UNIVERSIDAD DE SEVILLA
Facultad de Matemáticas
Departamento de Estadística e Investigación Operativa

APROXIMACIÓN A LA DISTRIBUCIÓN DE CIERTOS ESTADÍSTICOS EN CONTRASTES SOBRE MODELOS DE REGRESIÓN NO PARAMÉTRICA

Tesis Doctoral

Gustavo Ignacio Rivas Martínez

Directora

María Dolores Jiménez Gamero
Departamento de Estadística e Investigación Operativa
Facultad de Matemáticas
Universidad de Sevilla

Sevilla, Febrero de 2017

UNIVERSIDAD DE SEVILLA
Facultad de Matemáticas
Departamento de Estadística e Investigación Operativa

APROXIMACIÓN A LA DISTRIBUCIÓN DE CIERTOS ESTADÍSTICOS EN CONTRASTES SOBRE MODELOS DE REGRESIÓN NO PARAMÉTRICA

V° B°: DIRECTORA DEL TRABAJO

Memoria presentada por Gustavo Ignacio Rivas Martínez, para optar al grado de Doctor en Matemáticas.

Fdo. María Dolores Jiménez Gamero.
Profesora Titular de Universidad.

Fdo. Gustavo Ignacio Rivas Martínez.

Sevilla, Febrero de 2017

Resumen

Este trabajo se desarrolla en el contexto de modelos de regresión no paramétrica con diseño aleatorio para variable respuesta y covariable unidimensionales. Específicamente, sean (X_k, Y_k), $1 \leq k \leq d$, vectores aleatorios independientes e idénticamente distribuidos (IID), que satisfacen el modelo,

$$Y_k = m_k(X_k) + \sigma_k(X_k)\varepsilon_k,$$

donde $m_k(x) = E(Y_k \mid X_k = x)$ es la función de regresión, $\sigma_k^2(x) = Var(Y_k \mid X_k = x)$ es la varianza condicional y ε_k es el error de regresión o simplemente error. A lo largo de esta memoria se supone que ε_k y X_k son variables aleatorias independientes.

El conocimiento de la función de distribución del error puede mejorar varios procedimientos estadísticos realizados sobre el modelo considerado. Por otro lado, la hipótesis de igualdad de distribución de los errores es asumida en algunos procedimientos. En otros casos, cuando los errores tienen igual distribución, algunos procedimientos se simplifican considerablemente. Sobre el contraste de estas dos hipótesis se desarrolla este trabajo.

Para el test de bondad de ajuste de la distribución del error, nos centramos en un estadístico propuesto en la literatura. El estadístico es una norma L_2 de la diferencia entre una estimación no paramétrica de la función característica de la distribución del error y una estimación paramétrica de la función característica de la distribución del error bajo la hipótesis nula. La distribución nula asintótica del estadístico es desconocida. Por ese motivo, se ha propuesto un bootstrap paramétrico para aproximarla. Esta aproximación posee muy buenas propiedades, entre otras cosas, proporciona un estimador consistente de la distribución nula asintótica y además es muy fácil de implementarlo. Sin embargo, a medida que el número de parámetros aumenta y/o el tamaño muestral crece, el coste computacional que requiere su aplicación práctica, aumenta considerablemente.

Para el contraste de igualdad de las distribuciones del error, nos hemos centrado en dos estadísticos propuestos en la literatura. Los estadísticos están basados en la estimación

no paramétrica de la función de distribución de los errores en cada población y una estimación no paramétrica de la distribución común bajo la hipótesis nula. La distribución nula asintótica de los estadísticos es desconocida. Se recurre a una aproximación mediante un bootstrap suavizado para aproximarlas. Este estimador proporciona estimaciones consistentes de la distribución nula, pero desde el punto de vista computacional, es poco eficiente. Además, su aplicación requiere ciertas condiciones sobre la distribución de los errores: han de poseer distribución contínua satisfaciendo fuertes condiciones de suavidad.

En este trabajo, del estudio de los problemas mencionados, se derivan los siguientes resultados concretos. Se ha demostrado teóricamente la consistencia de una aproximación bootstrap ponderada a los estadísticos estudiados tanto para la bondad de ajuste de error como para la igualdad en las distribuciones del error. Esta consistencia es en el sentido de que asintóticamente son capaces de estimar correctamente el error tipo I. Así mismo, en los algoritmos de implementación de la aproximación propuesta, se evidencia que el remuestreo se hace con cálculos rápidos. Además, para contrastar la hipótesis de igualdad en las distribuciones del error, se ha propuesto un nuevo test. El estadístico de contraste está basado en comparar estimadores de la función característica de la distribución de los errores. Se construye bajo condiciones menos restrictivas que las asumidas para los dos estadísticos ya existentes considerados. La distribución nula asintótica del estadístico es desconocida y se estudia teóricamente la aproximación mediante un bootstrap ponderado. Al igual que en los casos anteriores, esta aproximación es consistente.

En el desarrollo de todo el trabajo se han utilizado los estimadores del tipo Nadaraya-Watson para la función de regresión y la varianza condicional. La metodología utilizada para la obtención de los resultados teóricos, pueden servir de base para ser extendida a otros tipo de estimadores (por ejemplo, polinómicos locales, Fan y Gijbels 1996). La posibilidad de conseguir resultados teóricos similares a los obtenidos, basados en otros tipo de estimadores para los residuos estará asociado a la consistencia del estimador y la imposicion de condiciones adecuadas a cada caso. En tal sentido, las expansiones asintóticas de los test tratados aquí podrían presentar otras expresiones. Sin embargo, varios de los argumentos, propiedades o resultados conseguidos pueden servir como una guía para derivar en resultados análogos.

Agradecimientos

Este trabajo no podría haberlo culminado sin la dirección de la Profesora Doctora María Dolores Jiménez Gamero. Sin duda, se ha convertido en mi modelo de investigador que espero algún día llegar a ser. Agradecerle por el tiempo dedicado, los conocimientos compartidos, las correcciones hechas y sobre todo por la paciencia ante mis limitaciones.

No encuentro las palabras exactas para expresar la gratitud que siento por el apoyo incondicional de mi querida esposa Marta. Ella ha hecho muchos sacrificios para acompañarme en este sueño. Si no he renunciado, ha sido por ella y por mis amados hijos Ezequiel y Analia.

Agradecer con el corazón el gran apoyo de mis padres, Aníbal y Mercedes. Sus testimonios de vida me motivan a superarme cada día. Espero que lleguen a estar al menos la mitad de orgulloso que yo estoy de ellos.

Así mismo, agradecer al Director del Departamento de Estadística e Investigación Operativa, Profesor Doctor Juan Luis Moreno Rebollo, con quien había contactado para iniciar este proyecto y desde ese entonces hasta ahora siempre me ha demostrado su entera predisposición y gran amabilidad.

Quiero destacar la gran ayuda que me ha brindado desde la secretaría del departamento, Francisco Méndez. Siempre estuvo dispuesto a encontrar alguna solución cuando lo necesitaba. Gracias por todo Paco.

A las instituciones que han solventado financieramente la concreción de este proyecto. A la Fundación Carolina, la Universidad de Sevilla, la Universidad Nacional de Asunción, CONACYT-Paraguay mediante el proyecto PROCIENCIA con fondos del FONACIDE y a la Asociación Universitaria Iberoamericana de Postgrado mediante su programa de Mobilidad Académica entre Universidades de Andalucía e Iberoamérica.

A mis padres, Aníbal y Mercedes.

Índice General

Índice de Cuadros

Capítulo 1

Introducción

Los modelos de regresión tratan de explicar el comportamiento de una variable respuesta, Y, mediante una o varias variables independientes o covariables, X. Esta memoria se centra en el contexto de los modelos de regresión no paramétrica con diseño para variable respuesta y covariable unidimensionales. Específicamente, sean (X_k, Y_k), $1 \leq k \leq d$, vectores aleatorios independientes e idénticamente distribuidos (IID) que satisfacen el modelo,

$$Y_k = m_k(X_k) + \sigma_k(X_k)\varepsilon_k, \tag{1.1}$$

donde $m_k(x) = E(Y_k \mid X_k = x)$ es la función de regresión, $\sigma_k^2(x) = Var(Y_k \mid X_k = x)$ es la varianza condicional y ε_k es el error de regresión o simplemente error. A lo largo de esta memoria supondremos que ε_k y X_k son variables aleatorias independientes. Nótese que por construcción la esperanza y la varianza de los errores son, $E(\varepsilon_k) = 0$ y $Var(\varepsilon_k) = 1$, respectivamente, $1 \leq k \leq d$.

1.1 Contrastes sobre modelos de regresión

Diferentes cuestiones pueden ser sometidas a un contraste de hipótesis en los modelos de regresión. A continuación se presentan algunas de ellas. Para una revisión más completa y detallada se recomienda consultar el artículo de González-Manteiga y Crujeiras (2013).

1.1.1 Tests de bondad de ajuste para una función de regresión paramétrica

En este caso la hipótesis de interés es,

$$H_0 : m \in \mathcal{M}_\theta = \{m_\theta\},$$

contra la hipótesis alternativa,

$$H_1 : m \notin \mathcal{M}_\theta,$$

donde $\mathcal{M}_\theta$ es una familia paramétrica de funciones de regresión, con $\theta \in \Theta \subset \mathbb{R}^p$. La literatura en este ámbito es muy abundante. Por lo cual, citaremos sólo algunos trabajos a modo de ejemplo. Kozek (1991) y Härdle y Mammen (1993), para covariable multivariante, proponen estadísticos basados en comparar una estimación no paramétrica de la función de regresión y una estimación paramétrica. Stute y González-Manteiga (1996) contrastan que la función de regresión proviene de una familia de funciones lineales contra la alternativa que $m(\cdot)$ es una función no lineal, para covariable univariante y suponiendo homocedasticidad. En Koul y Ni (2004) se propone un test para el caso de covariables multidimencionales y suponiendo heterocedasticidad. En Van Keilegom et. al (2008) proponen un test basado en una distancia entre la función de distribución empírica de los residuos y una estimación paramétrica de la misma bajo H_0. En el mismo escenario, Hušková y Meintanis (2009), proponen un test similar al anterior pero basado en la función característica empírica de los residuos.

1.1.2 Tests de bondad de ajuste para la varianza condicional paramétrica

En este caso, la hipótesis nula a tratar es la siguiente

$$H_0 : \sigma^2 \in \mathcal{M}_\theta = \{\sigma^2_\theta\},$$

contra la hipótesis alternativa,

$$H_1 : \sigma^2 \notin \mathcal{M}_\theta.$$

donde $\mathcal{M}_\theta$ es una familia paramétrica de funciones de varianza condicional con $\theta \in \Theta \subset \mathbb{R}^p$. Un caso especial de la hipótesis nula es $\sigma^2 = \theta$, es decir, la hipótesis de homocedasticidad, tratado en Diblasi y Bowman (1997), Dette y Munk (1998), Liero (2003) y Zhu et al. (2001), entre otros. Para una especificación más general, la literatura es más bien escasa.

En Dette et al. (2007) proponen un test cuyo estadístico se basa en la comparación entre la función de distribución empírica de los residuos y la función de distribución empírica de unos residuos calculados bajo la hipótesis nula.

1.1.3 Comparación de funciones de regresión

Cuando se tienen $d \geq 2$ poblaciones, se puede estar interesado en contrastar la igualdad de las d funciones de regresión, es decir, contrastar la siguiente hipótesis nula

$$H_0 : m_1 = \ldots = m_d,$$

contra la hipótesis alternativa,

$$H_1 : m_s \neq m_t, \text{ para algunos } 1 \leq s, t \leq d.$$

Delgado (1993) propone un test de igualdad entre dos funciones de regresión no paramétricas que no necesita parámetro de suavizado. En Koul y Schick (1997) y Neumeyer y Pardo-Fernández (2009) se considera el problema de la igualdad entre dos curvas de regresión en modelos no paramétricos homocedástico y heterocedástico, respectivamente, frente a una alternativa unidireccional. Neumeyer y Dette (2003) consideran la hipótesis de igualdad entre dos funciones de regresión en modelos heterocedásticos con covariable univariante, basados en procesos empíricos de los residuos. Pardo-Fernández et al. (2007) proponen un test para comparar d funciones de regresión. El test está basado en la diferencia entre una estimación no paramétrica de la distribución de los errores en cada población y una estimación no paramétrica de la distribución común de los errores bajo la hipótesis nula. Basados en la metología anterior, Pardo-Fernández et al. (2015a) proponen un test basado en la función característica. La ventaja con este planteamiento, frente al anterior, es que el método puede ser aplicado no sólo a distribuciones contínuas para el error, sino que para casos más generales. Park y Kang (2008) proponen un test gráfico para la igualdad de dos funciones de regresión y extienden al caso de más de dos curvas de regresión utilizando análisis similares al trabajo de Pardo-Fernández et al. (2007), basados en los residuos.

1.1.4 Comparación de la varianza condicional

De manera análoga a la subsección anterior, también se puede estar interesado en contrastar igualdad en las varianzas condicionales, esto es, en contrastar

$$H_0 : \sigma_1^2 = \ldots = \sigma_d^2,$$

contra la hipótesis alternativa,

$$H_1 : \sigma_s^2 \neq \sigma_t^2, \text{ para algunos } 1 \leq s, t \leq d.$$

Hasta donde conocemos, se cuenta con un sólo trabajo: el artículo de Pardo-Fernández et al. (2015b), que propone varios tests estadísticos: cuatro basados en la función de distribución empírica y dos basados en la función característica empírica de los residuos.

1.1.5 Tests de bondad de ajuste para la distribución del error

El conocimiento de la distribución del error puede mejorar varios procedimientos estadísticos sobre modelos de regresión. Por ejemplo, Eubank y Hart (1992) proponen un test para contrastar que la función de regresión es lineal para covariable univariante con diseño fijo. La distribución exacta del contraste se puede obtener cuando los errores tienen distribución normal. Por otro lado, unos de los modelos paramétricos más populares es el modelo de regresión lineal, donde comúnmente se asume que los errores tienen distribución normal. Sin embargo, no en todos los casos esta hipótesis se cumple. Por estas y otras razones varios autores, en diferentes contextos de los modelos de regresión, han propuesto tests de bondad de ajuste para la distribución del error con la siguiente hipótesis nula,

$$H_0 : F \in \mathcal{M}_\theta, \tag{1.2}$$

contra la hipótesis alternativa,

$$H_1 : F \notin \mathcal{M}_\theta,$$

donde F denota la función de distribución del error ε y $\mathcal{M}_\theta$ es una familia paramétrica, $\mathcal{M}_\theta = \{F(\cdot;\theta),\ \theta \in \Theta\},\ \Theta \subseteq \mathbb{R}^p$. En Pierce y Kopeckye (1979) se propone un test para H_0 en modelos paramétricos univariados homocedásticos. Una de las condiciones impuestas es considerar que el modelo admite una constante y que la distribución del error es contínua. Cuando el número de parámetros desconocidos del modelo se mantiene constante, a medida que el tamaño muestral n tiende a infinito, la distribución límite del estadístico propuesto es la misma que para el caso de tests clásicos para variables aleatorias IID de una familia de localización y escala. Para el caso multivariado y diseño fijo, Jiménez-Gamero et al. (2005), sin limitar que el error posea función de distribución contínua, estudian un test basado en la diferencia entre la función característica empírica de los residuos y la función característica de la distribución de los errores bajo la hipótesis nula. En Neumeyer et al. (2006), investigan tests en modelos de regresión lineal y no paramétricos multivariados homocedásticos, con diseño fijo. Los estadísticos propuestos están basados en procesos empíricos de los residuos. Para modelos lineales multivariados homocedásticos, Hušková y Meintanis (2007) desarrollan un test donde el estadístico de contraste es una norma L_2 de la diferencia entre una estimación paramétrica de la función característica del error bajo H_0 y la función característica empírica de los residuos. Heuchenne y Van Keilegom (2010) proponen un test para el caso de covariable multivariante. El estadístico se basa en la diferencia entre una estimación no paramétrica y una estimación paramétrica bajo H_0 de la distribución del error. Finalmente, describimos el trabajo propuesto en Hušková y Meintanis (2010). Estos autores proponen un estadístico que es una norma L_2 de la diferencia entre una estimación paramétrica bajo H_0 y una estimación no paramétrica de la función característica de la distribución de los errores. La distribución nula asintótica del estadístico es desconocida. Por ese motivo, proponen un bootstrap paramétrico para aproximarla. En el Capítulo 2 se propone otra aproximación.

1.1.6 Comparación de la distribución del error

La igualdad en las distribuciones del error puede simplificar algunos procedimientos estadísticos. Un ejemplo concreto se tiene en el test de igualdad en las funciones de regresión en varias poblaciones propuesto en Pardo-Fernández et al. (2015a), donde la distribución nula asintótica del estadístico coincide con el bien conocido test de ANOVA para comparación de varias medias (aunque también se debe dar una igualdad entre las densidades de las covariables). Otro, ejemplo es el test de igualdad de varianzas propuesto en Pardo-Fernández et al. (2015b), el cual cuando se cumple la igualdad en las distribuciones de los errores (y cuando las covariables son IID), tiene distribución nula asintótica que coincide con el clásico test de Levene de homogeneidad de varianzas. La hipótesis de igualdad de distribución de los errores también es asumida en algunos procedimientos como en los contrastes de igualdad de funciones de regresión propuestos en Young y Bowman (1995), Hall y Hart (1990) y Kulasequera y Wang (2001). Estas razones ponen de relieve la importancia de contrastar la siguiente hipótesis nula,

$$H_0 : F_1 = F_2 = \ldots = F_d, \tag{1.3}$$

contra la alternativa

$$H_1 : F_s \neq F_t, \text{ para algunos } 1 \leq s, t \leq d,$$

donde $F_1, \ldots, F_d$ son las funciones de distribución de los errores $\varepsilon_1, \ldots, \varepsilon_d$, respectivamente. Mora (2005) propone un test de igualdad en las distribuciones del error para $d = 2$ asumiendo una forma paramétrica de la función de regresión y homocedasticidad. Pardo-Fernández (2007) propone dos tests para contrastar la hipótesis nula (1.3). Ambos artículos estudian estadísticos basados en la estimación de la función de distribución de los errores. Sin embargo, la distribución nula asintótica es desconocia. Por lo que, en ambos trabajos se recurre a una aproximación mediante un bootstrap suavizado. En el Capítulo 4 se estudia otra estimación de la distribución nula de los estadísticos en Pardo-Fernández (2007).

1.2 Problema

Para contrastar la hipótesis nula (1.2), como ya se ha mencionado, Hušková y Meintanis (2010) proponen un bootstrap paramétrico para aproximar los valores críticos de la distribución nula asintótica del estadístico. Esta aproximación posee buenas propiedades, entre otras cosas, proporciona un estimador consistente de la distribución nula asintótica y además es muy fácil de implementarlo. Sin embargo, a medida que el número de parámetros aumenta y/o el tamaño muestral crece, el coste computacional que requiere su aplicación práctica, aumenta considerablemente. Por esta razón, en el Capítulo 2, se estudia teórica y empíricamente otro tipo de aproximación. Específicamente un bootstrap ponderado, basado en la metodología propuesta en Burke (2000). El bootstrap ponderado ya ha sido considerado en Kojadinovic y Yan (2012) y en Jiménez-Gamero y Kim (2015) para aproximar la distribución nula de tests de bondad de ajuste basados en la función de distribución empírica y la función característica empírica de las observaciones, respectivamente, entre otros. En ambos artículos se trabaja con variables observables IID. En nuestro contexto, se trata con variables no observables. Por lo que, la inferencia se basa en una aproximación de los errores, los residuos. Los residuos no son independientes. En consecuencia los procedimientos conllevan a un tratamiento teórico distinto.

Al considerar la hipótesis nula (1.3), en Pardo-Fernández (2007) se establecen ciertas condiciones sobre la distribución de los errores: con distribución contínua satisfaciendo fuertes condiciones de suavidad. Este problema se trata en el Capítulo 3, proponiendo un nuevo test, basado en la función característica empírica de los residuos de cada muestra y una estimación no paramétrica común de la función característica de la distribución de los errores bajo (1.3). El procedimiento es válido bajo condiciones menos restrictivas que en Pardo-Fernández (2007).

La distribución nula asintótica de los estadísticos propuestos en Pardo-Fernández (2007) es desconocida. Para aproximarla, el autor propone utilizar un bootstrap suavizado. Este procedimiento, al igual que el bootstrap paramétrico comparte el problema del coste computacional conforme el tamaño muestral aumenta. Este problema es abordado en el Capítulo 4, donde se estudia una aproximación mediante un bootstrap ponderado para estos estadísticos, que desde el punto de vista computacional, es más eficiente.

1.3 Objetivos

El objetivo principal de este trabajo es la de estudiar teórica y empíricamente una aproximación de la distribución nula asintótica de estadísticos para contrastar las hipótesis (1.2) y (1.3), que deriven en aproximaciones consistentes y computacionalmente eficientes. En concreto, se pueden identificar los objetivos específicos siguientes:

- Obtener una aproximación consistente y computacionalmente más eficiente al bootstrap paramétrico para la distribución nula del estadístico propuesto en Hušková y Meintanis (2010).

- Obtener una aproximación consistente y computacionalmente más eficiente al bootstrap suavizado para la distribución nula de los estadísticos propuestos en Pardo-Fernández (2007).

- Construir un test estadístico, basado en la función característica empírica de los residuos, para contrastar (1.3), válida bajo condiciones menos restrictivas que las asumidas en Pardo-Fernández (2007).

- Desarrollar una aproximación para el estadístico anterior que sea consistente y computacionalmente eficiente.

- Verificar empíricamente el comportamiento de los métodos propuestos, para todos los casos, para tamaños de muestras finitos y compararlos con métodos ya existentes.

1.4 Aportaciones

En el Capítulo 2, se considera un estadístico existente que es una norma L_2 de la diferencia entre la función característica empírica de los residuos y una estimación paramétrica de la función característica de la distribución del error bajo la hipótesis nula. Se estudia teóricamente una aproximación consistente de la distribución nula del estadístico. Los resultados empíricos obtenidos mediante una simulación confirman las buenas propiedades teóricas obtenidas para la aproximación propuesta en tamaños de muestra finito. Concluyendo que desde un punto de vista computacional, es más eficiente que la aproximación que proporciona el bootstrap paramétrico.

En el Capítulo 3, se propone un nuevo estadístico para contrastar la igualdad en la distribución de los errores. El estadístico está basado en una comparación de la función característica empírica de los residuos en cada muestra y una estimación no paramétrica de la función característica de la distribución común de los errores bajo H_0. La distribución nula asintótica del estadístico no puede ser utilizada para aproximar su distribución nula porque es desconocida y depende de la distribución común de los errores. Por lo cual, se estudia teóricamente una aproximación a dicha distribución. Esta aproximación es consistente y computacionalmente eficiente.

En el Capítulo 4, se tratan dos estadísticos ya existentes para contrastar la igualdad en la distribución de los errores. Estos estadísticos están basados en la comparación de la distribución empírica de los residuos en cada población y una estimación no paramétrica de la distribución común de los errores bajo la hipótesis nula. La distribución nula asintótica de estos estadísticos es desconocida. Se propone aproximarla mediante un bootstrap ponderado. Se demuestra la consistencia de la aproximación. Mediante estudios empíricos, para muestras de tamaño finito, se comprueba la consistencia del procedimiento propuesto y una mayor eficiencia computacional en comparación con el bootstrap suavizado.

En resumen, se ha demostrado teóricamente la consistencia de la aproximación bootstrap ponderado aplicado a los estadísticos estudiados. Esta consistencia es en el sentido de que asintóticamente son capaces de estimar correctamente el error tipo I. Así mismo, en los algoritmos de implementación de la aproximación propuesta, presentados en cada capítulo, se evidencia que el remuestreo se hace con cálculos rápidos. Los resultados teóricos, son acompañados con resultados empíricos. En tal sentido, tanto el bootstrap paramétrico, el bootstrap suavizado y el bootstrap ponderado comparten buenas propiedades. Sin embargo, desde el punto de vista computacional, la aproximación propuesta es más eficiente.

En el desarrollo de todo el trabajo se han utilizado los estimadores del tipo Nadaraya-Watson. La metodología utilizada para la obtención de los resultados teóricos, pueden servir de base para ser extendida a otros tipo de estimadores (por ejemplo, polinómicos locales, Fan y Gijbels 1996). La posibilidad de conseguir resultados teóricos similares a los obtenidos, basados en otros tipo de estimadores para los residuos estará asociado a la consistencia del estimador y la imposición condiciones adecuadas a cada caso. En tal sentido, las expansiones asintóticas de los test tratados aquí podrían presentar otras expresiones. Sin embargo, varios de los argumentos, propiedades o resultados conseguidos pueden servir como una guía para derivar en resultados análogos.

1.5 Organización del documento

Los Capítulos 2, 3 y 4 se desarrollan básicamente de una misma manera. El objetivo es, principalmente, que cada Capítulo conlleve a una lectura auto contenida. En tal afán, en primer lugar se realiza una Introducción. En la misma, se contextualiza el problema que se aborda, los objetivos, las notaciones utilizadas y se explica el contenido de las secciones involucradas en el capítulo. Seguidamente se presenta el test estadístico a estudiar y se describen cuestiones teóricas asociadas a la obtención de su distribución nula asintótica. Para luego, en la sección que le sigue, proponer soluciones concretas a los problemas que se detallan con anterioridad y que varían según caso en concreto. Así mismo, hacia el final de cada capítulo se comprueba empíricamente el comportamiento de las soluciones propuestas. En la última parte de cada capítulo se presentan aspectos más técnicos y se dan las pruebas a los resultados teóricos. Al inicio de cada capítulo se presenta un resumen de su contenido y al final las referencias utilizadas.

1.6 Referencias

1. Burke M.D., 2000. Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Statist Probab Lett 46, 13–20.

2. Delgado M., 1993. Testing the equality of nonparametric regression curves. Stat Probab Lett 17, 199–204.

3. Dette H., Munk A., 1998. Testing heterocedasticity in nonparametric regression. J R Stat Soc B 60, 693–708.

4. Dette H., Neumeyer N., Van Keilegom I., 2007. A new test for the parametric form of the variance function in non-parametric regression. J R Statist Soc B 69(5), 903–917.

5. Diblasi A. M., Bowman A. W., 1997. Testing for constant variance in a linear model. Stat Probab Lett 33, 95–103.

6. Eubank R. L., Hart J., 1992. Testing goodness of fit in regression via ordered selection criteria. Ann Stat 20, 1412–1425.

7. Fan J., Gijbels I., 1996. Local polynomial modelling and its applications. Chapman & Hall.

8. González-Manteiga W., Crujeiras R., 2013. An updated review of Goodness-of-Fit tests for regression models. Test, 22, 361–411.

9. Hall P., Hart J. D., 1990. Bootstrap test for difference between means in nonparametric regression. J Amer Statist Assoc 85, 1039–1049.

10. Härdle W., Mammen E., 1993. Comparing nonparametric versus parametric regression fits. Ann Stat 21, 1926–1947.

11. Heuchenne C., Van Keilegom I., 2010. Goodness of fit tests for the error distribution in nonparametric regression. Comput Statist Data Anal 54, 1942–1951.

12. Hušková M., Meintanis S.G., 2007. Omnibus tests for the error distribution in linear regression models. Statistics 41, 363–376.

13. Hušková M., Meintanis S.G., 2009. Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. Kybernetika 45(6), 960–971.

14. Hušková M., Meintanis S.G., 2010. Test for the error distribution in nonparametric possiby heteroscedatic regression models. Test, 19, 92–112.

15. Jiménez-Gamero M.D., Kim H-M., 2015. Fast goodness-of-fit test based on the characteristic function. Comput Statist Data Anal 89, 172–191.

16. Jimńez-Gamero, M.D., Muñoz-García, J., Pino-Mejías, R., 2005. Testing goodness of fit for the distribution of errors in multivariate linear models. J Multivariate Anal 95, 301-322.

17. Kojadinovic I., Yan J., 2012. Goodness-of-fit testing based on a weighted bootstrap: A fast sample alternative to the parametric bootstrap. Can J Statist 40, 480–500.

18. Koul H. L., Ni P., 2004. Minimum distance regression model checking. J Stat Plan Inference 119, 109–141.

19. Koul H. L., Schick A., 1997. Testing for the equality of two nonparametric regression curves. J Statis Plann Inference 65, 293–314.

20. Kozek A. S., 1991. A nonparametric test of fit of a parametric model. J Multivar Anal 37, 66–75.

21. Kulasekera K. B., Wang J., 2001. A test of the equality of regression curves using Gâteaux scores. Austral N Z J Statist 43, 89–99.

22. Liero H., 2003. Testing homocedasticity in nonparametric regression. J Nonparametr Stat 15, 31–51.

23. Mora J., 2005. Comparing distribution functions of errors in linear models: A nonparametric approach. Statist Probab Lett 73, 425–432.

24. Neumeyer N., Dette H., 2003. Nonparametric comparison of regression curves: an empirical process approach. Ann Stat 31, 880–920.

25. Neumeyer N., Pardo-Fernández J. C., 2009. A simple test for comparing regression curves versus one-sided alternatives. J Statis Plann Inference 139, 4006–4016.

26. Neumeyer N., Dette H., Nagel E-R., 2006. Bootstrap test for the error distribution in linear and nonparametric regression models. Aust N Z J Stat 48(2), 129–156.

27. Pardo-Fernández J. C., 2007. Comparison of error distributions in nonparametric regression. Statist Probab Lett 77, 350–356.

28. Pardo-Fernández J.C., Jiménez-Gamero M.D., El Ghouch A., 2015a. A nonparametric ANOVA-type test for regression curves based on characteristic functions. Scand J Stat 42, 197–213.

29. Pardo-Fernández J.C., Jiménez-Gamero M.D., El Ghouch A., 2015b. Tests for the equality of conditional variance functions in nonparametric regression. Electron J Stat 9, 1826–1851.

30. Pardo-Fernández J. C., Van Keilegom I., González-Manteiga W., 2007. Testing for the equality of k regression curves. Stat Sin 17, 1115–1137.

31. Park C., Kang K., 2008. Sizer analysis for the comparison of regression curves. Comput Stat Data Anal 52, 3954–3970.

32. Pierce D. A., Kopecky K. J., 1979. Testing goodness of fit for the distribution of errors in regression models. Biometrika 66, 1–5.

33. Stute W., González-Manteiga W., 1996. Nn goodness-of-fit tests for linear models. J Stat Plan Inference 53, 75–92.

34. Van Keilegom I., González-Manteiga W., Sánchez-Sellero C., 2008. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. Test 17, 401–415.

35. Young S. G., Bowman A. W., 1995. Non-parametric Analysis of Covariance. Biometrics 51, 920–931

36. Zhu L., Fujikoshi Y., Naito K., 2001. Heteroscedasticity checks for regression models. Sci China 44, 1236–1252.

Capítulo 2

Computationally efficient goodness-of-fit tests for the error distribution

Abstract

Several procedures have been proposed for testing goodness-of-fit to the distribution of the errors in nonparametric regression models. The null distribution of the associated test statistics is usually approximated by means of a parametric bootstrap which, under certain conditions, provides a consistent estimator. This chapter considers a goodness-of-fit test whose test statistic is an L_2 norm of the difference between the empirical characteristic function of the residuals and a parametric estimate of the characteristic function in the null hypothesis. It is proposed to approximate the null distribution through a weighted bootstrap which also produces a consistent estimator of the null distribution but, from a computational point of view, is more efficient than the parametric bootstrap.

2.1 Introduction

Let (X, Y) be a bivariate random vector satisfying the general nonparametric regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{2.1}$$

where $m(x) = E(Y \mid X = x)$ is the regression function, $\sigma^2(x) = Var(Y \mid X = x)$ is the conditional variance function and ε is the regression error, which is assumed to be independent of X. Note that, by construction, $E(\varepsilon) = 0$ and $Var(\varepsilon)$=1. The covariate X is continuous with density function f_X. The regression function, the variance function, the distribution of the errors and that of the covariate are unknown and no parametric models are assumed for them.

