\hat{m}_\gamma} = \tilde{\alpha}_{\hat{m}_\gamma}(\hat{\mathcal{T}}_{\hat{m}_\gamma}) = [\tilde{\alpha}_1(\hat{\mathcal{T}}_{\hat{m}_\gamma})', \dots, \tilde{\alpha}_{\hat{m}_\gamma+1}(\hat{\mathcal{T}}_{\hat{m}_\gamma})']$  indicates the post-LASSO estimates based on the break dates in  $\hat{\mathcal{T}}_{\hat{m}_\gamma} = \hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma)$ . In the analysis,  $\rho_2 = c \ln(\min(N, T)) / \min(N, T)$ , where  $c$  is a positive constant.

<sup>6</sup>We use Nadaraya-Watson regression with Gaussian kernel function.

Table 1: Panel estimation results, 1960-2015

	Model I	Model II	Model III
intercept	48.796*** (11.526)	40.410*** (9.597)	24.417* (13.501)
$\ln(Ypc_{it})$	-14.151*** (3.392)	-11.831*** (2.797)	-6.921* (3.967)
$\ln(Ypc_{it})^2$	1.385*** (0.333)	1.166*** (0.270)	0.663* (0.386)
$\ln(Ypc_{it})^3$	-0.045*** (0.011)	-0.038*** (0.009)	-0.021* (0.013)
$\ln(HCpc_{it})$		0.001 (0.006)	
$\ln(Wpc_{it})$			-0.007*** (0.003)

Notes: The optimal number of factors is selected using BIC criterion with a maximum number of factors set to 5. \*\*\*, \*\*, and \* denote significance at the 1%, 5% and 10% level. Standard errors are in parenthesis.

Equation (11) refers to model I (see Table 1). Further, by following the literature (see, for example, Leigh et al., 2009; Esping-Andersen and Myles, 2009), we also consider real per capita expenditure on health care and welfare as control variables

$$Gini_{it} = \beta_{0t} + \beta_{1t}\ln(Ypc_{it}) + \beta_{2t}\ln(Ypc_{it})^2 + \beta_{3t}\ln(Ypc_{it})^3 + \theta_t\ln(Z_{it}) + \lambda'_i f_t + u_{it}, \quad (12)$$

where  $Z_{it} = HCpc_{it}$  and  $Z_{it} = Wpc_{it}$  refer to model II and III, respectively (see Table 1).

The empirical findings, which are reported in Table 1, show a stable S-shaped relationship between Gini and real per capita GDP for model I, with all the estimated coefficients significant (see also Yang and Greaney, 2017). When considering models II and III, the coefficients of GDP remain statistically significant; health expenditure is positive but not statistically significant, whereas welfare spending is negative and significant (see, for example, d'Agostino et al., 2016).

In Figure 1, we plot the effects of per capita GDP on inequality (see also Desbordes and Verardi, 2012) using models I, II, and III. Similar S-shaped dynamics are observed, with slightly different turning points.

#### 4. Conclusions

This paper revisits the empirical relationship between inequality and real per capita GDP for the US using panel data for 50 states over the period 1960-2015. A preliminary nonparametric inspection of the data and the usage of a novel panel estimator with interactive fixed effects and multiple structural breaks seem to suggest a S-shaped curve for the empirical relationship, which may be interpreted as an extension of a U-shaped relationship. Further research may benefit from the empirical findings in this paper so to elaborate on the theoretical aspects behind the S-shape curve. In addition, whenever data is available, further investigation may be carried out at metropolitan-level so look into similar S-shaped pattern.

