

Consistent Estimation and Inference When Data Follows a Power Law*

David J. Price[†]

October 12, 2016

PRELIMINARY AND INCOMPLETE
COMMENTS AND SUGGESTIONS WELCOME

Abstract

Power laws are common in economic phenomena, such as the size of cities and firms. When they occur, these power laws can cause estimates of economic quantities to have extremely different variances when those quantities observed at an aggregate level (for example, at the city or firm level). I show that in general, estimators based on observations exhibiting this extreme heteroskedasticity may not be consistent or asymptotically normal, and may have unreliable confidence intervals. In fact, these problems can occur even when no heteroskedasticity is present in the original data if estimates are obtained using weighting (such as by city or firm size). I propose new estimators for these contexts that may help in determining the extent of extreme heteroskedasticity in the data, in forming more accurate estimates, and performing more reliable inference. Further work remains to be done in understanding the properties of these new estimators, and in applying them to empirical work.

JEL Classification: C12, C13, C18

*I am grateful for feedback from Han Hong, Davide Malacrino, Alonso Villacorta, and seminar participants at Stanford University.

[†]Stanford University; djprice@stanford.edu

1 Introduction

As documented by [Gabaix \(2009, 2016\)](#) and others, power laws are common in economic phenomena, from city and firm size to stock markets, income, and wealth. These power laws generally indicate that the variable of interest A , over some range, has a cumulative distribution of

$$P(A > x) = kx^{-\frac{1}{s}} \tag{1}$$

for some parameters k and s . This relationship approximately holds with $s = 1$ for the population of cities (see, for example, [Eeckhout \(2004\)](#) and [Rozenfeld et al. \(2011\)](#)) and for the size of firms (see, for example, [Axtell \(2001\)](#)). If n units are drawn from the power law shown in Equation 1, a “rank-size” rule approximately applies, where unit t is the t -ranked unit by A_t (often a size):

$$A_t \approx A_1 t^{-s} \tag{2}$$

Research has helped illuminate the cause of these power laws, and some econometric research has explored how to estimate the parameter s based on observational data. However, what has not been noted is the potential for these power laws to create an extreme form of heteroskedasticity. Unlike usual heteroskedasticity, this extreme heteroskedasticity can cause standard econometric tools, such as ordinary and weighted least squares (OLS and WLS, respectively), to fail to provide consistent estimation or inference.

To understand how such extreme heteroskedasticity can come about, consider an economic system of A_t units it within T groups t , where A_t follows a power law with parameter s . (For simplicity, assume that $t = 1$ indicates the largest group, $t = 2$ the second largest, and so on.) Now, consider a model where $y_{it} = \theta + \epsilon_{it}$ for some parameter θ , with $\epsilon = \eta_{it} + \nu_t$.¹ Suppose that the error terms $\{\eta_{it}\}$, $\{\nu_t\}$ are mean-zero and mutually independent, with identical variances for each error term (though possibly different variances between them): $\mathbb{V}[\eta_{it}] = \sigma_\eta$ for all it , and $\mathbb{V}[\nu_t] = \sigma_\nu$ for all t . Now, suppose that we wish to estimate θ , but only the group-wide average y_t is observed, where

$$y_t = \frac{1}{A_t} \sum_{i=1}^{A_t} y_{it}. \tag{3}$$

As an example, to make this more concrete, θ may be underlying ability on a test, which is equal for all individuals. However, when taking the test, individuals’ scores may vary from θ for two reasons: first, due to random variation at the individual level (because the test is not perfect, so there is measurement error); and due to factors that vary at the city level and affect test scores, such as the education system. Scores are only observed by the econometrician as an average at the city level.

¹Under reasonable regularity conditions, the problems encountered and techniques developed will likely be similar for other general method of moment estimators. For ease of exposition, however, I will study this simple environment.

We now define

$$\eta_t \equiv A_t^{\frac{1}{2}} \frac{1}{A_t} \sum_{i=1}^{A_t} \eta_{it}, \quad (4)$$

so that $\{\eta_t\}$ are homoskedastic with variance $\mathbb{V}[\eta_t] = \sigma_\eta$. Setting $A_1 = 1$ (without loss of generality, as a change in σ_η results in the same equation), we can now write

$$\epsilon_t = A_t^{-\frac{1}{2}} \eta_t + \nu_t \approx t^{-\frac{s}{2}} \eta_t + \nu_t. \quad (5)$$

In this case, observations of the largest groups will be much smaller variances than observations of the smallest groups. I show that this extreme heteroskedasticity can lead standard estimates of θ to be inconsistent, or to have incorrect asymptotic inference.² This can even be the case if $\sigma_\eta = 0$ —that is, if there is no actual heteroskedasticity at all in the original data—if the econometrician assumes there is and uses WLS (or if WLS is used for any other reason). The weighting involved in WLS induces extreme heteroskedasticity in the data.

I show these properties as conditions on the parameter s , where higher s indicates more extreme heteroskedasticity. I focus on the case of $s = 1$, because that is a common parameter (holding for cities and firms, for example), and because it is a particularly problematic one: for many of the propositions below, $s = 1$ is an edge case between consistency and the lack thereof. In this case, even if consistency holds for $s = 1$, estimates may converge very slowly to their probability limits, which can also cause problems in practice, where sample sizes are not infinite.

The remainder of this paper proceeds as follows. Section 2 discusses OLS estimation, while Section 3 discusses WLS estimation. Section 4 discusses how the analysis may differ with alternative asymptotics. Section 5 presents potential solutions when extreme heteroskedasticity arises. Section 6 concludes.

2 Standard unweighted parameter estimates

Often, a researcher will attempt to estimate θ as an unweighted mean,

$$\hat{\theta}_{uw} \equiv \frac{1}{T} \sum_{t=1}^T y_t = \theta + \frac{1}{T} \sum_{t=1}^T (t^{\frac{s}{2}} \eta_t + \nu_t). \quad (6)$$

This unweighted estimate may be made, for example, if the functional form of the heteroskedasticity (the “population” in this example) is not known a priori. However, this unweighted estimator may not be consistent or asymptotically normal. In fact, under general conditions, a necessary and sufficient condition for $\hat{\theta}_{uw} \xrightarrow{P} \theta$ is that $s < 1$.

Proposition 2.1. *Suppose $\hat{\theta}_{uw}$ is defined as in Equation 6, where η_t and ν_t are mean-zero, independently*

²If the group size only approximately obeys a power law, the following propositions can be used as approximations to the true behavior of the estimators, or guidelines for how to prove (lack of) convergence for actual population of interest.

distributed, and homoskedastic with finite variances σ_η^2 and σ_ν^2 , respectively. If $s < 1$, then $\hat{\theta}_{uw} \xrightarrow{P} \theta$. If additionally $\{\eta_t^2\}$ and $\{\nu_t^2\}$ are each uniformly integrable; $\sigma_\eta^2 > 0$; and $s \geq 1$; then $\hat{\theta}_{uw} \not\xrightarrow{P} \theta$.

