Society for the Study of Economic Inequality

Working Paper Series

The inequality of extreme incomes

Lidia Ceriani
Paolo Verme

ECINEQ WP 2019 - 490

ECINEQ 2019 - 490
January 2019

www.ecineq.org

The inequality of extreme incomes

Lidia Ceriani†
Georgetown University, USA

Paolo Verme
The World Bank, USA

Abstract

The paper derives the conditions under which income inequality measured with the Gini index is expected to increase or decrease if missing observations are added at the top or/and at the bottom of an income distribution. It shows that adding observations on the extremes of the income distribution does not necessarily result in an increase in inequality, but that meeting the conditions for obtaining a decrease in inequality is unlikely. It also shows that adding observations at the top weighs more on inequality than adding observations at the bottom. These findings are confirmed by an application to US states data. Adding observations on the extremes of an income distribution should be normally expected to increase inequality and recovering missing observations at the top should be prioritized.

Keywords: Income inequality, income distributions, migration, top incomes; bottom incomes.

JEL Classification: D31, D63, E64, O15.

†Corresponding authors, e-mail: lidia.ceriani@georgetown.edu.

1 Introduction

It is well aknowledged in the economic literature that top incomes are not well measured in household surveys resulting in a likely underestimation of income inequality (Cowell and Victoria-Feser, 1996, Atkinson et al., 2011, Hlasny and Verme, 2018b, Hlasny and Verme, 2018a). Less explored is the impact on inequality of missing observations at the bottom of the income distribution and also the direction of changes in inequality when observations are added at the top and at bottom of an income distribution.

Missing observations at the bottom are an increasing sizable problem. According to the International Organization for Migration (IOM), global migration has increased between 1990 and 2015 in absolute terms (from 153m to 244m people) and relatively to the world population (from 2.9% to 3.3%)[1] and most of these migrants should be expected to be poor.[2] Forced migrants are not captured in national censuses or surveys whereas national statistical offices typically struggle to capture internal or illegal migrants. This is a potentially relevant phnomenon for the meaurement of inequality for both countries of origin and destination.

From a purely mathematical perspective, it can be shown that inequality does not necessarily increase if observations are added at the bottom, at the top or at both ends of an income distribution. This paper clarifies this point and provides the conditions under which income inequality should be expected to increase or decrease when observations are added to the tails of the distribution (Section 2). It also provides an empirical application that shows how inequality should be expected to change if all poor and rich individuals are accounted for (Section 3). Results show that inequality almost invariably increases by adding observations on the tails of an income distribution but that missing a few observations at the top is much more relevant than missing many observations at the bottom. This justifies the scholarly focus on top incomes.

2 The Model

Let us assume there are n individuals in a society. Each i-th individuals is endowed with income $x_i \in \{\mathbb{R}_+ \cup 0\}$. The vector $X = [x_1, x_2, \ldots, x_n]$ summarizes the income distribution in the society and μ_X is the mean income of distribution X. For simplicity of exposition, let us assume that incomes are ordered, in such a way that $x_1 < x_2 < \ldots < x_n$.

[1]https://www.iom.int/wmr/chapter-2

[2]Most economic migrants should be expected to be low income relatively to destination countries and most of the 71.4m forced migrants estimated by the United Nations High Commissioner for Refugees (UNHCR) should be expected to be poor. For example, as many as 90% of the Syrian refugees in Jordan and Lebanon have been found to be poor (Verme et al., 2016). Note also that displacement crises can account for up to 50% of the host population (Verme and Schuettler, 2018) resulting in a significant addition to the bottom of the income distribution of host countries.

