Econometric Theory, **0**, 2026, 1–31.
doi:10.1017/S0266466626100450

HIGH-DIMENSIONAL NEWEY–POWELL TEST VIA APPROXIMATE MESSAGE PASSING

JING ZHOU
University of Manchester

HUI ZOU
University of Minnesota

We propose a high-dimensional extension of the heteroscedasticity test proposed in Newey and Powell (1987). Our test is based on expectile regression in the proportional asymptotic regime where $n/p \to \delta \in (0, 1]$. The asymptotic analysis of the test statistic uses the approximate message passing algorithm, from which we obtain the limiting distribution of the test and establish its asymptotic power. The numerical performance of the test is validated through an extensive simulation study. As real-data applications, we present the analysis based on "international economic growth" data (Belloni et al., 2013), which is found to be homoscedastic, and "supermarket" data (Lan et al., 2016), which is found to be heteroscedastic.

1. INTRODUCTION

In many modern economic applications, datasets often contain a large number of covariates p relative to the sample size n. When performing high-dimensional regression on such data, parameter estimation typically assumes homoscedastic errors for simplicity. However, heteroscedasticity is common in practice, and ignoring it can lead to inefficient estimates and invalid inference. Despite its practical importance, testing for heteroscedasticity remains both underexplored and challenging in high-dimensional settings.

Classical heteroscedasticity tests, such as those proposed in Breusch and Pagan (1979), White (1980), and Koenker and Bassett (1982), were developed under the assumption of fixed p, and are no longer valid when $p > n$, which is the primary focus of this article. To address this gap, we propose a high-dimensional

The authors would like to sincerely thank three anonymous referees, the associate editor, and the editor for their careful reading of the manuscript and for their constructive comments and valuable suggestions, which significantly improved the quality of this work. The research presented in this article was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia. The manuscript was written during a research visit to the University of Minnesota (Twin Cities). The first revision was completed at the University of East Anglia, and the second at the University of Manchester. Address correspondence to Jing Zhou, Department of Mathematics, University of Manchester, United Kingdom, e-mail: jing.zhou@manchester.ac.uk.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

heteroscedasticity test in the proportional regime where $n/p \to \delta \in (0,1]$. Our approach extends the classical test of Newey and Powell (1987), which considers the following heteroscedastic linear model:

$$Y = X\beta_0 + \sigma(X\gamma_0)\varepsilon, \tag{1.1}$$

where $X = (X_1, \ldots, X_p) \in \mathbb{R}^p$, $\beta_0, \gamma_0 \in \mathbb{R}^p$ are unknown parameter vectors, and $Y, \varepsilon \in \mathbb{R}$, with ε independent of X. The function $\sigma(\cdot)$ captures heteroscedasticity, allowing the error variance to vary with the linear combination $X\gamma_0$.

To detect nonconstant variance, we use the expectile as a distributional statistic, evaluated at various levels $\tau \in (0,1)$, defined as

$$u_\tau(Y) - \mathbb{E}(Y) = \frac{2\tau - 1}{1-\tau} \int_{[u_\tau(Y),\infty)} (y - u_\tau(Y))\, dF_Y(y). \tag{1.2}$$

In model (1.1), the τth expectile of Y given X is $u_\tau(Y \mid X) = X\beta_0 + u_\tau(\sigma(X\gamma_0)\varepsilon)$. A key insight is that under heteroscedasticity, the population parameter β_0 is affected by an additional contribution from $\sigma(X\gamma_0)$, scaled by the expectile of the error term ε. In contrast, under homoscedasticity, β_0 remains invariant across different values of τ. This observation becomes even clearer when a local linear approximation of $\sigma(\cdot)$ in (2.3) is considered, and it motivates a testing framework based on distributional statistics (Koenker and Bassett, 1982; Newey and Powell, 1987), where the hypothesis concerns the influence of γ_0.

Compared to previous work (Newey and Powell, 1987), this article addresses two additional challenges commonly encountered in contemporary datasets: (i) the number of parameters p exceeds the sample size n and diverges at a linear rate: $n/p \to \delta \in (0,1]$ and (ii) the p-vectors β_0 and γ_0 are sparse. This linear rate regime was first discussed in the early approximate message passing (AMP) literature (Donoho, Maleki, and Montanari, 2009, 2010) and then formally stated in Bayati and Montanari (2011a, Thm. 1). The AMP algorithm has attracted significant attention for the theoretical analysis of regularized M-estimators and low-rank matrix estimation. Subsequent work often separately discusses $\delta \in (0,1]$ for sparse regression with regularization (e.g., Bradic, 2016; Su, Bogdan, and Candès, 2017; Weinstein, Barber, and Candès, 2017; Weinstein et al., 2023; Zhou, Claeskens, and Bradic, 2020) and $\delta \in (1,\infty)$ for unstructured regression coefficient vectors (Bradic, 2016; Donoho and Montanari, 2016; Sur, Chen, and Candès, 2019; Sur and Candès, 2019; Candès and Sur, 2020). We refer to Feng et al. (2022) for a thorough literature review. In addition to the AMP framework, El Karoui et al. (2013) and El Karoui (2018) introduced an alternative way of analyzing the estimators in this n,p proportional regime, with conclusions agreeing with the AMP analysis. Specifically, the analysis addresses unregularized and ridge regression, where a sparse structure is not assumed for the parameter vector. In this work, we focus on sparse regression and address the $n/p \to \delta \in (0,1]$ setting.

In the past decade, fundamental research for high-dimensional linear regression has focused on sparse regularization (Tibshirani, 1996; Candès and Tao, 2007;

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

Zou, 2006; Fan and Lv, 2008; Zhang, 2010). We refer to Bühlmann and van De Geer (2011) and Fan et al. (2020) for a systematic review of this topic. Most existing literature in this field assumes homoscedastic regression errors, that is, $\sigma(\cdot)$ is a constant, as opposed to the fact that heteroscedasticity is commonly observed in real data, see, for example, Wang, Wu, and Li (2012) and Daye, Chen, and Li (2012) on genomic data and Belloni, Chernozhukov, and Wang (2014) on economic data.

There are natural solutions to address heteroscedasticity in model estimation. Cattaneo, Jansson, and Newey (2018) and Jochmans (2022) addressed this by splitting the p covariate vector X into a fixed low-dimensional subvector X_s of main interest with unknown parameter β_s and a nuisance subvector X_{p-s} with nonnegligible size $q = p - s$. Their high dimensionality refers to a setting where $n > p$ but p is considerably larger than in the classic setting, where $p/n \to 0$. The asymptotic normality is further derived for $\hat{\beta}_s$, and the main focus is on heteroscedastic-robust asymptotic covariance estimators. Addressed by Jochmans (2022, Sect. 3.2, 3.3), the consistency of the heteroscedasticity-robust covariance estimator requires the length of the nuisance parameter to satisfy $\lim_n \sup q/n < 1$, and handling the case of $1/2 < \lim_n \sup q/n < 1$ is already challenging. Another approach considers replacing the least-squares loss function with asymmetric loss functions, such as the quantile loss functions (Koenker and Bassett, 1978; Koenker, 2005) and the expectile loss functions (Newey and Powell, 1987). A thorough review of quantile regression can be found in Koenker (2017). High-dimensional quantile regression, addressing $p > n$ setting, has been studied by several authors (Wu and Liu, 2009; Belloni and Chernozhukov, 2011; Wang et al., 2012; Zheng, Gallagher, and Kulasekera, 2013; Gu et al., 2018; He et al., 2023). Gu and Zou (2016) studied high-dimensional expectile regression and proposed COSALES, which couples two expectile loss functions at expectile level 0.5 and another level of interest. By this construction, COSALES can simultaneously estimate parameter vectors β_0 and γ_0 in (2.3). Additionally, Zhou and Zou (2023) proposed a cross-fitted residual regression for estimating heteroscedasticity in (2.3).

Despite recent efforts in estimating heteroscedasticity through methods, such as quantile regression, expectile regression, or cross-refitted regression, testing for the presence of heteroscedasticity remains largely unexplored in the high-dimensional literature. Since quantile and expectile loss functions are robust to heteroscedastic regression errors, using these two asymmetric loss functions to construct heteroscedastic tests is natural and thus taken in this article. Moreover, as previously mentioned, the Newey–Powell test is shown to be superior to the quantile-based test in low-dimensional settings (Newey and Powell, 1987). Therefore, we focus on the high-dimensional Newey–Powell test in this article. In fact, the contrast between expectile regression and quantile regression is even more significant in high dimensions, owing to the smoother loss function in expectile regression. Gu and Zou (2016) show that fitting a high-dimensional expectile regression model is as simple as fitting Lasso regression. In contrast, serious efforts have been dedicated to handling the nonsmooth loss function in

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

high-dimensional quantile regression, including ADMM (Gu et al., 2018) and smoothing approximation (He et al., 2023).

Given the above considerations, we extend the Newey–Powell test (Newey and Powell, 1987) to sparse high-dimensional settings where $n/p \to \delta \in (0,1]$, $n,p \to \infty$, and ℓ_1-regularization or other sparse regularization is necessary to obtain good estimators. The heteroscedasticity of regression error is tested via $H_{0,j} : \gamma_{0,j} = 0$ versus $H_{1,j} : \gamma_{0,j} \neq 0$, $j = 1,\ldots,p$. By using the AMP theory, we establish the joint asymptotic normality of multiple expectile estimators, which we then use to construct the test statistic. We analytically examine the size and power of the proposed test.

The rest of the article is organized as follows. In Section 2, we first introduce the notation and assumptions for the high-dimensional heteroscedastic linear model in (2.3), following the expectile regression in Section 2.1. The AMP algorithm is introduced in Section 3, and the main technical results are included in Section 3.3. Section 4 presents the proposed heteroscedasticity test. The analytical expression of the power function is presented in Theorem 2. We discuss a practically important decorrelation step for using AMP in applications in Section 5. Section 6 contains simulation studies and real data examples from Belloni et al. (2013) and Lan et al. (2016). Section 7 concludes the article. The assumptions, proofs of lemmas and theorems, and auxiliary lemmas and definitions are provided in Appendixes A–C, respectively. Additional supporting results are included in the Supplementary Material (Zhou and Zou, 2025).

2. HETEROSCEDASTIC LINEAR REGRESSION

We restate the well-studied form of the heteroscedastic model from Koenker and Bassett (1982) and Newey and Powell (1987):

$$Y = X\beta_0 + \big(1 + X\gamma_0\big)\varepsilon, \tag{2.3}$$

where $\beta_0, \gamma_0 \in \mathbb{R}^p$ are unknown parameter vectors. In addition, we have n independent and identically distributed (i.i.d.) copies of the pair (Y,X) denoted by $(Y_i,X_i), i = 1,\ldots,n$. Further, we denote n-vectors $\boldsymbol{Y} = (Y_1,\ldots,Y_n)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)^\top$; the design matrix $\boldsymbol{X} = (X_1,\ldots X_n)^\top$. The components of the vector $\boldsymbol{\varepsilon}$ satisfy Assumption (A1).

(A1) $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d. random variables with mean zero, a finite $(2\kappa-2)$th moment for $\kappa \geq 2$.

The constant κ depends on the loss function used to estimate the unknown parameters β_0, γ_0; in our case, the expectile loss function in (2.6). This statement will become clear in Section 3.2 when we discuss the AMP in detail. Further, following Bayati and Montanari (2011a), the empirical measure of a p-vector is defined as putting $1/p$ weight on each component of the vector. Then, the unknown parameter vector γ_0 complies with Assumption (A2).

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

(A2) The empirical measure of the components of the p-vector γ_0, when $p \to \infty$, converges weakly to the probability measure of a random variable Γ_0 with bounded $(2\kappa - 2)$th moment for $\kappa \geq 2$.

This assumption precedes a standard Assumption (A3) on the parameter vector β_0 in the AMP literature. In practice, the proposed test is valid for β with finite p components (see Section 4), and the hypothesis in (4.33) tests the components γ_j's for finite p. We also demonstrate applying the proposed test to two real data examples with finite p in Section 6.3.

(A3) The empirical measure of the components of the p-vector β_0, when $p \to \infty$, converges weakly to the probability measure of a random variable B_0 with bounded $(2\kappa - 2)$th moment for $\kappa \geq 2$.

Notice $X\gamma_0 = \sqrt{n}X\frac{\gamma_0}{\sqrt{n}}$, where the scaled design $\sqrt{n}X$ and parameter vector $\frac{\gamma_0}{\sqrt{n}}$ agree with the assumptions in Koenker and Bassett (1982). Further, the sequence $\gamma_n = \frac{\gamma_0}{\sqrt{n}}$ restricts the calculation to local alternatives and guarantees well approximation of sequences of multiplicative heteroscedastic models.

