ECONSTOR Make Your Publications Visible.

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Wang, Shengquan; Gao, Yuan; Luo, Mingjin

Article

What drives income inequality? Bayesian evidence from the emerging market economies

Journal of Applied Economics

Provided in Cooperation with: University of CEMA, Buenos Aires

Suggested Citation: Wang, Shengquan; Gao, Yuan; Luo, Mingjin (2023) : What drives income inequality? Bayesian evidence from the emerging market economies, Journal of Applied Economics, ISSN 1667-6726, Taylor & Francis, Abingdon, Vol. 26, Iss. 1, pp. 1-35, https://doi.org/10.1080/15140326.2023.2237692

This Version is available at: https://hdl.handle.net/10419/314233

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.



https://creativecommons.org/licenses/by/4.0/

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU





Journal of Applied Economics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/recs20

What drives income inequality? Bayesian evidence from the emerging market economies

Shengquan Wang, Yuan Gao & Mingjin Luo

To cite this article: Shengquan Wang, Yuan Gao & Mingjin Luo (2023) What drives income inequality? Bayesian evidence from the emerging market economies, Journal of Applied Economics, 26:1, 2237692, DOI: 10.1080/15140326.2023.2237692

To link to this article: https://doi.org/10.1080/15140326.2023.2237692

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



6

Published online: 18 Jul 2023.

ſ	
ι	

Submit your article to this journal 🖸

Article views: 1731



View related articles

View Crossmark data 🗹



Citing articles: 2 View citing articles

DEVELOPMENT ECONOMICS

OPEN ACCESS Check for updates

Routledae

Taylor & Francis Group

What drives income inequality? Bayesian evidence from the emerging market economies

Shengguan Wang (D^{a,b}, Yuan Gao^c and Mingjin Luo^d

^aInstitute of Advanced Studies in Humanities and Social Sciences, Beijing Normal University, Zhuhai, China; ^bInternational School of Business and Finance, Sun Yat-Sen University, Zhuhai, China; Bay Area International Business School, Beijing Normal University, Zhuhai, China; ^dInternational Development Cooperation Academy, Shanghai University of International Business and Economics, Shanghai, China

ABSTRACT

What drives income inequality in emerging market economies? To end it, we approach this topic utilizing Bayesian model averaging and 27 emerging market economies panel dataset from 1990 to 2019. First, differently, we find a U-shaped Kuznets curve in emerging market economies rather than the notable inverted U-shaped curve. Besides, we find four strictly robust determinants of income inequality in emerging market economies, including the population aging, the female labor force participation, the unemployment level and the share of labor compensation in output. These results remain stable through a series of robustness checks. In most cases, the government expenditure and real exchange rate could also robustly drive the dynamics of income inequality in emerging market economies. Further, a comparative analysis suggests that financial development, inflation, human capital condition, total factor productivity and urbanization are distinct determinants of income inequality in the market economies.

ARTICLE HISTORY

Received 21 February 2022 Accepted 13 July 2023

KEYWORDS

Income inequality; drivers; Bayesian model averaging; emerging market economies

1. Introduction

To determine the law of income distribution is an essential issue in political economy. It has been widely recognized that the reduction in income inequality is crucial for economic and social development as unequal income distribution may lead to economic, political and social instabilities (Alesina & Rodrik, 1994; Inagaki, 2010). Even more, as suggested by Rajan (2010) and Kumhof et al. (2015), income inequality sows the seeds of the 2007-2008 global financial crisis and contributes to financial instability (Wang & Luo, 2023).

The classical Kuznets nexus (Kuznets, 1955) of income inequality and economic growth have been widely examined in the market economies (MEs) for the last decades. The regular-U structure (Piketty, 2014) vs an inverted-U structure, the positive relationship (List & Gallet, 1999) and even the S-shaped relationship (Costantini & Paradiso, 2018) are documented in the US and other MEs. In addition,

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http:// creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

CONTACT Mingjin Luo 🖾 luo_mingjin@163.com 🖃 International Development Cooperation Academy, Shanghai University of International Business and Economics, Shanghai 200336, China

2 😔 S. WANG ET AL.

financial development contributes to the income inequality in the US via financial firms' rent-seeking channels (Stiglitz, 2013). Beck et al. (2010) suggest that lowerincome shares benefit from financial development. Furthermore, the MEs have employed the fiscal measures, e.g., progressive taxes (Piketty et al., 2014), transfer payment (OECD, 2012) and public expenditure on education (Agnello & Tavares, 2014; Hidalgo-Hidalgo & Iturbe-Ormaetxe, 2018; Roine et al., 2009), to tackle the inequality. Besides, physical (Baldi, 2013; Piketty, 2014) and human capital accumulation (Sequeira et al., 2017), price stability (Menna & Tirelli, 2017) are regarded as the determinants of income inequality in the MEs.

However, fewer studies have been conducted on the income inequality of the emerging market economies (EMEs) probably due to the data unavailability. As shown in Figure 1 and Figures A1 and A2, almost all EMEs witness acute income inequality for the last quarter-century with the Gini coefficients always being above 0.40-the "alerting line" as advised by the United Nations. In addition, the patterns of income inequality diverge significantly across countries. A bullish trend of income inequality is for China while a slight drop for Brazil and an inverted-U shape for Argentina, Bolivia, Russia and Ukraine. Given the stylized facts of income inequality, however, no enough efforts



Figure 1. Dynamics of Gini coefficients in the EMEs. *Notes*: The Gini coefficient is calculated by using the net (post-tax and post-transfer) income. The data is available at http://fsolt.org/swiid/. The Gini coefficients of other 18 EMEs are shown in Figures A1 and A2 in Appendix 3.

have been taken by researchers to investigate the causes of income inequality of the EMEs.

Before revisiting the theories of income inequality of the EMEs in Section 2, we conclude methodologically that previous studies on the determinants of income inequality mainly employ the panel data models within the frequentist estimation framework, i.e., the static panel model estimated by the LSDV method and the dynamic panel model estimated by the system GMM. However, previous studies only focus on the specific variables with prior motivation.

This paper intends to fill these gaps by utilizing the Bayesian model averaging (BMA) approach to investigate the determinants of income inequality in 27 EMEs ranging from 1990 to 2019. By employing the BMA approach,¹ we can include as many independent variables mentioned in the literature as possible without any prior distinctions. Within the traditional method, the arbitrary selection of control variables due to lack of uniform theoretical guidance may cause the fragility of regression analysis, namely model uncertainty (E. Leamer & Leonard, 1983). The BMA approach can overcome this problem and selects the robust determinants of income inequality.

We investigate the determinants of income inequality of the EMEs using the BMA strategy and conduct a series of robustness checks, including excluding the recent crisis period, alternating the model and coefficient priors, dropping the variable with many missing data, reducing the business cycle, using the extreme bound analysis (EBA) and the stochastic search variable selection (SSVS), adapting two machine learning strategies (Lasso and Random forest). First, differently, we find a U-shaped Kuznets curve in emerging market economies rather than the notable inverted U-shaped curve. We also find four strictly robust determinants of income inequality in the EMEs, including the population aging, the female labor force participation, the unemployment level and the share of labor compensation in output. These results remain stable through above robustness checks. In most cases, the government expenditure and real exchange rate could also robustly drive the dynamics of income inequality in the EMEs. Further, we build a comparison analysis on the drivers of income inequality between the EMEs and the MEs and find that financial development, inflation, human capital condition, total factor productivity and urbanization are distinct determinants of income inequality in the market economies.

The contributions of this paper are four-fold. First, to the best of our knowledge, this is the first paper to investigate the determinants of income inequality in the EMEs under an extensive uniform framework. Prior literature always investigates a specific variable while leaving others as the control variables or even neglecting them. Second, by reviewing the literature, we find that bulks of variables have been considered as the driven-factors of income inequality. Nondiscriminatory inclusion of factors in the regression specification may incur model uncertainty. This paper overcomes the model uncertainty by utilizing the BMA approach. Third, we find many interesting conclusions which may provide new research directions and reflections for the theorists and empiricists, such as the U-shaped Kuznets curve, the roles of population aging and female labor force participation, and why financial development is not a robust determinant of income inequality in the

¹See Steel (2020) on the application of model averaging in economics.

EMEs? Fourth, in the robustness checks, we employ two machine learning approaches, i.e., Lasso and Random forest, which helps broaden the toolkits of economists.

The remainder of the paper is organized as follows. Section 2 reviews the existing theories on the determination of income inequality in the EMEs. Section 3 develops the panel data model, the BMA strategy and describes the data. Section 4 presents the empirical results and a series of robustness checks. Section 5 conducts a comparative study based on the EMs sample. Section 6 discusses the conclusions and reflects the potential directions and revisions in this field. Section 7 concludes.

2. Theories of income inequality of EMEs revisited

2.1. Economic growth and income inequality

The first theory of income inequality is its relationship with economic growth. In seminal work of Kuznets (1955), he argued that the effects of growth on income inequality is dependent on the labor movements from the agricultural sectors to industrial sectors. Before the nineteenth century (or before the first industrial revolution), the agricultural sectors dominated and the farmer, instead of the peasant, benefit from the expansion of agricultural production (i.e., economic growth), thus income inequality increased. After the nineteenth century, the industrial sectors began to inflate (i.e., economic growth), which attracted amounts of labors move from the agricultural sectors to industrial sectors and increased the per capita income, thus income inequality reduced.

However, many competing mechanisms may channel growth to inequality. First, the sociopolitical theories argue that income inequality may incentivize behaviors detrimental to economic growth, e.g., crime, political instability and revolution (Alesina & Rodrik, 1994). Besides, to avoid revolution or regime change, governments will intervene income redistribution as inequality increases, while these intervenes generate distortions (Barro, 2000) and cause the reductions in economic growth. Second, the saving rate theories presume that the marginal propensity to save increases as income increases, thus the low-income have a higher marginal propensity to save than the high-income. Therefore, this theory implies that as income inequality increases, the saving rate in the economy will increases, which will translate to more domestic investments and economic growth.

Besides, increasing literature starts to reconcile existing studies by modelling a nonlinear effect of income inequality on growth conditional on the economic development stages. Galor and Tsiddon (1997), Galor (2000) and Galor and Moav (2004) argued that in the early physical capital driven stage of economic development, income inequality could stimulate economic growth by shifting resources from the high-income because of their higher marginal propensity to save than the low-income; when economic development enters the human capital driven stage, income inequality will be harmful for growth because of credit constraints on human capital accumulation. Bandyopadhyay and Basu (2005) agreed that the inequality-growth nexus is conditional on the nature of technology and the extent of redistribution. If the knowledge could spill over with low barriers, the technology is highly skill-intensive and the degree of redistribution is high, the effect of income inequality on growth will be positive; otherwise, the effect will be negative. There will be a sign reversal if allowing the nature of technology and the extent of redistribution change. Empirically, Lin et al. (2009) and Lin et al. (2014) confirm the nonlinear nexus between income inequality and economic growth along the development process.

2.2. Globalization and income inequality

The distributive effects of globalization have been widely discussed but do not reach a consistent prediction about how globalization affects inequality (Dorn et al., 2018). Stolper and Samuelson (1941) employed the classical Heckscher-Ohlin (HO) model to analyze the effect of globalization on inequality. According to HO model, countries should specialize in production in their relatively abundant factor and export these products. Trade will induce relative changes in product prices and thus increase returns to factors used intensively. The Stolper and Samuelson theorem predicts that trade will increase income inequality in advanced countries where capital and skilled labor are relatively abundant; however, trade could decrease income inequality in developing countries because unskilled labor is relatively abundant.