We are interested in understanding the impact on inequality of adding to the original distribution more individuals at the top and/or at the bottom of the distribution. By *adding more individuals at the top (bottom)* we mean duplicating the richer (poorer) individual in the observed distribution $t-$times ($b-$times), with t, $b \in \mathbb{N}$. Starting from the observed distribution X, we can generate other three hypothetical new distributions: X^B, obtained from X by adding b new individuals having the same welfare level of the worse-off individual in the population; X^T, obtained from X by adding t new individuals having the same welfare level of the richest in the original distribution; and X^{TB}, obtained from X by adding both t individuals at the top and b individuals at the bottom of the distribution. The three new distributions are described as follows:

$$
\begin{aligned}
X^B &= [\,\underbrace{x_1, x_1, \ldots, x_1}_{\text{b times}}, \underbrace{x_1, x_2, \ldots, x_n}_{X}\,] \\
X^T &= [\,\underbrace{x_1, x_2, \ldots, x_n}_{X}, \underbrace{x_n, x_n, \ldots, n_n}_{\text{t times}}\,] \\
X^{BT} &= [\,\underbrace{x_1, x_1, \ldots, x_1}_{\text{b times}}, \underbrace{x_1, x_2, \ldots, x_n}_{X}, \underbrace{x_n, x_n, \ldots, n_n}_{\text{t times}}\,]
\end{aligned}
$$

Inequality is measured using the Gini index, described in following equation (1) (for a summary of the different possible formulation of the Gini index, see Ceriani and Verme, 2012):

$$G_X = \frac{1}{n^2 \mu_X} \sum_{i=1}^{n} \sum_{j>i} (x_j - x_i) \tag{1}$$

Since the Gini index gives more weight to changes occurring at the center of the distribution, and we are analyzing perturbations occurring at the tails, any conclusion we draw can be losely interpreted as the lower-bound estimate of the distributional impact of such changes.

Looking at equation (1), it is intuitive to notice that replicating observations at the tails of the distribution affects inequality via: (i) a *population* effect (change in n); (ii) a *mean* effect (change in μ_X), and (iii) a *deprivation* effect (changes in the arguments of the double sum). The three effects influence the index differently, and the overall impact of a change at the tails of the distribution is not straightforward.

We can express the number of individuals added at the top and at the bottom as a share of the total initial population: $t = \tau n$ and $b = \beta n$, with $\tau, \beta \in \mathbb{R}_+$. And the income level of the richest individual in the population and of the poorer individual in the population as a function of the original distribution mean income: $x_n = \lambda \mu_X$, $\lambda > 1$ and $x_1 = \gamma \mu_X$, $0 \leq \gamma \leq 1$. Then, as equations (2), (3), (4) summarize, the Gini index of the new distributions can be expressed only as functions of the original distribution's Gini index (G_X), the number of new individuals added at the bottom or top of the distribution as a share of the initial population (β and τ respectively), and the relative size of the welfare level of the poorer and richer individuals in the original distribution (γ and λ respectively).[3]

Table 1: Gini indices for different distributions where more observations are added at the extremes

Distribution	Gini Index	
X^B	$G_{X^B} = \dfrac{G_X + \beta(1-\gamma)}{(1+\beta)(1+\beta\gamma)}$	(2)
X^T	$G_{X^T} = \dfrac{G_X + \tau(\lambda-1)}{(1+\tau)(1+\tau\lambda)}$	(3)
X^{BT}	$G_{X^{BT}} = \dfrac{G_X + \tau(\lambda-1) + \beta(1-\gamma)}{(1+\beta+\tau)(1+\beta\gamma+\tau\lambda)}$	(4)

Next, we are interested in understanding how inequality is affected when more individuals are included at the extremes. In particular, we are interested in finding the conditions such that: (i) adding more individual at the bottom increases inequality with respect to the original distribution ($G_{X^B} \geq G_X$); (ii) adding more individual at the top increases inequality with respect to the original distribution ($G_{X^T} \geq G_X$); adding individual at the top of the distribution generates higher inequality than adding individuals at the bottom of the distribution ($G_{X^T} \geq G_{X^B}$); and (iv) adding individuals both at both extremes of the original distribution increases inequality with respect to the original distribution ($G_{X^{BT}} \geq G_X$). Table (2) summarizes the conditions for the four hypothesis to hold.

Hypothesis (H1) always holds whenever the welfare level of the poorest individual in society is zero (which is very likely when using a welfare aggregate such as income, consumption or expenditure). In this case, adding more individuals at the bottom always implies an increase in inequality. This is not necessarily true when the welfare level of the poorest individual in the society is different from zero. In this case, (H1) is most likely to hold the smallest is the share of individuals added at the bottom of the distribution (β), the smallest the level of wellbeing of the poorest individual in the population (γ), and the smallest the initial Gini index (G_X).