Following Newey and Powell (1987, Thm. 1(iii)), the τth expectile of Y given X can be expressed as

$$u_\tau(Y \mid X) = X\beta_0 + (1 + X\gamma_0)u_\tau(\varepsilon) = X\beta_\tau + u_\tau(\varepsilon), \tag{2.4}$$

where $\beta_\tau = \beta_0 + \gamma_0 u_\tau(\varepsilon)$ is the population parameter vector corresponding to the τth expectile of Y, approximated by $X\beta_\tau$. Here, $u_\tau(\varepsilon)$ is the τth expectile of the error term, as defined in (1.2), and will be denoted simply as u_τ henceforth. The parameter β_τ is fixed and depends on the scalar u_τ. In this spirit, we rewrite the model as follows:

$$Y = X\beta + \varepsilon, \tag{2.5}$$

where $\beta = \beta_0 + \varepsilon\gamma_0$ and the error ε satisfies Assumption (A1). In addition, we assume the ratio $n/p \to (0, 1]$ when $n, p \to \infty$. Let $s = \|\beta\|_0$, where $\|\cdot\|_0$ denotes the ℓ_0-norm of a vector. Obviously, $s \leq \|\beta_0\|_0 + \|\gamma_0\|_0$. The way of parameterization in (2.5) has two reasons: (1) we want to express the model to fit it in the AMP framework for asymptotic analysis, where an independent error term stands alone and (2) the error term ε determines the location change of β from β_0. Identifying γ_0 relies on making use of different spots of the error distribution and requires a pair of regularized estimators. This is technically challenging due not only to the nontrivial asymptotic behavior of regularized estimators when $p > n$, but also to the correlation structure among pairs of estimators whose population parameters do not have the same center.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

2.1. Expectile Regression under Heteroscedasticity

The expectile regression (Newey and Powell, 1987) estimates the τth conditional expectile of Y by minimizing an asymmetric quadratic loss function defined as

$$\rho_\tau(x; u_\tau) = |\tau - I\{x \le u_\tau\}|(x - u_\tau)^2, \tag{2.6}$$

where $\tau \in (0,1)$ is the expectile level and u_τ denotes the τth expectile of error ε. which is the solution to the nonlinear equation (Newey and Powell, 1987, Eq. (2.7)). A special case is $\tau = 1/2$, where the expectile loss function reduces to the least-squares loss. Under our model, the conditional expectile of Y at expectile level τ follows $X(\beta_0 + \gamma_0 u_\tau) + u_\tau$ (Newey and Powell, 1987) and the coefficient vector $\beta_\tau = \beta_0 + \gamma_0 u_\tau$, where u_τ is the expectile of error ε at τ in (1.2). Since ε has mean zero (Assumption (A1)), expectile regression at $\tau = 1/2$ does not identify the combination of γ_0 and β_0. Instead, $\beta_{1/2}$ only estimates β_0.

The proposed heteroscedasticity test is constructed based on the estimator of the parameter vector β via minimizing the empirical expectile loss function. Specifically, we consider the ℓ_1-regularized expectile estimator defined as follows:

$$\widehat{\beta}_\tau(\lambda) = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i\beta; u_\tau) + \lambda\|\beta\|_1, \tag{2.7}$$

where $\|\cdot\|_1$ denotes the ℓ_1-norm and λ denotes the tuning parameter. Due to technical restrictions of the AMP algorithm (see Section 3), we do not jointly optimize the intercept term u_τ and the slope vector β. Instead, we treat the term u_τ as a parameter of the expectile loss function. Recentering Y by the intercept u_τ is important to guarantee the following AMP analysis when considering robust loss functions. In our simplest homoscedastic error case, plugging in u_τ reduces the conditional τ-mean in (2.4) to only $X\beta_\tau$. This intercept term is only ignorable when we address the conditional mean of Y given X by Assumption (A1), the zero mean of ε. In homoscedastic settings, the error expectile u_τ can be estimated by solving

$$\hat{u}_\tau - \frac{1}{n}\sum_{i=1}^{n} r_i = [(2\tau - 1)/(1 - \tau)]\frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{u}_\tau)I\{r_i \ge \hat{u}_\tau\}, \tag{2.8}$$

where $r_i = Y_i - \widehat{\beta}_\tau X_i$'s are obtained using an initial estimator $\widehat{\beta}$, such as the regularized least-squares estimators, the regularized Huber estimator, etc. However, estimating u_τ becomes challenging under heteroscedasticity since solving (2.8) estimates the τ's expectile of $\sigma(X\gamma_0)\varepsilon$ instead of ε. We observe that u_τ is reflected in the intercept term of $\gamma_0 u_\tau$ in the composition $\beta_\tau = \beta_0 + \gamma_0 u_\tau$. Since the coupled sparse asymmetric least squares (COSALES) (Gu and Zou, 2016) estimates β_0 and $\gamma_0 u_\tau$ simultaneously for any $\tau \in (0,1), \tau \ne 1/2$, we use the intercept estimator of $\gamma_0 u_\tau$ from the COSALES as a surrogate for the true u_τ in (1.2). Further, to evaluate the estimation error of u_τ on the test, we compared the numerical results using the true and estimated u_τ. We observe that using the estimated u_τ from the COSALES

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

does not affect the test results. We also investigated the performance of the test when plugging in (2.8) under the null in Table S.1 in the Supplementary Material (Zhou and Zou, 2025). The numerical performance appears similar.

Importantly, when the regression errors are i.i.d., that is, $\sigma(\cdot) = 1$, the slope β does not vary across different expectile levels in the classical setting with fixed p and $n \to \infty$. We propose to test the violation of homoscedasticity under diverging n, p based on the ground of common slope under homoscedasticity at different expectile levels $\tau_k, k = 1, \ldots, K$.

3. APPROXIMATE MESSAGE PASSING FOR EXPECTILE REGRESSION

The AMP analysis starts by regularizing the loss functions ρ by using the proximal mapping operator defined as

$$\text{Prox}\ (z; b) = \arg\min_{x \in \mathbb{R}} \left\{ b\rho(x) + \frac{1}{2}(x - z)^2 \right\}, \tag{3.9}$$

where the parameter $b > 0$ is a positive scalar that controls the step size moving toward the minimum of the loss function ρ. In our case, the proximal operator for expectile loss (2.6) can be obtained following the derivation in Section B.1 and can be expressed as follows:

$$\begin{aligned} \text{Prox}_{\tau}(z; b) &= \begin{cases} \big(z + 2b(1-\tau)u_\tau\big)\big(2b(1-\tau)+1\big)^{-1}, & z \le u_\tau \\ \big(z + 2b\tau u_\tau\big)\big(2b\tau + 1\big), & z > u_\tau \end{cases} \\ &= \frac{z - 2bu_\tau(\tau - I\{z \le u_\tau\})}{1 - 2b(\tau - I\{z \le u_\tau\})}. \end{aligned} \tag{3.10}$$

With the analytical expressions of the proximal operator for the expectile loss function (3.10), we further work out the expression of the effective score function (Donoho and Montanari, 2016) defined as

$$\widetilde{G}(z; b) = b\partial\rho(x)|_{x = \text{Prox}(z; b)}, \text{ with } b > 0, \tag{3.11}$$

where $\partial\rho$ is the subgradient of ρ defined as $\partial\rho(x) = \{y : \rho(u) \ge \rho(x) + y(u - x), \text{for all } u\}$. This leads to the subgradient of the expectile loss function in (3.12) as

$$\partial\rho_\tau(z) = 2|I\{z \le u_\tau\} - \tau|(z - u_\tau). \tag{3.12}$$

One could interpret the effective score function by associating it with the score function. By plugging (3.10) in (3.12) and using the definition of the effective score function in (3.11), we obtain the effective score function as follows:

$$\widetilde{G}_\tau(z; b) = \begin{cases} \frac{2b(1-\tau)}{2b(1-\tau)+1}(z - u_\tau), & z \le u_\tau \\ \frac{2b\tau}{2b\tau+1}(z - u_\tau), & z > u_\tau. \end{cases} \tag{3.13}$$

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

To analyze regularized estimators under a sparse structure, specifically, when $s/p \to \omega$, where s denotes the number of nonzero components of β, Bradic (2016) and Zhou et al. (2020) incorporated sparsity by introducing the rescaled effective score function

$$G(z;b) = \delta\omega^{-1}\widetilde{G}(z;b). \tag{3.14}$$

3.1. The Approximate Message Passing Algorithm

We consider the AMP algorithm to approximate the estimators in (2.7) in the sense that the averaged difference between the estimators obtained from the AMP algorithm and the estimators by minimizing the loss functions converges to zero almost surely, see Zhou et al. (2020, p. 2562, Eq. (23) and below) for a summary of discussions in the relevant references (Bayati and Montanari, 2011b; Bradic, 2016; Donoho and Montanari, 2016; Sur et al., 2019; Huang, 2020).

The AMP algorithm is an iterative algorithm indexed by t consisting of three steps that belong to the general recursion in Bayati and Montanari (2011a). We will describe the general recursion when stating the theory in Section 3.3. Let the matrix $\boldsymbol{X}$ satisfy Assumption (A4).

(A4) A standard Gaussian design: $X_i, i = 1, \ldots, n,$ are i.i.d. copies of X, where $X^\top \in \mathbb{R}^p$ with $X_j \sim N(0, 1/n), j = 1, \ldots, p,$ i.i.d. components.

Since β_τ is a sparse vector with most components being zero, the algorithm starts with an initialization $\widehat{\beta}_{(0)} = 0, t = 0$ and proceed at $t \geq 1$ following steps:

Step 1. Adjust residuals The adjusted residuals $z_{(t)}$ are obtained by

$$z_{(t)} = Y - \boldsymbol{X}\widehat{\beta}_{(t)} + \tag{3.15}$$
$$n^{-1}G(z_{(t-1)}; b_{(t-1)}) \sum_{j=1}^{p} I\left\{\eta\left(\widehat{\beta}_{(t-1),j} + X_j G(z_{(t-1)}; b_{(t-1)}); \theta_{(t-1)}\right) \neq 0\right\},$$

where G is defined in (3.14) and $\eta(\cdot; \theta_{(t-1)})$ is the soft-thresholding operator with parameter $\theta_{(t-1)}$ defined as $\eta(\beta; \theta) = \text{ sgn } (\beta) \max(|\beta| - \theta, 0)$.

Step 2. Update the parameter for effective score function The scalar parameter $b_{(t)}$ is chosen such that the average of the effective score function has slope 1, that is, $\frac{1}{n}\sum_{i=1}^{n} \partial_1 G(z_{i,(t)}; b_{(t)}) = 1$, where $\partial_1 G(z; b)$ denotes the derivative of G w.r.t. the first argument z in case of differentiable function G. Hence, for the expectile regression with calculations in Section B.2, the parameter $b_{(t)}$ is updated by solving

$$\frac{\omega}{\delta} = \frac{2b(1-\tau)}{2b(1-\tau)+1} \frac{\#\{i : z_{(t),i} \leq u_\tau\}}{n} + \frac{2b\tau}{2b\tau+1} \frac{\#\{i : z_{(t),i} > u_\tau\}}{n}, \tag{3.16}$$

where ω is estimated by the fraction of the nonnull components in an initial estimator $\widehat{\beta}$, and u_τ is replaced by plugging in an estimator from the

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

COSALES algorithm (see discussions under (2.8)). The update $b_{(t)}$ at step t is taken to be the first value on a fine grid that solves (3.16).

Step 3. Update the estimators for β Update the coefficient estimator for β as follows:

$$\widehat{\beta}_{(t+1)} = \eta(\widetilde{\beta}_{(t)}; \theta_{(t)}), \text{ where } \widetilde{\beta}_{(t)} = \widehat{\beta}_{(t)} + \boldsymbol{X}^\top G(z_{(t)}; b_{(t)}). \tag{3.17}$$

The estimator $\widehat{\beta}$ is an approximation for regularized estimators in (2.7), where the soft-thresholding operator $\eta(\cdot;\theta)$ is the proximal operator for the regularizer $\|\beta\|_1$. The tuning parameter θ is chosen to minimize the mean squared error (MSE) of $\widehat{\beta}$. As suggested in Donoho and Johnstone (1994, 1998), Donoho et al. (2009), and Bayati and Montanari (2011b), we take $\theta_{(t)} = \alpha\bar{\zeta}_{\text{emp}(t)}$ for scalar α and

$$\bar{\zeta}^2_{\text{emp},(t)} = \frac{1}{n}\sum_{i=1}^{n} G(z_{i,(t)}; b_{(t)})^2. \tag{3.18}$$

In practice, one chooses a grid of values for α, then obtains the corresponding estimator $\widehat{\beta}$ and the estimated MSE for each value of α on the grid. The optimal α is chosen to be the one with the minimum estimated MSE. The details can be found in Zhou et al. (2020, Algorithm 1). We also implement practical stopping rules: the maximum number of iterations is set to $T = 100$, and convergence is declared when $\text{MSE}(\widehat{\beta}_{(t+1)}, \widehat{\beta}_{(t)}) = \frac{1}{p}\sum_{j=1}^{p}(\widehat{\beta}_{j,(t+1)} - \widehat{\beta}_{j,(t)})^2 < \text{tol}$. In the simulations, the tolerance level tol is set to 10^{-8} for high sparsity settings and 10^{-7} for medium sparsity settings.