The HO model focuses on between-sector reallocation in production while neglecting within-sector reallocation and vertical specializations across countries. Therefore, the offshored and outsourced activities along the value chain may be relatively skill-intensive for the developing countries (Feenstra & Hanson, 1999). For example, FDI could increase the demand for the skilled labor and wage of the skilled labor, thus may widen the income gap in the developing countries (Feenstra & Hanson, 1997).

Cragg and Epelbaum (1996) argued that because of the rising exposure to import competition, the traded sector in the developing countries will be more skill-intensive than before, thus the wage gap between skilled and unskilled labor broadens. Many other studies (Sampson, 2014; Yeaple, 2005) propose that the more productive firms are, the more probabilities of exporting firms have, thus exporting firms will pay higher wage to hire skilled labor.

2.3. Demographic variables and income inequality

Although different theoretical mechanisms have been proposed, it is widely agreed that population aging could increase income inequality. Two reasons contribute to the population aging, i.e., declining fertility rate and longer life expectancy. Santelli et al. (2017) find that as the population is aging, the wealthier a family is, the lower fertility rate is. The fertility gap between the poor and the rich could increase the intergenerational income inequality (Dietzenbacher, 1989). Besides, population aging could affect income inequality via wealth inheriting. Miyazawa (2006) modelled this mechanism with the assumption that income inequality in the same generation originates from the gap between the high-income and the low-income. Descendants of the high-income could inherit more wealth than those of the low-income, thus the former becomes richer, while the latter remains at the low-income level.

The effects of female labor force participation on income inequality are inconclusive (Sudo, 2017), the sign relies on which women work and which earn more. Female labor force participation could reduce income inequality when it improves the financial status of single-mother households. The disproportional increase in the labor supply of the poorly educated women would equalize income distribution (Esping-Andersen, 2007),

6 😉 S. WANG ET AL.

this is the case when their partners become unemployed or reduce work hours for other reasons. Female labor force participation could also equalize income distribution when women live with a high-income man who withdrew partially or fully from the labor market (Cancian & Schoeni, 1998; Raaum et al., 2008). However, other channels may predict a positive nexus between female labor force participation and income inequality, for example, women living in a high-earned family or with assortative mating worked more (Esping-Andersen, 2007).

Human capital accumulation (e.g., educational expansion) has an ambiguous effect on income inequality (Knight & Sabot, 1983). Educational expansion affects income distribution by two offsetting effects, i.e., the composition effect and the wage compression effect. The initial educational expansion could increase the wage of more-educated workers, which rises wage inequality, namely "the composition effect"; as education expands, the supply of educated workers exceeds the demand, thus the wage premium of educated workers will diminish, and wage inequality will reduce.

2.4. Financial development and income inequality

Three major theoretical papers explain the nexus between financial development and income inequality, i.e., Banerjee and Newman (1993), Galor and Zeira (1993), Greenwood and Jovanovic (1990). Banerjee and Newman (1993) and Galor and Zeira (1993) predict that financial development could reduce income inequality, the former argues households' occupational choice dependent on credit availability, while the latter argues human capital investment dependent on credit availability. Differently, Greenwood and Jovanovic (1990) predicted an inverted-U relation between financial development and income inequality. They highlighted that household could earn capital income by accessing the financial intermediaries, initially, the poor could not afford using banks for their savings, but the rich could, thus income inequality will exacerbate with financial development; as poor households become rich, who could access bank finance; thus, inequality will reduce with financial development.

2.5. Government intervene and income inequality

Government could redistribute resources via public expenditure to intervene income distribution (Doerrenberg & Peichl, 2014). For example, according to the human capital investment theory (Gruber, 2013) which argues education could increase a person's skill and productivity, government could provide subsidies to the low-income households for early-education investments to improve their education attainments, which will decrease inequality and increase intergenerational mobility (Yang & Qiu, 2016).

Monetary policy could affect income inequality in an ambiguous way (Coibion et al., 2017). First, expansionary monetary policy could increase inequality via asset prices and inflation. The rising of asset prices benefits the capital returns of the high-income, and the inflation will depreciate the income of the low-income because of their possessions of liquid assets. Second, expansionary monetary policy could reduce inequality. An unexpected expansion of monetary policy could benefit the borrowers, always the low-income (Doepke & Schneider, 2006). Labor earnings of the low-income are most affected by the economic activity (Heathcote et al., 2010), the expansionary monetary policy could

inflate the economy, thus could reduce income inequality. We will use inflation (Monnin, 2014; Székely, 2003) as a proxy of monetary policy because the direct measure of exogenous monetary policy is absent.

2.6. Other theories

Besides the above widely discussed theories, exchange rate (Jeanneney & Hua, 2001; Min et al., 2015), physical investment (Thorbecke & Charumilind, 2002), total factor productivity (Espoir & Ngepah, 2021), the share of labor compensation in output (Hémous & Olsen, 2022), unemployment level (Acemoglu, 1999; Dabla-Norris et al., 2015), urbanization (Oyvat, 2016) are often covered in the previous literature, we will also introduce them into our analysis.

Exchange rate could affect the relative price of tradable products and untradeable products, by which affect the income gap between skilled and unskilled workers. Skilled worker could benefit from the technology change, thus technology change increases income inequality by rising the wage premium of skilled workers (Dabla-Norris et al., 2015). Unemployment could increase inequality via directly affecting the share of labor income (Furceri & Ostry, 2019).

3. Methodology and data

3.1. Fixed effect model

We utilize the following two-way fixed effect panel data model to examine the effect of various factors on the income inequality. In practice, we conduct the Hausman Test (see Table 1) and accept the specifications of the fixed effect (FE) model:

	er meente meetaan		
Regressor	Coefficient	Regressor	Coefficient
OLD	-0.4280***	TFP	-1.3700
	(-4.46)		(-1.24)
CRE	0.0008	LAB	-9.4410***
	(0.15)		(-6.73)
FDI	0.0620	XR	-6.96e ^{-8**}
	(1.83)		(-2.71)
FEM	-0.0941	OPNESS	17.6000***
	(-1.93)		(7.25)
G	-0.2710***	GDP	-10.6800***
	(-7.02)		(-8.83)
I	0.0185	GDP ²	0.4570***
	(0.21)		(9.28)
INF	0.00005	UNE	0.0365
	(0.22)		(1.15)
HC	-0.4500	URB	0.0203
	(-0.51)		(0.76)
Country Fixed Effect		Yes	
Time Fixed Effect		Yes	
Observations		780	
Hausman test		25.96	
		(0.0172)	

Table 1. Determinants of income inequality in the EMEs: FE results.

p value in the bracket. *, ** and *** denote that the coefficients are significant at 10%, 5% and 1% levels respectively. The dependent variable is Gini_net here.

8 🔄 S. WANG ET AL.

$$Gini_{it} = \alpha_i + \beta_t + \gamma GDP_{it} + \delta GDP_{it}^2 + \mu XR_{it} + \nu INF_{it} + \zeta G_{it} + \eta I_{it} + \theta OPNESS_{it} + \kappa FDI_{it} + \lambda CRE_{it} + \pi HC_{it} + \rho TFP_{it} + \tau LAB_{it} + \chi UNE_{it} + \varphi OLD_{it} + \phi URB_{it} + \omega FEM_{it} + \varepsilon_{it}$$

(1)

where α_i and β_t are the country and time fixed effect, respectively, i = (1, 2, ..., N)denotes a country, $t(=1,2,\ldots,T)$ denotes time. α_i is used to control those timeinvariant variables but changing with countries. For example, the law system may affect the income distribution, but the law system doesn't change in a short time. β_t is used to control those country-invariant variables but changing with time. For example, the financial crisis shock may affect the income inequality of all countries but only change with time. ε_{it} is the disturbance term that is unobservable and uncorrelated with the independent variables and distributes identically and independently. We collect 16 potential determinants of income inequality by reviewing the existing literature extensively. GDP_{it} , the logarithm of real output per capita, and its quadratic term, GDP_{it}^{2} , are used to exploit the potential inverted U axis between economic development and income inequality. We expect that $\gamma > 0$ and $\delta < 0$. Two price variables, XR_{it} and INF_{it} , capture the effects of the real exchange rate and inflation on the income inequality, respectively, but the sign of the coefficients, μ and ν , are ambiguous. G_{it} is the government expenditure scaled by GDP and is expected to be a negative sign. I_{it} is the domestic investment scaled by GDP and represents the physical capital formation and is expected to be a negative sign. $OPNESS_{it}$, equaling the export scaled by GDP, is incorporated in the analysis on trade and inequality based on the literature and is expected to be a negative sign. We incorporate the growth rate of inward foreign direct investment, FDI_{it}, into the analysis and expect that it has an ambiguous sign. CRE_{it}, the bank credit scaled by GDP, is a proxy of financial development in the EMEs and is expected to be a negative sign. HC_{it} is the human capital based on years of schooling and returns to education and is expected to be a negative sign. TFP_{it} is the total factor productivity and mirrors the technological development and is expected to be an positive sign. LAB_{it} is the share of labor compensation in output and embodies the structure of factor income and is expected to be a negative sign. UNE_{it} is the unemployment rate and is expected a positive sign. Finally, three population structure variables are included, OLD_{it}, population over 65 years old scaled by the total population (with an expected positive sign), URB_{it} , urban population scaled by the total population (with an expected negative sign), FEM_{it} and the female workforce scaled by the total workforce (with an expected negative sign).

Due to the lack of the uniform theoretical framework, we cannot convincingly tell the determinants of income inequality in the EMEs. As the extant empirical literature often considers as many as possible variables in control variables, an econometric issue may arise, the model uncertainty which incurs the fragility of regressors (E. Leamer & Leonard, 1983). This paper intends to overcome this problem by employing the BMA strategy.

3.2. BMA approach

The basic idea of model averaging is to consider and estimate all candidate models and then report a weighted average as the estimate of the effect of interest variables (Moral-Benito, 2015). We select *Accessisdenied AccessisdeniedAccessisdenied* independent

variables through reviewing the literature. Thus, the number of candidate models is $2^{q}(=65536)$.

$$\hat{\Psi}_{MA} = \sum_{h=1}^{2^q} \omega_h \hat{\Psi}_h \tag{2}$$

where $\hat{\Psi}_{MA}$ is the coefficient vector estimated by model averaging, $\hat{\Psi}_h$ is the coefficient vector of the h^{th} model and ω_h is the posterior model probability related to the h^{th} model.

Each model will be estimated in a Bayesian way, and the coefficients' posterior distribution of the h^{th} model is as follows:

$$g(\Psi^{h}|Y, M_{h}) = \frac{f(Y|\Psi^{h}, M_{h})g(\Psi^{h}|M_{h})}{f(Y|M_{h})}$$
(3)

where *Y* is the dependent variable, $f(Y|\Psi^h, M_h)$ is the likelihood function, $g(\Psi^h|M_h)$ is the coefficients' prior distribution of the h^{th} model.