[3]The mathematical passages to obtain equations (2), (3), (4) are provided in the Appendix.

Table 2: Understanding inequality changes after adding extreme incomes

	Hypothesis	Condition
(H1)	$G_{X^B} \geq G_X$	$\beta \leq \frac{(1-\gamma) - G_X(\gamma+1)}{G_X\gamma}$, and $G_X \leq \frac{1-\gamma}{1+\gamma}$ if $\gamma \neq 0$ $G_X \leq 1$ if $\gamma = 0$
(H2)	$G_{X^T} \geq G_X$	$\tau \leq \frac{(\lambda-1) - G_X(\lambda+1)}{G_X\lambda}$, and $G_X \leq \frac{\lambda-1}{\lambda+1}$
(H3)	$G_{X^T} \geq G_{X^B}$	If $\gamma = 0$: $\beta \leq \frac{A - G_X B}{B - A}$, and where $A = G_X + \tau(\lambda-1)$ and $B = (1+\tau)(1+\tau\lambda)$
(H4)	$G_{X^{BT}} \geq G_X$	$G_X \leq \frac{\tau(\lambda-1) + \beta(1-\gamma)}{(1+\beta+\tau)(1+\beta\gamma+\tau\lambda) - 1}$

Adding more individuals at the top of the distribution is more likely to increase inequality (e.g. (H2) holds) the smallest the share of population added at the top of the distribution (τ), the smallest the initial inequality measure (G_X), and the largest the level of wellbeing of the richest individual in the population relative to the mean (λ).

Inequality is likely to be higher when individuals are added at the top rather than at the bottom of the distribution (hypothesis (H3)) the smallest the number of individual added at the bottom of the distribution (β), the smallest the share of individuals added at the top (τ), and the smallest the initial Gini index (G_X).

Finally, if the initial inequality level is high enough, adding more individuals at the bottom and at the top of the distribution may lead to a higher inequality level (e.g. (H4) holds).

Notice that the conditions such that (H1), (H2), (H3) and (H4) hold can be empirically tested using real household budget survey data.We only need few information on the initial income distribution: (i) the initial Gini index (G_X), how large is the income level of the poorest and the richest individual in the original distribution relative to the mean income (γ and λ respectively), and the number of added individuals at the bottom and/or at the top as a share of the initial population (β and τ respectively).

3 Empirical Application

For the empirical exercise, we use the 2016 wave of the United States' Current Population Survey and consider inequality in the 50 states of the Union. This is a unique data set in that we have comparable figures for 50 population groups characterized by a high degree of within group and between group inequality.[4] In 2016, the Gini index computed on total household per capita income in the US ranged from 0.38 in New Hampshire to 0.51 in the District of Columbia, with an average of 0.48. The average income per capita in the 50 US states was 32,245 dollars, ranging from 23,630 dollars in Mississippi to 52,532 dollars in the District of Columbia. The income of the richest individual was on average 21 times mean income, ranging from 8.4 times in Kentucky to 43 times in Vermont. The income of the poorest individuals was zero in all 51 states in our sample (see Table **??** for details).

Based on existing population statistics, it is also possible to provide gross estimates of the likely upper bounds of population changes at the bottom and at the top of the income distribution. The Census Bureau's estimates of population changes by state between 2015 and 2016 ranged from 0.01 to 2.02 percent.[5] The rate of natural increase of the population (the difference between the number of live births and the number of deaths occurring in a year) between July 2017 and July 2018 ranged between 0.06 and 0.59 percent across states. The annual internal interstate migration rate in the US between 2001 and 2010 was about 1.7 percent of the US population. This internal migration rate has been declining over the years and is similar in size for the top and bottom income 50 percent of the population (Molloy et al., 2017). Legal immigration in 2016 accounted for a migrant-population ratio of 0.37 percent.[6] The unauthorized migrant stock in the US in 2016 was estimated at 10.7m accounting for a share of 3.3 percent of the US population, which should be expected to be made of mostly low income individuals. Even if the full stock of unauthorized immigrants was legalized overnight, the maximum percentage variation in the US population would be below 4 percent with most of this people adding to the bottom of the income distribution. Therefore, upper bound estimates for population changes at the top and at the bottom of the income distribution in the US can be reasonably set at 1 and 3 percent respectively.