Proof. (Convergence if $s < 1$) Using Chung (1974), Theorem 5.4.1, Corollary i (p. 125), a sufficient condition for $\hat{\theta}_{uw} \xrightarrow{P} \theta$ is that

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{1}{t^2} \mathbb{E} \left[(t^{\frac{s}{2}} \eta_t + \nu_t)^2 \right] < \infty. \quad (7)$$

(In fact, this is sufficient for almost sure convergence.) Because η_t and ν_t are independent and have mean zero, we have

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{1}{t^2} \mathbb{E} \left[(t^{\frac{s}{2}} \eta_t + \nu_t)^2 \right] = \lim_{T \rightarrow \infty} \left\{ \sigma_\eta^2 \sum_{t=1}^T t^{s-2} + \sigma_\nu^2 \sum_{t=1}^T t^{-2} \right\}. \quad (8)$$

The sum on the left converges if $s < 1$, and the sum on the right always converges. To see this, note that by the definition of an integral as the sum of the area under a curve, for $p < -1$,

$$\sum_{t=1}^T t^p \leq 1 + \int_1^{T-1} t^p dt = \frac{1}{p+1} (T-1)^{p+1} + \frac{p}{p+1}. \quad (9)$$

□

Proof. (Non-convergence if $s \geq 1$) Using the methodology from the proof above, it is clear that the average of the ν_t terms converges. Thus a sufficient condition for non-convergence of $\hat{\theta}_{uw}$ is non-convergence of the average of the $t^{\frac{s}{2}} \eta_t$ terms.

Setting $\sigma_\nu^2 = 0$ for convenience, we have $\mathbb{V} \left[\hat{\theta}_{uw} \right] = \sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T t^s$. Thus for $s = 1$, we have $\mathbb{V} \left[\hat{\theta}_{uw} \right] \rightarrow C$ for some constant C ; and for $s > 1$, we have $\mathbb{V} \left[\hat{\theta}_{uw} \right] \rightarrow \infty$. In either case, if $\hat{\theta}_{uw} \xrightarrow{P} \theta$, then by Slutsky we would have $\left(\mathbb{V} \left[\hat{\theta}_{uw} \right] \right)^{-\frac{1}{2}} (\hat{\theta}_{uw} - \theta) \xrightarrow{P} 0$. But by Proposition 2.3, below (which does not use this result), $\left(\mathbb{V} \left[\hat{\theta}_{uw} \right] \right)^{-\frac{1}{2}} (\hat{\theta}_{uw} - \theta) \xrightarrow{d} \mathbb{N}(0, 1)$, a contradiction. □

Thus the unweighted estimator will not generally be consistent if $s \geq 1$. In fact, as shown in Figure 1, for $s > 1$, each additional observation actually makes the estimator worse. Even with $s = 1$, having an infinite number of observations leads to an estimator with about the same variance as if there are two observations with $s = 0$ (that is, errors are homoskedastic).

However, if both η_t^2 and ν_t^2 are uniformly integrable, then $\hat{\theta}_{uw}$ will be asymptotically normal for any s . For this proof, I will use the following lemma, which will also be used later in this paper. It may also be useful in applying these results to populations that do not obey a power law. (It is simple enough that it probably exists in some book, but I can't find it.)

Lemma 2.2. *Suppose $\{\epsilon_t\}$ are mean-zero, independently distributed, homoskedastic random variables with finite variance σ_ϵ^2 . Further, suppose $\{\epsilon_t^2\}$ are uniformly integrable, and that there is some function $g(T)$ such*

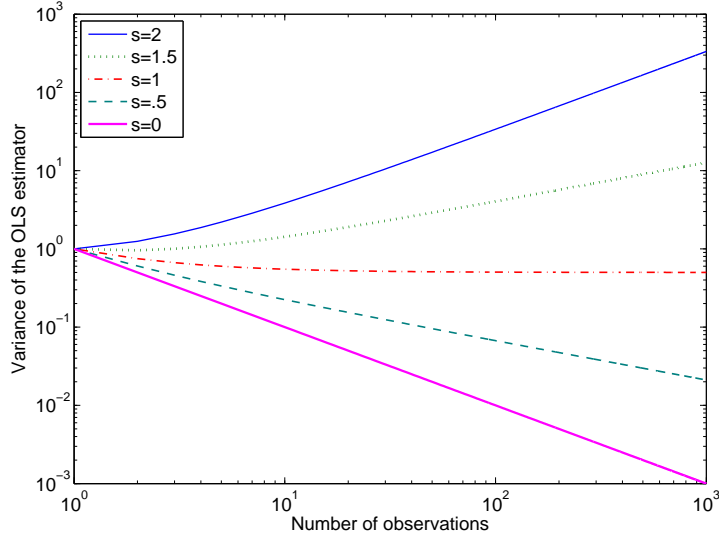


Figure 1: Variance of the OLS estimator: $\mathbb{V}[\hat{\theta}_{uw}] = \frac{\sigma_\eta^2}{T^2} \sum_{t=1}^T t^s + \frac{\sigma_\nu^2}{T}$. Variance is shown as sample size increases, with various values of s , setting $\sigma_\eta^2 = 1$ and $\sigma_\nu^2 = 0$. Additional data is assumed to be added from most accurate to least accurate. If $s < 1$, the variance converges to 0; if $s = 1$, the variance converges to a constant; and if $s > 1$, the variance grows without bound.

that, for all T , $g(T) \neq 0$; and a function $f(t)$ such that

$$\lim_{T \rightarrow \infty} \sup_{t \leq T} \frac{f(t)^2}{\sum_{s=1}^T f(s)^2} = 0. \quad (10)$$

Define $X_{Tt} \equiv g(T)f(t)\epsilon_t$; $S_T \equiv \sum_{t=1}^T X_{Tt}$; and $s_T^2 \equiv \sum_{t=1}^T \mathbb{V}[X_{Tt}]$. Then $\frac{S_T}{s_T} \xrightarrow{d} \mathbb{N}(0, 1)$.

Proof. It is sufficient to prove that the Lindeberg condition applies with the assumptions above. That is, for all $c > 0$, we must prove that

$$\mathbb{L} \equiv \lim_{T \rightarrow \infty} \sum_{t=1}^T \int \frac{X_{Tt}^2}{s_T^2} \mathbb{I} \left[\frac{X_{Tt}^2}{s_T^2} \geq c \right] dP = 0. \quad (11)$$

Note that $s_T^2 = \sigma_\epsilon^2 \sum_{t=1}^T g(T)^2 f(t)^2$. All $g(T)$ functions thus cancel out from Equation 11. Define $\alpha(t, T) \equiv$

$\left(\frac{f(t)^2}{c\sigma_\epsilon^2 \sum_{s=1}^T f(s)^2}\right)^{-1}$, and $\alpha(T) = \inf_{t \leq T} \alpha(t, T)$. Then

$$\begin{aligned}
\mathbb{L} &= \frac{1}{\sigma_\epsilon^2} \lim_{T \rightarrow \infty} \frac{1}{\sum_{t=1}^T f(t)^2} \sum_{t=1}^T f(t)^2 \int \epsilon_t^2 \mathbb{I}[\epsilon_t^2 \geq \alpha(t, T)] dP \\
&\leq \frac{1}{\sigma_\epsilon^2} \lim_{T \rightarrow \infty} \frac{1}{\sum_{t=1}^T f(t)^2} \sum_{t=1}^T f(t)^2 \int \epsilon_t^2 \mathbb{I}[\epsilon_t^2 \geq \alpha(T)] dP \\
&\leq \frac{1}{\sigma_\epsilon^2} \lim_{T \rightarrow \infty} \sup_t \int \epsilon_t^2 \mathbb{I}[\epsilon_t^2 \geq \alpha(T)] dP \\
&= \frac{1}{\sigma_\epsilon^2} \lim_{\alpha \rightarrow \infty} \sup_t \int \epsilon_t^2 \mathbb{I}[\epsilon_t^2 \geq \alpha] dP = 0.
\end{aligned} \tag{12}$$

