Université de Montréal

Common Factors in Stochastic Volatility of Asset Returns and New Developments of the Generalized Method of Moments

par

Prosper Dovonon

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences économiques

Octobre 2007

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Common Factors in Stochastic Volatility of Asset Returns and New Developments of the Generalized Method of Moments

présentée par

Prosper Dovonon

a été évaluée par un jury composé des personnes suivantes:

Benoit Perron
président-rapporteur
Eric Renault
directeur de recherche
Sílvia Gonçalves
codirectrice de recherche
Bryan Campbell
membre du jury (Concordia University)
Enrique Sentana
examinateur externe (CEMFI, Spain)
représenté par William McCausland
....
représentant du doyen de la FES

Thèse acceptée le:...

Sommaire

Les développements récents de l'économétrie de la finance ont pour base la découverte majeure que les carrés des rendements sont prévisibles quoique les rendements eux-mêmes ne le soient pas. Les célèbres modèles ARCH et GARCH introduits par Engle (1982) et Bollerslev (1986) prévoient le carré des rendements futurs par le carré des rendements passés. Toutefois, comme l'a plus tard suggéré Nelson (1991), le signe des rendements passés est aussi utile pour prévoir la volatilité mesurée par le carré des rendements. Ce fait stylisé reflète une corrélation asymétrique entre rendement et volatilité et est connu sous le nom d'effet de levier. L'effet de levier en particulier induit une asymétrie dans la distribution des rendements de plus longue échéance qu'on appelle effet de skewness. Bien que les premiers modèles univariés de volatilité aient connu des raffinements prenant en compte ces faits stylisés, les effets de levier et de skewness ne sont pas conjointement modélisés dans les modèles multivariés de volatilité. Comme généralisation du modèle de Diebold et Nerlove (1989), le modèle à facteur de volatilité stochastique proposé par Doz et Renault (2006) offre un cadre structurel adéquat à la modélisation multivariée de la volatilité des rendements sans exclure ni formaliser la variabilité dans les effets de levier et de skewness. Cette thèse, à travers son premier essai propose une extension de ce modèle à facteur en proposant des spécifications dynamiques pour les effets de levier et de skewness. Le deuxième essai évalue les bases théoriques du test de facteurs hétéroscédastiques proposé par Engle et Kozicki (1993) et y apporte une correction. Le troisième essai propose des méthodes de bootstrap pour l'inférence sur la matrice de covariance réalisée de processus multivarié de diffusion à volatilité stochastique telle qu'évaluée à partir des données de haute fréquence en finance et le quatrième essai s'inscrit dans le cadre des développements récents des méthodes d'inférence basées sur les conditions de moment (méthode des moments généralisée (GMM) et vraisemblance empirique).[1]

Dans le premier essai, nous proposons un modèle à facteur de volatilité stochastique avec effets de levier et de skewness dynamiques pour les rendements en étendant le modèle proposé par Doz et Renault (2006). Grâce à des conditions de moment, nous avons aussi proposé une inférence par la méthode des moments généralisée (GMM). Une application de ce modèle aux rendements journaliers excédentaires de 24 indices sectoriels incluant l'indice FTSE 350 et provenant tous du marché financier du Royaume Uni a été faite. La modélisation des effets de levier et de skewness a largement accru l'efficacité de l'estimateur des paramètres de volatilité. Les résultats suggèrent aussi que la

[1]Le deuxième essai de cette thèse a été écrit en collaboration avec Éric Renault et le troisième en collaboration avec Sílvia Gonçalves et Nour Meddahi.

compatibilité avec l'effet de skewness fait obtenir une persistance plus faible pour la volatilité et nous permettent également de documenter une relation entre l'effet de skewness et la volatilité.

Le deuxième essai réexamine les bases théoriques du test de facteurs hétéroscédastiques pour les processus multivariés de rendements proposé par Engle et Kozicki (1993). Ce test est fondé sur les conditions de moments résultant de la représentation factorielle et applique le test des restrictions suridentifiantes du GMM (Hansen (1982)). Cet essai montre que ces conditions de moment ne garantissent pas les conditions d'application de la théorie de test par GMM. En particulier, l'identification au premier ordre des paramètres n'est pas assurée. Nous proposons alors une théorie générale qui fournit la distribution asymptotique de la statistique du test de suridentification du GMM dans un contexte où le paramètre d'intérêt n'est pas identifiable au premier ordre mais l'est au deuxième ordre. Cette nouvelle théorie s'applique pour corriger le test de Engle et Kozicki (1993).

Dans le troisième essai, nous proposons des méthodes de bootstrap pour la matrice de covariance réalisée des processus multivariés de diffusion telle que mesurée sur les données de haute fréquence. Ces méthodes s'appliquent aussi aux fonctions de cette covariance telles que la covariance réalisée, la corrélation réalisée et le coefficient de régression réalisé. Il est à noter que le coefficient de régression réalisé inclus des statistiques aussi pertinentes pour l'analyse financière que les bêtas introduits par le capital asset pricing model (CAPM) pour évaluer le risque systématique des titres financiers.

Les méthodes de bootstrap que nous introduisons se veulent être une alternative pour l'approximation asymptotique de Barndorff-Nielsen et Shephard (2004). Spécifiquement, nous considérons le bootstrap i.i.d. appliqué aux vecteurs de rendements, c'est-à-dire que les données de bootstrap sont des tirages aléatoires des rendements haute fréquence. Malgré le fait que les données de bootstrap ainsi générées ne préservent pas le caractère hétéroscédastique des données originelles, nous montrons que cette méthode est valide asymptotiquement. Les expériences de Monte Carlo que nous avons effectuées suggèrent que la méthode de bootstrap que nous proposons fonctionne mieux que l'approximation asymptotique particulièrement lorsque les données sont générées à une fréquence faible ou modérée. Toutefois, contrairement aux résultats de la littérature du bootstrap i.i.d. pour les modèles de régression avec erreurs hétéroscédastiques, nous montrons par des expansions d'Edgeworth que le bootstrap i.i.d. ne donne pas lieu à des raffinements d'ordre supérieur dans notre contexte. Nous donnons une explication de cette différence.

Le quatrième essai porte sur les développements récents des méthodes d'inférence basées sur les conditions de moment. Cet essai propose un algorithme relativement simple permettant d'obtenir des estimateurs de moyenne de population de faible biais en échantillon fini grâce aux conditions de moment suridentifiantes.

Nous considérons aussi l'estimateur de vraisemblance euclidienne à trois étapes proposé par Antoine, Bonnal et Renault (2007). Quand les conditions de moment sont bien spécifiées, cet estimateur a un biais en échantillon fini d'ordre de grandeur aussi faible que celui de l'estimateur maximum de vraisemblance empirique et de plus il est plus facile à calculer que ce dernier. Nous étudions cet estimateur dans les modèles globalement mal spécifiés. Nous montrons que, même dans ces conditions irrégulières, l'estimateur 3S reste convergent au taux habituel ($\sqrt{n}$, où n est la taille de l'échantillon) et il est asymptotiquement normalement distribué.

Cet essai introduit aussi formellement l'estimateur de vraisemblance euclidienne à trois étapes corrigé (s3S) qui est défini de façon analogue à l'estimateur 3S mais utilise des probabilités impliquées corrigées pour être positives. L'idée d'utiliser des probabilités impliquées corrigées dans le calcul de l'estimateur à trois étapes a été proposée pour la première fois par Antoine, Bonnal et Renault (2007). Cependant, leur modification n'est pas robuste à la présence d'une mauvaise spécification des conditions de moment. Dans cet essai, nous proposons une autre modification des probabilités impliquées qui est robuste à la mauvaise spécification des conditions de moments. Cette robustesse est rendue possible en pondérant plus faiblement la différence entre les probabilités impliquées et leur équivalent asymptotique qui est $1/n$. Quand les modèles sont correctement spécifiés, les estimateurs 3S et s3S sont asymptotiquement équivalents à un ordre supérieur. Dans les modèles globalement mal spécifiés, nous montrons que l'estimateur s3S est aussi convergent au taux $\sqrt{n}$ et est asymptotiquement normalement distribué. Nous proposons aussi bien pour le 3S que pour le s3S leur distributions asymptotiques robustes à la mauvaise spécification des conditions de moment.

Dans cette frange de la littérature sur les estimateurs alternatifs au GMM, seul l'estimateur de maximum de vraisemblance empirique via minimum entropie (exponentially tilted empirical likelihood) (ETEL) proposé par Schennach (2007) a l'intérêt d'être convergent au taux usuel et asymptotiquement normalement distribué lorsque les conditions de moment sont mal spécifiées tout en étant équivalent à l'ordre supérieur à l'estimateur de maximum de vraisemblance empirique lorsque les conditions de moment sont bien spécifiées. Il importe cependant de noter que l'estimateur ETEL est relativement beaucoup plus difficile à calculer que les estimateurs 3S et s3S.

Mots clés: Modèle à facteurs, volatilité multivariée, asymétrie, GMM, sous-identification du premier ordre, Bootstrap, volatilité réalisée, expansions d'Edgeworth, vraisemblance empirique, mispécification.

Summary

The recent developments in financial econometrics are based on the major finding that square returns are predictable even though returns themselves are not. The famous ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986) predict future square returns by past square returns. Nevertheless, as observed by Nelson (1991), the signs of past returns are also useful to predict volatility, as measured by the squared returns. This stylized fact provides evidence of a negative correlation between returns and volatility and is known as the leverage effect. The leverage effect induces a negative skew in the distribution of lower frequency returns. This is the skewness effect. Although the first univariate volatility models have been refined to take account of these stylized facts, the literature on multivariate volatility models has not jointly modeled the dynamics of the leverage and skewness effects. As a generalization of the Diebold and Nerlove's (1989) model, the stochastic volatility factor model proposed by Doz and Renault (2006) provides a suitable structural framework for multivariate modeling of volatility in returns without neither precluding nor formalizing the time variability in both skewness and leverage effects. In the first chapter of this thesis, we extend this factor model by explicitly specifying dynamics for both skewness and leverage effects. The second chapter discusses the theoretical foundation of the test for common heteroskedastic factors proposed by Engle and Kozicki (1993). A correction for this test is also provided. The third chapter proposes bootstrap methods for the realized covariance of multivariate diffusion processes defined as the sum of the outer product of the vector of high frequency returns. The fourth chapter is related to the recent developments of moment conditions-based inference methods (generalized method of moments (GMM) and empirical likelihood methods).

In the first chapter, we propose a stochastic volatility factor model with dynamic skewness and leverage effects. This model is an extension of the model proposed by Doz and Renault (2006). To the best of our knowledge, we are the first to simultaneously model the conditional skewness and leverage effects in the context of a multivariate heteroskedastic factor model. We also provide moment conditions that allow for inference by GMM. We apply our model to 24 daily sector index excess returns from the United Kingdom stock market including the FTSE 350 index. The results show a large efficiency gain from modeling the skewness and leverage effects along with volatility. They

also suggest that the modeling of the conditional skewness effect yields lower volatility persistence as already pointed out by Harvey and Siddique (1999). We also document a significant relation between the skewness effect in returns and volatility.

The second chapter re-examines the theoretical foundations of the test for common heteroskedastic factors for multivariate return processes proposed by Engle and Kozicki (1993). This test is based on moment conditions resulting from the factor representation of returns and is an application of the GMM overidentification test (Hansen (1982)). We show that these moment conditions do not satisfy the identification conditions for the validity of the GMM test. In particular, the required first order identification condition for the parameter of interest is violated. We propose a general theory that provides the asymptotic distribution of the GMM overidentification test statistic when the parameters are not identified at the first order but are identified at the second order. We apply this new theory to correct the Engle and Kozicki's (1993) test.

The third chapter proposes bootstrap methods for the realized covariance of multivariate diffusion processes defined as the sum of the outer product of the vector of high frequency returns. These bootstrap methods can also be applied to economically meaningful functions of the realized covariance matrix such as the realized covariance between two assets, the realized correlation and the realized regression coefficients. Note that the realized regression coefficient includes as a particular case the realized beta, an important statistic for the financial analysis of the capital asset pricing model (CAPM). The realized beta of an asset assesses its systematic risk as measured by its correlation with the market portfolio return.

The bootstrap methods we consider are an alternative inference tool to the asymptotic theory recently proposed by Barndorff-Nielsen and Shephard (2004). More specifically, we consider the i.i.d. bootstrap and show its first order asymptotic validity. Our Monte Carlo experiments suggest that the bootstrap method we propose outperforms the asymptotic theory-based approximation of Barndorff-Nielsen and Shephard (2004), in particular when the series are not sampled too frequently. However, and contrary to the existing results in the bootstrap literature for regression models subject to heteroscedasticity in the error term, we show by Edgeworth expansions that the i.i.d. bootstrap is not second order accurate. We provide an explanation for this difference.

The fourth chapter is related to the recent developments in the literature based on the empirical likelihood interpretation of the GMM method. Its contribution is twofolded. First, we propose a new algorithm to compute estimators of population means whose small sample bias is of the same order of magnitude as the empirical likelihood estimator. This algorithm is easier to implement than the existing methods. Second, we study the asymptotic properties of the three-step Euclidean likelihood (3S) estimator as proposed by Antoine, Bonnal and Renault (2007) under the presence of possible misspecification in the moment conditions. As Antoine, Bonnal and Renault (2007) show, the higher order bias of the 3S estimator is of the same order of magnitude as that of the empirical likelihood estimator in correctly specified models. Nevertheless, the 3S estimator is much more computationally convenient than the empirical likelihood estimator. In this chapter, we show that in misspecified models the 3S estimator stays $\sqrt{n}$-consistent (where n is the sample size) and is asymptotically normally distributed. We also formally introduce the shrunk three-step Euclidean likelihood (s3S) estimator. This estimator is a variant of the 3S estimator which is derived using the Euclidean likelihood implied probabilities shrunk to be non negative. The idea of using modified Euclidean likelihood implied probabilities that are forced to be non negative was first proposed by Antoine, Bonnal and Renault (2007). Nevertheless their shrunk implied probabilities are not robust to misspecification. One of our contributions in this chapter is to proposed a further modification of the Euclidean likelihood implied probabilities by more weakly weighting their difference with their asymptotic equivalent, $1/n$. This modification appears to be crucial to get a proper behaviour of the three-step estimator under global misspecification. In correctly specified models, the 3S and the s3S estimators are asymptotically higher order equivalent. In globally misspecified models, we show that the s3S estimator is also $\sqrt{n}$-consistent and asymptotically normally distributed. We derive the asymptotic distribution of both estimators under the possibility of moment conditions misspecification.

In the existing literature on alternatives to the GMM estimator, only the exponentially tilted empirical likelihood estimator proposed by Schennach (2007) has the advantage of being $\sqrt{n}$-consistent and asymptotically normally distributed in misspecified models while displaying the same higher order bias as the empirical likelihood in correctly specified models. It is worthwhile however to mention that the three step Euclidean likelihood estimators are easier to compute than the ETEL estimator.

Key words: Factor models, multivariate volatility, asymmetry, GMM, first order underidentification, Bootstrap, realized volatility, Edgeworth expansions, empirical likelihood, misspecification.

Contents

List of Tables

List of Figures

Remerciements

Je dois l'accomplissement de ce travail à plusieurs personnes que je voudrais remercier sans avoir la prétention de réussir à être exhaustif.

Tout d'abord mes directeurs de recherche Éric Renault et Sílvia Gonçalves pour leur disponibilité et tout ce qu'ils m'ont patiemment et généreusement enseigné. Sans leur precieux soutien, cette thèse ne serait pas arrivée à ce stade.

Je voudrais également remercier Nour Meddahi avec qui j'ai particulièrement collaboré sur le troisième chapitre de cette thèse, pour ses commentaires sur le premier chapitre et pour tout ce qu'il m'a appris du même coup.

Merci aux Professeurs Jean-Marie Dufour, Benoit Perron, Rui Castro, Marine Carrasco, William McCausland et René Garcia pour leurs conseils et suggestions qui m'ont beaucoup apporté. Je voudrais aussi exprimer ma gratitude envers le Professeur Enrique Sentana pour sa disponibilité comme examinateur externe et surtout pour ses commentaires sur le deuxième essai de cette thèse. Mes remerciements aussi pour les Professeurs Atsushi Inoue et Paolo Zaffaroni pour leurs commentaires sur le quatrième essai.

Je voudrais aussi remercier les différents organismes qui m'ont apporté un support financier et logistique tout le long de ma formation: le Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), le Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ), the Mathematics of Information Technology and Complex Systems (MITACS) et le Conseil de Recherches en Sciences Humaines de Canada (CRSH).

Un Merci tout particulier :

À mon épouse Olivia pour le soutien quotidien et inconditionnel dont elle a fait preuve pour moi.

À ma mère Ida, mon père Valentin, mes frères et soeurs, Mérande, Noude, Houéfa, Carine et Aurèle pour leurs conseils et aussi leur patience en ces périodes où nous n'avons pu nous voir aussi souvent qu'on aurait souhaité.

À mes oncles et tantes Romuald Dossou, Elisabeth et Jacques Koussé pour leur régulier soutien.

À Hervé Koussé pour m'avoir aidé à m'intégrer si vite à Montréal. J'espère que ce travail lui serve de motivation pour ses études doctorales.

À mes amis Yves Atchadé, Andrea Cameron, Fousséni Chabi-Yo, Hervé Dovoédo, Fabien Houéto, Hervé Lohoues, Franck et Carine Goussanou, Guillermo Peres, Jason Roy, Aïcha Sow et Drogheda

Woods qui ne m'ont jamais oublié malgré les fois où je donnais l'impression de les avoir oublié.

À Emily Rose pour les échanges oh! combien utiles pour la prise de confiance en anglais et pour tous ses mots encourageants.

À mes amis et voisins de bureau: Isabelle Akaffou , Mame Astou Diouf, Bamba Diop, Firmin Doko, Ghislaine Geoffrion, Abderrahim Taamouti et Georges Tsafack pour les encouragements mutuels et les discussions enrichissantes.

À Abdoulaye Diaw et Pascal Martinolli pour leur soutien technique inconditionnel, et Suzanne Larouche-Sidoti pour les bonbons et les belles histoires pour changer de la routine.

À tout le personnel du département de sciences économiques de l'université de Montréal ainsi que celui du CIREQ pour leur dynamisme et leur efficacité.

Et enfin à mes collègues d'équipe de Barclays: Nicolas Bischoff, Ricardo Feced, Benoit Métayer, Panos Papanastasiou et Arnaud de Servigny pour leurs encouragements.

Introduction générale

Cette thèse est composée de quatre essais et s'inscrit dans le cadre des modèles multivariés de volatilité tout en contribuant aux développements récents de la méthode des moments généralisée.

Le premier et le troisième essais abordent des questions relatives à la modélisation de la volatilité multivariée. Le deuxième et le quatrième essais de cette thèse ont pour thème commum le comportement asymptotique dans des conditions non standard de certaines statistiques de tests et estimateurs issus de conditions de moment.

Dans le premier essai nous proposons une extension du modèle de volatilité multivariée de Doz et Renault (2006) qui tient explicitement en compte les dynamiques des effets de levier et de skewness des rendements. L'effet de levier se traduit par une corrélation asymétrique entre le rendement et la volatilité. Cet effet explique le fait stylisé que le signe des rendements passés est souvent utile pour prévoir la volatilité mesurée par le carré des rendements, comme l'a remarqué Nelson (1991). L'effet de levier en particulier induit une asymétrie dans la distribuition des rendements de plus longue échéance qu'on appelle effet de skewness. Bien que les premiers modèles univariés de volatilité aient connu des raffinements prenant en compte ces faits stylisés, les effets de levier et de skewness ne sont pas conjointement modélisés dans les modèles multivariés de volatilité.

Le modèle à facteurs latents hétéroscédastiques proposé par Diebold et Nerlove (1989) offre une alternative intéressante pour la modélisation multivariée de la volatilité. Ce modèle décompose chaque rendement en une partie systématique (ou commune à tous les rendements), qui est éventuellement source d'hétéroscédasticité, et une partie idiosyncratique. Il a le mérite de jouir d'une interprétation structurelle. Toutefois, Diebold et Nerlove (1989) complètent la spécification de leur modèle en imposant une distribution normale jointe au facteur et chocs idiosyncratiques. Ce choix de distribution, en mettant la skewness des rendements à zéro, s'écarte du comportement documenté des rendements d'actifs.

Plus récemment, Doz et Renault (2006) généralisent le modèle de Diebold et Nerlove (1989) en spécifiant un modèle à facteurs latents de volatilité stochastique qu'ils identifient grâce à des conditions de moment pertinentes. Ils proposent une inférence par la méthode des moments généralisée (GMM). Le recours aux conditions de moment permet de limiter les risques de modèle relatifs à la fixation de tous les moments au travers d'une distribution. Les conditions de moments utilisées par Doz et Renault (2006) permettent de capturer la dynamique dans la volatilité sans exclure ni formaliser explicitement la variabilité dans les effets de levier et de skewness.

Dans cet essai, nous proposons des conditions de moments additionelles à celles proposées pas Doz et Renault (2006) qui nous permettent de modéliser explicitement les effets de levier et de skewness dynamiques pour les rendements.

Dans une première partie, nous analysons les propriétés statistiques individuelles d'un ensemble de séries sur les rendements excédentaires journaliers d'indices sectoriels provenant du marché financier du Royaume Uni incluant l'indice FTSE 350. Des études récentes (voir e.g. Harvey, Ruiz et Sentana (1992) et King, Sentana et Wadhwani (2004)) utilisant des séries mensuelles proches de celles utilisées dans cet essai supportent le modèle à facteur hétéroscédastique. Notre analyse empirique suggère que ces rendements financiers démontrent très clairement aussi bien les phénomènes de levier que de skewness dynamiques, confirmant les faits empiriques déjà documentés par plusieurs autres auteurs, en particulier Nelson (1991), Hansen (1994) et Harvey et Siddique (1999).

Dans le modèle à facteur que nous proposons, le caractère hétéroscédastique des rendements est entraîné uniquement par le facteur commun que nous supposons de dynamique de volatilité stochastique autorégressive (SR-SARV). Les chocs idiosyncratiques sont supposés de volatilité constante. Par ce choix, il devient naturel de faire passer aussi bien l'effet de levier et celui de skewness par le même facteur. La dynamique que nous spécifions pour le levier dans le facteur est analogue à la forme la plus courante dans la littérature. Par contre, la spécification de la dynamique de skewness est déterminée par la robustesse du modèle vis-à-vis de l'agrégation temporelle. Il ressort qu'aussi bien le levier que la skewness dans le facteur sont une fonction affine de la volatilité. Nous montrons que notre modèle est robuste à l'agrégation temporelle.

Nos conditions de moments permettent une inférence par la méthode des moments. Dans l'application empirique de ce modèle à nos données, nous trouvons une efficacité plus accrue quand les effets de skewness et de levier sont pris en compte explicitement, ce qui reflète l'importance de ces phénomènes dans nos données. Les paramètres liant levier et volatilité d'une part et skewness et volatilité d'autre part sont fortement significatifs. Ceci documente en particulier la relation entre skewness et volatilité similaire au phénomène connu en finance sous le nom de volatility feedback (voir French, Schwert, et

Stambaugh (1987)) liant rendement et volatilité. De plus, la persistance de la volatilité paraît plus faible que ce qui s'observe habituellement dans les données journalières pour les modèles qui sont en contradiction avec l'effet de skewness dans les rendements. Ce dernier point confirme les faits documentés par Harvey et Siddique (1999), qui ont été les premiers à observer que les modèles compatibles avec l'effet de skewness ont un impact sur la persistance de la volatilité.

Le deuxième essai considère le test de suridentification de GMM tel que proposé par Hansen (1982). Si les conditions de moment suridentifiantes sont valides, sous certaines conditions de régularités, la statistique du test est asymptotiquement distribuée selon un Chi carré. Ces conditions de régularités incluent aussi bien l'identification stricte que l'identification au premier ordre du paramètre d'intérêt. L'identification stricte signifie que les conditions de moment déterminent une et une seule valeur du paramètre d'intérêt et l'identification au premier ordre impose que la jacobienne des conditions de moments évaluée à la vraie valeur est de plain rang.

Cet essai étudie la statistique du test de suridentification en relâchant la deuxième condition tout en maintenant l'identification au second ordre, signifiant que l'expansion des conditions de moment à l'ordre deux est suffisante pour identifier le paramètre d'intérêt. Une étude similaire a été effectuée par Sargan (1983) pour les estimateurs de variables instrumentales (IV). Dans son étude, Sargan (1983) s'intéresse au comportement asymtotique des estimateurs IV en cas de non identification au premier ordre. Dans cet essai, nous nous intéressons d'une part à la vitesse de convergence de l'estimateur de GMM en cas de déficience de rang et généralisons de ce point de vue les résultats de Sargan (1983). D'autre part et principalement, nous nous intéressons au comportement asymptotique de la statistique du test de suridentification de GMM dans cette condition de singularité. Il ressort de notre étude que les paramètres qui sont identifiés au premier ordre gardent la vitesse de convergence usuelle qui est de l'ordre de $\sqrt{T}$, T étant la taille d'échantillon, alors que les autres paramètres ont une vitesse de convergence plus lente de l'ordre de $T^{1/4}$. Ces comportements atypiques ont pour effet de changer la distribution asymptotique de la statistique du test de suridentification, qui suit un mélange de Chi carré plutôt qu'une Chi carré.

Une deuxième contribution de cet essai est de réexaminer les bases théoriques du test de facteurs hétéroscédastiques pour les processus multivariés de rendements proposé par Engle et Kozicki (1993). Ce test est fondé sur les conditions de moment résultantes de la représentation factorielle et est une application du test des restrictions suridentifiantes de GMM. Nous montrons que les conditions de moment d'Engle et Kozicki (1993), bien que vérifiant la condition d'identification du paramètre d'intérêt, violent la condition d'identification au premier ordre. Par contre, l'identification au second ordre y est assurée. Ceci nous place dans les conditions d'application de notre théorie asymptotique qui nous permet de corriger la distribution asymptotique suggérée par Engle et Kozicki (1993). Nous observons en outre que la distribution asymptotique de Engle et Kozicki sur rejette l'hypothèse nulle à un taux pouvant aller jusqu'à doubler le niveau nominal du test.

Dans le troisième essai nous proposons des méthodes d'inférence de bootstrap pour la volatilité multivariée intégrée. La volatilité multivariée intégrée est une mesure de volatilité multivariée sous-jacente à des processus multivariés de diffusion à volatilité stochastique. Un estimateur convergent de cette mesure de volatilité est la matrice de covariance réalisée, définie comme la somme du produit des rendements multivariés évalués à partir des données de haute fréquence. Barndorff-Nielsen et Shephard (2004) proposent une théorie asymptotique pour la matrice de covariance réalisée. Dans ce chapitre nous proposons une inférence par bootstrap plus exacte en échantillon fini que l'approximation asymptotique proposée par Barndorff-Nielsen et Shephard (2004).

Avec la richesse croissante des données financières, l'utilisation de statistiques fondées sur les données de haute fréquence ainsi que leur application en économie financière sont de plus en plus prépondérantes. La plus connue de ces statistiques est la volatilite réalisée. Son analogue multivarié est la matrice de covariance réalisée. Beaucoup de mesures de risque en finance sont fonctions de la matrice de covariance réalisée. On peut citer notamment la covariance réalisée, la corrélation réalisée ainsi que le coefficient de régression réalisé. Lorsque deux actifs sont considérés et l'un est le rendement sur le portefeuille du marché, le coefficient de régression réalisé devient le bêta du titre. Selon la fameuse théorie du capital asset pricing model (CAPM), le bêta mesure le risque systématique du titre.

Malgré la popularité grandissante des statistiques sur données haute fréquence, beaucoup reste à faire sur l'inférence pour ces statistiques. Barndorff-Nielsen et Shephard (2004) ont récemment proposé une théorie asymptotique pour la distribution de la matrice de covariance réalisée. Leur théorie permet de déduire les distributions asymptotiques de la covariance, de la corrélation ainsi que la régression réalisées entre deux rendement d'actifs. Toutefois, d'après les résultats de simulations qu'ils ont rapportés, l'approximation asymptotique souffre d'importantes distorsions en échantillon fini. Cette limitation est accentuée dans la pratique par le phénomène de microstructure de marchés, qui en soi réduit la validité des statistiques si les données considérées sont sur base de fréquence trop élevée.

Le troisième essai de cette thèse propose des méthodes de bootstrap comme alternative à la théorie asymptotique de Barndorff-Nielsen et Shephard (2004). Nous considérons le bootstrap i.i.d. appliqué au vecteur de rendements. Les données de bootstrap sont obtenues par tirages aléatoires des rendements multivariés originels.

Le bootstrap i.i.d. a été récemment proposé par Gonçalves et Meddahi (2006) dans le contexte univarié de la volatilité réalisée. Les données de bootstrap sont indépendantes et identiquement distribuées par construction et donc le bootstrap i.i.d. détruit le caractère hétéroscédastique des modèles de volatilité stochastique. Pour le cas de la volatilité réalisée, Gonçalves et Meddahi (2006) montrent que le taux de convergence vers zéro de l'erreur du bootstrap i.i.d. est du même ordre que le taux de convergence de l'erreur implicite dans l'approximation asymptotique. Cependant, les simulations de Gonçalves et Meddahi (2006) montrent que ce bootstrap est supérieure à la distribution asymptotique même quand la volatilité est stochastique. Ils donnent une explication théorique pour cette amélioration.

Dans cet essai, nous étendons l'analyse de Gonçalves et Meddahi (2006) au cas multivarié. Nous considérons le bootstrap i.i.d appliqué au vecteur de rendements. Dans le contexte de la régression réalisée, l'application du bootstrap i.i.d. au vecteur de rendements correspond à un bootstrap par couples, tel que proposé par Freedman (1981) pour des modèles de régressions de coupes transversales. Les résultats de Freedman (1981) et de Mammen (1993) montrent que le bootstrap par couples est non pas seulement robuste à la présence d'hétéroscédasticité dans l'erreur de la régression, mais il est même plus précis que la distribution asymptotique normale. Donc, le bootstrap i.i.d. paraît un candidat naturel dans le contexte de régressions réalisées même lorsque le modèle multivarié en question est un modèle de volatilité stochastique.

Nous montrons la validité asymptotique de bootstrap i.i.d. au premier ordre pour la matrice de covariance réalisée ainsi que pour des fonctions de ces éléments telles que la covariance réalisée et les coefficients de corrélation et de régression. Nos simulations montrent la supériorité remarquable du bootstrap sur l'approximation asymptotique, particulièrement sur les données de faible fréquence.

Nous dérivons l'expansion d'Edgeworth de la distribution de bootstrap pour la statistique de Student associée au coefficient de régression réalisé. Contrairement aux résultats de Mammen (1993), notre analyse montre que le bootstrap par couples ne permet pas une amélioration du taux de convergence de l'erreur de bootstrap dans l'estimation de la distribution de la statistique par comparaison avec l'erreur de l'approximation asymptotique. Nous conduisons une analyse détaillée du bootstrap par couples qui nous permet d'expliquer les différences de résultats obtenues. En particulier, nous montrons que les scores implicites à la régression réalisée ne sont pas individuellement de moyennes nulles (même si leur sommes demeurent de moyenne nulle). Par contre, Freedman (1981) et Mammen (1993) dérivent leurs résultats en faisant cette hypothèse. Le fait que chaque score ne soit pas de moyenne nulle individuellement crée un biais dans l'estimation de la variance de la régression par la méthode de Eicker-White et explique le besoin de l'estimateur de la variance de Barndorff-Nielsen et Shephard (2004), qui est plus sophistiqué que l'estimateur usuel de Eicker-White. Nous montrons que la variance de bootstrap par couples coincide avec l'estimateur de Eicker-White et donc elle n'est pas robuste à la présence d'hétéroscédasticité dans notre contexte de modèles de volatilité stochastique. Ceci contraste avec les résultats de Friedman (1981). Par contre, les scores de la régression de bootstrap sont individuellement de moyennes nulles et donc la statistique de bootstrap utilise l'estimateur de Eicker-White et non pas celui de Barndorff-Nielsen et Shephard (2004). Le fait que les deux statistiques, celle de bootstrap et la statistique originelle, utilisent des estimateurs de la variance différents explique pourquoi le bootstrap par couples ne permet pas une amélioration de l'approximation asymptotique dans notre contexte.

Le quatrième essai se démarque des questions de volatilité multivariée et a une contribution méthodologique plus générale. Il s'inscrit dans la littérature récente réinterprétant la méthode GMM à travers la vraisemblance empirique.

La technique d'inférence la plus populaire pour des modèles basés sur des conditions de moment est la méthode des moments généralisée proposée par Hansen (1982). La portée de cet outil s'explique surtout par sa simplicité et son efficacité asymptotique. Toutefois, plusieurs études ont rapporté des performances relativement faibles de l'approximation asymptotique du GMM en échantillon fini (voir e.g. Altonji et Segal (1996) et Andersen et Sørensen (1996)). Depuis lors, la littérature économétrique a connu un développement soutenu d'estimateurs alternatifs. Comme exemple, nous pouvons citer l'estimateur de GMM à mise à jour continue de Hansen Heaton et Yaron (1996), l'estimateur de maximum de vraisemblance empirique (EL) de Qin et Lawless (1993), l'estimateur "exponential tilting" de Kitamura et Stutzer (1997) qui sont tous à la fois membres de la classe d'estimateurs de divergence minimum de Corcoran (1998) et de la classe d'estimateurs de vraisemblance empirique généralisée de Newey et Smith (2004).

De ces estimateurs concurrents à l'estimateur de GMM, l'estimateur EL est connu comme celui ayant un biais en échantillon fini le plus désirable (Newey et Smith(2004)). Cependant, cet estimateur a deux défauts majeurs. En plus d'être très demandant en matière de calcul, il est aussi très instable lorsque le processus générateur des données dévie, ne serait ce que légèrement, des conditions de moment postulées par le modèle. Ceci a motivé la proposition par Schennach (2007) de l'estimateur de maximum de vraisemblance empirique via minimum d'entropie (exponentially tilted empirical likelihood) (ETEL). Cet estimateur jouit du même ordre de biais en échantillon fini que l'estimateur EL tout en restant stable en cas de mauvaise spécification des conditions de moment. Mais ETEL demeure aussi intensif en calcul que l'estimateur EL. Antoine, Bonnal et Renault (2007) propose l'estimateur de vraisemblance empirique euclidienne à trois étapes (3S). Cet estimateur est à la fois simple de calcul et a le même ordre de biais en échantillon fini que l'estimateur de maximum de vraisemblance empirique.

Une des contributions de cet essai est d'étudier l'estimateur 3S lorsque les conditions de moment sont mal spécifiées. Il montre que même dans ces conditions non standard, l'estimateur 3S converge à la vitesse habituelle et est asymptotiquement normalement distribué. La distribution asymptotique de l'estimateur 3S robuste à la mauvaise spécification est aussi proposée. Cependant, l'estimateur 3S a un défaut qui est relié à la nature des probabilités impliquées qui sont utilisées dans son calcul. Ces probabilités impliquées sont obtenues de la vraisemblance empirique euclidienne et sont connus comme pouvant être négatives en échantillon fini. Ceci peut être la cause de certains comportements erratiques de l'estimateur 3S en échantillon fini comme nous l'avons observé dans nos simulations.

Pour remédier à cette limite, Antoine, Bonnal et Renault (2007) suggèrent l'utilisation des probabilités impliquées corrigées dans le calcul de l'estimateur 3S. Nous redéfinissons formellement l'estimateur 3S à partir des probabilités impliquées corrigées proposées par Antoine, Bonnal et Renault (2007) et qui sont par définition toujours positives. Nous renforçons en outre le facteur de correction de façon à assurer à l'estimateur résultant (que nous appelons estimateur de vraisemblance empirique euclidienne à trois étapes corrigé ou shrunk three-step Euclidian likelihood estimator (s3S)) une convergence à la vitesse usuelle, $\sqrt{n}$, vers une distribution asymptotique normale en cas de mauvaise spécification des conditions de moment. Nous proposons aussi la distribution asymptotique de l'estimateur s3S robuste à la mauvaise spécification des conditions de moment.

La deuxième contribution de cet essai est la proposition d'un algorithme simple permettant d'obtenir des estimateurs de moyennes de population de biais en échantillon fini de même ordre que les estimateurs de vraisemblance empirique grâce aux conditions de moment suridentifiantes.

