

You've identified a key modern approach! Using explainability tools like SHAP/LIME with complex models is indeed a popular strategy, but it's not a perfect substitute for assumption checking. Let me explain why:

Yes, you can often skip linear assumptions with this approach

SHAP/LIME do provide:

- Feature importance rankings
- Local explanations for individual predictions
- Global model behavior insights
- Stakeholder-friendly visualizations

This makes complex models much more interpretable than before.

But there are important limitations

Explainability \neq Statistical Validity:

1. SHAP/LIME explain the model, not the data relationship

- They tell you what the model learned, not whether it learned correctly
- A model can be explainable but still fundamentally wrong

2. Assumption violations still cause problems:

- **Multicollinearity** → Unstable feature importances in SHAP
- **Non-linear relationships** → Linear models miss patterns even with SHAP
- **Heteroscedasticity** → Prediction intervals are unreliable
- **Outliers** → Can dominate SHAP explanations misleadingly

3. Different types of interpretability:

- **Model interpretability** (SHAP/LIME): "How does this black box work?"
- **Statistical interpretability** (linear regression): "What's the true relationship in the data?"

Real-world practical considerations

When SHAP/LIME approach works well:

- Prediction accuracy is the primary goal
- Stakeholders need explanations but not causal inference
- You have large datasets where assumption violations matter less

- Complex patterns exist that linear models can't capture

When assumption checking is still valuable:

- **Causal inference** - Understanding true relationships, not just predictions
- **Small datasets** - Assumption violations have bigger impact
- **Regulatory compliance** - Some industries require traditional statistical validation
- **Model stability** - Ensuring consistent behavior across different data conditions
- **Confidence intervals** - When you need reliable uncertainty quantification

Hybrid approach (often best in practice):

1. **Start with complex models + SHAP/LIME** for performance and initial insights
2. **Validate key findings** with simpler models where assumptions can be checked
3. **Use assumption checking** for the relationships that matter most to business decisions
4. **Deploy the complex model** but monitor it using insights from both approaches

Bottom line:

SHAP/LIME are powerful tools that can reduce the need for assumption checking in many scenarios, but they solve a different problem. If you only need predictive performance with some explainability, you can often skip assumptions. But if you need to understand the true underlying relationships or ensure statistical rigor, assumption checking remains valuable.

The choice depends on whether you're building a **prediction system** (SHAP/LIME sufficient) or doing **statistical inference** (assumptions still matter)