[4]For a full description of this data set, its variance within and across states and an analysis of missing observations see Hlasny and Verme, 2017.

[5]Estimated from Table 1. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2018 wired at: https://www.census.gov/data/tables/time-series.

[6]Or 1.183m people: https://www.migrationpolicy.org/programs/data-hub/us-immigration-trends.

These estimates provide the benchmarks for testing the hypotheses outlined in Table (2). Results are as follows:

- (H1) Since the poorest individuals in all states have zero income ($\gamma = 0$), (H1) is always satisfied: adding more individuals with zero income will always increase the Gini index with respect to the initial distribution;
- (H2) Given the initial distributions observed in the data, adding more top incomes will always increase inequality. The conditions under which (H2) holds ($G_{X^T} \geq G_X$) are in fact easily satisfied: the share of individuals added at the top must not exceed, on average, 104 percent, and that the initial Gini index must be, on average, smaller than 0.89;
- (H3) Adding 1 percent of the population at the top will result in a level if inequality always higher than inequality computed on any distribution where new individuals with zero incomes are added at the bottom, provided the share of individuals included at the bottom (β) does not exceed, on average, 18 percent. Such restriction on β ranges from 5 percent in Kentucky to 39 percent in Vermont;
- (H4) Finally, we search for the condition such that adding more observations at both extremes increases inequality (H4). Keeping $\tau = 0.01$, we let the share of individuals added at the bottom (β) to vary between 1 and 3 percent (similar results are found if we let β vary up to 100 percent). For inequality to increase when missing incomes are added at the top and at the bottom of the distribution, the initial Gini index must be smaller than 0.795 at the least (this is the most stringent condition, which refers to Kentucky, when both β and τ are set to 1 percent). This condition is always satisfied in the data.

4 Concluding Remarks

The paper has provided the mathematical treatment of the conditions under which income inequality increases or decreases if individuals are added at the top or at the bottom of the income distribution. The theoretical part showed that adding observations on the extremes of the income distribution does not necessarily result in an increase in inequality. It also showed that adding observations at the top is more likely to increase inequality than adding observations at the bottom and that the conditions to obtain a reduction in income inequality by adding observations on the tails of the distribution are unlikely to materialize with real data. The empirical application to the US 50 states confirms that adding observations at the top or at the bottom of the income distribution invariably results in an increase in inequality. Nevertheless, adding observations at the bottom has always a much more limited impact on increasing income inequality than adding observations at the top of the distribution. In order to have a complete picture of inequality, missing observations must be added at both extremes. But a few missing incomes at the top might generate much larger biases than a sizable number of missing incomes at the bottom, a fact that justifies the past focus on top incomes.

References

Atkinson, A., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature 49*, 3–71.

Ceriani, L. and P. Verme (2012). The origins of the gini index: Extracts from variabilita' e mutabilita' (1912) by corrado gini. *Journal of Economic Inequality 10*(3), 421–443.

Cowell, F. and M.-P. Victoria-Feser (1996). Poverty measurement with contaminated data: A robust approach. *European Economic Review 40*, 1761–1771.

Hlasny, V. and P. . Verme (2017). The impact of top incomes biases on the measurement of inequality in the united states. *Ecineq Working Paper* (452).

Hlasny, V. and P. . Verme (2018a). Top incomes and the measurement of inequality in egypt. *World Bank Economic Review 32*(32), 428–455.

Hlasny, V. and P. Verme (2018b). Top incomes and inequality measurement: A comparative analysis of correction methods using the eu silc data. *Econometrics 6*(2), 1–21.

Molloy, R., C. L. Smith, and A. K. . Wozniak (2017). Internal migration in the united states. *NBER Working Paper* (17307).

Verme, P., C. Gigliarano, C. Wieser, K. Hedlund, M. Petzoldt, and M. Santacroce (2016). *The Welfare of Syrian Refugees: Evidence from Jordan and Lebanon.* Washington DC: World Bank.