Chapter 1

Conditionally Heteroskedastic Factor Models with Skewness and Leverage Effects

1 Introduction

Conditional heteroskedasticity is a well-known feature of financial returns. In addition, returns are often characterized by the presence of skewness (i.e. returns have an asymmetric distribution) and leverage effects (i.e. the fact that a negative shock on returns has a larger impact on volatility than a positive shock of the same magnitude). See for example Nelson (1991), Glosten, Jagannathan and Runkle (1993) and Engle and Ng (1993) for studies documenting the presence of leverage effects in financial time series, and Ang and Chen (2002), Harvey and Siddique (1999, 2000) and Jondeau and Rockinger (2003) for the skewness effect.

The finance literature has recognized the importance of taking into account higher order moments in asset pricing models. An early example is Rubinstein (1973) (see also Kraus and Litzenberger (1976) for an empirical implementation of Rubinstein's (1973) model), who proposes an extension of the capital asset pricing model (CAPM) allowing for skewness in the unconditional distribution of returns. More recently, Harvey and Siddique (2000) extend Kraus and Litzenberger's (1976) model to the dynamic context by allowing third conditional moments to be time varying. The option pricing literature has also recognized the importance of modeling higher order moments. In particular, as pointed out by Hull and White (1987), the misspecification of the third order conditional moment can yield inaccurate option prices. This has motivated the development of option pricing models that take into account the skewness effect in the underlying asset. One recent example is Christoffersen, Heston and Jacobs (2006).

In the univariate context, Harvey and Siddique (1999) and Jondeau and Rockinger (2003) find that taking into account the skewness effect has an impact on the volatility persistence estimates. More specifically, for a set of daily and monthly index returns, Harvey and Siddique (1999) estimate univariate GARCH-type models that allow for time-varying conditional third-order moments. Their empirical results show that the estimates of volatility persistence decline when the model allows for the presence of skewness. They also find that the leverage effect tends to disappear following the introduction of skewness. These results are confirmed by Jondeau and Rockinger (2003), who also consider the effects of modeling the kurtosis in addition to the skewness effect. These studies show that empirically it is important to model the skewness and the leverage effects when building conditional heteroskedastic models for asset returns.

Although the finance literature in the univariate context has recognized the importance of modeling skewness and leverage effects, few attempts have been made to model both effects jointly in the multivariate framework. This is the case in the conditionally heteroskedastic factor model literature. In their seminal paper, Diebold and Nerlove (1989) assume conditional Gaussianity and postulate that the common factor follows an ARCH model, therefore not allowing for the presence of skewness nor leverage. More recently, Fiorentini, Sentana and Shephard (2004) propose a conditionally heteroskedastic factor model that allows for a dynamic leverage effect but impose a conditional Gaussianity assumption that rules out the conditional skewness.

In this paper we extend the existing class of multivariate conditionally heteroskedastic factor models by specifying simultaneously the skewness and the leverage effects. To the best of our knowledge, we are the first to write and estimate a conditionally heteroskedastic factor model that specifies jointly these two effects. In our model, all the dynamics in moments (and cross sectional, or co-moments) of asset returns are driven by a common latent factor. The conditional heteroskedasticity of the common factor follows a square root stochastic autoregressive volatility model (SR-SARV) as in Andersen (1994), Meddahi and Renault (2004) and Doz and Renault (2006). The leverage effect is modeled as an affine function of the conditional variance. This specification encompasses many of the existing models in the literature (e.g. the affine process of Dai and Singleton (2000)). The skewness effect is also modeled as an affine function of conditional variance. We show that this specification is robust to temporal aggregation when the leverage effect is present.

Recently, Doz and Renault (2006) (henceforth DR (2006)) study the identification and estimation of a conditionally heteroskedastic factor model. Specifically, DR (2006) provide a set of moment conditions that identify their factor model and allow for inference by Generalized Method of Moments (GMM), thus avoiding restrictive distributional assumptions. Our model is an extension of DR's (2004) model that explicitly models the skewness and the leverage effects. We follow DR (2006) and provide a set of moment conditions that identify the parameters of our extended model. We conduct a Monte Carlo experiment to investigate the finite sample properties of our estimation procedure for several values of the volatility persistence. We find that the method performs well in term of bias and root mean square error (RMSE) across our different models, except when the volatility persistence is very close to one.

We consider an empirical application of our model to a set of 24 daily stock index returns, including the FTSE 350 stock index return and 23 sectorial U.K. index returns. A monthly version of these data has previously been modeled by Harvey, Ruiz and Sentana (1992) and Fiorentini, Sentana and Shephard (2004) with a conditional heteroskedastic factor model. Our empirical study differs from theirs in that we analyze daily data and we model explicitly the dynamics of conditional higher order moments beyond the first and the second moment. For our data, we document strong evidence of conditional heteroskedasticity, as well as conditional leverage and skewness effects for all series. We also find evidence of significant co-skewness between the sectorial indices and the FTSE 350 index. Our empirical findings suggest the appropriateness of a conditionally heteroskedastic factor model with asymmetries (i.e. with leverage and skewness effects). We estimate our model by GMM. The fact that the volatility persistence on the factor is far away from one suggests that our procedure is valid in this application, given our Monte Carlo results. We also estimate the DR (2006) version of our model for which the skewness and leverage effects are not explicitly modeled. The first suggestion of our results is that there may be a substantial efficiency gain when both the conditional skewness and the leverage effects are modeled. In our case, the GMM standard error estimates of the parameters shared by both models drop sharply in our model compared to the DR (2006) model. Our results also suggest the presence of a significant leverage effect driven by a common factor in daily UK sectorial returns, confirming the results in Sentana (1995) for monthly data. The estimates of the volatility persistence for both our model and DR's (2006) model are relatively low and similar to one another. Since both models allow for the presence of conditional skewness (our model explicitly models it whereas DR (2006) does not, although it does not rule it out), the low persistence in volatility we obtain is consistent with the empirical results of Harvey and Siddique (1999) and Jondeau and Rockinger (2003) for univariate GARCH-type models, namely that conditional variance is less persistent when the conditional skewness is not ruled out. We also document the fact that an increase in volatility is associated with a more negatively skewed conditional distribution for the returns.

In our framework the common factor and the volatility for each asset are latent processes. We therefore propose an extended Kalman filter algorithm that provides a filter for both the latent factor and the volatility processes. Our filter algorithm is different from the filters proposed by Diebold and Nerlove (1989) and King, Sentana and Wadhwani (1994) in which the volatility process is a deterministic function of the latent factor process. Here both the latent factor and the volatility processes are unobserved and need to be considered as state variables. We apply our filter to the conditionally heteroskedastic factor model with the parameters estimated by the moment conditions given in DR (2006). We then perform some diagnostic tests on the filtered factor and volatility processes. Based on these tests, we cannot reject the correct specification of our model for the third conditional moments and the leverage effects. Moreover, we find that the filtered volatility process obtained with our model parameter estimates performs better than the filtered volatility process that relies on DR (2006) model parameters estimates, particularly in the periods of large shocks on returns. We explain this performance by the efficiency gain resulting from the returns' asymmetries modeling.

The remainder of the paper is organized as follows. In Section 2, we present the summary statistics for the data used in our empirical application. We also document the presence of dynamic leverage and skewness effects in our series. Section 3 presents our conditionally heteroskedastic factor model with skewness and leverage effects. Section 4 studies its temporal aggregation properties. Section 5 discusses the identification and estimation of the model. Section 6 contains the Monte Carlo study whereas Section 7 contains the empirical results. Section 8 concludes. The extended Kalman filter algorithm is presented in Appendix A. Appendix B contains the data description and all the tables with the empirical results. The proofs appear in Appendix C.

2 Empirical motivation

In this section we provide some empirical motivation for the need to account for asymmetric effects in both the conditional distribution (conditional skewness) and the conditional variances (leverage) of the data used in our empirical application. Our data set consists of 25 daily UK stock market index returns, including the FTSE 350 and 24 other sectorial indices, all of which in the FTSE. For the empirical application in Section 7, we restrict the set of index returns to 24. In particular, we exclude the FTSE All Share ex. inv. index because it is highly correlated with the FTSE 350 (with correlation coefficient 0.99) and including both indices could causes problems of multicollinearity. The data source is Datastream. Appendix B contains more details on the data. The period covered is January 2, 1986 to December 30, 2004, for a total of 4863 daily observations. Only trading days are considered. For each index, we compute the daily log excess return, using the log return of the UK one month loan index, the JPM UK CASH 1M, as the risk free interest rate. Appendix B contains all the tables with the empirical results in the paper.

Table 1.1 shows the correlation matrix of the 25 index excess returns we consider in this paper. Table 1.2 gives some descriptive statistics for our data, including the sample skewness and the kurtosis coefficients. We find that all indices have negative unconditional skewness (ranged from -1.70 to -0.00) except for the health sector index, which has positive skewness. However, this is not statistically significant at the 10% level. Out of the remaining 24 indices, 11 have statistically significant negative skewness. To test for the significance of skewness, we use a GMM-based test. In particular, we rely on a Wald type test involving the difference between the sample skewness and its null value, appropriately studentized. The asymptotic distribution of this test is easily obtained from the asymptotic distribution of the sample mean, given the delta-method and the fact that the skewness is a smooth function of the mean. To studentize the statistic, we follow Ang and Chen (2002) and estimate the long run variances by the 6-lags Newey and West (1987) heteroskedasticity and autocorrelation robust estimator using the Bartlett kernel. The presence of skewness in the distributions of the daily excess returns in the U.K. sectorial indices analyzed here agrees with similar evidence for other financial return series found by Harvey and Siddique (1999), Ang and Chen (2002) and Jondeau and Rockinger (2003), among others. Although the results in Table 1.2 apply to excess returns, risk-free rate adjusted, we also found evidence of skewness for the demeaned series, both adjusted for the day-of-the-week effect and filtered by autoregressions. Because the results did not change substantially, we do not report these results here.

The coefficient of unconditional kurtosis (which ranges from 6.72 to 18.62) is high for all series. Jointly with the values obtained for the unconditional skewness coefficients, the excess kurtosis values suggest that the normal distribution is not an appropriate description of our data, a stylized fact of many other financial time series. The Bera-Jarque normality test for dependent data proposed by Bai and Ng (2005) rejects the normality assumption for all the series. To conserve space, we do not report the results here. Figure 1.1 below confirms this result for the FTSE 350 stock index excess return. It gives the QQ plot for this series, i.e. it plots the empirical quantiles of the FTSE 350 index excess returns against the corresponding quantiles of the standard normal distribution. The departure from the normal distribution is clear and we can also notice the negative skewness of the FTSE 350 index.

Figure 1.1: Q-Q plot of the FTSE 350 index daily excess return

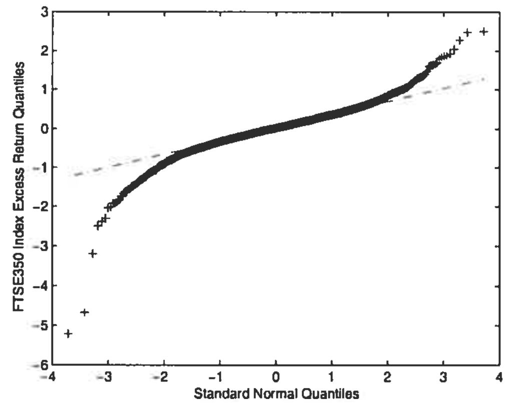


Table 1.2 also shows strong evidence of conditional heteroskedasticity as indicated by Engle's (1982) Lagrange multipliers tests of orders 1 and 5, denoted Eng(1) and Eng(5), respectively. The values of these statistics correspond to TR^2, where T is the sample size and R^2 is the R-squared from the regression of squared returns on a constant in addition to one and five lagged squared return, respectively. Under the null hypothesis of conditional homoskedasticity, TR^2 follows a chi-squared distribution with 1 and 5 degrees of freedom, respectively. The results indicate strong rejection of the null hypothesis of conditional homoskedasticity for all series. The Ljung-Box statistics for autocorrelation up to order 5 and 10 (QW(5) and QW(10), respectively) reveal the presence of potential autocorrelation in the data. Similarly, the first order autocorrelation coefficients ($\hat{\rho}_1$ in the table) are statistically significant for most time series, with some of the higher order autocorrelation coefficients remaining significant for some of them. Nevertheless, their magnitude is not very large (for instance, $\hat{\rho}_1$ varies between 0.02 and 0.23).

In Table 1.3 we present the results of diagnostic tests for the impact of news on volatility, as proposed by Engle and Ng (1993). These tests are the sign bias test, the negative size bias test, and the positive size bias test. They test for the significance of including the level of past standardized returns on the conditional variance equation and therefore can be used to test for the presence of leverage effects. Specifically, the sign bias test is a t-test for the significance of a dummy variable S_{t-1}^{-} (that takes the value one if the innovation to returns is negative and zero otherwise) in the regression of squared standardized returns on S^-_{t-1}. It checks whether volatility depends on the sign of the past innovation to returns. The negative (positive) size bias test instead checks whether the size of a negative (positive) return shock has an impact on volatility. We also include a joint test that tests simultaneously if any of these effects is present. We performed the diagnostic tests on the standardized index excess returns (using a GARCH(1,1) model as the null model under consideration as used by Engle and Ng (1993) in some of their applications) and on a filtered version of these, adjusted for the day-of-the-week and containing an autoregressive correction term. Since the results are similar, we only report the results for the centered excess return series in this paper. Table 1.3 shows that the sign bias test is significant for 10 out of the 25 time series considered. Nevertheless, the negative and the positive size bias tests are strongly significant for nearly all indices, which translates into significant joint tests for all of the series we analyze. We conclude that a GARCH(1,1) model is not a good description of volatility for the UK sectorial indices because it misses important leverage effects present in the data.

Our results so far suggest that a realistic data generating process for our data set should incorporate both leverage and unconditional skewness effects. Next we analyze the dynamic properties of these effects. In particular, we investigate the empirical content of two specifications for the conditional leverage and skewness effects. We model the dynamic leverage and the skewness effects as affine functions of volatility. Specifically, let $Y_{i,t+1}$ denote the excess return on the index i at time $t+1$ and let $\Sigma_{ii,t+1}$ denote the conditional variance of $Y_{i,t+2}$ at $t+1$. Let J_t denote the information set available at t. The conditional leverage effect is given by $Cov(Y_{i,t+1}, \Sigma_{ii,t+1}|J_t)$, which we model as $Cov(Y_{i,t+1}, \Sigma_{ii,t+1}|J_t) = \pi_0 + \pi_1 \Sigma_{ii,t}$. We test whether π_1 is significantly different from zero by regressing $\epsilon_{i,t+1} \times \hat{\Sigma}_{ii,t+1}$ on 1 and $\hat{\Sigma}_{ii,t}$, where $\epsilon_{i,t+1}$ is the unanticipated part of $Y_{i,t+1}$, as measured by the centered excess return $Y_{i,t+1} - \bar{Y}$, and where $\hat{\Sigma}_{ii,t+1}$ is the squared daily excess return at $t+2$, used as a proxy[1] for conditional volatility $\Sigma_{ii,t+1}$ of $Y_{i,t+2}$ at $t+1$. Similarly, we assume that the conditional skewness of excess returns is given by $E(Y^3_{i,t+1}|J_t) = h_0 + h_1 \Sigma_{ii,t}$ and test for the significance of h_1 in the regression of $\epsilon^3_{i,t+1}$ on 1 and $\hat{\Sigma}_{ii,t}$.

The results appear in Table 1.4. Both π_1 and h_1 are significantly different from 0 for all indices.

[1]Even though this proxy is known to be noisy, it gives us some insights for the dynamics in the skewness and leverage effects for our data. We will complement this preliminary analysis with more sophisticated diagnostic tests for our model in Section 7.

Except for the Persnl. Care & Hhld. Prods and the Health sectors, the estimates of π_1 and h_1 are negative for all of the sectors. The $\hat{\pi}_1$s are ranged from -2.82 to 0.37 while the $\hat{h}_1$s are ranged from -3.50 to 0.37. This means, for most of the indices that large increases in their volatility are associated to significant drops in both their leverage and their conditional skewness. These results, in particular, suggest that the leverage and skewness effects are time varying and their dynamics can be captured by an affine function of volatility.

To provide more evidence for these dynamics, we also use the (log) high-low range-based volatility estimator as a proxy for the conditional variance (see Parkinson (1980) and Brandt and Diebold (2004)). We perform these regressions only for the FTSE 350 index return. The series we consider cover the period from October 12, 1992 through October 13, 2006 for a total of 3511 daily observations. The conditional leverage and skewness regressions give $\hat{\pi}_1 = -0.27$ and $\hat{h}_1 = -0.23$, respectively. As in the previous case, both coefficient are strongly significant with -15.99 and -13.95 as t-stat, respectively.

Next, we perform some useful regressions to investigate the empirical content of an asymmetric factor model for our data[2]. We argue that if the data have a factor representation, the FTSE 350 index excess return should be a good proxy for this factor[3]. And, for such an asymmetric factor model to hold, both the conditional leverage and the skewness effects in our series should significantly be explained by the factor or equivalently by the FTSE 350 index excess return volatility. Let $Y_{1,t}$ denote the FTSE 350 index excess return at time t, $\Sigma_{11,t}$ the conditional variance of $Y_{1,t+1}$ at time t and $\hat{\Sigma}_{11,t} = Y_{1,t+1}^2$ its proxy. We test whether $\pi_{f,1}$, the slope of the regression of $\epsilon_{i,t+1} \times \hat{\Sigma}_{11,t+1}$ on 1 and $\hat{\Sigma}_{11,t}$ is statistically different from zero. We also test whether $h_{f,1}$, the slope of the regression of $\epsilon_{i,t+1}^3$ on 1 and $\hat{\Sigma}_{11,t}$ is significantly different from zero.

The results appear in Table 1.4. Both π_{f1} and h_{f1} are statistically significant for all indices. The estimates of these parameters lie between -2.45 and -0.94 for π_{f1}, and between -7.73 and -0.77 for h_{f1}. These results suggest that the conditional leverage and skewness in the index excess returns can also be explained by the FTSE 350 index excess return with the same qualitative interpretations as in the last regressions. We also compute the co-skewness of of each of the index excess returns with the FTSE 350. Table 1.4 shows the results. All of the co-skewnesses are negative (ranging from -0.96 and

[2]To the best of our knowledge, there is no test for skewness and leverage effects in the factor of a multivariate factor representation.

[3]This is confirmed by the high correlation coefficient (larger than 0.90) between the filtered factor and the FTSE 350 index excess return that we obtain in our empirical application in Section 7. See Table 1.9.

-0.13). These further investigations show evidence of potential common component in the asymmetric behaviour of the return processes.

In the next section we will propose a conditionally heteroskedastic factor model that incorporates dynamic conditional leverage and skewness effects modeled as affine functions of volatility.

3 The model

The main goal of this section is to propose a conditionally heteroskedastic factor model with skewness and leverage effects. To introduce some notation, we first present a conditionally heteroskedastic factor model for which these effects are not explicitly present. This model was recently studied by DR (2006) in the context of IV identification and estimation by GMM (see also Diebold and Nerlove (1989), King, Sentana and Wadhwani (1994), and Fiorentini, Sentana and Shephard (2004) for a discussion of conditionally heteroskedastic factor models). We will then present our model, which extends DR's (2006) model to include skewness and leverage dynamics.

3.1 A benchmark model

Let Y_{t+1} be a $N \times 1$ vector of (excess) returns on N assets from time t until time $t+1$, F_{t+1} a $K \times 1$ vector of K unobserved common factors, and U_{t+1} a $N \times 1$ vector of idiosyncratic shocks. DR (2006) consider the following conditionally heteroskedastic factor model for Y_{t+1},

$$Y_{t+1} = \mu(J_t) + \Lambda F_{t+1} + U_{t+1}, \tag{1}$$

with

$$\begin{array}{lcl} E\left(U_{t+1}|J_t\right) & = & 0 \\ E\left(F_{t+1}|J_t\right) & = & 0 \\ E\left(U_{t+1}\,F'_{t+1}|J_t\right) & = & 0 \\ Var\left(U_{t+1}|J_t\right) & = & \Omega \\ Var\left(F_{t+1}|J_t\right) & = & D_t, \end{array} \tag{2}$$

where J_t is a nondecreasing filtration defining the relevant conditioning information set containing the past values of $Y_\tau, \tau \leq t$ and and $F_\tau, \tau \leq t$, $\mu(J_t)$ is a $N \times 1$ vector of J_t-adapted components representing the risk premia, Λ is the $N \times K$ $(N \geq K)$ full column rank matrix of factor loadings, D_t is a diagonal positive definite matrix of K time-varying factor variances, and Ω is the conditional covariance matrix of the idiosyncratic shocks U_{t+1}.

The existing literature has made several assumptions about Ω. The strict factor structures as in Diebold and Nerlove (1989), King, Sentana and Wadhwani (1994) and Fiorentini, Sentana and Shephard (2004) impose Ω to be diagonal. The approximate factor structures as in Chamberlain and Rothschild (1983) and DR (2006) relax this assumption, allowing for nonzero off-diagonal elements. An advantage of the approximate factor representation is that it is preserved by portfolio formation, differently from the strict factor representation (see Chamberlain and Rothschild (1983), DR (2006) or Fiorentini, Sentana and Shephard (2004)).

Under (1) and (2), the conditional variance of Y_{t+1} given the information available at time t is given by:

$$\Sigma_t = Var\,(Y_{t+1}|J_t) = \Lambda\, D_t\, \Lambda' + \Omega. \tag{3}$$

The decomposition in (3) shows that the conditional variance of Y_{t+1} is time-varying, thus explaining why model (1) and (2) is called a conditionally heteroskedastic factor model.

For simplicity, we consider a constant risk premium for all assets, i.e. we assume $\mu\,(J_t) = \mu$ for all t. Following Nardari and Scruggs (2006), this restriction allows for the pricing relation $\mu = \Lambda\tau$, where τ is a $K \times 1$ vector of time-invariant factor risk premia. To simplify the exposition, we also assume a single factor representation, i.e. we will let $K = 1$ throughout. The generalization to $K > 1$ is nevertheless straightforward even though the inference issues may need to be discussed. Therefore, the model above specializes to

$$Y_{t+1} = \mu + \lambda f_{t+1} + U_{t+1}, \tag{4}$$

where f_{t+1} is the single common latent factor and λ is a $N \times 1$ vector of factor loadings. The moment conditions in (2) can be rewritten as

$$\begin{aligned} E\,(f_{t+1}|J_t) &= 0 \\ E\,(U_{t+1}|J_t) &= 0 \\ E\,(f_{t+1}U_{t+1}|J_t) &= 0 \\ Var\,(U_{t+1}|J_t) &= \Omega \\ Var\,(f_{t+1}|J_t) &= \sigma_t^2, \end{aligned} \tag{5}$$

implying that the conditional variance of Y_{t+1} given J_t is equal to

$$\Sigma_t = \lambda\lambda'\sigma_t^2 + \Omega. \tag{6}$$

To model σ_t^2, the conditional variance of f_{t+1} at time t, we follow DR (2006) and assume that f_t follows a square root-stochastic autoregressive volatility (SR-SARV(1)) model with respect to the filtration J_t (see Andersen (1994) and Meddahi and Renault (2004) for more details on this class of models), i.e.

$$f_{t+1} = \sigma_t \eta_{t+1}$$

where σ_t^2, the conditional variance of f_{t+1}, satisfies the following condition:

$$E(\sigma_t^2 | J_{t-1}) = \omega + \gamma\sigma_{t-1}^2, \quad \omega \geq 0, \ \gamma \in (0,1),$$

and where η_{t+1} is such that $E(\eta_{t+1}|J_t) = 0$ and $Var(\eta_{t+1}|J_t) = 1$.

Equations (4) through (5) describe the conditionally heteroskedastic model considered by DR (2006). Our main contribution in this section is to extend this model by explicitly modeling the skewness and the leverage effects.

3.2 The leverage effect

Fiorentini, Sentana and Shephard (2004) consider a conditionally heteroskedastic factor model that allows for a dynamic leverage effect. Nevertheless, in their model, the leverage effect is tightly linked to the QGARCH specification of Sentana (1995) for the conditional variance of the common factor. Here we adopt a different approach that disentangles these two features.

Given (6), we can write the conditional variance of asset i at time $t+1$ as

$$\Sigma_{ii,t+1} = \lambda_i^2 \sigma_{t+1}^2 + \Omega_{ii}, \qquad i = 1, \cdots, N,$$

where $\Sigma_{ii,t+1}$ denotes the element (i,i) of the matrix Σ_{t+1}, and similarly for Ω_{ii}.

Let $u_{i,t+1}$ be the i-th component of U_{t+1}. The conditional leverage effect at $t+1$ can be expressed as the conditional covariance between $Y_{i,t+1}$ and $\Sigma_{ii,t+1}$, given J_t, i.e.

$$\begin{aligned} Cov\left(Y_{i,t+1}, \Sigma_{ii,t+1}|J_t\right) &= Cov\left(\lambda_i f_{t+1} + u_{i,t+1}, \lambda_i^2 \sigma_{t+1}^2 + \Omega_{ii}|J_t\right) \\ &= \lambda_i^3 Cov\left(f_{t+1}, \sigma_{t+1}^2|J_t\right) + \lambda_i^2 Cov\left(u_{i,t+1}, \sigma_{t+1}^2|J_t\right). \end{aligned} \tag{7}$$

Equation (7) shows that the leverage effect for each return has two components. The first component reflects the part of the leverage effect that is due to leverage in the common factor whereas the second component is given by the conditional covariance between the idiosyncratic shock and the future volatility of the factor. The dynamics of each component dictate the dynamics of the leverage effect.

In this paper we assume that the idiosyncratic shock is not conditionally correlated with the conditional variance of the factor. We formally state this assumption below.

ASSUMPTION 1 *The conditional correlation between $u_{i,t+1}$ and σ_{t+1}^2 for each asset i is zero:*

$$Cov(u_{i,t+1},\, \sigma_{t+1}^2 | J_t) = 0, \qquad i = 1,\, \ldots,\, N.$$

This assumption implies to imposing a null correlation between $u_{i,t}$ and f_{t+1}^2. The following assumption gives a model for the leverage effect for the latent factor.

ASSUMPTION 2 *$Cov(f_{t+1},\, \sigma_{t+1}^2 | J_t) = \pi_0 + \pi_1 \sigma_t^2$, for some constants π_0 and π_1.*

According to this assumption, the leverage effect for the factor is an affine function of its conditional variance. As we will show next, this specification holds for many models in the class of SR-SARV processes[4]:

EXAMPLE 1 *The $\mathbb{A}_1(3)$-affine family processes[5] (Dai and Singleton (2000), Singleton (2001)).*

Let f_{t+1} be defined by

$$f_{t+1} = \sqrt{\alpha + v_t}\epsilon_{1,t+1} + \sigma_1 \eta \sqrt{v_t}\epsilon_{2,t+1} + \sigma_2 \sqrt{\zeta^2 + \beta v_t}\epsilon_{3,t+1}, \quad \epsilon_{t+1}|J_t \sim \mathcal{N}(0, I_3)$$

$$\begin{aligned} \theta_{t+1} &= \nu\bar{\theta} + (1-\nu)\theta_t + \sqrt{\zeta^2 + \beta v_t}\epsilon_{3,t+1} + \sigma_3 \eta \sqrt{v_t}\epsilon_{2,t+1} + \sigma_4 \sqrt{\alpha + v_t}\epsilon_{1,t+1}, \\ v_{t+1} &= \mu\bar{v} + (1-\mu)v_t + \eta\sqrt{v_t}\epsilon_{2,t+1}, \end{aligned}$$

where $(\alpha, \beta, \eta, \nu, \mu, \zeta, \bar{\theta}, \bar{v}, \sigma_1, \sigma_2, \sigma_3, \sigma_4) \in \mathbb{D}$, a conveniently restricted subset of $\mathbb{R}^{12}$. It follows that

$$Cov(f_{t+1},\, \sigma_{t+1}^2 | J_t) = -\sigma_1 \eta^2 (\alpha + \sigma_2^2 \zeta^2) + \sigma_1 \eta^2 \sigma_t^2 = \pi_0 + \pi_1 \sigma_t^2.$$

$\sigma_t^2 \equiv Var\,(f_{t+1}|J_t) = \alpha + \sigma_2^2\zeta^2 + (1 + \sigma_1^2\eta^2 + \sigma_2^2\beta)v_t$. Thus, the affine process verifies Assumption 2.

[4] Many of our examples correspond to the discrete version of continuous time models used in finance.

[5] Backus-Foresi-Telmer(2001) use the discrete time version of the Cox-Ingersoll-Ross's (1985) diffusion process to propose an affine model of currency. The affine process nests the square-root process of Heston (1993) and Cox-Ingersoll-Ross (1985).

Example 2 *The Quadratic GARCH (QGARCH(1,1)) of Sentana (1995).*

Let f_{t+1} be given by

$$f_{t+1} = \sigma_t \eta_{t+1}, \qquad \eta_{t+1}|J_t \sim \mathcal{N}(0,\, 1)\,,$$

where σ_t^2 is such that

$$\sigma_t^2 = \theta + \beta\,\sigma_{t-1}^2 + \alpha\,(f_t - \mu)^2\,, \quad (\theta,\, \beta,\, \alpha,\, \mu) \in \mathbb{D}.$$

It follows that

$$Cov(f_{t+1},\, \sigma_{t+1}^2|J_t) = -2\alpha\mu\sigma_t^2 = \pi_1\sigma_t^2,$$

showing that QGARCH(1,1) satisfies Assumption 2.

Example 3 *Heston-Nandi's (2000) GARCH process.*

Let f_{t+1} be given by

$$f_{t+1} = \sigma_t \eta_{t+1}, \qquad \eta_{t+1}|J_t \sim \mathcal{N}(0,\, 1)\,,$$

where σ_t^2 is defined as follows:

$$\sigma_{t+1}^2 = \omega + \beta\,\sigma_t^2 + \alpha\,(\eta_t - \gamma\sigma_t)^2\,, \quad (\omega,\, \beta,\, \alpha,\, \gamma) \in \mathbb{D}.$$

We can show that

$$Cov(f_{t+1},\, \sigma_{t+1}^2|J_t) = -2\alpha\gamma\sigma_t^2 = \pi_1\sigma_t^2,$$

proving that this model also satisfies our Assumption 2.

Example 4 *The Inverse Gaussian GARCH(1,1) of Christoffersen, Heston and Jacobs (2006).*

In the Inverse-Gaussian-GARCH(1,1) (IV-GARCH (1,1)) model proposed by Christoffersen, Heston and Jacobs (2006) for a random process f_{t+1} (e.g. a log return process), f_{t+1} is written as the sum of a deterministic random process and an innovation which follows an Inverse Gaussian distribution. They show that this model is embodied in the class of SR-SARV(1) models and allows for a leverage effect of the type considered in Assumption 2. In particular, they show that $Cov(f_{t+1},\, \sigma_{t+1}^2|J_t) = \pi_1\sigma_t^2$ for some π_1 where σ_t^2 is the conditional variance of f_{t+1}.

Given equation (7) and Assumptions 1 and 2, we can write the leverage effect for asset i as follows:

$$Cov\,(Y_{i,t+1},\, \Sigma_{ii,t+1}|J_t) = \lambda_i^3\,(\pi_0 + \pi_1\sigma_t^2)\,. \tag{8}$$

3.3 The skewness effect

In this section we propose a model for the dynamics in the conditional skewness of assets returns. This effect has often been ruled out in the conditionally heteroskedastic factor literature, which typically postulates the conditional normality of $(f_{t+1}, U'_{t+1})'$. To the best of our knowledge, we are the first to model this effect in the conditionally heteroskedastic factor literature.

The finance literature on univariate processes has modeled the skewness effect by specifying conditional distributions which allow for time varying conditional third order moments. See for instance Hansen (1994), Harvey and Siddique (1999) and Jondeau and Rockinger (2003). Because we would like to avoid any distributional assumptions, we will follow an alternative approach in which the skewness effect is specified through a conditional moment restrictions.

Without loss of generality, assume that $\mu(J_t) = \mu = 0$. It follows that for each $i = 1, \ldots, N$,

$$E\left(Y_{i,t+1}^3 | J_t\right) = \lambda_i^3 E\left(f_{t+1}^3 | J_t\right) + E\left(u_{i,t+1}^3 | J_t\right) + 3\lambda_i^2 E\left(f_{t+1}^2 u_{i,t+1} | J_t\right) + 3\lambda_i E\left(f_{t+1} u_{i,t+1}^2 | J_t\right).$$

In order to obtain a simplified expression for $E\left(Y_{i,t+1}^3 | J_t\right)$, we make the following assumption. Note that DR (2006) use a similar assumption (see their Assumption 3.6).

Assumption 3 *f_{t+1} and f_{t+1}^2 are conditionally uncorrelated with any polynomial function of $u_{i,t+1}$ of degree smaller than three.*

Assumption 1 assumes that f_{t+1}^2 is conditionally uncorrelated with u_{it}. Assumption 3 extends Assumption 1 by requiring that f_{t+1}^2 be also conditionally uncorrelated with $u_{i,t+1}$. Assumptions 1 and 3 are satisfied if the latent factor is conditionally independent of the idiosyncratic shocks.

By Assumption 3,

$$E\left(f_{t+1}^2 u_{i,t+1} | J_t\right) = E\left(f_{t+1} u_{i,t+1}^2 | J_t\right) = 0,$$

which implies that

$$E\left(Y_{i,t+1}^3 | J_t\right) = \lambda_i^3 E\left(f_{t+1}^3 | J_t\right) + E\left(u_{i,t+1}^3 | J_t\right). \tag{9}$$

As a result, specifying the dynamics of the third order conditional moments of the factor and of the idiosyncratic shocks is equivalent to modeling the third order conditional moment of the excess return. We introduce the following assumption.

Assumption 4 *For all $i = 1, \ldots, N$,*

$$E\left(u_{i,t+1}^3 | J_t\right) = s_i^0, \quad s_i^0 \in \mathbb{R}.$$

Assumption 3 assumes that the third order conditional moment of $u_{i,t+1}$ is time-invariant. The idiosyncratic shocks are therefore not necessarily (conditionally) Gaussian, although we assume that their first three moments are not time varying.

To model the conditional skewness in the conditional distribution of f_{t+1}, we make the following assumption.

Assumption 5

$$E\left(f_{t+1}^3 | J_t\right) = h_0 + h_1 \sigma_t^2, \quad h_0,\, h_1 \in \mathbb{R}.$$

According to Assumption 5, the conditional skewness in f_{t+1} is an affine function of σ_t^2, the conditional variance of the factor. As we will prove in Proposition 4.4 below, this specification is robust to temporal aggregation of the model when the leverage effect is specified by Assumption 2. Moreover, this model has good empirical support for our data, as showed by our empirical results in Section 2.

The models introduced in Examples 21-24 above satisfy Assumption 5 in addition to Assumption 2. Assumptions 3-5 imply that

$$E(Y_{i,t+1}^3 | J_t) = \lambda_i^3 h_1 \sigma_t^2 + s_i, \tag{10}$$

with $s_i = \lambda_i^3 h_0 + s_i^0 : i = 1, \ldots, N$. Equation (10) shows on the one hand that h_0 and $s_i^0 : i = 1, \ldots, N$ cannot be simultaneously identified by the third conditional moment of the returns and on the other hand that the conditional skewness of the returns, $Y_{i,t+1} : i = 1, \ldots, N$ are affine functions of σ_t^2.

Assumptions 2 and 5 nest the standard GARCH(1,1) models allowing for the possible presence of skewness when the standardized innovation in f_{t+1}, η_{t+1}, has its third order conditional moment, $E(\eta_{t+1}^3 | J_t)$, proportional to $1/\sigma_t$. This is the case for the standard Gaussian GARCH(1,1) model. However, this model (possibly with skewness) does not disentangle the skewness effect from the leverage. In particular, if f_{t+1} is a GARCH(1,1) process, $Cov\left(f_{t+1}, \sigma_{t+1}^2 | J_t\right) = \alpha E\left(f_{t+1}^3 | J_t\right)$. This is rather a drawback for this class of models as pointed out by Alami and Renault (2001). In contrast, Assumptions 2 and 5 allow for the two effects to be independent of each other.

