An Application of Hierarchical Clustering for Categorizing Producing Areas and Mitigating Variance Inflation Issues in Multivariate Predictive Analytics for the Eagle Ford Shale, Texas

Andrew C. Silver

Adret LLC

ABSTRACT

An enhanced understanding of production variables in the Eagle Ford Shale of the Texas Gulf Coast was developed using statistical modelling techniques. In building predictive multivariate models, variance inflation issues may be encountered with both autocorrelation and multicollinearity. To account for spatial variation, a surrogate for field designations was developed to overcome issues which otherwise would have undermined the usefulness of field names as a variable.

Field names registered with regulatory agencies may at times correspond to attributes at the surface instead of to attributes of the reservoirs being produced. This issue is exacerbated in unconventional plays which have large aerial extents. Therefore, an approach for building a field surrogate variable was adopted utilizing hierarchical clustering on a set of variables which related to rock quality, stratigraphic thickness, and horizontal well orientation. The resulting clusters served as reservoir-based field-name surrogates.

To choose the number of clusters used, cluster branch depth was increased until they delineated areas known to have differences based on first-hand experience. Excess clusters were then combined when they overlapped each other geographically. Some wells were reassigned to match the clusters dominating their proximity. Lastly, some of the larger clusters were arbitrarily split perpendicular to their geographical trends so that each cluster would have reasonably similar populations and hence similar variance representations.

The resulting clusters were able to represent much of the spatial variation in the Eagle Ford trend and did so without incurring strong multicollinearity issues. These clusters also handled the spatial variation without in-

Silver, A. C., 2019, An application of hierarchical clustering for categorizing producing areas and mitigating variance inflation issues in multivariate predictive analytics for the Eagle Ford Shale, Texas: GeoGulf Transactions, v. 69, p. 623–624.

troducing spatial autocorrelation effects which might otherwise have invalidated the modeling outputs.

The clusters substantially improved the stability and predictive capabilities of a multivariate regression model. The model was then used in an explanatory capacity to interpret the impact on production of specific geologic and engineering variables of interest.