Details of the population version of the parameter $\bar{\zeta}^2_{\text{emp},(t)}$ can be found in Section 3.2. The estimator $\widetilde{\beta}$ is of major interest in this project since it is connected to the debiased estimator (Javanmard and Montanari, 2014; van de Geer et al., 2014) and follows a Gaussian distribution asymptotically (Bayati and Montanari, 2011b; Bayati, Erdogdu, and Montanari, 2013). We will discuss this in Section 3.3.

3.2. State Evolution

When $p, n \to \infty$, the averaged-over-components performance of estimators $\widetilde{\beta}_{(t)}$ and $\widehat{\beta}_{(t)}$ from the AMP algorithm can be described using two important parameters $\bar{\zeta}^2_{(t)}, \bar{\sigma}^2_{(t)}$. Assume that the following limit exists $\bar{\sigma}^2_{(0)} = \lim_{p\to\infty}(p\delta)^{-1}\|q_{(0)}\|^2 < \infty$. With such initialization, the state evolution recursion for $t \in \{0\}\bigcup\mathbb{N}$ is defined as

$$\bar{\zeta}^2_{(t)} = \mathbb{E}[G(\varepsilon + \bar{\sigma}_{(t)}Z; b_{(t)})^2], \tag{3.19}$$

$$\bar{\sigma}^2_{(t)} = \delta^{-1}\mathbb{E}[(\eta(B + \bar{\zeta}_{(t-1)}Z; \theta_{(t-1)}) - B)^2], \tag{3.20}$$

where B denotes the limiting form of the components of β in (2.5) and $Z \sim N(0,1)$ is a standard Gaussian random variable that is independent of B and ε. The

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

parameter $\bar{\zeta}_{(t)}$ is the limiting version of (3.18) when $n, p \to \infty$ (Zhou et al., 2020, Sect. B.2.3).

The readers might wonder if (3.19) and (3.20) still hold under heteroscedasticity since the relevant literature has not discussed this issue and often assumes homoscedasticity. It is worth mentioning that the state evolution recursion in (3.19) and (3.20) is valid as long as the empirical distribution of ε and the components of β converge weakly (Bayati and Montanari, 2011a, Thm. 2). We will discuss this in detail in Section 3.3.

Using the two parameters, we can analyze the limiting performance of the estimator $\widehat{\beta}$ from the AMP algorithm. A typical example is analyzing the asymptotic MSE (AMSE), defined as $\text{AMSE}(\widehat{\beta}_{(t+1)}, \beta) = \lim_{p\to\infty} \frac{1}{p} \sum_{j=1}^{p} (\widehat{\beta}_{(t+1),j} - \beta_j)^2$. It is easy to show that AMSE complies with

$$\text{AMSE}(\widehat{\beta}_{(t+1)}, \beta) = \lim_{p\to\infty} \frac{1}{p} \sum_{j=1}^{p} \left(\eta(\widetilde{\beta}_{(t),j}; \theta_{(t)}) - \beta_j \right)^2 \overset{a.s.}{=} \mathbb{E}[\{\eta(B + \bar{\zeta}_{(t)} Z; \theta_{(t)}) - B\}^2],$$

which corresponds to Bradic (2016, Eq. (3.4)) with $\widetilde{\beta}_{(t),j}$ in (3.17). An important observation from (3.20) is that, at each iteration step t, the estimator $\widetilde{\beta}_{(t),j}$ is obtained by adding a Gaussian noise $N(0, \bar{\zeta}_{(t)}^2)$ to the jth component of the true parameter vector β_j. This suggests that $\widetilde{\beta}_{(t)} \sim N(\beta, \bar{\zeta}_{(t)}^2 I_{p\times p})$ for finite marginals. Similar results are also obtained for the Lasso estimator in Bayati and Montanari (2011b), which is a special case of the ℓ_1-regularized expectile regression when $\tau = 1/2$. We will come back to the asymptotic normality of $\widetilde{\beta}_{(t)}$ in Section 3.3 after introducing the main technical lemma for multiple estimators $\widetilde{\beta}_{k,(t)}, k = 1, 1 \ldots, K$.

3.3. Asymptotic Analysis for Multiple Estimators

The heteroscedasticity test for expectile regression is based on the difference of the estimators at different expectile levels $\tau_k, k = 1, \ldots, K$. The test statistic is based on asymptotic analysis for multiple estimators $\widehat{\beta}_k$'s among which the correlation is nontrivial when estimated using the same dataset (Zhou et al., 2020, Lemma 1). Since estimating the correlation requires information on the unknown parameter vector β, we further extend the discussion on the correlations to the adjusted residuals $z_{k,(t)}$'s in the main technical result—Lemma 1 and construct an adjusted-residual-based estimator for the covariance between $\widetilde{\beta}_k$'s. Further, we want to address that the technical results, including the general recursion and Lemma 1 in this section, hold for similar loss functions, such as the quantile loss function.

We describe the general recursions in Bayati and Montanari (2011a) for the completeness of the article and a better connection of Lemma 1 and the AMP algorithm in Section 3.1. Under the heteroscedastic model (2.5), the coefficient vector $\beta_k = \beta_0 + u_k \gamma_0$, where u_k is a constant featuring the strength of heteroscedasticity which depends on ε. For expectile regression, u_k is the error expectile. By Assumptions (A2) and (A3), the empirical distribution of the parameter vector β_k

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

converges weakly to $B_k = B_0 + u_k\Gamma_0$. In cases where $u_k = 0$ or homoscedasticity such that $\Gamma_0 = 0$ degenerates, $B_k = B_0$.

In addition, let a noise vector $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ and the coefficient vector $\beta_k \in \mathbb{R}^p$ satisfy Assumptions (A1) and (A3). By fixing the initial condition $q_{k,(0)}$ and recall the convention $m_{k,(-1)} = 0$, the general recursion for $t \geq 0$ is defined as

$$\begin{aligned} h_{k,(t+1)} &= \boldsymbol{X}^\top m_{k,(t)} - \xi_{1,k,(t)} q_{k,(t)}, \quad m_{k,(t)} = g_{1,k,(t)}(d_{k,(t)}, \boldsymbol{\varepsilon}), \\ d_{k,(t)} &= \boldsymbol{X} q_{k,(t)} - \xi_{2,k,(t)} m_{k,(t-1)}, \quad q_{k,(t)} = g_{2,k,(t)}(h_{k,(t)}, \beta_k), \end{aligned} \tag{3.21}$$

where

$$\xi_{1,k,(t)} = n^{-1} \sum_{i=1}^{n} \partial_1 g_{1,k,(t)}(d_{k,(t),i}, \varepsilon_i), \tag{3.22}$$

$$\xi_{2,(t)} = (\delta p)^{-1} \sum_{j=1}^{p} \partial_1 g_{2,k,(t)}(h_{k,(t),j}, \beta_{k,j}). \tag{3.23}$$

As a special case, the AMP algorithm takes

$$\begin{aligned} h_{k,(t+1)} &= \beta_k - \boldsymbol{X}^\top G_k(z_{k,(t)}; b_{k,(t)}) - \widehat{\beta}_{k,(t)}, \\ q_{k,(t)} &= \widehat{\beta}_{k,(t)} - \beta_k, \qquad z_{k,(t)} = \boldsymbol{\varepsilon} - d_{k,(t)}, \end{aligned}$$

which indicates

$$d_{k,(t)} = \boldsymbol{\varepsilon} - z_{k,(t)}, \qquad m_{k,(t)} = -G_k(z_{k,(t)}; b_{k,(t)}), \tag{3.24}$$

with the functions $g_{1,k,(t)}(x_1, x_2) = -G_k(x_2 - x_1; b_{k,(t)})$ and $g_{2,k,(t)}(x_1, \beta) = \eta(\beta - x_1; \theta_{k,(t)}) - \beta$. Notice $\widehat{\beta}_{k,(t)}$ the regularized estimator in (3.17), and

$$h_{k,(t+1)} = \beta_k - \widetilde{\beta}_{k,(t)}. \tag{3.25}$$

This detail is mathematically trivial but allows us to apply Lemma 1 directly to obtain the asymptotic normality of any pairs $(\widetilde{\beta}_{k_1,(t)}, \widetilde{\beta}_{k_2,(t)})$, as well as later obtain the power function for the proposed hypothesis test in Theorem 2.

Before presenting the main technical lemma, we assume a moment condition (Assumption (A5)), which also complements Assumptions (A1)–(A3). Recall in Assumptions (A1)–(A3), we state that the empirical measures of the components of the vectors β_0, γ_0, and $\boldsymbol{\varepsilon}$ converge weakly to the probability measures of random variables B_0, Γ_0, and ε, respectively. Let $\widehat{f}_{\beta_0(p)}, \widehat{f}_{\gamma_0(p)}$, and $\widehat{f}_{\boldsymbol{\varepsilon}(p)}$ denote the empirical measures of β_0, γ_0, and $\boldsymbol{\varepsilon}$, that is, putting equal weights on the components of the vectors. Specifically, for the sequence of vectors $\boldsymbol{\varepsilon}(p)$, we index the sequence by p to unify the notation following Bayati and Montanari (2011a). This sequence index is legit since our asymptotic theory is derived under $n/p \to \delta \in (0, 1]$. Then, Assumption (A5) can be stated as follows:

(A5) For some $\kappa > 1$, let the empirical measures $\widehat{f}_{\beta_0(p)}, \widehat{f}_{\gamma_0(p)}$, and $\widehat{f}_{\boldsymbol{\varepsilon}(p)}$ converge weakly to the probability measures f_{B_0}, f_{Γ_0}, and f_ε. In addition, assume the following $(2\kappa - 2)$th moment conditions hold:

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

(a) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{\beta_0(p)}}(B_0^{2\kappa-2}) = \mathbb{E}_{f_{B_0}}(B_0^{2\kappa-2}) < \infty$;
(b) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{\gamma_0(p)}}(\Gamma_0^{2\kappa-2}) = \mathbb{E}_{f_{\Gamma_0}}(\Gamma_0^{2\kappa-2}) < \infty$;
(c) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{\boldsymbol{\varepsilon}(p)}}(\varepsilon^{2\kappa-2}) = \mathbb{E}_{f_\varepsilon}(\varepsilon^{2\kappa-2}) < \infty$;
(d) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{q_k(p)}}(B_k^{2\kappa-2}) < \infty$.

The constant κ depends on the pseudo-Lipschitz order of function $g_{1,k,(t)}$, see Definition 1 for the definition of pseudo-Lipschitz functions and Lemma 1 for associating the order κ with the function $g_1 = \tilde{\psi}_{c'}, \tilde{\psi}_{c''}$. As described above, in the special case where the general recursion (3.21) is the AMP algorithm, function g_1 takes a more specific form, that is, the negative effective score function $-G_k(\cdot;\cdot)$.