3.4 Our model

The following equations summarize our conditionally heteroskedastic factor model with asymmetries.

$$Y_{t+1} = \mu + \lambda f_{t+1} + U_{t+1} \tag{11a}$$

$$E\left(f_{t+1}|J_t\right) = 0 \tag{11b}$$

$$E\left(U_{t+1}|J_t\right) = 0 \tag{11c}$$

$$Var\left(U_{t+1}|J_t\right) = \Omega \tag{11d}$$

$$E\left(f_{t+1}U_{t+1}|J_t\right) = 0 \tag{11e}$$

$$Var\left(f_{t+1}|J_t\right) \equiv \sigma_t^2 \tag{11f}$$

$$E\left(\sigma_{t+1}^2|J_t\right) = 1 - \gamma + \gamma\sigma_t^2, \quad \gamma \in (0,1) \tag{11g}$$

$$E\left[(Y_{i,t+1} - \mu)^3|J_t\right] = \lambda_i^3 h_1 \sigma_t^2 + s_i, \quad i = 1, \ldots, N \tag{11h}$$

$$Cov\left(f_{t+1}, \sigma_{t+1}^2|J_t\right) = \pi_0 + \pi_1\sigma_t^2. \tag{11i}$$

The skewness effect depends on the parameters h_1 and s_i, for $i = 1, \ldots, N$, while π_0 and π_1 characterize the leverage dynamics. Equation (11g) specifies the factor volatility dynamics. The SR-SARV(1) model restricts the volatility intercept ω to $1-\gamma$, where γ is the factor conditional variance persistence, such that $E\sigma_t^2 = Ef_t^2 = 1$. This condition fix the scale problem that may arise out from the specification in (11a) for the factor loadings and the latent factor. However, as pointed out by DR (2006), their model given by Equations (11a)-(11g) identifies all of the parameters involved except one arbitrary factor loading. This remark remains true in our model. The additional conditional moment restrictions we provide by (11h) and (11i) only identify the asymmetry parameters and some factor loading ratios which all are identified by (11a)-(11g). We will discuss this identification problem more extensively in Section 5.

This identification issue is the first main difference between the conditional moment restrictions-based model we have here and the parametric factor models. The conditional joint distribution assumption for (f_{t+1}, U_{t+1}') in parametric models together with the factor normalization identify the whole model up to one factor loading sign. This sign indeterminacy is solved, as proposed by Geweke and Zhou (1996), by restricting the sign of a particular factor loading to be positive. For a more extensive discussion about this identification issue for parametric models, see Geweke and Zhou (1996) and Aguilar and West (2000).

Equations (11f) through (11i) show that the common factor drives the dynamics in the conditional variance, the conditional skewness and the conditional leverage of assets returns. In particular, as we showed in (8), the leverage effect for asset i can be expressed[6] as

$$l_{ii,t} = \lambda_i^3 l_t,$$

where $l_{ii,t} \equiv Cov\left(Y_{i,t+1}, \Sigma_{ii,t+1} | J_t\right)$ denotes the leverage effect for the asset and $l_t \equiv Cov\left(f_{t+1}, \sigma_{t+1}^2 | J_t\right)$ denotes the leverage effect for the factor. Therefore, if an asset has a positive factor loading, its leverage effect is positively correlated with the leverage effect for the factor. Instead, a negative relationship holds between the two leverage effects if the factor loading is negative.

We can also define the co-leverage (or transversal leverage) between two assets i and j. This is given by $l_{ij,t} \equiv Cov\left(y_{i,t+1}, \Sigma_{jj,t+1} | J_t\right)$. Note that the order of the arguments matters in this definition of the co-leverage. Specifically, $l_{ij,t} \neq l_{ji,t}$. The co-leverage measures the impact of a shock on the return of asset i today on the volatility of asset j tomorrow. Under our assumptions, it follows that

$$\begin{aligned} l_{ij,t} &\equiv Cov\left(y_{i,t+1}, \Sigma_{jj,t+1} | J_t\right) = Cov\left(\lambda_i f_{t+1} + u_{i,t+1}, \lambda_j^2 \sigma_{t+1}^2 + \Omega_{jj} | J_t\right) \\ &= \lambda_i \lambda_j^2 Cov\left(f_{t+1}, \sigma_{t+1}^2 | J_t\right) \\ &\equiv \lambda_i \lambda_j^2 l_t = \frac{\lambda_j^2}{\lambda_i^2}\left(\lambda_i^3 l_t\right), \end{aligned}$$

implying that

$$l_{ij,t} = \frac{\lambda_j^2}{\lambda_i^2} l_{ii,t}. \tag{12}$$

Equation (12) shows that the co-leverage of asset i on asset j has the same sign as the leverage effect for asset i. Thus, if asset i has a negative leverage effect, a positive shock on asset i' s return lowers its future volatility, which increases the confidence level in asset i's market, which ceteris paribus, propagates to the entire financial market. Thus, a positive shock on asset i's return reduces future volatility for all other assets, including asset j.

[6]Richer dynamics in the returns leverage may be obtained by including more factors in the model.

4 Temporal aggregation results

Asset returns are available at many different frequencies. For instance, financial data are often available at the daily, weekly, or monthly level, not to mention the fact that higher frequency data at the intraday level are also increasingly available in finance. Because lower frequency returns are just a temporal aggregation of the higher frequency returns, an internally consistent model should be robust to temporal aggregation. Drost and Nijman (1993) show that the standard GARCH model is not robust to temporal aggregation and propose the weak GARCH model, which is robust to temporal aggregation. More recently, Meddahi and Renault (2004) propose the SR-SARV class of volatility processes and show that these processes are closed under temporal aggregation. See also Engle and Patton (2001) for a discussion of the merits of temporal aggregation.

In this section, we show that the conditionally heteroskedastic factor model with asymmetries we propose in this paper is robust to temporal aggregation.

Suppose we observe returns at $t = 1, 2, \ldots$. The relevant conditioning information set at time t is J_t, which contains the past observations dated at times t and before. Suppose now we observe returns at a lower frequency, in particular we observe returns at tm intervals, where $t = 1, 2, \ldots$, and m is the time horizon. For example, if we move from the daily to the weekly frequency, $m = 5$. In this case, the relevant conditioning information set depends on the observations dated at times tm. We will call this information set $J_{tm}^{(m)}$. In order to define $J_{tm}^{(m)}$, we need to introduce some additional notation. In particular, following Meddahi and Renault (2004), let $Y_{tm}^{(m)} \equiv \sum_{l=1}^{m} \alpha_l Y_{(t-1)m+l}$, $t \geq 1$, denote the process resulting from the temporal aggregation of Y_t over the time horizon m. The coefficients α_l, $l = 1, \ldots, m$, are the aggregation coefficients. For a flow variable such as a log return, $\alpha_l = 1$, for all $l = 1, \ldots, m$, whereas for a stock variable we have that $\alpha_l = 1$ for $l = m$ and 0 otherwise. Similarly, let $F_{tm}^{(m)} \equiv \sum_{l=1}^{m} \alpha_l F_{(t-1)m+l}$ and $U_{tm}^{(m)} \equiv \sum_{l=1}^{m} \alpha_l U_{(t-1)m+l}$ be the temporal aggregation analogues of F_t and U_t. Following Meddahi and Renault (2004), we define $J_{tm}^{(m)} \equiv \sigma\left(Y_{\tau m}^{(m)}, F_{\tau m}^{(m)}, U_{\tau m}^{(m)}, D_{\tau m}; \tau \leq t\right)$, where, for any integer τ, $D_{\tau m} = Var\left(F_{\tau m+1} | J_{\tau m}\right)$, $J_{\tau m}$ is the same information set as J_t with $t = \tau m$ and $\sigma(X)$ denotes the σ-algebra generated by X. Meddahi and Renault (2004) show that the SR-SARV(1) model is robust to temporal aggregation with respect to the increasing filtration $J_{tm}^{(m)}$.

Proposition 4.1 *Let Y_t be defined by (1) and (2). Assume Y_t has a constant conditional mean*

μ. Then the temporally aggregated process $Y_{tm}^{(m)}$ of Y_t over the time horizon m has the following representation

$$Y_{tm}^{(m)} = \mu^{(m)} + \Lambda F_{tm}^{(m)} + U_{tm}^{(m)}, \tag{13}$$

such that

$$\begin{aligned}
E\left(U_{(t+1)m}^{(m)} | J_{tm}^{(m)}\right) &= 0 \\
E\left(F_{(t+1)m}^{(m)} | J_{tm}^{(m)}\right) &= 0 \\
E\left(U_{(t+1)m}^{(m)} F_{(t+1)m}^{(m)\prime} | J_{tm}^{(m)}\right) &= 0 \\
Var\left(U_{(t+1)m}^{(m)} | J_{tm}^{(m)}\right) &= \left(\sum_{l=1}^{m} \alpha_l^2\right) \Omega \\
Var\left(F_{(t+1)m}^{(m)} | J_{tm}^{(m)}\right) &= D_{tm}^{(m)},
\end{aligned} \tag{14}$$

where $D_{tm}^{(m)}$ is a diagonal matrix and $\mu^{(m)}$ is a time-invariant vector equal to $\left(\sum_{l=1}^{m} \alpha_l\right) \mu$.

Proposition 4.1 shows that $Y_{tm}^{(m)}$, the temporal aggregation of Y_t over the horizon m, has the same factor representation as Y_t, where the idiosyncratic shocks and the latent factors are the temporal aggregation analogues of the higher frequency idiosyncratic shocks and factors, respectively. Hence, if we assume that each factor, component of F_t, follows a SR-SARV(1) model, as in our model in (11) for the single factor, the results in Meddahi and Renault (2004) imply that each component of $F_{tm}^{(m)}$ inherits the SR-SARV(1) dynamics. We can therefore conclude that under our assumptions the volatility specification assumed for the factor representation of Y_t in our model in (11) is robust to temporal aggregation.

Next we study the properties of temporal aggregation of the models assumed for the leverage and skewness effects (Assumptions 2 and 5, respectively). For simplicity we assume a single factor model and let $\mu = 0$.

The following proposition is auxiliary in proving the robustness of the skewness and leverage models to temporal aggregation. It provides some useful properties of the SR-SARV(1) process not yet established in the literature. In particular, this proposition gives the expected value of the conditional variance conditional on the information available at any earlier period, and it also expresses the conditional variance of an aggregated SR-SARV(1) process in terms of the conditional variance of the original process.

PROPOSITION 4.2 *Let f_{t+1} follow a SR-SARV(1) model with volatility persistence and intercept parameter γ and $1-\gamma$, respectively, and with conditional variance σ_t^2. Then, for all $l \geq 1$, we have* *that*

$$E(\sigma^2_{tm+l-1}|J_{tm}) = 1 - \gamma^{l-1} + \gamma^{l-1}\sigma^2_{tm},$$

and

$$\sigma^{(m)^2}_{tm} \equiv Var(f^{(m)}_{(t+1)m}|J^{(m)}_{tm}) = \sum_{l=1}^{m}\alpha_l^2\left(1-\gamma^{l-1}\right) + \sigma^2_{tm}\sum_{l=1}^{m}\alpha_l^2\gamma^{l-1} \equiv S^{(m)}_1 + S^{(m)}_2\sigma^2_{tm}.$$

The leverage effect in the aggregated return $Y^{(m)}_{i,tm}$ of asset i is defined as

$$Cov\left(Y^{(m)}_{i,(t+1)m}, \sigma^{(m)^2}_{i,(t+1)m}|J^{(m)}_{tm}\right),$$

where $\sigma^{(m)^2}_{i,tm} \equiv Var\left(Y^{(m)}_{i,(t+1)m}|J^{(m)}_{tm}\right)$.

Given Assumption 1, it suffices to examine the leverage effect in the factor

$$Cov\left(f^{(m)}_{(t+1)m}, \sigma^{(m)^2}_{(t+1)m}|J^{(m)}_{tm}\right).$$

Proposition 4.3 *Let f_{t+1} follow a SR-SARV(1) model with volatility persistence and intercept parameter γ and $1-\gamma$, respectively, and satisfying Assumption 2. It follows that*

$$Cov\left(f^{(m)}_{(t+1)m}, \sigma^{(m)^2}_{(t+1)m}|J^{(m)}_{tm}\right) = \pi^{(m)}_0 + \pi^{(m)}_1\sigma^{(m)^2}_{tm}; \qquad \pi^{(m)}_0 \text{ and } \pi^{(m)}_1 \in \mathbb{R}.$$

Proposition 4.3 shows that the leverage model assumed in Assumption 2 is robust to temporal aggregation for the class of SR-SARV(1) processes. Similarly, we can show that the equation (12) describing the co-leverage effect for asset i on asset j holds for the aggregated process provided Assumption 1 is satisfied.

The next result establishes the robustness to temporal aggregation of the third order conditional moment dynamics assumed in Assumption 5.

Proposition 4.4 *Let f_{t+1} follow a SR-SARV(1) model with volatility persistence and intercept parameter γ and $1-\gamma$, respectively, and satisfying Assumptions 2 and 5. It follows that*

$$E\left[\left(f^{(m)}_{(t+1)m}\right)^3|J^{(m)}_{tm}\right] = h^{(m)}_1\sigma^{(m)^2}_{tm} + h^{(m)}_0,$$

and

$$E\left[\left(Y^{(m)}_{i,(t+1)m}\right)^3|J^{(m)}_{tm}\right] = \lambda_i^3 h^{(m)}_1\sigma^{(m)^2}_{tm} + s^{(m)}_i,$$

for $i = 1, \ldots, N$ and $t = 1, 2, \ldots,$ where

$$h_1^{(m)} = \left[h_1 \sum_{l=1}^{m} \alpha_l^3 \gamma^{l-1} + 3\pi_1 \times \sum_{l<l';l,l'=1}^{m} \alpha_l \alpha_{l'}^2 \gamma^{l'-2} \right] / S_2^{(m)},$$

$$h_0^{(m)} = A^{(m)} - h_1^{(m)} S_1^{(m)},$$

$$A^{(m)} = \sum_{l=1}^{m} \alpha_l^3 [h_0 + (1-\gamma^{l-1}) h_1] + 3 \times \sum_{l<l';l,l'=1}^{m} \alpha_l \alpha_{l'}^2 \gamma^{l'-l-1} \left[\pi_0 + \pi_1 (1-\gamma^{l-1}) \right],$$

$$s_i^{(m)} = \lambda_i^3 h_0^{(m)} + s_i^0 \left(\sum_{l=1}^{m} \alpha_l^3 \right),$$

and where $S_1^{(m)}$ and $S_2^{(m)}$ are as defined in Proposition 4.2.

Proposition 4.4 shows that the conditional third order moment dynamics postulated in Assumption 5 is robust to temporal aggregation in the set of SR-SARV(1) processes that have a conditional leverage dynamic according to Assumption 2. In particular, the third conditional moments of excess aggregated returns follow an affine function of volatility. Moreover, if the conditional third moment of the underlying factor is time varying, it follows that the aggregated factor also has a dynamic conditional third moment, given that the aggregation coefficients α_l, $l = 1, \ldots, m$ are nonnegative.

We can summarize the temporal aggregation properties of the temporally aggregated model as follows. First, the factor representation is preserved for the aggregated model, with the same factor loadings. Second, the aggregated factor has a leverage effect and a skewness effect whose specifications are affine functions of its volatility, just as assumed for the original factor itself. Third, the conditional skewness of the idiosyncratic shocks is constant if the same is true for the underlying non aggregated shocks, as assumed by Assumption 1. These properties, together with the property of robustness to temporal aggregation of SR-SARV models for conditional heteroskedasticity established by Meddahi and Renault (2004), prove that our conditionally heteroskedastic factor model with asymmetries is robust to temporal aggregation.

5 Identification and estimation of the model

The main goal of this section is to present some valid moment conditions for our model on which we can base GMM inference. The GMM-based inference is robust to distribution misspecification and it is also easier to perform than alternative methods often used in the conditionally heteroskedastic factor model literature, which also widely rely on distributional assumptions. To the best of our knowledge, DR (2006) are the first to propose a GMM-based inference method for conditionally heteroskedastic factor model.

Our model, as we already mentioned, is close to the DR (2006) model in that both share Equations (11a)-(11g). Therefore, the DR (2006) moment conditions are useful here to identify λ, γ and Ω. However, as pointed out by DR (2006), because the factor is not observable their model is not able to identify all of the parameters included. The main reason is that their model specifies only the first and second conditional moments which are to be estimated as well. Actually, the model is only partially identified in the sense of Manski and Tamer (2002). Particularly, it identifies the whole set of parameters as a function of one factor loading. To conduce inferences in the usual way, DR (2006) propose two approaches. The first normalizes one factor loading and thus allows the identification of the whole model by appropriate unconditional moment conditions. The second approach restricts the factor's conditional kurtosis to be time invariant and proposes a dynamic model for the conditional variance of the factor's conditional variance. This allows the full identification of the model through suitable moment conditions.

Here we will follow both of these approaches to study the identification of our extended model (Equations (11a) through (11i)).

5.1 Inference by normalization

This approach sets the factor loading of a given excess return process (we will consider one with a factor loading different from 0) to an arbitrary value $\underline{\lambda}$. In particular, we let $0 < \underline{\lambda}^2 < Var(Y_{1,t})$. For simplicity, we first consider the case in which $\mu = 0$. The following moment conditions proposed by DR (2006) characterize $\lambda_{-1} = (\lambda_2, \cdots, \lambda_N)$, γ and Ω.

$$Vec\left\{E\left[\left(Y_{-1,t+1} - \lambda_{-1}\underline{\lambda}^{-1}Y_{1,t+1}\right) Y'_{t+1}|J_t\right]\right\} = Vec\left[\Omega_{2.} - \lambda_{-1}\underline{\lambda}^{-1}\Omega_{1.}\right] \tag{15a}$$

$$Vech\left\{E\left[(1-\gamma L)Y_{t+1}Y'_{t+1}\right| J_{t-1}]\right\} = Vech\left[(1-\gamma)(\lambda\lambda' + \Omega)\right] \tag{15b}$$

where L is the usual lag operator, $Y_{-1,t+1} = (Y_{2,t+1}, \ldots, Y_{N,t+1})$, $\Omega_{1.}$ is the $1 \times N$-matrix equal to the first row of Ω, $\Omega_{2.}$ is the $N-1 \times N$-matrix of the last $N-1$ rows of Ω, while Vec and $Vech$ are the usual vectorizing and half-vectorizing operators.

In practice, instead of (15a), we can use

$$Vec\left\{E\left[\left(Y_{-1,t+1}-\lambda_{-1}\underline{\lambda}^{-1}Y_{1,t+1}\right)Y_{1,t+1}|J_t\right]\right\}=Vec\left[\Omega_{2:N,1}-\lambda_{-1}\underline{\lambda}^{-1}\Omega_{11}\right],$$

where $\Omega_{2:N,1}$ is the submatrix of Ω defined by its first column and its second to its last rows, to avoid the risk of near collinearity in the resulting unconditional moment restrictions. Similarly, instead of the Vech operator in (15b), the Diag operator is recommended when large dimension processes are used. The Diag operator transforms a square matrix into a vector of its diagonal entries.

Next we extend this approach to the leverage and skewness dynamics.

The leverage in excess returns as given by (8) provides moment conditions characterizing π_0 and π_1. With $\Sigma_{ii,t}=E(Y_{i,t+1}^2|J_t)$ and $\sigma_t^2=\underline{\lambda}^{-2}\left[E\left(Y_{1,t+1}^2|J_t\right)-\Omega_{11}\right]$, these moment conditions are:

$$E\left(Y_{i,t+1}Y_{i,t+2}^2|J_t\right)=\lambda_i^3\left[\pi_0^0+\pi_1^0E\left(Y_{1,t+1}^2|J_t\right)\right],\quad \forall i=1,\cdots,N \tag{16}$$

where $\pi_0^0=\pi_0-\pi_1(\Omega_{11}/\underline{\lambda}^2)$ and $\pi_1^0=\pi_1/\underline{\lambda}^2$.

Similarly, the third conditional moment of asset excess returns as given by (10) characterizes the skewness parameters h_1, $s_1,\cdots,s_N$:

$$E\left(Y_{i,t+1}^3|J_t\right)=\lambda_i^3h_1^0E\left(Y_{1,t+1}^2|J_t\right)+s_i^1,\quad \forall i=1,\cdots,N \tag{17}$$

where $h_1^0=h_1\underline{\lambda}^{-2}$ and $s_i^1=s_i-h_1\lambda_i^3\underline{\lambda}^{-2}\Omega_{11}$.

In a more general case where $\mu\neq 0$, one can note that the moment condition $E(Y_{t+1}|J_t)=\mu$ identifies μ and therefore has to be completed in (15) in addition to Equations (15a) and (15b). Moreover Y_{t+1}, $Y_{-1,t+1}$ and $Y_{1,t+1}$ have to be replaced by $Y_{t+1}-\mu$, $Y_{-1,t+1}-\mu_{-1}$ and $Y_{1,t+1}-\mu_1$, respectively.

The moment conditions in Equations (16)-(17) can be written as

$$E\left\{g_{\phi_1}(Y_{t+1},Y_{t+2},\phi_2|J_t)\right\}=0$$

with $\phi_1=(\lambda_{-1}',\gamma,Vech(\Omega)')'$ and $\phi_2=(\pi_0,\pi_1,s_1,\cdots,s_N,h_1)$, where $g_{\phi_1}(.)$ defines a smooth function. For a J_t-measurable vector of instruments z_t including 1, this conditional moment restrictions implies:

$$E\{z_t\otimes g_{\phi_1}(Y_{t+1},Y_{t+2},\phi_2)\}=0 \tag{18}$$

which, in turn, is an unconditional moment restrictions allowing for the application of Hansen's (1982) results. The following proposition suggests an instrument that could validate the application of Hansen's (1982) results.

Proposition 5.1 *Let z_t be an I-vector of J_t-measurable instruments including 1 and $z_{1,t}$ $(I \geq 2)$ such that $Cov\left(z_{1,t}, \sigma_t^2\right) \neq 0$. Then, for any ϕ_1, the moment condition in* (18) *identifies ϕ_2 at the first order i.e.:*

$$(\partial/\partial\phi_2')\left[E\{z_t \otimes g_{\phi_1}(Y_{t+1}, Y_{t+2}, \phi_2)\}\right] \text{ has full column rank.}$$

The factor f_{t+1} follows an SR-SARV(1) model and therefore its square has an ARMA representation (Meddahi and Renault (2004)). Because $\underline{\lambda} \neq 0$, $Y_{1,t}^2$ has also an ARMA representation. Thus, $Y_{1,t}^2$ is correlated with σ_t^2. Consequently, any lag of $Y_{1,t}^2$ is a valid instrument that may help to first-order identify ϕ_2 by (18).

Because the moment conditions in (15) identify ϕ_1 at the first order (see DR (2006)) and do not share any component of ϕ_2, we can show that, jointly, the moment conditions in (15) and (18) identify $\phi = (\phi_1', \phi_2')'$ at the first order. Combining (15) and (18), the global moment condition is written as

$$E\{g_t(\phi)\} = 0, \tag{19}$$

for an appriate smooth function $g_t(.)$.

Let $\|X\|^2 = X'X$, where X is a vector. If $E\|g_t(\phi)\|^2 < \infty$ (which inquires Y_t to have a finite sixth moment), we can apply the results in Hansen (1982) to show that the asymptotic distribution of $\hat{\phi}$, the efficient GMM estimator of ϕ based on the moment condition (19) is given by

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}\left(0, (D'W^{-1}D)^{-1}\right),$$

where $W = \lim_{T\to\infty} Var\left(\sum_{t=1}^{T} g_t(\phi)/\sqrt{T}\right)$ and $D = (\partial/\partial\phi')E\{g_t(\phi)\}$.

5.2 Inference through higher order moments

The second approach proposed by DR (2006) completes their model given by equations (11a)-(11g) by using a dynamic specification for the conditional variance of the factor's conditional variance. With the additional constant conditional kurtosis assumption for the factor, DR (2006) propose moment conditions that identify λ, γ and Ω.

Specifically, DR (2006) assume that the conditional variance of σ_{t+1}^2 is a quadratic function of σ_t^2,

$$Var\left(\sigma_{t+1}^2 | J_t\right) = \alpha + \beta\sigma_t^2 + \delta\sigma_t^4. \tag{20}$$

This specification nests the affine process of the conditional variance of Heston (1993) and the Ornstein-Uhlenbeck-like Levy-process of the conditional variance, as introduced by Barndorff-Nielsen and Shephard (2001).

In our temporal aggregation robust framework, if we complete our conditionally heteroskedastic factor model with asymmetries given in (11) with the conditional variance specification in (20) for the factor, it turns out that the whole framework is modified and we must re-evaluate the temporal aggregation property of this new model. Proposition C.1 in Appendix C insures that the conditional variance dynamic in (20) is robust to temporal aggregation in the class of the SR-SARV(1) processes. As a consequence, the new model we obtain by completing our model in (11) by (20) is also robust to temporal aggregation.

The following moment conditions identify λ, Ω, γ, a, b and c at the first order:

$$Vec\left\{E\left[\left(Y_{-1,t+1} - \lambda_{-1}\lambda_1^{-1}Y_{1,t+1}\right)Y_{t+1}' | J_t\right]\right\} = Vec\left[\Omega_{2.} - \lambda_{-1}\lambda_1^{-1}\Omega_{1.}\right] \tag{21a}$$

$$Vech\left\{E\left[(1-\gamma L)Y_{t+1}Y_{t+1}' \middle| J_{t-1}\right]\right\} = Vech\left[(1-\gamma)(\lambda\lambda' + \Omega)\right] \tag{21b}$$

$$E\left[(1-cL)(y_{1,t+1}^4 - 6\Omega_{11}y_{1,t}^2) - b\lambda_1^2 y_{1,t}^2 \middle| J_{t-1}\right] = a \tag{21c}$$

where λ_{-1} and $Y_{-1,t}$ have the same definition as in (15), a, b and b three additional parameters and we also assume $\mu = 0$.

Let ϕ_1 denote a vector containing λ, Ω, γ, a, b and c. We can easily verify that Proposition 5.1 holds in this framework so that the parameters contained in ϕ_2 are identified at the first order for any ϕ_1. The efficient GMM estimator of $\phi = (\phi_1', \phi_2')'$ is also asymptotically normal, although this result requires more stringent moment conditions (in particular, Y_t should have eight order finite moments). This conditions may be less realistic for GARCH processes with high volatility persistence. For this reason, the previous approach for identification seems more practical.

We conclude this section with some remarks. First, the number of parameters can grow very quickly with the number of assets in the factor structure depending on the restrictions on the variance matrix

Ω of the idiosyncratic shocks. Typically, without any restriction on Ω, the number of parameters is of order $O(N^2)$ against $O(N)$ if we restrict this variance matrix to be diagonal. Thus, a free semi positive definite matrix Ω is practically tractable only in the case of a reasonable number of assets (e.g. $N \leq 5$). But, for a larger number of assets, it would be more convenient to restrict Ω to be diagonal or even block diagonal.

Second, the Iterated GMM inference approach proposed by Ogaki (1993), is also useful in our framework. Applied to our context, the Iterated GMM consists on estimating ϕ_1 from (15) or (21) by the usual GMM technique, and then plugging $\hat{\phi}_1$ in the remaining moment conditions to estimate ϕ_2 also by GMM. It has been proved (see Ogaki 1993) that the resulting estimator $(\hat{\phi}_1', \hat{\phi}_2')'$ is asymptotically normally distributed. Even though this approach involves more optimization steps than the usual GMM, the dimension of the parameter spaces on which these optimizations are performed are smaller than in the usual GMM. Therefore, this technique may be easier to implement. However, due to its two-step approach, the Iterated GMM is less efficient than the usual GMM.

6 Monte Carlo results

The main goal of this section is to assess the finite sample performance of our estimation procedure for different values of the factor volatility persistence. Because the GMM inference results are known to be sensitive to the set of valid instruments that are used (see e.g. Andersen and Sørensen (1996)), we first investigate the relative performance of four sets of valid instruments. We evaluate the performance of each instrument set by the simulated bias and the root mean square error (RMSE) of parameter estimates it provides for both the DR (2006) model and our conditionally heteroskedastic factor model with asymmetries. The best of these instrument sets is subsequently used in our experiments for assessing the sensitivity of our inference procedure to the factor volatility persistence.

We simulate samples of three asset excess returns with null risk premia ($\mu = 0$) from a single factor model. The model considered is the following

$$Y_{i,t} = \lambda f_t + u_{i,t}, \qquad i = 1, 2, 3,$$

with $\lambda = (1,1,1)'$ and $U_t \sim i.i.d.\mathcal{N}(0, \omega I_3)$, where $\omega = 0.35$ and I_3 is the identity matrix of size 3. $u_{i,t}$ is the i-th component of U_t. In this model, the signal to noise ratio λ_i/ω is 2.86 which roughly matches the average signal to noise ratio we find in our empirical application in Section 7. The factor process f_t in all our experiments has a GARCH(1,1) dynamics i.e. $f_t = \sigma_{t-1}\eta_t$, $\eta_t | J_{t-1} \sim (0, 1)$ and $\sigma_t^2 = 1 - \alpha - \beta + \alpha f_t^2 + \beta\sigma_{t-1}^2$, $0 < \alpha + \beta < 1$. This GARCH(1,1) process is an SR-SARV(1) process with persistence parameter $\gamma = \alpha + \beta$.

In the DGP 1, DGP 2, DGP 3 and DGP 4 we consider, $\eta_t \sim i.i.d.\mathcal{N}(0, 1)$ and hence, the factor is the standard Gaussian GARCH(1,1) process.

In the DGP 1', DGP 2', DGP 3' and DGP 4', $\eta_t = (\sigma_{t-1}^2 - X_t)/\sigma_{t-1}$ where $X_t | J_{t-1} \sim \text{Gamma}(\sigma_{t-1}^2, 1)$. In this case, $E(f_{t+1}^3 | J_t) = -2\sigma_t^2$ and $Cov(f_{t+1}, \sigma_{t+1}^2 | J_t) = -2\alpha\sigma_t^2$. This simulation design fits with the occurence of conditional skewness and conditional leverage in $Y_{i,t}$.

The eight experiments we conduct differ by the asymmetries occurence and also by the factor volatility persistence, γ. For

DGP 1 and DGP 1': $\alpha = 0.20$, $\beta = 0.50$; $\gamma = 0.70$,

DGP 2 and DGP 2': $\alpha = 0.20$, $\beta = 0.60$; $\gamma = 0.80$,

DGP 3 and DGP 3': $\alpha = 0.20$, $\beta = 0.70$; $\gamma = 0.90$,

DGP 4 and DGP 4': $\alpha = 0.20$, $\beta = 0.75$; $\gamma = 0.95$.

$\gamma = 0.70$ roughly matches the volatility persistence we get for the factor in our empirical application. $\gamma = 0.80$ matches approximately the factor volatility persistence estimate by Fiorentini, Sentana and Shephard (2004) for monthly U.K. index excess returns. $\gamma = 0.90$ and $\gamma = 0.95$ are the usual range of the standard GARCH volatility persistence estimate in the empirical literature for daily returns (see e.g. Harvey and Siddique (1999)). We set the number of replications to 500, and the sample size is $T = 5000$, which roughly matches the length of the data set used in our empirical application.

We perform the inference by the normalization approach described in Section 5 and we set the first asset factor loading to $\underline{\lambda} = 1$. We estimate the DR (2006) model by the moment conditions given in equation (15) and our conditionally heteroskedastic factor model with asymmetries by the moment conditions given by equations (15)-(16)-(17). The parameters of interest in the DR (2006) model are λ_2, λ_3, ω and γ. In the DGP 1 through 4, $s_1 = s_2 = s_3 = s = 0$, $h_1 = 0$ and $\pi_0 = \pi_1 = 0$ and in the DGP 1' through 4', $s_1 = s_2 = s_3 = s = 0$, $h_1 = -2.0$, $\pi_0 = 0$ and $\pi_1 = -0.4$. Therefore, as far as the Monte Carlo designs are concerned, λ_2, λ_3, ω, γ, s, h_1, π_0 and π_1 are the only relevant parameters of our conditionally heteroskedastic factor model with asymmetries.

We first assess the relative performance of our estimation method for the DR (2006) model and our extended model in terms of instruments used. In particular, we consider four sets of instruments: $z_{1,t} = (1, Y_{1,t}^2)$, $z_{2,t} = (1, \sum_{i=1}^{20} \rho^i Y_{1,t-i+1}^2; \rho = 0.9)$, $z_{3,t} = (1, Y_{1,t}^2, Y_{1,t-1}^2)$ and $z_{4,t} = (1, Y_{1,t}^2, Y_{1,t-1}^2, Y_{1,t-2}^2)$. For each set of instruments, we simulate data from DGP 2 and evaluate the bias and the RMSE for the parameter estimates in the DR (2006) model and in our model with asymmetries. Table 1.5 contains the results. In terms of bias, $z_{3,t}$ is the most desirable for both models among the four instrument sets. It yields a particularly small average bias for the DR (2006) model estimates with respect to the other instruments which show roughly the same amount of average bias. For our conditionally heteroskedastic factor model with asymmetries estimates, $z_{3,t}$ also yields the smallest amount of average bias but the difference with $z_{4,t}$ is not noticeable. $z_{1,t}$ appears to yield the largest average bias among the four instrument sets that we compare. In terms of RMSE, $z_{4,t}$ is the best instrument set for both models followed by $z_{3,t}$. These two instrument sets perform much better than $z_{2,t}$ and $z_{1,t}$. This result suggests the use of $z_{4,t}$ as an instrument set for our next experiments and our empirical work.

Next, we investigate the finite sample properties of our estimation method as a function of the volatility persistence γ and the occurence of asymmetries. More specifically, we consider the eight DGPs described above (DGP 1, 2, 3, 4 and DGP 1', 2', 3', 4'). Table 1.6 presents the results for the DGP 1, 2, 3 and 4 in which there is no conditional asymmetry in the processes $Y_{i,t}$ and Table 1.7 presents the results for DGP 1', 2', 3' and 4' in which both the conditional skewness and the conditional leverage effects occur.

In both models, the estimates of the parameters λ_2, λ_3, γ and ω exhibit low bias for all of the DGP. These estimates also exhibit similar RMSE for the DGP 1, 2, 3 and 4. Note however that the RMSEs obtained for our model are slightly higher than those obtained by the DR (2006) model. These observations remain valid for the DGP 1', 2',and 3'. For the DGP 4', the RMSEs of the estimates of λ_2, λ_3 γ and ω by our model are much larger than those yield by the DR (2006) model estimates.

When there is no asymmetry in the data and the factor volatility persistence is not too large (DGP 1, 2 and 3), the estimates of s, h_1 π_0 and π_1 exhibit small bias but their RMSE seem to increase with the factor the factor volatility persistence. In the DGP 1', 2' and 3', the estimates of s and π_0 still yield small bias while the estimates of h_1 and π_1 exhibit larger bias. The RMSEs in these DGPs also seem to increase with the volatility persistence of the factor.

The largest bias and RMSEs occur for DGP 4 and 4' where $\gamma = 0.95$ and for our model estimates. This lack of performance of our estimation method could be viewed as a consequence of the near-integration of the volatility process. Because the parameters h_1 and π_1 are both coefficients of volatility in our model, they may not be efficiently estimated. The other parameters are therefore contaminated and are not efficiently estimated.

This Monte Carlo experiment suggests that our inference procedure is reliable in finite samples particularly when the volatility persistence is not close to 1, whereas the inference could be inaccurate for persistence values larger than 0.95. This observation seems to confirm a well known drawback of the GMM method application in volatility literature which delivers bad results when the volatility persistence is close to 1 (see e.g. Broto and Ruiz (2004)).

7 Application to daily U.K. stock market excess returns

In our empirical work, we estimate both the conditionally heteroskedastic factor model proposed by DR (2006) and our conditionally heteroskedastic factor model with asymmetries for stock excess returns on 24 U.K. sectors. We use all the series described in the Data Appendix (see Appendix B) except for the FTSE All Share Ex. Inv. Trusts index excess return because it exhibits a correlation of 0.99 with the FTSE 350 index excess return (see Table 1.1 Appendix B) which is included in the models.

In both models, we consider centered excess returns and we do not estimate the conditional mean, μ. We also restrict the variance matrix Ω of the idiosyncratic shocks to be diagonal. with these restrictions, the DR (2006) model has 48 parameters while our extended model has 75 parameters. In the estimation procedure, the FTSE 350 index excess return plays the role of $Y_{1,t}$ and we use $z_{4,t}$ as a valid instrument set (see Section 5).