Verme, P. and K. Schuettler (2018). The impact of forced displacement on host communities: A review of the empirical literature in economics. World Bank, Mimeo.

Table 3: Descriptive Statistics and Hypotheses Conditions

	Original Distributions				**H1:** $G_B > G$	**H2:** $G_T > G$		**H3:** $G_T > G_B$ $\tau = 0.01$	**H4:** $G_{TB} \geq G$ $\tau = 0.01$ $\beta = 0.01$	$\tau = 0.01$ $\beta = 0.02$	$\tau = 0.01$ $\beta = 0.03$
State	G_X	μ_X	γ	λ	$G_X \leq$	$\tau \leq$	$G_X \leq$	$\beta \leq$	$G_X \leq$	$G_X \leq$	$G_X \leq$
AK	0.4539	35,880	0	25.3	1	1.0765	0.9241	0.216	0.9100	0.9051	0.9007
AL	0.4593	27,871	0	33.0	1	1.0809	0.9411	0.2876	0.9254	0.9191	0.9133
AR	0.4553	26,324	0	40.5	1	1.1175	0.9518	0.3592	0.9351	0.9281	0.9215
AZ	0.4930	29,591	0	37.5	1	0.9477	0.9480	0.3266	0.9317	0.9249	0.9186
CA	0.5025	34,141	0	32.5	1	0.8981	0.9404	0.2789	0.9247	0.9185	0.9128
CO	0.4376	35,637	0	10.6	1	0.9744	0.8271	0.0764	0.8269	0.8331	0.8384
CT	0.4456	40,919	0	14.6	1	1.0222	0.8719	0.1145	0.8644	0.8650	0.8654
DC	0.5102	52,532	0	26.4	1	0.8481	0.9271	0.2207	0.9127	0.9076	0.9029
DE	0.4070	32,612	0	10.7	1	1.1333	0.8287	0.0795	0.8283	0.8342	0.8393
FL	0.4784	30,654	0	36.6	1	1.0057	0.9468	0.3194	0.9305	0.9238	0.9176
GA	0.4658	28,531	0	12.5	1	0.8958	0.8522	0.0931	0.8477	0.8506	0.8531
HI	0.4370	31,566	0	35.3	1	1.1949	0.9449	0.3115	0.9288	0.9223	0.9162
IA	0.4439	33,719	0	26.9	1	1.1319	0.9284	0.2318	0.9139	0.9087	0.9039
ID	0.4546	28,197	0	19.4	1	1.0350	0.9021	0.1596	0.8905	0.8878	0.8853
IL	0.4612	35,803	0	17.3	1	0.9854	0.8909	0.1390	0.8807	0.8791	0.8777
IN	0.4228	28,897	0	10.5	1	1.0445	0.8260	0.0767	0.8260	0.8323	0.8377
KS	0.4501	31,192	0	9.3	1	0.8766	0.8064	0.0638	0.8102	0.8192	0.8268
KY	0.4604	26,527	0	8.4	1	0.7957	0.7880	0.0544	0.7954	0.8072	0.8170
LA	0.4829	29,131	0	38.6	1	0.9915	0.9495	0.3384	0.9330	0.9262	0.9198
MA	0.4510	39,158	0	19.4	1	1.0510	0.9017	0.1591	0.8902	0.8875	0.8850
MD	0.4211	37,967	0	9.7	1	1.0280	0.8136	0.0695	0.8160	0.8240	0.8308
ME	0.4048	30,557	0	13.1	1	1.2060	0.8584	0.1031	0.8530	0.8551	0.8569
MI	0.4241	31,488	0	12.1	1	1.0816	0.8479	0.0924	0.8441	0.8476	0.8505
MN	0.4272	36,085	0	8.9	1	0.9655	0.7980	0.0612	0.8034	0.8137	0.8223
MO	0.4251	32,112	0	14.4	1	1.1198	0.8703	0.1141	0.8631	0.8638	0.8644
MS	0.4625	23,630	0	24.0	1	1.0306	0.9200	0.2022	0.9063	0.9019	0.8978
MT	0.4277	30,034	0	10.2	1	1.0113	0.8217	0.0737	0.8225	0.8294	0.8353
NC	0.4638	30,056	0	20.0	1	0.9982	0.9048	0.1642	0.8929	0.8898	0.8871
ND	0.4323	34,018	0	16.9	1	1.1176	0.8884	0.1374	0.8786	0.8772	0.8761
NE	0.4392	31,107	0	35.6	1	1.1848	0.9453	0.3143	0.9292	0.9226	0.9165
NH	0.3790	38,226	0	14.5	1	1.3872	0.8707	0.1176	0.8634	0.8641	0.8646
NJ	0.4426	36,571	0	30.2	1	1.1515	0.9360	0.2633	0.9207	0.9149	0.9095
NM	0.4631	25,438	0	21.5	1	1.0126	0.9112	0.1787	0.8985	0.8949	0.8915
NV	0.4642	29,000	0	39.1	1	1.0739	0.9502	0.3452	0.9336	0.9267	0.9202
NY	0.4843	35,803	0	21.4	1	0.9214	0.9107	0.1755	0.8981	0.8945	0.8912
OH	0.4357	29,453	0	19.4	1	1.1253	0.9021	0.1610	0.8905	0.8877	0.8852
OK	0.4917	29,407	0	20.1	1	0.8828	0.9052	0.1626	0.8932	0.8902	0.8874
OR	0.4171	31,188	0	12.9	1	1.1352	0.8566	0.1006	0.8514	0.8538	0.8558
PA	0.4361	33,910	0	18.2	1	1.1118	0.8958	0.1493	0.8850	0.8829	0.8810
RI	0.4373	35,308	0	12.0	1	1.0127	0.8461	0.0901	0.8426	0.8463	0.8495
SC	0.4496	28,847	0	15.1	1	1.0109	0.8759	0.1189	0.8678	0.8679	0.8679
SD	0.4560	31,420	0	12.5	1	0.9367	0.8514	0.0932	0.8471	0.8501	0.8527
TN	0.4736	29,830	0	14.8	1	0.9019	0.8737	0.1144	0.8659	0.8662	0.8665
TX	0.4847	30,808	0	33.0	1	0.9705	0.9413	0.2856	0.9255	0.9192	0.9134
UT	0.4292	27,373	0	40.4	1	1.2474	0.9517	0.3608	0.9350	0.9280	0.9214
VA	0.4515	35,161	0	11.0	1	0.9236	0.8338	0.0799	0.8324	0.8377	0.8422
VT	0.4333	35,326	0	43.0	1	1.2310	0.9545	0.3848	0.9376	0.9304	0.9236
WA	0.4453	34,378	0	11.0	1	0.9508	0.8335	0.0802	0.8322	0.8375	0.8421
WI	0.4100	31,121	0	11.9	1	1.1510	0.8454	0.0914	0.8421	0.8458	0.8491
WV	0.4474	26,435	0	24.7	1	1.1041	0.9222	0.2105	0.9083	0.9037	0.8994
WY	0.4328	33,577	0	16.1	1	1.1043	0.8827	0.1292	0.8737	0.8730	0.8724