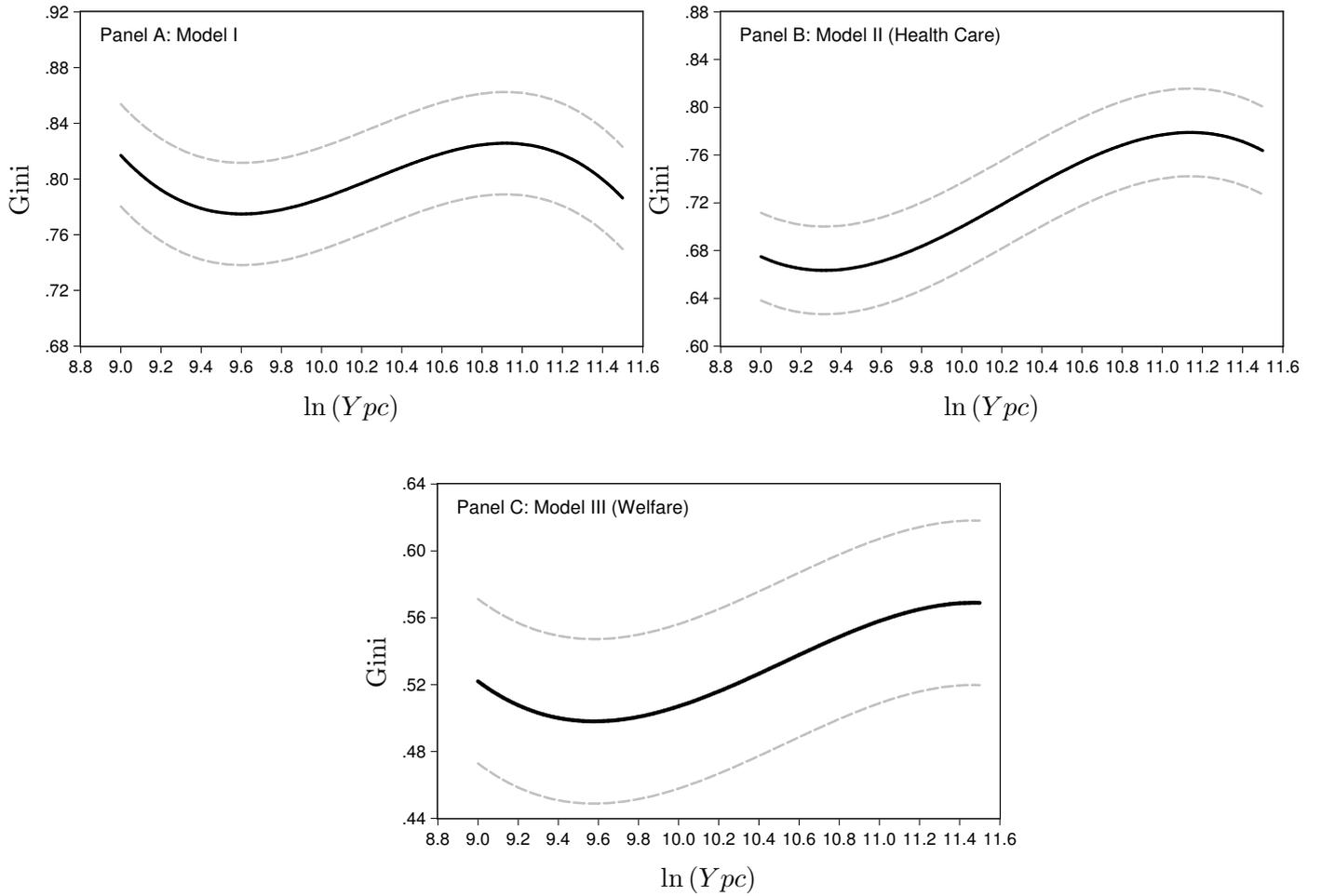


Figure 1: Fits of models I, II and III. Notes: Gini coefficients have been adjusted for the effects of the other explanatory variables and unobservable factors in the model (see Eq. (12)). Dotted lines corresponds to 95% confidence bands.

## References

- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70 (1), 191–221.
- Baltagi, B. H., Li, D., 2002. Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance* 3 (1), 103–116.
- d’Agostino, G., Pieroni, L., Procidano, I., 2016. Revisiting the relationship between welfare spending and income inequality in OECD countries. Tech. Rep. 72020, University Library of Munich, Germany.
- Desbordes, R., Verardi, V., 2012. Refitting the Kuznets curve. *Economics Letters* 116 (2), 258–261.
- Duncan, R., 2016. Does the US current account show a symmetric behavior over the business cycle? *International Review of Economics & Finance* 41, 202–219.
- Esping-Andersen, G., Myles, J., 2009. Economic inequality and the welfare State. In: Nolan, B., Salverda, W., Smeeding, T. M. (Eds.), *The Oxford Handbook of Economic Inequality*. Oxford University Press, Oxford, pp. 639–664.
- Frank, M. W., 2009. Inequality and growth in the United States: Evidence from a new state-level panel of income inequality measures. *Economic Inquiry* 47 (1), 55–68.
- Kim, D.-H., Huang, H.-C., Lin, S.-C., 2011. Kuznets hypothesis in a panel of states. *Contemporary Economic Policy* 29 (2), 250–260.
- Leigh, A., Jencks, C., Smeeding, T. M., 2009. Health and economic inequality. In: Nolan, B., Salverda, W., Smeeding, T. M. (Eds.), *The Oxford Handbook of Economic Inequality*. Oxford University Press, Oxford, pp. 384–405.
- Li, D., Qian, J., Su, L., 2016. Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association* 111 (516), 1804–1819.
- Mushinski, D. W., 2001. Using non-parametrics to inform parametric tests of Kuznets’ hypothesis. *Applied Economics Letters* 8 (2), 77–79.
- Yang, Y., Greaney, T. M., 2017. Economic growth and income inequality in the Asia-Pacific region: A comparative study of China, Japan, South Korea, and the United States. *Journal of Asian Economics* 48, 6–22.