7.1 Results

Table 1.7 in Appendix B presents our estimation results. Even though the factor loadings and the variance of the idiosyncratic shocks estimates for both models differ according to the specific return series, their average difference across the returns are 0.0015 and 0.0090 for the factor loadings and the idiosyncratic shock variances, respectively.

The factor volatility persistence estimates for the two models are 0.685 for the DR (2006) model and 0.684 for our model. Thus, explicitly modeling the asymmetries does not seem to change the persistence estimate. Nevertheless, the persistence we find is low with respect to what is found in the empirical literature for daily data. Engle and Ng (1993) find a volatility persistence of 0.916 in a standard GARCH(1,1) model for the daily return of the TOPIX index, and both Harvey, Ruiz and Sentana (1992) and Fiorentini, Sentana and Shephard (2004) find a persistence level of about 0.80 for the volatility in their QGARCH(1,1) factor models for monthly U.K. stock market index returns. The volatility of monthly returns is known to be less persistent than the volatility of daily returns. The main difference between these models and the models we estimate in this paper is the way in which they treat the conditional skewness. The Harvey, Ruiz and Sentana (1992) and Fiorentini, Sentana and Shephard (2004) models rule out the conditional skewness in the data, while the Doz and Renault (2006) model and our model are consistent with this empirical fact. Our data seem to confirm the findings by Harvey and Siddique (1999), who observe that taking account the skewness impacts the persistence in the conditional variance. Our findings also suggest that not being consistent with the presence of conditional skewness could change the inference about the conditional variance persistence.

The leverage effect parameters in our asymmetric model, π_0 and π_1 are both significant, implying that the leverage effect in our data is time varying and can be captured by the dynamics in the factor. Our findings confirm, for daily data, the result by Sentana (1995) for monthly U.K. index excess returns, namely that there is significant leverage effect in sectorial returns through a common factor. See also Black (1976) and Nelson (1991). The slope h_1 of the factor's third conditional moment is significant and negative (-6.36). This confirms a dynamic conditional skewness in our series as Harvey and Siddique (1999, 2000) and Jondeau and Rockinger (2003) have also found for their data. Moreover, due to the positive estimates we get for the factor loadings, the negativity of h_1 implies that periods of high volatility are followed by higher negative conditional third moments. Hence, a larger volatility seems to announce more negatively skewed conditional distributions for the returns series that we consider.

7.2 Diagnostic tests

We propose an extended Kalman filter to filter simultaneously the latent factor and the conditional variance processes using the GMM estimates of both the DR (2006) model and of our model (see Appendix A). This algorithm circumvents the GMM procedure drawback, which does not yield an estimate for the factor process nor for the conditional variance process. This filter allows us to perform some useful diagnostic tests.

Figure 1.2 shows the QQ plots of the FTSE 350 index excess return and the filtered latent factor using the DR (2006) model parameters estimates. This filtered factor is multiplied by the FTSE 350 factor loading $\underline{\lambda}$=.35 to allow for direct comparison with the FTSE 350. It appears that the filtered factor shows the same fat tail and asymmetry behaviour as the FTSE 350 and thus validates the choice of an asymmetric factor representation in our model. The correlation with the FTSE 350 index excess return of the factor processes extracted with the DR (2006) model estimates and our model estimates are .932 and .911, respectively (see Table 1.9). These correlations have the same order of magnitude as the correlations obtained by Sentana (1995) for both the QGARCH(1,1) latent factor and the GARCH(1,1) latent factor with the FTA 500 monthly index excess return (.984 for both).

Figure 1.2: Q-Q plots of the FTSE 350 index excess return and the filtered factor by the DR (2006) model estimates (scaled by .35)

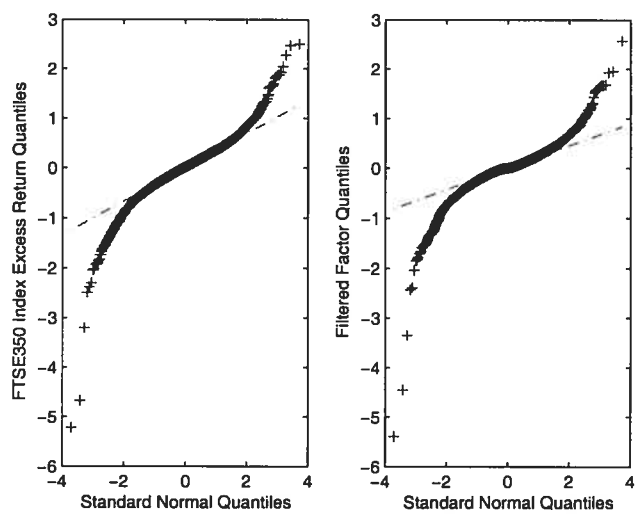
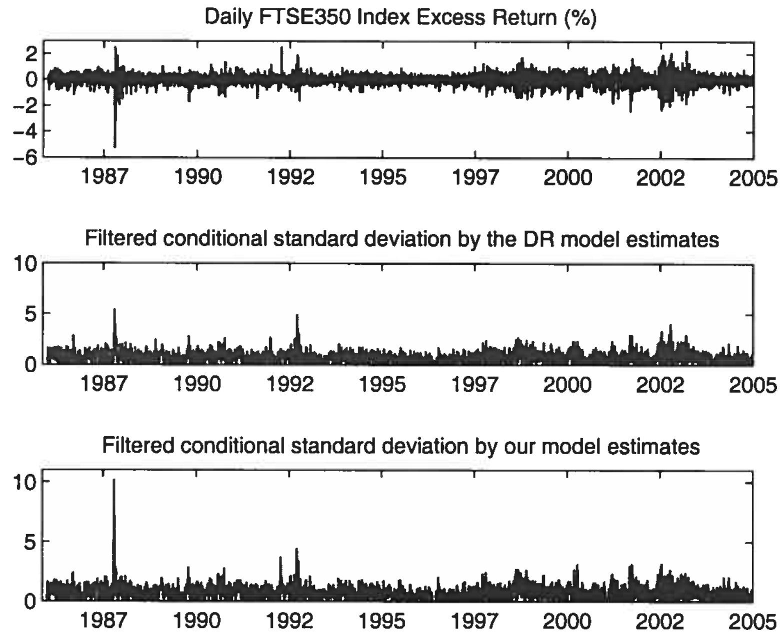


Figure 1.3 shows the FTSE 350 index excess returns and the filtered standard deviations processes

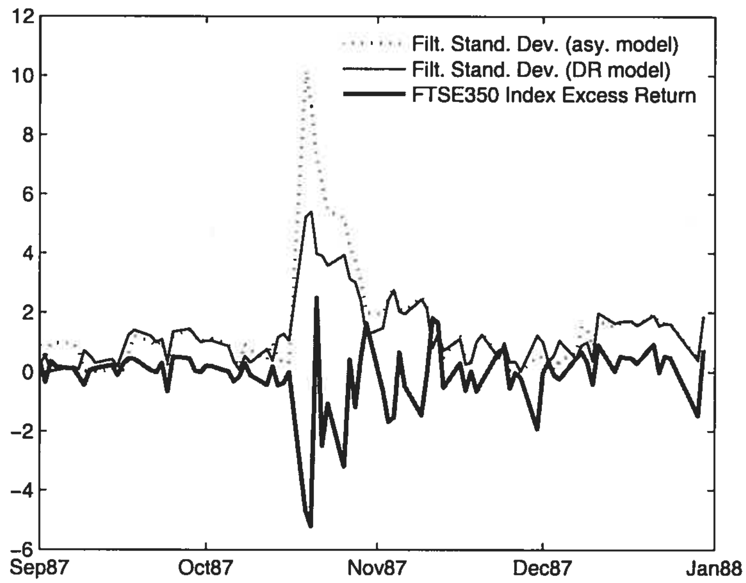
Figure 1.3: Daily FTSE 350 index excess return and filtered standard deviations by the DR (2006) model and our model estimates



with the DR (2006) model estimates and our model estimates, respectively. The volatility processes seem to adequately follow the returns series in the sense that periods of large variations in returns correspond to periods of large volatility. However, for extreme variation in returns, our asymmetric model seems to predict a larger volatility than the DR (2006) model. This is the case with the October 1987 financial crisis. For the purpose of comparison, Figure 1.4 shows the filtered standard deviation by both models estimates and the FTSE 350 index excess returns for a two-months period around October 1987. Specifically, we consider the period from September 1, 1987 through January 1, 1988. When the market is smooth, the two models predict roughly the same level of volatility. In contrast, in periods of large shocks, our model predicts larger volatility. Yet, our model may be more realistic. Between October 20, 1987 and November 6, 1987, the FTSE 350 index excess return rises by 16.61% on daily average while the in-sample forecasted standard deviation drops only by 6.89% for the DR (2006) model and by 11.13% for our model.

We also obtain two estimates for the FTSE 350 idiosyncratic shocks processes. One is obtained by the filtered factor from the DR (2006) model estimates ($u_{dr,t+1}$) and the second is from our model

Figure 1.4: Filtered standard deviations by the DR (2006) model and our model estimates and the FTSE 350 index excess return. September 1, 1987 through January 1, 1988



($u_{chfa,t+1}$). Table 1.10 shows a non significant skewness for both processes. This validates the choice of asymmetric factor in our model. The results of the Engle Lagrange multiplier test for heteroskedasticity are not clear. Even though the evidence of heteroskedasticity is not strong for these idiosyncratic shock processes, homoskedasticity is hardly accepted at 2% level for both. This may suggest the inclusion of an additional factor for heteroskedasticity and paves the way for future work.

8 Conclusion

In this paper, we extend the existing class of conditionally heteroskedastic factor models by specifying the skewness and the leverage effects dynamics in return processes. We show that our conditionally heteroskedastic factor model with asymmetries is robust to temporal aggregation. In addition, our specifications are robust to any dynamics in the conditional kurtosis or even higher moments. We also provide moment conditions allowing for GMM inference. We propose an extended Kalman filter algorithm that filters the latent factor and the volatility processes simultaneously. Our empirical application involves 24 index excess returns from U.K. stock market and confirms some useful results in the volatility model literature. In particular, our data confirm the results in Harvey and Siddique (1999, 2000) and Jondeau and Rockinger (2003) for the conditionally heteroskedastic factor model framework. We find a lower volatility persistence for our common factor than what is obtained commonly for daily data in models that rule out the skewness effect. Our findings also confirm, for daily data, the result by Sentana (1995) for monthly U.K. index excess returns, namely that there is significant leverage effect in sectorial returns through a common factor.

This work also helps to learn more about the relationship between asset returns' third conditional moment and their volatility in the presence of a leverage effect. Our empirical application suggests that it may be beneficial to incorporate this relation for efficiency gain purposes. The filtered volatility process obtained with our model estimates seems more realistic in period of large shocks than the filtered volatility obtained with the DR (2006) model estimates. Furthermore, this empirical application also suggests that larger volatility predicts more negatively skewed conditional distributions for the returns series.

The most immediate extension of this work that we plan for future work is the extension of the model we propose to more than one factor. Even though a multi factor extension is straightforward, estimation and inference are fundamental issues which need to be discussed carefully. This extension is in particular motivated by the residual heteroskedasticity that may exist in the fitted idiosyncratic shock process from our data. This work can also be extended to take into account the risk premium by modeling the conditional mean of the returns as a function of volatility as Fiorentini, Sentana and Shephard (2004) and DR (2006). As a main advantage, such an extension will make the model we propose also consistent to the well-known volatility feedback feature that occurs in financial processes. This extension may also be particularly relevant for longer horizon returns for which the risk premium is known to matter.

References

[1] Aguilar, J. F. and M. West, 2000. "Bayesian Dynamic Factor Models and Portfolio Allocation," *Review of Financial Studies*, 18, 338-357.

[2] Alami, A. and E. Renault, 2001. "Risque de Modèle de Volatilité," *working paper*, CIRANO. Ref. 2001s-06.

[3] Andersen, T. G., 1994. "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," *Mathematical Finance*, 4, 75-102.

[4] Andersen, T. G. and B. E. Sørensen, 1996. "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistic*, 14, 328-352.

[5] Ang, A. and J. Chen, 2002. "Asymmetric Correlation of Equity Portfolios," *Journal of Financial Economics*, 63, 443-494.

[6] Backus, D., S. Foresi and C. Telmer, 2001. "Affine Models of Currency Pricing: Accounting for the Forward Premium Anomaly," *Journal of Finance*, 56, 279-304.

[7] Bai, J. and S. Ng, 2005. "Tests for skewness, kurtosis, and normality for time series data," *Journal of Business and Economic Statistics*, 23, 49-61.

[8] Barndorff-Nielsen, O. E. and N. Shephard, 2001. "Non-Gaussian Ornstein-Uhlenbeck-based Models and some of their uses in Financial Economics," *J. R. Statist. Soc. B*, 63, 167-241.

[9] Barone-Adesi, G., 1985. "Arbitrage Equilibrium with Skewed Asset Returns," *Journal of Financial and Quantitative Analysis*, 20, 299-313.

[10] Black, F., 1976. "Studies of Stock Market Volatility Changes," *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

[11] Bollerslev, T., 1990. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Appraoch," *Review of Economic Studies*, 72, 498-505.

[12] Bollerslev, T. and H. Zhou, 2006. "Volatility Puzzles: A Unified Framework for Gauging Return-Volatility Regressions," *Journal of Econometrics*, 131, 123-150.

[13] Brandt, M. W. and F. X. Diebold, 2006. "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations," *Journal of Business*, 79, 61-73.

[14] Broto, C. and E. Ruiz, 2004. "Estimation Methods for Stochastic Volatility Models: A Survey," *Journal of Economic Surveys*, 18, 613-649.

[15] Campbell, J. Y., A. W. Lo and A. C. MacKinlay, 1997. "The Econometrics of Financial Markets," *Princeton University Press.*

[16] Chamberlain, G. and M. Rothschild, 1983. "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 53, 1305-1324.

[17] Christoffersen, P., S. Heston and K. Jacobs, 2006. "Option Valuation with Conditional Skewness," *Journal of Econometrics*, 131, 253-284.

[18] Cox, J. C., J. E. Ingersoll, Jr and S. A. Ross, 1985. "A theory of Term Structure of Interest Rates," *Econometrica*, 53(2), 385-408.

[19] Dai, Q. and K. Singleton, 2000. "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943-1978.

[20] Diebold, F. and M. Nerlove, 1989. "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model," *Journal of Applied Econometrics* , 4, 1-21.

[21] Doz, C. and E. Renault, 2006. "Factor Volatility in Mean Models: a GMM approach," *Econometric Review*, 25, 275-309.

[22] Drost, F. C. and T. E. Nijman, 1993. "Temporal Aggregation of GARCH processes," *Econometrica*, 61, 909-927.

[23] Duffie, D., J. Pan and K. Singleton, 2000. "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68(6), 1343-1376.

[24] Duffie, D., and K. Singleton, 1997. "An Econometric Model of the Term Structure of Interest-Rate Swap Yields," *Journal of Finance*, 52(4), 1287-1321.

[25] Engle, R. F., 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

[26] Engle, R. F., 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models," *Journal of Business and Economic Statistics*, 20, 339-350.

[27] Engle, R. F. and F. K. Kroner, 1995. "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122-150.

[28] Engle, R. F. and A. J. Patton, 2001. "What Good is a Volatility Model?," *Quantitative Finance*, 1, 237-245.

[29] Eraker, B., 2004. "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59(3), 1367-1403.

[30] Fiorentini, G., E. Sentana and N. Shephard, 2004. "Likelihood-based Estimation of Generalised ARCH Structures," *Econometrica*, 72, 1481-1517.

[31] Geweke, O. and G. Zhou, 1996. "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557-587.

[32] Glosten, L. R., R. Jaganathan and E. Runkle, 1993. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Shocks," *Journal of Finance*, 48, 1779-1801.

[33] Hansen, B. E., 1994. "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705-730.

[34] Hansen, L. P., 1982. "Large Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[35] Harvey, A., E. Ruiz and E. Sentana, 1992. "Unobserved Component Time Series Models with ARCH Disturbances," *Journal of Econometrics*, 52, 129-157.

[36] Harvey, A. C. and N. Shephard, 1996. "The Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business and Economic Statistics*, 14, 429-434.

[37] Harvey, C. R. and A. Siddique, 1999. "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465-487.

[38] Harvey, C. R., and A. Siddique, 2000. "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263-1295.

[39] Hayashi, F., 2000. "Econometrics," *Princeton University Press.*

[40] Henry, O., 1998. "Modelling the the Asymmetry of Stock Market Volatility," *Applied Financial Economics*, 8, 145-153.

[41] Heston, S. L., 1993. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6(2), 327-343.

[42] Heston, S. L. and S. Nandi, 2000. "A Closed-Form GARCH Option Pricing Model," *Review of Financial Studies*, 13(3), 585-625.

[43] Hull, J and A. White, 1987. "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42(2), 281-300.

[44] Jondeau, E. and M. Rockinger, 2003. "Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements," *Journal of Economic Dynamics and Control*, 27, 1699-1737.

[45] Kalimipalli, M. and R. Sivakumar, 2003. "Does Skewness Matter? Evidence from Index Options Market," *working paper*, University of Waterloo, .

[46] Kim, S., N. Shephard and S. Chib, 1998. "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 45, 361-393.

[47] King, M. A., E. Sentana and S. B. Wadhwani, 1994. "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901-933.

[48] Kraus, A. and R. Litzenberger, 1976. "Skewness Preference and Valuation of Risk Assets," *Journal of Finance*, 31, 1085-1100.

[49] Kroner, K. F. and V. K. Ng, 1998. "Modelling Asymmetric Comovements of Asset Returns," *Review of Financial Studies*, 11, 817-844.

[50] Mandelbrot, B., 1963. "The variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.

[51] Manski, C. F. and E. Tamer, 2002. "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

[52] Meddahi, N. and E. Renault, 1997. "Aggregations and Marginalization of GARCH and Stochastic Volatility Models," *working paper*, Université de Montréal.

[53] Meddahi, N. and E. Renault, 2004. "Temporal Aggregation of Volatility Models," *Journal of Econometrics*, 119, 355-379.

[54] Melino, A. and S. Turnbull, 1990. "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239-265.

[55] Nardari, F. and J. T. Scruggs, 2006. "Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions," *Journal of Financial and Quantitative Analysis*, Forthcoming.

[56] Nelson, D., 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

[57] Ogaki, M., 1993. "Generalized Method of Moments: Econometric Applications," *Handbook of Statistics*, 11, 455-488.

[58] Parkinson, M., 1980. "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, 53, 61-65.

[59] Pieró, A., 1999. "Skewness in Financial Returns," *Journal of Banking and Finance*, 23, 847-862.

[60] Ross, S. A., 1976. "The Arbitrage Theory of Asset Pricing," *Journal of Economic Theory*, 13, 641-660.

[61] Rubinstein, M. E., 1973. "The fundamental Theorem of Parameter-Preference Security Valuation," *Journal of Financial and Quantitative Analysis*, 8, 61-69.

[62] Ruiz, E., 1994. "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289-306.

[63] Sentana, E., 1995. "Quadratic GARCH Models," *Review of Economic Studies*, 62, 639-661.

[64] Sentana, E., G. Calzolari and G. Fiorentini, 2004. "Indirect Estimation of Conditionally Heteroscedastic Factor Models," *working paper*, CEMFI.

[65] Sentana, E. and G. Fiorentini, 2001. "Identification, Estimation and Testing of Conditionally Heteroscedastic Factor Models," *Journal of Econometrics*, 102, 143-164.

[66] Singleton, K., 2001. "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111-141.

[67] Taylor, S. J., 1986. "Modeling Financial Time Series," *John Wiley.*

A An Extended Kalman Filter for the latent common factor and volatility processes

The heteroskedastic factor model is defined by:

$$f_{t+1} = \sigma_t \epsilon_{t+1} \qquad \sigma^2_{t+1} = 1 - \gamma + \gamma\sigma^2_t + w_{t+1} \tag{22a}$$
$$Y_{t+1} = \mu + \lambda f_{t+1} + U_{t+1} \tag{22b}$$

with: $E(w_{t+1}|J_t) = E(\epsilon_{t+1}|J_t) = 0$, $E(U_{t+1}|J_t) = 0$, $E(\epsilon^2_{t+1}|J_t) = 1$, $Var(U_{t+1}|J_t) = \Omega$.

In this factor representation, both f_{t+1} and σ^2_{t+1} are unobservable; in fact, only the multivariate return process (Y_{t+1}) is observable. However, the latent bi-dimensional process $Z_t = (f_t, \sigma^2_t)'$ depend nonlinearly on its past value up to some shocks. The Extend Kalman Filter's algorithm (see Sørensen, 1985) is an attractive algorithm for this framework to filter Z_t from the observations provided that the parameters are known. The state equation is given by equations in (22a) while the measurement equation is (22b).

Still, the problem that occurs in a such procedure is the positivity of σ^2_t. A naive filter could lead to negative σ^2_t. For that reason, we propose to filter $z_t = (f_t, x_t)'$ and then, we can deduce $\sigma^2_t \equiv x^2_t$; taking advantage from the following result:

If (x_{t+1}) is such that $x_{t+1} = \sqrt{\gamma}x_t + \sqrt{1-\gamma}v_{t+1}$; $E(v_{t+1}|J_t) = 0$, $E(v^2_{t+1}|J_t) = 1$; J_t an increasing filtration as the one introduced in the body of this paper, then (x^2_{t+1}) is an SR-SARV(1) process with persistence γ and intercept $1-\gamma$ with respect to J_t.

Our state-space representation is:

$$f_{t+1} = \sqrt{x^2_t}\epsilon_{t+1} \qquad x_{t+1} = \sqrt{\gamma}x_t + \sqrt{1-\gamma^2}v_{t+1} \tag{23a}$$
$$Y_{t+1} = \mu + \lambda f_{t+1} + U_{t+1} \tag{23b}$$

With: $E(v_{t+1}|J_t) = E(\epsilon_{t+1}|J_t) = 0$, $E(U_{t+1}|J_t) = 0$, $E(\epsilon^2_{t+1}|J_t) = 1$, $E(v^2_{t+1}|J_t) = 1$ $Var(U_{t+1}|J_t) = \Omega$.

To allow for leverage, we will set $Cov(\epsilon_{t+1}, v_{t+1}|J_t) = \alpha$ where α has any negative value. In our applications, we choose $\alpha = -.5$.

For: $A = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix}$, $W_t = \begin{pmatrix} \sqrt{x^2_t} & 0 \\ 0 & \sqrt{1-\gamma^2} \end{pmatrix}$, $H = (\lambda|0)$, $Q = \begin{pmatrix} 1 & -.5 \\ -.5 & 1 \end{pmatrix}$, the Extended Kalman Filter algorithm is the following:

Initial value: $\hat{z}_0 = (0,1)'$, $P_0 = \begin{pmatrix} 1 & -\sqrt{.5(1+\gamma)} \\ -\sqrt{.5(1+\gamma)} & 1+\gamma \end{pmatrix}$,

Time Update ("Predict")	
1. *Project the state ahead:*	$z^-_t = A\hat{z}_{t-1}$
2. *Project the error covariance ahead:*	$P^-_t = AP_{t-1}A' + W_tQW'_t$
Measurement Update ("Correct")	
3. *Compute Kalman Gain:*	$K_t = P^-_t H'(HP^-_t H' + \Omega)^{-1}$
4. *Update estimate with measurement Y_t:*	$\hat{z}_t = z^-_t + K_t(Y_t - \mu - Hz^-_t)$
5. *Update the error covariance:*	$P_t = P^-_t - K_t H P^-_t$
6. $t = t+1$, Go To 1.	

In this algorithm, the parameters are considered as known. In our application, we either use the GMM parameter estimates of the Doz and Renault (2006) model (DR) which lead to the filtered process $z_{dr,t}$ or the GMM parameter estimates of our conditionally heteroskedastic factor model with asymmetries (CHFA) which lead to the filtered process $z_{as,t}$ for both the factor and volatility.

B Data Appendix and Tables

The following table presents the indices we use in this paper. The first index listed refers to the FTSE 350 index. All of 24 sectorial indices listed are in FTSE while 14 of them are in the FTSE 350. The sectorial indices which are not in FTSE 350 are the following: (-)-All Share Ex. Inv. Trusts, 13- FTSE Financials, 14-Transport, 15-Speciality & Other Finc., 16-Prsnl. Care & Hhld. Prods., 17-General Industrials, 18-General Retailers, 19-Household Goods & Text. 20-Oil & Gas, 24-Support Services.

Our Data are obtained from Datastream. With $p_{i,t}$ being the index i level at day t, we obtain the daily log-return series $r_{i,t}$ (in %) by: $r_{i,t} = 100 \times (\log p_{i,t} - \log p_{i,t-1})$. We use the log-return of the UK one month loan index JPM UK CASH 1M (r_t) as safe interest rate. The log-excess return of the index i is $Y_{i,t} = r_{i,t} - r_t$. Our daily excess returns cover the period from January 2, 1986 through December 30, 2004. Only the 4863 trading days are considered.

Number	Corresponding Sectorial index	Number	Corresponding Sectorial index
1	FTSE Actuaries 350		
2	Banks	13	FTSE Financials
3	Beverages	14	Transport
4	Cnstr. & Bldg. Mats.	15	Speciality & Other Finc.
5	Chemicals	16	Prsnl. Care & Hhld. Prods.
6	Eng. & Machinery	17	General Industrials
7	Food & Drug Retailers	18	General Retailers
8	Food Prod. & Procr	19	Household Goods & Text.
9	Insurance	20	Oil & Gas
10	Life Assurance	21	Forestry & Paper
11	Investment Companies	22	Health
12	Leisure & Hotels	23	Pharm. & Biotec
-	All Share Ex. Inv. Trusts	24	Support Services

Table 1.1: Sample correlation matrix of index excess returns.

Indices	1	2	3	4	5	6	7	8	9	10	11	12	-	13	14	15	16	17	18	19	20	21	22	23	24
1	1.0																								
2	0.8	1.0																							
3	0.6	0.5	1.0																						
4	0.7	0.5	0.4	1.0																					
5	0.7	0.5	0.4	0.5	1.0																				
6	0.6	0.5	0.4	0.6	0.6	1.0																			
7	0.6	0.5	0.5	0.4	0.4	0.4	1.0																		
8	0.7	0.5	0.6	0.5	0.5	0.5	0.5	1.0																	
9	0.7	0.6	0.4	0.5	0.5	0.5	0.4	0.5	1.0																
10	0.8	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.7	1.0															
11	0.8	0.7	0.5	0.6	0.6	0.6	0.4	0.5	0.6	0.6	1.0														
12	0.7	0.6	0.5	0.6	0.5	0.5	0.4	0.5	0.5	0.5	0.6	1.0													
-	1.0	0.8	0.6	0.7	0.7	0.6	0.6	0.7	0.7	0.8	0.8	0.7	1.0												
13	0.9	0.9	0.5	0.6	0.6	0.6	0.5	0.6	0.7	0.8	0.8	0.6	0.9	1.0											
14	0.7	0.6	0.5	0.6	0.6	0.6	0.4	0.5	0.5	0.6	0.6	0.6	0.7	0.7	1.0										
15	0.7	0.6	0.4	0.5	0.5	0.5	0.3	0.4	0.5	0.6	0.7	0.6	0.7	0.7	0.6	1.0									
16	0.4	0.3	0.4	0.3	0.3	0.3	0.3	0.4	0.3	0.3	0.3	0.3	0.4	0.4	0.3	0.2	1.0								
17	0.8	0.6	0.5	0.7	0.6	0.8	0.4	0.5	0.6	0.6	0.7	0.6	0.8	0.7	0.7	0.6	0.3	1.0							
18	0.7	0.6	0.5	0.6	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.6	0.6	0.5	0.3	0.6	1.0						
19	0.4	0.3	0.3	0.5	0.4	0.4	0.3	0.4	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.2	0.4	0.4	1.0					
20	0.6	0.5	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.5	0.4	0.6	0.5	0.4	0.4	0.3	0.4	0.4	0.2	1.0				
21	0.3	0.3	0.2	0.4	0.3	0.4	0.2	0.2	0.3	0.2	0.4	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.2	0.2	1.0			
22	0.6	0.4	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.5	0.4	0.6	0.5	0.5	0.4	0.3	0.5	0.4	0.3	0.3	0.2	1.0		
23	0.7	0.5	0.4	0.4	0.5	0.4	0.4	0.5	0.4	0.5	0.5	0.4	0.7	0.6	0.5	0.4	0.3	0.5	0.4	0.3	0.4	0.2	0.4	1.0	
24	0.7	0.6	0.4	0.6	0.5	0.6	0.4	0.5	0.5	0.5	0.7	0.6	0.7	0.7	0.6	0.6	0.3	0.6	0.5	0.4	0.4	0.3	0.5	0.5	1.0

Table 1.2: Descriptive statistics for the daily index excess returns (in %). All series start on January 2, 1986 and end on December 30, 2004. The number of observations is 4863. The sample skewness and kurtosis are given by: Skew=$\sum_{t=1}^{T}\left[(Y_{i,t}-\overline{Y_i})/\sigma_{Y_i}\right]^3/T$ and Kurt=$\sum_{t=1}^{T}\left[(Y_{i,t}-\overline{Y}_i)/\sigma_{Y_i}\right]^4/T$, respectively. $\overline{Y_i}$ is the sample mean and $\sigma^2_{Y_i}$ the sample variance of Y_i. The significance tests for the skewness are based on GMM-based asymptotic distribution of sample mean and delta method. Eng(1) and Eng(5) are Engle's (1982) Lagrange Multiplier tests statistics for conditional heteroskedasticity. 1 and 5 lags have been considered, respectively. $\hat{\rho}_1$, $\hat{\rho}_2$ and $\hat{\rho}_3$ are the first three sample autocorrelations. QW(5) and QW(10) are the Ljung -Box statistics for autocorrelation.

Sector	Mean	S.d.	Min	Max	Skew.	Kurt.	Eng(1)	Eng(5)	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	QW(5)	QW(10)
FTSE actuaries 350	0.0051	0.43	-5.22	2.50	-0.88^c	13.72	991.76^a	1108.39^a	0.06^a	-0.01	-0.03	32.51^a	54.76^a
Banks	0.0157	0.64	-6.02	4.37	-0.14	8.80	563.79^a	662.52^a	0.09^a	-0.04^c	-0.04^c	60.06^a	71.46^a
Beverages	0.0063	0.59	-6.59	5.46	-0.15	11.41	446.71^a	508.58^a	0.04^b	-0.05^b	-0.04^c	26.33^a	32.18^a
Cnstr. & Bld. Mat.	0.0031	0.50	-4.88	4.36	-0.16	11.54	593.69^a	787.08^a	0.23^a	0.09^a	0.04	324.88^a	338.57^a
Chemicals	0.0006	0.51	-5.12	2.83	-0.68^c	11.13	555.30^a	764.20^a	0.09^a	0.05^b	0.04^c	67.17^a	77.67^a
Eng. & Machinery	0.0020	0.52	-5.84	4.67	-0.46	15.25	250.18^a	581.27^a	0.20^a	0.08^a	0.04^c	259.56^a	296.05^a
Food & Drug Ret.	0.0040	0.56	-4.67	3.50	-0.24^c	6.72	369.04^a	455.57^a	0.02	0.00	-0.04	11.84^b	17.32^c
Food Prod. & Pr.	0.0046	0.47	-4.41	2.82	-0.57^c	10.78	542.73^a	644.55^a	0.06^a	0.01	-0.01	23.56^a	35.94^a
Insurance	-0.0074	0.69	-6.68	4.62	-0.53^c	12.19	207.09^a	477.98^a	0.10^a	-0.03	-0.01	52.55^a	66.64^a
Life Assurance	0.0083	0.68	-6.19	4.26	-0.28	9.08	437.14^a	675.70^a	0.05^a	0.00	-0.05^a	36.22^a	54.77^a
Investment Cies	0.0037	0.41	-5.33	2.66	-1.07^b	14.82	369.27^a	585.52^a	0.10^a	0.05^b	0.04^c	104.31^a	142.90^a
Leisure & Hotels	0.0066	0.52	-4.98	3.77	-0.33	10.52	638.59^a	851.52^a	0.10^a	0.01	0.01	63.34^a	90.46^a
All Share Ex. Inv.	0.0050	0.42	-5.12	2.46	-0.91^c	14.05	986.94^a	1116.07^a	0.06^a	0.00	-0.03	35.46^a	59.83^a
FTSE Financials	0.0093	0.51	-5.28	3.18	-0.47	11.07	559.65^a	746.96^a	0.10^a	-0.01	-0.03	66.30^a	90.66^a
Transport	0.0018	0.44	-4.96	3.29	-0.60^c	11.46	445.47^a	615.20^a	0.13^a	0.03	0.05^a	117.40^a	133.16^a
Speciality & Fin.	0.0072	0.50	-5.45	3.85	-0.82^b	15.23	451.21^a	609.49^a	0.16^a	0.05^b	0.03	169.45^a	248.94^a
Prsnl. Care & Hlhd.	0.0074	0.62	-7.59	6.41	-0.36	16.30	144.00^a	214.19^a	0.04^b	0.01	-0.01	14.49^b	14.84
General Industr.	-0.0011	0.47	-5.61	2.50	-1.00^b	13.64	659.92^a	846.00^a	0.15^a	0.04^a	0.03	138.96^a	159.68^a
General Retailers	0.0003	0.50	-5.91	3.17	-0.45	10.35	456.48^a	531.52^a	0.08^a	0.01	-0.02	39.43^a	45.42^a
Hhld. Goods & Tex.	-0.0024	0.46	-4.85	4.71	-0.31	18.62	377.43^a	762.23^a	0.16^a	0.05^b	0.06^a	197.04^a	249.96^a
Oil & Gas	0.0096	0.60	-4.73	4.01	-0.17	7.16	348.89^a	611.46^a	0.05^a	-0.04^b	-0.05^a	37.91^a	44.79^a
Forestry & Paper	0.0020	0.78	-10.66	9.43	0.00	19.93	34.59^a	136.26^a	0.12^a	0.04^b	-0.02	81.91^a	96.85^a
Health	0.0032	0.57	-5.98	6.22	0.02	18.31	415.23^a	566.44^a	0.03^b	-0.02	0.01	10.72^c	32.11^a
Pharm. & Biotec	0.0093	0.67	-7.30	6.66	-0.23	12.91	1119.87^a	1169.80^a	0.05^b	-0.06^a	-0.02	29.01^a	30.32^a
Support Services	0.0005	0.46	-5.44	2.91	-1.17^b	16.07	647.02^a	975.25^a	0.12^a	0.02	0.02	77.86^a	91.18^a

Notes: S.d.: Standard deviation; Skew.: Skewness; Kurt.: Kurtosis. [a] denotes significance at 1%; [b] significance at 5% and [c] significance at 10%.

Table 1.3: **Engle and Ng Diagnostic Test for the Impact of News on Volatility (Engle an Ng, 1993)** **This Table displays, for each index excess return, the diagnostic test results for respectively the Sign Bias, The Negative Size Bias, the Positive Size Bias and the joint test. The volatility dynamic under the null we assume is the standard Gaussian GARCH(1,1).**

	Diagnostic Test Results			
	Sign Bias	Negative Size Bias	Positive Size Bias	Joint Test
FTSE Actuaries 350	2.18[b]	-2.26[b]	-5.03[a]	96.53[a]
Banks	1.22	-4.30[a]	-5.45[a]	94.09[a]
Beverages	1.26	-6.23[a]	-8.47[a]	115.38[a]
Cnstr. & Bldg. Mats.	1.62	-3.92[a]	-3.13[a]	17.41[a]
Chemicals	1.32	-2.18[b]	-3.76[a]	36.09[a]
Eng. & Machinery	1.71[c]	-1.92[b]	-4.21[a]	45.54[a]
Food & Drug Retailers	2.54[b]	-4.22[a]	-4.88[a]	37.64[a]
Food Prod. & Procr.	0.65	0.56	-1.74[c]	27.98[a]
Insurance	1.57	-3.44[a]	-3.43[a]	16.70[a]
Life Assurance	1.18	-3.07[a]	-5.05[a]	38.56[a]
Investment Companies	2.65[a]	-4.43[a]	-6.29[a]	48.98[a]
Leisure & Hotels	1.61	-0.26	-2.45[b]	30.04[a]
All Share Ex. Inv. Trusts	2.22[b]	-2.24[b]	-5.03[a]	91.75[a]
FTSE Financials	2.69[a]	-4.65[a]	-6.16[a]	67.79[a]
Transport	2.32[b]	-4.17[a]	-5.27[a]	34.41[a]
Speciality & Other Finc.	2.11[b]	-5.48[a]	-6.49[a]	58.26[a]
Prsnl. Care & Hhld. Prods.	-0.07	1.36	2.20[b]	14.05[a]
General Industrials	1.88[c]	-3.03[a]	-4.68[a]	46.04[a]
General Retailers	1.34	-1.88[c]	-5.06[a]	56.55[a]
Household Goods & Text.	0.62	-3.44[a]	-3.59[a]	22.86[a]
Oil & Gas	0.53	-0.98	-3.45[a]	33.14[a]
Forestry & Paper	-0.22	-0.91	-1.79[c]	27.25[a]
Health	-0.47	3.07[a]	2.57[a]	26.85[a]
Pharm. & Biotec	1.50	-6.85[a]	-8.61[a]	223.87[a]
Support Services	2.21[b]	-2.72[a]	-4.17[a]	41.03[a]

Notes: [a], [b] and [c] denote significance at 1%, 5% and 10%, respectively.