Because the knowledge of the distribution of the errors will improve the statistical analysis of model (2.1), several authors have proposed tests for the distribution of the errors, that is, tests of the null hypothesis

$$H_0 : F \in \mathcal{F},$$

versus the alternative

$$H_1 : F \notin \mathcal{F},$$

where F stands for the cumulative distribution function (CDF) of ε and $\mathcal{F}$ is a parametric family, $\mathcal{F} = \{F(\cdot;\theta),\ \theta \in \Theta\},\ \Theta \subseteq \mathbb{R}^p$. Examples are the tests in Neumeyer et al. (2006) and Heuchenne and Van Keilegom (2010), which are based on comparing the empirical CDF of the residuals to a parametric estimator of the CDF under the null hypothesis. Since the equality of the CFDs can be also interpreted in terms of the associated characteristic functions (CFs), Hušková and Meintanis (2010) have proposed a test for H_0 that is based on comparing the empirical CF of the residuals to a parametric estimator of the CF under the null hypothesis. As commented in Jiménez-Gamero (2013), it is interesting to observe that the last paper requires weaker conditions for the validity of the procedures than the ones based on the CDF. Nevertheless, in all cases the limit distribution of the proposed test statistics is unknown, even under the null distribution, because it depends on the unknown value of the parameter θ. To overcome this difficulty, these papers propose to use a parametric bootstrap (PB) for approximating the null distribution of the test statistic. Although very easy to implement, the PB can become very computationally expensive as the sample size and/or the number of unknown parameters increase.

This work studies another method for estimating the null distribution of the test statistic $T_{n,w}(\hat{\theta})$ in Hušková and Meintanis (2010). Specifically, a weighted bootstrap (WB) approximation in the sense of Burke (2000) is considered (see also Zhu 2005). This method has been previously suggested in Kojadinovic and Yan (2012), to approximate the null distribution of goodness-of-fit (GOF) tests based on the empirical CDF, and in Jiménez-Gamero and Kim (2015), to approximate the null distribution of GOF tests based on the empirical CF (ECF), among others. Both papers assume observable independent and identically distributed (IID) data. They show that the properties of the WB are quite similar to those of the PB (it provides a consistent estimator of the null distribution and the resulting test is able to detect any alternative) but, from a computational point of view, it is more efficient. In view of the good properties of the WB in these and other papers, it is also expected to work satisfactorily for estimating the null distribution of the test statistic considered in this chapter. The purpose of the current study is to investigate, both theoretically and empirically, the use of the WB for approximating the null distribution of $T_{n,w}(\hat{\theta})$. A main difference between the setting in this work and the one in Jiménez-Gamero and Kim (2015) and Kojadinovic and Yan (2012) is that in our case the errors are not observable. Although they can be replaced by the residuals, they are not independent.

The chapter is organized as follows. Section 2.2 describes the test statistic and explains some problems with the WB approximation. Section 2.3 gives a solution to the problems described in the previous section and proves the consistency of the proposed WB approximation. It also shows that the resulting test is consistent, in the sense of being able to detect any alternative. The application of the proposed WB approximation requires the estimation of certain functions appearing in the linear expansion of the parameter estimators. The estimation of such functions is dealt with in Section 2.4. Section 2.5 reports the results of some simulation experiments designed to study the finite sample performance of the proposed approximation and to compare it to the PB. From this numerical study it is concluded that both approximations behave quite closely but, from a computational point of view, the WB outperforms the PB. Section 2.6 concludes and outlines possible extensions of the results presented in this chapter. All proofs and technical details are deferred to the last section.

The following notation will be used along the chapter: all vectors are column vectors; for any vector a, a_k denotes its kth coordinate and $\|a\|$ its Euclidean norm; the superscript ^{T} denotes transpose; E_θ and P_θ denote expectation and probability, respectively, assuming that the data has CDF $F(\cdot;\theta)$; P_* denotes the conditional probability law, given the data; all limits in this work are taken when $n \to \infty$; $\stackrel{\mathcal{L}}{\to}$ denotes convergence in distribution; $\stackrel{P}{\to}$ denotes convergence in probability; $\stackrel{a.s.}{\to}$ denotes the almost sure convergence; for any complex number $z = a + \mathrm{i}b$, $|z|$ is its modulus; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; for a given non-negative real-valued function w we denote $\|\cdot\|_w$ to the norm and $\langle\cdot,\cdot\rangle_w$ to the scalar product in the Hilbert space $L^2(w) = \{g : \mathbb{R} \to$

$\mathbb{C}$, $\int |g(t)|^2 w(t)dt < \infty\}$; if F is a CDF, then $L^2(F) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 dF(t) < \infty\}$; for any real function $f(t;\theta)$ differentiable at $t \in \mathbb{R}$ and at $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T \in \mathbb{R}^p$ the following notations will be used,

$$f'(t;\theta) = \frac{\partial}{\partial t} f(t;\theta), \qquad f_{(r)}(t;\theta) = \frac{\partial}{\partial \theta_r} f(t;\theta), \quad 1 \le r \le p,$$

$$\nabla f(t;\theta) = (f_{(1)}(t;\theta), f_{(2)}(t;\theta), \ldots, f_{(p)}(t;\theta))^T.$$

2.2 The test statistic

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be IID from model (2.1), that is, $Y_j = m(X_j) + \sigma(X_j)\varepsilon_j$, $1 \le j \le n$. Since the hypothesis H_0 is on the common distribution of the errors, $\varepsilon_1, \ldots, \varepsilon_n$, and the errors are not observable, the inference must be based on the residuals,

$$\hat{\varepsilon}_j = \frac{Y_j - \hat{m}(X_j)}{\hat{\sigma}(X_j)}, \quad 1 \le j \le n,$$

where $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ are estimators of $m(\cdot)$ and $\sigma(\cdot)$, respectively. Several choices are possible for $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$. Here, as in Hušková and Meintanis (2010), we use the following kernel estimators for the density function f_X of X, the regression function $m(\cdot)$ and the variance function $\sigma^2(\cdot)$,

$$\hat{f}_X(x) = \frac{1}{n} \sum_{j=1}^{n} K_{h_n}(X_j - x),$$

$$\hat{m}(x) = \frac{1}{n\hat{f}_X(x)} \sum_{j=1}^{n} K_{h_n}(X_j - x) Y_j,$$

$$\hat{\sigma}^2(x) = \frac{1}{n\hat{f}_X(x)} \sum_{j=1}^{n} K_{h_n}(X_j - x) \{Y_j - \hat{m}(x)\}^2,$$

where $K_{h_n}(\cdot) = \frac{1}{h_n} K(\frac{\cdot}{h_n})$, $K(\cdot)$ is a kernel and h_n is the bandwidth, satisfying certain conditions that will be specified later.

Hušková and Meintanis (2010) proposed the following test for testing H_0,

$$\Psi = \begin{cases} 1, & \text{if} \quad T_{n,w}(\hat{\theta}) \ge t_{n,\omega,\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $t_{n,\omega,\alpha}$ is the $1-\alpha$ percentile of the null distribution of $T_{n,\omega}(\hat{\theta})$,

$$T_{n,\omega}(\hat{\theta}) = n \int |c_n(t) - c(t,\hat{\theta})|^2 \omega(t)dt = n\|c_n(t) - c(t,\hat{\theta})\|_w^2, \tag{2.2}$$

$c_n(t)$ is the ECF of the residuals,

$$c_n(t) = \frac{1}{n}\sum_{j=1}^{n}\exp(\mathrm{i}t\hat{\varepsilon}_j) = \frac{1}{n}\sum_{j=1}^{n}\cos(t\hat{\varepsilon}_j) + \mathrm{i}\frac{1}{n}\sum_{j=1}^{n}\sin(t\hat{\varepsilon}_j),$$

$c(t;\theta)$ is the CF associated to $F(\varepsilon;\theta)$, that is, $c(t;\theta) = E_\theta\{\exp(\mathrm{i}t\varepsilon)\} = R(t;\theta) + \mathrm{i}I(t;\theta)$, $\omega(t)$ is a nonnegative function such that $\int \omega(t)dt < \infty$, which may depend on θ, and $\hat{\theta}$ is a consistent estimator of θ satisfying the following assumption.

(A.1) Under H_0, $\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi(\varepsilon_j;\theta_0) + o_p(1)$, where θ_0 is the true parameter value, $E_{\theta_0}\{\psi(\varepsilon_j;\theta_0)\} = 0$ and $E_{\theta_0}\{\|\psi(\varepsilon_j;\theta_0)\|^2\} < \infty$.

Assumption (A.1) implies that, when the null hypothesis is true and θ_0 denotes the true parameter value, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed. This assumption is satisfied by commonly used estimators such as maximum likelihood estimators and method of moment estimators when $\varepsilon_1, \ldots, \varepsilon_n$ are observable and, in such a case, the expression of the function ψ is well-known (see, for example, Chapter 5 in Bickel and Doksum 2001). In our setting, the errors are not observable and the expression of the function ψ differs from the observable case. This topic will be discussed in detail in Section 2.4.

Theorem 1 in Hušková and Meintanis (2010) states that if $\hat{\theta}$ satisfies (A.1), H_0 is true and θ_0 is the true parameter value, under certain additional conditions (assumptions (A.2)–(A.7) in Section 2.7),

$$T_{n,\omega}(\hat{\theta}) \xrightarrow{\mathcal{L}} \|Z(t;\theta_0)\|_\omega^2, \tag{2.3}$$

where $\{Z(t;\theta_0),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance structure of the form $Cov_{\theta_0}\{Z_1(\varepsilon;t,\theta_0,\psi), Z_1(\varepsilon;s,\theta_0,\psi)\}$,

$$\begin{aligned} Z_1(\varepsilon;t,\theta,\psi) \;&=\; \cos(t\varepsilon) + \sin(t\varepsilon) - R(t;\theta) - I(t;\theta) - t\varepsilon\{R(t;\theta) - I(t;\theta)\} \\ &\quad -t\tfrac{\varepsilon^2-1}{2}\{R'(t;\theta) + I'(t;\theta)\} - \psi^T(\varepsilon;\theta)\{\nabla R(t;\theta) + \nabla I(t;\theta)\}. \end{aligned} \tag{2.4}$$

Clearly, the asymptotic null distribution of $T_{n,\omega}(\hat{\theta})$ is unknown. It depends on the hypothetical distribution of the error, on the chosen estimator and the true unknown value of the parameter.

In order to try to approximate the null distribution of $T_{n,\omega}(\hat{\theta})$ we first observe that it resembles a degree-2 V-statistic, because

$$T_{n,\omega}(\hat{\theta}) = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\rho(\hat{\varepsilon}_j,\hat{\varepsilon}_k;\hat{\theta}),$$

with $\rho(\varepsilon, z; \theta) = u(\varepsilon - z) - u_0(\varepsilon; \theta) - u_0(z; \theta) + u_{00}(\theta)$, $u_0(\varepsilon; \theta) = \int u(\varepsilon - z) dF(z; \theta)$, $u_{00}(\theta) = \int u(\varepsilon - z) dF(\varepsilon; \theta) dF(z; \theta)$, and $u(t) = \int \cos(t\varepsilon)\omega(\varepsilon) d\varepsilon$.

Dehling and Mikosch (1994) (see also Hušková and Janssen 1993) showed that if $\varepsilon_1, \ldots, \varepsilon_n$ are IID, $\xi_1, \ldots, \xi_n$ are IID with $E(\xi_1) = 0$ and $var(\xi_1) = 1$, independent of $\varepsilon_1, \ldots, \varepsilon_n$ and $V_n = \frac{1}{n^2} \sum_{1 \le j,k \le n} g(\varepsilon_j, \varepsilon_k)$ is a degenerate degree-2 V-statistic, then the conditional distribution, given $\varepsilon_1, \ldots, \varepsilon_n$, of

$$\frac{1}{n} \sum_{1 \le j,k \le n} g(\varepsilon_j, \varepsilon_k) \xi_j \xi_k$$

consistently estimates that of nV_n. In the light of this result, since $\hat{\varepsilon}_j$ and $\hat{\theta}$ are approximations to ε_j and θ, respectively, one may be tempted to estimate the null distribution of $T_{n,\omega}(\hat{\theta})$ by means of the conditional distribution, given $(X_1, Y_1), \ldots, (X_n, Y_n)$, of

$$W^* = \frac{1}{n} \sum_{1 \le j,k \le n} \rho(\hat{\varepsilon}_j, \hat{\varepsilon}_k; \hat{\theta}) \xi_j \xi_k. \tag{2.5}$$

We will see that this approach is wrong. The next result gives the limit distribution of W^*. The required assumptions are listed in Section 2.7.

Theorem 1 *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, that assumptions (A.2)–(A.6) hold, that the first partial derivatives $R_{(r)}(t; \theta)$, $I_{(r)}(t; \theta)$, $1 \le r \le p$, exist and are continuous functions $\forall \theta \in U(\theta_1) \subseteq \Theta$, an open neighborhood of θ_1, and they are bounded by functions in $L_2(\omega)$, $\forall \theta \in U(\theta_1)$, then*

$$\sup_x |P_* \{W^* \le x\} - P\{W_0 \le x\}| \xrightarrow{P} 0,$$

where $W_0 = \|Z_0(t; \theta_1)\|_\omega^2$, $\{Z_0(t; \theta_1), t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance structure of the form $Cov\{Z_0(\varepsilon; t, \theta_1), Z_0(\varepsilon; s, \theta_1)\}$, $Z_0(\varepsilon; t, \theta) = \cos(t\varepsilon) + \sin(t\varepsilon) - R(t; \theta) - I(t; \theta)$.

From the result in Theorem 1 and (2.3), it is clear that the conditional distribution of W^* does not provide a consistent estimator of the asymptotic null distribution of $T_{n,\omega}(\hat{\theta})$ because replacing $m(\cdot)$, $\sigma(\cdot)$ and θ by $\hat{m}(\cdot)$, $\hat{\sigma}(\cdot)$ and $\hat{\theta}$, respectively, has an impact on the asymptotic null distribution of the test statistic that is not captured by the conditional distribution of W^*. The next Section shows how to deal with this problem.

Before ending this section we do some comments on the behaviour of $\hat{\theta}$ under the alternative. Theorem 1 assumes that $\hat{\theta}$ has a limit (in probability), θ_1. In practice, to estimate θ one proceeds as if H_0 were true. For example, θ is usually estimated by its quasi maximum likelihood estimator, which maximizes the likelihood under the null hypothesis (with the errors replaced by the residuals). If H_0 is true, under certain assumptions, the resulting estimator converges to the true parameter value (see Section 2.4); if H_0 is not true, then proceeding as in White (1982) for observable data, it can shown that, under certain conditions, the estimator also converges to a well-defined limit. Similar comments could be done for other estimators.

2.3 Consistency of the WB approximation

If assumptions (A.1)–(A.7) hold and H_0 is true, from the proof of Theorem 1 in Hušková and Meintanis (2010), it follows that

$$T_{n,\omega}(\hat{\theta}) = T_{1,n,\omega}(\theta_0) + o_p(1), \tag{2.6}$$

where

$$T_{1,n,\omega}(\theta) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\varepsilon_j; t, \theta, \psi)\|_\omega^2,$$

with $Z_1(\varepsilon; t, \theta, \psi)$ as defined in (2.4). Now, from (2.6) and applying the results in Dehling and Mikosch (1994), we get that the conditional distribution, given $(X_1, Y_1), \ldots, (X_n, Y_n)$, of

$$T^*_{1,n,\omega}(\theta_0) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\varepsilon_j; t, \theta_0, \psi)\xi_j\|_\omega^2,$$

provides a consistent estimator of the distribution of $T_{n,\omega}(\hat{\theta})$, when H_0 is true. From a practical point of view, this result is useless because $Z_1(\varepsilon_j; t, \theta_0, \psi)$ depends on the non-observable error ε_j, on the unknown value of θ_0 and on the function $\psi(\varepsilon_j; \theta_0)$, whose explicit expression is usually unknown. Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true. To overcome these difficulties we replace ε_j by $\hat{\varepsilon}_j$, θ_0 by $\hat{\theta}$ and $\psi(\varepsilon_j; \theta_0)$ by $\psi_n(\hat{\varepsilon}_j; \hat{\theta})$, where $\psi_n(\cdot; \hat{\theta})$ is a function of the data which approximates ψ in such a way that

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n} \|\psi_n(\hat{\varepsilon}_j; \hat{\theta}) - \psi_1(\varepsilon_j; \theta_1)\|^2 \xrightarrow{P} 0, \\ &\text{with } \; E\{\|\psi_1(\varepsilon; \theta_1)\|^2\} < \infty \; \text{ and } \; \psi_1(\varepsilon; \theta_1) = \psi(\varepsilon; \theta_1) \; \text{ if } H_0 \text{ is true.} \end{aligned} \tag{2.7}$$

The choice of ψ_n will depend on ψ, that is, on the estimator of θ considered. Section 2.4 studies some proposals for ψ_n satisfying (2.7) for two common choices for $\hat{\theta}$: the maximum likelihood estimator and the method of moments estimator, both based on the residuals. So, the null distribution of $T_{n,\omega}(\hat{\theta})$ is now estimated by means of the conditional distribution, given $(X_1, Y_1), \ldots, (X_n, Y_n)$, of

$$T^*_{2,n,\omega}(\hat{\theta}) = \| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_1(\hat{\varepsilon}_j; t, \hat{\theta}, \psi_n)\xi_j \|^2_\omega.$$

The next theorem gives the limit of the conditional distribution of $T^*_{2,n,\omega}(\hat{\theta})$, given (X_1, Y_1), $\ldots (X_n, Y_n)$.

Theorem 2 *Suppose that* $\|\hat{\theta} - \theta_1\| = o_p(1)$, *for some* $\theta_1 \in \Theta$, θ_1 *being the true parameter value if* H_0 *is true, and that assumptions (A.1)–(A.7) and* (2.7) *hold, then*

$$\sup_x \left| P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \leq x \right\} - P\left\{ T_2 \leq x \right\} \right| \xrightarrow{P} 0,$$

where $T_2 = \|Z_2(t;\theta_1)\|^2_\omega$, $\{Z_2(t;\theta_1),\, t \in \mathbb{R}\}$ *is a centered Gaussian process on* $L_2(\omega)$ *with covariance structure of the form* $Cov\{Z_1(\varepsilon; t, \theta_1, \psi_1), Z_1(\varepsilon; s, \theta_1, \psi_1)\}$.

The result in Theorem 2 is valid whether the null hypothesis H_0 is true or not. An immediate consequence of this fact and (2.3) is the following.

Corollary 1 *If* H_0 *is true and the assumptions in Theorem 2 hold, then*

$$\sup_x \left| P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \leq x \right\} - P_{\theta_1} \left\{ T_{n,\omega}(\hat{\theta}) \leq x \right\} \right| \xrightarrow{P} 0.$$

Let $\alpha \in (0, 1)$ and

$$\Psi_* = \begin{cases} 1, & \text{if} \quad T_{n,\omega}(\hat{\theta}) \geq t^*_{2,n,\omega,\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $t^*_{2,n,\omega,\alpha}$ is the $1 - \alpha$ percentile of the conditional distribution of $T^*_{2,n,\omega}(\hat{\theta})$, or equivalently, $\Psi_* = 1$ if $p^* \leq \alpha$, where $p^* = P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \geq T_{n,\omega}(\hat{\theta})_{obs} \right\}$ and $T_{n,\omega}(\hat{\theta})_{obs}$ is the observed value of the test statistic. The result in Corollary 1 states that Ψ_* is asymptotically correct, in the sense that its type I error is asymptotically equal to the nominal value α.

Corollary 2 *Suppose that* H_0 *is not true and let* $c(t)$ *denote the true CF of the errors. If the assumptions in Theorem 2 hold and* ω *is such that*

$$\kappa = \|c(t) - c(t;\theta_1)\|^2_\omega > 0, \tag{2.8}$$

then $P(\Psi_* = 1) \to 1$.

Corollary 2 shows that, if ω is such that (2.8) holds, then the test Ψ_* is consistent in the sense of being able to asymptotically detect any (fixed) alternative. Since two distinct characteristic functions can be equal in a finite interval (Feller 1971, p.506), a general way to ensure (2.8) is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$.

Remark 1 *If model* (2.1) *is homoscedastic, that is, if* $\sigma(x) = \sigma$, $\forall x$, *for some unknown* $\sigma > 0$, *we can use the residuals* $\tilde{\varepsilon}_j = Y_j - \hat{m}(X_j)$, $1 \le j \le n$, *and consider* σ *as a parameter of the family* $\mathcal{F}$. *In this framework, the result in Theorem 2 (with weaker assumptions) keeps on being true with the following simpler expression for* $Z_1(\varepsilon; t, \theta, \psi)$,

$$\begin{aligned} Z_1(\varepsilon; t, \theta, \psi) &= \cos(t\varepsilon) - R(t;\theta) + \sin(t\varepsilon) - I(t;\theta) - t\varepsilon R(t;\theta) + t\varepsilon I(t;\theta) \\ &\quad - \psi^T(\varepsilon;\theta)\{\nabla R(t;\theta) + \nabla I(t;\theta)\}. \end{aligned}$$

Remark 2 *If the null hypothesis is simple, then the result in Theorem 2 (with weaker assumptions) is also true with the following simpler expression for* $Z_1(\varepsilon; t, \theta, \psi) = Z_1(\varepsilon; t)$,

$$Z_1(\varepsilon; t) = \cos(t\varepsilon) - R(t) + \sin(t\varepsilon) - I(t) - t\varepsilon R(t) + t\varepsilon I(t) - t\frac{\varepsilon^2 - 1}{2}\{R'(t) + I'(t)\},$$

where $R(t)$ *and* $I(t)$ *denote the real and the imaginary parts of the CF of the law in the null hypothesis.*

Remark 3 *If model* (2.1) *is homoscedastic and the null hypothesis is simple, which implies that* $\sigma(x) = \sigma$, $\forall x$, *for some known* $\sigma > 0$, *as observed in Remark 1, we can use the residuals* $\tilde{\varepsilon} = Y_j - \hat{m}(X_j)$, $1 \le j \le n$. *In this setting, the result in Theorem 2 (with weaker assumptions) is also true with the following simpler expression for* $Z_1(\varepsilon; t, \theta, \psi) = Z_1(\varepsilon; t)$,

$$Z_1(\varepsilon; t) = \cos(t\varepsilon) - R(t) + \sin(t\varepsilon) - I(t) - t\varepsilon R(t) + t\varepsilon I(t),$$

where $R(t)$ *and* $I(t)$ *denote the real and the imaginary parts of the CF of the law in the null hypothesis.*

Remark 4 *When the null hypothesis is simple, the asymptotic null distribution of the test statistic* $T_{n,\omega}(\hat{\theta})$ *does not depend on unknown parameters. So, in this case the asymptotic null distribution could be used to approximate the null distribution. The simulations carried out (reported in Section 2.5) reveal that, for small to moderate sample sizes, the WB provides a better fit.*

Remark 5 *Theorem 3 in Hušková and Meintanis (2010) shows that the PB null distribution estimator of $T_{n,\omega}(\hat{\theta})$ satisfies a result which is similar to that stated in Corollary 1 for the WB estimator. Nevertheless, although the tests Ψ_* and the one obtained by approximating $t_{n,\omega,\alpha}$ through its PB estimator, are both of them consistent against all fixed alternatives, their powers will be different for finite sample sizes.*

So far we have assumed that the weight function does not depend on θ, but in some cases it does. Such dependence is motivated by the recommendations in Epps and Pulley (1983), who suggest to choose $\omega(t)$ giving high weight where the ECF is a relatively precise estimator of the population CF. It entails taking $\omega(t) = \nu\{|c(t;\hat{\theta})|\}$, for some ν, a nonnegative increasing function. For example, if $\int |c(t;\theta)|^2 dt < \infty$, one could choose $\omega(t) = |c(t;\hat{\theta})|^2 / \int |c(x;\hat{\theta})|^2 dx$, which is the choice for ω in Epps and Pulley (1983) (see also Epps 2005). In addition, as observed in Jiménez-Gamero et al. (2009), such choice for $\omega(t)$ may have some computational advantages when the density (under the null hypothesis) of $\varepsilon_1 - \varepsilon_2$, $\varepsilon_1 - \varepsilon_2 + \varepsilon_3$ and $\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4$ is known since from expression (14) in Jiménez-Gamero et al. (2009), the test statistic (2.2) can be expressed as

$$\frac{1}{f_{\varepsilon_1-\varepsilon_2}(0;\hat{\theta})}\left\{\frac{1}{n}\sum_{j,k=1}^{n} f_{\varepsilon_1-\varepsilon_2}(\hat{\varepsilon}_j - \hat{\varepsilon}_k;\hat{\theta}) - 2\sum_{j=1}^{n} f_{\varepsilon_1-\varepsilon_2+\varepsilon_3}(\hat{\varepsilon}_j;\hat{\theta}) + n f_{\varepsilon_1-\varepsilon_2+\varepsilon_3-\varepsilon_4}(0;\hat{\theta})\right\},$$

where $f_U(x;\theta)$ is the density function of U.

If the weight function ω depends on θ, $\omega(t) = \omega(t;\theta)$, then the test statistic (2.2) becomes

$$T_{n,\hat{\omega}}(\hat{\theta}) = n\int |c_n(t) - c(t;\hat{\theta})|^2 \omega(t;\hat{\theta})dt = n\|c_n(t) - c(t;\hat{\theta})\|^2_{\hat{\omega}},$$

where the subindex $\hat{\omega}$ means that the weight function depends on $\hat{\theta}$, that is, $\omega(t) = \omega(t;\hat{\theta})$. To deal with this case we will assume that the weight function is smooth as a function of θ, as expressed in the next assumption.

(A.8) $|\omega(t;\theta_1) - \omega(t;\theta)| \leq \omega_0(t;\theta_1)\|\theta - \theta_1\|$, $\forall\,\theta$ in an open neighborhood of θ_1, with $\omega_0(t;\theta_1)$ satisfying $\int \omega_0(t;\theta_1)dt < \infty$.

If assumption (A.8) holds, assumptions (A.2), (A.7) hold with $\omega(t) = \omega_0(t;\theta)$ and H_0 is true, then

$$T_{n,\hat{\omega}}(\hat{\theta}) = T^1_{n,\omega}(\hat{\theta}) + o_p(1),$$

with $T^1_{n,\omega}(\hat\theta) = n\int |c_n(t) - c(t,\hat\theta)|^2\omega(t;\theta_1)dt$. Let $T^*_{3,n,\omega}(\hat\theta) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_1(\hat\varepsilon_j; t, \hat\theta, \psi_n)\xi_j\|^2_{\hat\omega}$ and

$$\Psi_{1*} = \begin{cases} 1, & \text{if} \quad T_{n,\hat\omega}(\hat\theta) \ge t^*_{3,n,\omega,\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $t^*_{3,n,\omega,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of $T^*_{3,n,\omega}(\hat\theta)$. Now, proceeding as in the case where ω does not depend on the parameter θ, we state the following result.

Theorem 3 *Suppose that $\|\hat\theta - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true, that assumptions (A.1)–(A.8) and* (2.7) *hold, where both (A.2) and (A.7) hold with $\omega(t) = \omega_0(t;\theta_1)$ and $\omega(t) = \omega(t;\theta_1)$.*

(a) If H_0 is true, then

$$\sup_x \left| P_*\left\{T^*_{3,n,\omega}(\hat\theta) \le x\right\} - P_{\theta_1}\left\{T_{n,\hat\omega}(\hat\theta) \le x\right\}\right| \xrightarrow{P} 0.$$

(b) If H_0 is not true and (2.8) *holds with $\omega(t) = \omega(t;\theta_1)$, then $P(\Psi_{1*} = 1) \to 1$.*

The observation in Remark 1 also applies in this case.

Remark 6 *The results stated up to now keep on being true if instead of using the raw multipliers, $\xi_1, \ldots, \xi_n$, we use the centered multipliers, $\xi_1 - \bar\xi, \ldots, \xi_n - \bar\xi$, as suggested in Burke (2000) and Kojadinovic and Yan (2012), where $\bar\xi = \frac{1}{n}\sum_{j=1}^n \xi_j$.*

Remark 7 *In practice, to calculate the WB approximation to the null distribution of $T_{n,\omega}(\hat\theta)$ (analogously for $T_{n,\hat\omega}(\hat\theta)$) we proceed as follows:*

1. *Calculate the residuals $\hat\varepsilon_1, \ldots, \hat\varepsilon_n$ (or $\tilde\varepsilon_1, \ldots, \tilde\varepsilon_n$, if the model is homoscedastic).*
2. *Calculate $\hat\theta$ and the observed value of the test statistic $T_{n,\omega}(\hat\theta)_{obs}$.*
3. *Calculate $m_{jk} = \langle Z_1(\hat\varepsilon_j; t, \hat\theta, \psi_n),\, Z_1(\hat\varepsilon_k; t, \hat\theta, \psi_n)\rangle_\omega$, $1 \le j \le k \le n$, and take $m_{jk} = m_{kj}$.*
4. *For some large integer B, repeat the following steps for every $b \in \{1, \ldots, B\}$:*
   (a) Generate n IID variables $\xi_1, \ldots, \xi_n$ with mean 0 and variance 1.
   (b) Calculate $T^{*b}_{2,n,w}(\hat\theta) = \frac{1}{n}\sum_{j,k}\xi_j\xi_k m_{jk}$ (or $T^{*b}_{2,n,\omega}(\hat\theta) = \frac{1}{n}\sum_{j,k}(\xi_j - \bar\xi)(\xi_k - \bar\xi)m_{jk}$, as noted in Remark 6).
5. *Approximate the p-value by $\hat p = \frac{1}{B}\sum_{b=1}^B I\{T^{*b}_{2,n,\omega}(\hat\theta) > T_{n,\omega}(\hat\theta)_{obs}\}$.*

2.4 Parameter estimators

The maximum likelihood estimator (MLE) satisfies Assumption (A.1) for observable random variables. In our case, the errors are not observable. It seems reasonable to replace the errors by the residuals in the likelihood and then maximize in θ the resulting function. Specifically, assume that the CDF $F(x;\theta)$ has a Radon-Nikodym derivative $f(x;\theta)$ with respect to some σ-finite measure over $(\mathbb{R},\mathcal{B})$, where $\mathcal{B}$ is the class of Borel sets of $\mathbb{R}$. To estimate θ we treat the residuals as it they were the true errors and consider

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \sum_{j=1}^{n} \log f(\hat{\varepsilon}_j;\theta).$$

Theorem 3.1 in Heuchenne and Van Keilegom (2010) shows that (under certain conditions) $\hat{\theta}_{ML}$ satisfies (A.1) with $\psi(\varepsilon;\theta)=\psi_{ML}(\varepsilon;\theta)$ given by

$$\psi_{ML}(\varepsilon;\theta) = \rho(\varepsilon;\theta) + \varepsilon\rho_1(\theta) + \frac{\varepsilon^2-1}{2}\rho_2(\theta), \tag{2.9}$$

where $\rho_1(\theta)=E_\theta\{\rho'(\varepsilon;\theta)\}$, $\rho_2(\theta)=E_\theta\{\varepsilon\rho'(\varepsilon;\theta)\}$, $\rho(\varepsilon;\theta)=-A(\theta)^{-1}\nabla\log f(\varepsilon;\theta)$, $A(\theta)=(A_{rs}(\theta))$ and

$$A_{rs}(\theta) = E_\theta\left(\frac{\partial}{\partial\theta_r}\log f(\varepsilon;\theta)\frac{\partial}{\partial\theta_s}\log f(\varepsilon;\theta)\right), \quad 1\le s,r\le p.$$

In view of (2.9), a natural choice for $\psi_n(\varepsilon;\theta)$ is $\psi_n(\varepsilon;\theta)=\psi_{n,ML}(\varepsilon;\theta)$ with

$$\psi_{n,ML}(\varepsilon;\theta) = \rho_n(\varepsilon;\theta) + \varepsilon\hat{\rho}_1(\theta) + \frac{\varepsilon^2-1}{2}\hat{\rho}_2(\theta),$$

where

$$\begin{aligned}
\rho_n(\varepsilon;\theta) &= -\hat{A}_n(\theta)^{-1}\nabla\log f(\varepsilon;\theta),\\
\hat{\rho}_1(\theta) &= \frac{1}{n}\sum_{j=1}^{n}\rho_n'(\hat{\varepsilon}_j;\theta),\\
\hat{\rho}_2(\theta) &= \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j\rho_n'(\hat{\varepsilon}_j;\theta),\\
\rho_n'(\varepsilon;\theta) &= -\hat{A}_n(\theta)^{-1}\frac{\partial}{\partial\varepsilon}\nabla\log f(\varepsilon;\theta),\\
\hat{A}_n(\theta) &= (\hat{A}_{n,rs}(\theta)), \quad \hat{A}_{n,rs}(\theta) = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_r}\log f(\hat{\varepsilon}_j;\theta)\frac{\partial}{\partial\theta_s}\log f(\hat{\varepsilon}_j;\theta), \quad 1\le s,r\le p.
\end{aligned}$$

The next theorem shows that $\psi_{n,ML}(\varepsilon;\theta)$ satisfies (2.7). Let $A_F(\theta)=(A_{F,rs}(\theta))$, with $A_{F,rs}(\theta)=E\left(\frac{\partial}{\partial\theta_r}\log f(\varepsilon;\theta)\frac{\partial}{\partial\theta_s}\log f(\varepsilon;\theta)\right)$, $1\le s,r\le p$, $\rho_{1,F}(\theta)=E\{\rho_F'(\varepsilon;\theta)\}$, $\rho_{2,F}(\theta)=E\{\varepsilon\rho_F'(\varepsilon;\theta)\}$ and $\rho_F(\varepsilon;\theta)=-A_F(\theta)^{-1}\nabla\log f(\varepsilon;\theta)$.