Bayesian inference demonstrates that the posterior model probability can be regarded as the weight of model averaging. Given that the h^{th} prior probability is $P(M_h)$, the posterior probability can be calculated by the Bayesian approach as follows:

$$P(M_h|Y) = \frac{f(Y|M_h)P(M_h)}{f(Y)}$$
(4)

The implementation of BMA entails two different types of prior probability: one is the prior probability of coefficient space, and another is the prior probability of model space. To calculate Equation (4), $f(Y|M_h)$, which is the marginal likelihood function, should be calculated in advance. By combing Equation (3) and $g(\Psi^h|Y, M_h)d(\Psi^h) = 1$, we can derive.

$$f(Y|M_h) = \int f(Y|\Psi^h, M_h) g(\Psi^h|M_h) d(\Psi^h)$$
(5)

According to E. Leamer and Leonard (1983), the posterior coefficient probability is

$$g(\Psi|Y) = \sum_{h=1}^{2^q} P(M_h|Y)g(\Psi^h|Y, M_h)$$
(6)

Point estimation can be obtained by the expectation and variance of the coefficient distribution, namely the posterior mean (P.M) and the posterior standard deviation (P. SD), respectively:

$$E(\Psi|Y) = \sum_{h=1}^{2^q} P(M_h|Y) E(\Psi^h|Y, M_h)$$
(7)

$$V(\Psi|Y) = \sum_{h=1}^{2^{q}} P(M_{h}|Y) V(\Psi|Y, M_{h}) + \sum_{h=1}^{2^{q}} P(M_{h}|Y) (E(\Psi|Y, M_{h}) - E(\Psi|Y, M_{h}))^{2}$$
(8)

10 😉 S. WANG ET AL.

Besides, the BMA provides the posterior inclusion probability (P.IP) to convey the probability statement regarding the importance of a regressor. P.IP is the sum of the model posterior probability of the specific variable, $P.IP=P(\beta_k \neq 0|Y) = \sum_{\beta_k \neq 0} P(M_k|Y)$, β_k denotes the coefficient of the k^{th} regressor. Implementation of BMA procedures presents three challenges. First, we need to choose the prior model probability and the prior for coefficients. In the baseline study, following the specifications of Ley and Steel (2012), we use the random model prior probability, which means that the likelihood of the h^{th} model containing k^h variables is $P(M_h) = \theta^{k^h}(1-\theta)^{q-k^h}$ with θ subjective to *Beta* distribution. Thus, the prior model size is θq . Following Leon-Gonzàlez and Montolio (2015), the prior for the coefficients is specified as g type and the prior coefficients probability of the h^{th} model is $\beta_h | \sigma^2, M_h, g \sim N(0, \sigma^2 g(X'_h X_h)^{-1})$, where $g = \max(2^q, q^2)$. In the section of robustness check, we will alternate the prior probabilities. Second, the marginal likelihood $(f(Y|M_h))$ depends on an integral that cannot be solved analytically but can be calculated through a computationally intensive numerical approach. Finally, as the model space in our empirical application contains 65,536 models, a full estimation will be implemented.

3.3. Advantages of BMA

As mentioned before, BMA could be used to solve the model uncertainty issue. In this study, we regress the income inequality on as many as possible independent variables to identify the robust determinants of income inequality in the EMEs. Many competing theories of the determination of income inequality in the EMEs have been proposed as stated in Section 2. A natural question is which theory is robust? In traditional regression analysis, we often cover many potential independent variables in one model. However, it is problematic once we add the irrelevant factors to the model and omit the relevant factors, which is the model uncertainty issue (E. Leamer & Leonard, 1983).

The BMA strategy does not rely on any priors on the theories of income inequality, but robustly relies on two priors on the model space and the coefficient space under the Bayesian framework. In practice, the number of candidate models depends on the number of independent variables, any variable could be included or not, thus 2^{q} (*q*isthenumberofindependentvariables) models exist (in practice, it is often unnecessary to consider all models). For each model, we set a prior model probability distribution, and we could obtain the posterior distribution after the Bayesian sampling. For each variable, we set a prior coefficient probability distribution and get the posterior distribution through the Bayesian procedure. Finally, we could estimate the coefficients by averaging the estimates of the different models under consideration, each weighted by its model probability. Searching the literature, we find Furceri and Ostry (2019) is a similar paper. They also use the model averaging technique, specifically WALS (weighted-average least squares). However, their application is under the frequentist framework, ours is under the Bayesian framework.

EBA was ever a popular approach to investigate the determinants of variables of interest. However, EBA has been criticized for its lack of statistical foundations, and its implementation also restricted to a limited number of models (Chakrabarti, 2001).

Berger and Sellke (1987) have suggested that conventional sensitivity analyses (including EBA) overstate significance and confidence intervals in the absence of a full account of model uncertainty.

3.4. Data

The original data of the independent variables employed are mainly collected from two data libraries, Penn World Table 10.0 (PWT 10.0) and World Bank database (WB). The dependent variable is income inequality. Although there are diverse databases of income inequality, these databases have weaknesses of less comparability and limited coverage. The Standardized World Income Inequality Database (SWIID) provides a set of comparable income inequality data with the widest possible coverage across countries and over time (Solt, 2015). Many researchers have conducted studies using this database. For example, Acemoglu et al. (2015) use SWIID to examine the impacts of democracy on the inequality. Table A1.1 provides detailed information for each variable: definition, source and calculation. Given the availability of data, the sample covers a period from 1990 to 2019 with annual frequency. Although SWIID starts from 1960, at least two reasons constraint our sample from 1990. First, SWIID is severely data missing from 1960 to 1990. From 1960 to 2019, the number of missing of the SLOT database is 904, while the number of missing is 768 from 1960 to 1990. Second, the BMA approach we used require no more than three continuous missing observations. The sample contains 27 EMEs which constitute the J.P. Morgan Emerging Markets Bond Index Global (EMBI) or Morgan Stanley Capital International Emerging Market Index (MSCI-EMI) except for Israel. The list of EMEs is shown in Table A1.2. To avoid the influence of serial correlation, we adapt the interactive fixed effect (Bai, 2009) to deal with the original data.

Table A1.4 reports the basic statistics of the variables involved. Gini_net and Gini_market are the two alternative dependent variables. The former is measured by net income while the latter is measured by market income. We find that Gini_market (mean value 46.4691) is often higher than Gini_net (mean value 41.84346), but the standard error reverses in overall and between groups. This fact implies that personal income taxation flattens income inequality to certain degree (Agnello & Tavares, 2014). The volatility of real exchange rate, inflation, bank credit and urbanization are greater than the others, implying that the disparities of these variables across country and time are significant. Thus, the inclusion of country and time fixed effects is reasonable.

To favor the empirical analysis in Section 4, we present a correlation matrix in Table A1.5 as the preliminarily descriptive evidence. What we are interested in is the first column, which offers the correlation coefficients between income inequality and each independent variable. As we can see, real output per capita and its quadratic term, bank credits, female workforce ratio, government expenditure, human capital condition, the share of labor compensation in output, unemployment, urbanization and population aging are negatively correlated with the income inequality at least 5% significance level. Intuitively, the Kuznets curve may disappear because the coefficients of real output per capita and its quadratic term are both negative, implying the axis of symmetry locates

the negative space. However, we need more cautious inference rather than a simple statistical analysis.

4. Empirical results and analysis

In this section, we present the empirical results. In addition to the baseline results, we consider the potential effects of the recent crisis period on the findings by excluding the sample during the crises period. We also conduct a series of robustness checks.

4.1. Baseline results

Without considering model uncertainty, Table 1 presents the traditional panel estimation results under FE specification (1) and Hausman test statistics.² First, the Hausman test rejects the null hypothesis with χ^2 statistics of 25.96, implying that the FE model is preferred. Thus, we hereafter only report the FE results. Second, Table 1 finds the coefficients of real output per capita and its quadratic term are significant negative and positive, respectively, which implies that the nexus between real output per capita and income inequality is positive once the real output per capita is larger than 11.6849 $(=(-10.6800)/(2 \times 0.4570))$. Therefore, Table 1 supports the U-shaped instead of the inverted-U shaped Kuznets curve that exists in the EMEs, which is consistent with Bandyopadhyay and Basu (2005), but different from Tsount and Osueke (2014). Third, we find that only few factors in the left 14 potential drivers are statistically significant in shaping the inequality in the EMEs, i.e., the population aging, the government expenditure, the share of labor compensation in output, the real exchange rate and trade openness are significant at 1% and 5% levels, respectively. However, the traditional panel model provides limited information on the determination of inequality in the EMEs as some drivers advocated by the previous studies are not robust at least in the specification (1) due to model uncertainty.

Table 2 reports the baseline results using the BMA approach with both Gini_net and Gini_market as the dependent variables, respectively. Following Raftery (1995), evidence

	P.IP	P.M	P.SD		P.IP	P.M	P.SD
Regressor		Gini_net		Regressor		Gini_market	
FDI	1.0000	.1956	.0338	FDI	1.0000	.1841	.0328
G	1.0000	2582	.0410	G	1.0000	2213	.0394
OLD	1.0000	4596	.0769	LAB	1.0000	-9.9294	1.4239
LAB	1.0000	-1.5778	1.5219	GDP	1.0000	-1.6611	1.3095
GDP	1.0000	-11.6982	1.3418	GDP ²	1.0000	.4330	.0544
GDP ²	1.0000	.4764	.0525	UNE	1.0000	.2156	.0305
UNE	1.0000	.1939	.0312	OLD	.9463	2453	.0963
XR	.9703	-9×10^{-7}	3×10^{-8}	XR	.9363	-8×10^{-7}	3×10^{-8}
FEM	.5127	0065	.0730	I	.7273	.1671	.1253
				FEM	.7033	0895	.0703

 Table 2. Determinants of income inequality in the EMEs: baseline results.

²Random effect estimation results are presented in Table A2.1.

for a regressor with a P.IP of 0.50 to 0.75 is considered weak, of 0.75 to 0.95 positive, of 0.95 to 0.99 strong and above 0.99 very strong. Therefore, we only care about the independent variables with P.IP of 0.50 or above.

The BMA strategy identifies nine determinants of income inequality in the EMEs when the dependent variable is Gini_net. We observe that the results support the U-shaped Kuznets curve in the EMEs as Table 1. The coefficients of GDP and GDP^2 are -11.6982 and 0.4764, respectively. Then, we can deduce that the optimal point of logarithmic output per capita is $12.2777(=-(-11.6982)/[2\times (0.4764)])$. Combined with the statistics of GDP reported in Table A1.4, we speculate that the half of EMEs have crossed the threshold and positioned the right side of the U-shaped Kuznets curve, which means that income distribution will deteriorate as these economies perform better. Interestingly, if we select seven representative countries (i.e., Argentina, Bolivia, Chile, China, Colombia, Korea and Peru) as the sample, the results suggest that these countries have located the right side of the U-shaped Kuznets curve.

Combined with the results using Gini_market as the dependent variable, we find additional seven variables are robust determinants of income inequality in the EMEs, i.e., FDI, the government expenditure, population aging, the share of labor compensation in output, the unemployment, the real exchange rate and the female workforce ratio.

FDI is detrimental to income distribution in the EMEs, which is consistent with existing literature (Bogliaccini & Egan, 2017). Different from the previous studies (Cevik & Correacaro, 2015), we find that government expenditure (G) is a robust factor that could equalize the income distribution. Population aging (OLD) is negatively correlated with the income inequality robustly in the EMEs. This result is opposite to the existing literature, such as Dong et al. (2018). The increase of the share of labor compensation in output (LAB) reflects the reduction of functional income inequality. In line with Dabla-Norris et al. (2015), unemployment ratio (UNE) polarizes the income distribution in the EMEs because the low-income labours rather than the high-income labours become un-employees in the EMEs. Decrease in exchange rate (XR) increases the income inequality by increasing the cost of exported firms and decreases the wage of unskilled labors in the EMEs. A robust but often neglected variable is the female labor force participation (FEM), which could reduce the income inequality (Değirmenci, 2018).