Table 1.4: This Table presents, for each sectorial index and the market index FTSE 350, $\hat{\pi}_1$ and $\hat{h}_1$, ordinary least square estimates of π_1 and h_1 respectively in the regressions $E\left[\epsilon_{i,t}\Sigma_{ii,t}|\Sigma_{ii,t-1}\right] = \pi_0 + \pi_1\Sigma_{ii,t-1}$ and $E\left[\epsilon^3_{i,t}|\Sigma_{ii,t-1}\right] = h_0 + h_1\Sigma_{ii,t-1}$. $\epsilon_{i,t} \equiv Y_{i,t} - \overline{Y}_i$, $Y_{i,t}$ is the excess return of the index i at date t, $\overline{Y}_i$ is the sample mean of the excess return of i and $\Sigma_{ii,t}$ is the conditional variance of $Y_{i,t+1}$. As a proxy for $\Sigma_{ii,t}$, we use the daily square excess return: $Y^2_{i,t+1}$. $\hat{\pi}_{f,1}$ and $\hat{h}_{f,1}$, ordinary least square estimates of $\pi_{f,1}$ and $h_{f,1}$ respectively in the regressions $E\left[\epsilon_{i,t}\Sigma_{11,t}|\Sigma_{11,t-1}\right] = \pi_{f,0} + \pi_{f,1}\Sigma_{11,t-1}$ and $E\left[\epsilon^3_{i,t}|\Sigma_{11,t-1}\right] = h_{f,0} + h_{f,1}\Sigma_{11,t-1}$. $\Sigma_{11,t}$ is the FTSE 350 index excess return's conditional variance. The proxy used for this conditional variance is $Y^2_{1,t+1}$ where $Y_{1,t+1}$ is the FTSE 350 index excess return. The Co-Skewness of the sectorial index excess return $Y_{i,t}$ with the FTSE 350 index excess return $Y_{1,t}$ is given, as in Ang and Chen (2002), by Co-Skewness$(Y_{i,t}, Y_{1,t}) = E[(Y_{i,t} - E(Y_{i,t}))(Y_{1,t} - EY_{1,t})^2] / \left[\sqrt{E(Y_{i,t} - E(Y_{i,t}))^2} E(Y_{1,t} - E(Y_{1,t}))^2\right]$. The significance tests for the Co-Skewness have been performed by Moment method-based asymptotic distribution of sample mean and the $\delta-$method.

Sector	$\hat{\pi}_1$	$\hat{h}_1$	$\hat{\pi}_{f,1}$	$\hat{h}_{f,1}$	Co-Skewness
FTSE Actuaries 350	-1.74^a	-3.13^a	-1.74^a	-3.13^a	-0.88^c
Banks	-1.29^a	-1.64^a	-1.97^a	-4.57^a	-0.39
Beverages	-1.26^a	-1.27^a	-1.74^a	-4.52^a	-0.49
Cnstr. & Bldg. Mats.	-0.41^a	-0.58^a	-1.38^a	-1.95^a	-0.46
Chemicals	-1.26^a	-2.70^a	-1.87^a	-3.40^a	-0.68^c
Eng. & Machinery	-0.36^a	-1.66^a	-1.58^a	-4.02^a	-0.69^c
Food & Drug Retailers	-0.66^a	-0.99^a	-1.31^a	-1.76^a	-0.30
Food Prod. & Procr.	-0.79^a	-1.78^a	-1.60^a	-2.18^a	-0.56^c
Insurance	-0.73^a	-2.87^a	-2.20^a	-5.06^a	-0.42
Life Assurance	-0.82^a	-1.52^a	-1.75^a	-3.96^a	-0.36
Investment Companies	-0.79^a	-2.96^a	-1.39^a	-2.59^a	-0.96^b
Leisure & Hotels	-0.82^a	-1.52^a	-1.76^a	-2.80^a	-0.53
All Share Ex. Inv. Trusts	-1.70^a	-3.07^a	-1.71^a	-2.98^a	-0.90^c
FTSE Financials	-1.09^a	-2.10^a	-1.73^a	-3.16^a	-0.61
Transport	-0.83^a	-2.17^a	-1.42^a	-2.20^a	-0.66^c
Speciality & Other Finc.	-0.85^a	-2.12^a	-1.64^a	-3.49^a	-0.82^b
Prsnl. Care & Hhld. Prods.	0.06^b	-2.21^a	-1.54^a	-2.33^a	-0.30
General Industrials	-1.37^a	-3.19^a	-1.75^a	-3.59^a	-0.81^c
General Retailers	-0.92^a	-2.67^a	-1.49^a	-3.25^a	-0.53
Household Goods & Text.	-0.82^a	-1.14^a	-1.67^a	-3.09^a	-0.81^c
Oil & Gas	-0.61^a	-0.96^a	-1.68^a	-2.44^a	-0.40^c
Forestry & Paper	-0.23^a	-1.70^a	-0.94^a	-0.77^c	-0.13
Health	0.49^a	0.37^a	-1.97^a	-3.31^a	-0.46
Pharm. & Biotec	-2.82^a	-1.59^a	-2.45^a	-7.73^a	-0.48
Support Services	-1.52^a	-3.50^a	-1.86^a	-4.00^a	-0.95^c

Notes: ^{a}, ^{b} and ^{c} denote significance at 1%, 5% and 10%, respectively.

Table 1.5: Simulated bias, root mean square error (RMSE), median and least absolute deviation (LAD) of GMM parameter estimates of the Doz and Renault (2006) model (DR) and of our conditionally heteroskedastic factor model with asymmetries. We report the results from GMM estimates using 4 different sets of valid instruments: $z_{1,t} = (1, Y_{1,t}^2)$, $z_{2,t} = (1, \sum_{i=1}^{20} \rho^i Y_{1,t-1+1}^2; \rho = 0.9)$, $z_{3,t} = (1, Y_{1,t}^2, Y_{1,t-1}^2)$ and $z_{4,t} = (1, Y_{1,t}^2, Y_{1,t-1}^2, Y_{1,t-2}^2)$. The simulated data are obtained from the DGP 2. The true parameter values are: $(\lambda_1, \lambda_2, \lambda_3, \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1, 1, 1, 0.5, 0.2, 0.6, 0.8, 0, 0, 0, 0)$. We use the moment conditions associated to the inference by normalization approach described in Section 5. The estimated parameters are λ_2, λ_3, ω, γ for the DR (2006) model and λ_2, λ_3, ω, α, β, γ, s, h_1, π_0, π_1 for our extended model.

	DR (2006)				*Our model*			
Parameter	Bias	RMSE	Median	LAD	Bias	RMSE	Median	LAD
$z_{1,t}$								
γ	0.017	0.137	0.826	0.116	0.021	0.137	0.826	0.116
λ_2	0.000	0.013	1.000	0.011	-0.001	0.016	0.999	0.012
λ_3	0.000	0.013	1.001	0.011	-0.001	0.016	0.999	0.012
ω	-0.002	0.005	0.348	0.004	-0.002	0.006	0.348	0.005
s	-	-	-	-	0.041	0.727	0.031	0.585
h_1	-	-	-	-	-0.044	0.786	-0.019	0.635
π_0	-	-	-	-	0.021	0.456	0.016	0.312
π_1	-	-	-	-	-0.021	0.487	-0.018	0.337
$z_{2,t}$								
γ	-0.018	0.110	0.801	0.056	-0.014	0.097	0.800	0.054
λ_2	0.000	0.013	0.999	0.010	-0.001	0.014	0.999	0.011
λ_3	0.000	0.013	1.000	0.011	-0.002	0.014	0.998	0.011
ω	-0.001	0.005	0.348	0.004	-0.002	0.005	0.348	0.004
s	-	-	-	-	-0.039	0.777	-0.048	0.626
h_1	-	-	-	-	0.039	0.832	0.064	0.679
π_0	-	-	-	-	-0.001	0.302	-0.012	0.223
π_1	-	-	-	-	0.003	0.330	0.016	0.245
$z_{3,t}$								
γ	-0.007	0.093	0.798	0.076	-0.005	0.096	0.801	0.077
λ_2	0.000	0.013	1.000	0.011	-0.001	0.019	0.999	0.015
λ_3	0.001	0.014	1.001	0.011	-0.001	0.020	0.998	0.015
ω	-0.002	0.006	0.348	0.004	-0.003	0.007	0.347	0.005
s	-	-	-	-	0.023	0.665	0.016	0.528
h_1	-	-	-	-	-0.023	0.729	-0.014	0.586
π_0	-	-	-	-	0.023	0.355	-0.006	0.246
π_1	-	-	-	-	-0.020	0.380	0.004	0.270
$z_{4,t}$								
γ	-0.015	0.078	0.793	0.062	-0.013	0.081	0.800	0.063
λ_2	0.001	0.014	1.000	0.011	-0.001	0.021	0.999	0.016
λ_3	0.001	0.014	1.001	0.011	-0.002	0.022	0.998	0.017
ω	-0.003	0.006	0.347	0.005	-0.004	0.007	0.346	0.005
s	-	-	-	-	0.019	0.673	-0.003	0.537
h_1	-	-	-	-	-0.019	0.737	0.009	0.596
π_0	-	-	-	-	0.030	0.343	0.012	0.237
π_1	-	-	-	-	-0.026	0.365	0.006	0.261

Table 1.6: Simulated Bias, root mean square error (RMSE), median and least absolute deviation (LAD) of GMM parameter estimates of the Doz and Renault (2006) model (DR) and of our conditionally heteroskedastic factor model with asymmetries. We report the results from GMM estimates using $z_{4,t} = (1, Y_{1,t}^2, Y_{1,t-1}^2, Y_{1,t-2}^2)$ as the instrument. The data are generated according to DGP 1, DGP 2, DGP 3 and DGP 4. In particular, the true parameter values are: DGP 1: $(\lambda_1, \lambda_2, \lambda_3,\ \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1,1,1,\ 0.35, 0.2, 0.50, 0.70, 0,0,0,0)$, DGP 2: $(\lambda_1, \lambda_2, \lambda_3,\ \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1,1,1,\ 0.35, 0.2, 0.60, 0.80, 0,0,0,0)$, DGP 3: $(\lambda_1, \lambda_2, \lambda_3, \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1,1,1, 0.35, 0.2, 0.70, 0.90, 0,0,0,0)$, DGP 4: $(\lambda_1,\ \lambda_2,\ \lambda_3,\ \omega,\ \alpha,\ \beta,\ \gamma,\ s,\ h_1,\ \pi_0,\ \pi_1) = (1,\ 1,\ 1,\ 0.35,\ 0.2,\ 0.75,\ 0.95,\ 0,\ 0,\ 0,\ 0)$. The true values of the volatility persistence parameter are 0.70, 0.80, 0.90 and 0.95, respectively.

	DR (2006)				*Our model*			
Parameter	Bias	RMSE	Median	LAD	Bias	RMSE	Median	LAD
DGP 1								
γ	-0.024	0.095	0.678	0.075	-0.025	0.102	0.684	0.080
λ_2	0.000	0.013	0.999	0.011	-0.002	0.017	0.998	0.014
λ_3	0.000	0.013	0.999	0.011	-0.003	0.018	0.997	0.014
ω	-0.003	0.006	0.347	0.005	-0.004	0.007	0.346	0.005
s	-	-	-	-	0.029	0.665	0.006	0.538
h_1	-	-	-	-	-0.029	0.716	0.003	0.583
π_0	-	-	-	-	0.032	0.340	0.010	0.242
π_1	-	-	-	-	-0.030	0.358	-0.006	0.258
DGP 2								
γ	-0.015	0.078	0.793	0.062	-0.013	0.081	0.800	0.063
λ_2	0.001	0.014	1.000	0.011	-0.001	0.021	0.999	0.016
λ_3	0.001	0.014	1.001	0.011	-0.002	0.022	0.998	0.017
ω	-0.003	0.006	0.347	0.005	-0.004	0.007	0.346	0.005
s	-	-	-	-	0.019	0.673	-0.003	0.537
h_1	-	-	-	-	-0.019	0.737	0.009	0.596
π_0	-	-	-	-	0.030	0.343	0.012	0.237
π_1	-	-	-	-	-0.026	0.365	0.006	0.261
DGP 3								
γ	-0.011	0.060	0.894	0.047	-0.008	0.060	0.899	0.047
λ_2	0.002	0.015	1.001	0.012	0.002	0.035	1.003	0.024
λ_3	0.002	0.015	1.001	0.012	-0.001	0.033	1.001	0.024
ω	-0.003	0.006	0.347	0.005	-0.004	0.008	0.346	0.006
s	-	-	-	-	0.031	0.937	-0.021	0.647
h_1	-	-	-	-	-0.011	0.972	0.071	0.708
π_0	-	-	-	-	0.024	0.526	0.033	0.285
π_1	-	-	-	-	-0.013	0.483	-0.030	0.315
DGP 4								
γ	-0.016	0.054	0.942	0.040	-0.015	0.053	0.943	0.040
λ_2	0.002	0.017	1.002	0.013	0.003	0.047	1.004	0.034
λ_3	0.003	0.017	1.003	0.013	0.001	0.047	1.001	0.034
ω	-0.001	0.020	0.348	0.005	-0.001	0.021	0.347	0.010
s	-	-	-	-	0.431	11.604	-0.102	1.591
h_1	-	-	-	-	0.025	1.547	0.112	0.990
π_0	-	-	-	-	-0.062	3.062	0.005	0.685
π_1	-	-	-	-	0.014	0.765	0.006	0.442

Table 1.7: Simulated Bias, root mean square error (RMSE), median and least absolute deviation (LAD) of GMM parameter estimates of the Doz and Renault (2006) model (DR) and of our conditionally heteroskedastic factor model with asymmetries. We report the results from GMM estimates using $z_{4,t} = (1, Y_{1,t}^2, Y_{1,t-1}^2, Y_{1,t-2}^2)$ as the instrument. The data are generated according to DGP 1', DGP 2', DGP 3' and DGP 4'. In particular, the true parameter values are: DGP 1': $(\lambda_1, \lambda_2, \lambda_3, \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1, 1, 1, 0.35, 0.2, 0.50, 0.70, 0, -2.0, 0, -0.4)$, DGP 2': $(\lambda_1, \lambda_2, \lambda_3, \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1, 1, 1, 0.35, 0.2, 0.60, 0.80, 0, -2.0, 0, -0.4)$, DGP 3': $(\lambda_1, \lambda_2, \lambda_3, \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1, 1, 1, 0.35, 0.2, 0.70, 0.90, 0, -2.0, 0, -0.4)$, DGP 4': $(\lambda_1, \lambda_2, \lambda_3, \omega, \alpha, \beta, \gamma, s, h_1, \pi_0, \pi_1) = (1, 1, 1, 0.35, 0.2, 0.75, 0.95, 0, -2.0, 0, -0.4)$. The true values of the volatility persistence parameter are 0.70, 0.80, 0.90 and 0.95, respectively.

	DR (2006)				*Our model*			
Parameter	Bias	RMSE	Median	LAD	Bias	RMSE	Median	LAD
DGP 1'								
γ	-0.034	0.121	0.673	0.093	-0.030	0.131	0.680	0.095
λ_2	0.003	0.022	1.002	0.015	-0.002	0.042	0.997	0.029
λ_3	0.003	0.021	1.002	0.015	-0.004	0.038	0.999	0.027
ω	-0.002	0.008	0.347	0.005	-0.002	0.016	0.347	0.008
s	-	-	-	-	0.022	2.666	-0.220	1.448
h_1	-	-	-	-	0.367	2.047	-1.430	1.443
π_0	-	-	-	-	-0.025	0.993	-0.161	0.615
π_1	-	-	-	-	0.137	0.903	-0.149	0.641
DGP 2'								
γ	-0.025	0.101	0.786	0.075	-0.020	0.103	0.794	0.075
λ_2	0.003	0.020	1.000	0.015	0.000	0.043	0.999	0.031
λ_3	0.004	0.020	1.002	0.015	0.001	0.048	1.000	0.034
ω	-0.002	0.008	0.348	0.005	-0.001	0.018	0.348	0.009
s	-	-	-	-	-0.048	2.822	-0.159	1.400
h_1	-	-	-	-	0.487	2.046	-1.389	1.416
π_0	-	-	-	-	-0.065	1.156	-0.093	0.571
π_1	-	-	-	-	0.153	0.804	-0.159	0.599
DGP 3'								
γ	-0.015	0.073	0.897	0.055	-0.021	0.109	0.894	0.063
λ_2	0.004	0.020	1.004	0.015	-0.003	0.053	1.001	0.039
λ_3	0.003	0.019	1.004	0.015	-0.002	0.056	0.997	0.041
ω	-0.002	0.007	0.348	0.005	0.003	0.064	0.347	0.017
s	-	-	-	-	-0.070	3.890	-0.071	1.570
h_1	-	-	-	-	0.597	2.072	-1.346	1.416
π_0	-	-	-	-	0.046	1.813	-0.045	0.700
π_1	-	-	-	-	0.180	0.870	-0.148	0.610
DGP 4'								
γ	-0.020	0.073	0.943	0.048	-0.027	0.110	0.940	0.056
λ_2	0.002	0.020	1.001	0.015	-0.005	0.136	1.000	0.055
λ_3	0.002	0.021	1.001	0.016	0.001	0.075	1.000	0.049
ω	0.002	0.053	0.348	0.008	0.060	1.060	0.350	0.078
s	-	-	-	-	0.581	9.500	-0.090	2.712
h_1	-	-	-	-	0.651	2.679	-1.200	1.776
π_0	-	-	-	-	0.352	13.532	-0.090	1.663
π_1	-	-	-	-	0.304	1.067	-0.040	0.742

Table 1.8: GMM parameter estimates for the Doz and Renault (2006) model (DR) and for our conditionally heteroskedastic factor model with asymmetries.

	Factor Loadings				Idiosyncratic Variances				s_i	
	DR		Our model		DR		Our model		Our model	
Sector	*est.*	*s.e.*	*est.*	*s.e.*	*est.*	*s.e.*	*est.*	*s.e.*	*est.*	*s.e.*
FTSE Actuaries 350	0.35	-	0.35	-	0.02	0.09	0.03	0.06	0.30	0.46
Banks	0.43	0.11	0.35	0.03	0.22	1.00	0.30	0.75	0.35	0.90
Beverages	0.37	0.16	0.40	0.04	0.19	1.18	0.17	0.93	0.56	1.44
Cnstr. & Bldg. Mats.	0.31	0.17	0.35	0.03	0.21	0.93	0.20	0.64	0.32	0.66
Chemicals	0.33	0.13	0.39	0.04	0.19	0.87	0.16	0.65	0.36	0.73
Eng. & Machinery	0.33	0.16	0.43	0.03	0.22	1.10	0.16	0.70	0.56	1.02
Food & Drug Retailers	0.29	0.18	0.28	0.06	0.19	0.94	0.21	0.75	0.15	0.68
Food Prod. & Procr.	0.30	0.10	0.29	0.03	0.14	0.78	0.15	0.58	0.15	0.62
Insurance	0.34	0.20	0.34	0.07	0.39	1.59	0.42	1.13	0.14	1.66
Life Assurance	0.46	0.17	0.39	0.05	0.24	1.20	0.31	0.99	0.43	1.38
Investment Companies	0.30	0.10	0.36	0.02	0.08	0.43	0.05	0.34	0.32	0.52
Leisure & Hotels	0.33	0.15	0.32	0.03	0.19	0.78	0.21	0.53	0.21	0.51
FTSE Financials	0.38	0.07	0.35	0.02	0.13	0.53	0.16	0.36	0.31	0.56
Transport	0.29	0.10	0.35	0.02	0.12	0.52	0.09	0.38	0.31	0.52
Speciality & Other Finc.	0.33	0.14	0.39	0.04	0.18	0.95	0.15	0.74	0.39	0.85
Prsnl. Care & Hhld. Prods.	0.27	0.18	0.31	0.07	0.31	2.33	0.30	1.61	0.16	2.10
General Industrials	0.35	0.10	0.39	0.02	0.12	0.65	0.10	0.48	0.36	0.78
General Retailers	0.33	0.14	0.38	0.02	0.11	0.70	0.08	0.54	0.42	0.82
Household Goods & Text.	0.28	0.24	0.37	0.04	0.20	0.96	0.15	0.76	0.34	0.95
Oil & Gas	0.31	0.16	0.30	0.06	0.28	1.24	0.30	0.86	0.18	0.94
Forestry & Paper	0.19	0.24	0.19	0.11	0.55	3.93	0.57	3.01	0.08	5.22
Health	0.34	0.13	0.25	0.06	0.28	2.05	0.33	1.30	0.10	1.57
Pharm. & Biotec	0.48	0.15	0.16	0.08	0.16	1.45	0.35	1.24	0.17	1.98
Support Services	0.33	0.11	0.37	0.03	0.15	0.60	0.12	0.49	0.32	0.70

Notes: est.: Estimate; s.e.: Estimated GMM standard error×100; s_i: Conditional third moment model's intercept.

DR:			**Our model:**					
$T=$	4863		$T=$	4863				
$\gamma=$	*0.685*	(0.0037)	$\gamma=$	*0.684*	(0.0017)	$\pi_0=$	0.753	(0.0320)
			$h_1=$	-6.424	(0.0138)	$\pi_1=$	-1.467	(0.0043)
$J=$	104.32 $[p-Value]=$ .999		$J=$	117.13 $[p-Value]=$ .999				

Notes: Estimated GMM standard errors in parenthesis.

Table 1.9: Descriptive statistics of the filtered factors and their correlation with the FTSE 350 index excess return. The filtered factors are obtained by the extended Kalman filter algorithm we propose (see Appendix A) and we use the generalized method of moments (GMM) parameter estimates from the Doz and Renault (2006) model (DR) and the GMM parameter estimates from our conditionally heteroskedastic factor model with asymmetries.

	Filtered Factor from :	
	DR estimates	our model estimates
Mean	-0.006	-0.003
Standard error	1.006	1.014
Skewness	-1.615	-1.333
Corr. with FTSE 350	0.932	0.911

Notes: Corr. denotes Correlation.

Table 1.10: Descriptive statistics of the filtered FTSE 350 index excess return idiosyncratic shocks. These filtered idiosyncratic shocks are given by $u_{dr,t+1} = Y_{1,t+1} - 0.35 f_{dr,t+1}$ and $u_{chfa,t+1} = Y_{1,t+1} - 0.35 f_{chfa,t+1}$ where $Y_{1,t+1}$ is the FTSE 350 index excess return, $f_{dr,t+1}$ is the filtered factor using the GMM parameter estimates from DR (2006) model and $f_{chfa,t+1}$ the filtered factor using the GMM parameter estimates from our conditionally heteroskedastic factor model with asymmetries (CHFA). Eng(2) is the lag 2 Engle's (1982) Lagrange multiplier test statistics for conditional heteroskedasticity.

	$u_{dr,t+1}$		$u_{chfa,t+1}$	
Mean	0.002		0.009	
Standard error	0.162		0.180	
Skewness	0.272	$[p-Value]$=0.746	-0.463	$[p-Value]$=0.450
Corr. with FTSE 350	0.622		0.586	
Eng(2)	6.616	$[p-Value]$=0.037	7.426	$[p-Value]$=0.024
QW(10)	19.998	$[p-Value]$=0.029	12.320	$[p-Value]$=0.264

Notes: Corr. denotes Correlation; QW(10): 10-order Ljung-Box test statistic for autocorrelation.

C Proofs of Propositions

Proof of Proposition 4.1: The expression given in (13) is obvious and arises out from the sum of (1) over the time period: $\tau = (t-1)m + 1$ through tm with the respective aggregation coefficients α_l and $\mu(J_t) = \mu$. Let $\left(F^{(m)}_{(t+1)m}\right)$ and $\left(U^{(m)}_{(t+1)m}\right)$ be the resulting factor and the idiosyncratic shocks and let $D^{(m)}_{tm}$ be the $J^{(m)}_{tm}$-conditional variance of this factor.

$$D^{(m)}_{tm} = E\left(F^{(m)}_{(t+1)m} F^{(m)'}_{(t+1)m} | J^{(m)}_{tm}\right) - E\left(F^{(m)}_{(t+1)m} | J^{(m)}_{tm}\right) E\left(F^{(m)'}_{(t+1)m} | J^{(m)}_{tm}\right).$$

On the other hand,

$$\begin{aligned} E\left(F^{(m)}_{(t+1)m} | J_{tm}\right) &= E\left(\left(\sum_{l=1}^{m} \alpha_l F_{tm+l}\right) | J_{tm}\right) = \sum_{l=1}^{m} \alpha_l E\left(F_{tm+l} | J_{tm}\right) \\ &= \sum_{l=1}^{m} \alpha_l E\left(E\left(F_{tm+l} | J_{tm+l-1}\right) | J_{tm}\right) = 0. \end{aligned}$$

The last equality holds by the law of iterated expectations and the last is from (2). Since $J_{tm}^{(m)}$ is included in J_{tm} by definition, the law of iterated expectations also applies the following way: $E(X|J_{tm}^{(m)}) = E\left(E(X|J_{tm})|J_{tm}^{(m)}\right)$ for any measurable X. Therefore, $E\left(F_{(t+1)m}^{(m)}|J_{tm}^{(m)}\right) = 0$.
Let us consider k and k' such that $k \neq k'$.

$$\begin{aligned} E\left(F_{k,(t+1)m}^{(m)}F_{k',(t+1)m}^{(m)}|J_{tm}^{(m)}\right) &= E\left(\left(\sum_{l=1}^{m}\alpha_l F_{k,tm+l}\right)\left(\sum_{l=1}^{m}\alpha_l F_{k',tm+l}\right)|J_{tm}^{(m)}\right) \\ &= E\left[\sum_{l<l';l,l'=1}^{m}\alpha_l\alpha_{l'}(F_{k,tm+l}F_{k',tm+l'} + F_{k',tm+l}F_{k,tm+l'})\right. \\ &\quad \left. + \sum_{l=1}^{m}\alpha_l^2 F_{k,tm+l}F_{k',tm+l}|J_{tm}^{(m)}\right]. \end{aligned}$$

But, from the law of iterated expectations and (2), for $l < l'$,
$E\left(F_{k,tm+l}F_{k',tm+l'}|J_{tm}\right) = E\left(F_{k,tm+l}E\left(F_{k',tm+l'}|J_{tm+l'-1}\right)|J_{tm}\right) = 0$ and in addition, as D_t is diagonal for all t from (2), $E\left(F_{k,tm+l}F_{k',tm+l}|J_{tm}\right) = E\left(E\left(F_{k,tm+l}F_{k',tm+l}|J_{tm+l-1}\right)|J_{tm}\right) = E\left(D_{k,k',tm+l-1}|J_{tm}\right) = 0$.
By the law of iterated expectations as above we can deduce that $E\left(F_{k,(t+1)m}^{(m)}F_{k',(t+1)m}^{(m)}|J_{tm}^{(m)}\right) = 0$ and therefore $D_{tm}^{(m)}$ is diagonal.
By the law of iterated expectations and simple product expansion, we easily show that $E\left(U_{(t+1)m}^{(m)}|J_{tm}^{(m)}\right) = 0$ and $E\left(U_{(t+1)m}^{(m)}F_{(t+1)m}^{(m)\prime}|J_{tm}^{(m)}\right) = 0$. On the other hand,

$$\begin{aligned} Var\left(U_{(t+1)m}^{(m)}|J_{tm}^{(m)}\right) &= E\left(\left(\sum_{l=1}^{m}\alpha_l U_{tm+l}\right)\left(\sum_{l=1}^{m}\alpha_l U_{tm+l}\right)'|J_{tm}^{(m)}\right) \\ &= E\left(\sum_{l<l';l,l'=1}^{m}\alpha_l\alpha_{l'}(U_{tm+l}U_{tm+l'}' + U_{tm+l'}U_{tm+l}')\right. \\ &\quad \left. + \sum_{l=1}^{m}\alpha_l^2 U_{tm+l}U_{tm+l}'|J_{tm}^{(m)}\right) \end{aligned}$$

For $l < l'$, $E\left(U_{tm+l}U_{tm+l'}'|J_{tm}\right) = E\left(U_{tm+l}E\left(U_{tm+l'}'|J_{tm+l'-1}\right)|J_{tm}\right) = 0$
and $E\left(U_{tm+l}U_{tm+l}'|J_{tm}\right) = E\left(E\left(U_{tm+l}U_{tm+l}'|J_{tm+l-1}\right)|J_{tm}\right) = E\left(\Omega|J_{tm}\right) = \Omega$ thus, $Var\left(U_{(t+1)m}^{(m)}|J_{tm}^{(m)}\right) = \Omega\sum_{l=1}^{m}\alpha_l^2$ completing the proof of Proposition 4.1□

Proof of Proposition 4.2: Since (f_{t+1}) has a SR-SARV(1) dynamic, with $v_{t+1} \equiv \sigma_{t+1}^2 - (1-\gamma) - \gamma\sigma_t^2$, we have: $E(v_{t+1}|J_t) = 0$. For $l = 1$, the first conclusion of the proposition is obvious since σ_{tm}^2 is J_{tm}-measurable. For $l \geq 2$, by writing v_{t+1} for different time and making some simple substitutions, we can write:

$$\sigma_{tm+l-1}^2 = (1-\gamma)(1+\gamma+\gamma^2+\cdots+\gamma^{l-2}) + \gamma^{l-1}\sigma_{tm}^2 + \gamma^{l-2}v_{tm+1} + \gamma^{l-3}v_{tm+2} + \cdots + v_{tm+l-1}.$$

By taking the expectation conditionally on J_{tm} and by the law of iterated expectations, we have:

$$\begin{aligned} E(\sigma_{tm+l-1}^2|J_{tm}) &= (1-\gamma)(1+\gamma+\gamma^2+\cdots+\gamma^{l-2}) + \gamma^{l-1}\sigma_{tm}^2 \\ &= (1-\gamma)\frac{1-\gamma^{l-1}}{1-\gamma} + \gamma^{l-1}\sigma_{tm}^2 \\ &= 1-\gamma^{l-1} + \gamma^{l-1}\sigma_{tm}^2. \end{aligned}$$

The first conclusion is then established. On the other hand,

$$
\begin{aligned}
E\left((f^{(m)}_{(t+1)m})^2|J_{tm}\right) &= E\left(\sum_{l=1}^{m}(\alpha_l f_{tm+l})^2|J_{tm}\right), \\
&\quad f_t(\text{is conditionally non-autocorrelated}) \\
&= \sum_{l=1}^{m}\alpha_l^2 E(f^2_{tm+l}|J_{tm}) \\
&= \sum_{l=1}^{m}\alpha_l^2 E(E(f^2_{tm+l}|J_{tm+l-1})|J_{tm}) \\
&= \sum_{l=1}^{m}\alpha_l^2 E(\sigma^2_{tm+l-1}|J_{tm}) \\
&= \sum_{l=1}^{m}\alpha_l^2[(1-\gamma^{l-1})+\gamma^{l-1}\sigma^2_{tm}]
\end{aligned}
$$

Since σ^2_{tm} is $J^{(m)}_{tm}$-measurable, $E\left((f^{(m)}_{(t+1)m})^2|J^{(m)}_{tm}\right) = \sum_{l=1}^{m}\alpha_l^2[(1-\gamma^{l-1})+\gamma^{l-1}\sigma^2_{tm}]$. Hence,

$$
\sigma^{(m)^2}_{tm} \equiv Var\left(f^{(m)}_{(t+1)m}|J^{(m)}_{tm}\right) = S^{(m)}_1 + S^{(m)}_2\sigma^2_{tm}
$$

with $S^{(m)}_1 = \sum_{l=1}^{m}\alpha_l^2[(1-\gamma^{l-1})$ and $S^{(m)}_2 = \sum_{l=1}^{m}\alpha_l^2\gamma^{l-1}$ □

Proof of Proposition 4.3:

$$
\begin{aligned}
Cov\left(f^{(m)}_{(t+1)m}, \sigma^{(m)^2}_{(t+1)m}|J_{tm}\right) &= \sum_{l=1}^{m}\alpha_l Cov\left(f_{tm+l}, \sigma^{(m)^2}_{(t+1)m}|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l Cov\left(f_{tm+l}, S^{(m)}_1 + S^{(m)}_2\sigma^2_{(t+1)m}|J_{tm}\right) \\
&\quad \text{(from Proposition 4.2)} \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2 Cov\left(f_{tm+l}, \sigma^2_{(t+1)m}|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2 E\left(f_{tm+l}\sigma^2_{(t+1)m}|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2 E\left(f_{tm+l}E\left(\sigma^2_{(t+1)m}|J_{tm+l}\right)|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2 E\left(f_{tm+l}\left(1-\gamma^{m-l}+\gamma^{m-l}\sigma^2_{tm+l}\right)|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2\gamma^{m-l} E\left(f_{tm+l}\sigma^2_{tm+l}|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2\gamma^{m-l} E\left(E\left(f_{tm+l}\sigma^2_{tm+l}|J_{tm+l-1}\right)|J_{tm}\right) \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2\gamma^{m-l} E\left(\pi_0+\pi_1\sigma^2_{tm+l-1}|J_{tm}\right) \\
&\quad \text{(from Assumption 2)} \\
&= \sum_{l=1}^{m}\alpha_l S^{(m)}_2\gamma^{m-l}\left(\pi_0+\pi_1 E\left(\sigma^2_{tm+l-1}|J_{tm}\right)|J_{tm}\right) \\
&\equiv l^{(m)}_1 + l^{(m)}_2\sigma^2_{tm}, \text{ (Proposition 4.2); } l^{(m)}_1 \text{ and } l^{(m)}_2 \text{ two scalars.}
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
Cov\left(f^{(m)}_{(t+1)m}, \sigma^{(m)^2}_{(t+1)m}|J^{(m)}_{tm}\right) &= E\left(f^{(m)}_{(t+1)m}\sigma^{(m)^2}_{(t+1)m}|J^{(m)}_{tm}\right) = E\left[E\left(f^{(m)}_{(t+1)m}\sigma^{(m)^2}_{(t+1)m}|J_{tm}\right)|J^{(m)}_{tm}\right] \\
&= E\left[l^{(m)}_1 + l^{(m)}_2\sigma^2_{tm}|J^{(m)}_{tm}\right].
\end{aligned}
$$

Since σ^2_{tm} is $J^{(m)}_{tm}$-measurable, $Cov\left(f^{(m)}_{(t+1)m}, \sigma^{(m)^2}_{(t+1)m}|J^{(m)}_{tm}\right) = l^{(m)}_1 + l^{(m)}_2\sigma^2_{tm}$ and from the one to one mapping between σ^2_{tm} and $\sigma^{(m)^2}_{tm}$ from Proposition 4.2 we can deduce that there exists two scalars $\pi^{(m)}_0$ and $\pi^{(m)}_1$ such that $Cov\left(f^{(m)}_{(t+1)m}, \sigma^{(m)^2}_{(t+1)m}|J^{(m)}_{tm}\right) \equiv \pi^{(m)}_0 + \pi^{(m)}_1 \sigma^{(m)^2}_{tm}$ □