Theorem 4 *Suppose that* $\|\hat{\theta} - \theta_1\| = o_p(1)$*, for some* $\theta_1 \in \Theta$*,* θ_1 *being the true parameter value if* H_0 *is true, and that assumptions (A.3)–(A.6), (A.9) hold, then* $\psi_{n,ML}(\varepsilon;\theta)$ *satisfies*

$$\frac{1}{n}\sum_{j=1}^{n} \|\psi_{n,ML}(\hat{\varepsilon}_j;\hat{\theta}) - \psi_1(\varepsilon_j;\theta_1)\|^2 \xrightarrow{P} 0,$$

with $\psi_1(\varepsilon;\theta) = \rho_F(\varepsilon;\theta) + \varepsilon\rho_{1,F}(\theta) + \frac{\varepsilon^2-1}{2}\rho_{2,F}(\theta)$.

Clearly, $\psi_1(\varepsilon_j;\theta)$ in Theorem 4 satisfies $\psi_1(\varepsilon_j;\theta_1) = \psi_{ML}(\varepsilon;\theta_1)$ when H_0 is true.

Remark 8 *If model* (2.1) *is homoscedastic then the expressions for* $\psi_{ML}(\varepsilon;\theta)$ *and* $\psi_{n,ML}(\varepsilon;\theta)$ *simplify to* $\psi_{ML}(\varepsilon;\theta) = \rho(\varepsilon;\theta) + \varepsilon\rho_1(\theta)$ *and* $\psi_{n,ML}(\varepsilon;\theta) = \rho_n(\varepsilon;\theta) + \varepsilon\hat{\rho}_1(\theta)$*, respectively.*

Another estimator that is commonly used is the method of moment estimator (MME). Although these estimators are not usually optimal, they are frequently used because their calculation is less time consuming than that of MLEs. MMEs satisfy Assumption (A.1) for observable random variables. As noticed before, in our setting the errors are not observable. Next, we study if (A.1) still holds when the errors are replaced by the residuals. Assume that, under the null hypothesis, $\theta_0 = g(\mu_0)$, for some known function $g = (g_1, \ldots, g_p)^T$, $g_r : \mathbb{R}^{k-1} \to \mathbb{R}$, $1 \le r \le p$, $\mu_0 = (\mu_{0,2}, \ldots, \mu_{0,k})^T$ and $\mu_{0,s} = E_{\theta_0}(\varepsilon^s)$, $\forall s$. The first moment has not been included because, by construction, it is known and equal to 0. In heteroscedastic models the second order moment is also known (thus in this case $\mu_0 = (\mu_{0,3}, \ldots, \mu_{0,k})^T$), but it is not in homoscedastic models (thus in this case $\mu_0 = (\mu_{0,2}, \ldots, \mu_{0,k})^T$). Nevertheless, we will work with $\mu_0 = (\mu_{0,2}, \ldots, \mu_{0,k})^T$, by implicitly understanding that in heteroscedastic models $g(\mu_{0,2}, \ldots, \mu_{0,k}) = g(\mu_{0,3}, \ldots, \mu_{0,k})$. Let $\hat{\theta}_{MM} = g(\hat{\mu})$, with $\hat{\mu} = (\hat{\mu}_2, \ldots, \hat{\mu}_k)^T$, $\hat{\mu}_s = \frac{1}{n}\sum_{j=1}^{n} \hat{\varepsilon}_j^s$, $\forall s$. The next theorem states that, under certain conditions, assumption (A.1) holds for $\hat{\theta}_{MM}$. Let $\nabla g_r(x) = \left(\frac{\partial}{\partial x_2} g_r(x), \ldots, \frac{\partial}{\partial x_k} g_r(x)\right)^T$, $1 \le r \le p$, and let $\nabla g(x)$ be the $p \times (k-1)$–matrix with rows $\nabla g_1(x)^T, \ldots, \nabla g_p(x)^T$, for any $x = (x_2, \ldots, x_k)^T \in \mathbb{R}^{k-1}$.

Theorem 5 *Suppose that assumptions (A.3)–(A.6) hold, that* g *is continuously differentiable at* μ_0*, that* $\mu_{0,2k} < \infty$ *and that* H_0 *is true, then*

$$\sqrt{n}(\hat{\theta}_{MM} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \psi_{MM}(\varepsilon_j;\mu_0) + o_p(1),$$

where $\psi_{MM}(\varepsilon;\mu_0) = \nabla g(\mu_0) v$*,* $v = (v_2, \ldots, v_k)^T$*,* $v_s = \varepsilon^s - \mu_{0,s} - \mu_{0,s-1}\varepsilon - \mu_{0,s}\frac{\varepsilon^2-1}{2}$*,* $2 \le s \le k$*.*

In the light of the result in Theorem 5, to approximate $\psi_{MM}(\varepsilon;\mu)$ we could replace the population moments by their empirical counterparts based on the residuals. The next theorem shows that this approximation for $\psi_{MM}(\varepsilon;\theta)$ satisfies (2.7). Let $\mu_{F,s} = E(\varepsilon^s)$ and $\mu_F = (\mu_{F,2},\ldots,\mu_{F,k})^T$.

Theorem 6 *Suppose that assumptions (A.3)–(A.6), (A.10) hold and that $\mu_{F,2k} < \infty$, then*

$$\frac{1}{n}\sum_{j=1}^{n} \|\psi_{MM}(\hat{\varepsilon}_j;\hat{\mu}) - \psi_{MM}(\varepsilon_j;\mu_F)\|^2 \xrightarrow{P} 0.$$

Clearly, $\psi_{MM}(\varepsilon_j;\mu_F) = \psi_{MM}(\varepsilon_j;\mu_0)$ when H_0 is true.

Remark 9 *If model* (2.1) *is homoscedastic then the expressions for $\psi_{MM}(\varepsilon;\mu)$ simplifies to $\psi_{MM}(\varepsilon;\mu_0) = \nabla g(\mu_0)v$, $v = (v_2,\ldots,v_k)^T$, $v_s = \varepsilon^s - \mu_{0,s} - \mu_{0,s-1}\varepsilon$, $2 \le s \le k$.*

2.5 Finite sample performance

With the aim of studying the finite sample performance of the proposed procedure two simulation experiments were carried out: first, a homoscedastic regression model was considered, and then a heteroscedastic regression model. The main goal of these experiments is to compare the approximations provided by the asymptotic null distribution (when the null hypothesis is simple), the PB (as described in Hušková and Meintanis 2010) and the WB proposed in this chapter, in three senses: closeness of the approximation under the null, the power for fixed alternatives of the resulting test and the consumed time (for the PB and the WB). This section reports and summarizes the numerical results obtained. All computations were performed using programs written in the R language (R Core Team, 2015).

In both models the hypotheses $H_0 : \varepsilon \sim N(0,\theta)$, that corresponds to testing that the error distribution is normal with CF $\exp(-0.5\theta t^2)$, and $H_0 : \varepsilon \sim \mathcal{L}(0,\theta)$, that corresponds to testing that the error distribution is Laplace with CF $\frac{1}{1+\theta t^2}$, were studied. As in Hušková and Meintanis (2010), and following the recommendations in Epps and Pulley (1983), the weight functions considered were: $\omega(t;\theta) = \exp(-\lambda\theta t^2)$, when testing normality, and $\omega(t;\theta) = (1+\theta t^2)^4\exp(-\lambda t^2)$, when testing for the Laplace distribution. For the homoscedastic model two cases were considered: θ known and θ unknown. In this second case, the parameter was estimated by a MME. Specifically, $\hat{\theta} = \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j^2$, for testing normality, and $\hat{\theta} = \frac{1}{2n}\sum_{j=1}^n \hat{\varepsilon}_j^2$, for the Laplace distribution. To estimate the regression function and the conditional variance the Epanechnikov kernel $K(u) = 0.75 \times (1 - u^2)$ was employed.

As for the choice of the bandwidth, in a recent review about GOF problems in nonparametric regression, González-Manteiga and Crujeiras (2013) say that the bandwidth selection for tests based on smoothing is a “really tough problem” and “it is far from being solved” (see also the discussions of Sperlich 2013 and de Uña-Álvarez 2013, to the mentioned article). Because of this reason, to choose h, we proceeded as in the simulation study in Pardo-Fernández et al. (2015a): we took $h = c \times n^a$, where c and a are real constants and n is the sample size; to determine c, a and λ some preliminary simulations were performed with the purpose of finding values giving type I error close to the nominal. For all tried combinations of $c \in (1, 1.8)$, $a \in (-0.50, -0.25)$ and $\lambda \in (0.03, 0.54)$ good results were obtained for the WB. Here we only report the results for $c = 1.2$, $a = -0.375$ and $\lambda = 0.04$.

The distribution of the errors were generated from: the normal distribution (denoted as N in the tables), the Laplace distribution (denoted as LP), the logistic distribution (denoted as LG), the Gumbel distribution (denoted as G), the beta distribution with parameters $a = 1$ and $b = 0.5$ (denoted as β), the chi-squared distribution with 3 degrees of freedom (denoted as χ^2_3) and the Student t distribution with 5 degrees of freedom (denoted as t_5). All aforementioned distributions were conveniently centered and scaled to have mean 0 and variance 1.

To approximate the p-value, 1000 replications were generated for both the PB and the WB. For the WB, the raw multipliers and the centered multipliers were considered, denoted by WB1 and WB2 in the tables, respectively. The multipliers were generated from a univariate standard normal distribution. As for the asymptotic distribution (when the null hypothesis is simple, denoted as A in the tables), it is rather difficult to calculate because it coincides with that of $\sum_{j\geq 1} \lambda_j \chi^2_{1,j}$, where $\chi^2_{1,1}, \chi^2_{1,2}, \ldots$ are independent chi-squared variables with one degree of freedom, the set $\{\lambda_j,\, j \geq 1\}$ are the non-null eigenvalues of the integral equation $\int C(t,s) G_j(t) dt = \lambda_j G_j(s)$, with corresponding eigenfunctions $\{G_j(\cdot),\, j \geq 1\}$, $C(t,s)$ is the covariance kernel of $Z_1(\varepsilon; t)$ (see Remarks 2 and 3 for the expression of $Z_1(\varepsilon; t)$), and determining the eigenvalues of an integral equation is tricky. Because of this reason, we approximated it by generating 10000 samples of size 1000 obeying H_0 and calculated the test statistic at each sample, obtaining 10000 values. The empirical CDF of these 10000 values was taken as an approximation to the asymptotic null distribution.

1000 samples with size $n = 25$ were generated from each distribution and the fractions of p-values less than or equal to 0.05 and 0.1 were calculated. The experiment was repeated for $n = 50$, 100.

2.5.1 Homoscedastic model

The reported results correspond to the model

$$Y_j = X_j + X_j^2 + \varepsilon_j, \quad 1 \leq j \leq n,$$

where X_j follows the uniform $(0, 1)$ distribution. We first considered that θ is known. Since the model is homoscedastic and the null hypothesis is simple, the simplifications in Remark 3 can be applied. Table 2.1 displays the results obtained for the type I error and the power for testing normality and Table 2.2 for testing GOF to the Laplace distribution. Looking at these tables it can be concluded that, in terms of type I error, both the PB and the WB behave very close to the nominal levels, while the asymptotic approximation is a bit conservative, specially for testing GOF for the Laplace distribution. As for the power, the test based on the WB approximation seems to be a bit more powerful than one based on the PB. In most cases (all but alternatives β and χ_3^2 in Table 2.2), the WB approximation is also more powerful than one based on the asymptotic approximation.

Tables 2.3 and 2.4 show the results when θ is assumed to be unknown. In this case, the simplifications in Remark 1 can be applied. Looking at these tables it can be concluded that, in terms of the type I error, as before, both the PB and the WB behave very close to the nominal levels. As for the power, for $n = 25, 50$ in some cases the WB is more powerful than the PB, but in others cases the opposite is observed; for $n = 100$ the test based on the WB approximation seems to be a bit more powerful than one based on the PB.

2.5.2 Heteroscedastic model

The reported results correspond to the model

$$Y_j = X_j + X_j^2 + (X_j + 0.5)\varepsilon_j, \quad 1 \leq j \leq n,$$

where X_j follows the uniform $(0, 1)$ distribution. Since the model is heteroscedastic and the null hypothesis is simple, the simplifications in Remark 2 can be applied. Table 2.5 displays the results obtained for the type I error and the power for testing normality and

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	3.60	6.10	4.10	6.40	4.00	5.00	4.12	4.84	5.20	4.74	4.12	4.74
	8.20	11.50	10.20	12.30	9.00	10.04	9.24	10.48	10.20	9.64	9.34	10.40
LP	25.50	36.10	57.80	64.80	45.40	56.30	86.60	88.30	76.60	77.70	98.90	99.00
	35.90	48.70	70.60	74.70	57.40	68.50	90.40	91.00	83.60	83.20	99.60	99.70
LG	10.30	57.60	56.40	63.10	12.70	88.10	87.30	89.30	17.80	99.90	100.00	100.00
	18.10	70.40	72.00	76.00	20.60	93.20	94.30	95.10	27.80	99.90	100.00	100.00
G	18.40	33.50	45.80	52.00	36.70	61.80	87.30	89.30	71.70	90.70	100.00	100.00
	30.60	46.30	62.80	67.70	49.80	74.40	94.30	95.10	81.80	96.70	100.00	100.00
β	54.10	37.50	76.20	83.10	87.50	61.20	98.40	99.00	99.70	85.30	100.00	100.00
	65.20	49.40	87.60	89.60	92.70	69.90	99.60	98.80	99.90	90.70	100.00	100.00
χ^2_3	48.60	44.20	76.50	82.50	84.60	73.40	98.40	98.60	99.90	94.50	100.00	100.00
	61.30	57.30	87.80	89.60	92.70	83.10	99.10	99.30	99.90	97.00	100.00	100.00
t_5	15.50	44.50	49.10	55.00	24.50	74.00	87.30	89.20	39.30	97.50	99.90	99.90
	25.00	59.50	63.00	67.90	35.40	84.70	93.70	94.90	51.10	99.50	100.00	100.00

Table 2.1: (Homoscedastic model, simple null hypothesis) Percentage of rejections for the normality null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

Table 2.6 for testing GOF to the Laplace distribution. Similar conclusions to those given for Tables 2.1 and 2.2. can be also expressed in this case.

2.5.3 Time consumed

Table 2.7 compares the PB and the WB (with raw and centered multipliers) in terms of the required CPU time. This table shows the CPU time consumed in seconds to get a p-value for testing GOF for the normal and the Laplace distributions in the homoscedastic (for both single and composite null hypothesis) and the heteroscedastic models with sample sizes $n = 25$, 50, 100, 200. Looking at this table it becomes evident that the WB is more efficient than the PB, in terms of the required computing time, specially for larger sample sizes. The difference in time when using the raw and the centered multipliers is rather small.

The gain in computational efficiency of the WB over the PB stems from the fact that

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	3.70	22.60	17.20	19.10	4.20	42.60	38.10	39.30	8.30	69.70	68.60	69.20
	8.40	30.90	25.00	27.20	9.10	51.80	48.10	50.20	14.60	77.60	78.10	78.20
LP	2.70	4.70	3.60	4.20	3.80	4.80	3.80	3.80	3.90	5.50	4.40	4.50
	7.30	9.40	7.70	8.90	8.20	10.60	8.00	9.20	8.90	9.20	9.00	9.10
LG	4.20	25.60	18.90	20.60	4.70	40.60	36.90	37.50	5.90	69.90	70.00	70.70
	7.50	35.00	28.50	31.20	9.30	48.80	46.60	47.50	11.90	78.30	78.30	79.00
G	6.00	23.30	17.70	18.80	11.60	41.60	36.60	38.20	27.10	67.10	68.20	68.90
	10.90	31.70	25.90	28.20	20.70	50.60	47.50	48.60	40.40	77.10	77.10	77.80
β	35.50	12.80	13.40	15.30	78.60	19.30	30.90	32.60	99.30	36.20	66.00	66.60
	48.80	19.20	21.80	24.30	86.20	27.70	43.20	44.50	99.60	46.60	74.00	75.60
χ^2_3	17.50	20.00	16.00	17.80	44.40	34.00	32.60	33.80	92.30	61.50	65.60	66.50
	27.20	27.60	24.00	25.70	59.10	44.00	43.70	44.60	96.90	72.00	76.20	76.70
t_5	3.20	21.50	16.20	18.10	5.40	39.30	35.00	36.70	8.70	71.60	70.80	71.40
	8.00	31.10	24.60	27.10	10.00	49.60	45.90	48.00	14.10	79.80	80.10	80.70

Table 2.2: (Homoscedastic model, simple null hypothesis) Percentage of rejections for the Laplace null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

one does not have to re-estimate the parameters at each iteration, which slows down the process considerably. Note that in the WB the parameter θ, the regression function $m(.)$ and the conditional variance function $\sigma(\cdot)$ are estimated only one time. For the WB approximation, once the set $\{m_{jk},\ 1 \le j \le k \le n\}$ is computed, the WB replicates $T^{*1}_{2,n,\omega}(\hat{\theta}), \ldots, T^{*B}_{2,n,\omega}(\hat{\theta})$ can be calculated very rapidly.

2.6 Conclusions

This chapter proposes a WB approximation for the null distribution of a test statistic for testing GOF to the error distribution in nonparametric models. It provides a consistent estimator. The WB and the PB share this property. Nevertheless, from a computational point of view, the WB approximation is more efficient, in the sense of requiring less computation time. The numerical examples support these attributes. In addition, in cases were the asymptotic null distribution does not depend on unknown quantities, the simulations carried out declare that, for small to moderate sample sizes, the WB provides

	$n = 25$			$n = 50$			$n = 100$		
	PB	WB1	WB2	PB	WB1	WB2	PB	WB1	WB2
N	6.50	5.60	7.30	5.20	4.80	5.60	5.40	5.10	5.20
	10.70	10.90	14.50	10.00	9.90	11.10	9.20	9.20	9.60
LP	29.90	15.30	21.50	33.60	40.40	43.50	38.30	80.50	81.60
	40.50	26.20	30.60	44.10	56.30	58.70	54.70	90.40	91.00
LG	30.30	44.10	50.50	47.80	86.60	89.00	94.90	99.90	99.90
	40.30	60.60	65.80	63.90	93.80	94.60	98.50	99.99	99.99
G	29.10	18.50	21.50	35.70	42.50	43.50	51.80	80.50	83.60
	43.50	29.20	30.60	51.30	58.30	59.70	66.10	90.40	95.90
β	18.00	16.40	20.40	23.30	39.40	42.70	67.30	80.80	82.10
	25.40	27.10	30.50	32.80	55.30	56.60	72.10	89.70	91.60
χ^2_3	37.30	51.40	53.80	58.90	77.30	80.70	83.10	89.90	91.30
	48.50	63.20	64.20	67.80	85.40	87.20	91.50	97.70	98.80
t_5	40.40	14.50	21.50	52.90	38.90	42.40	76.80	82.30	83.10
	58.70	28.70	31.40	69.20	53.70	56.00	88.20	89.50	90.30

Table 2.3: (Homoscedastic model, composite null hypothesis) Percentage of rejections for the normality null hypothesis at significance levels 5% (upper entry) and 10% (lower entry).

a better fit than the asymptotic distribution.

To derive the results in this chapter we considered certain estimators for the regression function and the conditional variance function. In addition, we assumed that the covariate was univariate. The results could be extended by considering other estimators (such as other local polynomial estimators) as well as covariates with higher dimension. The null distribution of other test statistics (for example, those based on the empirical CDF) could be similarly approximated.

	$n=25$			$n=50$			$n=100$		
	PB	WB1	WB2	PB	WB1	WB2	PB	WB1	WB2
N	53.20	56.90	58.80	62.80	64.40	66.20	69.30	71.40	77.20
	66.30	68.20	71.10	74.50	75.40	76.60	80.60	80.90	81.20
LP	4.30	3.80	4.50	4.60	4.60	4.40	5.00	4.70	4.90
	9.20	8.30	9.20	10.30	9.30	10.40	9.50	9.80	9.50
LG	52.40	48.20	50.50	60.40	58.50	60.20	74.60	77.50	78.50
	65.30	62.00	65.70	72.10	71.70	73.90	90.80	93.20	93.70
G	52.20	47.20	50.30	50.40	51.10	58.70	63.80	65.50	66.90
	64.30	60.70	64.60	62.20	61.50	73.20	80.40	82.30	83.10
β	50.50	57.00	62.90	55.80	60.60	65.60	76.40	83.50	87.70
	63.60	71.50	76.50	72.30	74.20	77.30	87.70	95.60	98.80
χ^2_3	37.50	67.30	70.60	41.40	78.50	80.10	43.50	88.00	88.40
	51.50	79.60	82.30	54.10	91.30	93.20	59.60	97.30	98.30
t_5	33.30	42.20	44.60	38.10	44.80	44.80	44.10	51.20	52.00
	46.40	52.80	56.70	52.30	56.40	58.90	60.90	65.00	65.80

Table 2.4: (Homoscedastic model, composite null hypothesis) Percentage of rejections for the Laplace null hypothesis at significance levels 5% (upper entry) and 10% (lower entry).

2.7 Appendix

2.7.1 Assumptions

(A.2) The weight function ω satisfies

$$\omega(t) = \omega(-t), \quad \forall t, \tag{2.10}$$

$\omega(t) \geq 0$, $\forall t$, and $\int t^4 \omega(t)dt < \infty$.

There is no restriction in assuming that the weight function $\omega(t)$ satisfies (2.10) because otherwise by defining $\omega_1(t) = 0.5\{\omega(t) + \omega(-t)\}$, which satisfies (2.10), we have that $T_{n,\omega}(\hat{\theta}) = T_{n,\omega_1}(\hat{\theta})$.

(A.3) $\varepsilon_1, \ldots, \varepsilon_n$ are IID with $E(\varepsilon_j^4) < \infty$ and $\varepsilon_1, \ldots, \varepsilon_n$ and $X_1, \ldots, X_n$ are independent.

Recall that by construction we have that $E(\varepsilon_j) = 0$ and $Var(\varepsilon_j) = 1$.

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	4.50	6.00	5.40	6.50	4.90	5.36	4.82	5.92	4.90	5.32	5.08	5.74
	10.30	10.80	10.20	12.50	10.50	10.70	10.12	11.74	9.40	10.30	10.64	11.24
LP	16.40	43.00	60.00	64.30	34.20	60.00	87.10	88.50	65.00	75.60	99.50	99.50
	23.40	54.00	70.40	73.70	44.50	71.40	92.10	92.80	73.80	82.30	99.80	99.80
LG	7.40	57.60	56.40	63.10	8.50	91.90	94.70	95.20	12.60	99.80	100.00	100.00
	12.10	70.40	72.00	76.00	15.00	95.90	97.40	98.10	20.20	99.90	100.00	100.00
G	19.40	39.10	56.40	63.10	36.90	68.10	94.70	95.20	67.20	93.60	100.00	100.00
	29.90	55.00	72.00	76.00	49.90	80.30	97.40	98.10	76.10	97.60	100.00	100.00
β	43.00	16.10	57.60	63.30	86.20	77.00	99.80	99.80	99.90	95.20	100.00	100.00
	56.20	26.10	70.00	74.70	92.30	86.20	100.00	100.00	100.00	97.60	100.00	100.00
χ^2_3	50.90	41.60	85.50	89.10	83.00	71.30	99.70	99.70	99.20	95.70	100.00	100.00
	61.80	54.80	92.50	93.90	91.00	83.10	99.90	99.90	99.70	98.80	100.00	100.00
t_5	9.20	51.00	59.10	65.40	15.90	80.20	92.90	94.30	27.90	99.00	100.00	100.00
	16.20	65.70	71.60	76.50	23.40	89.30	97.70	98.00	36.80	99.90	100.00	100.00

Table 2.5: (Heteroscedastic model) Percentage of rejections for the normality null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

(A.4) (i) X has a compact support S. (ii) f_X, m and σ are twice continuously differentiable on S. (iii) $\inf_{x\in S} f_X(x) > 0$ and $\inf_{x\in S} \sigma(x) > 0$.

(A.5) $nh_n^4 \to 0$, $nh_n^2/\ln n \to \infty$.

(A.6) K is a twice continuously differentiable symmetric pdf with compact support.

Assumptions (A.4)–(A.6) are mainly needed to guarantee the uniform consistency of the kernel estimators $\hat{f}_X(\cdot)$, $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ for $f_X(\cdot)$, $m(\cdot)$ and $\sigma(\cdot)$, respectively.

(A.7) The first partial derivatives $R'(t;\theta)$, $I'(t;\theta)$, $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $1 \le r \le p$ exist and are continuous functions $\forall t \in \mathbb{R}$, $\forall\theta$ in an open neighborhood of θ_1. In addition, $R'(t;\theta)$, $I'(t;\theta)$, $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $tR'(t;\theta)$, $tI'(t;\theta)$, $tR_{(r)}(t;\theta)$, $tI_{(r)}(t;\theta)$, $1 \le r \le p$, are bounded by functions in $L_2(\omega)$, $\forall\theta$ in an open neighborhood of θ_1.

The following assumption will be used for the maximum likelihood estimator of the parameter.

	$n=25$				$n=50$				$n=100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	2.00	31.80	25.00	27.10	2.60	55.30	51.20	52.50	2.80	86.10	85.70	86.20
	4.90	40.30	34.70	37.10	7.40	64.80	61.80	62.90	7.80	90.50	91.20	91.40
LP	2.10	4.60	3.70	4.60	3.00	5.70	4.00	4.40	3.60	4.40	4.00	4.40
	6.80	10.00	8.00	9.60	7.30	11.50	9.20	10.20	7.80	9.10	8.40	9.00
LG	2.10	33.80	27.10	29.30	2.30	54.80	50.80	52.30	3.10	85.00	84.40	84.70
	6.30	43.80	37.60	40.20	6.80	64.40	61.60	62.50	7.00	89.30	89.60	89.90
G	2.10	31.30	23.50	25.50	2.80	53.90	50.20	51.50	3.00	85.30	85.10	85.60
	6.70	41.10	34.40	37.10	6.80	65.10	62.70	63.70	7.50	91.10	91.10	91.50
β	3.00	19.20	18.40	21.00	6.00	33.50	43.20	45.90	27.60	56.70	81.20	81.50
	8.00	27.40	29.10	31.70	14.60	43.70	55.30	56.80	39.60	68.50	87.30	87.80
χ^2_3	2.70	22.30	18.60	20.80	3.40	43.10	42.90	44.50	5.60	78.40	81.30	81.90
	7.10	30.80	27.30	30.10	7.60	54.50	54.10	56.60	12.70	84.10	87.30	87.70
t_5	2.90	30.60	22.80	24.50	3.90	56.80	53.20	53.90	4.60	84.30	83.90	84.30
	6.30	41.50	33.70	38.00	6.50	66.70	64.30	65.20	9.40	90.20	90.40	90.70

Table 2.6: (Heteroscedastic model) Percentage of rejections for the Laplace null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

(A.9) The following functions exist $\forall\theta$ in an open neighborhood of θ_1: $u_r(x;\theta) = \frac{\partial}{\partial\theta_r}\log f(x;\theta)$ $u_{1,r}(x;\theta) = \frac{\partial^2}{\partial x\partial\theta_r}\log f(x;\theta)$, $u_{0,r,s}(x;\theta) = \frac{\partial^2}{\partial\theta_r\partial\theta_s}\log f(x;\theta)$, $u_{2,r}(x;\theta) = \frac{\partial^3}{\partial x^2\partial\theta_r}\log f(x;\theta)$, $u_{1,r,s}(x;\theta) = \frac{\partial^3}{\partial x\partial\theta_r\partial\theta_s}\log f(x;\theta)$, and satisfy $|u_{1,r}(a_1+a_2x;\theta)| \le b_{1,r}(x)$, with $xb_{1,r}(x)$, $b_{1,r}(x) \in L_2(F)$, $|u_{0,r,s}(a_1+a_2x;\theta)| \le b_{0,r,s}(x) \in L_2(F)$, $|u_{2,r}(a_1+a_2x;\theta)| \le b_{2,r}(x) \in L_2(F)$, $|u_{1,r,s}(a_1+a_2x;\theta)| \le b_{1,r,s}(x) \in L_2(F)$, $\forall a_1,\ a_2,\theta$ such that $|a_1|, |a_2-1|, |\theta-\theta_1| \le \delta$, for some small δ, $1 \le r,s \le p$. In addition, the following expectations exist: $E\{u_r(\varepsilon;\theta_1)u_s(\varepsilon;\theta_1)\}$, $E\{\varepsilon u_{1,r}(\varepsilon;\theta_1)\}$, $1 \le r,s \le p$.

The following assumption will be used for the method of moment estimator of the parameter, which assumes that under the null hypothesis, $\theta_0 = g(\mu_0)$, for some known function $g = (g_1,\ldots,g_p)^T$, $g_r : \mathbb{R}^{k-1} \to \mathbb{R}$, $1 \le r \le p$

(A.10) g_r is twice continuously differentiable at a neighborhood of μ_F, $1 \le r \le p$.

	Normal distribution			Laplace distribution		
n	PB/WB1	WB1	WB2	PB/WB1	WB1	WB2
25	2.72	0.71	0.74	3.49	1.00	1.01
	7.45	0.33	0.35	7.17	0.54	0.60
	4.42	0.31	0.34	5.34	0.50	0.55
50	5.61	0.71	0.70	7.51	1.08	1.09
	30.88	0.17	0.22	38.15	0.26	0.25
	15.63	0.19	0.19	23.68	0.28	0.25
100	12.15	0.84	0.86	23.40	1.11	1.12
	52.80	0.25	0.27	74.33	0.42	0.45
	30.64	0.25	0.26	64.56	0.37	0.39
200	27.56	1.25	1.27	76.37	1.54	1.58
	66.19	0.59	0.62	127.80	0.83	0.83
	41.14	0.56	0.58	117.51	0.78	0.76

Table 2.7: CPU time consumed for the calculation of one p-value in seconds for testing normality and Laplace distribution for the homoscedastic model and composite null hypothesis (upper entry), the heteroscedastic model (middle entry) and the homoscedastic model and single null hypothesis (lower entry).

2.7.2 Proof

We now sketch the proofs of the results stated in the previous sections, as well as some preliminary results. Along this section M denotes a generic positive constant taking many different values.

Lemma 1 *Suppose that assumptions (A.3)–(A.6) hold, then*

(a) $\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j - \hat{\varepsilon}_j)^2 = o_p(1)$.

(b) $\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^2 - \varepsilon_j^2)^2 = o_p(1)$.

(c) $\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^2 - 1)^2 = O_p(1)$.

(d) $\frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j^2 = O_p(1)$.