The first inequality is by definition of $\alpha(T)$; the second inequality is from noting that the term is a convex combination of the positive-valued integrals; the final line's first equality is from noting that $\lim_{T \rightarrow \infty} \alpha(T) = \infty$ by the assumption in Equation 10 and using the fact that $f(t)^2 \geq 0$; and the final equality is from the assumption of uniform integrability. (Note that the first equality of the final line is only an equality on the condition that the second limit converges; but it does by assumption, so it works.) \square

Proposition 2.3. *Suppose $\hat{\theta}_{uw}$ is defined as in Equation 6, where η_t and ν_t are mean-zero, independently distributed, and homoskedastic with finite variances σ_η^2 and σ_ν^2 , respectively; and that $\{\eta_t^2\}$ and $\{\nu_t^2\}$ are uniformly integrable. Then $g(T) \left(\hat{\theta}_{uw} - \theta\right) \xrightarrow{d} \mathbb{N}(0, 1)$ for some function $g(T)$.*

Proof. First, note that $\hat{\theta}_{uw} - \theta = \sum_{t=1}^T \frac{1}{T} t^{\frac{s}{2}} \eta_t + \sum_{t=1}^T \frac{1}{T} \nu_t$. I will prove that each term converges to a normal distribution; the sum of independent normals is normal, so the sum also converges to a normal distribution. Using Lemma 2.2, I only need to prove that, for $s \geq 0$,

$$\lim_{T \rightarrow \infty} \sup_{t \leq T} \frac{t^s}{\sum_{s=1}^T t^s} = 0. \tag{13}$$

Note that for $s \geq 0$, we have $\sum_{t=1}^T t^s \geq \int_0^T t^s ds = \frac{1}{s+1} T^{s+1}$, so the limit in Equation 13 is $\lim_{T \rightarrow \infty} (s+1)T^{-1} = 0$. \square

Thus we know that $\left(\mathbb{V}[\hat{\theta}_{uw}]\right)^{-\frac{1}{2}} \left(\hat{\theta}_{uw} - \theta\right) \xrightarrow{d} \mathbb{N}(0, 1)$. However, we do not know $\mathbb{V}[\hat{\theta}_{uw}]$ a priori. Thus to perform inference, we must find some feasibly estimated \hat{V} such that $\hat{V} \left(\mathbb{V}[\hat{\theta}_{uw}]\right)^{-1} \xrightarrow{P} 1$; using Slutsky, we can then show that $\hat{V}^{-\frac{1}{2}} \left(\hat{\theta}_{uw} - \theta\right) \xrightarrow{d} \mathbb{N}(0, 1)$.

If $s \geq 1$, the estimates themselves will not converge, so the standard proof that heteroskedasticity-robust standard errors are consistent does not go through. In fact, if η_t and ν_t have finite kurtosis, then these estimated standard errors will indeed be consistent.

Proposition 2.4. *Define $\hat{\epsilon}_t \equiv y_t - \hat{\theta}_{uw}$, and $\hat{V} \equiv \frac{1}{T^2} \sum_{t=1}^T \hat{\epsilon}_t^2$. As above, suppose that $\hat{\theta}_{uw}$ is defined as in Equation 6, where η_t and ν_t are mean-zero, independently distributed, and homoskedastic with finite variances σ_η^2 and σ_ν^2 , respectively. If, additionally, η_t and ν_t have uniformly bounded kurtosis; then $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{uw}]} \xrightarrow{P} 1$.*

Proof. If $\sigma_\eta^2 = 0$, then this is a standard OLS estimator, so clearly standard errors are consistent. If $\sigma_\eta^2 > 0$, then the η_t term dominates both the variance of the estimator, and estimated variance; for simplicity, I will therefore only consider this term.

For estimated variance, we have

$$\hat{\epsilon}_t = t^{\frac{s}{2}} \eta_t - \frac{1}{T} \sum_{i=0}^T i^{\frac{s}{2}} \eta_i \quad (14)$$

$$\hat{\epsilon}_t^2 = t^s \eta_t^2 - 2t^{\frac{s}{2}} \eta_t \frac{1}{T} \sum_{i=1}^T i^{\frac{s}{2}} \eta_i + \left(\frac{1}{T} \sum_{i=1}^T i^{\frac{s}{2}} \eta_i \right)^2 \quad (15)$$

$$\hat{V} = \frac{1}{T^2} \sum_{t=1}^T t^s \eta_t^2 - \frac{1}{T} \left(\frac{1}{T} \sum_{t=1}^T t^{\frac{s}{2}} \eta_t \right)^2 \quad (16)$$

Now, note that the true variance is given by $\mathbb{V}[\hat{\theta}_{uw}] = \sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T t^s$. From Proposition 2.3, we know that

$$\left(\sigma_\eta^2 \frac{1}{T^2} \sum_{t=0}^T t^s \right)^{-\frac{1}{2}} \left(\frac{1}{T} \sum_{t=1}^T t^{\frac{s}{2}} \eta_t \right) \xrightarrow{d} \mathbb{N}(0, 1). \quad (17)$$

Thus

$$\left(\sigma_\eta^2 \frac{1}{T^2} \sum_{t=0}^T t^s \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T t^{\frac{s}{2}} \eta_t \right)^2 \xrightarrow{d} \chi^2, \quad (18)$$

and so $\frac{1}{T}$ times this $\xrightarrow{P} 0$ by Slutsky. I therefore only need to show that

$$\left(\mathbb{V}[\hat{\theta}_{uw}] \right)^{-1} \left(\frac{1}{T^2} \sum_{t=1}^T t^s \eta_t^2 \right) = \left(\sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T t^s \right)^{-1} \left(\frac{1}{T^2} \sum_{t=1}^T t^s \eta_t^2 \right) \xrightarrow{P} 1, \quad (19)$$

which I will do by showing that it has expectation 1 and variance converging to 0. Clearly,

$$\mathbb{E} \left[\left(\sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T t^s \right)^{-1} \left(\frac{1}{T^2} \sum_{t=1}^T t^s \eta_t^2 \right) \right] = 1. \quad (20)$$

Defining $\kappa_\eta^4 \equiv \sup_t \mathbb{V}[\eta_t^2]$, which exists by assumption, we also have that

$$\mathbb{V} \left[\frac{1}{T^2} \sum_{t=1}^T t^s \eta_t^2 \right] \leq \kappa_\eta^4 T^{-4} \sum_{t=1}^T t^{2s}, \quad (21)$$

and thus

$$\begin{aligned}
& \mathbb{V} \left[\left(\sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T t^s \right)^{-1} \left(\frac{1}{T^2} \sum_{t=1}^T t^s \eta_t^2 \right) \right] \leq \frac{\kappa_\eta^4}{\sigma_\eta^4} \left(\sum_{t=1}^T t^s \right)^{-2} \left(\sum_{t=1}^T t^{2s} \right) \\
& \leq \frac{\kappa_\eta^4}{\sigma_\eta^4} \left(\int_0^T t^s ds \right)^{-2} \left(\int_0^{T+1} t^{2s} ds \right) = \frac{\kappa_\eta^4}{\sigma_\eta^4} \frac{(s+1)^2}{2s+1} (T^{s+1})^{-2} (T+1)^{2s+1} \\
& = \frac{\kappa_\eta^4}{\sigma_\eta^4} \frac{(s+1)^2}{2s+1} \left(\frac{T+1}{T} \right)^{2s+1} T^{-1} \rightarrow 0,
\end{aligned} \tag{22}$$

where the inequality between summation and integration is because t^s is increasing in t . \square