Interestingly, we find that financial development (CRE) is not a robust determinant of income inequality in the EMEs. The idea that financial development reduces the income inequality by easing financial market constraints remains being suspected in the EMEs. Lack of diverse financial systems may pose an obstacle to financial access to the poor (Claessens & Perotti, 2007). We also find that trade (OPNESS) is not a key factor of income inequality in the EMEs, but FDI is.

4.2. Excluding the sample during the crisis period

The recent global financial crisis has substantially redistributive consequences (Goda et al., 2017), which may challenge the results above. To check the effect of financial crisis,

14 🛭 😔 S. WANG ET AL.

	P.IP	P.M	P.SD		P.IP	P.M	P.SD
Regressor		Gini_net	<u> </u>	Regressor		Gini_market	
FDI [†]	1.0000	.4619	.1056	FDI [†]	1.0000	.5624	.1173
FEM [†]	1.0000	.3082	.0342	FEM [†]	1.0000	.3622	.0385
OLD^{\dagger}	1.0000	-1.2588	.0935	OLD [†]	1.0000	6464	.1050
нс	1.0000	-5.6295	.9160	TFP	1.0000	7.7553	1.4809
LAB [†]	1.0000	-15.3313	4.1915	OPNESS	1.0000	13.9838	2.7911
OPNESS	1.0000	21.9777	2.5032	UNE [†]	1.0000	.6427	.0444
UNE [†]	1.0000	.4788	.0443	XR^{\dagger}	.9647	0005	.0002
G [†]	.9410	2016	.0819	HC	.9623	-2.9885	1.0081
URB	.7203	.0370	.0283	CRE	.6997	0174	.0139
CRE	.6603	.01437	.0128				
1	.5860	3245	.3260				

Table 3. Determinants of income inequality in the EMEs using the BMA: subtracting the crises.

[†]indicates the variable is also reported in baseline results in Table 2.

we exclude the sample during the crisis (from 2005 to 2009) from the original sample. The results are reported in Table 3.

The results suggest that the crises do affect the determination of income inequality in the EMEs to some extent. When the dependent variable is Gini_net, the BMA procedure selects 11 regressors as the driven-factors of income inequality in the EMEs. Different from the results above, bank credit (CRE) is a driver of income inequality with a posterior inclusion probability of 0.6603 in the EMEs. The Kuznets curve fails to be a robust component of income inequality, and exchange rate (XR) also play no roles in shaping the inequality. However, the human capital (HC), the trade openness (OPNESS) and urbanization (URB) become the determinants of income inequality in the EMEs. FDI, the female labor force participation (FEM), the population aging (OLD), the labor share (LAB) and the government expenditure (G) remain as the robust determinants of income inequality in the EMEs as Table 2. When the dependent variable is Gini_market, government expenditure (G), urbanization (URB) and physical capital investment (I) lose their explainable power as presented in Table 2.

Overall, no matter we subtract the crises period or not, we always find that foreign direct investment (FDI), female labor force ratio (FEM), the population aging (OLD), the unemployment level (UNE) are still the robust inequality drivers in the EMEs.

4.3. Robustness checks

We conduct a series of robustness checks, including alternating the model priors and the coefficient priors, excluding the variables with many missing values, using 5-year average sample to reduce business cycle, adapting the EBA approach, overcoming the multi-collinearity issue via SSVS, and two machine learning techniques (i.e., Lasso and Random forest).

4.3.1. Alternate the model priors

In the baseline analysis, we employ the random model prior. Here, we alternate the model priors with the fixed model prior, uniform model prior and Posterior Inclusion

Prior (PIP) model prior, which are the popular specifications in the BMA literature. Fixed model prior means that we set prior model size as 8 in mean, which implies that $\theta q = 8$, thus $\theta = \frac{1}{2}$. Uniform model prior assigns the equal probability to each candidate model, $P(M_h) = 2^{-q}$. As to PIP prior, we assign the prior probability of the robust factors in the benchmark model with a low value and the remaining variables with a high value.³ Explicitly, we assign FDI, G, OLD, LAB, GDP, GDP², UNE, XR, FEM with 0.01 of prior probability and others with 0.5 of prior probability.

The results are presented in Table 4. Compared to Table 2, we find that the baseline results are robust. Population aging (OLD), female workforce scaled by total workforce (FEM), unemployment (UNE) and labor income share (LAB) are identified as the most robust determinants under the three alternative specifications. Also, trade openness (OPNESS), human capital condition (HC) and urbanization (URB) are considered to be robust in models with different model priors.

Table 4. Determinants of income inequality in the EMEs using the BMA: alternative model priors.

	Fix		xed model prior		Uniform model prior		P	IP model prio	or
Regressor	P.IP	P.M	P.SD	P.IP	P.M	P.SD	P.IP	P.M	P.SD
FEM [†]	1.0000	0.3674	0.0299	1.0000	0.3691	0.0297	1.0000	0.3724	0.0299
OLD^{\dagger}	1.0000	-1.2890	0.0764	1.0000	-1.2908	0.07617	1.0000	-1.2997	0.0752
HC	1.0000	-7.3713	0.7745	1.0000	-7.4057	-0.7642	1.0000	-7.6948	0.6313
OPNESS	1.0000	44.2814	3.9711	1.0000	44.3850	3.9235	1.0000	44.6333	3.9030
UNE [†]	1.0000	0.4071	0.0406	1.0000	0.4052	0.0412	1.0000	0.3889	0.0341
LAB [†]	0.9997	-11.8799	2.9105	1.0000	-11.9457	2.8058	0.9937	-12.1630	3.3322
URB	0.9660	0.0584	0.0177	0.9873	0.0590	0.0158	0.5950	0.0626	0.0145

The dependent variable here is Gini_net. The similar results can be reached and thus are not reported when the dependent variable is Gini_market. [†] indicates the variable is also reported in baseline results in Table 2.



Figure 2. Model size distributions with alternative model priors. *Note*: from upper left to low right, the model priors are uniform, fixed, random and PIP, respectively.

³This practice is intuitive and counterfactual. If the results still hold when using a low prior probability for the selected robust variables in the baseline and a high prior probability for others, it implies that the results are robust.

16 👄 S. WANG ET AL.

Furthermore, we can check the robustness by examining the distribution of model size. As evidenced in Figure 2, the significant disparity in prior model size induced by different specifications of prior probability in model space converges to the similar posterior model size. The average model size is 8.10, 8.11, 8.26, 7.23, respectively, meaning that the model with 7 to 8 robust determinants can sufficiently explain the dynamics of income inequality in the EMEs, which is significantly different from the traditional frequentist estimation results as presented in Table 1.

4.3.2. Alternate the parameter priors

Similarly, we alternate the *g* type parameter priors in baseline specifications with three popular priors, the fixed-parameter prior, the empirical Bayesian local (EBL) parameter prior and Hyper-g parameter priors. The fixed parameter prior sets $\beta_h | \sigma^2, M_h, g \sim N\left(0, \sigma^2 g\left(X'_h X\right)^{-1}\right)$ with g = 8, which is different from the baseline specification with $g = \max(2^q, q^2)$. EBL parameter prior sets the distinctive *g* for each model, $g_h = \max(1, F_h^{OLS} - 1)$, and F_h^{OLS} is the standard OLS statistics for the h^{th} model (Hansen & Yu, 2001). The Hyper-g parameter prior set a shrinkage factor as $\frac{g}{1+g} \sim B\left(1, \frac{a}{2} - 1\right)$, where $a \in (2, 4]$ (Liang et al., 2008).

By alternating the specifications of prior probability in coefficient space, Table 5 shows the consistency of empirical results with the benchmark specifications. The robust determinants of income inequality in the EMEs are the female labor force ratio (FEM), the population aging (OLD), the unemployment levels (UNE), the labor income share (LAB) and the government expenditure (G) and other regressors as reported in Table 2.

As shown above, we investigate the model size distributions with three different parameter priors. Although the distributions of parameter priors diverge significantly, the posterior distributions are similar. The average model sizes are 11.69, 9.20, 9.12, respectively, which means that the model with 9 to 11 robust determinants can sufficiently explain the dynamics of income inequality in the EMEs. These results are consistent with Figure 3.

	Fixed parameter prior		Fixed parameter prior EBL parameter prior		Нуре	r-g parameter	r prior		
Regressor	P.IP	P.M	P.SD	P.IP	P.M	P.SD	P.IP	P.M	P.SD
FEM [†]	1.0000	0.3023	0.0339	1.0000	0.3618	0.0309	1.0000	0.3622	0.0307
OLD [†]	1.0000	-1.0672	0.0849	1.0000	-1.2765	0.0781	1.0000	-1.2782	0.0777
НС	1.0000	-5.3898	0.9435	1.0000	-6.9693	0.9267	1.0000	-7.0807	0.8739
OPNESS	1.0000	34.9591	4.5778	1.0000	43.4561	4.2000	1.0000	43.686	4.1244
UNE	1.0000	0.3647	0.0445	1.0000	0.4198	0.0044	1.0000	0.4161	0.0430
LAB [†]	0.9907	-10.3081	3.3762	0.9972	-11.7739	3.0722	0.9940	-11.6616	3.1037
\mathbf{G}^{\dagger}	0.8368	-0.1230	0.0811	0.6793	-0.1056	0.0886	0.6357	-0.0964	0.0877
URB	0.8188	0.0321	0.0221	0.9247	0.0516	0.0214			
GDP^{\dagger}	0.6621	-1.7528	2.2767						
1	0.6354	-0.1961	0.2100						
GDP ^{2†}	0.6206	0.0587	0.0867						
FDI [†]	0.5399	0.0658	0.0876						

Table 5. Determinants of income inequality in the EMEs using the BMA: alternative parameter priors.

The dependent variable here is Gini_net. The similar results can be reached and thus are not reported when the dependent variable is Gini_market. [†] indicates the variable is also robust in baseline results in Table 2.



Figure 3. Model size distributions with alternative parameter priors. *Note*: from upper left to low, the parameter priors are fixed, EBL and Hyper-g, respectively.

P.SD
.0325
.0392
1.4218
1.3060
.0520
.0395
.0947
3×10 ⁻⁸
.0365
.0681

Table 6. Determinants of income inequality in the EMEs using the BMA: excluding I.

[†]indicates the variable is also robust in baseline results in Table 2.

4.3.3. Excluding the variables with many missing values

We find the variable, physical capital investment (I), suffers from many missing data, which leads to large amount of data loss because of the requirement of balanced panel. The data loss may make the baseline results biased; thus, we drop this variable and reconduct the analysis in this section.

The results of excluding I are reported in Table 6. We find that, both when Gini_net and Gini_market as the dependent variable, FDI, the government expenditure (G), the population aging (OLD), the labor share (LAB), the unemployment levels (UNE), the real exchange rate (XR) and the female labor force participation (FEM) are robust determinants of income inequality in the EMEs, which is consistent with the baseline results in Table 2. Our results remain stable after dropping the variables with many missing observations.

18 😉 S. WANG ET AL.

4.3.4. Reducing the business cycle

Our baseline results may suffer from the impacts of the business cycle. Following the strategy of Moral-Benito (2012), to keep the data balance and avoid the serial correlation in the transitory component of the disturbance term, we split the sample into five-year periods. Besides, the stock variables are measured in the first year of each five years and the flow variables are measured as five-year averages. The results are reported in Table 7.