Proof First, observe that under the considered assumptions (see, for example, Masry 1996)

$$\sup_{x\in S}|\hat{m}(x) - m(x)| = o_p(n^{-1/4}), \tag{2.11}$$

$$\sup_{x \in S} |\hat{\sigma}(x) - \sigma(x)| = o_p(n^{-1/4}). \tag{2.12}$$

The difference between the residuals and the errors can be written as follows

$$\hat{\varepsilon}_j - \varepsilon_j = \varepsilon_j \left(\frac{\sigma(X_j) - \hat{\sigma}(X_j)}{\hat{\sigma}(X_j)} \right) + \left(\frac{m(X_j) - \hat{m}(X_j)}{\hat{\sigma}(X_j)} \right). \tag{2.13}$$

The results in (a)–(d) follow from (2.11)–(2.13). □

Lemma 2 *If* $\|\hat{\theta} - \theta_1\| = o_p(1)$ *and (A.7) holds, then*

(a) $\|t\{R'(t;\hat{\theta}) - R'(t;\theta_1)\}\|_\omega^2 = o_p(1)$, $\|t\{I'(t;\hat{\theta}) - I'(t;\theta_1)\}\|_\omega^2 = o_p(1)$.

(b) $\int \|\nabla R(t;\hat{\theta}) - \nabla R(t;\theta_1)\|^2 \omega(t)dt = o_p(1)$, $\int \|\nabla I(t;\hat{\theta}) - \nabla I(t;\theta_1)\|^2 \omega(t)dt = o_p(1)$.

(c) $\|R(t;\hat{\theta}) - R(t;\theta_1)\|_\omega^2 = o_p(1)$, $\|I(t;\hat{\theta}) - I(t;\theta_1)\|_\omega^2 = o_p(1)$.

(d) $\|t\{R(t;\hat{\theta}) - R(t;\theta_1)\}\|_\omega^2 = o_p(1)$, $\|t\{I(t;\hat{\theta}) - I(t;\theta_1)\}\|_\omega^2 = o_p(1)$.

Proof (a) From (A.7) $tR'(t;\theta) \in L_2(\omega)$, $\forall \theta$ in a neighborhood of θ_1. Since $\hat{\theta} \xrightarrow{P} \theta_1$, the integral $\int \{R'(t;\hat{\theta}) - R'(t;\theta_1)\}^2 t^2 \omega(t)dt$ is finite with probability tending to 1. Thus, $\forall \epsilon > 0, \exists M = M(\epsilon) > 0$ such that

$$\int_{\mathbb{R}\backslash[-M,M]} \{R'(t;\hat{\theta}) - R'(t;\theta_1)\}^2 t^2 \omega(t)dt < \epsilon, \tag{2.14}$$

with probability tending to 1. $tR'(t;\theta)$ is a uniformly continuous function in $[-M, M] \times B_\delta(\theta_1) = C$, where $B_\delta(\theta_1) = \{\theta : \|\theta - \theta_1\| \leq \delta\}$. Thus, $\forall \epsilon > 0$, $\exists\ \rho = \rho(\epsilon) > 0$ such that $\forall (t_a, \theta_a), (t_b, \theta_b) \in C$ satisfying $\|(t_a, \theta_a) - (t_b, \theta_b)\| < \rho$, we have $|t_1 R'(t_a;\theta_a) - t_2 R'(t_b;\theta_b)| < \epsilon/\iota$, with $\iota = \int \omega(t)dt$. As a consequence

$$\int_{-M}^{M} \{R'(t;\hat{\theta}) - R'(t;\theta_1)\}^2 t^2 \omega(t)dt < \epsilon, \tag{2.15}$$

with probability tending to 1. As ϵ is arbitrary, the result in (a) for the real part follows from (2.14) and (2.15). The proof for the imaginary part is parallel.
(b) The proof of this part is quite similar to that of part (a).
Parts (c) and (d) can be proven by applying the mean value theorem. □

Proof of Theorem 1 W^* can be expressed as $W^* = W_1 + W_2 + 2W_3$, where $W_3^2 \leq W_1 W_2$, $W_1 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^n Z_0(\varepsilon_j; t, \theta_1)\xi_j\|_\omega^2$, $W_2 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^n \{Z_0(\hat{\varepsilon}_j; t, \hat{\theta}) - Z_0(\varepsilon_j; t, \theta_1)\}\xi_j\|_\omega^2$. From the results in Dehling and Mikosch (1994),

$$\sup_x |P_* \{W_1 \leq x\} - P\{W_0 \leq x\}| \xrightarrow{a.s.} 0.$$

Thus, to show the result it suffices to see that $W_2 = o_{p_*}(1)$ in probability. With this aim, observe that W_2 can be expressed as $W_2 = \sum_{j=1}^{4} S_j + \sum_{j \neq k} S_{jk}$, with $S_{jk}^2 \leq S_j S_k$, $1 \leq j, k \leq 4$. In the proof of Theorem 2 it is given the expression of S_j and it is also proven that $S_j = o_{p_*}(1)$ in probability, $1 \leq j \leq 4$. This proves the result. □

Proof of Theorem 2 $T^*_{2,n,\omega}(\hat{\theta})$ can be expressed as $T^*_{2,n,\omega}(\hat{\theta}) = D_1 + D_2 + 2D_3$, where $D_3^2 \leq D_1 D_2$, $D_1 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_2(\varepsilon_j; t, \theta_1)\xi_j\|_\omega^2$, $D_2 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{Z_2(\hat{\varepsilon}_j; t, \hat{\theta}) - Z_2(\varepsilon_j; t, \theta_1)\}\xi_j\|_\omega^2$. From the results in Dehling and Mikosch (1994),

$$\sup_x |P_* \{D_1 \leq x\} - P\{T_2 \leq x\}| \xrightarrow{a.s.} 0.$$

Thus, to show the result it suffices to see that $D_2 = o_{p_*}(1)$ in probability. With this aim, observe that D_2 can be expressed as

$$D_2 = \sum_{j=1}^{10} S_j + \sum_{k<j} S_{jk},$$

with $S_{jk}^2 \leq S_j S_k$, $1 \leq j, k \leq 10$,

$$\begin{aligned}
&S_1 = \|\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^{n} \{\cos(t\varepsilon_j) - \cos(t\hat{\varepsilon}_j)\}\xi_j\|_\omega^2, \quad S_2 = \|\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^{n} \{\sin(t\varepsilon_j) - \sin(t\hat{\varepsilon}_j)\}\xi_j\|_\omega^2, \\
&S_3 = \|\tfrac{1}{\sqrt{n}} \{R(t;\hat{\theta}) - R(t;\theta_1)\} \left(\textstyle\sum_{j=1}^{n} \xi_j\right) \|_\omega^2, \quad S_4 = \|\tfrac{1}{\sqrt{n}} \{I(t;\hat{\theta}) - I(t;\theta_1)\} \left(\textstyle\sum_{j=1}^{n} \xi_j\right) \|_\omega^2, \\
&S_5 = \|\tfrac{t}{\sqrt{n}} \textstyle\sum_{j=1}^{n} \{\hat{\varepsilon}_j R(t;\hat{\theta}) - \varepsilon_j R(t;\theta_1)\}\xi_j\|_\omega^2, \quad S_6 = \|\tfrac{t}{\sqrt{n}} \textstyle\sum_{j=1}^{n} \{\hat{\varepsilon}_j I(t;\hat{\theta}) - \varepsilon_j I(t;\theta_1)\}\xi_j\|_\omega^2, \\
&S_7 = \|\tfrac{t}{2\sqrt{n}} \textstyle\sum_{j=1}^{n} \{(\hat{\varepsilon}_j^2 - 1)R'(t;\hat{\theta}) - (\varepsilon_j^2 - 1)R'(t;\theta_1)\}\xi_j\|_\omega^2, \\
&S_8 = \|\tfrac{t}{2\sqrt{n}} \textstyle\sum_{j=1}^{n} \{(\hat{\varepsilon}_j^2 - 1)I'(t;\hat{\theta}) - (\varepsilon_j^2 - 1)I'(t;\theta_1)\}\xi_j\|_\omega^2, \\
&S_9 = \|\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^{n} \{\psi_n^T(\hat{\varepsilon}_j;\hat{\theta})\nabla R(t;\hat{\theta}) - \psi_1^T(\varepsilon_j;\theta)\nabla R(t;\theta_1)\}\xi_j\|_\omega^2, \\
&S_{10} = \|\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^{n} \{\psi_n^T(\hat{\varepsilon}_j;\hat{\theta})\nabla I(t;\hat{\theta}) - \psi_1^T(\varepsilon_j;\theta)\nabla I(t;\theta_1)\}\xi_j\|_\omega^2.
\end{aligned}$$

We will show that $S_j = o_{p_*}(1)$ in probability, $1 \leq j \leq 10$. By the mean value theorem,

$$S_1 = \frac{1}{n} \sum_{j,k=1}^{n} \xi_j \xi_k (\varepsilon_j - \hat{\varepsilon}_j)(\varepsilon_k - \hat{\varepsilon}_k) \int t^2 \sin(t \overset{\sim}{\varepsilon}_j) \sin(t \overset{\sim}{\varepsilon}_k)\omega(t)dt,$$

where $\overset{\sim}{\varepsilon}_j = \alpha_j \varepsilon_j + (1 - \alpha_j)\hat{\varepsilon}_j$, for some $\alpha_j \in (0, 1)$. Then, from Lemma 1 (a),

$$E_*(S_1) \leq \frac{1}{n} \sum_{j=1}^{n} (\varepsilon_j - \hat{\varepsilon}_j)^2 \int t^2 \omega(t)dt = o_p(1),$$

which implies $S_1 = o_{p_*}(1)$ in probability. Analogously, $S_2 = o_{p_*}(1)$ in probability.

Since $S_3 = \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi_j\right)^2 \|R(t;\hat{\theta}) - R(t;\theta_1)\|_\omega^2$, the central limit theorem and Lemma 2 (c) imply that $S_3 = o_{p_*}(1)$ in probability. Analogously, $S_4 = o_{p_*}(1)$ in probability.

Observe that $S_5 = S_{51} + S_{52} + 2S_{53}$, with $S_{53}^2 \leq S_{51}S_{52}$,

$$S_{51} = \frac{1}{n}\sum_{j,k=1}^{n}(\hat{\varepsilon}_j - \varepsilon_j)(\hat{\varepsilon}_k - \varepsilon_k)\xi_j\xi_k\|tR(t;\hat{\theta})\|_\omega^2, \quad S_{52} = \frac{1}{n}\sum_{j,k=1}^{n}\varepsilon_j\varepsilon_k\xi_j\xi_k\|t\{R(t;\hat{\theta}) - R(t;\theta_1)\}\|_\omega^2.$$

From Lemma 1 (a) and Assumption (A.2), it follows that $E_*(S_{51}) = o_p(1)$ and thus $S_{51} = o_{p_*}(1)$, in probability. From Lemma 2 (d), it follows that $E_*(S_{52}) = o_p(1)$ and thus $S_{52} = o_{p_*}(1)$, in probability. Therefore, $S_5 = o_{p_*}(1)$, in probability. Analogously, $S_6 = o_{p_*}(1)$, in probability.

Observe that $S_7 = S_{71} + S_{72} + 2S_{73}$, with $S_{73}^2 \leq S_{71}S_{72}$,

$$S_{71} = \tfrac{1}{4}\tfrac{1}{n}\textstyle\sum_{j,k=1}^{n}(\hat{\varepsilon}_j^2 - 1)(\hat{\varepsilon}_k^2 - 1)\xi_j\xi_k\|t\{R'(t;\hat{\theta}) - R'(t;\theta_1)\}\|_\omega^2,$$
$$S_{72} = \tfrac{1}{4}\tfrac{1}{n}\textstyle\sum_{j,k=1}^{n}(\hat{\varepsilon}_j^2 - \varepsilon_j^2)(\hat{\varepsilon}_k^2 - \varepsilon_k^2)\xi_j\xi_k\|tR'(t;\theta_1)\|_\omega^2.$$

From Lemma 1 (c) and Lemma 2 (a), it follows that $E_*(S_{71}) = o_p(1)$ and thus $S_{71} = o_{p_*}(1)$, in probability. From Lemma 1 (b) and (A.7), it follows that $E_*(S_{72}) = o_p(1)$ and thus $S_{72} = o_{p_*}(1)$, in probability. Therefore, $S_7 = o_{p_*}(1)$, in probability. Analogously, $S_8 = o_{p_*}(1)$, in probability.

Observe that $S_9 = S_{91} + S_{92} + 2S_{93}$, with $S_{93}^2 \leq S_{91}S_{92}$,

$$S_{91} = \|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^{n}\{\psi_n(\hat{\varepsilon}_j;\hat{\theta}) - \psi_1(\varepsilon_j;\theta_1)\}^T\nabla R(t;\hat{\theta})\xi_j\|_\omega^2,$$
$$S_{92} = \|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^{n}\psi_1(\varepsilon_j;\theta_1)^T\{\nabla R(t;\hat{\theta}) - \nabla R(t;\theta_1)\}\xi_j\|_\omega^2.$$

From (2.7) and (A.7), it follows that $E_*(S_{91}) = o_p(1)$ and thus $S_{91} = o_{p_*}(1)$, in probability. From (A.1) and Lemma 2 (b), it follows that $E_*(S_{92}) = o_p(1)$ and thus $S_{92} = o_{p_*}(1)$, in probability. Therefore, $S_9 = o_{p_*}(1)$, in probability. Analogously, $S_{10} = o_{p_*}(1)$, in probability. This completes the proof. □

Proof of Corollary 2 From Theorem 2 it follows that $T_{2,n,\omega}^*(\hat{\theta}) = O_{p_*}(1)$ in probability. From Theorem 2 in Hušková and Meintanis (2010), $\frac{T_{n,\omega}(\theta)}{n} \xrightarrow{P} \kappa > 0$. These two facts imply the result. □

Lemma 3 *Suppose that* $\|\hat{\theta} - \theta_1\| = o_p(1)$*, for some* $\theta_1 \in \Theta$*, and that assumptions (A.3)–(A.6), (A.9) hold, then*

(a) $\frac{1}{n}\sum_{j=1}^{n}\|\nabla \log f(\hat{\varepsilon}_j;\hat{\theta}) - \nabla \log f(\varepsilon_j;\theta_1)\|^2 = o_p(1)$.

(b) $\hat{A}_{n,rs}(\hat{\theta}) = A_{F,rs}(\theta_1) + o_p(1)$, $1 \leq r, s \leq p$.

(c) $\hat{\rho}_1(\hat{\theta}) = \rho_{1,F}(\theta_1) + o_p(1)$.

(d) $\hat{\rho}_2(\hat{\theta}) = \rho_{2,F}(\theta_1) + o_p(1)$.

Proof (a) From the mean value theorem and (A.9),

$$\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_{j=1}^n \left\{\frac{\partial}{\partial\theta_r}\log f(\hat{\varepsilon}_j;\hat{\theta}) - \frac{\partial}{\partial\theta_r}\log f(\varepsilon_j;\theta_1)\right\}^2 \\
&\quad = \tfrac{1}{n}\textstyle\sum_{j=1}^n \left\{\frac{\partial^2}{\partial\varepsilon\partial\theta_r}\log f(\tilde{\varepsilon}_j;\tilde{\theta})(\hat{\varepsilon}_j-\varepsilon_j) + \sum_{s=1}^p \frac{\partial^2}{\partial\theta_r\partial\theta_s}\log f(\tilde{\varepsilon}_j;\tilde{\theta})(\hat{\theta}_s-\theta_{1s})\right\}^2 \\
&\quad \le S_{r,1} + S_{r,2} + 2S_{r,3},
\end{aligned}$$

with $S_{r,3}^2 \le S_{r,1}S_{r,2}$, $\tilde{\varepsilon}_j = (1-\alpha_j)\hat{\varepsilon}_j + \alpha_j\varepsilon_j$, for some $\alpha_j \in (0,1)$, $1 \le j \le n$, $\tilde{\theta} = (1-\alpha)\hat{\theta} + \alpha\theta_1$, for some $\alpha \in (0,1)$,

$$S_{r,1} = \|\hat{\theta} - \theta_1\|^2 \frac{1}{n}\sum_{j=1}^n\sum_{s=1}^p b_{0,r,s}^2(\varepsilon_j)$$

and

$$S_{r,2} = \frac{1}{n}\sum_{j=1}^n b_{1,r}^2(\varepsilon_j)(\hat{\varepsilon}_j - \varepsilon_j) = o_p(1).$$

From (A.9), (2.11)–(2.13), it follows that $S_{r,1} = o_p(1)$, $S_{r,2} = o_p(1)$, $1 \le r \le p$. This proves (a).

The proof of parts (b)–(d) follows similar steps to that of part (a). □

Proof of Theorem 4 Observe that $\frac{1}{n}\sum_{j=1}^n \|\psi_{1n}(\hat{\varepsilon}_j;\hat{\theta}) - \psi(\varepsilon_j;\theta_1)\|^2 \le D_1 + D_2 + D_3 + D_4$, with $D_4^2 \le \sum_{j\neq k} D_jD_k$,

$$\begin{aligned}
D_1 &= \frac{1}{n}\sum_{j=1}^n \|\hat{A}_n(\hat{\theta})^{-1}\nabla\log f(\hat{\varepsilon}_j;\hat{\theta}) - A_F(\theta_1)^{-1}\nabla\log f(\varepsilon_j;\theta_1)\|^2, \\
D_2 &= \frac{1}{n}\sum_{j=1}^n \|\hat{\varepsilon}_j\hat{\rho}_1(\hat{\theta}) - \varepsilon_j\rho_{F,1}(\theta_1)\|^2, \\
D_3 &= \frac{1}{n}\sum_{j=1}^n \|\frac{\hat{\varepsilon}_j^2 - 1}{2}\hat{\rho}_2(\hat{\theta}) - \frac{\varepsilon_j^2 - 1}{2}\rho_{F,2}(\theta_1)\|^2.
\end{aligned}$$

By using the results in Lemmas 1 and 3 one obtain $D_j = o_p(1)$, $1 \le j \le 3$, and hence the result. □

Proof of Theorem 5 From (2.11)–(2.13),

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \hat{\varepsilon}_j^s = \frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^s + \frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^{s-1}\frac{m(X_j) - \hat{m}(X_j)}{\hat{\sigma}(X_j)} + \frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^s\frac{\sigma(X_j) - \hat{\sigma}(X_j)}{\hat{\sigma}(X_j)} + o_p(1). \tag{2.16}$$

Taking into account the following facts

(m.1) $\sup_{x\in S}\left|\frac{\hat{m}(x)-m(x)}{\hat{\sigma}(x)}-\frac{\hat{m}(x)-m(x)}{\sigma(x)}\right|=o_p(n^{-1/2})$,

(m.2) $\sup_{x\in S}\left|\hat{m}(x)-m(x)-\frac{1}{nf_X(x)}\sum_{k=1}^{n_v}K_{h_n}(x-X_k)\sigma(X_k)\varepsilon_k\right|=o_p(n^{-1/2})$,

it follows that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s-1}\frac{m(X_j)-\hat{m}(X_j)}{\hat{\sigma}(X_j)}=\frac{-1}{n\sqrt{n}}\sum_{j,k=1}^{n}\varepsilon_j^{s-1}\varepsilon_k\frac{\sigma(X_k)}{f_X(X_j)\sigma(X_j)}K_{h_n}(X_j-X_k)+o_p(1).$$

Now, by using projections, we get (see, for example, the proof of Theorem 2 in Pardo-Fernández et al. (2015a) for a similar development)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s-1}\frac{m(X_j)-\hat{m}(X_j)}{\hat{\sigma}(X_j)}=-\mu_{F,s-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j+o_p(1). \tag{2.17}$$

Next we deal with the third term in the right-hand side of (2.16). Taking into account the following facts

(s.1) $\sup_{x\in S}\left|\frac{\hat{\sigma}(x)-\sigma(x)}{\hat{\sigma}(x)}-\frac{\hat{\sigma}(x)-\sigma(x)}{\sigma(x)}\right|=o_p(n^{-1/2})$,

(s.2) $\sup_{x\in S}\left|\hat{\sigma}(x)-\sigma(x)-\frac{\hat{\sigma}^2(x)-\sigma^2(x)}{2\sigma(x)}\right|=o_p(n^{-1/2})$,

(s.3) $\sup_{x\in S}\left|\hat{\sigma}^2(x)-\sigma^2(x)-\frac{1}{nf_X(x)}\sum_{j=1}^{n}K_{h_n}(X_j-x)\left[\{Y_j-m(x)\}^2-\sigma^2(x)\right]\right|=o_p(n^{-1/2})$,

it follows that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s}\frac{\sigma(X_j)-\hat{\sigma}(X_j)}{\hat{\sigma}(X_j)}$$

$$=\frac{1}{2n\sqrt{n}}\sum_{j,k=1}^{n}\varepsilon_j^{s}\frac{1}{f_X(X_j)\sigma^2(X_j)}K_{h_n}(X_j-X_k)\left[\sigma^2(X_j)-\{Y_k-m(X_j)\}^2\right]+o_p(1).$$

Now, by using projections, we get (see, for example, the proof of Lemma 11 in Pardo-Fernández et al. (2015b) for a similar development)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s}\frac{\sigma(X_j)-\hat{\sigma}(X_j)}{\hat{\sigma}(X_j)}=-\frac{\mu_{F,s}}{2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\varepsilon_j^2-1)+o_p(1). \tag{2.18}$$

The result follows from (2.16)–(2.18).

Proof of Theorem 6 Notice that

$$\hat{\mu}_s-\mu_{F,s}=\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^s-\varepsilon_j^s)+\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^s-\mu_{F,s}).$$

From (2.11)–(2.13), the first term in the right-hand side of the above equality is $o_p(1)$; from the SLLN, the second term in the right-hand side of the above equality is $o(1)$ a.s. Therefore $\hat{\mu}_s-\mu_{F,s}=o_p(1)$, $2\le s\le k$. The result follows from this fact and (A.10). □

2.8 References

1. Bickel P.J., Doksum K.A., 2001. Mathematical Statistics. Prentice Hall.

2. Burke M.D., 2000. Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Statist Probab Lett 46, 13–20.

3. Dehling H., Mikosch T., 1994. Random quadratic forms and the bootstrap for U-statistics. J Multivariate Anal 51, 392–413.

4. de Uña-Álvarez J., 2013. Comments on: An updated review of Goodness-of-Fit tests for regression models. Test 22, 414–418.

5. Epps T.W., 2005. Tests for location-scale families based on the empirical characteristic function. Metrika 62, 99–114.

6. Epps T.W., Pulley L.B., 1983. A test for normality based on the empirical characteristics function. Biometrika 70, 723-726.

7. Feller W., 1971. An Introduction to Probability Theory and its Applications. Vol 2. Wiley.

8. González-Manteiga W., Crujeiras R., 2013. An updated review of Goodness-of-Fit tests for regression models. Test 22, 361–411.

9. Heuchenne C., Van Keilegom I., 2010. Goodness of fit tests for the error distribution in nonparametric regression. Comput Statist Data Anal 54, 1942–1951.

10. Hušková M., Janssen, P., 1993. Consistency of the generalized bootstrap for degenerate U-statistics. Ann Statist 21, 1811–1823.

11. Hušková M., Meintanis S.G., 2010. Test for the error distribution in nonparametric possiby heteroscedatic regression models. Test 19, 92–112.

12. Jiménez-Gamero M.D., 2013. Comments on: An updated review of Goodness-of-Fit tests for regression models. Test 22, 412–413.

13. Jiménez-Gamero M.D., Alba-Fernández V., Muñoz-García J., Chalco-Cano Y., 2009. Goodness-of-fit tests based on empirical characteristics functions. Comput Statist Data Anal 53, 3957-3971.

14. Jiménez-Gamero M.D., Kim H-M., 2015. Fast goodness-of-fit test based on the characteristic function. Comput Statist Data Anal 89, 172–191.

15. Kojadinovic I., Yan J., 2012. Goodness-of-fit testing based on a weighted bootstrap: A fast sample alternative to the parametric bootstrap. Can J Statist 40, 480–500.

16. Masry E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. J Time Series Anal 17(6), 571–600.

17. Neumeyer N., Dette H., Nagel E-R., 2006. Bootstrap test for the error distribution in linear and nonparametric regression models. Aust N Z J Stat 48(2), 129–156.

18. Pardo-Fernández J.C., Jiménez-Gamero M.D., El Ghouch A., 2015a. A nonparametric ANOVA-type test for regression curves based on characteristic functions. Scand J Stat 42, 197–213.

19. Pardo-Fernández J.C., Jiménez-Gamero M.D., El Ghouch A., 2015b. Tests for the equality of conditional variance functions in nonparametric regression. Electron J Stat 9, 1826–1851.

20. R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna-Austria, . URL http://www.R-project.org/.

21. Sperlich S., 2013. Comments on: An updated review of Goodness-of-Fit tests for regression models. Test 22, 419–427.

22. White H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

23. Zhu LX., 2005. Lecture Notes in Statistics. Nonparametric Monte Carlo test and their applications. Springer.

Capítulo 3

Fast test for comparison of the error distributions based on the characteristic function

Abstract

A test for the equality of error distributions in two nonparametric regression models is proposed. The test statistic is based on comparing the empirical characteristic functions of the residuals calculated from independent samples of the models. The asymptotic null distribution of the test statistic cannot be used to estimate its null distribution because it is unknown, since it depends on the unknown common error distribution. To approximate the null distribution, a weighted bootstrap estimator is studied, providing a consistent estimator. The finite sample performance of this approximation as well as the power of the resulting test are evaluated by means of a simulation study. The procedure can be extended to testing for the equality of $d > 2$ error distributions.

3.1 Introduction

Let (X_1, Y_1) and (X_2, Y_2) be two independent random vectors. Assume that they satisfy the general nonparametric regression models,

$$Y_k = m_k(X_k) + \sigma_k(X_k)\varepsilon_k, \tag{3.1}$$

where $m_k(x) = E(Y_k \mid X_k = x)$ is the regression function, $\sigma_k^2(x) = Var(Y_k \mid X_k = x)$ is the conditional variance function and ε_k is the regression error, which is assumed to be independent of X_k, $k = 1, 2$. By construction $E(\varepsilon_k) = 0$ and $Var(\varepsilon_k)$=1. The regression functions, the variance functions, the distributions of the error and that of the covariates are unknown and no parametric models are assumed for them. We are interested in testing for the equality of the error distributions, that is, in tests of the null hypothesis

$$H_0 : F_1 = F_2,$$

versus the alternative

$$H_1 : F_1 \neq F_2,$$

where F_1, F_2 stand for the cumulative distribution function (CDF) of ε_1 and ε_2, respectively. In view of the uniqueness of the characteristic function (CF), the null hypothesis may equivalently be stated as

$$H_0 : C_1 = C_2,$$

versus the alternative

$$H_1 : C_1 \neq C_2,$$

where C_k denotes the CF corresponding to F_k, that is, $C_k(t) = \int \exp(\mathrm{i}tx)\ dF_k(x) = R_k(t) + \mathrm{i}I_k(t)$, $k = 1, 2$.

The equality of the error distributions is a usual assumption in several statistical problems such as that of testing for the equality of regression curves (see, for example, Young and Bowman 1995; Hall and Hart 1990 and Kulasequera and Wang 2001). The equality of the error distributions may considerably simplify some procedures. For instance, under H_0, the asymptotic null distribution of the test statistic for the equality of regression functions in Pardo-Fernández et al. (2015a) coincides with the one of the classical ANOVA test for comparing means (it also requires the equality of the densities of the covariates). Another example is given by the asymptotic null distribution of the test statistic for the equality of variance functions in Pardo-Fernández et al. (2015b) based on CFs. The authors prove that, when the covariates are identically distributed and the error distributions are equal, the asymptotic distribution coincides with that of the classical Levene test for comparing variances. Thus, the problem of testing for H_0 is of considerable practical interest.

The problem of testing whether two samples come from the same population has generated a considerable amount of research. Many different approaches have been proposed to deal with this problem when the data are observable (see, for example, the references in Meintanis 2005; Alba-Fernández et al. 2008; Hobza et al. 2014; Baringhaus and Kolbe 2015 and Modarres 2016). In our setting the errors are not observable and the inference must be based on the residuals, which are not independent even if the original data are. The number of proposals to deal with this case is not so big. Mora (2005) proposed tests for testing H_0 when the regression models are linear; for the more general model given in (1.1), Pardo-Fernández (2007) -PF07 from now on- also proposed tests for H_0. These two papers study Kolgomorov-Smirnov (KS) and Cramér-von Misses (CvM) type test statistics based on the empirical CDFs of the residuals. Since the null distribution of these test statistics is unknown, these papers use a smooth bootstrap to approximate the critical values. Two main problems with these procedures are: they assume strong conditions on the distributions of the errors which, among other things, are supposed to have a smooth density; in addition, although quite easy to implement, the bootstrap can become computationally expensive as the sample sizes of the data increase.

In this chapter is proposes a test for H_0 that is based on comparing the empirical CF (ECF) of the residuals in samples from the models. It can be seen as a residual version of the test in Alba-Fernández et al. (2008), designed for the two-sample problem when observable independent and identically distributed (IID) data are available from each population. A weighted bootstrap (WB) estimator, in the sense of Burke (2000), is proposed to approximate the critical values. This method has been previously used in Kojadinovic and Yan (2012) and Ghoudi and Rémillard (2014), to approximate the null distribution of goodness-of-fit tests based on the empirical CDF, in Jiménez-Gamero and Kim (2015), to approximate the null distribution of goodness-of-fit tests based on the ECF, in Quessy and Éthier (2012), for the two-sample problem for dependent data, and in Jiménez-Gamero et al. (2016), for the two-sample problem for observable independent data, among others. In view of the good properties of the WB in these and other papers, it is also expected to work well for approximating the null distribution of the test statistic proposed in this chapter.

Compared to the tests in PF07, the procedure suggested in this chapter has two main advantages. First, it assumes less stringent assumptions on the distribution of the regression errors. Specifically, the we do not assume that the error distribution has a probability density function. Thus, the method can be applied when such distribution is arbitrary: continuous, discrete or mixed. Secondly, the WB approximation is computationally more efficient than that based on the smooth bootstrap. The chapter is organized as follows.

Section 3.2 describes the test statistic. The problem of approximating the null distribution of the proposed test statistic is dealt with in Section 3.3, where the use of a WB estimator it is studied. The consistency of the resulting null distribution estimator is proved. It is also shown that the resulting test is consistent, in the sense of being able to detect any alternative. Some practical issues are addressed in Section 3.4. Section 3.5 reports the results of some simulation experiments designed to study the finite sample performance of the proposed approximation, to compare it with other methods as well as a real data application. From this numerical study it is concluded that the WB approximation works, in the sense of providing levels close to the nominal values, and that the power of the test is comparable or even greater than the power of the test based on the empirical CDF. Section 3.6 shows how the proposed test can be extended to the general case of comparison of $d > 2$ error distributions. All proofs and technical details are deferred to the last section.

The following notation will be used along the chapter: all vectors are column vectors; the superscript ^{T} denotes transpose; $1_n \in \mathbb{R}^n$ has all its components equal to 1; if $x \in \mathbb{R}^k$, with $x' = (x_1, \ldots, x_k)$, then $diag(x)$ is the $k \times k$ diagonal matrix whose (i,i) entry is x_i, $1 \leq i \leq k$; P_0, E_0 and Cov_0 denote probability, expectation and covariance, respectively, by assuming that the null hypothesis is true; P_*, E_* and Cov_* denote the conditional probability law, expectation and covariance, given the data, respectively; $\stackrel{\mathcal{L}}{\to}$ denotes convergence in distribution; $\stackrel{P}{\to}$ denotes convergence in probability; $\stackrel{a.s.}{\to}$ denotes the almost sure convergence; for any complex number $z = a + ib$, $|z|$ is its modulus; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; for a given non-negative real-valued function ω we denote $\|\cdot\|_\omega$ to the norm and $\langle \cdot, \cdot \rangle_\omega$ to the scalar product in the Hilbert space $L^2(\omega) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 \omega(t) dt < \infty\}$.