3 Standard weighted parameter estimates

In empirical work, researchers often weight using population weighting. That is, they estimate

$$\hat{\theta}_{pw} \equiv \frac{1}{\sum_{t=1}^T \text{pop}_t} \sum_{t=1}^T \text{pop}_t y_t. \tag{23}$$

With the same assumptions as above on the structure of the error term, this becomes

$$\hat{\theta}_{pw} = \theta + \frac{1}{\sum_{t=1}^T t^{-s}} \sum_{t=1}^T (t^{-\frac{s}{2}} \eta_t + t^{-s} \nu_t). \tag{24}$$

As noted by [Solon et al. \(2015\)](#), there are many possible reasons to use weighting. One justification is that, if the errors are only due to measurement error (i.e., $\nu_t = 0$), then this estimator will be equivalent to GLS, and will thus be efficient. However, we still may not have consistency; and furthermore, the weighted estimate may not be asymptotically normal or have consistent standard errors.

Under reasonable regularity conditions, a necessary and sufficient condition for consistency is that $s \leq 1$.

Proposition 3.1. *Suppose $\hat{\theta}_{pw}$ is defined as in Equation 24, where η_t and ν_t are mean-zero, independently distributed, and homoskedastic with finite variances σ_η^2 and σ_ν^2 , respectively, at least one of which is non-zero. Then $\hat{\theta}_{pw} \xrightarrow{P} \theta$ if and only if $s \leq 1$.*

Proof. (Convergence if $s \leq 1$) Clearly, $\mathbb{E}[\hat{\theta}_{pw}] = \theta$. I will prove that $\mathbb{V}[\hat{\theta}_{pw}] \rightarrow 0$, which is sufficient. We

then have

$$\begin{aligned}
\mathbb{V}[\hat{\theta}_{pw}] &= \sigma_\eta^2 \left(\sum_{t=1}^T t^{-s} \right)^{-1} + \sigma_\nu^2 \left(\sum_{t=1}^T t^{-s} \right)^{-2} \sum_{t=1}^T t^{-2s} \\
&\leq \sigma_\eta^2 \left(\int_{t=1}^T t^{-s} dt \right)^{-1} + \sigma_\nu^2 \left(\int_{t=1}^T t^{-s} dt \right)^{-2} \left(1 + \int_{t=1}^T t^{-2s} dt \right) \\
&= \sigma_\eta^2 (1-s) (T^{1-s} - 1)^{-1} \\
&\quad + \sigma_\nu^2 (1-s)^2 (T^{1-s} - 1)^{-2} \left(1 + \frac{1}{1-2s} (T^{1-2s} - 1) \right) \\
&\rightarrow 0,
\end{aligned} \tag{25}$$

with a similar expression if $s = 1$ (in which case the integral for t^{-s} becomes a logarithm); or if $s = \frac{1}{2}$ (in which case the integral for t^{-2s} becomes a logarithm). \square

Proof. (Non-convergence if $s > 1$) Note that $\sum_{t=1}^T t^{-s} \rightarrow C$ for some constant C . Now, define $X_{lim} \equiv C^{-1} \lim_{T \rightarrow \infty} \sum_{t=1}^T (t^{-\frac{s}{2}} \eta_t + t^{-s} \nu_t)$. (In fact, X_{lim} will be a non-degenerate random variable with finite variance, but I need not prove that here.) We now have that $\hat{\theta}_{pw} \xrightarrow{p} \theta + C^{-1}(\eta_1 + \nu_1) + X_{lim}$. Note that $\mathbb{V}[C^{-1}(\eta_1 + \nu_1) + X_{lim}] = C^{-1}(\sigma_\eta^2 + \sigma_\nu^2) + \mathbb{V}[X_{lim}] > 0$, where no covariance is needed by independence. Thus $\hat{\theta}_{pw}$ does not converge to a constant. \square

Some caution should be taken in understanding the lack of convergence of the η_t term. This is because the η_t terms themselves often come from a measured variable, which might be quite accurate due to the law of large numbers. In fact, if the values of y_t are measured as described above, then the WLS estimate is numerically identical to the mean of all observations in all cities; so if the only error is idiosyncratic, rather than at the city level, then the law of large numbers may approximately apply to this error. Lack of convergence is therefore more interesting and useful for the ν_t term.

Despite the lack of convergence for the weighted estimator, it has some properties that are desirable. As shown in Figure 2, adding more observations always improves the accuracy (as measured by the variance of the estimator). In fact, if $\sigma_\eta^2 = 1$ and $\sigma_\nu^2 = 0$, then this weighted estimator is GLS, so $\hat{\theta}_{pw}$ is also the best linear unbiased estimator, in the mean-square sense. Note that this means that there does not exist a consistent (linear) estimator of θ .

In addition to the lack of convergence, the weighted estimator will often fail to be asymptotically normal. The following lemma, a partial converse of Lemma 2.2, will be useful in proving this.

Lemma 3.2. *Suppose $\{\epsilon_t\}$ are mean-zero, independently distributed, homoskedastic random variables with finite variance $\sigma_\epsilon^2 > 0$. Further, suppose there is some function $g(T)$ such that for all T , $g(T) \neq 0$; and a finite-valued function $f(t) > 0$ such that*

$$\lim_{T \rightarrow \infty} \frac{f(1)^2}{\sum_{t=1}^T f(t)^2} = C^2 \tag{26}$$

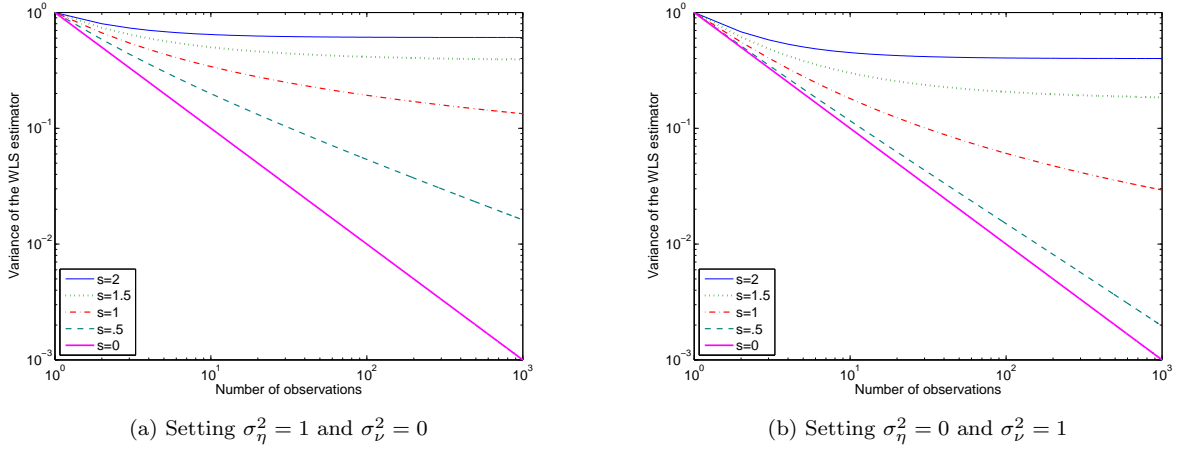


Figure 2: Variance of the WLS estimator: $\mathbb{V}[\hat{\theta}_{pw}] = \sigma_\eta^2 \left(\sum_{t=1}^T t^{-s} \right)^{-1} + \sigma_\nu^2 \left(\sum_{t=1}^T t^{-s} \right)^{-2} \sum_{t=1}^T t^{-2s}$. Variance is shown as sample size increases, with various values of s . Additional data is assumed to be added from most accurate to least accurate. If $s \leq 1$, the variance converges to 0; if $s > 1$, the variance converges to a constant.