Compared with the baseline results, the BMA strategy reduces the number of common variables selected under different dependent variables to four, suggesting that the use of 5-year averages may indeed reduce some of the information contained in the original data. The coefficients of the population aging (OLD) variable are opposite to the baseline results, suggesting that increasing population aging increases income inequality. The impacts of unemployment (UNE) and real exchange rate (XR) on income inequality are consistent with the benchmark results. Trade openness (OPNESS) reduces the income inequality on average, which is consistent with the Stolper-Samuelson theorem and Jaumotte et al. (2013). The low-skilled labors in the developing countries benefit from the trade liberalization, while competition leads to a reduction in the compensation of the high-skilled workers.

Overall, the population aging (OLD), the real exchange rate (XR), the unemployment level (UNE) and the government expenditure (G) remain as the robust determinant of income inequality in the EMEs as the baseline results of Table 2.

4.3.5. Eba

EBA was a popular procedure used to probe the determinants of the variable of interest. In this section, we employ the EBA strategy to check the baseline results although the EBA strategy has been criticized for many reasons as mentioned in Section 3.3. In practice, there are two representative methods of EBA, i.e., E. E. Leamer (1985) and Sala-I-Martin (1997). E. E. Leamer (1985) considered a variable to be robust only if the upper and lower bounds of its coefficients both lie on the same side of zero. However, this standard is stringent and generally lead to results where all variables that we concern are not robust. In response to the shortcomings of E. E. Leamer (1985), Sala-I-Martin (1997) proposes an alternative approach that focuses on the entire distribution of estimates and assigns a certain level of confidence to the robustness of each variable. Specifically, a variable is robust if the major parts of its coefficient distribution lie simultaneously to

				-			
	P.IP	P.M	P.SD		P.IP	P.M	P.SD
Regressor		Gini_net		Regressor		Gini_market	
OLD†	1.0000	.9142	.2166	UNE†	1.0000	.6089	.1028
OPNESS	1.0000	1332	.0214	OPNESS	.9620	0689	.0236
XR†	.9933	0010	.0003	XR†	.8997	0009	.0004
UNE†	.9883	.3473	.1043	OLD†	.8103	.5447	.3257
CRE	.9243	.0611	.0261	FDI †	.7413	.4970	.3581
GDP ² †	.9057	-7.7768	4.0409	FEM †	.7023	.1546	.1180
GDP †	.8517	51.8278	22.8376				
URB	.7333	.0651	.0495				
Gt	.6013	1098	.1096				

Table 7. Determinants of income inequality in the EMEs using the BMA: 5-year averages.

[†]indicates the variable is also robust in baseline results in Table 2.

Variable	Mean β	Mean SE	Upper bound	Lower bound	CDF(0)
OLD [†]	-1.364	0.079	113	-0.679	100.00
G [†]	-0.369	0.063	-1.293	0.344	100.00
HC	-7.556	0.642	-15.83	0.264	100.00
LAB [†]	-17.030	3.259	-54.253	7.847	100.00
OPNESS	16.639	2.707	-15.268	34.542	100.00
GDP^\dagger	-1.717	1.178	-17.666	14.753	100.00
UNE [†]	0.468	0.043	.183	0.877	100.00
FEM [†]	0.150	0.035	362	0.496	99.999
1	-0.795	0.193	-2.259	0.642	99.998
FDI [†]	0.145	0.095	594	0.580	92.921
CRE	0.011	0.009	073	0.086	88.891
GDP ²	0.030	0.03	596	0.669	86.500
XR	0.000	0.000	.000	0.000	82.976
TFP	1.086	1.284	-13.707	13.523	79.953
INF	0.000	0.001	004	0.002	68.726
URB	-0.003	0.016	272	0.157	57.314

Table 8. Determinants of income inequality in the EMEs using EBA.

[†]indicates the variable is also robust in baseline results in Table 2. "Mean β" and "Mean SE" indicate the mean over all regressions of the coefficient and the standard error, respectively. "CDF(0)" gives the proportion of all of the cumulative distribution functions to one side of zero, and variables are sorted according to this criterion.

the left or right of zero. Following Sala-I-Martin (1997), the cut-off point is a CDF(0) value of 90 or higher. The EBA results are shown in Table 8.

The robust variables selected by the EBA procedure are fairly identical to the baseline results of Table 2. Besides, the EBA additionally reveals the significant effects of human capital condition (HC) and the trade openness (OPNESS) on income inequality in the EMEs.

4.3.6. Ssvs

Table A1.5 indicates that the correlations between competing predictors are quite high in some cases. Stochastic Search Variable Selection (SSVS, George & McCulloch, 1993, 1997) is suggested to be less sensitive to multicollinearity issues than competing Bayesian alternatives based on Bayes factors. Therefore, we use SSVS as a robustness check to the baseline results. SSVS is another commonly used method for Bayesian variable selection. It identifies the set of models that may have better predictive performance by estimating the level of model uncertainty among all possible candidate models. Jetter et al. (2022) summarize that SSVS is a methodological improvement over EBA in terms of dealing with multicollinearity, relative fit of different models and the uncertainty in the extreme-bound estimation (Brock & Durlauf, 2001; McAleer & Veall, 1989).

Like the BMA, SSVS uses marginal inclusion probabilities (MIP) to select variables, and we usually only care about the independent variables with MIP of 0.50 or above. SSVS results in Figure 4 shows that almost all the variables selected in BMA, including the unemployment level (UNE), the population aging (OLD), the female workforce as a ratio of total workforce (FEM), are proved to be important in the SSVS approach. In addition, like the results in Tables 7 and 8, SSVS approach reveals that the trade openness (OPNESS) has a significant effect on income inequality.

4.3.7. Lasso: evidence from machine learning

Machine learning approaches are often used to select a significant predictor variable. It has been proven that machine learning performs significantly better than traditional

20 👄 S. WANG ET AL.



Figure 4. Determinants of income inequality in the EMEs using the SSVS. *Note*: "MIP" denotes marginal inclusion probabilities of each predictor variables. Regressor with a MIP of 0.50 to 0.75 is considered weak, of 0.75 to 0.95 positive, of 0.95 to 0.99 strong and 0.99 very strong.

econometric models in many aspects, especially in prediction (Varian, 2014). In this case, we additionally choose Least Absolute Shrinkage and Selection Operator models (Lasso, Tibshirani, 2011), which is often used as a comparison to SSVS and BMA in the previous literature (Bainter et al., 2020), as one of the robustness tests.

The Lasso model mainly focuses on the β_j which minimizes the following objective function.

$$\sum_{i=1}^{N} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right|$$
(9)

In Equation (9), the left part is the RSS (sum of squared residuals) and the right part is the penalty term that is adjusted by the hyperparameter λ . The hyperparameter is exogenously generated by the researcher via a cross-validation process. When a new variable is introduced into the Lasso but has a negligible contribution to the reduction of RSS, the effect of the shrinkage penalty grows, which means that the Lasso considers the coefficient of this variable to be zero.

As shown in Figure 5, the optimal parameter after 10-fold cross-validation is 0.2625, and the corresponding model size is 9. In this case, the variables chosen by the Lasso are FDI, the female labor force participation (FEM), the physical capital investment (I), the population aging (OLD), the human capital (HC), the total factor productivity (TFP), the labor share (LAB), the trade openness (OPNESS) and the real output per capita (GDP), which is basically identical to



Figure 5. Determinants of income inequality in the EMEs using the Lasso.

Note: Graph A reports how the MSE of the Lasso model varies with $\log(\lambda)$. The optimal Lambda we choose here corresponds to the model with the best performance and the cleanest (minimum number of variables). $\lambda_{opt} \in [\lambda_{minMSE} - se, \lambda_{minMSE} + se]$. Model sizes (number of predictor variables in the model) for different λ are also reported, and the size of the model corresponding to the optimal lambda is 9. The graph B reports how coefficients of candidate set of variables varies with $\log(\lambda)$.

the baseline results of Table 2. Consistent with the results of other robustness tests, Lasso also places additional emphasis on three variables, i.e., the physical capital investment (I), the human capital condition (HC) and the trade openness (OPNESS). In addition, Lasso also chooses the total factor productivity (TFP) in the optimal model, but this finding does not receive any stronger evidence in the baseline results and other robustness tests.

4.3.8. Random forest: evidence from machine learning

Random forest (Breiman, 2001) is another popular machine learning strategy used to select the predictor, which is a supervised learning and ensemble algorithm and constructed based on multiple decision trees. Random forest has attracted many attention in economics due to its versatility and superior ability in reducing overfitting and multi-collinearity issues (Lindner et al., 2022).

In terms of variable selection, random forests can filter predictor variables in two aspects. First, MSE (mean square error) of model. In a random forest, each predictor is randomly assigned a value, and if the predictor is more important, the error in the model prediction increases when its value is randomly replaced. "% IncMSE", as shown in Figure 6, denotes the increase in MSE if one variable is randomly replaced. Thus, a larger "%IncMSE" indicates a higher importance of the variable. Second, the node purity (Gini index), which represents the heterogeneity effect of each variable on the observations at the nodes (Genuer et al., 2010). The higher the node purity (i.e., the smaller the Gini index), the more important the variable. The results of both indicators above are reported in Figure 6.

After 10-fold cross-validations, the model with 10 or 12 variables, whose error is 2.0013 and 1.8307, respectively, was preferable. The unemployment level (UNE), the human capital (HC), the real exchange rate (XR), the population aging (OLD) and other variables remain robust. However, the predictive capacity of the trade openness (OPNESS) for income inequality, which performs favorably in other robustness tests,



Figure 6. Determinants of income inequality in the EMEs using the random forest. *Note*: "%IncMSE" is short for "increase in mean squared error", "IncNodePurity" is short for "increase in node purity". The points marked in blue represent the variables selected by the optimal random forest model.

is questioned in the random forest model, suggesting that the effect of the trade openness on income inequality is still subject to model uncertainty, and the mechanism of the effect of the trade openness on EMEs remains to be further discussed and tested.

5. A comparative study with the MEs

We extend the BMA strategy to exploit the determinants of income inequality by employing the data from 27 widely accepted MEs, which is collected similarly to Section 2.3. A list of the MEs is presented in Table A1.3.

	P.IP	P.M	P.SD		P.IP	P.M	P.SD
Regressor		Full Sample		Regressor	Su	btract crises per	iod
CRE	1.0000	0.0095	0.0021	CRE	1.0000	0.0182	0.0030
Gt	1.0000	-0.4393	0.0277	G†	1.0000	-0.4917	0.0372
INF	1.0000	0.1146	0.0260	OLD†	1.0000	-0.1971	0.0414
OLD†	1.0000	-0.4820	0.0371	HC	1.0000	-2.5781	0.3520
НС	1.0000	-3.4524	0.2775	TFP	1.0000	-7.4564	0.7418
TFP	1.0000	-9.8615	0.5691	LAB†	1.0000	-19.4051	1.5976
LAB†	1.0000	-10.5461	1.4103	UNE†	1.0000	0.1216	0.0257
XR†	1.0000	-0.0103	0.0017	FEM†	0.9933	0.1611	0.0420
OPNESS	1.0000	-10.4441	2.4481	GDP†	0.9337	1.4649	0.6155
GDP ² †	1.0000	0.0780	0.0152	INF	0.8340	0.0852	0.0494
UNE†	1.0000	0.1116	0.0213	XR†	0.7877	-0.0046	0.0030
1	0.7890	0.1582	0.1073	URB	0.7273	0.0177	0.0133
URB	0.7680	0.0141	0.0102				

Table 9. Determinants of income inequality in the MEs using the BMA.

The dependent variable here is Gini_net. The similar results can be reached and thus are not reported when the dependent variable is Gini_market. † indicates the variable is also robust in baseline results in Table 2.