3.2 The test statistic

Let (X_{kj}, Y_{kj}), $j = 1, 2, \ldots, n_k$, be an IID sample from (X_k, Y_k) satisfying (3.1), and let $\varepsilon_{k1}, \ldots, \varepsilon_{kn_k}$ denote the associated errors, $k = 1, 2$. Since the hypothesis H_0 establishes the equality of the distributions of the errors ε_{kj} and they are not observable, the inference must be based on the residuals,

$$\hat{\varepsilon}_{kj} = \frac{Y_{kj} - \hat{m}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}, \ j = 1, 2, \ldots, n_k, \tag{3.2}$$

where $\hat{m}_k$ and $\hat{\sigma}_k$ are estimators of m_k and σ_k, respectively, $k = 1, 2$. Several choices are possible for $\hat{m}_k$ and $\hat{\sigma}_k$. Here, as in PF07, we use the following kernel estimators for the regression function m_k and the variance function σ_k^2,

$$\hat{m}_k(x) = \sum_{j=1}^{n_k} W_{kj}(x; h_k) Y_{kj}, \quad x \in S,$$

$$\hat{\sigma}_k^2(x) = \sum_{j=1}^{n_k} W_{kj}(x; h_k) \left\{Y_{kj} - \hat{m}_k(x)\right\}^2, \quad x \in S,$$

where

$$W_{kj}(x; h_k) = \frac{K_{h_k}(X_{kj} - x)}{\sum_{s=1}^{n_k} K_{h_k}(X_{ks} - x)}, \quad x \in S,$$

$K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$, $K(\cdot)$ is a kernel and h_k is the bandwidth, satisfying certain conditions that will be specified later. The proposed test statistic takes the form

$$T_{n_1,n_2} = \|\hat{C}_1 - \hat{C}_2\|_\omega^2,$$

where

$$\hat{C}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \exp(\mathrm{i}t\hat{\varepsilon}_{kj}) = \hat{R}_k(t) + \mathrm{i}\hat{I}_k(t), \tag{3.3}$$

$$\hat{R}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \cos(t\hat{\varepsilon}_{kj}), \quad \hat{I}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \sin(t\hat{\varepsilon}_{kj}),$$

$k = 1, 2$, and $\omega(t)$ is a non-negative function.

In order to study properties of T_{n_1,n_2} some assumptions will be required. Next we list them.

(A.1) All limits in this work are taken when $n_1, n_2 \to \infty$ in such a way that

$$\lim \frac{n_1}{N} = \tau, \quad \text{for some} \quad \tau \in (0, 1),$$

where $N = n_1 + n_2$.

(A.2) The weight function $\omega(t)$ is a non-negative symmetric function,

$$\omega(t) = \omega(-t), \quad \forall t,$$

and $\int t^4 \omega(t) dt < \infty$.

There is no restriction in assuming that the weight function $\omega(t)$ is symmetric because otherwise by defining $\omega_1(t) = 0.5\{\omega(t) + \omega(-t)\}$, which is is clearly symmetric, we have that

$$\|C_1 - C_2\|_\omega = \|C_1 - C_2\|_{\omega_1},$$

for any two CFs C_1 and C_2. Note that the symmetry of ω implies that

$$T_{n_1,n_2} = \|\hat{R}_1 - \hat{R}_2 + \hat{I}_1 - \hat{I}_2\|_\omega^2.$$

The following assumption will be required to ensure that $\hat{m}_k$ and $\hat{\sigma}_k$ provide consistent estimators of m_k and σ_k, respectively.

(A.3) For $k = 1, 2$,

(i) $E(\varepsilon_k^4) < \infty$.

(ii) X_k is absolutely continuous with compact support S and density function f_k.

(iii) f_k, m_k and σ_k are two times continuously differentiable.

(iv) $\inf_{x\in S} f_k > 0$ and $\inf_{x\in S} \sigma_k > 0$.

(v) $n_k h_k^4 \to 0$ and $n_k h_k^2 / \ln n_k \to \infty$.

(vi) K is a symmetric density function with compact support and twice continuously differentiable.

For simplicity we assume that the same kernel function, K, is used in both populations. Nevertheless, the results to be stated remain true if different kernels are used, whenever they satisfy Assumption (A.3)(vi).

In order to give a sound justification of T_{n_1,n_2} as a test statistic for testing H_0 we next derive its limit.

Theorem 1 *Suppose that Assumptions (A.1)–(A.3) hold, then* $T_{n_1,n_2} \xrightarrow{P} \kappa = \|C_1 - C_2\|_\omega^2$.

Note that $\kappa \geq 0$. If H_0 is true then $\kappa = 0$. Since two distinct characteristic functions can be equal in a finite interval (Feller 1971, p. 506), a general way to ensure that $\kappa = 0$ iff H_0 is true is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$. Thus, a reasonable test for testing H_0 should reject the null hypothesis for large values of T_{n_1,n_2}. Now, to determine what are large values we must calculate its null distribution, or at least an approximation to it. This is the topic of the next section.

3.3 Approximating the null distribution

The null distribution of T_{n_1,n_2} is clearly unknown, so it must be approximated. We first try to estimate it by means of its asymptotic null distribution.

Let

$$R_k'(t) = -\int x \sin(tx) dF_k(x), \quad I_k'(t) = \int x \cos(xt) dF_k(t),$$

$k = 1, 2$. Note that under the null hypothesis $R_1(t) = R_2(t) = R(t)$, $I_1(t) = I_2(t) = I(t)$, $R_1'(t) = R_2'(t) = R'(t)$ and $I_1'(t) = I_2'(t) = I'(t)$.

Theorem 2 *Suppose that Assumptions (A.1)–(A.3) hold. If H_0 is true, then*

$$\frac{n_1 n_2}{N} T_{n_1,n_2} \xrightarrow{\mathcal{L}} \|Z\|_\omega^2,$$

where $\{Z(t),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance kernel $\varrho_0(t,s) = Cov_0\{Z_0(\varepsilon;t), Z_0(\varepsilon;s)\}$ and

$$\begin{aligned} Z_0(\varepsilon;t) &= \cos(t\varepsilon) + t\varepsilon I(t) - t\frac{\varepsilon^2 - 1}{2} R'(t) - R(t) \\ &\quad + \sin(t\varepsilon) - t\varepsilon R(t) - t\frac{\varepsilon^2 - 1}{2} I'(t) - I(t). \end{aligned}$$

Remark 1 *The asymptotic null distribution of $\frac{n_1 n_2}{N} T_{n_1,n_2}$ can be expressed as*

$$\|Z\|_\omega^2 \stackrel{d}{=} \sum_{j \geq 1} \lambda_j Z_j^2, \tag{3.4}$$

where $\stackrel{d}{=}$ stands from the equality in distribution, $Z_1, Z_2, \ldots$ are independent standard normal variables and the set $\{\lambda_j\}_{j\geq 1}$ are the non-zero eigenvalues of the integral equation

$$\int \varrho_0(t,s) g_j(t) w(t) dt = \lambda_j g_j(s)$$

with corresponding eigenfunctions $\{g_j(\cdot)\}_{j\geq 1}$.

From Remark 1 it becomes evident that the asymptotic null distribution of T_{n_1,n_2} does not provide a useful approximation to its null distribution since it depends on the unknown common distribution. So, we next study another way of approximating it by means of a WB estimator.

Let

$$C_\tau(t) = \tau C_1(t) + (1-\tau) C_2(t) = R_\tau(t) + \mathrm{i} I_\tau(t).$$

Let $\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2}$ be IID random variates with mean 0 and variance 1, which are independent of $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1}), (X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$. We define the following WB version of T_{n_1,n_2},

$$T^*_{1,n_1,n_2} = \|C^*_1 - C^*_2\|^2_\omega,$$

where

$$C^*_k(t) = \frac{1}{n_k}\sum_{j=1}^{n_k} \xi_{kj} Z_{k,\tau}(\varepsilon_{kj}; t),$$

$$\begin{aligned} Z_{k,\tau}(\varepsilon; t) &= \cos(t\varepsilon) + t\varepsilon I_k(t) - t\frac{\varepsilon^2-1}{2} R'_k(t) - R_\tau(t) \\ &\quad + \sin(t\varepsilon) - t\varepsilon R_k(t) - t\frac{\varepsilon^2-1}{2} I'_k(t) - I_\tau(t), \end{aligned} \tag{3.5}$$

$k = 1, 2$. The next result gives the weak limit of the conditional distribution of T^*_{1,n_1,n_2}, given the data $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1}), (X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$.

Theorem 3 *Suppose that Assumptions (A.1)–(A.3) hold, then*

$$\sup_x \left| P_* \left\{ \frac{n_1 n_2}{N} T^*_{1,n_1,n_2} \le x \right\} - P\left\{ T_\tau \le x \right\} \right| \xrightarrow{P} 0,$$

where $T_\tau = \|Z_\tau\|^2_\omega$, $\{Z_\tau(t),\, t \in \mathbb{R}\}$ *is a centered Gaussian process on* $L_2(\omega)$ *with covariance kernel* $\varrho_\tau(t,s) = (1-\tau)\varrho_{1,\tau}(t,s) + \tau\varrho_{2,\tau}(t,s)$ *and* $\varrho_{k,\tau}(t,s) = E\{Z_{k,\tau}(\varepsilon_k; t) Z_{k,\tau}(\varepsilon_k; s)\}$, $k = 1, 2$.

The result in Theorem 3 is valid whether the null hypothesis is true or not. If H_0 holds, then the kernels $\varrho_0(t,s)$ and $\varrho_\tau(t,s)$ coincide. Therefore, a direct consequence of Theorems 2 and 3 is that the conditional distribution of T^*_{1,n_1,n_2}, given the data, provides a consistent estimator of the distribution of T_{n_1,n_2} when H_0 is true. However, from a practical point of view, this result is useless because the function $Z_{k,\tau}(\varepsilon_{kj}; t)$ depends on the non-observable error ε_{kj} and on the unknown values of the functions $R_k(t)$, $I_k(t)$, $R'_k(t)$ and $I'_k(t)$, $j = 1, \ldots, n_j$, $k = 1, 2$. To overcome these difficulties we replace ε_{kj} by $\hat{\varepsilon}_{kj}$, $R_k(t)$ by $\hat{R}_k(t)$, $I_k(t)$ by $\hat{I}_k(t)$, $R'_k(t)$ by $\hat{R}'_k(t)$ and $I'_k(t)$ by $\hat{I}'_k(t)$, with

$$\hat{R}'_k(t) = -\frac{1}{n_k}\sum_{j=1}^{n_k} \hat{\varepsilon}_{kj}\sin(t\hat{\varepsilon}_{kj}), \quad \hat{I}'_k(t) = \frac{1}{n_k}\sum_{j=1}^{n_k} \hat{\varepsilon}_{kj}\cos(t\hat{\varepsilon}_{kj}).$$

So, instead of T^*_{1,n_1,n_2}, now we consider

$$T^*_{2,n_1,n_2} = \|\hat{U}^*_1 - \hat{U}^*_2\|^2_\omega.$$

where

$$\begin{aligned}\hat{U}_k^*(t) = & \ \tfrac{1}{n_k}\sum_{j=1}^{n_k}\Big\{\cos(t\hat{\varepsilon}_{kj}) + t\hat{\varepsilon}_{kj}\hat{I}_k(t) - t\tfrac{\hat{\varepsilon}_{kj}^2-1}{2}\hat{R}_k'(t) - \hat{R}_\tau(t) \\ & +\sin(t\hat{\varepsilon}_{kj}) - t\hat{\varepsilon}_{kj}\hat{R}_k(t) - t\tfrac{\hat{\varepsilon}_{kj}^2-1}{2}\hat{I}_k'(t) - \hat{I}_\tau(t)\Big\}\xi_{kj},\end{aligned} \tag{3.6}$$

$k = 1, 2$, and

$$\hat{R}_\tau(t) = \frac{n_1}{N}\hat{R}_1(t) + \frac{n_2}{N}\hat{R}_2(t), \quad \hat{I}_\tau(t) = \frac{n_1}{N}\hat{I}_1(t) + \frac{n_2}{N}\hat{I}_2(t).$$

The next theorem states that replacing ε_{kj} by $\hat{\varepsilon}_{kj}$, …, $I_k'(t)$ by $\hat{I}_k'(t)$ in the expression of T^*_{1,n_1,n_2} has no asymptotic effect, in the sense that T^*_{1,n_1,n_2} and T^*_{2,n_1,n_2} both have the same conditional asymptotic distribution, given the data. Observe that all quantities involved in the definition of T^*_{2,n_1,n_2} are known, thus, in principle, one could be able to know, or at least to accurately approximate its conditional distribution, given the data. This practical issue will be handled in Section 3.4.

Theorem 4 *Suppose that Assumptions (A.1)–(A.3) hold, then*

$$\sup_x \left| P_*\left\{\frac{n_1 n_2}{N} T^*_{2,n_1,n_2} \le x\right\} - P\left\{T_\tau \le x\right\}\right| \xrightarrow{P} 0,$$

where T_τ is as defined in Theorem 3.

The result in Theorem 4 is valid whether the null hypothesis H_0 is true or not. As observed before for T^*_{1,n_1,n_2}, an immediate consequence of this fact and Theorem 2 is the following.

Corollary 1 *If H_0 is true and the assumptions in Theorem 4 hold, then*

$$\sup_x \left| P_*\left\{\frac{n_1 n_2}{N} T^*_{2,n_1,n_2} \le x\right\} - P\left\{\frac{n_1 n_2}{N} T_{n_1,n_2} \le x\right\}\right| \xrightarrow{P} 0.$$

Let $\alpha \in (0,1)$. For testing H_0 we consider

$$\Psi_* = \begin{cases} 1, & \text{if} \quad T_{n_1,n_2} \ge t^*_{2,n_1,n_2,\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $t^*_{2,n_1,n_2,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of T^*_{2,n_1,n_2}, or equivalently, $\Psi_* = 1$ if $p^* \le \alpha$, where $p^* = P_*\left\{T^*_{2,n_1,n_2} \ge T_{n_1,n_2,obs}\right\}$ and $T_{n_1,n_2,obs}$ is the observed value of the test statistic. The result in Corollary 1 states that Ψ_* is asymptotically correct, in the sense that its type I error probability is asymptotically equal to the nominal value α.

From Theorems 1, 2 and 4, it readily follows the next result.

Corollary 2 *Suppose that H_0 is not true, the assumptions in Theorem 4 hold and ω is such that*

$$\kappa = \|C_1 - C_2\|_\omega^2 > 0, \tag{3.7}$$

then $P(\Psi_* = 1) \to 1$.

Corollary 2 shows that, if ω is such that (3.7) holds, then the test Ψ_* is consistent in the sense of being able to asymptotically detect any (fixed) alternative. As discussed before, a general way to ensure (3.7) is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$.

Remark 2 *The results so far stated keep on being true if instead of using the raw multipliers, $\xi_{11}, \ldots, \xi_{1n_1}$, $\xi_{21}, \ldots, \xi_{2n_2}$, we use the centered multipliers, $\xi_{11} - \bar{\xi}_1, \ldots, \xi_{1n_1} - \bar{\xi}_1$, $\xi_{21} - \bar{\xi}_2, \ldots, \xi_{2n_2} - \bar{\xi}_2$, where $\bar{\xi}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} \xi_{jk}$, $k = 1, 2$, as suggested in Burke (2000) and Kojadinovic and Yan (2012).*

Remark 3 *In Remark 1 we saw that the null distribution of T_{n_1,n_2} is a linear combination of independent χ^2 variables, the weights in that linear combination being the eigenvalues $\{\lambda_j\}_{j\geq 1}$ of certain functional. Routine algebra shows that the conditional distribution of T^*_{2,n_1,n_2}, given the data, can be also expressed as a linear combination of a certain finite set of variables, $\sum_{j=1}^N \hat{\lambda}_j W_j^2$, where $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$ are the eigenvalues of the symmetric $N \times N$-matrix M_2, that will be defined in next section (see equations* (3.9) *and* (3.10)*), and $(W_1, \ldots, W_N) = (\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2})H$, H being the matrix containing the eigenvectors associated to the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$, that is, $M_2 = H\, diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)H^T$. What really happens is that the set $\{\hat{\lambda}_j\}_{j=1}^N$ converges to $\{\lambda_j\}_{j\geq 1}$ (see Delhing and Mikosch, 1994).*

Remark 4 *From Remark 3 it becomes evident that the conditional distribution of T^*_{2,n_1,n_2}, given the data, depends on the distribution of $(W_1, \ldots, W_N)$. The distribution of this random vector is, in general, unknown. For the special case where the multipliers come from a standard normal distribution, the vector $(W_1, \ldots, W_N)$ has independent components distributed according to a standard normal distribution, and thus the conditional distribution of T^*_{2,n_1,n_2}, given the data, is a finite linear combination of independent χ^2 variables, where the weights in the linear combination are $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$. Note that in this special case the WB distribution of the test statistic is of the same type as its asymptotic null distribution.*

3.4 On the practical calculation

This section describes some computational issues related to the calculation of the test statistic T_{n_1,n_2} and the WB approximation to its null distribution.

3.4.1 Calculation of the test statistic

Let

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad A = \begin{pmatrix} \frac{1}{n_1^2} 1_{n_1} 1_{n_1}^T & \frac{-1}{n_1 n_2} 1_{n_1} 1_{n_2}^T \\ \frac{-1}{n_1 n_2} 1_{n_2} 1_{n_1}^T & \frac{1}{n_2^2} 1_{n_2} 1_{n_2}^T \end{pmatrix},$$

with $M_{rs} = (\varphi_\omega(\hat{\varepsilon}_{rj} - \hat{\varepsilon}_{sv}))_{1\le j\le n_r,\, 1\le v\le n_v}$, $r, s = 1, 2$, and

$$\varphi_\omega(x) = \int \cos(t'x)\omega(t)dt. \tag{3.8}$$

Let v be the vector of $\mathbb{R}^N$ with the first n_1 components equal to $1/n_1$ and the rest equal to $-1/n_2$. In practice, the test statistic T_{n_1,n_2} can be computed by using the following expression (see Lemma 1 in Alba et al., 2008)

$$\begin{aligned} T_{n_1,n_2} &= \frac{1}{n_1^2}\sum_{j,r=1}^{n_1} \varphi_\omega(\hat{\varepsilon}_{1j} - \hat{\varepsilon}_{1r}) + \frac{1}{n_2^2}\sum_{l,v=1}^{n_2} \varphi_\omega(\hat{\varepsilon}_{2l} - \hat{\varepsilon}_{2v}) \\ &\quad - \frac{2}{n_1 n_2}\sum_{j=1}^{n_1}\sum_{l=1}^{n_2} \varphi_\omega(\hat{\varepsilon}_{1j} - \hat{\varepsilon}_{2l}) \\ &= v^T M v \\ &= 1_N^T M_1 1_N, \end{aligned}$$

with $M_1 = M \odot A$, $\odot$ denoting the Hadamard product.

The WB version of T_{n_1,n_2}, T^*_{2,n_1,n_2}, can be expressed as

$$T^*_{2,n_1,n_2} = \xi^T M_2 \xi,$$

with $\xi^T = (\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2})$ and

$$M_2 = M_3 \odot A, \quad M_3 = \begin{pmatrix} M_{3,11} & M_{3,12} \\ M_{3,21} & M_{3,22} \end{pmatrix}, \tag{3.9}$$

$$M_{3,rs} = \left(\int Z_{r,\tau}(\hat{\varepsilon}_{rj}; t) Z_{s,\tau}(\hat{\varepsilon}_{sv}; t)\omega(t)dt \right)_{1\le j\le n_r,\, 1\le v\le n_v}, \tag{3.10}$$

$r, s = 1, 2$. An explicit expression for M_3 is given in the Appendix.

3.4.2 Calculation of the WB distribution of the test statistic

Normal multipliers. As observed in Remark 4, if the multipliers has a normal distribution then, conditional on the data, T^*_{2,n_1,n_2} is distributed as $W = \sum_{j=1}^{N} \hat{\lambda}_j \chi^2_{1,j}$, where $\hat{\lambda}_1, \cdots, \hat{\lambda}_N$ are the eigenvalues of M_2 and $\chi^2_{1,1}, \cdots, \chi^2_{1,N}$ are independent variables having a chi-squared distribution with 1 degree of freedom. The law of W can be numerically approximated by using, for example, Imhof's method (Imhof, 1961). In this case, the WB estimator of the p-value can be calculated as follows:

Algorithm 1

1. Calculate the residuals $\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{2n_2}$.
2. Calculate the observed value of the test statistic T_{n_1,n_2}, $T_{n_1,n_2,obs}$.
3. Calculate the eigenvalues of M_2, $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$.
4. Approximate the p-value by $\hat{p}^* = P_* \left(\sum_{j=1}^{N} \hat{\lambda}_j \chi^2_{1,j} > T_{n_1,n_2,obs} \right)$.

Arbitrary multipliers. As also observed in Remark 4, the WB distribution of T_{n_1,n_2} is unknown for arbitrary multipliers. Nevertheless, the WB p-value estimator can be easily approximated by simulation as follows. Let $\Delta(u) = 1$ if $u > 0$ and $\Delta(u) = 0$ if $u \leq 0$.

Algorithm 2

1. Calculate the residuals $\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{2n_2}$.
2. Calculate the observed value of the test statistic T_{n_1,n_2}, $T_{n_1,n_2,obs}$.
3. Calculate M_2.
4. For some large integer B, repeat the following steps for every $b \in \{1, \ldots, B\}$:
   (a) Generate $\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2}$ IID variables with mean 0 and variance 1.
   (b) Calculate $T^{*b}_{2,n_1,n_2} = \xi^T M_2 \xi$.
5. Approximate the p-value by $\hat{p}^* = \frac{1}{B} \sum_{b=1}^{B} \Delta(T^{*b}_{2,n_1,n_2} - T_{n_1,n_2,obs})$.

3.5 Numerical results

3.5.1 Finite sample performance

The properties so far studied are asymptotic. In order to empirically investigate the performance of the proposed test for finite sample sizes, we carried out a simulation experiment. The objective of this experiment is fourfold: first, to study the goodness of the WB approximation to the null distribution of the test statistic; second, to analyze the WB approximation in terms of power, comparing it to the power that results when the bootstrap employed in PF07 is used to approximate the null distribution of the proposed test statistic (denoted as Boot in the tables); third, to compare the power of the proposed test to the CvM type test in PF07 (denoted as CM in the tables) (the KS test is not considered in our simulation study because, in the simulations carried out in PF07, it was less powerful than the CvM test); and finally, to compare the WB and the bootstrap approximations in terms of the CPU time required. This section reports and summarizes the results obtained. All computations have been performed by using programs written in the R language (R Core Team, 2015). Specifically, to numerically approximate the WB p-value by Imhof's method the R package `CompQuadForm` (Duchesne and Lafaye de Micheaux, 2010) was used.

Three specifications for the functions m_k and σ_k were considered:

S1: $Y_{kj} = X_{kj} + X_{kj}^2 + (X_{kj} + 0.5)\varepsilon_{kj}$, $1 \leq j \leq n_k$, $k = 1, 2$,

S2: $Y_{kj} = X_{kj} + 0.5\varepsilon_{kj}$, $1 \leq j \leq n_k$, $k = 1, 2$,

S3: $Y_{1j} = X_{1j} + X_{1j}^2 + (X_{1j} + 0.5)\varepsilon_{1j}$, $1 \leq j \leq n_1$, and $Y_{2j} = X_{2j} + 0.5\varepsilon_{2j}$, $1 \leq j \leq n_2$,

with $X_{kj} \sim U(0, 1)$, $1 \leq j \leq n_k$, $k = 1, 2$. For each of these specifications for m_k and σ_k, the following three cases were considered for the error distribution:

(i) $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$,

(ii) $\varepsilon_1 \sim N(0, 1)$, $\varepsilon_2 \sim E(1) - 1$,

(iii) $\varepsilon_1 \sim N(0, 1)$, $\varepsilon_2 \sim U(-\sqrt{3}, \sqrt{3})$,

where $E(1)$ stands for a negative exponential law with mean 1. Case (i) corresponds to the null hypothesis, while cases (ii) and (iii) are alternatives.

To estimate the regression function and the conditional variance the Epanechnikov kernel $K(u) = 0.75 \times (1 - u^2)$ was employed.

As weight function for T_{n_1,n_2} we took $\omega(t) = \exp(-\beta t^2)$. This weight function has been considered in many other test statistics involving ECFs (see, for example, the tests in Alba et al., 2008, Meintanis, 2005, 2015, Pardo-Fernández et al., 2015a,b, among others).

Another issue is the choice of h. Since the choice of the bandwidth for tests based on smoothing remains an open issue (see, for example, de Uña-Álvarez, 2013, González-Manteiga and Crujeiras, 2013, Sperlich, 2013), we proceeded as in the simulation studies in PF07 and Pardo-Fernández et al. (2015a): we took $h_k = c \times n_k^{-a}$, where c and a are real constants. To decide the values for a and c, we performed an extensive simulation experiment with the aim of selecting those values giving type I error probabilities closest to the nominal values. We also tried several values for β, specifically $\beta = \{0.05, 0.10, 0.15, 0.20, 0.25\}$. In general, better results -in the sense of agreement between the observed type I error probabilities and the target values- were obtained for $\beta = 0.15$. Because of this reason we fixed $\beta = 0.15$ in all simulations.

1000 samples with sizes $n_1, n_2 \in \{50, 100\}$ were generated for each case and each specification for m_k and σ_k. For each sample, to approximate the WB p-value of the observed value of the test statistic, we applied Algorithm 2, with raw and centered multipliers generated from a standard normal distribution and $B = 1000$, and Algorithm 1. In simulations we observed that, as expected, these approximations provided quite similar values. Nevertheless, the WB with centered multipliers gives slightly better results, in the sense of yielding type I error probabilities which are a bit closer to the nominal values than the other two. Because of this reason, we recommend its use. All results displayed in the tables were obtained by using Algorithm 2 with centered multipliers. To approximate the bootstrap p-value we proceeded as in PF07, generating 200 bootstrap samples. The tables report the fractions of p-values less than or equal to 0.05 and 0.10.

Tables 3.1–3.3 display the results for the level. Looking at them it can be concluded that for n_1, $n_2 = 100$ all choices for a and c in these tables give values very similar to the true value of α, for all specifications and for all tests. In general, $a = 0.3$ and $c = 1.0$ give quite reasonable results, so we set these values for a and c to study the power.

Table 3.4 displays the results for the power. In case (ii) all tests have similar power for all considered specifications; in case (iii) the test proposed in this chapter exhibits larger power than the one based on the empirical CDF. As for the WB and bootstrap approximations to the null distribution of T_{n_1,n_2}, the bootstrap test is slightly more powerful than the one based on the WB approximation. Nevertheless, as the sample size increases, the power of both tests become closer. This was also observed in Kojadinovic and Yan (2012) and Ghoudi and Rémillard (2014) for goodness-of-fit tests based on the empirical CDF. As will be seen a bit later, the practical importance of this fact resides in that for larger sample sizes the bootstrap becomes extremely time consuming.

We also compared the bootstrap and the WB approximations in terms of the required computing time. To calculate the WB approximation we used Algorithm 1 (denoted as WB1 in Table 3.5) and Algorithm 2 (denoted as WB2 in Table 3.5). For the comparisons

a	n_1, n_2	$c = 1.00$			$c = 1.25$			c=1.50		
		CM	Boot	WB	CM	Boot	WB	CM	Boot	WB
0.30	50,50	4.60	4.50	6.40	6.00	6.00	6.10	6.00	6.10	5.90
		8.60	14.50	10.90	11.00	12.50	11.30	10.50	11.00	11.80
	50,100	5.40	6.50	5.10	4.50	4.00	5.10	4.00	4.40	5.10
		10.40	9.00	9.60	9.00	8.90	10.00	9.50	10.20	10.30
	100,100	4.40	4.20	5.60	6.00	5.50	5.30	4.00	5.40	4.80
		9.40	9.30	10.20	10.50	11.50	10.00	9.00	10.50	9.60
0.35	50,50	6.50	6.50	6.30	6.00	4.50	6.40	6.00	6.00	6.20
		10.50	13.50	11.10	10.50	12.50	10.80	11.50	12.50	11.20
	50,100	4.00	4.00	5.10	4.50	5.50	5.00	5.50	5.00	4.90
		11.00	9.50	9.50	10.50	9.50	9.80	9.00	8.00	10.00
	100,100	5.50	4.00	5.20	5.00	5.20	5.60	4.50	5.50	5.50
		11.50	9.20	9.60	11.50	9.00	10.00	11.00	12.00	9.60
0.40	50,50	6.00	5.00	6.10	4.00	6.00	6.40	6.00	4.50	6.40
		9.00	13.00	10.10	11.00	14.00	10.20	11.50	14.50	10.80
	50,100	5.00	6.50	5.20	4.30	4.00	5.10	6.50	4.50	4.90
		12.00	12.50	9.30	10.50	9.00	9.60	9.00	9.00	9.50
	100,100	6.00	4.10	5.70	5.70	5.50	5.20	4.70	5.50	5.60
		9.50	9.00	9.40	10.90	11.50	9.70	11.00	8.50	9.90
0.45	50,50	6.00	3.50	5.70	6.00	6.00	5.90	4.50	6.50	6.40
		8.50	13.00	11.80	9.50	13.50	11.40	11.50	13.50	11.20
	50,100	6.50	5.50	5.00	5.00	6.00	5.40	7.00	6.50	5.10
		10.00	12.50	10.30	12.50	13.00	9.00	11.00	10.00	9.30
	100,100	6.00	6.00	5.80	5.80	5.50	5.70	4.60	4.50	5.30
		10.90	11.00	10.00	9.50	9.00	9.40	11.50	9.00	9.70

Table 3.1: Percentage of rejections at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specification S1.

to be fair, we took $B = 1000$ for the the bootstrap and the WB2 estimators. As for the raw and the centered multipliers, the difference in the required computing time is negligible. Table 3.5 shows the CPU time consumed in seconds to get a p-value for testing the equality of the error distribution for several sample sizes. Looking at this table it becomes evident that WB2 is more efficient than the bootstrap approximation, in terms of the required

		$c = 1.00$			$c = 1.25$			c=1.50		
a	n_1, n_2	CM	Boot	WB	CM	Boot	WB	CM	Boot	WB
0.30	50,50	5.90	6.00	5.90	7.00	6.50	6.50	6.50	6.00	6.20
		11.30	11.00	11.20	12.00	11.00	11.40	12.00	9.50	11.60
	50,100	5.30	4.50	4.50	4.50	4.00	5.10	5.50	4.00	4.40
		10.60	9.50	10.50	11.50	9.00	10.10	9.00	9.00	9.30
	100,100	5.90	4.00	5.50	4.50	4.50	5.60	5.50	5.50	5.30
		10.90	11.50	10.30	9.00	9.10	10.20	9.00	11.00	10.10
0.35	50,50	5.00	5.00	6.20	7.00	5.50	5.90	6.50	6.50	6.20
		9.00	13.00	11.60	10.00	11.00	11.10	11.50	11.00	11.60
	50,100	5.50	5.50	4.90	5.00	4.00	4.60	4.00	6.50	4.90
		11.50	10.50	10.60	11.50	10.50	10.30	12.50	10.00	10.00
	100,100	4.50	4.50	5.90	4.50	5.00	5.40	4.00	5.50	5.50
		9.50	9.00	9.60	9.50	9.00	10.10	9.00	9.50	9.80
0.40	50,50	5.00	4.50	6.30	5.50	5.00	6.30	6.00	6.00	5.90
		8.00	11.50	11.70	10.00	13.00	11.60	10.00	11.00	11.10
	50,100	5.00	4.50	4.90	5.50	5.50	5.00	4.50	4.00	4.60
		12.00	11.00	9.60	10.00	10.50	10.20	11.00	9.50	10.10
	100,100	4.50	5.50	5.70	4.50	5.50	5.90	5.00	6.00	5.50
		9.00	9.50	9.30	9.50	9.50	9.50	10.00	10.00	10.20
0.45	50,50	5.00	4.50	6.00	5.00	4.50	6.40	5.50	5.00	6.30
		8.50	10.00	12.00	11.00	12.00	11.60	9.00	13.00	11.40
	50,100	5.50	4.00	4.80	6.50	4.50	4.80	5.50	5.00	4.80
		12.50	11.00	10.00	12.00	12.00	9.20	11.50	10.50	10.10
	100,100	4.50	5.50	6.60	4.00	4.00	5.80	5.50	6.00	6.00
		9.00	9.50	10.00	9.00	9.00	9.20	9.00	11.00	9.50

Table 3.2: Percentage of rejections at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specification S2.

computing time, specially for larger sample sizes. The difference between WB1 and WB2 is really small. The gain in computational efficiency of the WB over the bootstrap stems from the fact that one does not have to re-estimate m and σ at each iteration, which slows down the process considerably; by contrast, for the WB approximation, once the matrix M_3 is calculated, the WB replicates $T^{*1}_{2,n_1,n_2}, \dots, T^{*B}_{2,n_1,n_2}$ are calculated very rapidly.

a	n_1, n_2	$c = 1.00$			$c = 1.25$			c=1.50		
		CM	Boot	WB	CM	Boot	WB	CM	Boot	WB
0.30	50,50	4.00	7.00	6.40	7.50	6.50	6.10	6.00	7.00	5.90
		10.00	12.50	10.90	11.50	13.50	11.30	12.50	10.50	11.80
	50,100	5.20	5.00	5.10	6.50	5.50	5.10	6.00	4.00	5.10
		11.60	11.00	9.60	10.50	9.00	10.00	11.50	9.00	10.30
	100,100	5.80	4.00	5.60	4.50	4.50	5.30	4.50	5.50	4.80
		9.80	9.50	10.20	10.50	11.50	10.00	11.50	11.00	9.60
0.35	50,50	4.00	4.00	6.20	5.50	6.50	6.60	7.00	6.50	6.80
		11.50	13.50	11.40	13.50	12.00	11.10	11.50	13.00	12.40
	50,100	6.50	6.50	5.00	7.00	7.50	5.20	4.50	5.50	4.90
		11.00	11.50	9.50	10.00	9.50	9.80	9.00	11.50	10.20
	100,100	4.00	4.00	5.50	4.00	6.00	5.40	6.50	6.50	5.50
		9.00	9.10	10.10	10.00	12.50	10.30	9.50	11.50	10.40
0.40	50,50	5.50	5.50	6.10	7.50	5.00	6.30	5.50	5.50	6.40
		9.50	12.00	11.70	11.50	13.50	11.40	13.00	12.50	11.20
	50,100	5.00	5.50	5.10	6.50	5.00	4.90	6.00	4.00	5.00
		12.50	12.50	9.50	11.00	12.00	9.60	11.50	10.00	9.70
	100,100	4.50	5.50	5.90	4.00	4.00	5.50	4.80	4.00	5.30
		9.50	11.50	9.50	9.00	9.50	10.10	11.00	12.00	10.30
0.45	50,50	5.50	4.00	6.00	6.00	6.00	6.50	4.50	4.50	6.30
		11.50	14.50	12.00	9.50	12.00	11.80	11.50	13.50	11.30
	50,100	7.00	5.50	5.40	5.00	5.00	5.10	4.50	5.00	4.70
		13.00	11.50	10.40	12.00	13.50	9.50	10.50	11.50	9.70
	100,100	5.50	5.00	6.10	5.50	5.50	5.90	4.30	6.00	5.60
		10.50	11.50	10.00	9.00	9.00	9.60	11.50	9.00	10.20

Table 3.3: Percentage of rejections for at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specification S3.