LEMMA 1. *Let the sequences of design matrices $\{\boldsymbol{X}(p)\}$, location coefficient vectors $\{\beta_0(p)\}$, heteroscedasticity parameter vectors $\{\gamma_0(p)\}$, error vectors $\{\boldsymbol{\varepsilon}(p)\}$, and initial condition vectors $\{q_k(p)\}$ be weakly convergent sequences for K recursions satisfying Assumptions (A1)–(A5). Let $\{\bar{\sigma}^2_{k,(t)}, \bar{\zeta}^2_{k,(t)}\}$ be defined uniquely by the recursions in (3.19) and (3.20). These are the state evolution parameters for the kth estimation with initialization $\bar{\sigma}^2_{k,(0)} = \lim_{n\to\infty}(p\delta)^{-1}\|q_{k,(0)}\|^2$. Further, we assume a constant u_k for each of the K recursions specifying the signal strength of the heteroscedasticity parameter, such that the parameter vector (2.5) has the composition $\beta_k = \beta_0 + u_k\gamma_0$. Assume Lemma 1 in Bayati and Montanari (2011a) holds individually for each of the K recursions. Additionally, for all pseudo-Lipschitz functions $\tilde{\psi}_c : \mathbb{R}^{t+2} \to \mathbb{R}$ of order κ_c for some $1 \le \kappa_c = \kappa/2$ with κ as in (A5) and the iteration index $t \ge 0, t \in \mathbb{Z}$,*

$$\lim_{p\to\infty} \frac{1}{p}\sum_{j=1}^{p} \tilde{\psi}_c(h_{k_1,(1),j}, \ldots, h_{k_1,(t+1),j}, \beta_{k_1,j})\tilde{\psi}_c(h_{k_2,(1),j}, \ldots, h_{k_2,(t+1),j}, \beta_{k_2,j}) \overset{a.s.}{=}$$
$$\mathbb{E}[\tilde{\psi}_c(\bar{\zeta}_{k_1,(0)}Z_{k_1,(0)}, \ldots, \bar{\zeta}_{k_1,(t)}Z_{k_1,(t)}, B_{k_1})\tilde{\psi}_c(\bar{\zeta}_{k_2,(0)}Z_{k_2,(0)}, \ldots, \bar{\zeta}_{k_2,(t)}Z_{k_2,(t)}, B_{k_2})]. \tag{3.26}$$

And, for all pseudo-Lipschitz functions $\tilde{\psi}_{c'}, \tilde{\psi}_{c''} : \mathbb{R}^{t+2} \to \mathbb{R}$ of order $\kappa_{c'}, \kappa_{c''}$ for some $1 \le \kappa_{c'}, \kappa_{c''} \le \kappa/2$, where $\kappa'_c + \kappa''_c = \kappa$ with κ as in (A5),

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \tilde{\psi}_{c'}(d_{k_1,(0),i}, \ldots, d_{k_1,(t),i}, \varepsilon_i)\tilde{\psi}_{c''}(d_{k_2,(0),i}, \ldots, d_{k_2,(t),i}, \varepsilon_i) \overset{a.s.}{=}$$
$$\mathbb{E}[\tilde{\psi}_{c'}(\bar{\sigma}_{k_1,(0)}\hat{Z}_{k_1,(0)}, \ldots, \bar{\sigma}_{k_1,(t)}\hat{Z}_{k_1,(t)}, \varepsilon)\tilde{\psi}_{c''}(\bar{\sigma}_{k_2,(0)}\hat{Z}_{k_2,(0)}, \ldots, \bar{\sigma}_{k_2,(t)}\hat{Z}_{k_2,(t)}, \varepsilon)], \tag{3.27}$$

where $(Z_{k,(0)}, \ldots, Z_{k,(t)}), (\hat{Z}_{k,(0)}, \ldots, \hat{Z}_{k,(t)})$ are multivariate Gaussian with, $k = k_1, k_2$, is a $(t+1)$-dimensional zero-mean multivariate standard normal vector independent of B_k, ε; at iteration t, $(Z_{k_1,(t)}, Z_{k_2,(t)})$ and $(\hat{Z}_{k_1,(t)}, \hat{Z}_{k_2,(t)})$ are standard normal 2-vectors with covariance not necessarily equal to zero.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

Further, for $0 < t_1, t_2 < t$, *with* $\xrightarrow{P}$ *denoting convergence in probability, it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} m_{k_1,(t_1),i} m_{k_2,(t_2),i} \xrightarrow{P} \lim_{p\to\infty} \frac{1}{p}\sum_{j=1}^{p} h_{k_1,(t_1+1),j} h_{k_2,(t_2+1),j}. \tag{3.28}$$

Lemma 1 generalizes Bayati and Montanari (2011a, Lemma 1(b)) by considering the correlation structure between multiple recursions. Further, Lemma 1 extends the analysis to account for error heteroscedasticity, which has not been addressed in the relevant literature. When $k_1 = k_2$ and the heteroscedastic parameter vector $\gamma_0 = \mathbf{0}$ is a null vector, Lemma 1 reduces to Bayati and Montanari (2011a, Lemma 1(b)). Further, the covariance matrices of the standard normal 2-vectors $(Z_{k_1,(t)}, Z_{k_2,(t)})$ and $(\hat{Z}_{k_1,(t)}, \hat{Z}_{k_2,(t)})$ will be involved in constructing the test statistic in Section 4. The proof of Lemma 1 is provided in the Supplementary Material (Zhou and Zou, 2025).

Remark 1. We thank a reviewer for drawing our attention to Theorem 1 and Lemma 2 in Javanmard and Montanari (2013). Their result agrees with our conclusion in (3.26), which shows the asymptotic approximation of the covariance of the structure of multiple estimators. Our additional contributions lie in establishing (3.27) and (3.28), which are of independent interest to statistics and econometrics, as they are important for establishing Theorem 1, (3.30), and demonstrate that the correlation between estimators is linked to the correlation of residuals.

To guarantee an estimable test statistic, we propose an adjusted-residual-based estimator in Theorem 1, which is directly obtainable from the AMP algorithm. The construction of the covariance matrix estimators is included in Section B.4 by directly applying (3.28). It is worth noting that (3.28) is a contribution to the existing AMP literature. The significance of (3.28) is that it shows the correlation between estimators (noticing $h_{k,(t+1)} = \beta - \widetilde{\beta}_{k,(t)}$ in (3.25)) from two sequences is asymptotically equivalent to the product of two effective score functions. This avoids the problem of lacking information about the true parameter vector β. Additionally, Theorem 1 validates that Z_{k_1} and Z_{k_2} are correlated by (3.29).

THEOREM 1. *Under the assumptions in Lemma 1 and* $\kappa = 2$ *for any two recursions* $k_1, k_2 = 1, \ldots, K$, *with probability 1, it holds that*

$$\lim_{p\to\infty} \frac{1}{p}\sum_{j=1}^{p} \left(\widetilde{\beta}_{k_1,(t),j} - \beta_{k_1,j}\right)\left(\widetilde{\beta}_{k_2,(t),j} - \beta_{k_2,j}\right) = \bar{\zeta}_{k_1,(t)} \bar{\zeta}_{k_2,(t)} \mathrm{Cov}(Z_{k_1}, Z_{k_2}), \tag{3.29}$$

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} & G_{k_1}(z_{k_1,(t),i}; b_{k_1,(t)}) G_{k_2}(z_{k_2,(t),i}; b_{k_2,(t)}) \\ &= \mathbb{E}[G_{k_1}(\varepsilon + \bar{\sigma}_{k_1}\hat{Z}_{k_1}; b_{k_1,(t)}) G_{k_2}(\varepsilon + \bar{\sigma}_{k_2}\hat{Z}_{k_2}; b_{k_2,(t)})].\end{aligned} \tag{3.30}$$

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

The estimator constructed as follows:

$$\widehat{\zeta}_{k_1,k_2,(t)} = \frac{1}{n}\sum_{i=1}^{n} G_{k_1}(z_{k_1,(t),i}; b_{k_1,(t)}) G_{k_2}(z_{k_2,(t),i}; b_{k_2,(t)}) \tag{3.31}$$

is a consistent estimator of the covariance $\bar{\zeta}_{k_1,(t)}\bar{\zeta}_{k_2,(t)}\mathrm{Cov}(Z_{k_1,(t)}, Z_{k_2,(t)})$.

Theorem 1 first validates our claim in Lemma 1 that $Z_{k_1,(t)}, Z_{k_2,(t)}$ are correlated. The correlation between $Z_{k_1,(t)}, Z_{k_2,(t)}$ describes the averaged correlation between the components of $\widetilde{\beta}_{k_1,(t)}$ and $\widetilde{\beta}_{k_2,(t)}$. A similar conclusion holds for $\hat{Z}_{k_1,(t)}, \hat{Z}_{k_2,(t)}$. Further, the covariance between $Z_{k_1,(t)}, Z_{k_2,(t)}$ can be approximated by using the adjusted residuals without any information about β.

The last piece of information before moving to the test statistic is the joint asymptotic distribution of the pair $(\widetilde{\beta}_{k_1}, \widetilde{\beta}_{k_2})$.

LEMMA 2 (Joint normality). *Under the assumptions in Lemma 1 for any two recursions* $k_1, k_2 = 1, \ldots, K$ *and for any pseudo-Lipschitz function* $\psi : \mathbb{R} \to \mathbb{R}$ *of order* κ_c *in (3.26), the following convergence for estimators from the AMP algorithm holds with probability 1:*

$$\lim_{p\to\infty} \frac{1}{p}\sum_{j=1}^{p} \psi(\widetilde{\beta}_{k_1,(t),j})\psi(\widetilde{\beta}_{k_2,(t),j}) = \mathbb{E}\big[\psi(B_{k_1} + \bar{\zeta}_{k_1,(t)}Z_{k_1})\psi(B_{k_2} + \bar{\zeta}_{k_2,(t)}Z_{k_2})\big], \tag{3.32}$$

where (Z_{k_1}, Z_{k_2}) *is a standard normal random 2-vector.*

The proof of Lemma 2 is a direct application of Lemma 1 (see Section B.3). When writing in the matrix form, the conditional joint asymptotic normality $\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi} \sim N\Big((u_{k_1}, u_{k_2}) \otimes \gamma_0, \Sigma_{(t)} \otimes \boldsymbol{I}_{p\times p}\Big)$ holds for $k_1, k_2 = 1, \ldots, K$, where $\widetilde{\boldsymbol{\xi}} = (\widetilde{\beta}_{k_1,(t)}, \widetilde{\beta}_{k_2,(t)})$, $\boldsymbol{\xi} = (1,1) \otimes \beta_0$, and $\boldsymbol{I}_{p\times p}$ is a $p \times p$ identity matrix. The covariance matrix $\Sigma_{(t)}$ has (k_1, k_2)th component $\Sigma_{k_1,k_2,(t)} = \bar{\zeta}_{k_1,k_2,(t)}$, which can be estimated by (3.31).

Under homoscedasticity where $\gamma_0 = \mathbf{0}$, $\widetilde{\beta}_{k_1}$ and $\widetilde{\beta}_{k_2}$ have a common center β_0. In addition, variance components $\bar{\zeta}^2_{k,(t)}, k = 1, \ldots, K$, are inflated under heteroscedasticity. This phenomenon can be observed by the inflation of the standard deviation $\bar{\sigma}^2_{k,(t)}$ in (3.19), which is caused by the extra heteroscedastic parameter in (3.20).

4. HETEROSCEDASTICITY TEST

Although the asymptotic normality of $\widetilde{\boldsymbol{\xi}}$ holds for all iterations $t = 0, 1, \ldots$ of the AMP algorithm, we use the estimators $\widetilde{\boldsymbol{\xi}}$ at $t \to \infty$ to construct test statistics. When the iteration $t \to \infty$, the l_2-norm of the difference of the AMP output and the estimator in (2.7) converges to zero almost surely (see Bradic 2016, Thm. 2). In practice, we consider a stopping rule and define the algorithm's convergence in

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

Section 3.1 under (3.18). At the stop, we obtain an estimator approximating the estimator defined in (2.7). Hence, the iteration index t is dropped in the following sections. We consider testing the components of the vector γ_0

$$H_{0,j} : \gamma_{0,j} = 0 \text{ versus } H_{1,j} : \gamma_{0,j} \neq 0, j = 1, \ldots, p. \tag{4.33}$$

Lemma 2 states that the K expectile estimators follow a multivariate Gaussian distribution asymptotically with a common center β_0 under homoscedasticity. And the rows of $\widetilde{\boldsymbol{\xi}}$ are i.i.d. following a joint Gaussian distribution with covariance matrix Σ.

For $\gamma_{0,j}$, the test statistic is defined as

$$T_j = \frac{\Delta \widetilde{\boldsymbol{\xi}}_j^\top}{\sqrt{\Delta \Sigma \Delta^\top}}, j = 1, \ldots, p. \tag{4.34}$$

Following Newey and Powell (1987, Sect. 4.1), we consider the contrast between two expectile levels for the testing problem, that is, $\widetilde{\boldsymbol{\xi}} = (\widetilde{\beta}_{\tau_1}, \widetilde{\beta}_{\tau_2})$, $\Delta = (1, -1)$, and the components of the covariance matrix Σ are replaced by their estimators in (3.31). Then, by Lemma 2, the test statistic T_j follows a standard normal distribution $N(0, 1)$. For the two-sided alternative in (4.33), the p-value can be calculated by

$$P_j = 2 \left\{ 1 - \Phi \left(\left| \frac{\Delta \widetilde{\boldsymbol{\xi}}_j^\top}{\sqrt{\Delta \Sigma \Delta^\top}} \right| \right) \right\}, \tag{4.35}$$

with rejection region $R = (-\infty, \Phi^{-1}(\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), \infty)$, where Φ denotes the cumulative distribution function of a standard normal distribution $N(0, 1)$. The null hypothesis is rejected if $P_j \leq \alpha$ at nominal level α. To analyze the power function on the population scale when $n, p \to \infty$, we consider for any jth component $\vartheta(\gamma_{0,j}) = I\big(T_j \in R, \Gamma_0 = \gamma_{0,j}\big)$. Then, the limiting expression of the function ϑ is obtained in Theorem 2 with proof in Appendix B.5.