for some finite constant $C > 0$. Define $X_{Tt} \equiv g(T)f(t)\epsilon_t$; $S_T \equiv \sum_{t=1}^T X_{Tt}$; and $s_T^2 \equiv \sum_{t=1}^T \mathbb{V}[X_{Tt}]$. Then $\frac{S_T}{s_T}$ is not generally asymptotically normal, in the sense that, for all t , holding fixed the distribution of $\{\epsilon_k\}$ for $k \neq t$, there is at most one distribution of ϵ_t for which $\frac{S_T}{s_T} \xrightarrow{d} \mathbb{N}(0, 1)$.

Proof. First, note that all $g(T)$ expressions cancel out from $\frac{S_T}{s_T}$, so they will be ignored. Next, note that $s_\infty \equiv \lim_{T \rightarrow \infty} s_T$ is finite, or else Equation 26 would not hold. Because of this, $\left(\frac{S_T}{s_T} - \frac{S_T}{s_\infty} \right) \xrightarrow{p} 0$ by Slutsky; so I will prove the above proposition about $\frac{S_T}{s_\infty}$. We now have that

$$\begin{aligned} \frac{S_T}{s_\infty} &= (s_\infty)^{-1} f(t)\epsilon_t + (s_\infty)^{-1} \sum_{k \neq t}^T f(k)\epsilon_k \\ &= C \frac{f(t)}{f(1)} \epsilon_t + (s_\infty)^{-1} \sum_{k \neq t}^T f(k)\epsilon_k. \end{aligned} \quad (27)$$

Define ϕ_A as the characteristic function of $(s_\infty)^{-1} \sum_{k \neq t}^T f(k)\epsilon_k$; ϕ_t as the characteristic function of ϵ_t ; and ϕ_N as the characteristic function of a standard normal random variable. If $\frac{S_T}{s_\infty} \xrightarrow{d} \mathbb{N}(0, 1)$, then ϕ_t is uniquely defined by

$$\phi_t(x) = \frac{\phi_N \left(\frac{f(1)}{Cf(t)} x \right)}{\phi_A \left(\frac{f(1)}{Cf(t)} x \right)}. \quad (28)$$

This may not be a valid characteristic function; if it is not, then there is no distribution of ϵ_t that would lead to asymptotic normality. \square

Essentially, this lemma says that convergence to a normal distribution will only occur under very specific conditions (for example, if all terms are already normal). The following proposition shows that this lemma applies to the weighted estimator for certain values of s .

Proposition 3.3. *Suppose $\hat{\theta}_{pw}$ is defined as in Equation 24, where $\{\eta_t\}$ and $\{\nu_t\}$ are mean-zero, independently distributed, uniformly integrable, and homoskedastic with finite variances σ_η^2 and σ_ν^2 , respectively, with at least one strictly greater than zero. If $s \leq \frac{1}{2}$, then $\hat{\theta}_{pw}$ is asymptotically normal (with suitable normalization). If $\frac{1}{2} < s \leq 1$, then $\hat{\theta}_{pw}$ is generally (in the sense of Lemma 3.2) asymptotically normal if and only if $\sigma_\eta^2 > 0$. If $s > 1$, then $\hat{\theta}_{pw}$ is generally not asymptotically normal in the sense of Lemma 3.2.*

Proof. Asymptotic normality of $\hat{\theta}_{pw}$ will depend on the normality of the sum $\sum_{t=1}^T (t^{-\frac{s}{2}}\eta_t + t^{-s}\nu_t)$, where the normalization term can be dropped for the same reason as the $g(T)$ term is dropped in Lemma 3.2. Combining Lemma 2.2 and Lemma 3.2, and because $\sup_{t \geq 1} t^{-s} = 1$, a necessary and sufficient condition for (general) normality of the η_t term is that $\sum_{t=1}^T t^{-s} \rightarrow \infty$, i.e. that $s \leq 1$; and for (general) normality of the ν_t term is that $\sum_{t=1}^T t^{-2s} \rightarrow \infty$, i.e. that $s \leq \frac{1}{2}$. Furthermore, if $\frac{1}{2} < s \leq 1$ and $\sigma_\eta^2 > 0$, then the variance of the η_t term goes to ∞ , while the variance of the ν_t term goes to a constant; thus, after normalization, the ν_t term disappears, and we are only left with the η_t term, which is asymptotically normal. \square

The extent to which the central limit theorem fails to apply in this case is shown in Figure 3. This figure shows excess kurtosis for various values of s . Excess kurtosis gives us a rough understanding of the weight on the tails of the distribution, which determines the extent to which p-values based on a normal approximation will be accurate (assuming accurate standard errors). When the central limit theorem applies, excess kurtosis will converge to 0, the value for a normal distribution. However, for $s = 1$, excess kurtosis of the ν_t term with an infinite amount of data is approximately equal to excess kurtosis where $s = 0$ (i.e., homoskedasticity) and 3 data points. Thus if we do not think that the sum of 3 homoskedastic variables will be sufficiently normal to perform robust inference, we should not think that the sum of an infinite number of $s = 1$ extreme heteroskedastic errors will lead to a sufficiently normal estimator.

One important note is that, although the distribution of the η_t term may fail to be generally normal, it may in fact be close to normal. This is due to the fact that the η_t terms themselves often come from a measured variable, which might be approximately normal due to the central limit theorem. (This is related to the note, above, that consistency for the η_t term is somewhat misleading.) As with consistency, the result of a lack of asymptotic normality is more interesting and useful for the ν_t term.

Of course, asymptotically, non-normality with $\frac{1}{2} < s \leq 1$ only occurs when $\sigma_\eta^2 = 0$, because the η_t term dominates the error otherwise. However, it is possible that σ_η^2 will be non-zero, but small enough that the $\{\nu_t\}$ terms will dominate the error for the finite number of observations we have. In this case, it is likely that the asymptotics based on $\sigma_\eta^2 = 0$ will come closest to approximating the true distribution.