Finally, we ran simulations when the error distributions come from a mixed distribution. Specifically, we considered the following cases:

(iv)

$$\varepsilon_1, \varepsilon_2 \sim \begin{cases} 0, & \text{with probability } 0.20, \\ N(0, \sqrt{5/4}) & \text{with probability } 0.80, \end{cases}$$

	case	n_1, n_2	$\alpha = 0.05$			$\alpha = 0.10$		
			CM	Boot	WB	CM	Boot	WB
S1	(ii)	50,50	76.00	75.60	75.10	85.60	85.30	84.00
		50,100	88.20	87.60	88.30	92.60	92.70	90.80
		100,100	95.60	97.70	96.80	98.00	99.40	98.70
	(iii)	50,50	14.40	23.20	20.60	27.60	35.20	29.10
		50,100	27.20	29.70	26.20	40.60	44.20	42.00
		100,100	33.40	49.00	46.60	47.60	64.20	61.10
S2	(ii)	50,50	76.20	75.90	71.00	81.00	83.90	82.70
		50,100	83.60	87.50	86.90	89.20	92.80	91.80
		100,100	96.70	97.70	97.90	97.90	99.10	98.70
	(iii)	50,50	16.50	24.40	18.70	23.00	35.80	29.30
		50,100	20.80	29.00	24.60	31.90	43.60	38.00
		100,100	40.50	50.00	47.00	55.00	64.90	61.80
S3	(ii)	50,50	72.20	73.00	70.10	82.80	84.00	78.90
		50,100	85.60	84.50	83.50	90.00	91.80	90.20
		100,100	98.00	97.00	96.00	99.20	98.80	97.70
	(iii)	50,50	18.40	27.30	20.30	31.60	38.90	32.80
		50,100	28.60	31.90	27.30	43.20	48.00	43.60
		100,100	37.60	53.60	51.20	54.40	68.00	65.40

Table 3.4: Percentage of rejections at the significance levels 5% and 10% for cases (ii) and (iii) and specifications S1–S3.

n_1, n_2	Boot/WB2	WB1	WB2
50,50	13.12	1.10	1.20
50,100	19.43	1.20	1.35
100,100	25.05	1.32	1.45
100,150	36.13	1.41	1.70
150,150	35.63	1.53	1.85

Table 3.5: CPU time in seconds for the calculation of a p-value.

(v)

$$\varepsilon_1 \sim \begin{cases} 0, & \text{with probability } 0.20, \\ N(0, \sqrt{5/4}) & \text{with probability } 0.80, \end{cases}$$

$$\varepsilon_2 \sim \begin{cases} 0, & \text{with probability } 0.50, \\ N(0, \sqrt{2}) & \text{with probability } 0.50. \end{cases}$$

Case (iv) corresponds to the null hypothesis and case (v) is an alternative. In practice, these cases could model a situation where the observations come from two devices, one of them with no measurement error. The test in PF07 cannot be applied in this setting because it requires the error distribution to have a smooth density. Table 3.6 displays the obtained results for the test proposed in this chapter for $n_1, n_2 = 100$. Again the empirical levels are close to the target values and the test has power against the alternative.

		$c = 1.00$		$c = 1.25$		c=1.50	
	a	(v)	(vi)	(v)	(vi)	(v)	(vi)
S1	0.30	5.00	52.00	5.50	51.00	5.00	47.50
		10.50	70.00	10.00	68.50	10.00	66.50
	0.35	5.50	52.00	5.50	51.50	5.00	50.50
		10.00	69.50	10.50	70.00	9.50	69.50
	0.40	5.40	50.50	5.50	51.50	5.50	52.00
		10.00	69.50	10.00	69.50	10.00	70.50
S2	0.30	5.50	47.50	5.00	44.00	5.00	42.50
		10.00	65.50	9.00	60.00	9.00	56.50
	0.35	5.00	50.50	5.00	48.00	5.50	45.00
		9.00	68.00	10.50	65.50	9.50	60.50
	0.40	5.00	51.50	5.00	50.00	5.00	49.00
		10.00	71.50	10.00	68.00	10.00	65.00
S3	0.30	5.00	51.50	5.00	48.00	5.00	46.50
		10.00	69.50	10.50	66.00	11.50	62.50
	0.35	5.00	51.50	5.00	51.00	5.00	49.00
		10.50	69.50	10.00	69.50	10.50	67.50
	0.40	5.50	55.00	5.00	52.00	5.50	51.50
		10.50	70.00	10.50	69.50	9.50	69.00

Table 3.6: Percentage of rejections at the significance levels 5% (upper entry) and 10% (lower entry) for cases (v) and (vi) and specifications S1–S3.

3.5.2 Real data analysis

Finally, we applied the proposed test to a real data set. To estimate the p-value we applied Algorithm 2 with $B = 1000$. Several values for a and c were tried. Next we briefly describe it.

Young and Bowman (1995) proposed a method for testing the equality and parallelism of two or more smooth curves. Their method assumes that the errors are equally distributed in each population. They applied their method to a data set consisting of the yield (g/plant) and density (plants/m^2) of White Spanish Onions from two South Australian localities, namely Purnong Landing (first group, 42 observations) and Virginia (second group, 42 observations). This data set is available in the R package `sm` (Bowman and Azzalini, 2014). Table 3.7 displays the estimated p-values when using the test proposed in this chapter for testing the equality of the error distributions. As in Young and Bowman (1995), the test was applied on the logarithm of the data. Looking at this table we see that the equality of the error distribution cannot be rejected.

		$\beta = 0.05$			$\beta = 0.15$			$\beta = 0.25$		
a	c	1.00	1.25	1.50	1.00	1.25	1.50	1.00	1.25	1.50
0.30		0.961	0.743	0.496	0.944	0.570	0.393	0.920	0.521	0.363
0.35		0.959	0.943	0.760	0.989	0.850	0.579	0.976	0.840	0.529
0.40		0.863	0.970	0.946	0.861	0.996	0.883	0.790	0.989	0.848
0.45		0.570	0.903	0.971	0.422	0.915	0.996	0.345	0.846	0.988

Table 3.7: p-values for the data set.

3.6 Testing for the equality of $d > 2$ error distributions

The proposed test can be extended to testing for the equality of $d > 2$ error distributions as follows. Let (X_k, Y_k), $1 \le k \le d$, be d independent random vectors satisfying the general nonparametric regression model (3.1), $1 \le k \le d$. Let F_k and $C_k = R_k + \mathrm{i}I_k$ denote de CDF and the CF of ε_k, respectively, $1 \le k \le d$. Suppose that independent samples are available from each population: $(X_{k1}, Y_{k1}), \ldots, (X_{kn_k}, Y_{kn_k})$, $1 \le k \le d$. Let

$N = n_1 + \ldots + n_d$. For testing

$$H_{0d}:\ F_1 = \ldots = F_d \quad \Leftrightarrow \quad C_1 = \ldots = C_d,$$

against the general alternative

$$H_{1d}:\ H_{0d}\,\text{is not true},$$

for observable data, Hušková and Meintanis (2008) have proposed to compare the ECF associated to the sample from each population to the ECF of all available data which, under H_{0d}, estimates the common CF, say $C = C_1 = \ldots = C_d$. A residual version of such test can be used for testing H_{0d} in our setting. Specifically, let $\hat{\varepsilon}_{kj}$, $1 \le j \le n_k$, $1 \le k \le d$, be defined as in (3.2) and let

$$T_N = \sum_{k=1}^{d} n_k \|\hat{C}_k - \hat{C}\|_\omega^2,$$

where $\hat{C}_k$ is as defined in (3.3) and

$$\hat{C} = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{C}_k.$$

Analogue results to those given in Theorems 1, 2 and 4 can be given for T_N. Next we state them without proofs because they closely follows those provided for $d = 2$.

Theorem 5 *Suppose that $n_k/N \to \tau_k > 0$, $1 \le k \le d$, Assumptions (A.2) and (A.3) hold for all $1 \le k \le d$, then $\frac{1}{N}T_N \xrightarrow{P} \sum_{k=1}^{d} \tau_k \|C_k - C_0\|_\omega^2$, with $C_0 = \sum_{k=1}^{d} \tau_k C_k$.*

Theorem 6 *Suppose that assumptions in Theorem 5 hold. If H_0 is true, then*

$$T_N \xrightarrow{\mathcal{L}} \sum_{k=1}^{d} \|Z_k - \sqrt{\tau_k} Z_0\|_\omega^2,$$

where $\{Z_k(t),\, t \in \mathbb{R}\}$, $k = 1, \ldots, d$, are d IID centered Gaussian processes on $L_2(\omega)$ with covariance kernel $\varrho_0(t,s)$ as defined in Theorem 2 and $Z_0 = \sum_{k=1}^{d} \sqrt{\tau_k} Z_k$.

Now let $\xi_{1,1}, \ldots, \xi_{1,n_1}, \ldots, \xi_{d,1}, \ldots, \xi_{d,n_d}$ be IID random variates with mean 0 and variance 1, which are independent of the data, $(X_{k1}, Y_{k1}), \ldots, (X_{kn_k}, Y_{kn_k})$, $1 \le k \le d$. Let

$$T_N^* = \sum_{k=1}^{d} n_k \|\hat{U}_k^* - \hat{U}_0^*\|_\omega^2,$$

where $\hat{U}_k^*$ is as defined in (3.6) with $\hat{R}_\tau$ and $\hat{I}_\tau$ replaced by $\hat{R}_{\tau_1,\ldots,\tau_d}$ and $\hat{I}_{\tau_1,\ldots,\tau_d}$, respectively,

$$\hat{R}_{\tau_1,\ldots,\tau_d}(t) = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{R}_k(t), \quad \hat{I}_{\tau_1,\ldots,\tau_d}(t) = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{I}_k(t),$$

and

$$\hat{U}_0^* = \frac{1}{N}\sum_{k=1}^{d} n_k \hat{U}_k^*.$$

Theorem 7 *Suppose that assumptions in Theorem 5 hold, then*

$$\sup_x \left| P_* \left\{ T_N^* \le x \right\} - P\left\{ T_{\tau_1,\ldots,\tau_d} \le x \right\} \right| \xrightarrow{P} 0,$$

where $T_{\tau_1,\ldots,\tau_d} = \sum_{k=1}^d \|Z_{k,\tau_1,\ldots,\tau_k} - \sqrt{\tau_k} Z_{0,\tau_1,\ldots,\tau_k}\|_\omega^2$, $\{Z_{k,\tau_1,\ldots,\tau_k}(t),\, t \in \mathbb{R}\}$, $1 \le k \le d$, *are independent centered Gaussian processes on* $L_2(\omega)$ *with covariance kernel*

$$\varrho_{k,\tau_1,\ldots,\tau_k}(t,s) = E\{Z_{k,\tau_1,\ldots,\tau_k}(\varepsilon_k;t) Z_{k,\tau_1,\ldots,\tau_k}(\varepsilon_k;s)\},$$

$Z_{k,\tau_1,\ldots,\tau_k}(\varepsilon_k;t)$ *is defined as in* (3.5) *with* R_τ *and* I_τ *replaced by* $R_{\tau_1,\ldots,\tau_d}$ *and* $I_{\tau_1,\ldots,\tau_d}$, *respectively,*

$$R_{\tau_1,\ldots,\tau_d}(t) = \sum_{k=1}^{d} \tau_k R_r(t), \quad I_{\tau_1,\ldots,\tau_d}(t) = \sum_{k=1}^{d} \tau_k I_r(t),$$

and $Z_{0,\tau_1,\ldots,\tau_k} = \sum_{k=1}^d \sqrt{\tau_k} Z_{k,\tau_1,\ldots,\tau_k}$.

Similar results to those stated in Corollaries 1 and 2 for T_{n_1,n_2}^* can be given for T_N^*. To save space we omit them.

3.7 Appendix

3.7.1 An expression for M_3

Let

$$\varphi_\omega'(t) = \frac{\partial}{\partial t}\varphi_\omega(t), \quad \varphi_\omega''(t) = \frac{\partial^2}{\partial t^2}\varphi_\omega(t),$$

where φ_ω is as defined in (3.8). Let D_1, D_2 the matrices defined similarly to M with φ_ω replaced by φ'_ω and φ''_ω, respectively. Let

$$\begin{aligned}
e_1 &= (\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{2n_2})^T, \\
e_2 &= \left(\frac{\hat{\varepsilon}_{11}^2 - 1}{2}, \ldots, \frac{\hat{\varepsilon}_{1n_1}^2 - 1}{2}, \frac{\hat{\varepsilon}_{21}^2 - 1}{2}, \ldots, \frac{\hat{\varepsilon}_{2n_2}^2 - 1}{2}\right)^T, \\
d_1^T &= 1_N^T D_1/N, \\
d_2^T &= 1_N^T (e_1 1_N^T \odot D_1)/N, \\
c^T &= 1_N^T M/N, \\
a_1 &= 1_N^T D_2 1_N/N^2, \\
a_2 &= 1_N^T (e_1 1_N^t \odot D_2) 1_N/N^2, \\
a_3 &= -1_N^T (e_1 1_N^t \odot D_1) 1_N/N^2, \\
a_4 &= e_1^T D_2 e_1/N^2, \\
a_5 &= 1_N^T M 1_N/N^2.
\end{aligned}$$

With this notation,

$$\begin{aligned}
M_3 &= M - e_1 d_1^T - d_1 e_1^T - e_2 d_2^T - d_2 e_2^T - 1_N c^T - c 1_N^T - a_1 e_1 e_1^T \\
&\quad - a_2(e_1 e_2^T + e_2 e_1^T) - a_3(e_2 1_N^T + 1_N e_2^T) - a_4 e_2 e_2^T + a_5 1_N 1_N^T.
\end{aligned}$$

3.7.2 Proofs

We now sketch the proofs of the results stated in the previous sections, as well as some preliminary results. Observe that under Assumption (A.3) (see, for example, Masry, 1996)

$$\begin{aligned}
&\sup_{x \in S} |\hat{m}_k(x) - m_k(x)| = o_P(n_k^{-1/4}), \\
&\sup_{x \in S} |\hat{\sigma}_k(x) - \sigma_k(x)| = o_P(n_k^{-1/4}),
\end{aligned} \tag{3.11}$$

$k = 1, 2$. Let

$$\tilde{C}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \exp(\mathrm{i} t \varepsilon_{kj}) = \tilde{R}_k(t) + \mathrm{i} \tilde{I}_k(t),$$

$$\tilde{R}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \cos(t \varepsilon_{kj}), \quad \tilde{I}_k(t) = \frac{1}{n_k} \sum_{j=1}^{n_k} \sin(t \varepsilon_{kj}),$$

$k = 1, 2$.

Proof of Theorem 1 From Lemma 10(i) in Pardo-Fernández et al. (2015a) and (3.11),

$$T_{n_1,n_2} = \|\tilde{C}_1 - \tilde{C}_2\|_\omega^2 + o_P(1). \tag{3.12}$$

Theorem 2 in Alba–Fernández et al. (2008) asserts that

$$\|\tilde{C}_1 - \tilde{C}_2\|_\omega^2 \xrightarrow{a.s.} \kappa. \tag{3.13}$$

The result follows from (3.12) and (3.13). □

Lemma 1 *If Assumptions (A.2) and (A.3) hold, then*

$$\frac{n_1 n_2}{N} T_{n_1,n_2} = \|Z_{n_1,n_2}\|_\omega^2,$$

with

$$Z_{n_1,n_2}(t) = \sqrt{\frac{n_2}{N}} U_1(t) - \sqrt{\frac{n_1}{N}} U_2(t),$$

$$U_k(t) = U_{0k}(t) + t\varrho_{k,1}(t) + t^2 \varrho_{k,2}(t),$$

$$U_{0k}(t) = \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \Big\{ \cos(t\varepsilon_{kj}) + t\varepsilon_{kj} I_k(t) - t\frac{\varepsilon_{kj}^2-1}{2} R_k'(t) - R_\tau + \sin(t\varepsilon_{kj}) - t\varepsilon_{kj} R_k(t) - t\frac{\varepsilon_{kj}^2-1}{2} I_k'(t) - I_\tau \Big\},$$

$\sup_t |\varrho_{k,s}(t)| = o_P(N^{-1/2})$, *$k, s = 1, 2$.*

Proof We have that

$$\frac{n_1 n_2}{N} T_{n_1,n_2} = \|Z^0_{n_1,n_2}\|_\omega^2,$$

where $Z^0_{n_1,n_2}(t) = \sqrt{\frac{n_2}{N}} U^0_1(t) - \sqrt{\frac{n_1}{N}} U^0_2(t)$, $U^0_k(t) = \sqrt{n_k}\{\hat{C}_k(t) - C_\tau(t)\}$, $k = 1, 2$. From Lemma 10 in Pardo-Fernández et al. (2015b),

$$U^0_k(t) = \sqrt{n_k}\tilde{C}_k(t) + A_{k,1}(t) + A_{k,2}(t) - \sqrt{n_k} C_\tau(t) + t\rho_{k,1}(t) + t^2 \rho_{k,2}(t), \quad k = 1, 2,$$

where

$$A_{k,1}(t) = \mathrm{i}t \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \exp(\mathrm{i}t\varepsilon_{kj}) \left(\frac{\sigma_k(X_{kj}) - \hat{\sigma}_k(X_{kj})}{\sigma_k(X_{kj})} \right) \varepsilon_{kj},$$

$$A_{k,2}(t) = \mathrm{i}t \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \exp(\mathrm{i}t\varepsilon_{kj}) \left(\frac{m_k(X_{kj}) - \hat{m}_k(X_{kj})}{\sigma_k(X_{kj})} \right)$$

and $\sup_t |\rho_{k,s}(t)| = o_P(N^{-1/2})$, $k, s = 1, 2$. From Lemma 11 in Pardo-Fernández et al. (2015b),

$$A_{k,1}(t) = -\frac{t}{2} C_k'(t) \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} (\varepsilon_{kj}^2 - 1) + t\rho_{k,3}(t), \quad \sup_t |\rho_{k,3}(t)| = o_P(1),$$

with $C_k'(t) = R_k'(t) + \mathrm{i}I_k'(t)$. From the proof of Theorem 1 in Pardo-Fernández et al. (2015a),

$$A_{k,2}(t) = -\mathrm{i}tC_k(t)\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\varepsilon_{kj} + t\rho_{k,4}(t), \quad \sup_t|\rho_{k,4}(t)| = o_P(1).$$

All above facts and Assumption (A.2) imply that $\|Z^0_{n_1,n_2}\|^2_\omega = \|Z_{n_1,n_2}\|^2_\omega$. This completes the proof. □

Proof of Theorem 2 Let us continue with the notation in the statement of Lemma 1. By the central limit theorem for IID random elements in Hilbert spaces, $\{U_{01}(t),\, t \in \mathbb{R}\}$ converges to a centered Gaussian process $U^{(1)}$ on $L^2(w)$ with covariance structure $\varrho_0(t,s)$. By the independence of the two samples, $\{U_{02}(t),\, t \in \mathbb{R}\}$ converges in distribution to an independent copy $U^{(2)}$ of $U^{(1)}$. As, for constants a and b satisfying $a^2 + b^2 = 1$, the centered process $Z(t) = aU^{(1)}(t) + bU^{(2)}(t)$ has covariance structure $\varrho_0(t,s)$, and since $(\sqrt{1-n_1/N})^2 + (\sqrt{n_1/N})^2 = 1$ and n_1/N converges to τ, it follows that $\{Z_{n_1,n_2}(t),\, t \in \mathbb{R}\}$ converges in law to $\{Z(t),\, t \in \mathbb{R}\}$, under H_0. Finally, the result follows from the continuous mapping theorem. □

Proof of Theorem 3 Note that

$$\frac{n_1 n_2}{N}T^*_{1,n_1,n_2} = \|Z^*_{n_1,n_2}\|^2_\omega,$$

where

$$Z^*_{n_1,n_2} = \sqrt{\frac{n_2}{N}}U_1^*(t) - \sqrt{\frac{n_1}{N}}U_2^*(t),$$

where $U_k^*(t) = \frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\xi_{kj}Z_{k,\tau}(\varepsilon_{kj};t),\ k = 1,2.$

First, it will be shown that conditional on $(X_{11},Y_{11}),\ldots,(X_{1n_1},Y_{1n_1})$, $\{U_1^*(t),\, t \in \mathbb{R}\}$ converges in law to $\{U_{1\tau}(t),\, t \in \mathbb{R}\}$ on $L_2(\omega)$, where $\{U_{1\tau}(t),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance kernel $\varrho_{1,\tau}(t,s)$. To achieve this result we will apply Theorem 1.1 in Kundu et al. (2000). Next we will show that conditions (i)-(iii) in that theorem hold.

Note that $E_*\{\xi_{1j}Z_{1,\tau}(\varepsilon_{1j};t)\} = 0,\ 1 \le j \le n_1$. Denote

$$c_{n_1}(t,s) = Cov_*\{U_1^*(t), U_1^*(s)\} = \frac{1}{n_1}\sum_{j=1}^{n_1}Z_{1,\tau}(\varepsilon_{1j};t)Z_{1,\tau}(\varepsilon_{1j};s).$$

From the strong law of large numbers,

$$c_{n_1}(t,s) \xrightarrow{a.s.} \varrho_{1,\tau}(t,s),\ \forall s,t \in \mathbb{R}. \tag{3.14}$$

Note also that

$$|c_{n_1}(t,s)| \leq g(t,s),\ \forall s,t \in \mathbb{R}, \tag{3.15}$$

with

$$g(t,s) = \varpi_1 + \varpi_2(|t|+|s|) + \varpi_3|t||s|, \quad a.s.$$

for certain positive constants ϖ_1, ϖ_2, ϖ_3.

Let $\{e_k, k \geq 0\}$ be an orthonormal basis of $L_2(\omega)$. Let V_1 denote the covariance operator of U_1^* and let V_τ denote the covariance operator of $U_{1\tau}(t)$. From (3.14) and (3.15), by the dominated convergence theorem,

$$\begin{aligned} \lim\langle V_1 e_k, e_l\rangle_\omega &= \lim \textstyle\int c_{n_1}(s,t)e_k(t)e_l(s)\omega(t)\omega(s)dtds \\ &= \textstyle\int \varrho_{1,\tau}(t,s)e_k(t)e_l(s)\omega(t)\omega(s)dtds = \langle V_\tau e_k, e_l\rangle_\omega \quad a.s. \end{aligned}$$

Thus taking $a_{kl} = \langle V_\tau e_k, e_l\rangle_\omega$, the condition (i) in the aforementioned Theorem 1.1 holds. To check the condition (ii), by monotone convergence theorem, Parseval's relation and dominated convergence theorem, it follows

$$\begin{aligned} \lim \textstyle\sum_{k=0}^\infty \langle V_1 e_k, e_l\rangle_\omega &= \lim \textstyle\sum_{k=0}^\infty E_*\{\langle U_1^*(t), e_k\rangle_\omega^2\} = \lim E_*\{\|U_1^*(t)\|_\omega^2\} \\ &= \textstyle\int \lim c_{n_1}(t,t)\omega(t)dt = \int \varrho_{1,\tau}(t,t)\omega(t)dt \\ &= E_*\|U_{1\tau}\|_\omega^2 < \infty \quad a.s. \end{aligned}$$

Before verifying condition (iii), we first notice that

$$\left|\langle \frac{1}{\sqrt{n_1}}\xi_{1j}Z_{1,\tau}(\varepsilon_{1j};t), e_k\rangle_\omega\right| \leq \frac{|\xi_{1j}|}{\sqrt{n_1}}\left(\int Z_{1,\tau}^2(\varepsilon_{1j};t)\omega(t)dt\right)^{1/2} \leq g(t,t)\frac{|\xi_{1j}|}{\sqrt{n_1}}.$$

From the above inequality,

$$\begin{aligned} &\sum_{j=1}^{n_1} E_*\left[\langle \frac{1}{\sqrt{n_1}}\xi_{1j}Z_{1,\tau}(\varepsilon_{1j};t), e_k\rangle_\omega^2 I\left\{\left|\langle \frac{1}{\sqrt{n_1}}\xi_{1j}Z_{1,\tau}(\varepsilon_{1j};t), e_k\rangle_\omega\right| > \epsilon\right\}\right] \\ &\quad \leq \tfrac{g^2(t,t)}{n_1}\textstyle\sum_{j=1}^{n_1} E_*[\xi_{1j}^2 I\{|\xi_{1j}| > g^{-1}(t,t)\epsilon\sqrt{n_1}\}] \\ &\quad = g^2(t,t)E_*[\xi_{1j}^2 I\{|\xi_{1j}| > g^{-1}(t,t)\epsilon\sqrt{n_1}\}] \to 0, \end{aligned}$$

$\forall \epsilon > 0$, $\forall k \geq 0$.

Analogously, conditional on $(X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$, $\{U_2^*(t),\, t \in \mathbb{R}\}$ converges in law to $\{U_{2\tau(t)},\, t \in \mathbb{R}\}$ on $L_2(\omega)$, where $\{U_{2\tau}(t),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance kernel $\varrho_{2,\tau}(t,s)$. A similar argument to that in the proof of Theorem 2 shows that, conditional on $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1}), (X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$, $\{Z_{1,n_1,n_2}^*(t),\, t \in \mathbb{R}\}$ converges in law to $\{Z_\tau(t),\, t \in \mathbb{R}\}$ on $L_2(\omega)$. □

Lemma 2 *Suppose that assumption (A.3) holds, then*

(a) $\frac{1}{n_k}\sum_{j=1}^{n_k}(\varepsilon_{kj}-\hat{\varepsilon}_{kj})^2 = o_P(1),\ k=1,2.$

(b) $\frac{1}{n_k}\sum_{j=1}^{n_k}(\hat{\varepsilon}_{kj}^2-\varepsilon_{kj}^2)^2 = o_P(1),\ k=1,2.$

(c) $\frac{1}{n_k}\sum_{j=1}^{n_k}(\hat{\varepsilon}_{kj}^2-1)^2 = O_P(1),\ k=1,2.$

(d) $\frac{1}{n_k}\sum_{j=1}^{n_k}\hat{\varepsilon}_{kj}^2 = O_P(1),\ k=1,2.$

Proof The difference between the residuals and the errors can be written as follows

$$\hat{\varepsilon}_{kj}-\varepsilon_{kj} = \varepsilon_{kj}\left(\frac{\sigma_k(X_{kj})-\hat{\sigma}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}\right)+\left(\frac{m_k(X_{kj})-\hat{m}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}\right), \tag{3.16}$$

$k=1,2$. The results in (a)–(d) follow from (3.11) and (3.16). □

Lemma 3 *Suppose that Assumptions (A.1)–(A.3) hold, then*

(a) $\|t(\hat{R}_k-R_k)\|_\omega^2 = o_P(1)$, $\|t(\hat{I}_k-I_k)\|_\omega^2 = o_P(1), \quad k=1,2.$

(b) $\|R_\tau-\hat{R}_\tau\|_\omega^2 = o_P(1), \quad \|I_\tau-\hat{I}_\tau\|_\omega^2 = o_P(1),$

(c) $\|t(R_k'-\hat{R}_k')\|_\omega^2 = o_P(1)$, $\|t(I_k'-\hat{I}_k')\|_\omega^2 = o_P(1). \quad k=1,2.$

Proof (a) By the mean value theorem,

$$t\{\hat{R}_k(t)-\tilde{R}_k(t)\} = -t^2 r_k(t).$$

From Lemma 2(a) and the Cauchy–Schwarz inequality,

$$\sup_t |r_k(t)| \le \frac{1}{n_k}\sum_{j=1}^{n_k}|\hat{\varepsilon}_{kj}-\varepsilon_{kj}| \le \left(\frac{1}{n_k}\sum_{j=1}^{n_k}(\hat{\varepsilon}_{kj}-\varepsilon_{kj})^2\right)^{1/2} = o_P(1).$$

Therefore,

$$\|t(\hat{R}_k-\tilde{R}_k)\|_\omega^2 \le \sup_t |r_k(t)|^2 \int t^4\omega(t)dt = o_P(1). \tag{3.17}$$

We also have that

$$\|t(R_k-\tilde{R}_k)\|_\omega^2 = o_P(1). \tag{3.18}$$

Finally, (3.17) and (3.18) both imply that $\|t(\hat{R}_k-R_k)\|_\omega^2 = o_P(1)$. The proof for $\|t(\hat{I}_k - I_k)\|_\omega^2$ is parallel. The proof of parts (b) and (c) follow similar steps. □

Proof of Theorem 4 $\frac{n_1n_2}{N}T^*_{2,n_1,n_2}$ can be expressed as $\frac{n_1n_2}{N}T^*_{2,n_1,n_2} = D_1+D_2+2D_3$, where $D_3^2 \le D_1D_2$, $D_1 = \frac{n_1n_2}{N}T^*_{1,n_1,n_2}$, $D_2 = \frac{n_1n_2}{N}\|(\hat{U}_1^*-\hat{U}_2^*)-(C_1^*-C_2^*)\|_\omega^2$. From Theorem 3,

$$\sup_x |P_*(D_1\le x)-P(T_\tau\le x)| \xrightarrow{P} 0,$$

Thus, to show the result it suffices to see that $D_2 = o_{p_*}(1)$ in probability. With this aim, observe that D_2 can be expressed as

$$D_2 = \sum_{k=1,2} \sum_{j=1}^{8} S_{k,j} + \sum_{k,s=1,2} \sum_{1 \le l \ne j \le 8} S_{k,s,j,l},$$

with $S^2_{k,s,j,l} \le S_{k,j} S_{s,l}$, $1 \le j, l \le 8$, $k, s = 1, 2$,

$$\begin{aligned}
S_{k,1} &= \tfrac{N-n_k}{N} \| \tfrac{1}{\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} \{\cos(t\hat{\varepsilon}_{kj}) - \cos(t\varepsilon_{kj})\}\xi_{k,j} \|^2_\omega, \\
S_{k,2} &= \tfrac{N-n_k}{N} \| \tfrac{1}{\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} \{\sin(t\hat{\varepsilon}_{kj}) - \sin(t\varepsilon_{kj})\}\xi_{k,j} \|^2_\omega, \\
S_{k,3} &= \tfrac{N-n_k}{N} \| \tfrac{t}{\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj}\hat{R}_k - \varepsilon_{kj} R_k)\xi_{k,j} \|^2_\omega, \\
S_{k,4} &= \tfrac{N-n_k}{N} \| \tfrac{t}{\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} (\hat{\varepsilon}_{kj}\hat{I}_k - \varepsilon_{kj} I_k)\xi_{k,j} \|^2_\omega, \\
S_{k,5} &= \tfrac{N-n_k}{N} \| \tfrac{1}{\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} (R_\tau - \hat{R}_\tau)\xi_{1,k} \|^2_\omega, \\
S_{k,6} &= \tfrac{N-n_k}{N} \| \tfrac{1}{\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} (I_\tau - \hat{I}_\tau)\xi_{k,j} \|^2_\omega, \\
S_{k,7} &= \tfrac{N-n_k}{N} \| \tfrac{t}{2\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} \{(\hat{\varepsilon}^2_{kj} - 1)\hat{I}'_k - (\varepsilon^2_{kj} - 1)I'_k\}\xi_{k,j} \|^2_\omega, \\
S_{k,8} &= \tfrac{N-n_k}{N} \| \tfrac{t}{2\sqrt{n_k}} \textstyle\sum_{j=1}^{n_k} \{(\varepsilon^2_{kj} - 1)R'_k - (\hat{\varepsilon}^2_{kj} - 1)\hat{R}'_k\}\xi_{k,j} \|^2_\omega,
\end{aligned}$$

$k = 1, 2$. We will show that $S_{k,j} = o_{p_*}(1)$ in probability, $1 \le j \le 8$, $k = 1, 2$.