THEOREM 2. *The function denoted by* $\vartheta(\gamma_{0,j}), j = 1, \ldots, p$ *complies with*

$$\lim_{p \to \infty} \frac{1}{p} \sum_{j=1}^{p} \vartheta(\gamma_{0,j}) \overset{a.s.}{=} \tag{4.36}$$

$$\mathbb{E}_{\Gamma_0} \Big[1 - \Phi\big(\Phi^{-1}(1-\alpha/2) - \Gamma_0 \frac{\Delta(u_{\tau_1}, u_{\tau_2})^\top}{\sqrt{\Delta \Sigma \Delta^\top}}\big) + \Phi\big(\Phi^{-1}(\alpha/2) - \Gamma_0 \frac{\Delta(u_{\tau_1}, u_{\tau_2})^\top}{\sqrt{\Delta \Sigma \Delta^\top}}\big) \Big].$$

Theorem 2 can be used to confirm the size and power of the proposed test. Assume the nominal significance level of the test is α. The size of the test $P(T_j \in R | \Gamma_0 = 0)$ is approximated by

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

$$\frac{\lim_{p\to\infty}\frac{1}{p}\sum_{j=1}^{p} I(T_j \in R, \Gamma_0 = 0)}{P(\Gamma_0 = 0)}$$
$$\overset{a.s.}{=} \frac{\mathbb{E}_{\Gamma_0}\Big[1-\Phi\big(\Phi^{-1}(1-\alpha/2)-0\big)+\Phi\big(\Phi^{-1}(\alpha/2)-0\big) \mid \Gamma_0 = 0\Big]P(\Gamma_0 = 0)}{P(\Gamma_0 = 0)}$$
$$= \alpha,$$

where the conditional expectation is obtained by (4.36) at $\Gamma_0 = 0$. This confirms that the size of a nominal α test is α asymptotically.

Further, the empirical size of the test is defined in (6.37) and used to evaluate the performance of the proposed test. As presented in Tables 1 and 2, the empirical size, averaged over the components of γ_0, is approximately equal to the nominal level α. Similarly, the power of the test is obtained by considering the nonnull components $\gamma_{0,j}$'s. However, from the right-hand side of (4.36), we notice that the power of the test depends on the averaged signal strength of Γ_0, the expectiles (u_{τ_1}, u_{τ_2}), and the covariance matrix Σ. The empirical power of the test defined in (6.38) is also used to evaluate the numerical performance of the test in Section 6.

5. DECORRELATION IN AMP

It is well known that the standard Gaussian design assumption, which is stated in Assumption (A4), is important for the AMP framework. Researchers have made several attempts to relax this assumption. One direction is to relax the Gaussian distribution assumption. For example, Bayati, Lelarge, and Montanari (2015) and Chen and Lam (2021) showed that the AMP framework could be extended to other random matrices; Rangan, Schniter, and Fletcher (2019) and Fan (2022) discussed the extension of AMP algorithms to right rotationally invariant design. Li and Sur (2026) showed the statistical applications of these algorithms. However, the independence assumption is still essential in these works. A more relevant direction for this work is to relax the independence assumption in order to handle general design matrices with arbitrary covariance matrices. Consider a general design matrix in the sparse high-dimensional heteroscedastic linear model: $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\beta + \tilde{\boldsymbol{\varepsilon}}$, where $\tilde{\boldsymbol{Y}} \in \mathbb{R}^n$, $\tilde{\boldsymbol{X}} \in \mathbb{R}^{n\times p}$, $\tilde{\boldsymbol{\varepsilon}} \in \mathbb{R}^n$, and the rows of $\tilde{\boldsymbol{X}}$ are i.i.d. from a multivariate Gaussian distribution $N_p(0, \tilde{\Sigma})$. Several papers have discussed incorporating a general Gaussian design with strictly positive-definite covariance matrices. For example, Donoho and Montanari (2016, Def. 4.5, Cor. 4.6), Candès and Sur (2020, Sect. 3.1.1), Zhao et al. (2022, Sect. 2), and Huang (2022) have proposed a method to decorrelate the design matrix using the nonsingular covariance matrix $\tilde{\boldsymbol{\Sigma}}$ by applying the transformation $(\tilde{\boldsymbol{X}}\tilde{\Sigma}^{-1/2})/\sqrt{n}$. This decorrelation allows for satisfying Assumption (A4) and expands the applicability of AMP-based methods and theories. However, this decorrelation approach has limitations when dealing with sparse parameter vectors β because the null components of β become nonnull under the linear transformation $\sqrt{n}\tilde{\boldsymbol{\Sigma}}^{1/2}\beta$ and the corresponding back

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

transformation. Additionally, in practical applications, the covariance matrix $\tilde{\boldsymbol{\Sigma}}$ is often unknown, and estimating it becomes challenging when $p > n$ unless certain structure assumptions are made. These challenges restrict the use of this approach in real applications.

In applications, we consider a different method for decorrelation in the AMP. The validity of this type of decorrelation has been tested numerically in Zhou and Claeskens (2023, Table 3). We utilize a "generalized puffer transformation" on the data, resulting in transformed variables denoted as $\boldsymbol{Y} = F\tilde{\boldsymbol{Y}}$ and $\boldsymbol{X} = F\tilde{\boldsymbol{X}}$, where the transformed data satisfy the model $\boldsymbol{Y} = \boldsymbol{X}\beta + F\tilde{\boldsymbol{\varepsilon}}$. It is important to note that in our setting, $p > n$. To perform the transformation, we utilize matrices $\boldsymbol{U} \in \mathbb{R}^{n\times n}$, $\boldsymbol{D} \in \mathbb{R}^{n\times n}$, and $\boldsymbol{V}^\top \in \mathbb{R}^{n\times p}$ obtained from the singular value decomposition $\tilde{\boldsymbol{X}} = UDV^\top$. The "generalized puffer transformation" was originally proposed in Jia and Rohe (2015), where the authors aimed to decorrelate the design matrix to bypass the so-called "irrepresentable condition" for the Lasso regression. The "generalized puffer transformation" uses $F = \sqrt{p/n}\boldsymbol{U}\hat{\boldsymbol{D}}U^\top$, where $\hat{\boldsymbol{D}}$ is a diagonal matrix with diagonal elements $\hat{\boldsymbol{D}}_{ii} = I(\boldsymbol{D}_{ii} \le 1/\sqrt{n})\sqrt{n} + I(\boldsymbol{D}_{ii} > 1/\sqrt{n})/\boldsymbol{D}_{ii}, i = 1, \ldots, n$. Subsequently, we apply the AMP algorithm and conduct the hypothesis test using the transformed data $(\boldsymbol{Y}, \boldsymbol{X})$.

Unlike the decorrelation step in Donoho and Montanari (2016), Candès and Sur (2020), and Zhao et al. (2022), our approach does not involve a transformation of β. Thus, it can preserve sparse structures. When $p > n$, we cannot guarantee that $\boldsymbol{X}$ precisely satisfies Assumption (A4), but the correlation is largely reduced, as observed in the numerical study. Moreover, on the transformed data, the error term becomes $F\tilde{\boldsymbol{\varepsilon}}$. Under the null hypothesis that the model (2.3) is homoscedastic, we have verified that $F\tilde{\boldsymbol{\varepsilon}}$ is independent of the transformed design matrix $\boldsymbol{X} = \sqrt{p/n}\boldsymbol{U}V^\top$. This observation suggests that the "generalized puffer transformation" does not alter the null hypothesis. This explains why the proposed test retains its validity size, as demonstrated in our simulation study. For further details, please refer to Section 6.2.

6. NUMERICAL PERFORMANCE

6.1. Simulation

We consider simulation settings where $p = 500$ and $n = 250$. The samples X_i are generated from a standard Gaussian design, as specified in Assumption (A4). The nonnull components of the true parameter vector β_0 are generated using a standard normal $N(0, 1)$ distribution and then multiplied by 5. Each setting is replicated $R = 400$ times. The parameter vector β_0 and design matrix $\boldsymbol{X}$ are fixed across R replications using the same seed number, while the regression errors $\boldsymbol{\varepsilon}$ vary in each replication. We evaluate the test under six different error distributions: $N(0, 1)$, t_3, Lognormal $(0, 1)$, $0.9N(-0.2, 0.25) + 0.1N(1.8, 0.01)$, $0.9N(0.2, 0.25) + 0.1N(-1.8, 0.01)$, and $0.95N(0, 0.25) + 0.05N(0, 4)$. The first three error distributions are randomly generated, centered, and scaled to have a

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

standard deviation of 0.5. We aim to assess the test's performance using symmetric errors (represented by $N(0,1)$), heavy-tailed errors (represented by t_3), and fat-tailed errors (represented by Lognormal $(0,1)$). Additionally, we consider three mixture normal distributed errors. The first two, $0.9N(-2,0.25)+0.1N(18,0.01)$ and $0.9N(2,0.25)+0.1N(-18,0.01)$, model clustered outliers on the right and left sides of the main error distribution, respectively. The last distribution—$0.95N(0,0.25)+0.05N(0,4)$—adds additional variability to $N(0,0.25)$. For heteroscedasticity, we assume five components of γ_0 are nonnulls $(3,1,-5,-5,-3)^\top$, and the rest components are all zeros. Similar to β_0, the components of γ_0 are fixed throughout R replications using the same seed number.

The individual test is evaluated by the averaged size (or false positive) proportion and power (or true positive proportion). Both proportions are obtained by averaging over simulation replications and the components of γ_0, which agrees with the construction of the power function in Theorem 2. At significance level α, the size can be estimated via

$$\text{FP}(\alpha) = (p-s)^{-1}\sum_{j\in S^c}\sum_{r=1}^{R} I\{P_{r,j}\le\alpha\}/R \qquad \text{(empirical size)}, \tag{6.37}$$

and the power can be estimated by

$$\text{TP}(\alpha) = s^{-1}\sum_{j\in S}\sum_{r=1}^{R} I\{P_{r,j}\le\alpha\}/R \qquad \text{(empirical power)}. \tag{6.38}$$

For homoscedastic cases, the components of γ_0 are all zeros and $\|\gamma_0\|_0 = 0$; hence only FP values are calculated. For heteroscedasticity cases, five components of γ_0 are nonnull with magnitude $(3,1,-5,-5,-3)^\top$. The signal strength of heteroscedasticity $u_\tau\gamma_0$ depends on the error expectile u_τ. Since the errors are scaled to a standard deviation of 0.5, the heteroscedasticity signal strength is largely weakened by u_τ.

We consider two sparsity levels for fixed ratio $\delta = 0.5$, $p = 500$, and $n = 250$.

1. (High-sparsity) β_0 has $s = 5$ nonnull components.
2. (Medium-sparsity) β_0 has $s = 40$ nonnull components.

The above two settings intend to investigate the impact of sparsity. Further, we consider homoscedastic and heteroscedastic regression errors for each sparsity level, where the errors are generated from six different distributions. Simulation results are presented in Tables 1 and 2. The test statistic in (4.34) is constructed using pairs of expectile estimators $\boldsymbol{\xi} = (\widetilde{\beta}_{\tau_1}, \widetilde{\beta}_{\tau_2})$, where seven pairs—$\tau_2 = 0.8$, $\tau_1 = 0.1, 0.2, 0.6$, and $\tau_2 = 0.9$, $\tau_1 = 0.1, 0.2, 0.6, 0.8$—are tested. Since the test results for four pairs are similar, we only report two pairs—$(\tau_1,\tau_2) = (0.2, 0.8)$ and $(0.6, 0.8)$. Table 1 uses expectile levels at $(\tau_1,\tau_2) = (0.2, 0.8)$, and Table 2 uses expectile levels at $(\tau_1,\tau_2) = (0.6, 0.8)$.