A final question is whether standard errors will be consistent—that is, whether $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]} \xrightarrow{P} 1$, where \hat{V} is the standard heteroskedasticity-robust standard errors. Where the estimator is asymptotically normal, this is clearly an important question, as outlined above. If the estimator is non-normal, even perfect standard errors should be interpreted cautiously, as the usual interpretation of standard errors as indicating p-values is based on (asymptotic) normality. Still, correctly-estimated standard errors could at least give an idea of

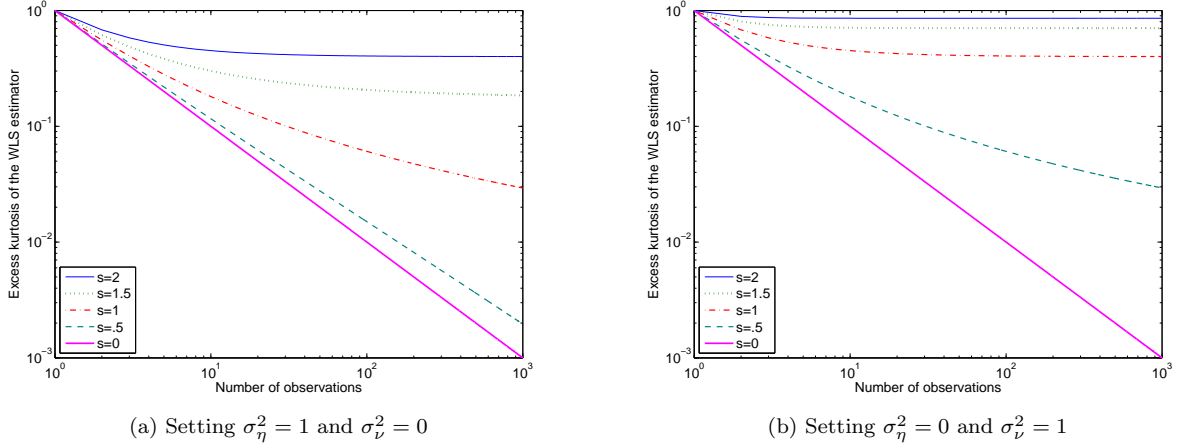


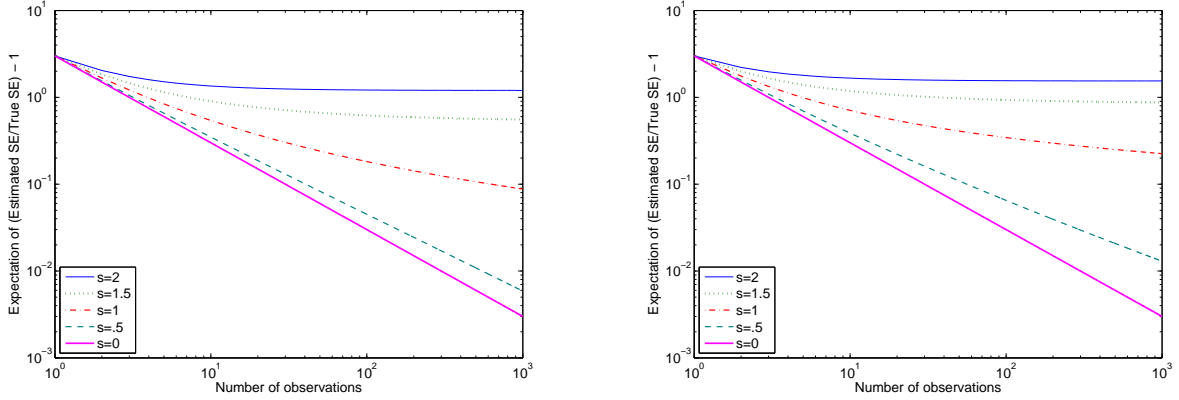
Figure 3: Excess kurtosis of the WLS estimator: $\mathbb{E} \left[\hat{\theta}_{pw}^4 \right] - 3 \left(\mathbb{V} \left[\hat{\theta}_{pw} \right] \right)^2$. Excess kurtosis is shown as sample size increases, with various values of s . Additional data is assumed to be added from most accurate to least accurate. When the central limit theorem applies, excess kurtosis converges to 0, the value for a normal distribution. In both graphs, excess kurtosis of each η or ν is assumed to be 1. For panel (a), the η_t term, excess kurtosis converges to 0 if and only if $s \leq 1$; for panel (b), the ν_t term, excess kurtosis converges to 0 if and only if $s \leq \frac{1}{2}$. In all other cases, excess kurtosis converges to a constant.

the dispersion of the distribution of the estimator.

In fact, standard errors may not be consistent either. The true variance of the weighted estimator is shown in the first line of Equation 25. To estimate the variance, however, the standard equations lead to

$$\begin{aligned}
\hat{\epsilon}_t &\equiv Y_t - \hat{\theta}_{pw} = t^{\frac{s}{2}} \eta_t + \nu_t - \left(\sum_{k=1}^T k^{-s} \right)^{-1} \sum_{k=1}^T (k^{-\frac{s}{2}} \eta_k + k^{-s} \nu_k) \\
\hat{\epsilon}_t^2 &= (t^{\frac{s}{2}} \eta_t + \nu_t)^2 + 2 (t^{\frac{s}{2}} \eta_t + \nu_t) \left(\sum_{k=1}^T k^{-s} \right)^{-1} \sum_{k=1}^T (k^{-\frac{s}{2}} \eta_k + k^{-s} \nu_k) \\
&\quad + \left(\sum_{k=1}^T k^{-s} \right)^{-2} \left(\sum_{k=1}^T (k^{-\frac{s}{2}} \eta_k + k^{-s} \nu_k) \right)^2 \\
\hat{V} &\equiv \left(\sum_{t=1}^T t^{-s} \right)^{-2} \sum_{t=1}^T t^{-2s} \hat{\epsilon}_t^2 \\
&= \left(1 + \left(\sum_{t=1}^T t^{-s} \right)^{-2} \left(\sum_{t=1}^T t^{-2s} \right) \right) \left(\sum_{t=1}^T t^{-s} \right)^{-2} \sum_{t=1}^T (t^{-\frac{s}{2}} \eta_t + t^{-s} \nu_t)^2 \\
&\quad + 2 \left(\sum_{t=1}^T t^{-s} \right)^{-3} \left(\sum_{t=1}^T (t^{-\frac{s}{2}} \eta_t + t^{-s} \nu_t) \right) \left(\sum_{t=1}^T (t^{-\frac{3s}{2}} \eta_t + t^{-2s} \nu_t) \right) \tag{29}
\end{aligned}$$

The random variable $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]}$ will have a complicated distribution even for seemingly simple $\{\eta_t\}$ and $\{\nu_t\}$.



(a) Expectation of standard errors of the WLS estimator, setting $\sigma_\eta^2 = 1$ and $\sigma_\nu^2 = 0$.

(b) Expectation of standard errors of the WLS estimator, setting $\sigma_\eta^2 = 0$ and $\sigma_\nu^2 = 1$.

Figure 4: Expectation of standard errors of the WLS estimator, minus 1: $\mathbb{E} \left[\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]} \right] - 1$. Expectation is shown as sample size increases, with various values of s . Additional data is assumed to be added from most accurate to least accurate. In both panels, expectation converges to 1 if and only if $s \leq 1$. In all other cases, the expectation converges to a constant greater than 1.

Defining $H_n(T) \equiv \sum_{t=1}^T t^n$ (and suppressing the argument for simplicity), we have that

$$\begin{aligned} \mathbb{E} \left[\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]} \right] &= 1 + H_{-s}^{-2} H_{-2s} \\ &\quad + 2H_{-s}^{-1} (\sigma_\eta^2 H_{-s} + \sigma_\nu^2 H_{-2s})^{-1} (\sigma_\eta^2 H_{-2s} + \sigma_\nu^2 H_{-3s}). \end{aligned} \quad (30)$$

This expectation, for various sample sizes and with different values of s , is shown in Figure 4. The expected value only converges to 1 if $s \leq 1$. Future work involves proving that the variance of $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]}$ is bounded by assuming finite fourth moments, and that the variance goes to 0 if $s \leq 1$, which should prove that $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]} \xrightarrow{P} 1$ if and only if $s \leq 1$, but that will require a lot of algebra. I will do this at a later date.