Table 9 reports the results based on the MEs sample. It is evident that the determinants of income inequality in the MEs significantly differentiate from those in the EMEs but share some variables. In the full sample, the BMA strategy selects 13 robust determinants of income inequality in the MEs. In the sample subtracted crises period, 12 variables are selected as the robust factors of income inequality. Compared with Tables 2 and 3, we find the government expenditure (G), the population aging (OLD), the labor share (LAB), the real exchange rate (XR) and the unemployment level (UNE) are also robust determinants of inequality in the MEs.

The results based on the full sample indicate that share of labor compensation, human capital condition (Sequeira et al., 2017), physical capital and inflation are the determinants of income inequality in the MEs. Share of labor compensation in GDP could remit the income inequality on average, implying that the MEs can reduce the income inequality by improving the factor structure. Similarly, human capital development is beneficial to the reduction of income inequality in the MEs. Accessing to education and professional training is easier than ever (Baldi, 2013) in the MEs and thus reduces the income inequality. Physical capital alleviates the income inequality on average in the MEs. Higher inflation is associated with the worsening income distribution (Walsh & Yu, 2012). Government expenditure and population aging are the robust determinants of income inequality in the MEs even when we exclude the data during the crises period. Government expenditure can reduce the income inequality on average, which is consistent with Agnello and Tavares (2014) and Roine et al. (2009). Population aging has the impact on the income inequality in the MEs.

Overall, different with the results of the EMEs in Table 2, the bank credit (CRE), the inflation (INF), the human capital (HC), the total factor productivity (TFP) and the urbanization (URB) are important factors affecting inequality in the MEs. Female labor force participation is no longer a robust determinant of income inequality in the MEs.

6. Discussions

Using the BMA procedure, we get many interesting results on the determination of income inequality in the EMEs and a comparative analysis with the MEs. These results may provide some new research directions and may push the theorists to revise the existing theories.

First, we find that in the EMEs, the notable inverted-U shaped Kuznets curve doesn't exist, alternatively we find the evidence that support the U-shaped curve. This result indicates that many EMEs have surmounted the threshold value of economic development, the further development will worsen the income distribution. The classical Kuznets theory said that after the threshold, as the country performs better in economy and the governance ability improves, the country could flatten the income inequality in many delicate ways. Why the opposite nexus is found in the EMEs? Although the prior studies have found the similar results, no theoretical explanation have been suggested.

Second, we find the population aging is a strictly robust negative determinant of income inequality in both the EMEs and the MEs. Prior literature has probed the effects of population aging on income inequality mainly using empirical methods in the EMEs, such as Zhong (2011), Dong et al. (2018). Intuitively, the increase of population aging

24 😉 S. WANG ET AL.

indicates the productivity heterogeneity between economic agents decreases, which means that the wage disparity will decrease and income distribution tends to be equalized. However, except the productivity effect, population aging many produce many economic effects. How about these channels affect the income inequality? We need more delicate empirical works and build theoretical model to assist the analysis.

Third, we find the female labor force participation is a robust determinant of income inequality in the EMEs. This is an interesting finding, although prior studies have investigated the nexus between female labor force participation and income inequality (Değirmenci, 2018). As female labor force participation increases, the gender wage gap decreases, suggesting that income tends to be equal. Given the importance of female labor force participation on income inequality, we need more delicate studies, such as checking the nexus using the firm-level data.

Fourth, we find financial development is not a robust determinant of income inequality in the EMEs, but a key determinant in the MEs. The idea that financial development reduces the income inequality by easing financial market constraints remains being suspected in the EMEs. Lack of diverse financial systems may pose an obstacle to financial access to the poor (Claessens & Perotti, 2007). Thus, we need a detailed comparative analysis about the different roles of financial development in income inequality between the EMEs and the MEs. In practice, what kind of financial system could improve income distribution?

7. Conclusion

We use the BMA strategy to overcome the model uncertainty and identify the robust determinants of income inequality in the EMEs by employing the panel data over a period from 1990 to 2019. First, differently, we find a U-shaped Kuznets curve in emerging market economies rather than the widely accepted inverted U-shaped curve. Besides, we find four strictly robust determinants of income inequality in emerging market economies, including the population aging, the female labor force participation, the unemployment level and the share of labor compensation in output. These results remain stable through a series of robustness checks, including excluding the recent crisis period, alternating the model and coefficient priors, dropping the variable with many missing data, reducing the business cycle, using the extreme bound analysis and the Stochastic Search Variable Selection, adapting two machine learning strategies (Lasso and Random forest). In most cases, the government expenditure and real exchange rate could also robustly drive the dynamics of income inequality in emerging market economies.

Further, we build a comparison analysis on the drivers of income inequality between the EMEs and the MEs and find that financial development, inflation, human capital condition, total factor productivity and urbanization are distinct determinants of income inequality in the market economies.

Finally, we reflect our main findings and argue that both in theory and in empirics, we need more delicate studies. For example, the Kuznets effect of economic growth on income inequality, the roles of population aging and female labor force participation and a design of financial development beneficial to the income equality.

Acknowledgments

This work is supported by the Natural Science Foundation of China (72003205), the Humanities and Social Sciences Foundation of Chinese Ministry of Education (20YJC790142) and the General Project of Social Science Planning in Guangdong Province (GD22CYJ12).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Natural Science Foundation of China [72003205]; General Project of Social Science Planning in Guangdong Province [GD22CYJ12]; Humanities and Social Sciences Research of the Ministry of Education of China [20YJC790142].

Notes on contributors

Shengquan Wang is a distinguished research associate fellow at the Institute of Advanced Studies in Humanities and Social Sciences, Beijing Normal University. His research interests are income inequality and macro-finance.

Yuan Gao is an undergraduate at Bay Area International Business School, Beijing Normal University. His research interests are income inequality and marketing.

Mingjin Luo is an assistant research fellow at the International Development Cooperation Academy, Shanghai University of International Business and Economics. Her research interests are corporate finance and international finance.

ORCID

Shengquan Wang (D) http://orcid.org/0000-0003-4138-7371

References

- Acemoglu, D. (1999). Changes in unemployment and wage inequality: An alternative theory and some evidence. *American Economic Review*, 89(5), 1259–1278. https://doi.org/10.1257/aer.89.5. 1259
- Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2015). Democracy, redistribution and inequality. In *Handbook of income distribution* (pp. 1885–1966). Elsevier. https://doi.org/10. 1016/B978-0-444-59429-7.00022-4
- Agnello, L., & Tavares, R. M. S. (2014). How does fiscal consolidation impact on income inequality? *Review of Income and Wealth*, 60(4), 702–726.
- Alesina, A., & Rodrik, D. (1994). Distributive politics and economic growth. *The Quarterly Journal* of *Economics*, 109(2), 465–490. https://doi.org/10.2307/2118470
- Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica, 77(4), 1229-1279.
- Bainter, S. A., McCauley, T. G., & Wager, T. (2020). Improving practices for selecting a subset of important predictors in psychology: An application to predicting pain. Advances in Methods and Practices in Psychological Science, 3(1), 66–80. https://doi.org/10.1177/2515245919885617
- Baldi, G. (2013). Physical and human capital accumulation and the evolution of income and inequality. *Journal of Economic Development*, 38(3), 57–83. https://doi.org/10.35866/caujed.2013.38.3.003

26 🛭 😔 S. WANG ET AL.

- Bandyopadhyay, D., & Basu, P. (2005). What drives the cross-country growth and inequality correlation? *Canadian Journal of Economics*, 38(4), 1272–1297. https://doi.org/10.1111/j.0008-4085.2005.00325.x
- Banerjee, A. V., & Newman, A. F. (1993). Occupational choice and the process of development. Journal of political Economy, 101, 274–298. https://doi.org/10.1086/261876
- Barro, R. J. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth*, 5 (1), 5–32. https://doi.org/10.1023/A:1009850119329
- Beck, T., Levine, R., & Levkov, A. (2010). Big bad banks? The winners and losers from bank deregulation in the United States. *The Journal of Finance*, 65(5), 1637–1667. https://doi.org/10. 1111/j.1540-6261.2010.01589.x
- Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association*, 82(397), 112–122. https://doi.org/ 10.1080/01621459.1987.10478397
- Bogliaccini, J. A., & Egan, P. J. W. (2017). Foreign direct investment and inequality in developing countries: Does sector matter? *Economics & Politics*, 29(3), 209–236. https://doi.org/10.1111/ecpo.12098
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/ A:1010933404324
- Brock, W. A., & Durlauf, S. N. (2001). What have we learned from a decade of empirical research on growth? Growth empirics and reality. *World Bank Economic Review*, 15(2), 229–272. https://doi.org/ 10.1093/wber/15.2.229
- Cancian, M., & Schoeni, R. F. (1998). Wives' earnings and the level and distribution of married couples' earnings in developed countries. *Journal of Income Distribution*, 8(1), 45–61. https://doi.org/10.1016/S0926-6437(99)80003-1
- Cevik, S., & Correacaro, C. (2015). Growing (un) equal: Fiscal policy and income inequality in China and BRIC. IMF Working Paper.
- Chakrabarti, A. (2001). The determinants of foreign direct investments: Sensitivity analyses of cross-country regressions. *Kyklos*, *54*(1), 89–114. https://doi.org/10.1111/1467-6435.00142
- Claessens, S., & Perotti, E. (2007). Finance and inequality: Channels and evidence. Journal of Comparative Economics, 35(4), 748-777. https://doi.org/10.1016/j.jce.2007.07.002
- Coibion, O., Gorodnichenko, Y., Kueng, L., & Silvia, J. (2017). Innocent Bystanders? Monetary policy and inequality. *Journal of Monetary Economics*, 88, 70–89. https://doi.org/10.1016/j. jmoneco.2017.05.005
- Costantini, M., & Paradiso, A. (2018). What do panel data say on inequality and GDP? New evidence at US state-level. *Economics Letters*, *168*, 115–117. https://doi.org/10.1016/j.econlet.2018.04.019
- Cragg, M. I., & Epelbaum, M. (1996). Why has wage dispersion grown in Mexico? Is it the incidence of reforms or the growing demand for skills? *Journal of Development Economics*, 51 (1), 99–116. https://doi.org/10.1016/S0304-3878(96)00427-0
- Dabla-Norris, M. E., Kochhar, M. K., Suphaphiphat, M. N., Ricka, M. F., & Tsounta, E. (2015). Causes and consequences of income inequality: A global perspective. IMF Working Papers.
- Değirmenci, S. (2018). Female labor force participation and income inequality: The case of Turkey. *Institutions, National Identity, Power, and Governance in the 21st Century*, 2017, June 28–30, Belgrade, Serbia.
- Dietzenbacher, E. (1989). The dynamics of population growth, differential fertility, and inequality: Comment. *American Economic Review*, *79*(3), 584–587.
- Doepke, M., & Schneider, M. (2006). Inflation and the redistribution of nominal wealth. *Journal of Political Economy*, 114(6), 1069–1097. https://doi.org/10.1086/508379
- Doerrenberg, P., & Peichl, A. (2014). The impact of redistributive policies on inequality in OECD countries. *Applied Economics*, 46(17), 2066–2086. https://doi.org/10.1080/00036846.2014.892202
- Dong, Z., Tang, C., & Wei, X. (2018). Does population aging intensify income inequality? Evidence from China. *Journal of the Asia Pacific Economy*, 23(1), 1–12. https://doi.org/10.1080/13547860. 2017.1354270
- Dorn, F., Fuest, C., & Potrafke, N. (2018). Globalization and income inequality revisited. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3143398