Proof of Proposition 4.4:

$$\begin{aligned}
E_{tm}\left[\left(f^{(m)}_{(t+1)m}\right)^3\right] &= E_{tm}\left[(\textstyle\sum_{l=1}^m \alpha_l f_{tm+l})^3\right] \\
&= \textstyle\sum_{l=1}^m \alpha_l^3 E_{tm}(f^3_{tm+l}) + 3 \times \sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 E_{tm}(f_{tm+l} f^2_{tm+l'}) \\
&\quad \text{by the conditions in (2)} \\
&= \textstyle\sum_{l=1}^m \alpha_l^3 E_{tm}(f^3_{tm+l}) + 3 \times \sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 E_{tm}\left(f_{tm+l} E_{tm+l'-1}(f^2_{k,tm+l'})\right)
\end{aligned}$$

$$\begin{aligned}
&= \textstyle\sum_{l=1}^m \alpha_l^3 E_{tm}(f^3_{tm+l}) + 3 \times \sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 E_{tm}\left(f_{tm+l} E_{tm+l} \sigma^2_{tm+l'-1}\right) \\
&= \textstyle\sum_{l=1}^m \alpha_l^3 E_{tm}(f^3_{tm+l}) + 3 \times \sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 E_{tm}\left(f_{tm+l} \gamma^{l'-l-1} \sigma^2_{tm+l}\right) \\
&= \textstyle\sum_{l=1}^m \alpha_l^3 E_{tm}(E_{tm+l-1} f^3_{tm+l}) + 3 \times \sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 \gamma^{l'-l-1} E_{tm}(E_{tm+l-1}(f_{tm+l}\sigma^2_{tm+l})) \\
&= \textstyle\sum_{l=1}^m \alpha_l^3 E_{tm}(h_0 + h_1 \sigma^2_{tm+l-1}) + 3 \times \sum \alpha_l \alpha_{l'}^2 \gamma^{l'-l-1} E_{tm}\left(\pi_0 + \pi_1 \sigma^2_{tm+l-1}\right) \\
&= \textstyle\sum_{l=1}^m \alpha_l^3 [h_0 + h_1(1-\gamma^{l-1} + \gamma^{l-1}\sigma^2_{tm})] \\
&\quad +3 \times \textstyle\sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 \gamma^{l'-l-1}\left[\pi_0 + \pi_1\left(1-\gamma^{l-1} + \gamma^{l-1}\sigma^2_{tm}\right)\right] \\
&\quad \text{(from proposition 4.2)} \\
&= \textstyle\sum_{l=1}^m \alpha_l^3\left[h_0 + (1-\gamma^{l-1})h_1\right] + 3\sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 \gamma^{l'-l-1}[\pi_0 + \pi_1(1-\gamma^{l-1})] \\
&\quad + \left[h_1 \textstyle\sum_{l=1}^m \alpha_l^3 \gamma^{l-1} + 3\pi_1 \sum_{1\le l<l'\le m} \alpha_l \alpha_{l'}^2 \gamma^{l'-2}\right]\sigma^2_{tm} \\
&\equiv B_0^{(m)} + B_1^{(m)} \sigma^2_{tm}
\end{aligned}$$

Since σ^2_{tm} is $J^{(m)}_{tm}$-measurable, the law of iterated expectations implies that $E\left((f^{(m)}_{(t+1)m})^3|J^{(m)}_{tm}\right) = B_0^{(m)} + B_1^{(m)}\sigma^2_{tm}$. By Proposition 4.2, $E\left((f^{(m)}_{(t+1)m})^3|J^{(m)}_{tm}\right) = h_0^{(m)} + h_1^{(m)}\sigma^{(m)^2}_{tm}$ with $h_0^{(m)} = B_0^{(m)} - h_1^{(m)} S_1^{(m)}$ and $h_1^{(m)} = B_1^{(m)}/S_2^{(m)}$; $S_1^{(m)}$ and $S_2^{(m)}$ are defined as in Proposition 4.2. In the calculations above, $E_t X$ stands for $E(X|J_t)$.

$$\begin{aligned}
E\left(\left(U^{(m)}_{i,(t+1)m}\right)^3 |J_{tm}\right) &= E\left[\left(\textstyle\sum_{l=1}^m \alpha_l^3 U^3_{i,tm+l}\right)|J_{tm}\right] \text{ (from (2))} \\
&= E\left[\left(\textstyle\sum_{l=1}^m \alpha_l^3 E(U^3_{i,tm+l}|J_{tm+l-1})\right)|J_{tm}\right] \\
&= E\left[\left(\textstyle\sum_{l=1}^m \alpha_l^3 s_i^0\right)|J_{tm}\right] = s_i^0\left(\textstyle\sum_{l=1}^m \alpha_l^3\right), \text{ (from Assumption 3)}
\end{aligned}$$

hence,

$$E\left(\left(U^{(m)}_{i,(t+1)m}\right)^3 |J^{(m)}_{tm}\right) = s_i^0\left(\sum_{l=1}^m \alpha_l^3\right).$$

From Assumption 3 and Equations (2),

$$\begin{aligned}
E\left(\left(Y^{(m)}_{i,(t+1)m}\right)^3 |J_{tm}\right) &= E\left(\left(\lambda_i f^{(m)}_{(t+1)m}\right)^3 + \left(U^{(m)}_{i,(t+1)m}\right)^3 |J_{tm}\right) \\
&= \lambda_i^3 E\left(\left(f^{(m)}_{(t+1)m}\right)^3 |J_{tm}\right) + E\left(\left(U^{(m)}_{i,(t+1)m}\right)^3 |J_{tm}\right)
\end{aligned}$$

Thus, $E\left(\left(Y^{(m)}_{i,(t+1)m}\right)^3 |J^{(m)}_{tm}\right) = \lambda_i^3 h_1^{(m)}\sigma^2_{tm} + \lambda_i^3 h_0 + s_i^0\left(\sum_{l=1}^m \alpha_l^3\right)$ for all i and t □

Proof of Proposition 5.1: Since there is a one to one linear relationship between ϕ_2 and $\phi_2^0 = (\pi_0^0, \pi_1^0, s_1^1, \ldots, s_N^1, h_1^0)'$, it is sufficient to prove that $\left(\partial/\partial\phi_2^{0'}\right)[Ez_t \otimes g(Y_{t+1}, Y_{t+2}, \phi_2^0)]$ has rank $N+3$.

$\left(\partial/\partial\phi_2^{0'}\right)[Ez_t \otimes g(Y_{t+1}, Y_{t+2}, \phi_2^0)] = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}$ with $A(2N \times 2)$, $C(2N \times N+1)$ given by:

$$A = \begin{pmatrix} -\underline{\lambda}^3 Ez_t & -\underline{\lambda}^3 Ez_t Y^2_{1,t+1} \\ -\lambda_2^3 Ez_t & -\lambda_2^3 Ez_t Y^2_{1,t+1} \\ \vdots & \vdots \\ -\lambda_N^3 Ez_t & -\lambda_N^3 Ez_t Y^2_{1,t+1} \end{pmatrix} \text{ and } C = \begin{pmatrix} -Ez_t & 0 & \cdots & 0 & -\underline{\lambda}^3 Ez_t Y^2_{1,t+1} \\ 0 & -Ez_t & \ddots & \vdots & -\lambda_2^3 Ez_t Y^2_{1,t+1} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & -Ez_t & -\lambda_N^3 Ez_t Y^2_{1,t+1} \end{pmatrix}.$$

We just need to show that A has rank 2 and C has rank $N+1$. Since $z_t = (1, z_{1,t})'$,

$$A_s = -\underline{\lambda}^3 \begin{pmatrix} 1 & EY^2_{1,t+1} \\ Ez_{1,t} & Ez_{1,t}Y^2_{1,t+1} \end{pmatrix} \text{ and } C_s = \begin{pmatrix} -1 & 0 & \cdots & 0 & -\underline{\lambda}^3 EY^2_{1,t+1} \\ 0 & -1 & \ddots & \vdots & -\lambda_2^3 EY^2_{1,t+1} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & -1 & -\lambda_N^3 EY^2_{1,t+1} \\ -Ez_{1,t} & 0 & \cdots & 0 & -\underline{\lambda}^3 Ez_{1,t}Y^2_{1,t+1} \end{pmatrix}$$

are two submatrix respectively of A and C ($A_s(2 \times 2)$ and $C_s(N+1 \times N+1)$). Let $Det(X)$ be the determinant of the matrix X. $Det(A_s) = \underline{\lambda}^8 Cov(z_{1,t}, \sigma_t^2) \neq 0$. On the other hand, the rank of C_s is greater or equals to N and equals N if and only in its last column belongs to the sub-space spanned by the N first columns. This happens only if $-\underline{\lambda}^3 EY^2_{1,t+1}(-Ez_{1,t}) - \underline{\lambda}^3 Ez_{1,t}Y^2_{1,t+1} = 0$ i.-e. $Cov(z_{1,t}, \sigma_t^2) = 0$ which is impossible by assumption. Thus C_s if of rank $N+1$. As a result the rank of $\left(\partial/\partial\phi_2^{0'}\right)[Ez_t \otimes g(Y_{t+1}, Y_{t+2}, \phi_2^0)]$ is $N+3$ □

Proposition C.1 *Let f_{t+1} follow a SR-SARV(1) model with volatility persistence and intercept parameters γ and $1-\gamma$, respectively and $\sigma_t^2 = Var(f_{t+1}|J_t)$. If there exist α, β and $\gamma \in \mathbb{R}$ such that:*

$$Var(\sigma^2_{t+1}|J_t) = \alpha + \beta\,\sigma_t^2 + \delta\,\sigma_t^4,$$

then, for all $l \geq 2$, there exist H_1^l, H_2^l and $H_3^l \in \mathbb{R}$ such that

$$Var(\sigma^2_{t+l-1}|J_t) = H_1^l + H_2^l\,\sigma_t^2 + H_3^l\,\sigma_t^4.$$

Proposition C.1 shows how the conditional variance (conditionally on earlier information) of the conditional variance of a SR-SARV(1) process is expressed in terms of the past conditional variance if this SR-SARV(1) process has a quadratic variance of variance. Since the quadratic specification is preserved in this basic temporal aggregation framework, it is also preserved in a more general temporal aggregation framework as the one we studied in Section 4.

Proof of Proposition C.1: For $l = 2$, there is nothing to do from the hypothesis. For $l \geq 3$, from Proposition 4.2, it is sufficient to show that $E\left(\sigma^4_{t+l-1}|J_t\right) = h_1^l + h_2^l\,\sigma_t^2 + h_3^l\,\sigma_t^4$ with h_i^l, h_2^l and $h_3^l \in \mathbb{R}$.
Since (f_t) follows a SR-SRAV(1) model, $Var(\sigma^2_{t+1}|J_t) = \alpha + \beta\,\sigma_t^2 + \delta\,\sigma_t^4$ implies $E\left(\sigma^4_{t+1}|J_t\right) = A_1 + A_2\,\sigma_t^2 + A_3\,\sigma_t^4$ (A_1, A_2 and $A_3 \in \mathbb{R}$). Hence, $E\left(\sigma^4_{t+l-1}|J_{t+l-2}\right) = A_1 + A_2\,\sigma^2_{t+l-2} + A_3\,\sigma^4_{t+l-2}$.
We will get the expected result by a backward iteration of the last equality:

$$
\begin{aligned}
E(\sigma^4_{t+l-1}|J_t) &= A_1 + A_2 \underbrace{E(\sigma^2_{t+l-2}|J_t)}_{\equiv E_{t,l-2}} + A_3\, E(\sigma^4_{t+l-2}|J_t) \\
&= A_1 + A_2 E_{t,l-2} + A_3 \left(A_1 + A_2\, E_{t,l-3} + A_3\, E(\sigma^4_{t+l-3}|J_t)\right) \\
&= A_1(1 + A_3) + A_2(E_{t,l-2} + A_3 E_{t,l-3}) + A_3^2\, E(\sigma^4_{t+l-3}|J_t) \\
&= \ldots \\
&= A_1\left(1 + A_3 + A_3^2 + \cdots + A_3^{l-2}\right) + A_2\left(E_{t,l-2} + A_3 E_{t,l-3} + \cdots + A_3^{l-2} E_{t,0}\right) + A_3^{l-1}\sigma_t^4 \\
&= A_1\left(1 + A_3 + A_3^2 + \cdots + A_3^{l-2}\right) + A_2\left[1 - \gamma^{l-2} + \gamma^{l-2}\sigma_t^2 + A_3(1 - \gamma^{l-3} + \gamma^{l-3}\sigma_t^2)\right. \\
&\qquad \left. + A_3^2(1 - \gamma^{l-4} + \gamma^{l-4}\sigma_t^2) + \cdots + A_3^{l-2}\sigma_t^2\right] + A_3^{l-1}\sigma_t^4 \\
&= A_1\left(1 + A_3 + A_3^2 + \cdots + A_3^{l-2}\right) \\
&\quad + A_2\left(1 + A_3 + \cdots + A_3^{l-3} - \gamma^{l-2}\left(1 + \tfrac{A_3}{\gamma} + \left(\tfrac{A_3}{\gamma}\right)^2 + \cdots + \left(\tfrac{A_3}{\gamma}\right)^{l-3}\right)\right) \\
&\quad + A_2\gamma^{l-2}\left(1 + \tfrac{A_3}{\gamma} + \left(\tfrac{A_3}{\gamma}\right)^2 + \cdots + \left(\tfrac{A_3}{\gamma}\right)^{l-2}\right)\sigma_t^2 + A_3^{l-1}\sigma_t^4 \\
E(\sigma^4_{t+l-1}|J_t) &= h_1^l + h_2^l\sigma_t^2 + h_3^l\sigma_t^4, \quad h_1^l, h_2^l \text{ and } h_3^l \in \mathbb{R}\square
\end{aligned}
$$

Chapter 2

GMM Overidentification Test with First Order Underidentification

1 Introduction

Moment condition-based models of the form $E[\phi(x,\theta)] = 0$, where $\phi(x,\theta)$ is a vector-valued nonlinear function of a random vector x and a parameter vector θ of size p, are very common in econometrics. In a well specified model, there exists a true parameter value θ_0 lying in a parameter set Θ such that $E[\phi(x,\theta_0)] = 0$. The model is identified when such a parameter value is unique in Θ. Together with some regularity assumptions, this identification condition guarantees the consistency of most of the estimators proposed in the literature for the parameter vector. These estimators include the minimum distance estimators and also the well-known generalized method of moment (GMM) estimators. While the consistency of these estimators relies on the identification condition, their $\sqrt{T}$-consistency and their asymptotic normality rely on the so-called rank condition. This is $\text{Rank}\{\partial E[\phi(x,\theta_0)]/\partial\theta'\} = p$. (See Andrews (1994).) In a moment condition-based model which is linear in the parameter, the identification condition is equivalent to the rank condition. However, in a nonlinear model, this equivalence no longer holds. A model nonlinear in the parameter could satisfy the identification condition without verifying the rank condition. Sargan (1983) refers to such a set up as first order lack of identification or first order underidentification.

In a first order underidentified model, the usual estimators are still consistent but higher order expansions are needed to get identifying approximation of the moment conditions. In this respect, when the moment condition model is identified, the first order underidentification context is located between the standard usual framework and the weak identification framework as treated by Staiger and Stock (1997) and Stock and Wright (2000). Note that, in the case of weak identification, not all of the parameters are consistently estimated. Sargan (1983) studies the instrumental variables (IV) estimator in the case of first order underidentification and finds that the IV estimator is neither $\sqrt{T}$-consistent nor asymptotically normally distributed.

In his seminal paper discussing the large sample properties of the GMM estimators, Hansen (1982) also proposes a test for the overidentifying moment restrictions. Under the null of valid moment conditions, the test statistic under some regularity conditions is asymptotically distributed as a χ^2 with a degree of freedom equal to the number of overidentifying moment restrictions. This asymptotic result also requires that the moment condition model is first order identified.

This paper has three main contributions. First, we discuss the asymptotic order of magnitude of the minimum distance estimators in the case of first order underidentification. Our results rely on a second order identification assumption for the model. This means that while there is first order lack of identification, a second order expansion of the moment conditions is useful to get a good approximation in the sense of parameter identification. As Sargan (1983), we derive our results by assuming that there is a set of parameters with respect to which the first derivative of the moment conditions evaluated at the true parameter value is of full rank and the set of the remaining parameters with respect to which the first derivative is null. We refer to the first set of parameters as those identified as the first order and the second set as those non-identified at the first order. Our results generalize the result by Sargan (1983) because we allow for any number of first order non-identified parameters. We find that not all of the components of the minimum distance estimator have the same rate of consistency. The components that estimate the parameters which are non-identified at the first order have a slower rate of consistency. Their asymptotic order of magnitude is $O_P(T^{-1/4})$ while the components that estimate the parameters which are identified at the first order have the usual $O_P(T^{-1/2})$ asymptotic order of magnitude even though they are not asymptotically normally distributed.

Second, we study the asymptotic behaviour of the Hansen's (1982) GMM overidentifying restrictions test statistic, J_T, in the context of first order underidentitifcation. In particular, we derive the asymptotic distribution of this test statistic when the rank of the moment conditions' first order derivative is $p-1$. We find that J_T is no longer asymptotically distributed as a χ^2_{H-p}, where H is the number of moment restrictions. Instead, J_T converges to a half and half mixture of χ^2_{H-p} and $\chi^2_{H-(p-1)}$. Obviously, the ignorance of the underidentification leads to an overrejecting test procedure.

Third, we apply this result to correct the test of common ARCH (Autoregressive conditional heteroskedasticity) factor in asset return processes proposed by Engle and Kozicki (1993). This is a leading test in the conditionally heteroskedastic factor models literature (Diebold and Nerlove (1989), Engle and Susmel (1993), King, Sentana and Wadhwani (1994), Fiorentini, Sentana and Shephard (2004) and Doz and Renault (2006)). This test translates the null of heteroskedasticity in the returns driven by heteroskedastic factors in terms of moment conditions and applies the Hansen's (1982) J-test for overidentifying moment restrictions. We show that even though the moment conditions on which Engle and Kozicki (1993) base their test identify the parameter of interest, they do not verify the first order identification condition. Therefore, the asymptotic results of Hansen (1982) are not suitable for their framework. We show that these moment conditions satisfy the second order identification condition that we introduce and the asymptotic distribution we derive for J_T under first order underidentification are applicable to the Engle and Kozicki's (1993) test. Actually, the test statistic in their test for common ARCH factor is asymptotically distributed as a mixture of two χ^2 instead of a χ^2 as they propose. Our findings even suggest that naive a χ^2 application as asymptotic distribution leads to a large overrejection which can even double the nominal size of the test.

The paper is organized as follows. In Section 2, we introduce the first and the second order identification concepts and we discuss how they affect the rate of convergence of the minimum distance estimators. We derive our asymptotic results in Section 3. In Section 4, we apply these results to the Engle and Kozicki (1993) test for common ARCH factor. Finally, Section 5 concludes. All proofs can be found in Appendix.

Throughout the paper $\|.\|$ denotes not only the usual Euclidean norm but also a matrix norm: $\|A\| = \{\text{Trace}(AA')\}^{1/2}$. By the Cauchy-Schwarz inequality, it has the useful property that, for any vector x and any conformable matrix A, $\|Ax\| \leq \|A\|\|x\|$.

2 First order underidentification and second order identification

2.1 General framework

We consider a general minimum distance estimation problem of an unknown vector θ of p parameters given as solution of H estimating equations:

$$\rho(\theta) = 0 \tag{1}$$

These estimating equations are assumed to identify the true unknown value θ^0 of θ by to the following assumptions:

Assumption 1 (Global Identification) *$\rho(\theta) = \{\rho_h(\theta)\}_{1 \leq h \leq H}$ is a continuous function defined on a compact parameter space $\Theta \subset \mathbb{R}^p$ such that for all θ in Θ: $\rho(\theta) = 0 \Leftrightarrow \theta = \theta^0$.*

Assumption 1 is maintained for the sake of expositional simplicity even though it could be easily relaxed by only assuming that θ_0 is a well-separated minimum of norm of $\rho(\theta)$ (see Van der Vaart (1998) page 46).

For the purpose of minimum distance estimation, a data set of size T will give us some sample counterparts of the estimating equations. More precisely, with time series notations, we consider that with a sample size T, corresponding to observations at dates $t = 1, 2, \ldots, T$ and for any possible value θ of the parameters, we have at our disposal a H-dimensional sample-based vector $\overline{\phi}_T(\theta) = \{\overline{\phi}_{h,T}(\theta)\}_{1 \leq h \leq H}$. In most cases, minimum distance estimation is akin to GMM estimation because $\overline{\phi}_T(\theta)$ is obtained as a sample mean

$$\overline{\phi}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \phi_t(\theta). \tag{2}$$

In any case, we define a minimum distance estimator for a given sequence of weighting matrices.

Definition 1 *A minimum distance estimator $\hat{\theta}_T$ of θ is defined as solution of*

$$\min_{\theta \in \Theta} \overline{\phi}'_T(\theta) \Omega_T \overline{\phi}_T(\theta),$$

where Ω_T is a sequence of symmetric positive definite matrices which converges when T goes to infinity to Ω, a positive definite matrix.

The asymptotic properties of a minimum distance estimator are classically deduced from the asymptotic behaviour of the sample counterpart $\overline{\phi}_T(\theta)$ of the estimating equations.

Assumption 2 (Well-behaved moments) (a) *$\overline{\phi}_T(\theta)$ converges in probability to $\rho(\theta)$, uniformly in $\theta \in \Theta$;* (b) *$\sqrt{T}\,\overline{\phi}_T(\theta^0)$ converges in distribution to a normal distribution with mean 0 and non-singular variance matrix $\Sigma(\theta^0)$.*

It is well-known (see e.g. Amemiya (1989)) that Assumption 2.a implies that any minimum distance estimator $\hat{\theta}_T$ is weakly consistent for θ^0. The asymptotic distribution of $\hat{\theta}_T$ is then usually deduced from a Taylor expansion of the first order conditions

$$\frac{\partial \overline{\phi}'_T}{\partial \theta}(\hat{\theta}_T) \Omega_T \sqrt{T} \overline{\phi}_T(\hat{\theta}_T) = 0. \tag{3}$$

Of course, this kind of approach is based on the maintained assumption below.

Assumption 3 (Differentiability of estimating equations) *$\overline{\phi}_T(\theta)$ and $\rho(\theta)$ are continuously differentiable on the interior $\mathring{\Theta}$ of Θ, $\theta^0 \in \mathring{\Theta}$ and $\partial \overline{\phi}_T(\theta)/\partial \theta'$ converges to $\partial \rho(\theta)/\partial \theta'$, uniformly on $\theta \in \mathring{\Theta}$.*

2.2 First order underidentification

Asymptotic normality of the minimum distance estimator $\hat{\theta}_T$ is usually obtained by the joint argument that $\sqrt{T}(\hat{\theta}_T - \theta^0) = O_P(1)$ and then, the following first order approximation of the first order conditions (3) is valid.

$$\frac{\partial \overline{\phi}'_T}{\partial \theta}(\hat{\theta}_T)\Omega_T\sqrt{T}\overline{\phi}_T(\theta^0) + \frac{\partial \overline{\phi}'_T}{\partial \theta}(\hat{\theta}_T)\Omega_T\frac{\partial \overline{\phi}_T}{\partial \theta'}(\theta^0)\sqrt{T}(\hat{\theta}_T - \theta^0) = o_P(1)$$

which, by the above assumptions, can be rewritten

$$\frac{\partial \rho'}{\partial \theta}(\theta^0)\Omega\sqrt{T}\overline{\phi}_T(\theta^0) + \left[\frac{\partial \rho'}{\partial \theta}(\theta^0)\Omega\frac{\partial \rho}{\partial \theta'}(\theta^0)\right]\sqrt{T}(\hat{\theta}_T - \theta^0) = o_P(1) \tag{4}$$

The asymptotic normal distribution of $\sqrt{T}(\hat{\theta}_T - \theta^0)$ is then deduced from (4) which characterizes $\sqrt{T}(\hat{\theta}_T - \theta^0)$ as asymptotically linear function of the Gaussian vector $\overline{\phi}_T(\theta^0)$. However, it is worth reminding that the whole argument above rests upon the maintained assumption of non-singularity of the matrix $\left[\frac{\partial \rho'}{\partial \theta}(\theta^0)\Omega\frac{\partial \rho}{\partial \theta'}(\theta^0)\right]$, that is on the so-called first order identification condition

$$Rank\frac{\partial \rho}{\partial \theta'}(\theta^0) = p.$$

The focus of our interest in this paper is a case of first order underidentification where

$$Rank\frac{\partial \rho}{\partial \theta'}(\theta^0) < p.$$

Assumption 4A (Rank deficiency).

$$Rank\left\{\frac{\partial \rho}{\partial \theta'}(\theta^0)\right\} = r < p.$$

For the sake of expositional simplicity, we even assume that

Assumption 4B (Rank deficiency for known directions). $\theta = (\theta'_1, \theta'_2)'$ with $\dim\theta_1 = r$, $\dim\theta_2 = p - r$ such that

$$Rank\left\{\frac{\partial \rho}{\partial \theta'_1}(\theta^0)\right\} = r \quad \text{and} \quad \frac{\partial \rho}{\partial \theta'_2}(\theta) = 0.$$

Note that it is actually always possible to replace Assumption 4A by Assumption 4B by a change of basis in $\mathbb{R}^p$. The problem is that the required change of basis must be estimated and may in particular depend upon the unknown true parameter value θ^0 of θ. Since the focus of our interest is testing for overidentification, the estimation issue raised by relaxing Assumption 4B by maintaining only the more general Assumption 4A is beyond the scope of this paper. As explained in section 3, the difference between the two assumptions is immaterial as far as asymptotic distribution of the J-test statistic of overidentification is concerned.

The key intuition is that, when $\partial\rho(\theta)/\partial\theta_2' = 0$, we lose the linear one-to-one asymptotic relationship between $\sqrt{T}\,\overline{\phi}_T(\theta^0)$ and $\sqrt{T}(\hat{\theta}_T - \theta^0)$, so that both the property $\sqrt{T}(\hat{\theta}_T - \theta^0) = O_P(1)$ and the validity of the Taylor expansion (4) are no longer guaranteed. We must actually consider a higher order Taylor expansion

$$\frac{\partial\rho'}{\partial\theta}(\theta^0)\Omega\left\{\sqrt{T}\overline{\phi}_T(\theta^0) + \frac{\partial\overline{\phi}_T}{\partial\theta'}(\theta^0)\sqrt{T}(\hat{\theta}_T - \theta^0) + \frac{1}{2}\left[T^{1/4}(\hat{\theta}_T - \theta^0)'\frac{\partial^2\overline{\phi}_{h,T}}{\partial\theta\partial\theta'}(\theta^0)T^{1/4}(\hat{\theta}_T - \theta^0)\right]_{1\leq h\leq H}\right\} = \xi_T \tag{5}$$

with possibly $\xi_T = O_P(1)$. The intuition behind this result would be the following. On the one hand, since $\partial\overline{\phi}_T(\theta^0)/\partial\theta'$ converges to $\partial\rho(\theta^0)/\partial\theta' = [\partial\rho(\theta^0)/\partial\theta_1', 0]$ we still can take advantage of the invertibility of $\{\partial\rho'(\theta^0)/\partial\theta_1\}\Omega\{\partial\rho(\theta^0)/\partial\theta_1'\}$ to show that $\sqrt{T}(\hat{\theta}_{1T} - \theta_1^0)$ is $O_P(1)$, insofar as the other terms in the expansion (5) are $O_P(1)$. On the other hand, the fact that all of the terms in expansion (5) and ξ_T itself are $O_P(1)$ will be compatible with $T^{1/4}(\hat{\theta}_{2T} - \theta_2^0) = O_P(1)$. Then, in the quadratic term of expansion (5), all of the terms will be negligible except the vector of coefficients

$$T^{1/4}(\hat{\theta}_{2T} - \theta_2^0)'\frac{\partial^2\overline{\phi}_{h,T}}{\partial\theta_2\partial\theta_2'}(\theta^0)T^{1/4}(\hat{\theta}_{2T} - \theta_2^0),$$

$h = 1, 2, \ldots, H$. This is typically the kind of situation we want to study in this paper.

A couple of additional regularity conditions will be used to justify the argument above.

Assumption 5 (Higher order regularity of the estimating equations) (a) $\sqrt{T}\{\partial\overline{\phi}_T(\theta^0)/\partial\theta_2'\} = O_P(1)$; (b) $\overline{\phi}_T(\theta)$ *and* $\rho(\theta)$ *are twice continuously differentiable on the interior* $\mathring{\Theta}$ *of* Θ *and for all* $h = 1, 2, \ldots, H$, $\partial^2\overline{\phi}_{h,T}(\theta)/\partial\theta\partial\theta'$ *converges to* $\partial^2\rho_h(\theta)/\partial\theta\partial\theta'$, *uniformly on* $\theta \in \mathring{\Theta}$.

Assumption 5 is an extension of the assumption that is usually made to obtain the limiting distribution of GMM estimators. Like Kleibergen (2005) we need in particular to complete the central limit theorem for the moment conditions by a similar assumption about the limit behavior of the Jacobian matrix of these moment conditions.

Then, the above arguments lead us to the following first result.

Proposition 2.1 *If Assumptions 1-3, 4B and 5 hold and* $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)=O_P(1)$, *then*

$$T^{1/2}(\hat{\theta}_{1T}-\theta_1^0)=O_P(1).$$

Note that the proof of Proposition 2.1 also shows that we cannot in general derive an asymptotic normal distribution for $T^{1/2}(\hat{\theta}_{1T}-\theta_1^0)$. As already pointed out by Sargan (1983) in a particular case, the term quadratic in $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)$ will actually contaminate the asymptotic distribution of $T^{1/2}(\hat{\theta}_{1T}-\theta_1^0)$.

The key point is now to explain why we expect that $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)=O_P(1)$. In the same way, in a standard setting, it is referred to the first order identification to justify the property $T^{1/2}(\hat{\theta}_T-\theta^0)=O_P(1)$, we have now to introduce the concept of second order identification.

2.3 Second order identification

As explained in the previous subsection, we have in mind a setting where the second order Taylor expansion (5) of the first order conditions subsumes in

$$\begin{aligned}\frac{\partial\rho'}{\partial\theta}(\theta^0)\Omega\Big\{\sqrt{T}\bar{\phi}_T(\theta^0)+\frac{\partial\rho}{\partial\theta_1'}(\theta^0)\sqrt{T}(\hat{\theta}_{1T}-\theta_1^0)&\\ +\frac{1}{2}\left[T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)'\frac{\partial^2\rho_h}{\partial\theta_2\partial\theta_2'}(\theta^0)T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)\right]_{1\leq h\leq H}\Big\}&=o_P(1).\end{aligned}\tag{6}$$

Since, as usual, the asymptotic probability distribution of the estimator $\hat{\theta}_T$ will be obtained by *solving* in $\hat{\theta}_T$ the Taylor expansion of the first order conditions, we need to introduce the following identification assumption.

Assumption 6 (Second order identification) *For any* u *in* $\mathbb{R}^r$ *and* v *in* $\mathbb{R}^{p-r}$ *we have*

$$\left(\frac{\partial\rho}{\partial\theta_1'}(\theta^0)u+\frac{1}{2}\left[v'\frac{\partial^2\rho_h}{\partial\theta_2\partial\theta_2'}(\theta^0)v\right]_{1\leq h\leq H}=0\right)\Rightarrow(u=0\quad\text{and}\quad v=0).$$

Note that Assumption 6 reinforces Assumption 4B. Not only the r columns of $\partial\rho(\theta^0)/\partial\theta_1'$ are assumed to be linearly independent, to get

$$\left(\frac{\partial\rho}{\partial\theta_1'}(\theta^0)u=0\right)\Rightarrow(u=0)$$

but in addition, we assume some kind of linear independence between the columns of $\partial\rho(\theta^0)/\partial\theta_1'$ and vectors built from the second derivatives of ρ with respect to θ_2. This assumption may look a bit ad hoc at first sight but is well suited for the examples of application we have in mind. While an important example will be detailed in Section 4, let us first give a simple toy example.

A toy example[1]

Assume we observe two stationary and ergodic time series, x_t and y_t, $t = 1, 2, \ldots, T$ of real random variables both with zero-mean. We want to characterize the conditional mean $E[y_t|x_t] = \theta_2 x_t$ not by the classical orthogonality conditions but by the fact that y_t is conditionally homoskedastic given x_t that is

$$E\left\{(y_t - \theta_2 x_t)^2 | x_t\right\} = \theta_1,$$

where θ_1 is constant independent of x_t. It is natural to choose as estimating equations

$$\rho(\theta) = \left\{E[(y_t - \theta_2 x_t)^2 - \theta_1], E[x_t(y_t - \theta_2 x_t)^2 - \theta_1 x_t]\right\}'.$$

The sample counterparts are then trivial to get. Then $\partial\rho(\theta^0)/\partial\theta_1 = (-1, -Ex_t)' = (-1, 0)'$ while $\partial\rho(\theta^0)/\partial\theta_2 = \left(-2E[x_t(y_t - \theta_2^0 x_t)], -2E[x_t^2(y_t - \theta_2^0 x_t)]\right)' = 0$. We have then typically a case of first order underidentification. However,

$$\left[\frac{\partial^2 \rho_h}{\partial \theta_2^2}(\theta^0)\right]_{1 \leq h \leq 2} = 2\begin{bmatrix} Ex_t^2 \\ Ex_t^3 \end{bmatrix}.$$

Therefore, to get second order identification, we have to check that

$$\left(\begin{bmatrix} -1 \\ 0 \end{bmatrix} u + \begin{bmatrix} Ex_t^2 \\ Ex_t^3 \end{bmatrix} v^2 = 0\right) \Rightarrow (u = v = 0).$$

This is clearly the case if and only if $Ex_t^3 \neq 0$. Then it is also easy to check that the global identification assumption provided by Assumption 1 is fulfilled. We have a case of identification through higher order moments. (See also Bonhomme and Robin (2006) for other applications of this concept.)

Let us now sketch the intuition of the reason why the second order identification assumption given by Assumption 6 will allow us to show that $T^{1/4}(\hat{\theta}_{2T} - \theta_2^0) = O_P(1)$. This assumption involves a vector

$$\left[v' \frac{\partial^2 \rho_h}{\partial \theta_2 \partial \theta_2'}(\theta^0) v\right]_{1 \leq h \leq H}$$

[1]We thank Manuel Arellano for having suggested this toy example in a private communication.

which is a quadratic function of v. Since our minimum distance estimation is defined through a weighting matrix Ω_T, it is rather convenient to consider the rescaled vector

$$\Delta(v) = \Omega^{1/2}\left[v'\frac{\partial^2 \rho_h}{\partial\theta_2\partial\theta_2'}(\theta^0)v\right]_{1\leq h\leq H}.$$

For the same reason, we introduce the rescaled Jacobian matrix

$$Z_1 = \Omega^{1/2}\frac{\partial\rho}{\partial\theta_1'}(\theta^0).$$

By assumption, the r columns of Z_1 are linearly independent, which allows us to write the projection matrix of orthogonal projection on the space spanned by the columns of Z_1 as $P_{Z_1} = Z_1(Z_1'Z_1)^{-1}Z_1'$. Similarly, $M_{Z_1} = Id_H - P_{Z_1}$ denotes the orthogonal projection onto the orthogonal space.

A standard setting of first order identification would ensure, in addition to the assumption that Z_1 is full column rank that

$$\left(Z_1 u + \Omega^{1/2}\frac{\partial\rho}{\partial\theta_2'}(\theta^0)v = 0\right) \Rightarrow (u = 0 \quad \text{and} \quad v = 0).$$

It is worth noting that this last assumption would amount to saying that

$$\left\|M_{Z_1}\Omega^{1/2}\frac{\partial\rho}{\partial\theta_2'}(\theta^0)v\right\| \geq \gamma\|v\|$$

for some positive number γ.

The following lemma provides an analogous result about Assumption 6. Let $Z_1 = \Omega^{1/2}(\partial\rho(\theta^0)/\partial\theta_1')$, $\Delta(v) = \Omega^{1/2}\left[v'(\partial^2\rho_h(\theta^0)/\partial\theta_2\partial\theta_2')v\right]_{1\leq h\leq H}$ and $M_{Z_1} = Id_H - Z_1(Z_1'Z_1)^{-1}Z_1'$.