Interestingly, because $H_n > 0$ for all n , this expectation will be greater than 1 for all s . Thus standard errors will, on average, overestimate the variance. However, because $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]}$ is a random variable with nonzero dispersion even in the limit, standard errors may still understate the variance often enough to cause problems. Simulations, to be completed later, will demonstrate when this can occur.

4 Alternative asymptotics

Suppose we now assume that the least accurate observation's variance is held fixed. In other words, the error term of Equation 5 will be rewritten as

$$\epsilon_t = \left(\frac{\text{pop}_t}{\text{pop}_T} \right)^{-\frac{1}{2}} \eta_t + \nu_t = \left(\frac{t}{T} \right)^{\frac{s}{2}} \eta_t + \nu_t, \quad (31)$$

This formulation is a valid description of the data; for any given sample size, it is equivalent to Equation 5, but with a different value for σ_η^2 .

In this formulation, the unweighted estimator will be consistent. The unweighted estimator is now

$$\hat{\theta}_{uw} \equiv \frac{1}{T} \sum_{t=1}^T y_t = \theta + \frac{1}{T} \sum_{t=1}^T \left(\left(\frac{t}{T} \right)^{\frac{s}{2}} \eta_t + \nu_t \right) \quad (32)$$

To see that this is consistent, note that the variance of the (unbiased) estimator is given by

$$\begin{aligned} \mathbb{V}[\hat{\theta}_{uw}] &= \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{\frac{s}{2}} \eta_t + \nu_t \right)^2 \right] = \sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T \left(\frac{t}{T} \right)^s + \frac{1}{T} \sigma_\nu^2 \\ &\leq \sigma_\eta^2 \frac{1}{T^2} \sum_{t=1}^T 1 + \frac{1}{T} \sigma_\nu^2 = \frac{1}{T} (\sigma_\eta^2 + \sigma_\nu^2) \rightarrow 0. \end{aligned} \quad (33)$$

However, it remains true that, for $s > 1$, the unweighted estimator would be more efficient if only the most accurate observation is used. It is simply that each observation becomes arbitrarily accurate as sample size increases.

The weighted estimator is now given by

$$\begin{aligned} \hat{\theta}_{pw} &\equiv \frac{1}{\sum_{t=1}^T \left(\frac{t}{T} \right)^{-s}} \sum_{t=1}^T \left(\frac{t}{T} \right)^{-s} y_t = \frac{1}{\sum_{t=1}^T t^{-s}} \sum_{t=1}^T t^{-s} y_t \\ &= \theta + \frac{1}{\sum_{t=1}^T t^{-s}} \sum_{t=1}^T \left(\frac{1}{T^{\frac{s}{2}}} t^{-\frac{s}{2}} \eta_t + t^{-s} \nu_t \right) \end{aligned} \quad (34)$$

Note that the ν_t term in this expression is the same as in Equation 24; thus, all results from above for the ν_t term in the weighted estimator will remain valid. Focusing instead on the η_t term, we see that it converges; if $\sigma_\nu^2 = 0$, then

$$\mathbb{V}[\hat{\theta}_{pw}] = T^{-s} \sigma_\eta^2 \left(\sum_{t=1}^T t^{-s} \right)^{-1} \rightarrow 0 \quad (35)$$

$$(36)$$

because $T^{-s} \rightarrow 0$ for all $s > 0$, and $\left(\sum_{t=1}^T t^{-s} \right)^{-1}$ is bounded (and goes to 0 if $s = 0$).

Results for asymptotic normality of the η_t term will not change. The extra $T^{\frac{s}{2}}$ will function as the $g(T)$ term in Lemma 2.2 and Lemma 3.2, and therefore will not affect the conclusions.

Results on standard errors for the η_t term will also not be affected, as the extra $T^{\frac{s}{2}}$ will factor out of $\frac{\hat{V}}{\mathbb{V}[\hat{\theta}_{pw}]}$ if $\nu_t = 0$.

There are, of course, infinite other asymptotic assumptions that agree with any finite amount of data, many of which would lead to consistent, asymptotically normal estimators with consistent standard errors.

For example, if we assume that the errors in all remaining observations will be homoskedastic, then of course the usual assumptions apply and estimators are consistent, asymptotically normal, and have consistent standard errors. Simulations may be the best guide to understanding which asymptotic assumptions come closest to approximating the truth. In addition, the figures in this paper are all based on small-sample statistics; so the conclusions reached by looking at them remain valid for any asymptotic assumptions.

5 Potential solutions

5.1 Imperfect solutions

One intuitively attractive option would be to use a resampling method (such as bootstrapping) to estimate confidence intervals. Using resampling methods to estimate standard errors would not be useful; the problem here is not inconsistently estimated standard errors, but the fact that standard errors are not meaningful. Resampling to estimate confidence intervals, however, would not be a valid resolution to this problem. The validity of resampling is based on the assumption that a sample of the population has similar properties to the whole. In these cases, however, a very small number of observations are leading to a large amount of the variance; thus the behavior of an estimator in a sample may depend strongly on the composition of that sample.

A second potential solution would be to estimate the variance of the error terms, then estimate the result with FGLS. This solution is similar to the one proposed below, and is generally valid if the initial estimator is consistent. A standard method of estimating the variance of the error terms would be to use OLS, noting that

$$\mathbb{E} \left[\left(t^{\frac{s}{2}} \eta_t + \nu_t \right)^2 \right] = t^s \sigma_\eta^2 + \sigma_\nu^2, \quad (37)$$

where $\sigma_\eta \equiv \mathbb{E} [\eta^2]$ and $\sigma_\nu \equiv \mathbb{E} [\nu^2]$. We might then naively regress the initial residual squared (assumed to be consistent for the error squared) on a constant and t^s . However, such a regression would not always be consistent, for two reasons. The first is that one of the regressors grows unboundedly, so that the $X'X$ matrix is not invertible in the limit, so standard asymptotic assumptions do not apply. The second reason is that the error terms of this regression may also display extreme heteroskedasticity for many values of s .

5.2 QML procedure

One potential new estimator of the parameter of interest and of the variance of the error terms that may have better properties than OLS or WLS (and, indeed, may be optimal in some sense). We can estimate $(\hat{\theta}, \hat{\sigma}_\eta^2, \hat{\sigma}_\nu^2)$ as those values that maximize the log-likelihood, assuming normality:

$$\mathcal{L} = \sum_{t=1}^T \left[-\frac{1}{2} \log (\text{pop}_t^{-1} \hat{\sigma}_\eta^2 + \hat{\sigma}_\nu^2) - \frac{1}{2} \frac{(y_t - \hat{\theta})^2}{\text{pop}_t^{-1} \hat{\sigma}_\eta^2 + \hat{\sigma}_\nu^2} \right] \quad (38)$$