- Esping-Andersen, G. (2007). Sociological explanations of changing income distributions. *American Behavioral Scientist*, 50(5), 639–658. https://doi.org/10.1177/0002764206295011
- Espoir, D. K., & Ngepah, N. (2021). Income distribution and total factor productivity: A cross-country panel cointegration analysis. *International Economics & Economic Policy*, 18(4), 661–698. https://doi.org/10.1007/s10368-021-00494-6
- Feenstra, R. C., & Hanson, G. H. (1997). Foreign direct investment and relative wages: Evidence from Mexico's maquiladoras. *Journal of International Economics*, 42(3–4), 371–393. https://doi. org/10.1016/S0022-1996(96)01475-4
- Feenstra, R. C., & Hanson, G. H. (1999). The impact of outsourcing and high-technology capital on wages: Estimates for the United States, 1979–1990. *Quarterly Journal of Economics*, *114*(3), 907–940. https://doi.org/10.1162/003355399556179
- Furceri, D., & Ostry, J. D. (2019). Robust determinants of income inequality. Oxford Review of Economic Policy, 35(3), 490–517.
- Galor, O. (2000). Income distribution and the process of development. *European Economic Review*, 44(4–6), 706–712. https://doi.org/10.1016/S0014-2921(99)00039-2
- Galor, O., & Moav, O. (2004). From physical to human capital accumulation: Inequality and the process of development. *Review of Economic Studies*, 71(4), 1001–1026. https://doi.org/10.1111/0034-6527.00312
- Galor, O., & Tsiddon, D. (1997). Technical progress, mobility, and economic growth. *American Economic Review*, 87(3), 363–382.
- Galor, O., & Zeira, J. (1993). Income distribution and macroeconomics. *The Review of economic studies*, 60, 35–52. https://doi.org/10.2307/2297811
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014
- George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881–889. https://doi.org/10.1080/01621459.1993.10476353
- George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7, 339–373. https://doi.org/10.1007/s004400050098
- Goda, T., Onaran, Ö., & Stockhammer, E. (2017). Income inequality and wealth concentration in the recent crisis. *Development and Change*, 48(1), 3–27. https://doi.org/10.1111/dech.12280
- Greenwood, J., & Jovanovic, B. (1990). Financial Development, Growth, and the Distribution of Income. *Journal of political Economy*, 98(5), 1076–1107. https://doi.org/10.1086/261720
- Gruber, J. (2013). Public finance and public policy (4th ed.). Worth Publishers.
- Hansen, M. H., & Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, *96*(454), 746–774. https://doi.org/10.1198/ 016214501753168398
- Heathcote, J., Perri, F., & Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics*, 13(1), 15–51. https://doi.org/10.1016/j.red.2009.10.010
- Hémous, D., & Olsen, M. (2022). The rise of the machines: Automation, horizontal innovation, and income inequality. American Economic Journal: Macroeconomics, 14(1), 179–223. https:// doi.org/10.1257/mac.20160164
- Hidalgo-Hidalgo, M., & Iturbe-Ormaetxe, I. (2018). Long-run effects of public expenditure on poverty. Journal of Economic Inequality, 16(1), 1–22. https://doi.org/10.1007/s10888-017-9360-z
- Inagaki, K. (2010). Income inequality and the suicide rate in Japan: Evidence from cointegration and LA-VAR. *Journal of Applied Economics*, 13(1), 113–133. https://doi.org/10.1016/S1514-0326(10)60006-2
- Jaumotte, F., Lall, S., & Papageorgiou, C. (2013). Rising income inequality: Technology, or trade and financial globalization? *IMF Economic Review*, 61(2), 271–309. https://doi.org/10.1057/imfer.2013.7
- Jeanneney, S. G., & Hua, P. (2001). How does real exchange rate influence income inequality between urban and rural areas in China? *Journal of Development Economics*, 64(2), 529–545. https://doi.org/10.1016/S0304-3878(00)00149-8

28 🕳 S. WANG ET AL.

- Jetter, M., Mahmood, R., Parmeter, C. F., & Ramírez-Hassan, A. (2022). Post-Cold War civil conflict and the role of history and religion: A stochastic search variable selection approach. *Economic Modelling*, 114, 105907. https://doi.org/10.1016/j.econmod.2022.105907
- Knight, J. B., & Sabot, R. H. (1983). Educational expansion and the Kuznets effect. American Economic Review, 73(5), 1132–1136.
- Kumhof, M., Rancière, R., & Winant, P. (2015). Inequality, leverage, and crises. American Economic Review, 105(3), 1217–1245. https://doi.org/10.1257/aer.20110683
- Kuznets, S. (1955). Economic growth and income inequality. American Economic Review, 45(1), 1–28.
- Leamer, E. E. (1985). Sensitivity analyses would help. American Economic Review, 75(3), 308-313.
- Leamer, E., & Leonard, H. (1983). Reporting the fragility of regression estimates. Review of Economics & Statistics, 65(2), 306-317. https://doi.org/10.2307/1924497
- Leòn-Gonzàlez, R., & Montolio, D. (2015). Endogeneity and panel data in growth regression: A Bayesian model averaging approach. *Journal of Macroeconomics*, 46, 23–39. https://doi.org/ 10.1016/j.jmacro.2015.07.003
- Ley, E., & Steel, M. F. (2012). Mixtures of g-priors for Bayesian model averaging with economic applications. *Journal of Econometrics*, 171(2), 251–266. https://doi.org/10.1016/j.jeconom. 2012.06.009
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*(481), 410–423. https://doi.org/10.1198/016214507000001337
- Lindner, T., Puck, J., & Verbeke, A. (2022). Beyond addressing multicollinearity: Robust quantitative analysis and machine learning in international business research. *Journal of International Business Studies*, 53(7), 1307–1314. https://doi.org/10.1057/s41267-022-00549-z
- Lin, S. C., Huang, H. C., Kim, D. H., & Yeh, C. C. (2009). Nonlinearity between inequality and growth. *Studies in Nonlinear Dynamics & Econometrics*, 13(2). https://doi.org/10.2202/1558-3708.1635
- Lin, Y. C., Huang, H. C., & Yeh, C. C. (2014). Inequality-growth nexus along the development process. *Studies in Nonlinear Dynamics & Econometrics*, 18(3), 237–252. https://doi.org/10. 1515/snde-2012-0037
- List, J. A., & Gallet, C. A. (1999). The Kuznets curve: What happens after the inverted-U? *Review of Development Economics*, 3(2), 200–206. https://doi.org/10.1111/1467-9361.00061
- McAleer, M., & Veall, M. R. (1989). How fragile are fragile inferences? A re-evaluation of the deterrent effect of capital punishment. *The Review of Economics and Statistics*, 71(1), 99–106. https://doi.org/10.2307/1928056
- Menna, L., & Tirelli, P. (2017). Optimal inflation to reduce inequality. *Review of Economic Dynamics*, 24, 79-94. https://doi.org/10.1016/j.red.2017.01.004
- Min, H. G., Shin, S. O., & Mcdonald, J. A. (2015). Income inequality and the real exchange rate: Linkages and evidence. *Annals of Economics and Finance*, *16*(1), 115–141.
- Miyazawa, K. (2006). Growth and inequality: A demographic explanation. *Journal of Population Economics*, 19(3), 559–578. https://doi.org/10.1007/s00148-005-0047-6
- Monnin, P. (2014). Inflation and income inequality in developed economies. CEP Working Paper Series.
- Moral-Benito, E. (2012). Determinants of economic growth: A Bayesian panel data approach. *Review of Economics and Statistics*, 94(2), 566–579. https://doi.org/10.1162/REST_a_00154
- Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29(1), 46–75. https://doi.org/10.1111/joes.12044
- OECD. (2012). *Economic policy reforms 2012 going for growth* (pp. 181–202). OECD Publishing. https://doi.org/10.1787/growth-2012-en
- Oyvat, C. (2016). Agrarian structures, urbanization, and inequality. *World Development*, 83, 207–230. https://doi.org/10.1016/j.worlddev.2016.01.019
- Piketty, T. (2014). Capital in the twenty-first century. Harvard University Press.
- Piketty, T., Saez, E., & Stantcheva, S. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1), 230–271. https://doi.org/10.1257/pol. 6.1.230

- Raaum, O., Bratsberg, B., Røed, K., Österbacka, E., Eriksson, T., Jäntti, M., & Naylor, R. A. (2008). Marital sorting, household labor supply, and intergenerational earnings mobility across countries. *The BE Journal of Economic Analysis & Policy*, 7(2). https://doi.org/10.2202/1935-1682.1767
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. https://doi.org/10.2307/271063
- Rajan, R. (2010). Fault lines: How hidden fractures still threaten the world economy. Princeton University Press. https://doi.org/10.1515/9781400839803
- Roine, J., Vlachos, J., & Daniel, W. (2009). The long-run determinants of inequality: What can we learn from top income data? *Journal of Public Economics*, 93(7–8), 974–988. https://doi.org/10. 1016/j.jpubeco.2009.04.003
- Sala-I-Martin, X. (1997). I just ran four million regressions. NBER working paper.
- Sampson, T. (2014). Selection into trade and wage inequality. American Economic Journal: Microeconomics, 6(3), 157-202. https://doi.org/10.1257/mic.6.3.157
- Santelli, J. S., Song, X., Garbers, S., Sharma, V., & Viner, R. M. (2017). Global trends in adolescent fertility, 1990–2012, in relation to national wealth, income inequalities, and educational expenditures. *Journal of Adolescent Health*, 60(2), 161–168. https://doi.org/10.1016/j.jadohealth.2016.08.026
- Sequeira, T. N., Santos, M., & Ferreira-Lopes, A. (2017). Income inequality, TFP, and human capital. Economic Record. *Economic Record*, 93(300), 89–111. https://doi.org/10.1111/1475-4932.12316
- Solt, F. (2015). On the assessment and use of cross-national income inequality datasets. *Journal of Economic Inequality*, 13(4), 683–691. https://doi.org/10.1007/s10888-015-9308-0
- Steel, M. F. J. (2020). Model averaging and its use in economics. *Journal of Economic Literature*, 58 (3), 644–719. https://doi.org/10.1257/jel.20191385
- Stiglitz, J. (2013). The price of inequality. W. W. Norton & Company.
- Stolper, W. F., & Samuelson, P. A. (1941). Protection and real wages. The Review of Economic Studies, 9(1), 58–73. https://doi.org/10.2307/2967638
- Sudo, N. (2017). The effects of women's labor force participation: An explanation of changes in household income inequality. Social Forces, 95(4), 1427–1450. https://doi.org/10.1093/sf/sox011
- Székely, M. (2003). The 1990s in Latin America: Another decade of persistent inequality, but with somewhat lower poverty. *Journal of Applied Economics*, 6(2), 317–339. https://doi.org/10.1080/ 15140326.2003.12040597
- Thorbecke, E., & Charumilind, C. (2002). Economic inequality and its socioeconomic impact. *World Development*, 30(9), 1477–1495. https://doi.org/10.1016/S0305-750X(02)00052-9
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273–282. https://doi.org/ 10.1111/j.1467-9868.2011.00771.x
- Tsount, E., & Osueke, A. (2014). What is behind Latin America's declining income inequality? IMF Working Paper.
- Varian, H. R. (2014). Big data New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. https://doi.org/10.1257/jep.28.2.3
- Walsh, J. P., & Yu, J. (2012). Inflation and income inequality: Is food inflation different? IMF Working Paper.
- Wang, S., & Luo, R. (2023). Income distribution, financial liberalisations and banking stability: Theory and international evidence. *International Journal of Finance & Economics*. https://doi. org/10.1002/ijfe.2809
- Yang, J., & Qiu, M. (2016). The impact of education on income inequality and intergenerational mobility. *China Economic Review*, 37, 110–125. https://doi.org/10.1016/j.chieco.2015.12.009
- Yeaple, S. R. (2005). A simple model of firm heterogeneity, international trade, and wages. *Journal of International Economics*, 65(1), 1–20. https://doi.org/10.1016/j.jinteco.2004.01.001
- Zhong, H. (2011). The impact of population aging on income inequality in developing countries: Evidence from rural China. *China Economic Review*, 22(1), 98–107. https://doi.org/10.1016/j. chieco.2010.09.003