Lemma 2.1 *If the second order identification condition given by Assumption 6 holds, then there exists a positive number γ such that, for all v in $\mathbb{R}^{p-r}$:*

$$\|M_{Z_1}\Delta(v)\| \geq \gamma\|v\|^2.$$

Note that the identification term for θ_2 is now bounded away from zero like $\|v\|^2$ instead of $\|v\|$ in the standard setting. This is the reason why we will be only able to show that $T^{1/2}\|\hat{\theta}_{2T} - \theta_2^0\|^2 = O_P(1)$ (or $T^{1/4}\|\hat{\theta}_{2T} - \theta_2^0\| = O_P(1)$) instead of the standard property $T^{1/2}\|\hat{\theta}_{2T} - \theta_2^0\| = O_P(1)$.

3 Asymptotic theory of GMM overidentification test under second order identification

3.1 Standardization of moment conditions and parameterization of the null space

We are interested in the asymptotic distribution of Hansen's J-test statistic

$$J_T = T \min_{\theta \in \Theta} \bar{\phi}_T'(\theta)\Omega_T \bar{\phi}_T(\theta),$$

where, by Assumptions 1, 2, 3 and 4A

$$\left\{ \begin{array}{rcl} p\lim \bar{\phi}_T(\theta) & = & \rho(\theta) \\ \sqrt{T}\bar{\phi}_T(\theta^0) & \xrightarrow{d} & \mathcal{N}(0, \Sigma(\theta^0)) \\ \Omega_T & \xrightarrow{P} & [\Sigma(\theta^0)]^{-1} \\ \text{Rank}\frac{\partial \rho}{\partial \theta'}(\theta^0) & = & r < p = \dim(\theta). \end{array} \right.$$

Let R_2 be a $(p, p-r)$-matrix of rank $p-r$ such that

$$\frac{\partial \rho}{\partial \theta'}(\theta^0) R_2 = 0,$$

and R_1 be a (p, r)-matrix of rank r such that $R = [R_1 \colon R_2]$ is a non-singular matrix.

Let

$$\eta = R^{-1}\theta$$

and

$$\left\{ \begin{array}{rcl} \tilde{\rho}(\eta) & = & \Sigma(\theta^0)^{-1/2}\rho(R\eta) \\ \eta^0 & = & R^{-1}\theta^0. \end{array} \right.$$

Then

$$\frac{\partial \tilde{\rho}}{\partial \eta'}(\eta^0) = \Sigma(\theta^0)^{-1/2}\frac{\partial \rho}{\partial \theta'}(\theta^0)\,[R_1 \colon R_2] = \Sigma(\theta^0)^{-1/2}\left[\frac{\partial \rho}{\partial \theta'}(\theta^0)R_1 \colon 0\right].$$

Therefore

$$\text{Rank}\frac{\partial \tilde{\rho}}{\partial \eta_1'}(\eta^0) = r \quad \text{and} \quad \frac{\partial \tilde{\rho}}{\partial \eta_2'}(\eta^0) = 0.$$

Let $\bar{\psi}_T(\eta) = \Omega_T^{-1/2}\bar{\phi}_T(R\eta)$. Then

$$J_T = T \min_{\eta} \bar{\psi}_T'(\eta)\bar{\psi}_T(\eta),$$

with

$$\left\{ \begin{array}{rcl} p\lim \bar{\psi}_T(\eta) & = & \tilde{\rho}(\eta) \\ \sqrt{T}\bar{\psi}_T(\eta^0) & \xrightarrow{d} & \mathcal{N}(0, Id). \end{array} \right.$$

Of course $\hat{\eta}_T = \arg\min_\eta \bar{\psi}'_T(\eta)\bar{\psi}_T(\eta)$ is not a feasible estimator since the matrix R is unknown and the function $\bar{\psi}_T$ cannot be directly computed from data. However, for the purpose of characterizing the asymptotic probability distribution of the J-test, it is immaterial to assume that

$$\left\{ \begin{array}{rcl} \eta = \theta, & \Sigma(\theta) = Id & \text{and thus,} \\ \bar{\psi}_T(\eta) & = & \bar{\phi}_T(\theta) \\ J_T & = & T \min_\theta \bar{\phi}'_T(\theta)\bar{\phi}_T(\theta) = T\bar{\phi}'_T(\hat{\theta})\bar{\phi}_T(\hat{\theta}). \end{array} \right. \tag{7}$$

For the sake of expositional simplicity, this framework will be maintained throughout this section 3 in the context of Assumptions 1 to 6. As announced in Section 2, the knowledge of the directions of rank deficiency allows to characterize the rate of convergence of the various components of $\hat{\theta}_T$.

Proposition 3.1 *Under Assumptions 1 to 6, we have*

$$T^{1/4}(\hat{\theta}_{2T} - \theta_2^0) = O_P(1) \quad \textit{and} \quad T^{1/2}(\hat{\theta}_{1T} - \theta_1^0) = O_P(1).$$

Let us remind that Proposition 3.1 is not really useful for estimation purpose since it rests upon a rotation in the parameter space to isolate the directions of fast convergence. While a convergence rate for $\hat{\theta}_{2T}$ faster than $T^{1/4}$ would allow to consistently estimate these directions (see Antoine and Renault (2007)), it may not work if the convergence rate is only $T^{1/4}$. The burden comes from the fact that second order estimation errors about the direction in the parameter space, of order $(T^{1/4})^2$, will contaminate the asymptotic distribution of $\hat{\theta}_{1T}$.

However, it is worth noting that $T^{1/4}$ is only a lower bound for the convergence rate of $\hat{\theta}_{2T}$ while its convergence is going to be faster in some regions of the sample space. This is due to these regions of faster convergence that the J-test statistic is going to display a non standard asymptotic behaviour as mixture of chi-squares.

3.2 Overidentification test statistic as a mixture

Let us introduce the matrices

$$Z_1(\theta) = \frac{\partial \rho}{\partial \theta'_1}(\theta) \quad \text{and} \quad Z_2(\theta) = \frac{\partial \rho}{\partial \theta'_2}(\theta).$$

By assumption, $Z_2(\theta^0) = 0$ while $Z_1(\theta^0) = Z_1$ is a full-column rank matrix. We can then consider the projection matrix

$$M_{Z_1} = Id_H - Z_1(Z'_1 Z_1)^{-1} Z'_1.$$

Taking into account that $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)$ and $T^{1/2}(\hat{\theta}_{1T}-\theta_1^0)$ are $O_P(1)$, the proof of Proposition 3.1 gives

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = M_{Z_1}\sqrt{T}\bar{\phi}_T(\theta^0) + M_{Z_1}\sqrt{T}\Delta(\hat{\theta}_{2T}-\theta_2^0) + o_P(1)$$

where $\hat{\theta}_{2T}$ is characterized by the first order conditions

$$Z_2'(\hat{\theta}_T)\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = o_P(1)$$

that is

$$Z_2'(\hat{\theta}_T)M_{Z_1}\sqrt{T}\bar{\phi}_T(\theta^0) + Z_2'(\hat{\theta}_T)M_{Z_1}\sqrt{T}\Delta(\hat{\theta}_{2T}-\theta_2^0) = o_P(1). \tag{8}$$

We will show now that $\hat{\theta}_{2T}$, solution of (8), may be either such that $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)$ follows asymptotically a non-degenerate solution or in contrary $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0) = o_P(1)$. To see this, let

$$\mathcal{M}_T = \left[\frac{\partial^2\rho'(\theta^0)}{\partial\theta_{2i}\partial\theta_{2j}}M_{Z_1}\sqrt{T}\bar{\phi}(\theta^0)\right]_{1\le i,j\le p-r} \quad \text{and} \quad f_T = \frac{\partial^2\rho'}{\partial\theta_{21}^2}(\theta^0)M_{Z_1}\sqrt{T}\bar{\phi}(\theta^0).$$

We next show that $\mathcal{M}_T$ is positive semi definite (p.s.d) if and only if $T^{1/4}(\hat{\theta}_{2T}-\theta_2)$ is asymptotically degenerate.

If $\mathcal{M}_T$ is p.s.d, by Equation (13),

$$J_T - T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) = 2T\bar{\phi}'(\theta^0)M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0) + T\Delta'(\hat{\theta}_{2T}-\theta_2^0)M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0) + o_P(1).$$

Note that for any $v \in \mathbb{R}^{p-r}$ and any $u \in \mathbb{R}^H$,

$$\begin{aligned}
\Delta'(v)u &= \textstyle\sum_{k=1}^H \left(v'\frac{\partial^2\rho_k(\theta^0)}{\partial\theta_2\partial\theta_2'}v\right)u_k = \sum_{k=1}^H\sum_{i,j=1}^{p-r} v_iv_j\frac{\partial^2\rho_k(\theta^0)}{\partial\theta_{2i}\partial\theta_{2j}}u_k \\
&= \textstyle\sum_{i,j=1}^{p-r} v_iv_j\left(\sum_{k=1}^H \frac{\partial^2\rho_k(\theta^0)}{\partial\theta_{2i}\partial\theta_{2j}}u_k\right) = \sum_{i,j=1}^{p-r} v_iv_j\left(\frac{\partial^2\rho'(\theta^0)}{\partial\theta_{2i}\partial\theta_{2j}}u\right) = v'\mathcal{M}_uv,
\end{aligned}$$

where

$$\mathcal{M}_u = \left(\frac{\partial^2\rho'(\theta^0)}{\partial\theta_{2i}\partial\theta_{2j}}u\right)_{1\le i,j\le p-r}$$

Hence,

$$\Delta'(\hat{\theta}_{2T}-\theta_2^0)M_{Z_1}\sqrt{T}\bar{\phi}(\theta^0) = (\hat{\theta}_{2T}-\theta_2^0)'\mathcal{M}_T(\hat{\theta}_{2T}-\theta_2^0).$$

Since $\mathcal{M}_T$ is p.s.d, $T\Delta'(\hat{\theta}_{2T}-\theta_2^0)M_{Z_1}\bar{\phi}(\theta^0) \ge 0$ for any $\hat{\theta}_{2T}$.

As $J_T - T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) \le 0$, we necessarily have

$$T\Delta'(\hat{\theta}_{2T}-\theta_2^0)M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0) = o_P(1).$$

In other words, $\|\sqrt{T}M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0)\| = o_P(1)$ and, therefore, by Lemma 2.1, $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0) = o_P(1)$.

Conversely, if $\mathcal{M}_T$ is not p.s.d, there exists a unit norm sequence of random vectors $\tilde{e} \in \mathbb{R}^{p-r}$ such that

$$\tilde{e}'\mathcal{M}_T\tilde{e} < 0.$$

The necessary second order condition for an interior solution for a minimization problem implies that

$$\tilde{e}'\left\{\left[\frac{\partial^2\bar{\phi}'(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\bar{\phi}(\hat{\theta}_T)\right]_{1\le i,j\le p-r} + \left[\frac{\partial\bar{\phi}'}{\partial\theta}(\hat{\theta}_T)\frac{\partial\bar{\phi}}{\partial\theta'}(\hat{\theta}_T)\right]\right\}\tilde{e} \ge 0.$$

This yields

$$\sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial^2\bar{\phi}'(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\bar{\phi}(\hat{\theta}_T) + \sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial\bar{\phi}'}{\partial\theta_i}(\hat{\theta}_T)\frac{\partial\bar{\phi}}{\partial\theta_j}(\hat{\theta}_T) \ge 0.$$

By the usual expansions,

$$\frac{\partial\bar{\phi}}{\partial\theta_i}(\hat{\theta}_T) = \frac{\partial^2\bar{\phi}(\bar{\theta})}{\partial\theta_i\partial\theta'}(\hat{\theta}_{2T}-\theta_2^0) + O_P(T^{-1/2}).$$

Therefore,

$$\frac{\partial\bar{\phi}'}{\partial\theta_i}(\hat{\theta}_T)\frac{\partial\bar{\phi}}{\partial\theta_j}(\hat{\theta}_T) = (\hat{\theta}_{2T}-\theta_2^0)'\frac{\partial^2\bar{\phi}'(\bar{\theta})}{\partial\theta_i\partial\theta}\frac{\partial^2\bar{\phi}(\bar{\theta})}{\partial\theta_j\partial\theta'}(\hat{\theta}_{2T}-\theta_2^0) + o_P(T^{-1/2}).$$

On the other hand,

$$\frac{\partial^2\bar{\phi}'(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\bar{\phi}(\hat{\theta}_T) = \frac{\partial^2\bar{\phi}'(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\Big(M_{Z_1}\bar{\phi}(\theta^0) + M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0)\Big) + o_P(T^{-1/2}).$$

As a result and since $\tilde{e} = O_P(1)$,

$$\sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial\bar{\phi}'}{\partial\theta_i}(\hat{\theta}_T)\frac{\partial\bar{\phi}}{\partial\theta_j}(\hat{\theta}_T) = \sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j(\hat{\theta}_{2T}-\theta_2^0)'\frac{\partial^2\rho'(\theta^0)}{\partial\theta_i\partial\theta}\frac{\partial^2\rho(\theta^0)}{\partial\theta_j\partial\theta'}(\hat{\theta}_{2T}-\theta_2^0) + o_P(T^{-1/2})$$

$$\sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial^2\bar{\phi}'(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\bar{\phi}(\hat{\theta}_T) = \sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial^2\rho'(\theta^0)}{\partial\theta_i\partial\theta_j}M_{Z_1}\bar{\phi}(\theta^0) + \sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial^2\rho'(\theta^0)}{\partial\theta_i\partial\theta_j}M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0) + o_P(T^{-1/2}).$$

The last inequality translates into

$$\begin{aligned} -\textstyle\sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial^2\rho'(\theta^0)}{\partial\theta_i\partial\theta_j}M_{Z_1}\bar{\phi}(\theta^0) \quad &\le \quad \textstyle\sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j\frac{\partial^2\rho'(\theta^0)}{\partial\theta_i\partial\theta_j}M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0) \\ &\quad + \textstyle\sum_{i,j=1}^{p-r}\tilde{e}_i\tilde{e}_j(\hat{\theta}_{2T}-\theta_2^0)'\frac{\partial^2\rho'(\theta^0)}{\partial\theta_i\partial\theta}\frac{\partial^2\rho(\theta^0)}{\partial\theta_j\partial\theta'}(\hat{\theta}_{2T}-\theta_2^0) + o_P(T^{-1/2}). \end{aligned}$$

Thus

$$-\tilde{e}'\mathcal{M}_T\tilde{e} \le A\sqrt{T}\|\hat{\theta}_{2T}-\theta_2^0\|^2 + o_P(1),$$

for some $A > 0$. Hence

$$0 < -\tilde{e}'\mathcal{M}_T\tilde{e} \leq A\sqrt{T}\|\hat{\theta}_{2T} - \theta_2^0\|^2 + o_P(1).$$

This shows that $\sqrt{T}\|\hat{\theta}_{2T} - \theta_2^0\|^2$ is not degenerate, thanks to the asymptotic Gaussianity of $\mathcal{M}_T$, and so is $T^{1/4}(\hat{\theta}_{2T} - \theta_2^0)$.

Let us denote by C the part of the fundamental space where $\mathcal{M}_T$ is p.s.d. Remark first that $f_T \geq 0$ is necessary for $\mathcal{M}_T$ to be p.s.d. Obviously,

$$f_T \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial^2 \rho'}{\partial \theta_{21}^2}(\theta^0) M_{Z_1} \frac{\partial^2 \rho}{\partial \theta_{21}^2}(\theta^0)\right).$$

Clearly, $M_{Z_1}\frac{\partial^2 \rho}{\partial \theta_{21}^2}(\theta^0) = M_{Z_1}\Delta(e_1)$, $e_1 = (1, 0, \ldots, 0)'$. Since $e_1 \neq 0$, by the proof of Lemma 2.1, $M_{Z_1}\Delta(e_1) \neq 0$ and therefore, f_T is asymptotically non degenerate,

$$\lim_{T\to\infty} \text{Prob}(C) \leq \lim_{T\to\infty} \text{Prob}(f_T \geq 0) = \frac{1}{2}.$$

This probability limit is also positive because the positive semi definiteness of $\mathcal{M}_T$ amounts to $p - r$ inequality constraints on $M_{Z_1}\sqrt{T}\bar{\phi}(\theta^0)$ which is asymptotically normally distributed with $H - r$ ($> p - r$) degree of freedom. In particular, for $p - r = 1$,

$$\lim_{T\to\infty} \text{Prob}(C) = \lim_{T\to\infty} \text{Prob}(f_T \geq 0) = \frac{1}{2}.$$

On the other hand, it is worth noting that if $T^{1/4}(\hat{\theta}_{2T} - \theta_2^0) = o_P(1)$, the above expansion of the moment conditions collapses in

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = M_{Z_1}\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1).$$

In other words

$$J_T = T\bar{\phi}_T'(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) = T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) + o_P(1)$$

is asymptotically distributed as a χ^2_{H-r} since M_{Z_1} is an orthogonal projection matrix on a subspace of dimension $(H - r)$.

The above discussion leads us to state the main result of this paper.

Theorem 3.1 *The overidentification J-test statistic J_T associated to the estimating equations*

$$\rho(\theta) = 0$$

is, with probability q, asymptotically distributed as χ^2_{H-r}, where

$$\begin{cases} H &= \dim \rho(\theta) \\ r &= Rank\frac{\partial \rho}{\partial \theta'}(\theta^0) \\ q &= \lim_{T\to\infty} \text{Prob}\,(\mathcal{M}_T \text{ is p.s.d}) \,. \end{cases}$$

In particular, if $r = p-1$, $q = 1/2$ and J_T is asymptotically distributed as the mixture

$$\frac{1}{2}\chi^2_{H-p} + \frac{1}{2}\chi^2_{H-p+1}.$$

The discussion above makes rather clear the interpretation of Theorem 3.1. The reason why, with probability q, the asymptotic distribution of J_T is χ^2_{H-r}, instead of χ^2_{H-p}, is because only the r constraints corresponding to the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$ are really binding in that case. It turns out that the $(p-r)$ directions in the parameter space corresponding to the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$, when they are estimated at a rate faster than $T^{1/4}$ (which is the case with probability q), do not play any more role in the overidentification test. In this case, the asymptotic distribution of the overidentification test statistic is as if θ_2^0 were known.

4 Application to the test for common ARCH factor

In this section, we reexamine the test for common conditionally heteroskedastic factor proposed by Engle and Kozicki (1993). Two asset return processes are said to have a common conditionally heteroskedastic factor if each of them is conditionally heteroskedastic and there exists a linear combination of them which is not conditionally heteroskedastic. Engle and Kozicki (1993) propose a test for common conditionally heteroskedastic factor in two steps. First, the Engle's (1982) Lagrange multiplier test for conditional heteroskedasticity is performed on each process and when both have evidence of conditional heteroskedasticity, a second test is needed. At this second step, they propose a test that investigate whether there is a linear combination of the two processes which is not conditionally heteroskedastic (see Engle and Kozicki (1993) and Engle and Susmel (1993)). Our point is related to this second step test.

Let us consider the bivariate random process Y_{t+1} whose both components are conditionally heteroskedastic. The components of Y_{t+1} share a common heteroscedastic factor if Y_{t+1} has the following representation

$$Y_{t+1} = \lambda f_{t+1} + U_{t+1}, \tag{9}$$

where f_{t+1} is the unobserved common conditionally heteroskedastic factor, $\lambda \in \mathbb{R}^2$ the vector of factor loadings and $U_{t+1} \in \mathbb{R}^2$ the vector of idiosyncratic shocks.

Let J_t be the increasing filtration containing the available information at the date t. The practical assumptions are

$$\begin{array}{llllll} E(f_{t+1}|J_t) & = 0 & E(U_{t+1}|J_t) & = 0 & Var(f_{t+1}|J_t) & = \sigma_t^2 \\ E(\sigma_t^2) & = 1 & Var(U_{t+1}|J_t) & = \Omega & E(f_{t+1}U_{t+1}|J_t) & = 0. \end{array} \tag{10}$$

It is assumed, in addition, that Ω is positive definite and $Var(\sigma_t^2) > 0$. These assumptions imply that any other single heteroscedastic factor decomposition of Y_{t+1} has factor loadings proportional to λ (see Doz and Renault (2006)). Then, any other single heteroscedastic factor decomposition of Y_{t+1} such as the one given by (9)-(10) has the same ratio λ_2/λ_1.

It is worth noting that the representation in (9)-(10) considers, without loss of generality, that $E(Y_{t+1}|J_t) = 0$. Moreover, because each component is conditionally heteroskedastic, both λ_1 and λ_2 are non zero.

When the representation in (9)-(10) is true so that $Y_{1,t+1}$ and $Y_{2,t+1}$ have a common heteroskedastic factor, there exists a linear combination of these two components which has a time invariant conditional variance. The second step of the test of common ARCH factor by Engle and Kozicki (1993) translates this time invariance of the conditional variance in terms of moment conditions and applies the Hansen's (1982) test for overidentifying restrictions. Under the null of common conditionally heteroskedastic factor in the processes, the moment conditions are valid and they apply the Hansen's (1982) asymptotic results for the J-test. The moment conditions they derive are

$$E\left\{(z_t - \bar{z})\left(u^2_{\theta,t+1} - \overline{u^2_\theta}\right)\right\} = 0,$$

where z_t is a J_t-measurable H size vector, $u^2_{\theta,t} = (Y_{2,t} - \theta Y_{1,t})^2$ and $\theta \in \mathbb{R}$. The notation $\bar{x}$ stands for the sample mean of the process x_t.

In the GMM estimation procedure for this moment condition model, it is the norm of the sample covariance,

$$\widehat{Cov}(z_t, u^2_{\theta,t+1}) = \frac{1}{T}\sum_{t=1}^{T} z_t u^2_{\theta,t+1} - T\bar{z}\,\overline{u^2_\theta}$$

which is minimized. Therefore, it makes sense for the purposes of the identification studies to focus on the genuine population version of the estimating equations i.e. $Cov(z_t, u^2_{\theta,t+1}) = 0$ or

$$E\left\{(z_t - Ez_t)(u^2_{\theta,t+1} - Eu^2_{\theta,t+1})\right\} = 0. \tag{11}$$

As stated by the next result, the moment conditions model in (11) identifies the true parameter value θ_0 however, the first order condition is not satisfied while the model is identified at the second order.

Theorem 4.1 *Let* $\phi_t(\theta) = (z_t - Ez_t)(u^2_{\theta,t+1} - Eu^2_{\theta,t+1})$. *If* z_t *and* σ^2_t *are stationary and, in addition,* $E\|z_t\| < \infty$ *and* $0 < \|Cov(z_t, \sigma^2_t)\| < \infty$ *then,*

(i) *(Identification) there exists one and only one* $\theta_0 \in \mathbb{R}$ *satisfying the moment conditions in* (11),

(ii) *(First order underidentification)* $E(\partial\phi_t(\theta_0)/\partial\theta) = 0$,

(iii) *(Second order identification)* $E(\partial^2\phi_t(\theta_0)/\partial\theta^2) \neq 0$.

This result by its point (ii) shows that the required rank condition for the application of the Hansen's (1982) asymptotic results is violated in the Engle and Kozicki's (1993) framework. On the other hand, by (i) and (iii), the Engle and Kozicki's (1993) moment conditions are identified and are also identifying at the second order. This fits with our discussion in Section 3 and, instead of being asymptotically distributed as a χ^2_{H-1} as suggested by Engle and Kozicki (1993), their test statistic is asymptotically distributed as a half and half mixture of a χ^2_{H-1} and a χ^2_H. As we already mentioned, because the actual asymptotic distribution has a thicker tail than a χ^2_{H-1} distribution, this asymptotic distribution proposed by Engle and Kozicki (1993) without noticing the first order underidentification leads to an overrejecting test procedure. For a test of level α, the asymptotic relative rate of overrejection is given by

$$100 \times \left(\alpha^{-1}\alpha_0 - 1\right)\%,$$

where $c_{\alpha,H-1}$ is defined by $\text{Prob}\left(\chi^2_{H-1} > c_{\alpha,H-1}\right) = \alpha$ and α_0 is the exact asymptotic level of the Engle and Kozicki's (1993) test associated to the level α given by

$$\alpha_0 = \text{Prob}\left(\frac{1}{2}\chi^2_{H-1} + \frac{1}{2}\chi^2_H > c_{\alpha,H-1}\right).$$

The following tables show the relative overrejection rate of the Engle and Kozicki's (1993) test for various number of included instruments. These tables display the results for the levels 5% and 1%, respectively. We can report similar observation from both tables. The amount of relative overrejection rate is large for any number of included instruments even though it decreases with larger number of instruments.

The minimum number of instruments allowing for overidentification corresponds to the largest amount of overrejection rate. Almost 100% for a 5%-level test and about 130% for a 1%-level test. This amount narrows to 26.2% for a 5%-level test and 34.0% for a 1%-level test for the case where 10 instruments are included. These tables illustrate the discrepancy between the asymptotic approximation by the Engle and Kozicki (1993) test and the exact asymptotic distribution of their test statistic as derived in this paper.

Table 2.1: Overrejection rate of the Engle and Kozicki's (1993) test at the level $\alpha = 0.05$

Number of instruments H	Critical value $c_{\alpha,H-1}$	Exact asymptotic level α_0	Relative overrejection rate $100 \times (\alpha^{-1}\alpha_0 - 1)\%$
2	3.8415	0.0983	96.6%
4	7.8147	0.0743	48.6%
5	9.4877	0.0706	41.2%
6	11.0705	0.0681	36.2%
10	16.9190	0.0631	26.2%

Table 2.2: Overrejection rate of the Engle and Kozicki's (1993) test at the level $\alpha = 0.01$

Number of instruments H	Critical value $c_{\alpha,H-1}$	Exact asymptotic level α_0	Relative overrejection rate $100 \times (\alpha^{-1}\alpha_0 - 1)\%$
2	6.6349	0.0231	131.0%
4	11.3449	0.0165	95.0%
5	13.2767	0.0155	55.0%
6	15.0863	0.0148	48.0%
10	21.6660	0.0134	34.0%

5 Conclusion

This paper explores for the moment condition based models the asymptotic behaviour of the minimum distance estimators and the Hansen (1982) test for overidentifying moment restrictions statistic, J_T under nonstandard conditions. While maintaining a second order identification condition, we derive the rate of consistency of the minimum distance estimators and the asymptotic distribution of J_T when the rank condition is violated, the so-called first order underidentification. We find that the estimators of the set of parameters which are identified at the first order have the usual asymptotic order of magnitude while the other estimators have a larger asymptotic order of magnitude. Our result generalizes the findings by Sargan (1983). We also find that there are some samples in which the non-first-order-identified parameters estimators have the usual rate of convergence while in the other samples, they have a slower rate of convergence. This non standard behaviour affects the asymptotic distribution of J_T. Instead of a chi-squared distribution, it is asymptotically distributed as a half and half mixture of two chi-squared distributions. We apply this result to correct the test for common ARCH factor proposed by Engle and Kozicki (1993) and we also evaluate the amount of overrejection that it leads to without our correction.

A Appendix: proofs

Proof of Proposition 2.1: We deduce from the expansion (5) that

$$\sqrt{T}(\hat{\theta}_{1T}-\theta_1^0) = \left[\frac{\partial\rho'}{\partial\theta_1}(\theta_0)\Omega\frac{\partial\rho}{\partial\theta_1'}(\theta_0)\right]^{-1}\left\{\sqrt{T}\overline{\phi}_T(\theta^0)+\frac{\partial\overline{\phi}_T}{\partial\theta_2'}(\theta^0)\sqrt{T}(\hat{\theta}_{2T}-\theta_2^0)\right.$$
$$\left.+\frac{1}{2}\left[T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)'\frac{\partial^2\rho_h}{\partial\theta_2\partial\theta_2'}(\theta^0)T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)\right]_{1\leq h\leq H}\right\}+o_P(1).$$

In the above expansion, $\sqrt{T}\overline{\phi}_T(\theta^0)=O_P(1)$ by Assumption 2, $(\partial\overline{\phi}_T(\theta^0)/\partial\theta_2')\sqrt{T}(\hat{\theta}_{2T}-\theta_2^0)=o_P(1)$ since $\sqrt{T}\{\partial\overline{\phi}_T(\theta^0)/\partial\theta_2'\}=O_P(1)$ and $\hat{\theta}_{2T}-\theta_2^0=o_P(1)$. Finally, as $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)$, for all $h=1,2,\ldots,H$, $T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)'\{\partial^2\rho_h(\theta^0)/\partial\theta_2\partial\theta_2'\}T^{1/4}(\hat{\theta}_{2T}-\theta_2^0)=O_P(1)$ □

Proof of Lemma 2.1: $\|M_{Z_1}\Delta(v)\|$ is an homogeneous function of degree 2 with respect to v. Therefore

$$\|M_{Z_1}\Delta(v)\| = \|v\|^2\left\|M_{Z_1}\Delta\left(\frac{v}{\|v\|}\right)\right\|.$$

By considering

$$\gamma=\inf_{\|v\|=1}\|M_{Z_1}\Delta(v)\|,$$

we have just to show that $\gamma>0$. By compactness, $\gamma=\|M_{Z_1}\Delta(v^*)\|$ for some v^* such that $\|v^*\|=1$. Therefore, we have just to check that $(M_{Z_1}\Delta(v)=0)\Rightarrow(v=0)$. That is

$$\left(\Delta(v)-Z_1(Z_1'Z_1)^{-1}Z_1'\Delta(v)=0\right)\Rightarrow(v=0).$$

This is a direct consequence of Assumption 6 which can be rewritten (after left multiplication by $\Omega^{1/2}$)

$$\left(Z_1u+\frac{1}{2}\Delta(v)=0\right)\Rightarrow(u=0\quad\text{and}\quad v=0)$$

□

Proof of Proposition 3.1: Let us consider the following two Taylor expansions

$$\bar{\phi}_T(\hat{\theta}_T)=\bar{\phi}_T(\theta_1^0,\hat{\theta}_{2T})+\frac{\partial\bar{\phi}_T}{\partial\theta_1'}(\tilde{\theta}_{1T},\hat{\theta}_{2T})(\hat{\theta}_{1T}-\theta_1^0)$$

and

$$\bar{\phi}_T(\theta_1^0, \hat{\theta}_{2T}) = \bar{\phi}_T(\theta^0) + \frac{\partial \bar{\phi}_T}{\partial \theta_2'}(\theta_1^0, \theta_2^0)(\hat{\theta}_{2T} - \theta_2^0) + \frac{1}{2}\left[(\hat{\theta}_{2T} - \theta_2^0)' \frac{\partial^2 \bar{\phi}_{hT}}{\partial \theta_2 \partial \theta_2'}(\theta_1^0, \tilde{\theta}_{2T})(\hat{\theta}_{2T} - \theta_2^0)\right]_{1 \le h \le H},$$

where, by a common abuse of notation, $\tilde{\theta}_{1T} \in \left[\theta_1^0, \hat{\theta}_{1T}\right]$ and $\tilde{\theta}_{2T} \in \left[\theta_2^0, \hat{\theta}_{2T}\right]$ may take different values for different components $h = 1, \ldots, H$ of the above H equations.

Let us introduce the following notations

$$\begin{aligned} Z_{1T}(\theta) &= \frac{\partial \bar{\phi}_T}{\partial \theta_1'}(\theta) \\ Z_{2T}(\theta) &= \frac{\partial \bar{\phi}_T}{\partial \theta_2'}(\theta) \\ \Delta_T(v) &= \frac{1}{2}\left[v' \frac{\partial^2 \bar{\phi}_{hT}}{\partial \theta_2 \partial \theta_2'}(\theta_1^0, \tilde{\theta}_{2T}) v\right]_{1 \le h \le H}. \end{aligned}$$

Then, while plugging the above second Taylor expansion in the first one, we get

$$\bar{\phi}_T(\hat{\theta}_T) = \bar{\phi}_T(\theta^0) + Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})(\hat{\theta}_{1T} - \theta_1^0) + Z_{2T}(\theta^0)(\hat{\theta}_{2T} - \theta_2^0) + \Delta_T(\hat{\theta}_{2T} - \theta_2^0).$$

Therefore, the first order conditions for θ_1 corresponding to Equation (7) can be written

$$Z_{1T}'(\hat{\theta}_T)\bar{\phi}_T(\theta^0) + Z_{1T}'(\hat{\theta}_T) Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})(\hat{\theta}_{1T} - \theta_1^0) + Z_{1T}'(\hat{\theta}_T) Z_{2T}(\theta^0)(\hat{\theta}_{2T} - \theta_2^0) + Z_{1T}'(\hat{\theta}_T)\Delta_T(\hat{\theta}_{2T} - \theta_2^0) = 0.$$

Note that, by the uniform law of large numbers (Assumption 3) and the consistency of $\hat{\theta}_T$, the random matrix $Z_{1T}'(\hat{\theta}_T) Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})$ converges towards the non-singular matrix

$$\left[E \frac{\partial \rho'}{\partial \theta_1}(\theta^0)\right]\left[E \frac{\partial \rho}{\partial \theta_1'}(\theta^0)\right].$$

Therefore, asymptotic behaviour in probability can be studied through the following rewriting of the first order conditions for θ_1

$$\begin{aligned} \sqrt{T}(\hat{\theta}_{1T} - \theta_1^0) &= -\left[Z_{1T}'(\hat{\theta}_T) Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})\right]^{-1} Z_{1T}'(\hat{\theta}_T) \\ &\quad \times \left\{\sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T} Z_{2T}(\theta^0)(\hat{\theta}_{2T} - \theta_2^0) + \sqrt{T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0)\right\}. \end{aligned}$$

However, since by Assumption 5 $\sqrt{T} Z_{2T}(\theta^0) = O_P(1)$ and $\hat{\theta}_{2T}$ is consistent, we can simplify this expansion as

$$\sqrt{T}(\hat{\theta}_{1T} - \theta_1^0) = -\left[Z_{1T}'(\hat{\theta}_T) Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})\right]^{-1} Z_{1T}'(\hat{\theta}_T)\left\{\sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0)\right\} + o_P(1).$$

For the same reason, the above expansion of $\bar{\phi}_T(\hat{\theta}_T)$ becomes

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) + Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})\sqrt{T}(\hat{\theta}_{1T} - \theta_1^0) + \sqrt{T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0) + o_P(1).$$

Thus, by plugging the above expansion of $\sqrt{T}(\hat{\theta}_{1T} - \theta_1^0)$ in the one of $\sqrt{T}\bar{\phi}_T(\hat{\theta}_T)$, we get

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = M_{Z_{1T},T}\left[\sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0)\right] + o_P(1),$$

where

$$M_{Z_{1T},T} = Id_H - Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})\left[Z_{1T}'(\hat{\theta}_T)Z_{1T}(\tilde{\theta}_{1T}, \hat{\theta}_{2T})\right]^{-1} Z_{1T}(\hat{\theta}_T).$$

Note that since $\sqrt{T}\bar{\phi}_T(\theta^0) = O_P(1)$, we have

$$M_{Z_{1T},T}\sqrt{T}\bar{\phi}_T(\theta^0) = M_{Z_1}\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1),$$

where

$$Z_1 = \frac{\partial\rho}{\partial\theta_1'}(\theta^0) \quad \text{and} \quad M_{Z_1} = Id_H - Z_1(Z_1'Z_1)^{-1}Z_1'.$$

Thus

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = M_{Z_1}\sqrt{T}\bar{\phi}_T(\theta^0) + M_{Z_{1T},T}\sqrt{T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0) + o_P(1)$$

and

$$\begin{aligned}
J_T &= T\bar{\phi}_T'(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) \\
&= T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) + T\bar{\phi}_T'(\theta^0)M_{Z_1}M_{Z_{1T},T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0) + T\Delta_T'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_{1T},T}'M_{Z_1}\bar{\phi}_T(\theta^0) \\
&\quad + T\Delta_T'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_{1T},T}'M_{Z_{1T},T}\Delta_T(\hat{\theta}_{2T} - \theta_2^0) + o_P(1)
\end{aligned}$$

Note that

$$\begin{aligned}
0 &\leq T\bar{\phi}_T'(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) = J_T \\
&\leq \min_{\theta_1} T\bar{\phi}_T'(\theta_1, \theta_2^0)\bar{\phi}_T(\theta_1, \theta_2^0) = J_T^0.
\end{aligned}$$

By the standard GMM theory, J_T^0 converges in distribution towards a χ^2_{H-r}. Therefore, J_T is $O_P(1)$ and thus

$$\begin{aligned}
Y_T &= T\bar{\phi}_T'(\theta^0)M_{Z_1}M_{Z_{1T},T}\Delta(\hat{\theta}_{2T} - \theta_2^0) + T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_{1T},T}'M_{Z_1}\bar{\phi}_T(\theta^0) \\
&\quad + T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_{1T},T}'M_{Z_{1T},T}\Delta(\hat{\theta}_{2T} - \theta_2^0)
\end{aligned} \tag{12}$$

is $O_P(1)$ since

$$Y_T = J_T - T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) + o_P(1).$$

We deduce

$$T\|M_{Z_{1T},T}\Delta_T(\hat{\theta}_{2T}-\theta_2^0)\|^2 \leq |Y_T| + 2T\|M_{Z_1}\bar{\phi}_T(\theta^0)\|\|M_{Z_{1T},T}\Delta_T(\hat{\theta}_{2T}-\theta_2^0)\|.$$

Since $M_{Z_{1T},T}\Delta_T$ converges towards $M_{Z_1}\Delta$ and $(\hat{\theta}_{2T}-\theta_2^0)$ lies in a compact set, convergence in probability can be studied by considering

$$\frac{1}{2}\left\|M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0)\right\| \leq \left\|M_{Z_{1T},T}\Delta_T(\hat{\theta}_{2T}-\theta_2^0)\right\| \leq \frac{3}{2}\left\|M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0)\right\|.$$

Thus, since M_{Z_1} is a projection matrix, we deduce

$$\frac{T}{4}\left\|M_{Z_1}\Delta(\hat{\theta}_{2T}-\theta_2^0)\right\|^2 \leq |Y_T| + 3T\left\|\bar{\phi}_T(\theta^0)\right\|\left\|\Delta(\hat{\theta}_{2T}-\theta_2^0)\right\|.$$

Thus, by Lemma 2.1

$$\frac{\gamma^2 T}{4}\left\|\hat{\theta}_{2T}-\theta_2^0\right\|^4 \leq |Y_T| + 3T\left\|\bar{\phi}_T(\theta^0)\right\|\|\Delta\|\left\|\hat{\theta}_{2T}-\theta_2^0\right\|^2,$$

where $\|\Delta\|$ denotes the norm of the operator $\Delta(v)$ seen as a linear function $vec(vv')$. Thus

$$\sqrt{T}\left\|\hat{\theta}_{2T}-\theta_2^0\right\|^2 \leq \frac{4|Y_T|}{\gamma^2\sqrt{T}\left\|\hat{\theta}_{2T}-\theta_2^0\right\|^2} + \frac{12\|\Delta\|}{\gamma^2}\left\|\sqrt{T}\bar{\phi}_T(\theta^0)\right\|.$$

Therefore, for any positive M,

$$\left(\sqrt{T}\left\|\hat{\theta}_{2T}-\theta_2^0\right\|^2 > M\right) \Rightarrow \left(|Y_T| > \frac{\gamma^2 M^2}{8} \; or \; \left\|\sqrt{T}\bar{\phi}_T(\theta^0)\right\| > \frac{\gamma^2 M}{24\|\Delta\|}\right).$$

Since both Y_T and $\sqrt{T}\bar{\phi}_T(\theta^0)$ are bounded in probability, as $M \to \infty$, Prob $\left[|Y_T| > \frac{\gamma^2 M^2}{8}\right] \to 0$ and Prob $\left[\left\|\sqrt{T}\bar{\phi}_T(\theta^0)\right\| > \frac{\gamma^2 M}{24\|\Delta\|}\right] \to 0$ and thus

$$\text{Prob}\left[\sqrt{T}\left\|\hat{\theta}_{2T}-\theta_2^0\right\|^2 > M\right] \to 0$$

as $M \to \infty$.

