ε	synthetic-RMSE		realworld-RMSE	
	CGLSE	Uni	CGLSE	Uni
0.1	**.005**	.007	**.008**	.014
0.2	**.011**	.017	**.021**	.029
0.3	**.022**	.026	**.065**	.073
0.4	**.032**	.056	**.116**	.136
0.5	**.051**	.099	**.167**	.207

Table 1: existing simulations - RMSE

ε	max. emp. err.		avg. (std.) of emp. err.		RMSE		size
	CGLSE	Uni	CGLSE	Uni	CGLSE	Uni	
0.1	.029	1.574	.026(.002)	.984(.305)	**.026**	1.030	68220
0.2	.042	1.465	.015(.011)	1.066(.206)	**.019**	1.085	10977
0.3	.046	.998	.023(.012)	.976(.012)	**.026**	.976	3356
0.4	.088	.997	.055(.017)	.995(.057)	**.057**	.995	1428
0.5	.071	.996	.067(.003)	.985(.067)	**.067**	.985	741

Table 2: new simulations - synthetic data (Cauchy)

We sincerely thank all the reviewers for their insightful comments. We take these comments seriously and address them below. This paper was first submitted to ICML 2020. The paper got **one accept and two weak accepts**. The main comments were to focus on a single regression problem (GLSE) in the main body of the paper and include empirical results with real world data – we did both. In doing so – we moved GLSE$_k$ to the appendix – this seems to have diluted the technical novelty of the current paper for some of the current reviewers. While this is disappointing, we fully understand – we can easily bring back the emphasis on GLSE$_k$ where there is significant technical novelty.

Technical novelty for GLSE$_k$. Novelty 1: Coreset definition. The first difficulty is that, unlike GLSE, due to the $\min$ operation, the objective function $\psi^{(G,q,k)}$ of GLSE$_k$ can only be decomposed into sub-functions $\psi_i^{(G,q,k)}$ instead of individual-time pairs. We address this by incorporating $\min$ operations in the computation function $\psi_S^{(G,q,k)}$ over the coreset S. The second difficulty is that the clustering centers are *subspaces* induced by regression vectors $\beta^{(1)}, \ldots, \beta^{(k)}$, instead of *points* as in Gaussian mixture models or k-means. So it is unclear how GLSE$_k$ can be reduced to projective clustering used in Gaussian mixture models (see also [Feldman et al., 2019, Coresets for Gaussian mixture models of any shape]). We address this by treating observation vectors of an individual $(x_{i1}, \ldots, x_{iT})$ as one entity while constructing coresets. **Novelty 2: Coreset construction/upper bounding total sensitivity.** This involves two steps. In **Step 1**, we reduce the sensitivity function from GLSE$_k$ to OLSE$_k$ (Lemma D.10), based on two observations: for any $\zeta = (\beta, \rho) \in \mathcal{P}_\lambda$ (recall that $\mathcal{P}_\lambda = \mathbb{R}^d \times B^q_{1-\lambda}$ for some constant $\lambda \in (0,1)$ where $B^q_{1-\lambda} = \left\{\rho \in \mathbb{R}^q : \|\rho\|_2^2 \le 1-\lambda\right\}$) we have $\psi_i^{(G,q)}(\zeta) \ge \lambda \cdot \psi_i^{(O)}(\beta)$ that provides an upper bound of the individual objective gap between GLSE and OLSE, and for any $\zeta = (\beta^{(1)}, \ldots, \beta^{(k)}, \rho^{(1)}, \ldots, \rho^{(k)}) \in \mathcal{P}^k$, $\psi_i^{(G,q,k)}(\zeta) \le 2(q+1) \cdot \min_{l\in[k]} \psi_i^{(O)}(\beta^{(l)})$; and for any $\zeta = (\beta^{(1)}, \ldots, \beta^{(k)}, \rho^{(1)}, \ldots, \rho^{(k)}) \in \mathcal{P}^k_\lambda$, $\psi_i^{(G,q,k)}(\zeta) \le 2(q+1) \cdot \min_{l\in[k]} \psi_i^{(O)}(\beta^{(l)})$, that provides a lower bound of the individual objective gap between GLSE$_k$ and OLSE$_k$. **Step 2** upper bounds the total sensitivity of OLSE$_k$. This key step for coreset construction (Lines 3-4 in Algorithm 2) is done by showing that the max. influence of individual i is at most $\frac{u_i}{u_i + \sum_{j\ne i} \ell_j}$ where u_i is the largest eigenvalue of $(Z^{(i)})^\top Z^{(i)}$ and ℓ_j is the smallest eigenvalue of $(Z^{(j)})^\top Z^{(j)}$, where $Z^{(i)} \in \mathbb{R}^{T\times(d+1)}$ is the matrix whose t-th row is $z_t^{(i)} = (x_{it}, y_{it}) \in \mathbb{R}^{d+1}$ (Definition D.3 and Lemma D.9).

Empirical performance of our coresets. Reviewers note weak performance of coresets relative to uniform on average error. As noted by R1, uniform may have lower average error, but has higher std (more large errors), i.e., many observations may be poorly represented. Std is also higher in real data, where errors may not be "regular" as in Gaussian noise. To illustrate this, we will include RMSE (root mean square error)–a standard metric of performance. Given a set of errors $e_1, \ldots, e_n$, **RMSE** $:= \sqrt{\frac{1}{n}\sum_{i\in[n]} e_i^2}$. In Table 1, RMSE(coresets) < RMSE(uniform) in both datasets—60%-90% of uniform on real data and 50%-85% of uniform on synthetic data with Gaussian errors. Further, unlike coresets, uniform also has no performance bounds on max error. For instance, in the real-world data with $\varepsilon = 0.5$, the max. error of uniform is .775 which exceeds the desired error bound. R1 asks about role of leverage. Coresets should perform better with high leverage observations, given max. error guarantees. We earlier presented Gaussian error (low leverage, few outliers) because it is a "hard" benchmark to beat. We illustrate this by replacing Gaussian errors in Eq. (2) with the **Cauchy** (0,2) distribution that has heavy tails as in $e_{it} = \sum_{a=1}^{\min\{t-1,q\}} \rho_a e_{i,t-a} + \textbf{Cauchy}(0,2)$, Table 2 shows that coreset performance relative to uniform is now even better for max/avg/RMSE errors. The max error for uniform exceeds the desired bound for all values of ε; and is at least 10x that of our coreset. In summary, we greatly appreciate the issues raised. We note that the issues raised (performance on real data, outliers/leverage points) strengthen evidence in favor of our coreset. We will add these points to the final version.

R1. Thank you for your constructive feedback. The final version will account for your expositional suggestions. Hopefully we have addressed your concerns and we hope you will support our paper.

R2. Thank you for appreciating our paper, providing positive feedback, and supporting it. The code is already on github (not included for anonymity); a link will be added in final version.

R3. Thanks for appreciating the novelty of our GLSE result. We hope the above discussion on empirical results and technical novelty (for GLSE$_k$) addresses your concerns. We hope you will strengthen support for our paper.

R4. Thank you for your detailed feedback. We clarify that our GLSE coreset with AR(q) **works for any** $q' \le q$ and any $\rho \in B^q$. On the real-world use of GLSE$_k$, it is a basic problem with applications in many fields; as accounting for *unobserved heterogeneity* in panel regressions is critical for unbiased estimates. See, Arellano, M. (2003). *Panel data econometrics*. Halaby, C. N. (2004). *Panel models in sociological research*. Annual Review of Sociology. We will add a discussion on the importance of GLSE$_k$. We will add boxplots in the full version. We hope you will support our paper.