By the mean value theorem,

$$S_{1,1} = \frac{n_2}{N} \frac{1}{n_1} \sum_{j,m=1}^{n_1} \xi_{1j}\xi_{1m}(\hat{\varepsilon}_{1j} - \varepsilon_{1j})(\hat{\varepsilon}_{1m} - \varepsilon_{1m}) \int t^2 \sin(t\tilde{\varepsilon}_{1j}) \sin(\tilde{\varepsilon}_{1m})\omega(t)dt,$$

where $\tilde{\varepsilon}_{1j} = \alpha_{1j}\varepsilon_{1j} + (1 - \alpha_{1j})\hat{\varepsilon}_{1j}$, for some $\alpha_{1j} \in (0, 1)$. Then, from Lemma 2(a),

$$E_*(S_{1,1}) \le \frac{n_2}{N} \frac{1}{n_1} \sum_{j=1}^{n_1} (\varepsilon_{1j} - \hat{\varepsilon}_{1j})^2 \int t^2 \omega(t)dt = o_p(1),$$

which implies $S_{1,1} = o_{p_*}(1)$ in probability. Analogously, $S_{2,1} = o_{p_*}(1)$, $S_{1,2} = o_{p_*}(1)$, $S_{2,2} = o_{p_*}(1)$ in probability.

Observe that $S_{1,3} = S_{13} + S_{23} + 2S_{33}$, with $S^2_{33} \le S_{13}S_{23}$,

$$S_{13} = \frac{n_2}{N} \frac{1}{n_1} \sum_{j,m=1}^{n_1} (\hat{\varepsilon}_{1j} - \varepsilon_{1j})(\hat{\varepsilon}_{1m} - \varepsilon_{1m})\xi_{1j}\xi_{1m} \|tR_1\|^2_\omega,$$

$$S_{23} = \frac{n_2}{N} \frac{1}{n_1} \sum_{j,m=1}^{n_1} \hat{\varepsilon}_{1j}\hat{\varepsilon}_{1m}\xi_{1j}\xi_{1m} \|t(\hat{R}_1 - R_1)\|^2_\omega.$$

From Lemma 2(a) it follows that $E_*(S_{13}) = o_P(1)$ and thus $S_{13} = o_{p_*}(1)$ in probability. From Lemma 2(d) and Lemma 3(a) it follows that $E_*(S_{23}) = o_P(1)$ and thus $S_{23} = o_{p_*}(1)$ in probability. Therefore, $S_{1,3} = o_{p_*}(1)$ in probability. Analogously, $S_{2,3} = o_{p_*}(1)$, $S_{1,4} = o_{p_*}(1)$ and $S_{2,4} = o_{p_*}(1)$ in probability.

Since $S_{1,5} = \frac{n_2}{N}\left(\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\xi_{kj}\right)^2\|R_\tau - \hat{R}_\tau\|_\omega^2$, the central limit theorem and Lemma 3(b) imply that $E_*(S_{1,5}) = o_P(1)$ and thus $S_{1,5} = o_{p_*}(1)$ in probability. Analogously, $S_{2,5} = o_{p_*}(1)$, $S_{1,6} = o_{p_*}(1)$ and $S_{1,6} = o_{p_*}(1)$ in probability.

Observe that $S_{1,7} = S_{17} + S_{27} + 2S_{37}$, with $S_{37}^2 \leq S_{17}S_{27}$,

$$\begin{aligned}
S_{17} &= \tfrac{n_2}{N}\tfrac{1}{4}\tfrac{1}{n_1}\textstyle\sum_{j,m=1}^{n_1}(\hat{\varepsilon}_{1j}^2 - 1)(\hat{\varepsilon}_{1m}^2 - 1)\xi_{1j}\xi_{1m}\|t(\hat{I}_1' - I_1')\|_\omega^2,\\
S_{27} &= \tfrac{n_2}{N}\tfrac{1}{4}\tfrac{1}{n_1}\textstyle\sum_{j,m=1}^{n_1}(\hat{\varepsilon}_{1j}^2 - \varepsilon_{1j}^2)(\hat{\varepsilon}_{1m}^2 - \varepsilon_{1m}^2)\xi_{1j}\xi_{1m}\|tI_k'\|_\omega^2.
\end{aligned}$$

From Lemma 2(c) and Lemma 3(c), it follows that $E_*(S_{17}) = o_P(1)$ and thus $S_{17} = o_{p_*}(1)$ in probability. From Lemma 2(b) it follows that $E_*(S_{27}) = o_P(1)$ and thus $S_{27} = o_{p_*}(1)$, in probability. Therefore, $S_{1,7} = o_{p_*}(1)$, in probability. Analogously, $S_{2,7} = o_{p_*}(1)$, $S_{1,8} = o_{p_*}(1)$, and $S_{2,8} = o_{p_*}(1)$ in probability. This completes the proof. □

3.8 References

1. Alba-Fernández V., Jiménez-Gamero M. D., Muñoz-García J., 2008. A test for the two-sample problem based on empirical characteristic functions. Comput Statist Data Anal 52, 3730–3748.

2. Baringhaus L., Kolbe D., 2015. Two-sample tests based on empirical Hankel transforms. Stat Papers 56, 597-617.

3. Bowman A. W., Azzalini A., 2014. R package 'sm': nonparametric smoothing methods (version 2.2-5.4)

   URL http://www.stats.gla.ac.uk/~adrian/sm, http://azzalini.stat.unipd.it/Book_sm

4. Burke M. D., 2000. Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Statist Probab Lett 46, 13–20.

5. Cleveland W. S., 1993. Visualizing Data. Summit, New Jersey.

6. Delhing H., Mikosch T., 1994. Random quadratic forms and the bootstrap for U-statistics. J Multivariate Anal 51, 392–413.

7. de Uña-Álvarez J., 2013. Comments on: An updated review of Goodness-of-Fit tests for regression models. Test 22, 414–418.

8. Duchesne P., Lafaye de Micheaux P., 2010. Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. Comput Statist Data Anal 54, 858–862.

9. Feller W., 1971. An Introduction to Probability Theory and its Applications. Vol 2. Wiley. New York.

10. Ghoudi K., Rémillard B., 2014. Comparison of specification tests for GARCH models. Comput Statist Data Anal 76, 291–300.

11. González-Manteiga W., Crujeiras R., 2013. An updated review of Goodness-of-Fit tests for regression models. Test 22, 361–411.

12. Hall P., Hart J. D., 1990. Bootstrap test for difference between means in nonparametric regression. J Amer Statist Assoc 85, 1039–1049.

13. Hobza T., Morales D., Pardo L., 2014. Divergence-based test of homogeneity for spatial data. Stat Papers 55, 1059–1077.

14. Hušková M., Meintanis S. G., 2008. Tests for the multivariate k-sample problem based on the empirical characteristic function. J Nonparametr Stat 20, 263–277.

15. Imhof J. P., 1961. Computing the Distribution of Quadratic Forms in Normal Variables. Biometrika 48, 419–426.

16. Jiménez-Gamero M. D., Alba-Fernández M. V., Jodrá P., Barranco-Chamorro I., 2016. Fast tests for the two-sample problem based on the empirical characteristic function. Math Comp Simulation, http://dx.doi.org/10.1016/j.matcom.2016.09.007.

17. Jiménez-Gamero M. D., Kim H-M., 2015. Fast goodness-of-fit test based on the characteristic function. Comput Statist Data Anal 89, 172–191.

18. Kojadinovic I., Yan J., 2012. Goodness-of-fit testing based on a weighted bootstrap: A fast sample alternative to the parametric bootstrap. Can J Statist 40, 480–500.

19. Kulasekera K. B., Wang J., 2001. A test of the equality of regression curves using Gâteaux scores. Austral N Z J Statist 43, 89–99.

20. Kundu S., Majumdar S., Mukherjee K., 2000. Central Limit Theorems revisited. Statist Probab Lett 47, 265–275.

21. Masry E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. J Time Series Anal 17, 571–600.

22. Meintanis S. G., 2005. Permutation tests for homogeneity based on the empirical characteristic function. J Nonparametr Stat 17, 583–592.

23. Meintanis S. G., Ngatchou-Wandji J., Taufer E., 2015. Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function. J Multivariate Anal 140, 171–192.

24. Modarres R., 2016. Multinomial interpoint distances. Stat Papers DOI 10.1007/s00362-016-0766-7

25. Mora J., 2005. Comparing distribution functions of errors in linear models: A nonparametric approach. Statist Probab Lett 73, 425–432.

26. Pardo-Fernández J. C., 2007. Comparison of error distributions in nonparametric regression. Statist Probab Lett 77, 350–356.

27. Pardo-Fernández J. C., Jiménez-Gamero M. D., El Ghouch A., 2015a. A nonparametric ANOVA-type test for regression curves based on characteristic functions. Scand J Stat 42, 197–213.

28. Pardo-Fernández J. C., Jiménez-Gamero M. D., El Ghouch A., 2015b. Tests for the equality of conditional variance functions in nonparametric regression. Electron J Stat 9, 1826–1851.

29. Quessy J. F., Éthier F., 2012. Cramér-von Mises and characteristic function tests for the two and k-sample problems with dependent data. Comput Stat Data Anal 56, 2097–2111.

30. R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna-Austria.URL http://www.R-project.org/

31. Sperlich S., 2013. Comments on: An updated review of Goodness-of-Fit tests for regression models. Test 22,419–427.

32. Young S. G., Bowman A. W., 1995. Non-parametric Analysis of Covariance. Biometrics 51, 920–931.

Capítulo 4

A weighted bootstrap approximation for comparing the error distributions

Abstract

Several procedures have been proposed for testing the equality of error distributions in two or more nonparametric regression models. Here we deal with methods based on comparing estimators of the cumulative distribution function of the errors in each population to an estimator of the common cumulative distribution function under the null hypothesis. The null distribution of the associated test statistics has been approximated by means of a smooth bootstrap estimator. In this chapter is proposed to approximate their null distribution through a weighted bootstrap. It is shown that it produces a consistent estimator. The finite sample performance of this approximation is assessed by means of a simulation study, where it is also compared to the smooth bootstrap. This study reveals that, from a computational point of view, the proposed approximation is more efficient than the one provided by the smooth bootstrap.

4.1 Introduction

Let (X_k, Y_k), $1 \leq k \leq d$, be d independent random vectors and assume that they satisfy the following nonparametric regression models,

$$Y_k = m_k(X_k) + \sigma_k(X_k)\varepsilon_k, \tag{4.1}$$

where the error variable ε_k, with cumulative distribution function (CDF) and probability density function (PDF) F_k and f_k, respectively, is independent of X_k, $m_k(x) = E(Y_k|X_k = x)$ is the regression function and $\sigma_k^2(x) = Var(Y_k|X_k = x)$ is the conditional variance function. By construction $E(\varepsilon_k) = 0$ and $Var(\varepsilon_k) = 1$. Along the chapter it will be assumed that all the covariates $X_1, \ldots, X_d$ have the same support, S. The equality of the error distributions is a usual assumption in certain statistical problems such as that of testing for the equality of regression curves (see, for example, Young and Bowman 1995; Hall and Hart 1990 and Kulasequera and Wang 2001). Likewise, Mora (2005) has provided some examples in economics and in applied medical studies where testing for the equality of the error distributions play an important role. Because of the above reasons, several authors have suggested tests for such hypothesis, that is, tests of the null hypothesis

$$H_0 : F_1 = F_2 = \ldots = F_d,$$

versus the general alternative

$$H_1 : F_s \neq F_t, \text{ for some } 1 \leq s, t \leq d.$$

Mora (2005) has proposed tests for testing the equality of $d = 2$ the error distributions by assuming parametric specifications for the regression function and homoscedastic models. Along this chapter the regression functions, the variance functions, the error distributions and that of the covariates are unknown and no parametric models are assumed for them. In this setting Pardo-Fernández (2007) -PF07 from now on- has proposed two tests for H_0 which are based con comparing estimators of the CDFs of the error in each population to an estimator of the common CDF under the null hypothesis. Since the asymptotic null distribution of these test statistics depends on some unknown quantities, this author has suggested to use a smooth bootstrap (SB) to approximate the critical values. Although very easy to implement, the SB can become very computationally expensive as the sample sizes of the data increase.

In the previous chapter is propose a test for H_0 which are based on comparing estimators of the characteristic function of the error distribution in each population to an estimator of the common characteristic function under the null hypothesis. As in PF07, the asymptotic null distribution of the test statistic cannot be used to estimate its null distribution because it is unknown. In contrast to PF07, in Chapter 3, we have proposed to approximate the null distribution by means of a weighted bootstrap (WB) estimator, in the sense of Burke (2000). Also numerically comparison between the WB and the SB approximations for the test statistic studied in Chapter 3, shown that the WB approximation is computationally more efficient than that approximation based on the SB.

For the problem considered in this work, the variables of interest -the errors- are not observable. For observable variables the papers by Quessy and Éthier (2012), Jiménez-Gamero et al. (2016) and Alba-Fernández et al. (2017) deal with tests for H_0 and propose to approximate the null distribution of the considered test statistics by means of a WB estimator. In view of the good properties of the WB approximation in the previous chapters and in these and other papers, it is also expected to work satisfactorily for estimating the null distribution of the test statistics proposed in PF07. The objective of this chapter is to study, both theoretically and empirically, the use of the WB for approximating the null distribution of them.

The chapter is organized as follows. The tests in PF07 are functions of certain empirical process. Section 4.2 describes the test statistics, the empirical process they are based on and explains some problems with the asymptotic null distribution of such process. Section 4.3 proposes and studies a WB approximation to the previously mentioned process. It is shown that the proposed approximation yield a consistent estimator when the null hypothesis is true. It is also shown that the resulting tests are consistent, in the sense of being able to detect any alternative. Some practical issues are addressed in Section 4.4. Section 4.5 reports the results of some simulation experiments designed to study the finite sample performance of the proposed approximation and to compare it to the smooth bootstrap approximation and a real data set application. All technical details such as the require assumptions, some preliminary results needed for the proofs of the stated results as well as the proofs are deferred to the last section.

The following notation will be used along the chapter: all vectors are column vectors; $\sup_y$ denotes $\sup_{y\in\mathbb{R}}$; the superscript ^{T} denotes transpose; $I(\cdot)$ denotes the indicator function; P_0 and E_0 denote probability and expectation, respectively, by assuming that the null hypothesis is true; P_* denotes the conditional probability law, given the data; all limits in this chapter are taken when $n \to \infty$, where n stands for the total number of observations; for any real function $f(t)$ differentiable at $t \in \mathbb{R}$ the following notation will be used, $f'(t) = \frac{\partial}{\partial t} f(t)$,

4.2 Test statistics

Let (X_{kj}, Y_{kj}), $1, \leq j \leq n_k$, be an independent and identically distributed (IID) sample from (X_k, Y_k), $1 \leq k \leq d$, and denote $n = \sum_{k=1}^{d} n_k$. Along this chapter it will be assumed that $n_k/n \to p_k > 0$, $1 \leq k \leq d$. The errors associated to the available data are not observable since m_k and σ_k are unknown. Because of this reason the inference on the error distributions must be based on the residuals, which are obtained by replacing m_k and σ_k by adequate estimators, say $\hat{m}_k$ and $\hat{\sigma}_k$, respectively, that is,

$$\hat{\varepsilon}_{kj} = \frac{Y_{kj} - \hat{m}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}, \quad 1 \leq j \leq n_k, \quad 1 \leq k \leq d. \tag{4.2}$$

As in PF07, the following kernel estimators for the regression function m_k and the variance function σ_k^2 will be used,

$$\hat{m}_k(x) = \sum_{j=1}^{n_k} W_{kj}(x; h_n) Y_{kj}, \quad x \in S,$$

$$\hat{\sigma}_k^2(x) = \sum_{j=1}^{n_k} W_{kj}(x; h_n) \left\{ Y_{kj} - \hat{m}_k(x) \right\}^2, \quad x \in S,$$

where

$$W_{kj}(x; h_n) = \frac{K_{h_n}(X_{kj} - x)}{\sum_{s=1}^{n_k} K_{h_n}(X_{ks} - x)}, \quad x \in S,$$

$K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$, $K(\cdot)$ is a kernel and h_n is the bandwidth.

To test for H_0, FP07 has proposed to compare estimators of the CDFs of the error in each population to an estimator of the common CDF under the null hypothesis. With this aim, he considers functions of the process

$$\hat{U}(y) = (\hat{U}_1(y), \ldots, \hat{U}_d(y))^T,$$

where

$$\hat{U}_k(y) = n_k^{1/2} \left\{ \hat{F}(y) - \hat{F}_k(y) \right\}, \quad 1 \leq k \leq d,$$

with

$$\hat{F}_k(y) = \frac{1}{n_k} \sum_{j=1}^{n_k} I(\hat{\varepsilon}_{kj} \leq y),$$

an estimator of the CDF of ε_k, F_k, $1 \leq k \leq d$, (see Akritas and Van Keilegom, 2001) and

$$\hat{F}(y) = \frac{1}{n} \sum_{k=1}^{d} \sum_{j=1}^{n_k} I(\hat{\varepsilon}_{kj} \leq y).$$

an estimator of the common CDF under H_0. Specifically, PF07 considers a Kolgomorov-Smirnov type statistic,

$$S_{KS} = \sum_{k=1}^{d} \sup_y |\hat{U}_k(y)|,$$

and a Cramér-von Mises type statistic,

$$S_{CM} = \sum_{k=1}^{d} \int \hat{U}_k^2(y) d\hat{F}(y),$$

and proposed to reject the null hypothesis for large values of these test statistics. Although quite reasonable, no theoretical justification was supplied for these critical regions. In order give a sound justification of them, next we derive the limit of S_{KS} and S_{CM}. With this aim, let

$$F(y) = \sum_{k=1}^{d} p_k F_k(y).$$

Notice that if H_0 is true, then $F_1 = F_2 = \ldots = F_d = F$.

Proposition 1 *If Assumptions (A1)-(A4) hold, then*

(a) $\frac{1}{\sqrt{n}} S_{KS} \xrightarrow{P} \kappa_{KS} = \sum_{k=1}^{d} \sup_y \sqrt{p_k} |F(y) - F_k(y)|$.

(b) $\frac{1}{n} S_{CM} \xrightarrow{P} \kappa_{CM} = \sum_{k=1}^{d} p_k \int \{F(y) - F_k(y)\}^2 dF(y)$.

Note that $\kappa_{KS} \geq 0$ ($\kappa_{CM} \geq 0$), with $\kappa_{KS} = 0$ ($\kappa_{CM} = 0$) if and only if the null hypothesis is true. Thus, a reasonable test for H_0 should reject the null hypothesis for large values of S_{KS} (or S_{CM}). Now, to determine what are large values one must calculate its null distribution, or at least an approximation to it. The null distribution of these test statistics is clearly unknown. As at first approximation, PF07 derived their asymptotic null distribution. Specifically, Theorem 1 in PF07 states that if (A1)-(A4) hold and H_0 is true, then the d-dimensional process $\hat{U}(y)$ converges weakly to a centered k-dimensional Gaussian process $U_0(y) = (U_{0,1}(y), \ldots, U_{0,d}(y))^T$ with covariance structure given by

$$Cov_0 \left(U_s(y), U_t(z) \right) = p_s^{1/2} p_t^{1/2} E_0 \left\{ \varphi_0(\varepsilon, y) \varphi_0(\varepsilon, z) \right\} \sum_{k=1}^{d} \left\{ 1 - \frac{I(k = s)}{p_s} \right\} \left\{ 1 - \frac{I(k = t)}{p_t} \right\},$$

$1 \leq s, t \leq d$, where

$$\varphi_0(\varepsilon, y) = I(\varepsilon \leq y) - F_0(y) + f_0(y)\varepsilon + \frac{1}{2} y f_0(y)(\varepsilon^2 - 1), \tag{4.3}$$

with f_0 and F_0 denoting the common probability density function (PDF) and CDF, respectively, of the errors under the null hypothesis. Since the statistics S_{KS} and S_{CM} are functions of $\hat{U}(y)$, their asymptotic null distributions can be derived from the asymptotic null distribution of $\hat{U}(y)$ (see Corollary 2 in PF07). The distribution of $U_0(y)$ is unknown because it depends on the common PDF and CDF of the errors under the null. Because of this reason, PF07 suggested to use a SB to approximate the critical values. Next section studies another way of approximating it by means of a WB estimator.

4.3 Weighted Bootstrap approximation

From Theorem 1 in Akritas and Van Keilegom (2001),

$$\hat{F}_k(y) - F_k(y) = \frac{1}{n_k}\sum_{j=1}^{n_k} \varphi_k(\varepsilon_{kj}, y) + o_P(n_k^{-1/2}),$$

uniformly in y, with

$$\varphi_k(\varepsilon, y) = I(\varepsilon \le y) - F_k(y) + f_k(y)\varepsilon + \frac{1}{2} y f_k(y)(\varepsilon^2 - 1),$$

$1 \le k \le d$, and therefore

$$\hat{F}(y) - \hat{F}_k(y) = \sum_{l=1}^{d} \left\{ \frac{n_l}{n} - I(l = k) \right\} \frac{1}{n_l} \sum_{s=1}^{n_l} \varphi_l(\varepsilon_{ls}, y) + o_P(n^{-1/2}), \tag{4.4}$$

uniformly in y. Under the null hypothesis $\varphi_k(\varepsilon, y) = \varphi_0(\varepsilon, y)$, $1 \le k \le d$, with φ_0 as defined in (4.3). Because of this reason, to define a WB version of $\hat{U}_k(y)$ yielding a consistent null distribution estimator we consider

$$\varphi(\varepsilon, y) = I(\varepsilon \le y) - F(y) + f(y)\varepsilon + \frac{1}{2} y f(y)(\varepsilon^2 - 1), \tag{4.5}$$

with

$$f(y) = \sum_{k=1}^{d} p_k f_k(y).$$

Under H_0, $F = F_0$ and $f = f_0$, and thus, $\varphi = \varphi_0$.

Let ξ_{kj}, $1 \le j \le n_k$, $1 \le k \le d$, be IID random variates with mean 0, variance 1 and $\|\xi_{kj}\|_{2,1} < \infty$, where $\|\xi_{kj}\|_{2,1} = \int_0^\infty \sqrt{P(|\xi_{kj}| > x)}dx$, which are independent of $(X_{11}, Y_{11}), \dots, (X_{dn_d}, Y_{dn_d})$. Based on (4.4), we define the following WB version of $\hat{U}_k(y)$,

$$U_k^*(y) = \sqrt{n_k} \sum_{l=1}^{d} \left\{ \frac{n_l}{n} - I(l = k) \right\} \frac{1}{n_l} \sum_{s=1}^{n_l} \varphi(\varepsilon_{ls}, y)\xi_{ls}, \quad 1 \le k \le d,$$

The next result gives the weak limit of the conditional distribution of $U^*(y) = (U_1^*(y), \ldots, U_d^*(y))^T$ given the data, $(X_{11}, Y_{11}), \ldots, (X_{dn_d}, Y_{dn_d})$.

Theorem 1 *Suppose that Assumptions (A1)-(A4) hold. Then, the conditional distribution of the d-dimensional process $U^*(y) = (U_1^*(y), \ldots, U_d^*(y))^T$, given the data, converges to a centered d-dimensional Gaussian process $U(y) = (U_1(y), \ldots, U_d(y))^T$ with covariance structure given by*

$$Cov\left(U_s(y), U_t(z)\right) = p_s^{1/2} p_t^{1/2} \sum_{l=1}^{d} \left\{1 - \frac{I(l=s)}{p_s}\right\} \left\{1 - \frac{I(l=t)}{p_t}\right\} E\left\{\varphi(\varepsilon_l, y)\varphi(\varepsilon_l, z)\right\}, \tag{4.6}$$

$1 \leq s, t \leq d$, *with* $\varphi(\varepsilon, y)$ *as defined in* (4.5).

The result in Theorem 1 is valid whether the null hypothesis is true or not. If the null hypothesis is true, then, as observed before, $\varphi = \varphi_0$ and therefore the conditional distribution of $U^*(y)$, given the data, provides a consistent estimation of $U_0(y)$. Nevertheless, from a practical point of view, this result is useless because the expression of $\varphi(\varepsilon_{kj}, y)$ depends on some unknowns. To overcome this difficulty, we replace the unknown quantities by consistent estimators. Specifically, we replace the errors ε_{kj} by the residuals (4.2), $F(y)$ by $\hat{F}(y)$ and $f(y)$ by $\hat{f}(y)$, defined as

$$\hat{f}(y) = \sum_{k=1}^{d} \frac{n_k}{n} \hat{f}_k(y),$$

with

$$\hat{f}_k(y) = \frac{1}{n_k a_n} \sum_{j=1}^{n_k} L\left(\frac{y - \hat{\varepsilon}_{kj}}{a_n}\right),$$

where L is a kernel and a_n an appropriate bandwidth. Therefore, instead of $U^*(y)$ we consider $\hat{U}^*(y) = (\hat{U}_1^*(y), \ldots, \hat{U}_d^*(y))^T$, with

$$\hat{U}_k^*(y) = \sqrt{n_k} \sum_{l=1}^{d} \left\{\frac{n_l}{n} - I(l=k)\right\} \frac{1}{n_l} \sum_{s=1}^{n_l} \hat{\varphi}(\hat{\varepsilon}_{ls}, y)\xi_{ls}, \quad 1 \leq k \leq d,$$

where

$$\hat{\varphi}(\varepsilon, y) = I(\varepsilon \leq y) - \hat{F}(y) + \hat{f}(y)\varepsilon + \frac{1}{2} y \hat{f}(y)(\varepsilon^2 - 1) \tag{4.7}$$

The next theorem states that replacing ε_{ls} by $\hat{\varepsilon}_{ls}$, $F(y)$ by $\hat{F}(y)$ and $f(y)$ by $\hat{f}(y)$ has no asymptotic effect, in the sense that both of U^* and $\hat{U}^*$ have the same conditional asymptotic distribution, given the data. Observe that all quantities involved in the definition of $\hat{U}^*$ are known, thus, in principle, one could be able to know, or at least to accurately approximate its conditional distribution, given the data. This practical issue will be handled in Section 4.4.

Theorem 2 *Suppose that Assumptions (A1)-(A5) hold. Then, the conditional distribution of the d-dimensional process* $\hat{U}^*(y) = (\hat{U}_1^*(y), \ldots, \hat{U}_d^*(y))^T$*, given the data, converges to a centered d-dimensional Gaussian process* $U(y) = (U_1(y), \ldots, U_d(y))^T$ *with covariance structure given by* (4.6)*.*

The WB versions of S_{KS} and S_{CM}, say S_{KS}^* and S_{CM}^*, are analogously defined with $\hat{U}_k$ replaced by $\hat{U}_k^*$, $1 \le k \le d$, that is,

$$S_{KS}^* = \sum_{k=1}^{d} \sup_y |\hat{U}_k^*(y)|, \quad S_{CM}^* = \sum_{k=1}^{d} \int \hat{U}_k^{*2}(y) d\hat{F}(y),$$

The next Corollary gives the limit of the conditional distribution of S_{KS}^* and S_{CM}^*, given $(X_{11}, Y_{11}), \ldots, (X_{dn_d}, Y_{dn_d})$.

Corollary 1 *Suppose that Assumptions (A1)-(A5) hold. Then,*

(a) $\sup_x |P_*(S_{KS}^* \le x) - P(\sum_{k=1}^d \sup_y |U_j(y)| \le x)| \xrightarrow{P} 0,$

(b) $\sup_x |P_*(S_{CM}^* \le x) - P(\sum_{k=1}^d \int \{U_j(y)\}^2 dF(y) \le x)| \xrightarrow{P} 0.$

The result in Corollary 1 is valid whether the null hypothesis H_0 is true or not. An immediate consequence of this fact and the result in Corollary 2 in PF07 is the following.

Corollary 2 *If* H_0 *is true and the assumptions in Corollary 1 hold. Then,*

(a) $\sup_x |P_* \{S_{KS}^* \le x\} - P_0 \{S_{KS} \le x\}| \xrightarrow{P} 0.$

(b) $\sup_x |P_* \{S_{CM}^* \le x\} - P_0 \{S_{CM} \le x\}| \xrightarrow{P} 0.$

Let $\alpha \in (0, 1)$ and

$$\Psi_{KS} = \begin{cases} 1, & \text{if} \quad S_{KS} \ge s_{KS,\alpha}^*, \\ 0, & \text{otherwise}, \end{cases}$$

where $s^*_{KS,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of S^*_{KS}, or equivalently, $\Psi_{KS} = 1$ if $p^*_{KS} \leq \alpha$, where $p^*_{KS} = P_* \{S^*_{KS} \geq S_{KS}(obs)\}$ and $S_{KS}(obs)$ is the observed value of the test statistic S_{KS}. Analogously, let

$$\Psi_{CM} = \begin{cases} 1, & \text{if} \quad S_{CM} \geq s^*_{CM,\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $s^*_{CM,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of S^*_{CM}, or equivalently, $\Psi_{CM} = 1$ if $p^*_{CM} \leq \alpha$, where $p^*_{CM} = P_* \{S^*_{CM} \geq S_{CM}(obs)\}$ and $S_{CM}(obs)$ is the observed value of the test statistic S_{CM}.

The result in Corollary 2 states that Ψ_{KS} and Ψ_{CM} are asymptotically correct, in the sense that their type I error probabilities are asymptotically equal to the nominal value α.

From Proposition 1 and Corollary 1, it readily follows the next result.

Corollary 3 *Suppose that H_0 is not true and the assumptions in Corollary 1 hold, then $P(\Psi_{KS} = 1) \to 1$ and $P(\Psi_{CM} = 1) \to 1$.*

Corollary 3 shows that the tests Ψ_{KS} and Ψ_{CM} are consistent in the sense of being able to asymptotically detect any (fixed) alternative.