In addition, under homoscedasticity of the regression error, the test statistic in (4.34) can be seen as samples from a standard normal $N(0,1)$ distribution. To

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

TABLE 1. Test ($\alpha = 0.05$) results for different distributions of ε under homoscedastic and heteroscedastic variance

$\alpha = 0.05, \tau_1 = 0.2, \tau_2 = 0.8$									
	Homoscedastic			Heteroscedastic					
ε	$N(0,1)$	t_3	Lognormal$(0,1)$	$N(0,1)$	t_3	Lognormal$(0,1)$	$N(0,1)$	t_3	Lognormal$(0,1)$
	FP(size)			FP(size)			TP(power)		
High-sparsity	0.05	0.05	0.05	0.05	0.05	0.05	0.50	0.32	0.34
Medium-sparsity	0.06	0.06	0.05	0.05	0.06	0.06	0.09	0.09	0.09
	Homoscedastic			Heteroscedastic					
ε	mixNormal 1	mixNormal 2	mixNormal 3	mixNormal 1	mixNormal 2	mixNormal 3	mixNormal 1	mixNormal 2	mixNormal 3
	FP(size)			FP(size)			TP(power)		
High-sparsity	0.06	0.05	0.05	0.06	0.05	0.05	0.60	0.32	0.34
Medium-sparsity	0.06	0.05	0.06	0.06	0.05	0.07	0.19	0.08	0.10

Note: The pair of expectile levels is $(\tau_1, \tau_2) = (0.2, 0.8)$. The top half presents results for ε following $N(0,1)$, t_3, Lognormal$(0,1)$; the bottom half for $0.9N(-0.2, 0.25) + 0.1N(1.8, 0.01)$ (mixNormal 1), $0.9N(0.2, 0.25) + 0.1N(-1.8, 0.01)$ (mixNormal 2), $0.95N(0, 0.25) + 0.05N(0, 4)$ (mixNormal 3). Each simulation setting is replicated $R = 400$ times. The average FP proportions are calculated for homoscedastic variance, where the heteroscedastic parameter vector $\gamma_0 = \mathbf{0}$. For heteroscedastic variance, where γ_0 has five nonnull components $(3, 1, -5, -5, -3)$, the average FP and TP proportions are calculated.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

TABLE 2. Test ($\alpha = 0.05$) results for different distributions of ε under homoscedastic and heteroscedastic variance

$\alpha = 0.05, \tau_1 = 0.6, \tau_2 = 0.8$									
	Homoscedastic			Heteroscedastic					
ε	$N(0,1)$	t_3	Lognormal$(0,1)$	$N(0,1)$	t_3	Lognormal$(0,1)$	$N(0,1)$	t_3	Lognormal$(0,1)$
	FP(size)			FP(size)			TP(power)		
High-sparsity	0.05	0.05	0.05	0.05	0.05	0.05	0.24	0.17	0.17
Medium-sparsity	0.06	0.06	0.05	0.06	0.06	0.06	0.05	0.06	0.06
	Homoscedastic			Heteroscedastic					
ε	mixNormal 1	mixNormal 2	mixNormal 3	mixNormal 1	mixNormal 2	mixnormal 3	mixNormal 1	mixNormal 2	mixnormal 3
	FP(size)			FP(size)			TP(power)		
High-sparsity	0.05	0.04	0.05	0.05	0.04	0.04	0.28	0.13	0.16
Medium-sparsity	0.06	0.04	0.04	0.06	0.04	0.06	0.10	0.05	0.09

Note: The pair of expectile levels is $(\tau_1, \tau_2) = (0.6, 0.8)$. The top half presents results for ε following $N(0,1)$, t_3, Lognormal$(0,1)$; the bottom half for $0.9N(-0.2, 0.25) + 0.1N(1.8, 0.01)$ (mixNormal 1), $0.9N(0.2, 0.25) + 0.1N(-1.8, 0.01)$ (mixNormal 2), $0.95N(0, 0.25) + 0.05N(0, 4)$ (mixNormal 3). Each simulation setting is replicated $R = 400$ times. The average FP proportions are calculated for homoscedastic variance, where the heteroscedastic parameter vector $\gamma_0 = \mathbf{0}$. For heteroscedastic variance, where γ_0 has five nonnull components $(3, 1, -5, -5, -3)$, the average FP and TP proportions are calculated.

confirm the theory, we also inspected the normality of the test statistic constructed using different pairs of expectile estimators. Example QQ-plots are included in Figure S.1 in the Supplementary Material (Zhou and Zou, 2025); the examples are randomly chosen among simulation replications to avoid "cherry-picking."

We make several observations from Tables 1 and 2. When comparing Tables 1 and 2, we observe that using the expectile level pair $(\tau_1, \tau_2) = (0.2, 0.8)$ leads to higher test power, which agrees with Theorem 2 that shows the power of the test depends on the difference $(u_{\tau_2} - u_{\tau_1})$. Placing two expectile levels far enough would preserve the signal strength of γ_0 to a certain extent. However, we do not suggest using two very extreme expectiles, since estimating them is more challenging than estimating those closer to the center.

When comparing the results for different error distributions, we note that the average FP proportions are close to the nominal α, which complies with Theorem 2. The average FP proportions are slightly greater than the nominal level for the mixture normal distribution $0.9N(-0.2, 0.25) + 0.1N(1.8, 0.01)$. At the same time, this error distribution tends to have the highest empirical power among all six error distributions. In medium-sparsity settings, where the number of nonnull components of β_0 is 40, estimating β_τ becomes more challenging, and it is not surprising that the power of the test in this setting dropped compared to Setting 2, where the number of nonnull components of β_0 is 5. The estimation variances diminish the signal strength when $s = 40$. This setting is very challenging for all tests, not just for ours.

6.2. Check the Decorrelation Proposal

We use simulations to check the validity of the decorrelation step proposed in Section 5. We assume an autoregressive ($AR(1)$) covariance matrix for $\tilde{\boldsymbol{X}} \sim N(0, \tilde{\boldsymbol{\Sigma}})$, where the components $\tilde{\boldsymbol{\Sigma}}_{ij} = \rho^{|i-j|}$ for $i, j = 1, \ldots, p$. The error distributions considered are $N(0, 1)$ and t_3. To assess the impact of correlation on test performance, we consider $\rho = 0.3$ and 0.7.

We see in Tables 3 and 4 that the test is still valid numerically using the transformed data by the "generalized puffer transformation." We have checked that on the transformed data, the error is independent of the covariates, which suggests that the null hypothesis remains unchanged after the transformation. Simulation results demonstrate that the size of the test is approximately equal to the nominal α. The power of the test decreases as the "generalized puffer transformation" reduces the signal-to-noise ratio. Furthermore, the power of the test decreases when the correlation parameter ρ increases from 0.3 to 0.7.

6.3. Real Data Examples

6.3.1. *International Economic Growth Data.* We apply the proposed test on the "international economic growth" data used in Belloni et al. (2013), where the authors mainly focused on variable selection and estimation. We now want

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

TABLE 3. Test ($\alpha = 0.05$) results are presented for the decorrelation procedure under both homoscedastic and heteroscedastic variance

$\alpha = 0.05, \tau_1 = 0.2, \tau_2 = 0.8$												
	$\rho = 0.3$						$\rho = 0.5$					
	Homoscedastic		Heteroscedastic				Homoscedastic		Heteroscedastic			
ε	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3
	FP(size)		FP(size)		TP(power)		FP(size)		FP(size)		TP(power)	
High-sparsity	0.05	0.05	0.05	0.05	0.40	0.26	0.05	0.05	0.05	0.05	0.33	0.23
Medium-sparsity	0.06	0.05	0.06	0.06	0.13	0.13	0.07	0.07	0.06	0.06	0.12	0.11

Note: The pair of expectile levels is $(\tau_1, \tau_2) = (0.2, 0.8)$. The covariance matrix for $\tilde{\boldsymbol{X}} \sim N(0, \tilde{\boldsymbol{\Sigma}})$ has components $\tilde{\boldsymbol{\Sigma}}_{ij} = \rho^{|i-j|}$ for $i, j = 1, \ldots, p$. The error distributions considered are $N(0,1)$ and t_3. Each simulation setting is replicated $R = 400$ times. The average FP proportions are calculated for homoscedastic variance, where the heteroscedastic parameter vector $\gamma_0 = \mathbf{0}$. For heteroscedastic variance, where γ_0 has five nonnull components $(3, 1, -5, -5, -3)$, the average FP and TP proportions are calculated.

TABLE 4. Test ($\alpha = 0.05$) results are presented for the decorrelation procedure under both homoscedastic and heteroscedastic variance

$\alpha = 0.05, \tau_1 = 0.6, \tau_2 = 0.8$												
	$\rho = 0.3$						$\rho = 0.5$					
	Homoscedastic		Heteroscedastic				Homoscedastic		Heteroscedastic			
ε	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3	$N(0,1)$	t_3
	FP(size)		FP(size)		TP(power)		FP(size)		FP(size)		TP(power)	
High-sparsity	0.05	0.06	0.05	0.06	0.16	0.12	0.06	0.05	0.05	0.05	0.13	0.10
Medium-sparsity	0.06	0.05	0.06	0.06	0.08	0.09	0.07	0.07	0.06	0.07	0.07	0.08

Note: The pair of expectile levels is $(\tau_1, \tau_2) = (0.6, 0.8)$. The covariance matrix for $\tilde{\boldsymbol{X}} \sim N(0, \tilde{\boldsymbol{\Sigma}})$ has components $\tilde{\boldsymbol{\Sigma}}_{ij} = \rho^{|i-j|}$ for $i, j = 1, \ldots, p$. The error distributions considered are $N(0,1)$ and t_3. Each simulation setting is replicated $R = 400$ times. The average FP proportions are calculated for homoscedastic variance, where the heteroscedastic parameter vector $\gamma_0 = \mathbf{0}$. For heteroscedastic variance, where γ_0 has five nonnull components $(3, 1, -5, -5, -3)$, the average FP and TP proportions are calculated.

to test whether heteroscedasticity is present in this data. The dataset is available on the website https://stuff.mit.edu/~vchern/NBER/ and consists of $n = 90$ samples. Further, we exclude the intercept term in the unprocessed dataset. The original dataset consists of 61 covariates measuring aspects, such as education, trading openness, and science policies. We further expand the covariate set by including the polynomial terms up to order 4 of all 61 covariates. The expanded dataset consists of 244 covariates with a ratio δ approximately 0.37. The primary interest is the response variable—the national growth rate of GDP per capita. We apply

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

the decorrelation procedure outlined in Section 5 before constructing the test. Similar to Section 6.1, the test statistic is calculated using two ℓ_1-regularized expectile estimators. We consider expectile estimators at level $\tau_2 = 0.8$ and at $\tau_1 = 0.2$. According to Belloni et al. (2013), the dataset is homoscedastic since no unmeasured underlying variables exist. Our test results show that none of the covariates cause heteroscedasticity across the different pairs. To assess the stability of the method when using pairs at different expectile levels, we also test $\tau_1 = 0.4, 0.5, 0.6$. The test results agree with $\tau_1 = 0.2$ for this dataset.

6.3.2. *Supermarket Data.* We also apply the proposed heteroscedasticity test to the "supermarket data" used in Lan et al. (2016). The authors extended the t-test in their study and focused on testing $\beta_{0,j} \neq 0$ in a high-dimensional setting ($p \gg n$). The dataset consists of $n = 464$ daily records, where the response variable is the number of customers, and the covariates represent the sales volume of 6,398 different products. Although the dataset is potentially heteroscedastic, the test proposed in Lan et al. (2016) is robust to heteroscedastic errors.

Similarly, we utilize the fused Kolmogorov filter (Mai and Zou, 2015) to reduce the number of covariates to $p = 928$, achieving $\delta = n/p = 0.5$. The ratio $\delta = n/p = 0.5$. Note that this reduced dimension is well beyond the theoretical limit $O(n/\log(n))$, which ensures the sure screening property of the fused Kolmogorov filter. Thus, we have high confidence that the screening step only removes noisy features and retains all important features. After applying the decorrelation procedure, we employ a pair of ℓ_1-regularized expectile estimators at $(\tau_1, \tau_2) = (0.2, 0.8)$ to construct the test. The test results indicate that 61 covariates potentially contribute to heteroscedasticity. We also tested $\tau_1 = 0.4, 0.5, 0.6$ and $\tau_2 = 0.8$. When using the pair $(\tau_1, \tau_2) = (0.6, 0.8)$, potentially 43 covariates contribute to heteroscedasticity. In addition, there are 34 common covariates detected by both $\tau_1 = 0.2$ and $\tau_1 = 0.6$. The other two pairs using $\tau_1 = 0.4, 0.6$ have more stable performance, overlapping 43 and 47 covariates with the pair using $\tau_1 = 0.2$.

7. CONCLUDING REMARKS

In this article, we propose a high-dimensional Newey–Powell heteroscedasticity test based on contrasting expectile regression at two different expectile levels. The testing procedure is based on a bias-corrected estimator directly obtained from the AMP algorithm in Section 3, which is tracked by the state evolution recursion asymptotically when $n, p \to \infty$. We clarify that we want to offer an alternative debiasing framework in addition to the existing one, which requires adding a bias-correction term to the regularized estimators. Such a well-explored debiasing framework requires estimating sparse sample precision matrices and a thorough theoretical investigation for the bias-corrected estimator, which requires a considerable amount of work. The proposed test has stable performance when γ_0 is highly sparse, meaning that only a few covariates X_j contribute to the

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

heteroscedasticity. The numerical performance suggests that the error distributions also have an impact on the size and power of the test, which is worth further investigation. Although the AMP theory is developed for multiple levels, we did not explore contrasting multiple expectiles for testing heteroscedasticity in this work. This could be interesting for future research.

Upon examining Theorem 2, we observe that the strength of the signal in γ_0 and the covariance of the two expectile estimators also impact the power of the test. An open question remains regarding the optimal choice of expectile levels.

While the current literature on the AMP framework assumes an independent Gaussian design matrix, it is important to acknowledge that this assumption can be relaxed. However, a significant challenge in AMP remains unresolved: the lack of known theories to relax the assumption of independent covariates, accounting for both numerical and theoretical aspects. It is crucial to emphasize that this challenge is not unique to our work but is a broader issue in the AMP literature. To address this limitation in practical applications, we propose a "generalized puffer transformation" to enhance the applicability of the proposed test to real-world data. Through a simulation study, we demonstrate that the resulting test maintains its validity by preserving its nominal size. However, a further theoretical investigation is necessary to understand and address this challenge, and it remains an open question for future research.

