The high level of participation by demand response sets ERCOT apart from other operating electricity markets. Figure 51 shows that the amount of responsive reserves provided by LaaRs gradually increased from about 900 MW at the beginning of 2004 to an average of 1,147 MW in 2006. The majority of this increase was procured through self-provision and bilateral agreements rather than the ERCOT administered auction. Currently, LaaRs are permitted to supply up to 1,150 MW of the responsive reserves requirement. In 2005 and 2006, it became commonplace for the 1,150 MW restriction to limit the set of demand resources that could provide responsive reserves. This has highlighted a flaw with the way that the ancillary services auction selects demand resources to provide responsive reserves.

The auction ranks responsive reserves providers according to their offer price from lowest to highest.[27] The auction goes up the offer stack until it reaches the 2,300 MW required quantity of reserves. However, if the auction reaches the 1,150 MW limit before meeting the 2,300 MW requirement, the offers of any additional LaaRs cannot be used and are discarded. In such cases, the marginal generator resource sets the clearing price for responsive reserves at a level that exceeds the offer prices of some of the unaccepted offers from LaaRs.

This mechanism for selecting providers and determining clearing prices for responsive reserves is inefficient and leads to excessive reliability costs for consumers. Routinely, the quantity of LaaRs willing to supply responsive reserves at the clearing price exceeds the demand for this service (*i.e.*, 1,150 MW). When supply exceeds demand for a product at the prevailing price, it should cause the price of the product to decrease until the market reaches a level where the supply equals demand. Under the current market design, there is no mechanism for this to happen since there is only one price for all responsive reserves. Since ERCOT limits the amount of responsive reserves that can be provided by LaaRs, the price of reserves provided by LaaRs should clear below the price of reserves provided by synchronized generators.

---

27 In October 2005, ERCOT began to use a simultaneous clearing model for regulation up, regulation down, responsive reserves, and non-spinning reserves. This selection mechanism is conceptually similar since resources are selected in merit order. However, a resource with a low-priced responsive reserves offer may be selected to provide another product, such as regulation up, if the reduced cost of the other product exceeds the added cost of not using the resource to provide responsive reserves. In this case, the clearing price for responsive reserves is the marginal cost to the system of meeting the reserves requirement. This is always equal to the marginal reserves provider's offer price plus the opportunity cost of not providing an alternate product in the auction.

The design of this market encourages inefficient behavior by QSEs that want to sell responsive reserves from their demand resources. Under current market conditions, the clearing price for responsive reserves is usually set by a generator. In order to be selected, it is not sufficient for LaaRs to submit an offer price that is below the clearing price. The LaaR's offer must also be included among the lowest priced 1,150 MW of LaaRs. This gives QSEs an incentive to offer LaaRs at arbitrarily low (even negative) prices. Under these incentives, competition does not lead to having the most efficient resources provide responsive reserves. This also raises the concern that a negative LaaR offer could set the responsive reserves clearing price in the event that 1,150 MW of generators are bilaterally scheduled for reserves. In this unlikely event, LaaRs might receive large invoices to provide reserves, raising potential credit issues.

To improve the efficiency of responsive reserves pricing and incentives for suppliers, we recommend that ERCOT set separate prices for the two types of responsive reserves. The best way to accomplish this would be by having two responsive reserves constraints in the ancillary services auction: (i) that the responsive reserves procurement (including bilateral schedules) be greater than or equal to 2,300 MW and (ii) that the responsive reserves procurement from LaaRs (including bilateral schedules) be less than or equal to 1,150 MW. The clearing price paid to generators would be equal to the shadow price of the first constraint only, while the clearing price paid to LaaRs would be equal to the shadow price of the first constraint minus the shadow price of the second constraint.

Under this proposal, whenever the 1,150 MW limit on LaaRs providing responsive reserves was binding, the clearing price for responsive reserves from LaaRs would be determined by the offer of the marginal LaaR. Whenever the 1,150 MW limit did not affect the selection of resources (*i.e.*, the shadow price of the second constraint equals $0), the clearing prices would be identical for both types of responsive reserves providers. This recommendation would likely require some slight changes to the ancillary services market clearing engine software.

ERCOT stakeholders considered this change in 2006 and, due to resource constraints, decided not to implement it in the current market and instead drafted a protocol revision to implement it in the nodal market. However, this protocol revision failed to receive the necessary two-thirds vote at the ERCOT Technical Advisory Committee in 2007; thus, there is currently no plan to

implement any of the changes described above for the RRS market. As previously discussed, the current mechanism for selecting providers and determining clearing prices for responsive reserves is inefficient and leads to excessive reliability costs for consumers. Therefore, we recommend that these changes be reconsidered for implementation in the nodal market design.

Although LaaRs are active participants in the responsive reserves market, they did not offer into the balancing energy or regulation services markets and their participation in the non-spinning reserves market averaged only 14 MW in 2006. This is not surprising because the value of curtailed load tends to be very high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, providing non-spinning reserves introduces a much higher probability of being curtailed. Participation in the regulation services market requires technical abilities that most LaaRs cannot meet at this point. Finally, prices in the balancing energy market have not been high enough to attract active load participation in that market. Hence, most LaaRs will have a strong preference for providing responsive reserves over regulation services, non-spinning reserves, or balancing energy.