Remark 1 *The results stated in Theorems 1 and 2 and Corollaries 1–3 are also true if instead of using the raw multipliers ξ_{ij} we use the centered multipliers $\xi_{ls} - \bar{\xi}_{l\cdot}$, where $\bar{\xi}_{l\cdot} = \frac{1}{n_l}\sum_{s=1}^{n_l} \xi_{ls}$, $1 \leq s \leq n_l$, $1 \leq l \leq d$.*

4.4 Some practical issues

4.4.1 Calculation of the test statistic

Clearly,

$$S_{CM} = \frac{1}{n}\sum_{k=1}^{d}\sum_{i=1}^{d}\sum_{l=1}^{n_i}\{\hat{U}_k(\hat{\varepsilon}_{il})\}^2, \quad S^*_{CM} = \frac{1}{n}\sum_{k=1}^{d}\sum_{i=1}^{d}\sum_{l=1}^{n_i}\{\hat{U}^*_k(\hat{\varepsilon}_{il})\}^2.$$

As for S_{KS},

$$S_{KS} = \sum_{k=1}^{d} \max_{y\in\{\hat{\varepsilon}_{11},\ldots,\hat{\varepsilon}_{dn_d}\}} |\hat{U}_k(y)|. \tag{4.8}$$

The calculation of S^*_{KS} is not so easy. As an approximation to the suprema involved in the definition of S^*_{KS} and on the analogy of (4.8), in practice we calculate

$$S^*_{1,KS} = \sum_{k=1}^{d} \max_{y \in \{\hat{\varepsilon}_{11},\dots,\hat{\varepsilon}_{dn_d}\}} |\hat{U}^*_k(y)|$$

as an approximation to S^*_{KS}. This approximation was also considered in Kojadinovik and Yan (2012) for the problem of testing goodness-of-fit.

Let $0_s \in \mathbb{R}^s$ denote the vector with all its components equal to 0, let $1_s \in \mathbb{R}^s$ denote the vector with all its components equal to 1, let $e_1^T = (1_{n_1}^T, 0_{n_2}^T, 0_{n_3}^T, \dots, 0_{n_d}) \in \mathbb{R}^n$, $e_2^T = (0_{n_1}^T, 1_{n_2}^T, 0_{n_3}^T, \dots, 0_{n_d}) \in \mathbb{R}^n$, ..., $e_d^T = (0_{n_1}^T, 0_{n_2}^T, 0_{n_3}^T, \dots, 1_{n_d}) \in \mathbb{R}^n$, $v_k = \sqrt{n_k}\left(\frac{1}{n}1_n - \frac{1}{n_k}e_k\right)$, $1 \le k \le d$, and let M be the following $n \times n$-matrix

$$M = (M_{11}, M_{12}, \dots, M_{dn_d})$$

with $M_{kj}^T = (\hat{\varphi}(\hat{\varepsilon}_{11}, \hat{\varepsilon}_{kj}), \hat{\varphi}(\hat{\varepsilon}_{12}, \hat{\varepsilon}_{kj}), \dots, \hat{\varphi}(\hat{\varepsilon}_{dn_d}, \hat{\varepsilon}_{kj})) \in \mathbb{R}^n$, $1 \le j \le n_k$, $1 \le k \le d$. Note that once M has been calculated, by denoting $\xi^T = (\xi_{11}, \xi_{12}, \dots \xi_{dn_d})$, then $\hat{U}^*_k(\hat{\varepsilon}_{ij})$ can be easily calculated as follows

$$\hat{U}^*_k(\hat{\varepsilon}_{ij}) = v_k^T(M_{ij} \odot \xi),$$

$1 \le j \le n_i$, $1 \le k, i \le d$, where $\odot$ denotes the Hadamard product.

4.4.2 Calculation of the WB approximation

The calculation of the exact WB distribution of S_{CM} (S_{KS}) is, in general, unaffordable. Let S denote S_{CM} or S_{KS}. Let $\Delta(u) = 1$ if $u > 0$ and $\Delta(u) = 0$ if $u \le 0$. The WB p-value estimator can be approximated by simulation as follows:

1. Calculate the residuals $\hat{\varepsilon}_{11}, \dots, \hat{\varepsilon}_{dn_d}$.
2. Calculate the observed value of the test statistic, say S_{obs}.
3. Calculate M.
4. For some large integer B, repeat the following steps for every $b \in \{1, \dots, B\}$:
   (a) Generate $\xi_{11}, \dots, \xi_{dn_d}$ IID variables with mean 0 and variance 1.
   (b) Calculate S^{*b}.
5. Approximate the p-value by $\hat{p}^* = \frac{1}{B}\sum_{b=1}^{B} \Delta(S^{*b} - S_{obs})$.

4.5 Experimental results

The properties so far studied are asymptotic. In order to check the performance of the proposed procedure in practice, we ran a numerical simulation and a real data application. This section summarizes the obtained results. All computations in this work were performed by using programs written in the R language (R Core Team, 2015).

4.5.1 Simulated data

To empirically investigate the behavior of the proposed WB approximation for finite sample sizes, we carried out a simulation experiment. This experiment attempts to evaluate three main features: first, the goodness of the proposed WB approximation to the null distribution of the considered test statistics; second, to analyze the WB approximation in terms of power, comparing it to the power that results when the SB approximation is used (denoted as CvM_Boot and KS_Boot for the Cramér-von Mises and Kolgomorov-Smirnov type tests, respectively, in the tables); and finally, to compare the WB and the SB approximations in terms of the required CPU time.

In our simulations we took $d = 2$. Three specifications for the functions m_k and σ_k were considered:

S1: $Y_{kj} = X_{kj} + X_{kj}^2 + (X_{kj} + 0.5)\varepsilon_{kj}$, $1 \leq l \leq n_k$, $k = 1, 2$,

S2: $Y_{kj} = X_{kj} + 0.5\varepsilon_{kj}$, $1 \leq l \leq n_k$, $k = 1, 2$,

S3: $Y_{1j} = X_{1j} + X_{1j}^2 + (X_{1j} + 0.5)\varepsilon_{1j}$, $1 \leq j \leq n_1$, and $Y_{2j} = X_{2j} + 0.5\varepsilon_{2j}$, $1 \leq j \leq n_2$,

with $X_{kj} \sim U(0, 1)$, $1 \leq j \leq n_k$, $k = 1, 2$. For each of these specifications for m_k and σ_k, the following three cases were considered for the error distribution:

(i) $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$,

(ii) $\varepsilon_1 \sim N(0, 1)$, $\varepsilon_2 \sim E(1) - 1$,

(iii) $\varepsilon_1 \sim N(0, 1)$, $\varepsilon_2 \sim U(-\sqrt{3}, \sqrt{3})$,

where $E(1)$ stands for a negative exponential law with mean 1. Case (i) corresponds to the null hypothesis, while cases (ii) and (iii) are alternatives.

The Epanechnikov kernel $K(u) = 0.75 \times (1 - u^2)$ was employed to estimate the regression function, the conditional variance and f. For the choice of the bandwidths we proceeded as in the simulation study in PF07: we took $h_k = c \times n_k^{-a}$ to estimate m_k and

σ_k, with $a = 0.30$ and $c = \{0.50, 1.00, 1.50\}$ (although PF07 only displayed results for $c = \{0.50, 1.00\}$, in our simulations we tried a wider range of values for c, and decided to display the values for $c = 1.50$ since they yielded better results than for $c = 0.50$). To estimate f_k we took $a_k = h_k$.

1000 samples with sizes $n_1, n_2 \in \{50, 100\}$ were generated for each case and each specification for m_k and σ_k. For the WB approximation, the raw multipliers and the centered multipliers were considered, denoted in the tables by CvM_WB1, CvM_WB2, for the Cramér-von Mises type test, respectively, and analogously KS_WB1, KS_WB2 for the Kolgomorov-Smirnov type test. To approximate the SB p-value we proceeded as in PF07 by generating 200 bootstrap samples and using the same amount of smoothing, $\lambda_k = 2n_k^{-1/4}$; while for the WB approximation 1000 bootstrap samples were generated. The tables report the percentages of p-values less than or equal to $\alpha = 0.05$ and 0.10.

Tables 4.1–4.2 display the results for the level. Since the results for $c = 0.50$ were not as good as for the other values, these tables only present the results for $c \in \{1.00, 1.50\}$. Looking at them it can be seen that most values are quite close to the true value of α, for all specifications and for all tests. In general, $c = 1.0$ gives better results, in the sense of closeness to α, so we set this value to study the power.

Tables 4.3–4.4 display the results for the power. For the sake of brevity, we do not include the results for the specifications S2 and S3 because they are quite similar to the ones obtained for S1. As in the simulations in PF07, it is again observed that the Cramér-von Mises type test gives higher power than the Kolgomorov-Smirnov one. There are small differences in power for the WB with raw and centered multipliers. For smaller sample sizes, the tests based on the SB approximation exhibit higher power than the ones based on the WB approximation. Nevertheless, as the sample size increases, the power of both tests become closer. This is important because, as the sample sizes increase, the SB approximation becomes very time-consuming. This can be appreciated by looking at Tables 4.5–4.6, which display the CPU time consumed in seconds to get a p-value for several sample sizes by generating 1000 bootstrap samples in all cases. The difference in the required computing time with raw and the centered multipliers is negligible. It is evident that the WB approximation is more efficient than the SB one, in terms of the required computing time. The gain in computational efficiency of the WB over the SB stems from the fact that one does not have to re-estimate m and σ at each iteration (see Subsection 4.4.2), which slows down the process considerably.

		$c = 1.00$			c=1.50		
	n_1, n_2	CvM_Boot	CvM_WB1	CvM_WB2	CvM_Boot	CvM_WB1	CvM_WB2
	50,50	4.50	5.00	5.60	6.10	5.20	5.50
		8.60	9.70	10.20	11.00	10.60	11.30
S1	50,100	5.40	4.40	4.80	4.40	4.90	5.40
		10.40	10.30	10.80	10.20	10.10	10.90
	100,100	4.20	4.80	5.10	5.40	5.50	5.60
		9.30	9.20	9.50	10.50	10.00	10.60
	50,50	5.90	5.40	5.70	6.50	4.90	5.50
		11.30	10.70	11.00	12.00	11.30	12.00
S2	50,100	5.30	4.60	5.50	5.50	5.10	5.20
		10.60	10.10	10.70	9.00	10.00	10.40
	100,100	5.90	4.80	4.80	5.50	4.60	5.20
		10.90	9.50	9.40	9.00	9.00	9.40
	50,50	4.00	5.30	5.60	6.00	7.00	8.00
		10.00	10.30	10.60	12.50	12.00	12.60
S3	50,100	5.20	4.70	5.10	6.00	5.50	6.00
		11.60	10.60	11.00	11.50	11.00	12.00
	100,100	5.80	4.80	5.10	4.50	5.50	5.60
		9.80	9.30	9.80	11.50	11.00	12.00

Table 4.1: Percentage of rejections for the Cramér-von Mises type test at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specifications S1-S3.

		$c = 1.00$			c=1.50		
	n_1, n_2	KS_Boot	KS_WB1	KS_WB2	KS_Boot	KS_WB1	KS_WB2
	50,50	5.50	5.40	5.40	7.50	5.60	5.90
		12.00	11.50	11.90	13.00	12.00	13.00
S1	50,100	3.00	4.90	5.60	6.50	5.00	5.50
		8.00	11.00	11.50	9.00	11.50	12.00
	100,100	4.50	4.50	4.50	5.50	5.30	5.50
		9.00	9.60	9.80	12.50	10.10	10.30
	50,50	6.00	5.40	5.40	5.00	5.60	5.90
		12.00	11.50	11.90	14.00	12.00	13.00
S2	50,100	3.20	5.30	5.90	4.50	5.20	5.70
		7.50	10.50	11.00	9.50	11.00	12.00
	100,100	5.00	4.50	4.50	5.50	5.30	5.50
		9.00	9.60	9.80	10.00	10.10	10.30
	50,50	7.00	5.50	5.70	7.50	6.00	6.40
		14.00	11.20	12.00	13.50	12.70	13.50
S3	50,100	4.00	5.00	5.40	6.50	6.00	5.50
		9.00	10.90	11.80	12.00	12.00	12.30
	100,100	5.00	4.10	4.50	6.00	5.20	5.30
		9.00	9.10	9.20	11.00	9.30	10.00

Table 4.2: Percentage of rejections for the Kolgomorov-Smirnov type test at the significance levels 5% (upper entry) and 10% (lower entry) for case (i) and specifications S1-S3.

	$\alpha = 0.05$			$\alpha = 0.10$		
	n_1, n_2			n_1, n_2		
Method	50,50	50,100	100,100	50,50	50,100	100,100
CvM_Boot	76.00	88.20	95.60	85.60	92.60	98.00
CvM_WB1	65.60	73.40	95.90	79.00	85.40	98.10
CvM_WB2	67.00	75.00	96.00	79.60	86.40	98.20
KS_Boot	62.50	75.00	90.00	82.50	86.50	95.50
KS_WB1	50.90	60.20	88.30	66.90	74.60	94.40
KS_WB2	53.70	62.50	88.80	69.30	76.30	94.50

Table 4.3: Percentage of rejections at the significance levels 5% and 10% for case (ii) and specification S1.

	$\alpha = 0.05$			$\alpha = 0.10$		
	n_1, n_2			n_1, n_2		
Method	50,50	50,100	100,100	50,50	50,100	100,100
CvM_Boot	23.20	29.70	49.00	35.20	44.20	64.20
CvM_WB1	14.30	20.00	38.30	22.20	34.00	51.50
CvM_WB2	15.20	20.30	38.70	23.70	34.80	52.80
KS_Boot	11.50	17.50	23.00	25.50	30.50	40.50
KS_WB1	9.80	14.50	26.20	21.40	25.00	40.50
KS_WB2	10.10	15.00	27.20	22.60	26.00	41.20

Table 4.4: Percentage of rejections at the significance levels 5% and 10% for case (iii) and specification S1.

n_1, n_2	CvM_Boot/CvM_WB1	CvM_WB1	CvM_WB2
50,50	17.51	0.87	0.92
50,100	20.75	1.22	1.25
100,100	19.76	1.85	1.87
100,150	19.69	2.48	2.56
150,150	19.24	3.26	3.30

Table 4.5: CPU time in seconds for the calculation of a p-value for the Cramér-von Mises type test.

n_1, n_2	KS_Boot	KS_WB1	KS_WB2
50,50	17.53	0.92	0.93
50,100	21.89	1.23	1.39
100,100	20.61	1.84	1.88
100,150	20.96	2.48	2.59
150,150	20.69	3.20	3.29

Table 4.6: CPU time in seconds for the calculation of a p-value for the Kolgomorov-Smirnov type test.

4.5.2 Real data application

We applied the tests Ψ_{SK} and Ψ_{CM} to a real data set. To estimate the p-value we applied the WB approximations (with raw multipliers) for the two statistics considered with 1000 bootstrap samples. Next we briefly describe the application.

Kulasekera and Wang (2001) tested for the equality of two regression curves with data coming from a exhaust gas study using ethanol as fuel in an experimental design (the data set is available from Cleveland 1993). The equality of the error distributions in both populations is a crucial assumption in their statistical analysis, and thus it must be checked. From the first population (engines with low compression ratio) there are 39 observations, and 49 observations from the second one (engines with high compression ratio). The response Y is the concentration of nitric oxide plus the concentration of nitrogen dioxide in the exhaust of an experimental engine and the covariate X is the equivalence ratios at which the engine is set. Table 4.7 displays the estimated p-values for several values for a and c. Looking at this table we see that the equality of the error distribution cannot be rejected. This

		CvM_WB1			CvM_WB2			KS_WB1			KS_WB2		
a	c	1.00	1.25	1.50	1.00	1.25	1.50	1.00	1.25	1.50	1.00	1.25	1.50
0.30		0.367	0.229	0.248	0.346	0.218	0.225	0.341	0.182	0.141	0.324	0.159	0.120
0.35		0.467	0.314	0.215	0.415	0.294	0.204	0.651	0.336	0.182	0.616	0.616	0.159
0.40		0.470	0.406	0.302	0.416	0.366	0.287	0.478	0.545	0.338	0.425	0.425	0.320
0.45		0.375	0.496	0.394	0.333	0.449	0.347	0.250	0.597	0.543	0.217	0.217	0.504

Table 4.7: p-values for the real data analysis.

4.6 Appendix

4.6.1 Assumptions

Let $F_{X_k}(x) = P(X_k \leq x)$ and $F_{Y_k}(y|x) = P(Y_k \leq y|X_k = x)$ stand for the CDF of the covariate and the conditional CDF of the response given the covariate, respectively, $1 \leq k \leq d$.

(A1) For $1 \leq k \leq d$,

- **(i)** X_k is absolutely continuous with compact support S and density f_{X_k}.
- **(ii)** f_{X_k}, m_k and σ_k are two times continuously differentiable.
- **(iii)** $\inf_{x\in S} f_{X_k}(x) > 0$ and $\inf_{x\in S} \sigma_k(x) > 0$.

(A2) For $1 \leq k \leq d$,

- **(i)** $n_k/n \to p_k > 0$, with $\sum_{k=1}^{d} p_k = 1$.
- **(ii)** $n_k h_n^4 \to 0$ and $n_k h_n^{3+\delta} \to \infty$ for some $\delta > 0$.

(A3) For $1 \leq k \leq d$,

- **(i)** $E(\varepsilon_k^4) < \infty$.
- **(ii)** $F_{Y_k}(y|x)$ is continuous in (x, y), differentiable with respect to y, $F'_{Y_k}(y|x)$ is continuous in (x, y) and $\sup_{x,y} |y^2 F'_{Y_k}(y|x)| < \infty$. The same holds for all other partial derivatives of $F_{Y_k}(y|x)$ with respect to x and y up to order two.

(A4) K is a symmetric probability density function with compact support $[-1, 1]$, say, twice continuously differentiable and such that $K(-1) = K(1) = 0$.

(A5)

- **(i)** L is a symmetric probability density function with compact support and twice continuously differentiable. $L'(y)$, $yL(y)$ and $yL'(y)$ are of bounded variation.
- **(ii)** $a_n \to 0$, $na_n^4 \to 0$, $a_n^2/h_n^{1+\alpha/4} \to 0$, $nh_n a_n^{3+\alpha/3}/\log(h_n^{-1}) \to \infty$, $h_n = o(a_n^{1+\alpha/9})$.
- **(iii)** The first derivative of K is of bounded variation.

Assumptions (A1)–(A4) were employed in PF07 to derive the asymptotic null distribution of S_{KS} and S_{CM}. Assumption (A5) was required in Neumeyer (2006) to show the uniform convergence of $\hat{f}_k$ to f_k. Such result will be used in the proof of Theorem 2. For simplicity in notation we are assuming that the bandwidths (and the kernels) are the same in all populations. Nevertheless, all results remain true if different bandwidths (kernels) are considered for each population whenever they satisfy the required assumptions.

4.6.2 Preliminary results

Lemma 1 *Suppose that Assumptions (A1)-(A5) hold, then*

(a) $\frac{1}{n_k}\sum_{j=1}^{n_k}(\varepsilon_{kj} - \hat{\varepsilon}_{kj})^2 = o_P(1)$, $1 \leq k \leq d$.

(b) $\frac{1}{n_k}\sum_{j=1}^{n_k}(\hat{\varepsilon}_{kj}^2 - \varepsilon_{kj}^2)^2 = o_P(1)$, $1 \leq k \leq d$.

(c) $\frac{1}{n_k}\sum_{j=1}^{n_k}\hat{\varepsilon}_{kj}^2 = O_P(1)$, $1 \leq k \leq d$.

(d) $\frac{1}{n_k}\sum_{j=1}^{n_k}(\hat{\varepsilon}_{kj}^2 - 1)^2 = O_p(1)$, $1 \leq k \leq d$.

(e) $\hat{F}_k(y) = \frac{1}{n_k}\sum_{j=1}^{n_k} I(\varepsilon_{kj} \leq y) + o_P(n^{-1/2})$, *uniformly in* $y \in \mathbb{R}$, $1 \leq k \leq d$.

(f) $\hat{F}_k(y) = F_k(y) + o_P(1)$, *uniformly in* $y \in \mathbb{R}$, $1 \leq k \leq d$.

(g) $\hat{F}(y) = F(y) + o_P(1)$, *uniformly in* $y \in \mathbb{R}$.

(h) $\hat{f}_k(y) = f_k(y) + o_P(1)$, *uniformly in* $y \in \mathbb{R}$, $1 \leq k \leq d$.

(i) $\hat{f}(y) = f(y) + o_P(1)$, *uniformly in* $y \in \mathbb{R}$.

(j) $y\hat{f}(y) = yf(y) + o_P(1)$, *uniformly in* $y \in \mathbb{R}$.

Proof (a)–(d) Under Assumptions (A1),(A2) and (A.4) (see, for example, Masry 1996),

$$\begin{aligned}
&\sup_{x\in S}|\hat{m}_k(x) - m_k(x)| = o_P(n_k^{-1/4}),\\
&\sup_{x\in S}|\hat{\sigma}_k(x) - \sigma_k(x)| = o_P(n_k^{-1/4}), \quad 1 \leq k \leq d.
\end{aligned} \tag{4.9}$$

The difference between the residuals and the errors can be written as follows

$$\hat{\varepsilon}_{kj}-\varepsilon_{kj} = \varepsilon_{kj}\left(\frac{\sigma_k(X_{kj})-\hat{\sigma}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}\right)+\left(\frac{m_k(X_{kj})-\hat{m}_k(X_{kj})}{\hat{\sigma}_k(X_{kj})}\right), \quad 1 \le j \le n_k, \quad 1 \le k \le d. \tag{4.10}$$

The results in (a)–(d) follow from (4.9)–(4.10).

(e) From Theorem 1 in Akritas and Van Keilegom (2001) can write

$$\left|\hat{F}_k(y) - \frac{1}{n_k}\sum_{j=1}^{n_k} I(\varepsilon_{kj} \le y)\right| \le r_{k1}(y) + r_{k2}(y) + r_{k3}(y), \tag{4.11}$$

with

$$r_{k1}(y) = f(y)\left|\frac{1}{n_k}\sum_{j=1}^{n_k}\varepsilon_{kj}\right|, \quad r_{k2}(y) = \frac{1}{2}|y|f(y)\left|\frac{1}{n_k}\sum_{j=1}^{n_k}(\varepsilon_{kj}^2-1)\right|,$$

and $\sup_y r_{k3}(y) = o_P(n^{-1/2})$, $1 \le k \le d$. From the SLLN and (A3)(ii), we get that $\sup_y r_{kv}(y) = o_P(1)$, $v = 1, 2$, $1 \le k \le d$. Thus the result follows.

(f) The result follows from part (e) and the Glivenko–Cantelli theorem.

(g) The result follows from part (f) and (A2)(i).

(h) The result follows from Lemma 2.19 in Neumeyer (2006). Although Neumeyer (2006) considers the centered residuals, a closed look at her proof reveals that result is also true for the raw (noncentered) residuals.

(i) The result follows from part (h) and (A2)(i).

(j) The result follows from Lemma 2.20 in Neumeyer (2006). Although Neumeyer (2006) considers the centered residuals, a closed look at her proof reveals that result is also true for the raw (noncentered) residuals. □

4.6.3 Proof of main results

Proof of Proposition 1 (a) Note that

$$\frac{1}{\sqrt{n}}S_{KS} = \sum_{k=1}^{d}\left(\frac{n_k}{n}\right)^{1/2}\sup_y|\hat{F}(y)-\hat{F}_k(y)|.$$

From Lemma 1(f),

$$\hat{F}(y)-\hat{F}_k(y) = F(y)-F_k(y)+r_k(y), \quad \sup_y|r_k(y)| = o_P(1), \quad 1 \le k \le d. \tag{4.12}$$

The result for S_{KS} follows from (4.12) and (A2)(i).

(b) From (A2)(i) and (4.12), we can write

$$\frac{1}{n}S_{CM} = \sum_{k=1}^{d}\{p_k + o(1)\}\{I_{k1} + I_{k2} + I_{k3}\}, \tag{4.13}$$

where $I_{k1} = \int\{F(y) - F_k(y)\}^2 d\hat{F}(y)$, $I_{k2} = \int r_k^2(y)d\hat{F}(y)$, $I_{k3}^2 \leq I_{k1}I_{k2}$, $1 \leq k \leq d$. We have that $\{F(y) - F_k(y)\}^2$ is a continuous and bounded function. Also from Lemma 1(g), $\hat{F}(y) = F(y) + o_P(1)$. Thus, by the Helly-Bray theorem, it follows that

$$I_{k1} = \int\{F(y) - F_k(y)\}^2 dF(y) + o_P(1), \quad 1 \leq k \leq d. \tag{4.14}$$

As for I_{k2} we have that

$$I_{k2} \leq \sup_y r_k^2(y) \int d\hat{F}(y) = \sup_y r_k^2(y) = o_P(1), \quad 1 \leq k \leq d. \tag{4.15}$$

The result for S_{CM} follows from (4.13)–(4.15). □

Let $l^\infty(\mathcal{F})$ denote the space of all bounded functions from a set $\mathcal{F}$ to $\mathbb{R}$ equipped with the supremum norm $\|z\|_{\mathcal{F}} = \sup_{a\in\mathcal{F}}|z(a)|$.

Proof of Theorem 1 In the proof of Theorem 1 in PF07, it is shown that the class

$$\mathcal{F} = \left\{ \begin{array}{ll} g: & \mathbb{R} \to l^\infty(\mathbb{R}) \\ & x \to \{I(x \leq y) + xf(y) + \frac{1}{2}yf(y)(x^2 - 1),\ y \in \mathbb{R}\} \end{array} \right\}$$

$\mathcal{F}$ is Donsker. Theorem 1 in PF07 also shows that that $\hat{U}(y) \xrightarrow{\mathcal{L}} U(y)$. Applying Theorem 10.4 in Kosorov (2008) (see also, Theorem 2.9.6 in van der Vaart and Wellner, 1996) it follows that, conditionally on the data, $U^*(y) \xrightarrow{\mathcal{L}} U(y)$ in probability. □

Proof of Theorem 2 From Theorem 1, to show the result it suffices to prove that $D_k(y) = o_{P_*}(1)$, in probability, uniformly in $y \in \mathbb{R}$, $1 \leq k \leq d$, where

$$D_k(y) = \frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\hat{\varphi}(\hat{\varepsilon}_{kj}, y)\xi_{kj} - \frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\varphi(\varepsilon_{kj}, y)\xi_{kj}.$$

With this aim note that $D_k(y) = D_{1k}(y) + D_{2k}(y) + D_{3k}(y) + D_{4k}(y)$ with

$$D_{1k}(y) = \frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\{I(\hat{\varepsilon}_{kj} \leq y) - I(\varepsilon_{kj} \leq y)\}\,\xi_{kj}, \quad D_{2k}(y) = \{F(y) - \hat{F}(y)\}\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\xi_{kj},$$

$$D_{3k}(y) = \frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\left\{\hat{f}(y)\hat{\varepsilon}_{kj} - f(y)\varepsilon_{kj}\right\}\xi_{kj}, \quad D_{4k}(y) = \frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\left\{y\hat{f}(y)\hat{\varepsilon}_{kj} - yf(y)\varepsilon_{kj}\right\}\xi_{kj}.$$

It will be shown that $D_{sk}(y) = o_{P_*}(1)$, in probability, uniformly in $y \in \mathbb{R}$, $1 \leq s \leq 4$, $1 \leq k \leq d$.

Lemma 1 and Theorem 1 in Akritas and Van Keilegon (2001) and Theorem 10.4 in Kosorov (2008) (see also, Theorem 2.9.6 in van der Vaart and Wellner, 1996) imply that, conditionally on the data,

$$D_{1k}(y) = f(y)\frac{1}{n_k}\sum_{j=1}^{n_k}\varepsilon_{kj}\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\xi_{kj} + \frac{1}{2}f(y)\frac{1}{n_k}\sum_{j=1}^{n_k}(\varepsilon_{kj}^2 - 1)\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\xi_{kj} + o_P(1),$$

in probability, uniformly in $y \in \mathbb{R}$, $1 \leq k \leq d$. By the SLLN, $\frac{1}{n_k}\sum_{j=1}^{n_k}\varepsilon_{kj} = o(1)$ a.s. and $\frac{1}{n_k}\sum_{j=1}^{n_k}(\varepsilon_{kj}^2 - 1) = o(1)$ a.s.; by the CLT, $\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\xi_{kj}$ is bounded in probability. Since $f(y)$ and $yf(y)$ are bounded, it follows that $D_{1k}(y) = o_{P_*}(1)$, in probability, uniformly in $y \in \mathbb{R}$, $1 \leq k \leq d$.

As for $D_{2k}(y)$, from Lemma 1(g) and the CLT,

$$\sup_{y\in\mathbb{R}}|D_{2k}(y)| \leq \sup_{y\in\mathbb{R}}|F(y) - \hat{F}(y)|\frac{1}{\sqrt{n_j}}\sum_{j=1}^{n_k}|\xi_{kj}| = o_P(1),$$

showing that $D_{2k}(y) = o_{P_*}(1)$, in probability, uniformly in $y \in \mathbb{R}$, $1 \leq k \leq d$.

Now,

$$E_*\left[\left(\frac{1}{\sqrt{n_k}}\sum_{j=1}^{n_k}\left\{\hat{f}(y)\hat{\varepsilon}_{kj} - f(y)\varepsilon_{kj}\right\}\xi_{kj}\right)^2\right] = f(y)^2\frac{1}{n_k}\sum_{j=1}^{n_k}(\hat{\varepsilon}_{kj} - \varepsilon_{kj})^2.$$

Since f is bounded and by Lemma 1 (a), if follows that $D_{3k}(y) = o_{P_*}(1)$, in probability, uniformly in $y \in \mathbb{R}$, $1 \leq k \leq d$. Proceeding analogously, it can be seen that $D_{4k}(y) = o_{P_*}(1)$, in probability, uniformly in $y \in \mathbb{R}$, $1 \leq k \leq d$, which completes the proof. □

Proof of Corollary 1 (a) The convergence of the statistic S_{KS}^* follows directly from Theorem 2 and the continuous mapping theorem.

(b) The proof follows similar steps to that of Proposition 1 (b), so we omit it. □

4.7 References

1. Akritas M. G., Van Keilegom I., 2001. Non-parametric Estimation of the Residual Distribution. Scand J Stat 28, 549-567.

2. Alba-Fernández M. V., Batsidis A., Jiménez-Gamero M. D., Jodrá P., 2017. A class of tests for the two-sample problem for count data. J Comput Appl Math 318, 220–229.

3. Burke M. D., 2000. Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Statist Probab Lett 46, 13–20.

4. Cleveland W. S., 1993. Visualizing Data. Summit, New Jersey.

5. Hall P., Hart J. D., 1990. Bootstrap test for difference between means in nonparametric regression. J Amer Statist Assoc 85,1039–1049.

6. Jiménez-Gamero M. D., Alba-Fernández M. V., Jodrá P., Barranco-Chamorro I., 2016. Fast tests for the two-sample problem based on the empirical characteristic function. Math Comp Simulation, http://dx.doi.org/10.1016/j.matcom.2016.09.007

7. Kojadinovic I., Yan J., 2012. Goodness-of-fit testing based on a weighted bootstrap: A fast sample alternative to the parametric bootstrap. Can J Statist 40, 480–500.

8. Kosorok M. R., 2008. Introduction to empirical processes and semiparametric inference. Dordrecht: Springer.

9. Kulasekera K. B., Wang J., 2001. A test of the equality of regression curves using Gâteaux scores. Austral N Z J Statist 43, 89–99.

10. Masry E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. J Time Series Anal 17, 571–600.

11. Mora J., 2005. Comparing distribution functions of errors in linear models: A nonparametric approach. Statist Probab Lett 73, 425–432.

12. Neumeyer N., 2006. Bootstrap procedures for empirical processes based on nonparametric regression residuals. Habilitationsschrift, Ruhr-Universitä Bochum.

13. Pardo-Fernández J.C., 2007. Comparison of error distributions in nonparametric regression. Statist Probab Lett 77, 350–356.

14. Quessy J. F., Éthier F., 2012. Cramér-von Mises and characteristic function tests for the two and k-sample problems with dependent data. Comput Stat Data Anal 56, 2097–2111.

15. R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna-Austria, URL http://www.R-project.org/

16. van der Vaart A. W., Wellner J. A., 1996. Weak convergence and empirical processes. Springer, New York.

17. Young S. G., Bowman A. W., 1995. Non-parametric Analysis of Covariance. Biometrics 51, 920–931.