Source: Authors' elaboration on data from the Current Population Survey, 2016.

A1 Appendix

A1.1 Derivation of Gini Index for distributions X^B, X^T, X^{BT}

Let us define $x_0 = \gamma\mu_X$, and $x_n = \lambda\mu_X$

$$\begin{aligned}
G_{X^B} &= \frac{1}{(\beta n+n)^2\mu_{X^B}}\sum_{i=1}^{\beta n+n}\sum_{j>i}(x_j-x_i)\\
&= \frac{1}{(n+\beta n)^2\frac{1}{(n+\beta n)}\sum_{i=1}^{n+\beta n}x_i}\left[\sum_{i=1}^{n}\sum_{j>i}(x_j-x_i)+n\beta\sum_{j>0}(x_j-x_0)\right]\\
&= \frac{1}{(n+\beta n)(n\mu_X+\beta n x_0)}\left(n^2\mu_X G_X+n^2\beta\mu_X-n^2\beta x_0\right)\\
&= \frac{n^2\left(\mu_X G_X+\beta\mu_X-\beta\gamma\mu_X\right)}{n^2(1+\beta)(\mu_X+\beta\gamma\mu_X)}\\
&= \frac{\mu_X\left(G_X+\beta-\beta\gamma\right)}{\mu_X(1+\beta)(1+\beta\gamma)}\\
&= \frac{G_X+\beta(1-\gamma)}{(1+\beta)(1+\beta\gamma)}
\end{aligned}$$