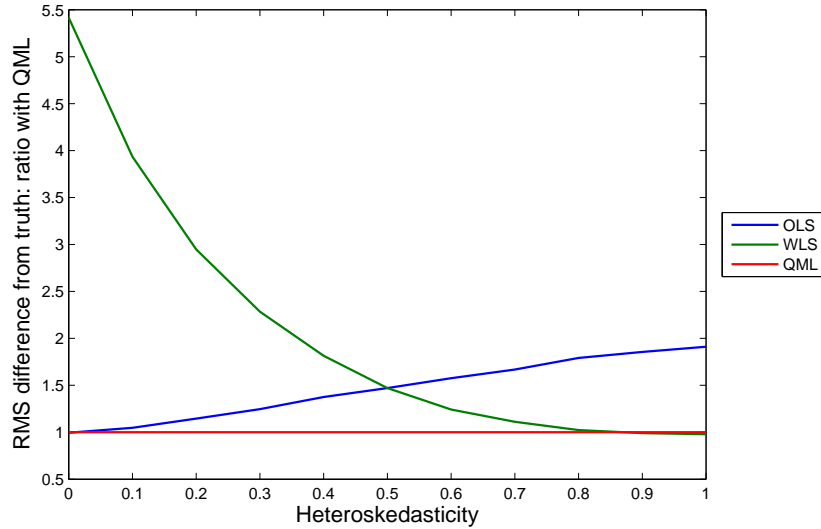


Figure 5: Each graph shows the root mean squared error of estimators relative to QML for $s = 1$. Each point includes 10,000 simulations of an estimate based 1,000 observations where η_t is Gaussian and ν_t is exponential. Heteroskedasticity is defined by $h = \Phi\left(\frac{1}{2} \log_2\left(\frac{\sigma_\eta/\sigma_\nu}{R}\right)\right)$, where R^2 is the variance ratio that would cause $\mathbb{V}\left[\hat{\theta}_{OLS}\right] = \mathbb{V}\left[\hat{\theta}_{WLS}\right]$, and $\Phi(\cdot)$ is the CDF of a normal distribution.

Based on simulations, this estimator performs better than OLS or WLS under a range of values for the nuisance parameter. Figure 5 shows the root mean squared error of OLS and WLS relative to QML under a wide range of values for heteroskedasticity, while Figure 6 shows the actual size of estimated 95% confidence intervals, where those intervals are estimated with a variety of techniques. However, there is no reason to believe that it is consistent, and more work is still needed to determine how to use this estimator to perform accurate inference. Solving this problem (or, indeed, creating any procedure that can generate valid inference) is the main theoretical issue still needed in this project.

5.3 Using the original data to estimate heteroskedasticity

Another potential solution may be possible if more data is available. In particular, consider the original example, where each observation represents a city. If we have the individual-level data that was used to estimate city-wide averages, we could potentially estimate the variance on the city-wide estimates, thus allowing us to estimate σ_η/σ_ν and perform consistent inference. More work is needed, though, to show the properties of such an estimator. More work is also needed in understanding the way that power laws can affect estimation based on observation-level data.

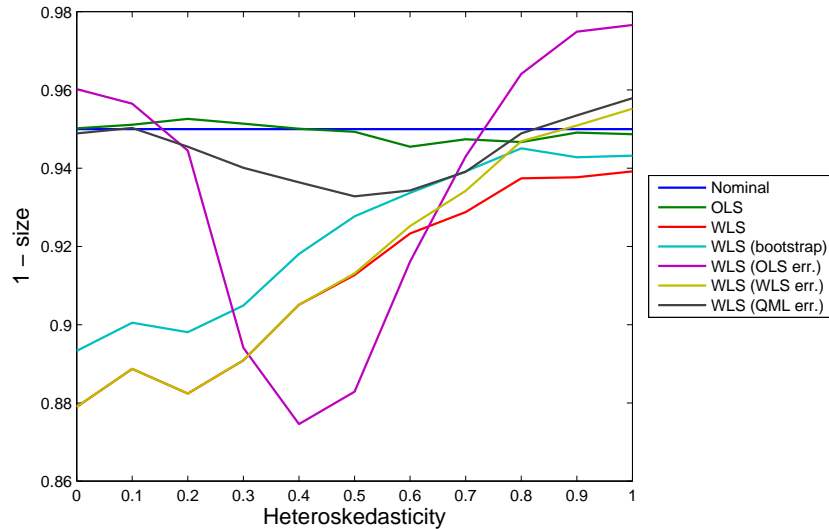


Figure 6: Each graph shows the size of nominally 95% confidence intervals for $s = 1$, using different estimation techniques. For WLS, standard errors can be estimated using the standard method, or with a bootstrap technique, or by estimating σ_η/σ_ν with OLS, WLS, or the QML procedure presented here. Each point includes 10,000 simulations of an estimate based 1,000 observations where η_t is Gaussian and ν_t is exponential. Heteroskedasticity is defined by $h = \Phi\left(\frac{1}{2} \log_2\left(\frac{\sigma_\eta/\sigma_\nu}{R}\right)\right)$, where R^2 is the variance ratio that would cause $\mathbb{V}[\hat{\theta}_{OLS}] = \mathbb{V}[\hat{\theta}_{WLS}]$, and $\Phi(\cdot)$ is the CDF of a normal distribution.

6 Concluding remarks

Power laws occur frequently in economic applications. The size of firms and cities approximate obey power laws, and many other systems do as well. If accuracy of observations is related to size—or if the econometrician estimates parameters assuming it is—then the variance of observations will also obey a power law. Traditional least squares tools can lead to inconsistent estimators and unreliable inference if this extreme heteroskedasticity is present.

In practice, this may be a particular concern for studies that use observations at the level of cities or firms, where size obeys Zipf’s law (that is, a power law with $s = 1$). In this case, OLS estimates will not generally converge to the truth; instead, the variance of the estimator converges to a non-zero constant. On the other hand, the WLS estimator will converge to the parameter of interest, but slowly, because $s = 1$ is an edge case: if $s = 1 + \epsilon$ for any $\epsilon > 0$, then the WLS estimator does not converge, even for $\epsilon \rightarrow 0$. Additionally, WLS for $s = 1$ can cause particular problems with inference if the actual heteroskedasticity is small; for example, if we weight by city size, but estimates are just as accurate for small as large cities (this can occur if estimates are not based on a survey where each person is equally likely to appear). In this case, asymptotic normality no longer holds in general. In fact, estimated standard errors are only consistent for the true standard deviation of the estimator as an edge case; it is not true if $s > 1$. Because of this, estimated confidence intervals will no longer have the correct size.

Two procedures are outlined that may be able to address this problem: a QML procedure, and a procedure

using the original data to estimate the extent of the heteroskedasticity. More work is needed, though, to understand the properties of these estimators. Further research is also needed to understand where extreme heteroskedasticity may have lead to incorrect inference in real-world studies. As researchers find more examples of systems that obey power laws across economic domains, it is important to account for extreme heteroskedasticity in estimates and inferences based on these systems.

References

- Axtell, Robert L. Zipf Distribution of U.S. Firm Sizes. *Science*, 293:1818–1820, September 2001.
- Chung, Kai Lai. *A Course in Probability Theory*. Academic Press, second edition, 1974.
- Eeckhout, Jan. Gibrat’s Law for (All) Cities. *American Economic Review*, 94(5):1429–1451, 2004.
- Gabaix, Xavier. Power Laws in Economics and Finance. *Annual Review of Economics*, 1(1):255–294, 2009.
- Gabaix, Xavier. Power Laws in Economics: An Introduction. *Journal of Economic Perspectives*, 30(1): 185–206, 2016.
- Rozenfeld, Hernan D., Diego Rybski, Xavier Gabaix, and Hernan A. Makse. The Area and Population of Cities: New Insights from a Different Perspective on Cities. *American Economic Review*, 101(5):2205–2225, 2011.
- Solon, Gary, Steven J. Haider, and Jeffrey Wooldridge. What Are We Weighting For? *Journal of Human Resources*, 50(2):301–316, 2015.