Appendix 1 List of variables, countries, descriptive statistics

Definition Variable Calculation Source **Dependent Variables** SWIID Gini_net Inequality in net income Based on post-tax and post-transfer income 9.3 SWIID Gini_market Inequality in market income Based on pre-tax and pre-transfer income 9.3 Independent Variables GDP Logarithmic Real GDP per capita Deflated by 2011 US dollars at PPP PWT 10.0 GDP² Quadratic GDP Calculation by author PWT 10.0 XR Real exchange rate National currency/USD PWT 10.0 INF Inflation Growth ratio of CPI PWT 10.0 G Government expenditure Government expenditure scaled by GDP PWT 10.0 I Physical capital investment Investment scaled by GDP PWT 10.0 **OPNESS** Export scaled by GDP PWT Trade openness 10.0 FDI Foreign direct investment FDI scaled by GDP PWT 10.0 CRE Bank credit to nonfinancial Credit scaled by GDP WB sectors HC Based on years of schooling and returns to education PWT Human capital condition 10.0 TFP Total factor productivity TFP level at current PPPs (USA = 1) PWT 10.0 LAB Labor income share Share of labour compensation in GDP at current national PWT prices 10.0 UNE PWT Unemployment Unemployment as a ratio of workforce 10.0 OLD Population aging Population aged above 65 as a ratio of total population WB URB Urbanization Urban population as a ratio of total population WB FEM Female labor force participation Female workforce as a ratio of total workforce WB

Table A1.1 List of variables.

Table A1.2 List of EMEs.

Argentina [‡]	Czech Republic [†]	Mexico ^{†‡}	Thailand ^{†‡}
Bolivia [‡]	Ecuador‡	Nigeria [‡]	Turkey ^{†‡}
Brazil ^{†‡}	Egypt†‡	Peru ^{†‡}	Ukraine [‡]
Chile ^{†‡}	Indonesia†‡	Philippines ^{†‡}	Uruguay [‡]
China ^{†‡}	India†‡	Poland ^{†‡}	Venezuela [‡]
Côte d'Ivoire [‡]	Israel	Korea ^{†‡}	South Africa ^{†‡}
Colombia ^{†‡}	Jordan‡	Russia ^{†‡}	

[†]indicates the countries covered by MACI-EMI. [‡] indicates the countries covered by EMBI.

Table A1.3 List of MEs.

Australia	Spain	Iceland	New Zealand
Austria	Finland	Italy	Portugal
Belgium	France	Japan	Singapore
Canada	United Kingdom	Luxembourg	Slovenia
Switzerland	Greece	Latvia	Sweden
Germany	Hungary	Netherlands	United States
Denmark	Ireland	Norway	

ive st	atistics of v	/ariables.												
Mean		S.D	Min	Max		Mean	S.D	Min	Мах		Mean	S.D	Min	Мах
41.84346		8.4482	20.2000	63.1000	J	13.9544	4.6390	0.9112	27.3989	TFP	0.5899	0.2112	0.0544	4.3463
		8.3077	24.4133	61.6520			4.3424	4.2905	24.2915			0.1974	0.3268	1.1100
		2.2190	31.7418	48.0718			1.8345	3.7365	22.3005			0.0841	0.1854	0.8296
46.4691		8.4482	32.3000	72.7000	_	1.7615	1.2744	-2.1067	8.3756	LAB	0.4979	0.0846	0.3101	0:0030
		8.3077	33.7433	70.6287			1.0235	0.0496	3.8178			0.0721	0.3756	0.6576
		2.2190	37.3491	51.1846			0.7846	-1.7000	7.6033			0.0465	0.3257	0.8282
12.8971		1.4569	8.8772	16.8140	OPNESS	0.5488	0.0926	0.1073	0.7585	UNE	7.7000	5.6828	0.2500	33.29
		10.4742	15.9331	15.9331			0.0405	0.4577	0.6074			5.3301	1.3633	28.08
		9.7120	14.2847	14.2847			0.0837	0.1754	0.7905			2.2230	-0.7617	16.9833
168.4554		37.5758	78.8044	282.7120	FDI	2.5680	2.4024	-2.7574	23.5374	OLD	7.4954	3.8973	2.6626	19.8007
		35.6501	110.2655	254.2620			1.3459	0.8656	6.0239			3.7516	2.8016	15.0487
		13.7231	100.5851	203.4189			2.0068	-3.6958	20.6567			1.2800	3.5933	13.4206
8420.8400		2721237	0.00002	76000000	CRE	44.1892	31.5681	1.3839	166.5040	URB	65.0702	18.9787	25.5470	95.4260
		499276	0.7048	2546310			28.6856	10.0330	116.4342			18.8909	29.4060	92.8404
		2676777	-244789	7360000			14.2939	-9.8768	102.1003			4.0756	49.1013	82.9673
34.4912		280.9986	-26.3000	6261.2400	Ч	2.6469	0.5534	1.2226	3.8915	FEM	39.3354	7.8151	16.2893	51.3744
		63.7330	3.0097	234.4750			0.5198	1.4758	3.5327			7.7889	18.8524	49.8756
		273.9517	-195.6751	6066.427			0.2147	1.9296	3.1360			1.6332	33.4211	44.7758
	ŀ													
-	and deviatio	in, minimum,	. and maximu	im, respective	ly.									
				-										

varia
q
statistics
ptive
Descri
A1.4
b

JOURNAL OF APPLIED ECONOMICS 🛞 31

Table A	1.5 Pearsc	on's correl	ation mat	trix.													
	Gini_net	CRE	FDI	FEM	9	-	INF	OLD	Н	TFP	LAB	XR	OPNESS	GDP	GDP ²	UNE	URB
Gini_net	1.000																
CRE	-0.173***	1.000															
	(0000:0)																
FDI	-0.061	0.185***	1.000														
	(0.087)	(000.0)															
FEM	-0.224***	0.064	0.015	1.000													
	(000:0)	(0.073)	(0.669)														
0	-0.396***	0.356***	0.188***	0.147***	1.000												
	(000:0)	(000.0)	(0000)	(0000)													
_	-0.020	0.059	0.065	-0.243***	-0.034	1.000											
	(0.577)	(0.102)	(0.071)	(0000)	(0.338)												
INF	-0.026	-0.107***	-0.087	0.044	-0.049	-0.051	1.000										
	(0.472)	(0.003)	(0.015)	(0.220)	(0.176)	(0.153)											
OLD	-0.670***	0.187***	0.160***	0.525***	0.508***	-0.186***	0:030	1.000									
	(000:0)	(0000)	(0000)	(0000)	(0000)	(0000)	(0.401)										
Ϋ́	-0.603***	0.321***	0.245***	0.397***	0.601***	-0.101***	-0.007	0.772***	1.000								
	(000:0)	(0000)	(0000)	(0000)	(0000)	(0.005)	(0.854)	(0000)									
TFP	0.061	-0.032	0.031	-0.401***	0.127***	0.103***	-0.075	0.035	0.059	1.000							
	(0.087)	(0.373)	(0.394)	(0000)	(000.0)	(0.004)	(0.037)	(0.329)	(0.100)								
LAB	-0.274***	0.389***	-0.022	0.395***	0.322***	-0.116***	-0.045	0.266***	0.162***	433***	1.000						
	(00:0)	(0000)	(0.534)	(0000)	(0000)	(0.001)	(0.205)	(0000)	(0000)	(000)							
XR	-0.024	-0.026	-0.025	-0.019	-0.012	0.000	0.000	0.001	0.022	091	-0.030	1.000					
	(0.511)	(0.470)	(0.482)	(0.600)	(0.730)	(0.998)	(0.995)	(0.978)	(0.530)	(110)	(0.409)						
OPNESS	-0.069	0.331***	0.162***	0.043	0.191***	-0.091	-0.098***	0.294***	0.415***	.225***	-0.025	0.013	1.000				
	(0.053)	(000.0)	(0000)	(0.235)	(000.0)	(0.011)	(900:0)	(0000)	(0.000)	(000)	(0.490)	(0.727)					
GDP	-0.162***	0.308***	-0.147***	0.096***	-0.050	-0.333 ***	-0.025	0.174***	0.141***	008	0.104***	-0.099***	.242***	1.000			
	(000:0)	(0000)	(0000)	(0.007)	(0.160)	(0000)	(0.487)	(0000)	(0000)	(.824)	(0.004)	(0.006)	(000)				
GDP ²	-0.157***	0.329***	-0.145***	0.080	-0.047	-0.333***	-0.027	0.166***	0.127***	028	0.110***	-0.086	.244***	.997***	1.000		
	(00:0)	(000:0)	(0000)	(0.025)	(0.189)	(0.000)	(0.456)	(0000)	(0000)	(.430)	(0.002)	(0.017)	(000)	(000)			
UNE	0.255***	-0.051	-0.003	-0.146***	0.359***	-0.101***	-0.050	-0.028	-0.016	.336***	-0.018	-0.017	112***	119***	129***	1.000	
	(000:0)	(0.151)	(0:630)	(0000)	(000.0)	(0.005)	(0.165)	(0.429)	(0.648)	(000)	(0.619)	(0.645)	(.002)	(100.)	(000)		
URB	-0.226***	-0.037	0.225***	0.153***	0.391***	-0.157***	0.031	0.449***	0.606***	.278***	-0.227***	0.044	.175***	257***	267***	0.230***	1.000
	(000.0)	(0.308)	(0.000)	(0000)	(000.0)	(0.000)	(0.388)	(0000)	(0000)	(000)	(000.0)	(0.221)	(000)	(000)	(000)	(0.000)	
*** denoi	tes that the	coefficient	is at 1% sig	inificance le	vel. The de	spendent va	riable is Gir	ni_net here	. The result	s are simili	ar and thus	not reporte	ed when u	sing Gini_n	narket.		

results. RE RE Regressor Gini_net Regressor Gini_net GDP -10.73 CRE 0.004** (1.324)(0.006)GDP² 0.441 HC 0.489 (0.188)(0.684)XR 0.000 TFP -1.861 (0.000) (1.156) INF 0.000 I AB -11.107* (0.000) (1.552) G -0.27 UNE 0.188*** (0.042) (0.033) OLD -0.497*** Т 0.064 (0.095) (0.091) **OPNESS** -0.788 URB 0.010** (1.377) (0.027) FDI 0.193*** FEM -0.134*** (0.035) (0.051)

Table A2.1 Determinants of income inequality in the EMEs: RE

Appendix 2 Random effects results

p value in the bracket. *, **, **** denote that the coefficients are significant at 10%, 5%, 1% levels, respectively. The dependent variable is Gini_net here. The results are similar and thus not reported when using Gini_market.

Appendix 3 Dynamics of Gini coefficients in the EMEs



Figure A1. Dynamics of Gini coefficients in the EMEs.

S. WANG ET AL.



Figure A2. Dynamics of Gini coefficients in the EMEs.

34