$$\begin{aligned}
G_{X^T} &= \frac{1}{(n+\tau n)^2\mu_{X^T}}\sum_{i=1}^{n+\tau n}\sum_{j>i}(x_j-x_i)\\
&= \frac{1}{(n+\tau n)^2\frac{1}{(n+\tau n)}\sum_{i=1}^{n+\tau n}x_i}\left[\sum_{i=1}^{n}\sum_{j>i}(x_j-x_i)+n\tau\sum_{i=1}^{n}(x_n-x_i)\right]\\
&= \frac{1}{(n+\tau n)(n\mu_X+\tau n x_0)}\left(n^2\mu_X G_X+n^2\tau x_n-n^2\tau\mu_X\right)\\
&= \frac{n^2\left(\mu_X G_X+\tau\lambda\mu_X-\tau\mu_X\right)}{n^2(1+\tau)(\mu_X+\tau\lambda\mu_X)}\\
&= \frac{\mu_X\left(G_X+\tau\lambda-\tau\right)}{\mu_X(1+\tau)(1+\tau\lambda)}\\
&= \frac{G_X+\tau(\lambda-1)}{(1+\tau)(1+\tau\lambda)}
\end{aligned}$$

$$
\begin{aligned}
G_{X^{BT}} &= \frac{1}{(\beta n+n+\tau n)^2 \mu_{X^T}} \sum_{i=1}^{\beta n+n+\tau n} \sum_{j>i} (x_j - x_i) \\
&= \frac{1}{\frac{(\beta n+n+\tau n)^2}{(\beta n+n+\tau n)} \sum_{i=1}^{\beta n+n+\tau n} x_i} \left[n\beta \sum_{j>0} (x_j - x_0) + \sum_{i=1}^{n} \sum_{j>i} (x_j - x_i) + n\tau \sum_{i=1}^{n} (x_n - x_i) \right] \\
&= \frac{1}{(\beta n+n+\tau n)(\beta n x_0 + n\mu_X + \tau n x_n)} \left(n^2 \beta \mu_X - n^2 \beta x_0 + n^2 \mu_X G_X + n^2 \tau x_n - n^2 \tau \mu_X \right) \\
&= \frac{n^2 \left(\beta \mu_X - \beta x_0 + \mu_X G_X + \tau x_n - \tau \mu_X \right)}{n^2 (\beta + 1 + \tau)(\beta \gamma \mu_X + \mu_X + \tau \lambda \mu_X)} \\
&= \frac{\mu_X \left(\beta - \beta\gamma + G_X + \tau\lambda - \tau \right)}{\mu_X (\beta + 1 + \tau)(\beta\gamma + 1 + \tau\lambda)} \\
&= \frac{\beta(1-\gamma) + G_X + \tau(\lambda - 1)}{(\beta + 1 + \tau)(\beta\gamma + 1 + \tau\lambda)}
\end{aligned}
$$

A1.2 Testing the Hypothesis

H1

If $\gamma \neq 0$:

$$
\begin{aligned}
G_{X^B} &\geq G_X \\
\frac{G_X + \beta(1-\gamma)}{(1+\beta)(1+\beta\gamma)} &\geq G_X \\
G_X + \beta(1-\gamma) &\geq G_X(1+\beta)(1+\beta\gamma) \\
G_X - G_X - G_X\beta\gamma - G_X\beta - G_X\beta^2\gamma &\geq -\beta + \beta\gamma \\
-G_X\gamma - G_X - G_X\beta\gamma &\geq -1 + \gamma \\
\beta &\leq \frac{(1-\gamma) - G_X(\gamma+1)}{G_X\gamma}
\end{aligned}
$$