APPENDIX

A. ASSUMPTIONS

For the sake of completeness, we summarize the assumptions used in our theory.

(A1) $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. random variables with mean zero, a finite $(2\kappa - 2)$th moment for $\kappa \geq 2$.

(A2) The empirical measure of the components of the p-vector γ_0, when $p \to \infty$, converges weakly to the probability measure of a random variable Γ_0 with bounded $(2\kappa - 2)$th moment for $\kappa \geq 2$.

(A3) The empirical measure of the components of the p-vector β_0, when $p \to \infty$, converges weakly to the probability measure of a random variable B_0 with bounded $(2\kappa - 2)$th moment for $\kappa \geq 2$.

(A4) A standard Gaussian design: $X_i, i = 1, \ldots, n,$ are i.i.d. copies of $X \in \mathbb{R}^p$ with $X_j \sim N(0, 1/n), j = 1, \ldots, p,$ i.i.d. components.

(A5) For some $\kappa > 1$,

(a) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{\beta_0(p)}}(B_0^{2\kappa-2}) = \mathbb{E}_{f_{B_0}}(B_0^{2\kappa-2}) < \infty$;

(b) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{\gamma_0(p)}}(\Gamma_0^{2\kappa-2}) = \mathbb{E}_{f_{\Gamma_0}}(\Gamma_0^{2\kappa-2}) < \infty$;

(c) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{\boldsymbol{\varepsilon}(p)}}(\varepsilon^{2\kappa-2}) = \mathbb{E}_{f_\varepsilon}(\varepsilon^{2\kappa-2}) < \infty$;

(d) $\lim_{p\to\infty} \mathbb{E}_{\widehat{f}_{q_k(p)}}(B_k^{2\kappa-2}) < \infty$.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

B. PROOFS

B.1. Proof of (3.10)

Proof. The proximal operator is defined as the minimizer of the objective function $b\rho(x) + \frac{1}{2}(x-z)^2$, which, by elementary calculus, is the root of the subgradient of the objective function. The proximal operator is then obtained by solving an equation of x as follows:

$$b\partial\rho_\tau(x;u_\tau) + x - z = 0,$$

where $\partial\rho_\tau$ is the gradient of the differentiable expectile loss function ρ_τ in (3.12). Calculations follow similar steps and arguments for the two cases, that is, $x \le u_\tau$ and $x > u_\tau$. We present the calculation steps for $x \le u_\tau$ in detail. When $x \le u_\tau$, $\partial\rho_\tau(x;u_\tau) = 2(1-\tau)(x-u_\tau)$. Then, by plugging in $\partial\rho_\tau$ and rearranging $2b(1-\tau)(x-u_\tau) + x - z = 0$, we obtain

$$x = \frac{z + 2b(1-\tau)u_\tau}{2b(1-\tau)+1}.$$

The domain of z is obtained by $\frac{z+2b(1-\tau)u_\tau}{2b(1-\tau)+1} \le u_\tau$ which is $z \le u_\tau$. □

B.2. Proof of (3.16)

Proof. By (3.14), the subgradient of the effective score function $\tilde{G}_\tau(z;b)$ in (3.13) w.r.t. z follows

$$\partial_1 G_\tau(z;b) = \frac{\delta}{\omega}\left\{\frac{2b(1-\tau)}{2b(1-\tau)+1} I\{z \le u_\tau\} + \frac{2b\tau}{2b\tau+1} I\{z > u_\tau\}\right\}.$$

The parameter b is updated by finding solutions $b > 0$, s.t.

$$1 = \frac{1}{n}\sum_{i=1}^{n} \partial_1 G_\tau(z_i;b) = \frac{\delta}{\omega}\left\{\frac{1}{n}\sum_{i=1}^{n} \frac{2b(1-\tau)}{2b(1-\tau)+1} I\{z_i \le u_\tau\} + \frac{2b\tau}{2b\tau+1} I\{z_i > u_\tau\}\right\}.$$

□

B.3. Proof of Lemma 2

Proof. We first take $\tilde{\psi}_c$ in (3.26) to be $\tilde{\psi}_c(x_1,\ldots,x_{t+1},y) = \tilde{\psi}(x_{t+1},y)$. Then, (3.32) can be obtained by taking $\tilde{\psi}(x_{t+1},y) = \psi(y - x_{t+1})$ and using (3.25). □

B.4. Proof of Theorem 1

We first prove (3.29) and (3.30). By choosing $\widetilde{\psi}_c(x_1,\ldots,x_{t+1},y) = -x$ in (3.26), we obtain

$$\begin{aligned}\lim_{p\to\infty} \frac{1}{p}(\widetilde{\beta}_{k_1,(t+1),j} - \beta_{k_1,j})(\widetilde{\beta}_{k_2,(t+1),j} - \beta_{k_2,j}) &\overset{a.s.}{=} \mathbb{E}[(-\bar{\zeta}_{k_1,(t)} Z_{k_1,(t)})(-\bar{\zeta}_{k_2,(t)} Z_{k_2,(t)})] \\ &= \bar{\zeta}_{k_1,(t)}\bar{\zeta}_{k_2,(t)}\mathbb{E}[Z_{k_1,(t)} Z_{k_2,(t)}] = \bar{\zeta}_{k_1,(t)}\bar{\zeta}_{k_2,(t)}\mathrm{Cov}(Z_{k_1}, Z_{k_2}).\end{aligned}$$

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

The last equality holds since $(Z_{k_1,(t)}, Z_{k_2,(t)})$ is a standard normal random 2-vector. And we drop the iteration index t for convenience.

Similarly, we choose $\tilde{\psi}_{c'}(x_1,\ldots,x_t,y)$ and $\tilde{\psi}_{c''}(x_1,\ldots,x_t,y)$ in (3.27) to be the rescaled effective score functions $G_{k_1}(y-x_t;\cdot)$ and $G_{k_2}(y-x_t;\cdot)$. By (3.24), we obtain

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} G_{k_1}(z_{k_1,i,(t)};b_{k_1,(t)})G_{k_2}(z_{k_2,i,(t)};b_{k_2,(t)})\\
&\qquad\overset{a.s.}{=}\mathbb{E}[G_{k_1}(\varepsilon-\bar{\sigma}_{k_1}\hat{Z}_{k_1};b_{k_1,(t)})G_{k_2}(\varepsilon-\bar{\sigma}_{k_2}\hat{Z}_{k_2};b_{k_2,(t)})]\\
&\qquad=\mathbb{E}[G_{k_1}(\varepsilon+\bar{\sigma}_{k_1}\hat{Z}_{k_1};b_{k_1,(t)})G_{k_2}(\varepsilon+\bar{\sigma}_{k_2}\hat{Z}_{k_2};b_{k_2,(t)})].
\end{aligned}$$

The last equality holds by $Z\overset{d}{=}-Z$ for $Z\sim N(0,1)$.

To obtain (3.31), we first combine (3.24), (3.25), and (3.28) which leads to

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} G_{k_1}(z_{k_1,i,(t)};b_{k_1,(t)})G_{k_2}(z_{k_2,i,(t)};b_{k_2,(t)})\\
&\qquad\xrightarrow{P}\lim_{p\to\infty}\frac{1}{p}(\widetilde{\beta}_{k_1,(t+1),j}-\beta_{k_1,j})(\widetilde{\beta}_{k_2,(t+1),j}-\beta_{k_2,j}).
\end{aligned}$$

Then, when n,p are sufficiently large, we use (3.31) to approximate the covariance of $\widetilde{\beta}_{k_1,(t+1)}$ and $\widetilde{\beta}_{k_2,(t+1)}$.

B.5. Proof of Theorem 2

Proof.

$$\begin{aligned}
&\vartheta(\gamma_{0,j}) = I\{T_j\in R_j,\Gamma_0=\gamma_{0,j}\}\\
&=I\Big\{\frac{\Delta\tilde{\boldsymbol{\xi}}_j^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}>\Phi^{-1}(1-\alpha/2)\Big\}+I\Big\{\frac{\Delta\tilde{\boldsymbol{\xi}}_j^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}<\Phi^{-1}(\alpha/2)\Big\}\\
&=I\Big\{\frac{\Delta(\tilde{\boldsymbol{\xi}}_j-(u_{\tau_1},u_{\tau_2})\gamma_{0,j})^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}>\Phi^{-1}(1-\alpha/2)-\gamma_{0,j}\frac{\Delta(u_{\tau_1},u_{\tau_2})^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}\Big\}\\
&\quad+I\Big\{\frac{\Delta(\tilde{\boldsymbol{\xi}}_j-(u_{\tau_1},u_{\tau_2})\gamma_{0,j})^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}<\Phi^{-1}(\alpha/2)-\gamma_{0,j}\frac{\Delta(u_{\tau_1},u_{\tau_2})^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}\Big\}.
\end{aligned}$$

By Lemma 1 and define $\widetilde{\Xi}=(B_0+u_{\tau_1}\Gamma_0+\bar{\zeta}_1Z_1,B_0+u_{\tau_2}\Gamma_0+\bar{\zeta}_2Z_2)$, where $(\bar{\zeta}_1Z_1,\bar{\zeta}_2Z_2)\sim N(0,\Sigma)$, we obtain

$$\begin{aligned}
&\lim_{p\to\infty}\frac{1}{p}\sum_{j=1}^{p}\vartheta(\gamma_{0,j})\\
&\overset{a.s.}{=}\mathbb{E}I\Big\{\frac{\Delta(\widetilde{\Xi}-(u_{\tau_1},u_{\tau_2})\Gamma_0)^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}>\Phi^{-1}(1-\alpha/2)-\Gamma_0\frac{\Delta(u_{\tau_1},u_{\tau_2})^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}\Big\}\\
&\quad+\mathbb{E}I\Big\{\frac{\Delta(\widetilde{\Xi}-(u_{\tau_1},u_{\tau_2})\Gamma_0)^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}<\Phi^{-1}(\alpha/2)-\Gamma_0\frac{\Delta(u_{\tau_1},u_{\tau_2})^{\top}}{\sqrt{\Delta\Sigma\Delta^{\top}}}\Big\}
\end{aligned}$$

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

$$= \mathbb{E}I\{Z_3 > \Phi^{-1}(1-\alpha/2) - \Gamma_0 \frac{\Delta(u_{\tau_1}, u_{\tau_2})^\top}{\sqrt{\Delta\Sigma\Delta^\top}}\Big\} + \mathbb{E}I\{Z_3 < \Phi^{-1}(\alpha/2) - \Gamma_0 \frac{\Delta(u_{\tau_1}, u_{\tau_2})^\top}{\sqrt{\Delta\Sigma\Delta^\top}}\Big\}$$

$$= \mathbb{E}_{\Gamma_0}\Big[1 - \Phi\big(\Phi^{-1}(1-\alpha/2) - \Gamma_0 \frac{\Delta(u_{\tau_1}, u_{\tau_2})^\top}{\sqrt{\Delta\Sigma\Delta^\top}}\big) + \Phi\big(\Phi^{-1}(\alpha/2) - \Gamma_0 \frac{\Delta(u_{\tau_1}, u_{\tau_2})^\top}{\sqrt{\Delta\Sigma\Delta^\top}}\big)\Big].$$

□

C. AUXILIARY DEFINITIONS AND LEMMAS

DEFINITION 1 (Pseudo Lipschitz function). *A function $\phi : \mathbb{R}^m \to \mathbb{R}$ is pseudo-Lipschitz of order $\kappa \geq 1$, if there exists a constant $L > 0$, such that for all $x, y \in \mathbb{R}^m$,*

$$|\phi(x) - \phi(y)| \leq L(1 + \|x\|^{\kappa-1} + \|y\|^{\kappa-1})\|x - y\|.$$

It follows that if ϕ is a pseudo Lipschitz function of order κ, then there exists a constant L' such that for all $x \in \mathbb{R}^m : |\phi(x)| \leq L'(1 + \|x\|^\kappa)$.

LEMMA 3. *For any pseudo Lipschitz functions $\phi_{c'}, \phi_{c''} : \mathbb{R}^m \to \mathbb{R}$ of order $\kappa_{c'}$ and $\kappa_{c''}$, their product function $\phi_{c'}\phi_{c''}$ is a pseudo Lipschitz function of order $\kappa_{c'} + \kappa_{c''}$.*

LEMMA 4 (Bayati and Montanari (2011a, Lemma 2)). *For any deterministic unit vectors $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^n$, and a random matrix $\tilde{\boldsymbol{X}}$ distributed as $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, the following conclusions hold:*

1. $v^\top \tilde{\boldsymbol{X}} u \stackrel{d}{=} Z/\sqrt{n}$, *where* $Z \sim N(0,1)$.
2. $\lim_{n\to\infty} \|\tilde{\boldsymbol{X}} u\|^2 = 1$ *almost surely.*
3. *Consider a d-dimensional subspace $\mathbb{S} \in \mathbb{R}^n$ with orthonormal basis $s_1, \ldots, s_d$ and $\|s_i\|^2 = n, i = 1, \ldots, d$. Denote the matrix $\boldsymbol{D} = (s_1, \ldots, s_d)$ and the orthogonal projection onto $\mathbb{S}$ as $P_{\mathbb{S}}$. Then, we have $P_{\mathbb{S}} \tilde{\boldsymbol{X}} u \stackrel{d}{=} \boldsymbol{D}x$ with $x \in \mathbb{R}^d$ satisfying $\lim_{n\to\infty} \|x\| \stackrel{a.s.}{=} 0$ and $x = o_d(1)$.*

LEMMA 5 (Jameson (2014, Thm. 1)). *If $x_i \geq 0$ where $i = 1, \ldots, n$ and $p \geq 1$, then $\sum_{i=1}^n x_i^p \leq (\sum_{i=1}^n x_i)^p \leq n^{p-1} \sum_{i=1}^n x_i^p$. The reversed inequality holds for $p \in (0,1)$.*

LEMMA 6 (Bayati and Montanari (2011a, Lemma 4)). *Let $\kappa \geq 2$ and a sequence of vectors $\{v(N)\}_{N\geq 0}$ whose empirical distribution converges weakly to probability measure f_V on $\mathbb{R}$ with bounded κth moment; additionally, assume that $\lim_{p\to\infty} \mathbb{E}_{\widehat{f_v}}(V^\kappa) = \mathbb{E}_{f_V}(V^\kappa)$. Then for any pseudo Lipschitz function $\psi : \mathbb{R} \to \mathbb{R}$ of order κ,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^N \psi(v_j) \stackrel{a.s.}{=} \mathbb{E}[\psi(V)].$$

LEMMA 7. *For any fixed unit vectors $u \in \mathbb{R}^p$, $v \in \mathbb{R}^n$ and random matrices $\tilde{\boldsymbol{X}}$ distributed as $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, it holds that $\tilde{\boldsymbol{X}} u = \left(\sum_{j=1}^p \tilde{\boldsymbol{X}}_{1,j} u_j, \ldots, \sum_{j=1}^p \tilde{\boldsymbol{X}}_{n,j} u_j\right)^\top \stackrel{d}{=} \frac{1}{\sqrt{n}}(Z_1, \ldots, Z_n)^\top \sim N(0, \frac{1}{n} I_n)$, where Z_i's are i.i.d. samples of a standard normal $N(0,1)$ distributed random*

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

variable Z. *Similarly,* $\tilde{\boldsymbol{X}}^\top v = \left(\sum_{i=1}^n \tilde{\boldsymbol{X}}_{i,1} u_i, \ldots, \sum_{i=1}^n \tilde{\boldsymbol{X}}_{i,p} u_i\right)^\top \overset{d}{=} \frac{1}{\sqrt{n}}(Z'_1, \ldots, Z'_p)^\top \sim N(0, \frac{1}{n} I_p)$, *where* Z'_j*'s are i.i.d. samples of a standard normal* $N(0,1)$ *distributed random variable* Z'.

LEMMA 8 (Shi et al. (2012, Thm. 5)). *Assume any* $m_1, m_2 \in \mathbb{R}^n$ *and any random Gaussian matrix* $\tilde{\boldsymbol{X}}$ *distributed as* $\tilde{\boldsymbol{X}} \in \mathbb{R}^{n\times p}$. *For any* $\epsilon \in (0,1)$, *if the inner product* $\langle m_1, m_2\rangle > 0$, *then with probability at least* $1 - 6\exp(-\frac{n}{2}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}))$, *the following holds:*

$$\frac{1+\epsilon}{1-\epsilon}\frac{\langle m_1, m_2\rangle}{\|m_1\|\|m_2\|} - \frac{2\epsilon}{1-\epsilon} \le \frac{\langle \tilde{\boldsymbol{X}}^\top m_1, \tilde{\boldsymbol{X}}^\top m_2\rangle}{\|\tilde{\boldsymbol{X}}^\top m_1\|\|\tilde{\boldsymbol{X}}^\top m_2\|} \le 1 - \frac{\sqrt{1-\epsilon^2}}{1+\epsilon} + \frac{\epsilon}{1+\epsilon} + \frac{1-\epsilon}{1+\epsilon}\frac{\langle m_1, m_2\rangle}{\|m_1\|\|m_2\|}.$$

COMPETING INTEREST STATEMENT

The authors declare that no competing interests exist.

FUNDING STATEMENT

H.Z. is supported in part by NIH 1R01GM163244-01.

SUPPLEMENTARY MATERIAL

Zhou, J., & Zou, H. (2025). Supplement to "High-dimensional Newey–Powell test via approximate message passing," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466626100450.

REFERENCES

Bayati, M., Erdogdu, M., & Montanari, A. (2013). Estimating LASSO risk and noise level. In C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 944–952). Curran Associates.

Bayati, M., Lelarge, M., & Montanari, A. (2015). Universality in polytope phase transitions and message passing algorithms. *The Annals of Applied Probability*, 25(2), 753–822.

Bayati, M., & Montanari, A. (2011a). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2), 764–785.

Bayati, M., & Montanari, A. (2011b). The LASSO risk for Gaussian matrices. *IEEE Transactions on Information Theory*, 58(4), 1997–2017.

Belloni, A., & Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*, 39(1), 82–130.

Belloni, A., Chernozhukov, V., & Hansen, C. (2013). Inference for high-dimensional sparse econometric models. In D. Acemoglu, M. Arellano, & E. Dekel (Eds.), *Advances in economics and econometrics: Tenth world congress* (Vol. 3, pp. 245–295). Cambridge University Press.

Belloni, A., Chernozhukov, V., & Wang, L. (2014). Pivotal estimation via square-root lasso in nonparametric regression. *The Annals of Statistics*, 42(2), 757–788.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

Bradic, J. (2016). Robustness in sparse high-dimensional linear models: Relative efficiency and robust approximate message passing. *Electronic Journal of Statistics*, 10(2), 3894–3944.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 47(5), 1287–1294.

Bühlmann, P., & van De Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media.

Candès, E., & Sur, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *The Annals of Statistics*, 48(1), 27–42.

Candès, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6), 2313–2351.

Cattaneo, M. D., Jansson, M., & Newey, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association*, 113(523), 1350–1361.

Chen, W.-K., & Lam, W.-K. (2021). Universality of approximate message passing algorithms. *Electronic Journal of Probability*, 26, 1–44.

Daye, Z. J., Chen, J., & Li, H. (2012). High-dimensional heteroscedastic regression with an application to eQTL data analysis. *Biometrics*, 68(1), 316–326.

Donoho, D., & Johnstone, I. (1994). Minimax risk over ℓ_p-balls for ℓ_q-error. *Probability Theory and Related Fields*, 99(2), 277–303.

Donoho, D., & Johnstone, I. (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26(3), 879–921.

Donoho, D., Maleki, A., & Montanari, A. (2009). Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45), 18914–18919.

Donoho, D., Maleki, A., & Montanari, A. (2010). Message passing algorithms for compressed sensing: I. Motivation and construction. In *2010 IEEE information theory workshop on information theory (ITW 2010, Cairo)* (pp. 1–5). IEEE. https://arxiv.org/pdf/0911.4219.

Donoho, D., & Montanari, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166(3), 935–969.

El Karoui, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probability Theory and Related Fields*, 170, 95–175.

El Karoui, N., Bean, D., Bickel, P., Lim, C., & Yu, B. (2013). On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36), 14557–14562.

Fan, J., Li, R., Zhang, C.-H., & Zou, H. (2020). *Statistical foundations of data science*. Chapman and Hall/CRC.

Fan, J., & Lv, J. (2008). Sure Independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849–911.

Fan, Z. (2022). Approximate message passing algorithms for rotationally invariant matrices. *Annals of Statistics*, 50(1), 197–224.

Feng, O., Venkataramanan, R., Rush, C., & Samworth, R. J. (2022). A unifying tutorial on approximate message passing. *Foundations and Trends® in Machine Learning*, 15(4), 335–536.

Gu, Y., Fan, J., Kong, L., Ma, S., & Zou, H. (2018). ADMM for high-dimensional sparse penalized quantile regression. *Technometrics*, 60(3), 319–331.

Gu, Y., & Zou, H. (2016). High-dimensional generalizations of asymmetric least squares regression and their applications. *The Annals of Statistics*, 44(6), 2661–2694.

He, X., Pan, X., Tan, K. M., & Zhou, W. (2023). Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2), 367–388.

Huang, H. (2020). Asymptotic risk and phase transition of l_1-penalized robust estimator. *Annals of Statistics*, 48(5), 3090–3111.

Huang, H. (2022). LASSO risk and phase transition under dependence. *Electronic Journal of Statistics*, 16(2), 6512–6552.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

Jameson, G. (2014). Some inequalities for $(a+b)p$ and $(a+b)p+(a-b)p$. *The Mathematical Gazette*, 98(541), 96–103.

Javanmard, A., & Montanari, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2), 115–144.

Javanmard, A., & Montanari, A. (2014). Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory*, 60(10), 6522–6554.

Jia, J., & Rohe, K. (2015). Preconditioning the lasso for sign consistency. *Electronic Journal of Statistics*, 9(1), 1150–1172. https://doi.org/10.1214/15-JS1029.

Jochmans, K. (2022). Heteroscedasticity-robust inference in linear regression models with many covariates. *Journal of the American Statistical Association*, 117(538), 887–896.

Koenker, R. (2005). *Quantile regression*, Econometric Society Monographs. Cambridge University Press.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, 9, 155–176.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1), 33–50.

Koenker, R., & Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, 50(1), 43–61.

Lan, W., Zhong, P.-S., Li, R., Wang, H., & Tsai, C.-L. (2016). Testing a single regression coefficient in high dimensional linear models. *Journal of Econometrics*, 195(1), 154–168.

Li, Y., & Sur, P. (2026). Spectrum-aware debiasing: A modern inference framework with applications to principal components regression. *The Annals of Statistics*, 54(2), 745–770. https://doi.org/10.1214/25-AOS2586.

Mai, Q., & Zou, H. (2015). The fused Kolmogorov filter: A nonparametric model-free screening method. *Annals of Statistics*, 43(4), 1471–1497.

Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, 55(4), 819–847.

Rangan, S., Schniter, P., & Fletcher, A. K. (2019). Vector approximate message passing. *IEEE Transactions on Information Theory*, 65(10), 6664–6684.

Shi, Q., Shen, C., Hill, R., & Van Den Hengel, A. (2012). Is margin preserved after random projection? In J. Langford & J. Pineau (Eds.), *Proceedings of the 29th international conference on machine learning, Edinbrugh, Scotland, June 27-July 3* (pp. 643–650). Omnipress.

Su, W., Bogdan, M., & Candès, E. (2017). False discoveries occur early on the lasso path. *The Annals of Statistics*, 45(5), 2133–2150.

Sur, P., & Candès, E. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29), 14516–14525.

Sur, P., Chen, Y., & Candès, E. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probability Theory and Related Fields*, 175(1), 487–558.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

van de Geer, S., Bühlmann, P., Ritov, Y., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3), 1166–1202.

Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107(497), 214–222.

Weinstein, A., Barber, R., & Candès, E. (2017). *A power and prediction analysis for knockoffs with lasso statistics*. Preprint, arXiv:1712.06465.

Weinstein, A., Su, W. J., Bogdan, M., Barber, R. F., & Candès, E. J. (2023). A power analysis for model-X knockoffs with ℓ_p-regularized statistics. *The Annals of Statistics*, 51(3), 1005–1029. https://doi.org/10.1214/23-AOS2274.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4), 817–838.

Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 801–817.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894–942.

Zhao, Q., Sur, P., & Candès, E. (2022). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli*, 28(3), 1835–1861.

Zheng, Q., Gallagher, C., & Kulasekera, K. (2013). Adaptive penalized quantile regression for high dimensional data. *Journal of Statistical Planning and Inference*, 143(6), 1029–1038.

Zhou, J., & Claeskens, G. (2023). Automatic bias correction for testing in high-dimensional linear models. *Statistica Neerlandica*, 77(1), 71–98.

Zhou, J., Claeskens, G., & Bradic, J. (2020). Detangling robustness in high dimensions: Composite versus model-averaged estimation. *Electronic Journal of Statistics*, 14(2), 2551–2599.

Zhou, J., & Zou, H. (2025). Supplement to "High-dimensional Newey–Powell Test Via Approximate Message Passing". *Econometric Theory Supplementary Material*. https://doi.org/10.1017/S0266466626100450.

Zhou, L., & Zou, H. (2023). Cross-fitted residual regression for high-dimensional heteroscedasticity pursuit. *Journal of the American Statistical Association*, 118(542), 1056–1065.

Zou, H. (2006). The adaptive lasso and its Oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.

https://doi.org/10.1017/S0266466626100450 Published online by Cambridge University Press