And since $\beta \geq 0$, $(1-\gamma) - G_X(\gamma+1) \geq 0$, or $G_X \leq \frac{1-\gamma}{1+\gamma}$.
If $\gamma = 0$:

$$
\begin{aligned}
G_{X^B} &\geq G_X \\
\frac{G_X + \beta}{(1+\beta)} &\geq G_X \\
G_X + \beta - G_X - G_X\beta &\geq 0 \\
G_X &\leq 1
\end{aligned}
$$

H2

$$\begin{aligned}
G_{X^T} &\geq G_X \\
\frac{G_X + \tau(\lambda - 1)}{(1+\tau)(1+\tau\lambda)} &\geq G_X \\
G_X + \tau(\lambda - 1) &\geq (G_X + \tau G_X)(1+\tau\lambda) \\
G_X + \tau\lambda - \tau &\geq G_X + G_X\tau\lambda + G_X\tau + G_X\tau^2\lambda \\
\lambda - 1 &\geq G_X\lambda + G_X + G_X\tau\lambda \\
\tau &\leq \frac{(\lambda - 1) - G_X(\lambda + 1)}{G_X\lambda}
\end{aligned}$$

And since $\tau \geq 0$, $(\lambda - 1) - G_X(\lambda + 1) \geq 0$, or $G_X \leq \frac{\lambda-1}{\lambda+1}$.

H3

$$\begin{aligned}
G_{X^T} &\geq G_{X^B} \\
\frac{G_X + \tau(\lambda - 1)}{(1+\tau)(1+\tau\lambda)} &\geq \frac{G_X + \beta(1-\gamma)}{(1+\beta)(1+\beta\gamma)}
\end{aligned}$$

If $\gamma = 0$:

$$\begin{aligned}
\frac{G_X + \tau(\lambda - 1)}{(1+\tau)(1+\tau\lambda)} &\geq \frac{G_X + \beta}{(1+\beta)} \\
G_X + \tau(\lambda - 1) - G_X(1+\tau)(1+\tau\lambda) &\geq \beta(1+\tau)(1+\tau\lambda) - \beta\left[G_X + \tau(\lambda - 1)\right] \\
\beta &\leq \frac{G_X + \tau(\lambda - 1) - G_X(1+\tau)(1+\tau\lambda)}{(1+\tau)(1+\tau\lambda) - G_X + \tau(\lambda - 1)}
\end{aligned}$$

Since $\beta \geq 0$, the right hand side of the inequality must be greater or equal than zero. The denominator is by definition greater than zero. For the numerator to be greater than zero:

$$\begin{aligned}
G_X + \tau(\lambda - 1) - G_X(1+\tau)(1+\tau\lambda) &\geq 0 \\
G_X + \tau\lambda - \tau - G_X - G_X\tau - G_X\tau\lambda - G_X\tau^2\lambda &\geq 0 \\
\tau &\leq \frac{\lambda - 1 - G_X - G_X\lambda}{G_X\lambda} \\
\tau &\leq \frac{(\lambda - 1) - G_X(\lambda + 1)}{G_X\lambda}
\end{aligned}$$

Finally, $\tau \geq 0$, therefore $(\lambda - 1) - G_X(\lambda + 1) \geq 0$, or $G_X \leq \frac{\lambda-1}{\lambda+1}$.

H4

$$\begin{aligned}
G_{X^{BT}} &\geq G_X \\
\frac{G_X + \tau(\lambda - 1) + \beta(1-\gamma)}{(1+\beta+\tau)(1+\beta\gamma+\tau\lambda)} &\geq G_X \\
G_X + \tau(\lambda - 1) + \beta(1-\gamma) &\geq G_X(1+\beta+\tau)(1+\beta\gamma+\tau\lambda) \\
G_X\left[1 - (1+\beta+\tau)(1+\beta\gamma+\tau\lambda)\right] &\geq -\left[\tau(\lambda-1) + \beta(1-\gamma)\right] \\
G_X\left[(1+\beta+\tau)(1+\beta\gamma+\tau\lambda) - 1\right] &\leq \tau(\lambda-1) + \beta(1-\gamma) \\
G_X &\leq \frac{\tau(\lambda-1) + \beta(1-\gamma)}{(1+\beta+\tau)(1+\beta\gamma+\tau\lambda) - 1}
\end{aligned}$$