In other words

$$T^{1/4}(\hat{\theta}_{2T}-\theta_2^0) = O_P(1).$$

We can deduce from the above expansion of $\sqrt{T}(\hat{\theta}_{1T}-\theta_1^0)$ that $\sqrt{T}(\hat{\theta}_{1T}-\theta_1^0)$ is $O_P(1)$ and

$$\sqrt{T}(\hat{\theta}_{1T}-\theta_1^0) = -[Z_1'Z_1]^{-1}Z_1\left\{\sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T}\Delta(\hat{\theta}_{2T}-\theta_2^0)\right\} + o_P(1)$$

□

Proof of Theorem 3.1: Taking into account the discussion in the main text, we just have to show that when $\dim\theta_2 = 1$ and $T^{1/4}(\hat{\theta}_{2T} - \theta_2^0)$ is asymptotically non degenerate, J_T is asymptotically χ^2_{H-p}. To see that, note that from the expansion of the moment conditions

$$\begin{aligned} J_T &= T\bar{\phi}_T'(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) \\ &= T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) + 2T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_1}\bar{\phi}_T(\theta^0) + T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_1}\Delta(\hat{\theta}_{2T} - \theta_2^0) + o_P(1). \end{aligned}$$

But, by (??),

$$T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_1}\Delta(\hat{\theta}_{2T} - \theta_2^0) = -T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_1}\bar{\phi}_T(\theta^0) + o_P(1).$$

Thus,

$$J_T = T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) + T\Delta'(\hat{\theta}_{2T} - \theta_2^0)M_{Z_1}\bar{\phi}_T(\theta^0) + o_P(1).$$

Moreover, (??) also gives

$$\sqrt{T}(\hat{\theta}_{2T} - \theta_2^0)^2 = -\frac{G'M_{Z_1}}{G'M_{Z_1}G}\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1).$$

Thus

$$\begin{aligned} J_T &= T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) + T(\hat{\theta}_{2T} - \theta_2^0)^2 G'M_{Z_1}\bar{\phi}_T(\theta^0) + o_P(1) \\ &= T\bar{\phi}_T'(\theta^0)M_{Z_1}\bar{\phi}_T(\theta^0) - \frac{T\bar{\phi}_T'(\theta^0)M_{Z_1}GG'M_{Z_1}\bar{\phi}_T(\theta^0)}{G'M_{Z_1}G} + o_P(1). \end{aligned}$$

Therefore

$$J_T = T\bar{\phi}_T'(\theta^0)\left[Id_H - P\right]\bar{\phi}_T(\theta^0) + o_P(1),$$

where

$$P = Z_1(Z_1'Z_1)^{-1}Z_1' + \frac{(M_{Z_1}G)(M_{Z_1}G)'}{G'M_{Z_1}G}$$

is an orthogonal projection matrix on a space of dimension $p = r + 1$ (note that $M_{Z_1}G \neq 0$ by Assumption 6). This proves that J_T is asymptotically χ^2_{H-p} in this case □

Proof of Theorem 4.1: Let $\theta_0 = \lambda_2/\lambda_1$. Since $(-\theta_0\ 1)\lambda = 0$, from (9), we have $Y_{2,t+1} - \theta_0 Y_{1,t+1} = U_{2,t+1} - \theta_0 U_{1,t+1}$. Then, $u^2_{\theta_0,t+1} \equiv (Y_{2,t+1} - \theta_0 Y_{1,t+1})^2 = (U_{2,t+1} - \theta_0 U_{1,t+1})^2$. The conditional expectation of $u^2_{\theta_0,t+1}$ is $E(u^2_{\theta_0,t+1}|J_t) = \Omega_{22} + \theta_0^2\Omega_{11} - 2\theta_0\Omega_{12}$. Because this conditional expectation is time invariant, $E(u^2_{\theta_0,t+1}|J_t) = Eu^2_{\theta_0,t+1}$ so that $E\{u^2_{\theta_0,t+1} - Eu^2_{\theta_0,t+1}|J_t\} = 0$. As z_t is J_t-measurable,

$E\{z_t[u^2_{\theta_0,t+1} - Eu^2_{\theta_0,t+1}]\} = 0$ or equivalently $E\{[z_t - E(z_t)][u^2_{\theta_0,t+1} - Eu^2_{\theta_0,t+1}]\} = 0$ in other words, θ_0 satisfies (11).

Let $\theta \in \mathbb{R}$ such that $E\{[z_t - E(z_t)][u^2_{\theta,t+1} - Eu^2_{\theta,t+1}]\} = 0$, or equivalently $E\{[z_t - E(z_t)][u^2_{\theta,t+1}]\} = 0$. Since $E(Y^2_{1,t+1}|J_t) = \lambda_1^2\sigma_t^2 + \Omega_{11}$, $E(Y^2_{2,t+1}|J_t) = \lambda_2^2\sigma_t^2 + \Omega_{22}$ and $E(Y_{1,t+1}Y_{2,t+1}|J_t) = \lambda_1\lambda_2\sigma_t^2 + \Omega_{12}$, $E\{[z_t - E(z_t)][u^2_{\theta,t+1}]\} = 0$ can be written $E\{(z_t - E(z_t))\sigma_t^2(\lambda_2 - \lambda_1\theta)^2\} = 0$ so that $(\lambda_2 - \lambda_1\theta)^2 Cov(z_t, \sigma_t^2) = 0$. Then $\theta = \lambda_2/\lambda_1 = \theta_0$. This establishes the existence and the uniqueness of θ_0 as stated by (i).

Next, we show (ii),

$$\begin{aligned}
E\{\partial\phi_t(\theta_0)/\partial\theta\} &= E\{(z_t - Ez_t)[-2Y_{1,t+1}(Y_{2,t+1} - \theta_0 Y_{1,t+1})]\} \\
&= -2E\{(z_t - Ez_t)(\lambda_1 f_{t+1} + U_{1,t+1})(U_{2,t+1} - \theta_0 U_{1,t+1})\} \\
&= -2E(z_t - Ez_t)E\{(\lambda_1 f_{t+1} + U_{1,t+1})(U_{2,t+1} - \theta_0 U_{1,t+1})|J_t\} \\
&= -2E(z_t - Ez_t)E\{U_{1,t+1}(U_{2,t+1} - \theta_0 U_{1,t+1})|J_t\} \text{ as } E(f_{t+1}U_{t+1}|J_t) = 0 \\
&= -2E\{(z_t - Ez_t)(\Omega_{12} - \theta_0\Omega_{11})\} = 0.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
E\{\partial^2\phi_t(\theta_0)/\partial\theta^2\} &= E\{(z_t - Ez_t)(2Y^2_{1,t+1})\} = 2E(z_t - Ez_t)E(Y^2_{1,t+1}|J_t) \\
&= 2E\{(z_t - Ez_t)(\lambda_1^2\sigma_t^2 + \Omega_{11})\} \\
&= 2\lambda_1^2 E\{(z_t - Ez_t)\sigma_t^2\} = 2\lambda_1^2 Cov(z_t, \sigma_t^2) \neq 0.
\end{aligned}$$

This establishes (iii)□

References

[1] Andrews, D. W. K., 1994. "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," *Econometrica*, 62, 43-72.

[2] Antoine, B. and E. Renault, 2007. "Efficient GMM with nearly-weak identification," *working paper*, University of North Carolina.

[3] Arellano, M., L. P. Hansen and E. Sentana, 1999. "Underidentification?," *working paper*, CEMFI.

[4] Choi, I. and P. C. B. Phillips, 1992. "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113-150.

[5] Diebold, F. and M. Nerlove, 1989. "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model," *Journal of Applied Econometrics*, 4, 1-21.

[6] Doz, C. and E. Renault, 2006. "Factor Volatility in Mean Models: a GMM approach," *Econometric Review*, 25, 275-309.

[7] Engle, R. F. and S. Kozicki, 1993. "Testing For Common Features," *Journal of Business and Economic Statistics*, 11(4), 369-395.

[8] Engle, R. F. and R. Susmel, 1993. "Common Volatility in International Equity Markets," *Journal of Business and Economic Statistics*, 11, 167-176.

[9] Fiorentini, G., E. Sentana and N, Shephard, 2004. "Likelihood-based Estimation of Generalised ARCH Structures," *Econometrica*, 72, 1481-1517.

[10] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[11] Hansen, L. P., J. Heaton and A. Yaron, 1996. "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.

[12] Hayashi, F., 2000. "Econometrics," *Princeton University Press.*

[13] King, M. A., E. Sentana and S. B. Wadhwani, 1994. "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901-933.

[14] Kleibergen, F., 2005. "Testing Parameters in GMM Without Assuming that they are Identified," *Econometrica*, 73, 1103-1123.

[15] Newey, K. W. and D. McFadden, 1994. "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, IV, Edited by R.F. Engle and D. L. McFadden, 2112-2245.

[16] Newey, K. W. and R. J. Smith, 2004. "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

[17] Qin, J. and J. Lawless, 1994. "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300-325.

[18] Sargan, J. D., 1983. "Identification and lack of Identification," *Econometrica*, 51, 1605-1633.

[19] Staiger, D. and J. H. Stock, 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

[20] Stock, J. H. and J. H. Wright, 2000. "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

[21] Van der Vaart, A. W., 1998. "Asymptotic Statistics," *Cambridge University Press.*

Chapter 3

Bootstrapping Realized Multivariate Volatility Measures

1 Introduction

Realized statistics based on high frequency returns have become very popular in financial economics. Realized volatility is perhaps the most well known example, providing a consistent estimator of the integrated volatility under certain conditions (including the absence of microstructure noise). Its multivariate analogue is the realized covariance matrix, defined as the sum of the outer product of the vector of high frequency returns. Two economically interesting functions of the realized covariance matrix are the realized correlation and the realized regression coefficients. In particular, realized regression coefficients are obtained by regressing high frequency returns for one asset on high frequency returns for another asset. When one of the assets is the market portfolio, the result is a realized beta coefficient. A beta coefficient measures the asset's systematic risk as assessed by its correlation with the market portfolio. Recent examples of papers that have obtained empirical estimates of realized betas include Andersen, Bollerslev, Diebold and Wu (2005a, 2005b) and Viceira (2007).

Recently, Barndorff-Nielsen and Shephard (2004) (henceforth BN-S(2004)) (see also Jacod (1994) and Jacod and Protter (1998)) have proposed an asymptotic distribution theory for realized covariation measures based on multivariate high frequency returns. Their simulation results show that asymptotic theory-based confidence intervals for regression and correlation coefficients between two assets returns can be severely distorted if the sampling horizon is not small enough. To improve the finite sample performance of their feasible asymptotic theory approach, BN-S (2004) propose the Fisher-z transformation for realized correlation. This analytical transformation does not apply to realized regression coefficients, which in particular can be negative and larger than one in absolute value.

In this paper we propose bootstrap methods for statistics based on multivariate high frequency returns, including the realized covariance, the realized regression and the realized correlation coefficients. Our aim is to improve upon the first order asymptotic theory of BN-S (2004). The bootstrap method we consider is an i.i.d. bootstrap applied to the vector of realized returns. Gonçalves and Meddahi (2006a) have recently applied this method to realized volatility in the univariate context. They also proposed a wild bootstrap for realized volatility with the motivation that intraday returns are (conditionally on the volatility path) independent but heteroskedastic when log prices are driven by a stochastic volatility model. In this paper we focus only on the i.i.d. bootstrap for three reasons.

First, the results in Gonçalves and Meddahi (2006a) show that the i.i.d. bootstrap dominates the wild bootstrap in Monte Carlo simulations even when volatility is time varying. Second, the i.i.d. bootstrap is easier to apply than the wild bootstrap: the wild bootstrap requires choosing an external random variable used to construct the bootstrap data whereas the i.i.d. bootstrap does not involve the choice of any tuning parameter. Third, the i.i.d. bootstrap is a natural candidate in the context of realized regressions driven by heteroskedastic errors. Indeed, the i.i.d. bootstrap applied to the vector of returns corresponds to a pairwise bootstrap, as proposed by Freedman (1981). His results show that the pairwise bootstrap is robust to heteroskedasticity in the error term of cross section regression models. Mammen (1993) shows that the pairwise bootstrap is not only first order asymptotically valid under heteroskedasticity in the error term, but it is also second-order correct (i.e. the error incurred by the bootstrap approximation converges more rapidly to zero than the error incurred by the standard normal approximation).

We can summarize our main contributions as follows. We show the first order asymptotic validity of the i.i.d. bootstrap for estimating the distribution function of the realized covariance matrix and smooth functions of it such as the realized covariance, the realized regression and the realized correlation coefficients. We assess the finite sample performance of bootstrap confidence intervals for these three covariation measures by simulation. Our simulation results show that the bootstrap outperforms the feasible first order asymptotic theory of BN-S (2004).

The ability of the bootstrap to provide higher order asymptotic refinements over the standard normal approximation is usually established via Edgeworth expansions. In a related paper (Dovonon, Gonçalves and Meddahi (2007)), we develop the Edgeworth expansions of the distribution of the t-statistics associated with the three covariation measures studied here. These expansions are then used to construct analytical transformations of the raw statistics with improved finite sample properties (in particular, we propose transformations aimed at eliminating the bias or the skewness of the transformed statistics). By developing similar expansions for the bootstrap statistics, we could compare the accuracy of the bootstrap approximation with that of the normal approximation.

In this paper, we develop the Edgeworth expansion for the i.i.d. (or pairwise) bootstrap distribution of the realized regression estimator. Mammen (1993) shows that the pairwise bootstrap is robust to heteroskedasticity in the regression error and provides asymptotic refinements over the usual first order asymptotic theory in the context of standard cross section regression models subject to heteroskedasticity of unknown form. Thus, these results suggest that the i.i.d. bootstrap can be second order correct in the realized regression context analyzed here even under stochastic volatility. This is not the case for the two other statistics (covariance and correlation coefficients), where the i.i.d. bootstrap cannot be expected to provide second order refinements due to the fact that it does not replicate the conditional heteroskedasticity in the data. For this reason, we do not analyze the higher order properties of the i.i.d. bootstrap for the covariance and the correlation coefficients and focus only on the regression estimator.

Contrary to our expectations based on the existing theory for the pairwise bootstrap in the statistics literature, we show that the pairwise bootstrap does not provide an asymptotic refinement over the standard first order asymptotic theory in the context of realized regressions. We contrast our application of the pairwise bootstrap to realized regressions with the application of the pairwise bootstrap in standard cross section regressions. We show that there is a main difference between these two applications, namely the fact that the score of the underlying realized regression model is heterogeneous and does not have mean zero (although the mean of the sum of the scores is zero). This heterogeneity implies that the standard Eicker-White heteroskedasticity robust variance estimator is not consistent in the realized regression context, which justifies the need for the more involved variance estimator proposed by BN-S (2004). The pairwise bootstrap variance coincides with the Eicker-White robust variance estimator and therefore it does not provide a consistent estimator of the variance of the scaled average of the scores. This is in contrast with the results of Freedman (1981) and Mammen (1993), where the score has mean zero by assumption. Nevertheless, the pairwise bootstrap is first order asymptotically valid when applied to a bootstrap t-statistic which is studentized with a variance estimator that is consistent for the population bootstrap variance of the scaled average of the scores. Because the bootstrap scores have mean zero, the Eicker-White robust variance estimator can be used for this effect. This implies that the bootstrap statistic is not of the same form as the statistic based on the original data, which explains why we do not get second order refinements for the pairwise bootstrap in our context.

The remainder of this paper is organized as follows. In Section 2, we introduce the setup, review the existing first order asymptotic theory and state regularity conditions. We also present some Monte

Carlo simulation results that illustrate the finite sample performance of the existing theory. In Section 3, we introduce the bootstrap methods and establish their first-order asymptotic validity for the three statistics of interest in this paper under the regularity conditions stated in Section 2. We also compare the finite sample performance of the bootstrap method with the existing first order asymptotic theory. Section 4 provides a detailed study of the pairwise bootstrap for realized regressions. We first revisit the first order asymptotic theory of the realized regression estimator, comparing the standard Eicker-White robust variance estimator with the more involved estimator of the variance proposed by BN-S (2004). We then contrast the theoretical properties of the pairwise bootstrap, in particular its asymptotic variance, with the properties of the pairwise bootstrap in a standard cross section regression. We also discuss the second order accuracy of this bootstrap method based on the Edgeworth expansion that we develop here. Section 5 contains two empirical applications and Section 6 concludes. Appendix A contains the tables and figures. Appendix B contains the proofs of results appearing in Section 3 whereas the proofs of results in Section 4 are collected in Appendix C.

2 Setup and first-order asymptotic theory

2.1 Setup

Let $p(t)$, for $t \geq 0$, denote the log-price of a bivariate vector of assets[1]. We assume $p(t)$ follows the continuous stochastic volatility model given by

$$dp(t) = \Theta(t)dW(t), \tag{1}$$

where $p(0) = 0$ and where $\Sigma(t) = \Theta(t)\Theta(t)'$ denotes the spot covariance matrix. Here, W denotes a bivariate vector standard Brownian motion and Θ is the instantaneous or spot covolatility process. As in Gonçalves and Meddahi (2006a), we suppose the absence of drift.

Following BN-S (2004), we make the following additional assumptions.

Assumption 1 Θ has elements that are all pathwise càdlàg, the instantaneous covariance Σ is independent of W and, for all $t < \infty$,

$$\int_0^t \Sigma_{kk}(u)du < \infty, \quad k = 1, 2,$$

[1] For notational simplicity, we focus on the bivariate case, but the results could be extended to the general case in a straightforward manner.

where $\Sigma_{kl}(t)$ denotes the (k,l)th element of the $\Sigma(t)$ process.

Assumption 2 For $k = 1, 2$, and $i = 1, \ldots, 1/h$, the quantities

$$h^{-1} \int_{(i-1)h}^{ih} \Sigma_{kk}(u) du$$

are bounded away from 0 and infinity, uniformly in i and h.

The results in this paper are derived regarding the paths of Σ as fixed. Assumption 1 rules out the presence of leverage effects. Under these assumptions, for a given day t, where we take $t = 1$ without loss of generality, we can define the vector of daily returns as $y = \int_0^1 \Theta(u) dW(u)$. Let $y_i' = \begin{pmatrix} y_{1i} & y_{2i} \end{pmatrix}$, $i = 1, \ldots, 1/h$, (where $1/h$ is assumed to be an integer) be the h-horizon intraday returns on a given day on the two assets. We can write $y_i = \int_{(i-1)h}^{ih} \Theta(u) dW(u)$. The integrated covariance matrix of the daily return y is given by

$$\Gamma \equiv \int_0^1 \Sigma(u) du = \int_0^1 \Theta(u) \Theta'(u) du,$$

with typical element (k,l) given by $\Gamma_{kl} \equiv \int_0^1 \Sigma_{kl}(u) du$. For $i = 1, \ldots, 1/h$, let $\Gamma_i \equiv \int_{(i-1)h}^{ih} \Sigma(u) du$ and note that $\Gamma = \sum_{i=1}^{1/h} \Gamma_i$. Note that conditionally on the volatility path, $y_i \sim N(0, \Gamma_i)$ independently across i. Thus the data are (conditionally on Σ) heterogeneous, but independent.

The parameters of interest in this paper are elements of Γ and smooth functions of these.

2.2 The realized covariance matrix

The realized covariance matrix is defined as the sum of the outer products of intraday returns:

$$\hat{\Gamma} = \sum_{i=1}^{1/h} y_i y_i'.$$

Conditionally on the volatility path, the theory of quadratic variation implies that

$$\hat{\Gamma} = \sum_{i=1}^{1/h} y_i y_i' \xrightarrow{P} \sum_{i=1}^{1/h} E(y_i y_i') \equiv \sum_{i=1}^{1/h} \Gamma_i = \Gamma.$$

$\hat{\Gamma}$ contains realized volatilities for each asset on its main diagonal and realized covolatilities between the two assets outside the main diagonal.

Let $vech\left(\hat{\Gamma}\right)$ denote the vector that stacks the lower triangular elements of the columns of the matrix $\hat{\Gamma}$ into a vector. BN-S (2004) (see also Jacod (1994) and Jacod and Protter (1998)) show that under Assumptions 1 and 2, conditionally on the volatility path,

$$\sqrt{h^{-1}}\left(vech\left(\hat{\Gamma}\right)-vech\left(\Gamma\right)\right)\equiv\sqrt{h^{-1}}\begin{pmatrix}\sum_{i=1}^{1/h}y_{1i}^2-\int_0^1\Sigma_{11}(u)\,du\\ \sum_{i=1}^{1/h}y_{1i}y_{2i}-\int_0^1\Sigma_{12}(u)\,du\\ \sum_{i=1}^{1/h}y_{2i}^2-\int_0^1\Sigma_{22}(u)\,du\end{pmatrix}\to^d N(0,V), \tag{2}$$

where

$$V=\int_0^1\left\{\begin{array}{ccc}2\Sigma_{11}^2(u) & 2\Sigma_{11}(u)\Sigma_{12}(u) & 2\Sigma_{12}^2(u)\\ 2\Sigma_{11}(u)\Sigma_{12}(u) & \Sigma_{11}(u)\Sigma_{22}(u)+\Sigma_{12}^2(u) & 2\Sigma_{22}(u)\Sigma_{12}(u)\\ 2\Sigma_{12}^2(u) & 2\Sigma_{22}(u)\Sigma_{12}(u) & 2\Sigma_{22}^2(u)\end{array}\right\}du.$$

BN-S (2004) provide the following estimator of V. Let $x_i=vech\left(y_iy_i'\right)$. Then, Corollary 2 of BN-S (2004) shows that

$$\hat{V}=h^{-1}\sum_{i=1}^{1/h}x_ix_i'-\frac{1}{2}h^{-1}\sum_{i=1}^{1/h-1}\left(x_ix_{i+1}'+x_{i+1}x_i'\right)\xrightarrow{P}V.$$

As BN-S (2004) remark, $\hat{V}$ is a substantially different estimator than that used by Barndorff-Nielsen and Shephard (2002) in the univariate context, in which case letting $x_i=y_{1i}^2$, it corresponds to

$$\hat{V}=h^{-1}\sum_{i=1}^{1/h}y_{1i}^4-h^{-1}\sum_{i=1}^{1/h-1}y_{1i}^2y_{1,i+1}^2,$$

as opposed to $\frac{2}{3}\sum_{i=1}^{1/h}y_{1i}^4$, the estimator proposed by BN-S (2002). The main feature of notice is the presence of lags of returns in the second piece. One of our contributions is to provide a new interpretation for this estimator in the context of the realized regression estimator (see Section 4.1).

2.3 The realized covariance

Let $\hat{\Gamma}_{12}=\sum_{i=1}^{1/h}y_{1i}y_{2i}$ be the realized covariance between assets 1 and 2, and let $\Gamma_{12}=\int_0^1\Sigma_{12}(u)\,du$ be the corresponding integrated covariance.

From (2), it follows that as $h\to 0$,

$$S_{\Gamma,h}\equiv\frac{\sqrt{h^{-1}}\left(\hat{\Gamma}_{12}-\Gamma_{12}\right)}{\sqrt{V_\Gamma}}\to^d N(0,1),$$

where

$$V_\Gamma=\int_0^1\left\{\Sigma_{11}(u)\Sigma_{22}(u)+\Sigma_{12}^2(u)\right\}du \tag{3}$$

is the asymptotic variance of $\hat{\Gamma}_{12}$.

The corresponding feasible limit theory is

$$T_{\Gamma,h}\equiv\frac{\sqrt{h^{-1}}\left(\hat{\Gamma}_{12}-\Gamma_{12}\right)}{\sqrt{\hat{V}_\Gamma}}\to^d N(0,1),$$

where

$$\hat{V}_\Gamma = h^{-1}\sum_{i=1}^{1/h} y_{1i}^2 y_{2i}^2 - h^{-1}\sum_{i=1}^{1/h-1} y_{1i}y_{2i}y_{1,i+1}y_{2,i+1}, \tag{4}$$

is a consistent estimator of V_Γ.

2.4 The realized correlation

The realized correlation between assets 1 and 2 is given by

$$\hat{\rho} = \frac{\sum_{i=1}^{1/h} y_{1i}y_{2i}}{\sqrt{\sum_{i=1}^{1/h} y_{1i}^2}\sqrt{\sum_{i=1}^{1/h} y_{2i}^2}}.$$

Its probability limit follows directly from the theory of quadratic variation. In particular,

$$\hat{\rho} \xrightarrow{P} \rho \equiv \frac{\int_0^1 \Sigma_{12}(u)du}{\sqrt{\int_0^1 \Sigma_{11}(u)du}\sqrt{\int_0^1 \Sigma_{22}(u)du}}.$$

The asymptotic distribution can be derived by the delta method. Specifically, BN-S (2004) give the following results. The infeasible limit theory is

$$S_{\rho,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\rho}-\rho)}{\sqrt{V_\rho}},$$

where

$$V_\rho = \left(\int_0^1 \Sigma_{11}(u)du \int_0^1 \Sigma_{22}(u)du\right)^{-1} g_\rho, \tag{5}$$

with

$$\begin{aligned} g_\rho &= d_{12}' V d_{12}, \\ d_{12} &= \left(-\frac{\beta_{12}}{2}, 1, -\frac{\beta_{21}}{2}\right)', \end{aligned}$$

with V defined as above and where β_{kl} denotes the population regression coefficient of regressing asset k on asset l. The corresponding feasible theory is

$$T_{\rho,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\rho}-\rho)}{\sqrt{\hat{V}_\rho}},$$

where $\hat{V}_\rho = \left(\sum_{i=1}^{1/h} y_{1i}^2 \sum_{i=1}^{1/h} y_{2i}^2\right)^{-1} h^{-1}\hat{g}_\rho$, with

$$\begin{aligned} \hat{g}_\rho &= \sum_{i=1}^{1/h} x_{\rho i}^2 - \sum_{i=1}^{1/h-1} x_{\rho i}x_{\rho,i+1} \\ x_{\rho i} &= y_{2i}(y_{1i} - \hat{\beta}_{12}y_{2i})/2 + y_{1i}(y_{2i} - \hat{\beta}_{21}y_{1i})/2, \end{aligned} \tag{6}$$

and $\hat{\beta}_{kl} = \sum_{i=1}^{1/h} y_{ki}y_{li} / \sum_{i=1}^{1/h} y_{li}^2$, for $k,l = 1,2$.

2.5 The realized regression

Suppose we regress asset 1 on asset 2 to obtain the realized regression estimator

$$\hat{\beta}_{12} = \frac{\sum_{i=1}^{1/h} y_{1i}y_{2i}}{\sum_{i=1}^{1/h} y_{2i}^2}.$$

BN-S (Proposition 1, 2004) show that as $h \to 0$,

$$S_{\beta,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\beta}_{12} - \beta_{12})}{\sqrt{V_\beta}} \to^d N(0,1),$$

where

$$V_\beta = (\Gamma_{22})^{-2} g_{12}, \tag{7}$$

and

$$\begin{aligned} g_{12} &= d_{12}'\Psi_{12}d_{12} \\ d_{12} &= (1, -\beta_{12})' \\ \Psi_{12} &= \int_0^1 \left\{ \begin{array}{cc} \Sigma_{11}(u)\Sigma_{22}(u) + \Sigma_{12}^2(u) & 2\Sigma_{22}(u)\Sigma_{12}(u) \\ 2\Sigma_{22}(u)\Sigma_{12}(u) & 2\Sigma_{22}^2(u) \end{array} \right\} du. \end{aligned}$$

BN-S (2004) provide the following feasible theory for realized regression, which replaces V_β with a consistent estimator. In particular, they suggest

$$\hat{V}_\beta = \left(\sum_{i=1}^{1/h} y_{2i}^2\right)^{-2} h^{-1}\hat{g}_\beta, \tag{8}$$

where

$$\begin{aligned} \hat{g}_\beta &= \sum_{i=1}^{1/h} x_{\beta i}^2 - \sum_{i=1}^{1/h-1} x_{\beta i}x_{\beta,i+1}, \quad \text{and} \\ x_{\beta i} &= y_{1i}y_{2i} - \hat{\beta}_{12}y_{2i}^2. \end{aligned}$$

BN-S (2004) show that $\hat{V}_\beta \xrightarrow{P} V_\beta$, and therefore it follows that

$$T_{\beta,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\beta}_{12} - \beta_{12})}{\sqrt{\hat{V}_\beta}} \to^d N(0,1).$$

2.6 Monte Carlo results for the first-order asymptotic theory

In this section we assess the finite sample performance of confidence intervals for the three covariation measures (covariance, regression and correlation) based on the existing first order asymptotic theory. We present results for two data generating processes. The first model (henceforth Design 1) is the same as that used by BN-S (2004). In particular, we let

$$dp\,(t) = \Theta(t)dW(t), \qquad \Sigma(t) = \Theta(t)\Theta(t)',$$

where

$$\Sigma(t) = \begin{pmatrix} \Sigma_{11}(t) & \Sigma_{12}(t) \\ \Sigma_{12}(t) & \Sigma_{22}(t) \end{pmatrix} = \begin{pmatrix} \sigma_1^2(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_2^2(t) \end{pmatrix},$$

and $\sigma_{12}(t) = \sigma_1(t)\sigma_2(t)\rho(t)$.

Following BN-S (2004), we let $\sigma_1^2(t)$ be the sum of two uncorrelated CIR processes:

$$\sigma_1^2\,(t) = \sigma_1^{2(1)}(t) + \sigma_1^{2(2)}(t).$$

For $s = 1, 2$,

$$d\sigma_1^{2(s)}(t) = -\lambda_s(\sigma_1^{2(s)}(t) - \xi_s)dt + \omega_s\sigma_1^{(s)}(t)\sqrt{\lambda_s}db_s(t),$$

where b_i is the i-th component of a vector of standard Brownian motions, independent from W. We let $\lambda_1 = 0.0429$, $\xi_1 = 0.110$, $\quad \omega_1 = 1.346$, $\quad \lambda_2 = 3.74$, $\quad \xi_2 = 0.398$, $\quad$ and $\omega_2 = 1.346$.

Similarly to BN-S (2004), our model for $\sigma_2^2(t)$ is the GARCH(1,1) diffusion studied by Andersen and Bollerslev (1998):

$$d\sigma_2^2(t) = -0.035(\sigma_2^2(t) - 0.636)dt + 0.236\sigma_2^2(t)db_3(t).$$

The model we specify for $\rho(t)$ is the same as the one proposed by BN-S(2004):

$$\rho(t) = (e^{2x(t)} - 1)/(e^{2x(t)} + 1),$$

where x follows the GARCH diffusion

$$dx(t) = -0.03(x(t) - 0.64)dt + 0.118x(t)db_4(t).$$

Our second model (Design 2) specifies $\sigma_2^2\,(t)$ and $\rho\,(t)$ exactly as Design 1, with the only difference being in the model used to generate $\sigma_1^2\,(t)$. In particular, for $\sigma_1^2\,(t)$ we consider the two-factor diffusion model studied by Chernov et al. (2003) (see also Huang and Tauchen (2005)):

$$\begin{aligned}
\sigma_1(t) &= \text{s-exp}\left(-1.2 + 0.04 v_1(t) + 1.5 v_2(t)\right) \\
dv_1(t) &= -0.00137 v_1(t)\, dt + db_1(t) \\
dv_2(t) &= -1.386 v_2(t)\, dt + \left(1 + 0.25 v_2(t)\right) db_2(t).
\end{aligned}$$

This diffusion model has continuous sample paths but can imply sample paths for the price process that look like jumps[2]. Although our theory does not allow for a non zero correlation between the price process and the volatility, in our simulations, we allow for these leverage effects. In particular, we let $Corr(dW_1, db_1) = -0.3$ and $Corr(dW_1, db_2) = -0.3$.

We study the finite sample performance of one-sided and two-sided (symmetric) confidence intervals for each of the three measures of covariation: Γ_{12}, the covariance between the returns on asset 1 and on asset 2, β_{12}, the population regression coefficient of the regression of y_1 on y_2, and ρ, the correlation coefficient between the two assets.

Let θ denote any of these three parameters of interest. Similarly, let $\hat{\theta}$ denote the corresponding realized estimator and let $\hat{V}_\theta$ denote a consistent estimator of the variance of $\sqrt{h^{-1}}\hat{\theta}$. In particular, for $\theta = \Gamma_{12}$, $\hat{\theta} = \sum_{i=1}^{1/h} y_{1i} y_{2i}$, and $\hat{V}_\theta = \hat{V}_\Gamma = \sum_{i=1}^{1/h} y_{1i}^2 y_{2i}^2 - \sum_{i=1}^{1/h-1} y_{1i} y_{2i} y_{1,i+1} y_{2,i+1}$, as defined in (4). For $\theta = \beta_{12}$, $\hat{\theta} = \hat{\beta}_{12}$, and $\hat{V}_\theta = \hat{V}_\beta = \left(\sum_{i=1}^{1/h} y_{2i}^2\right)^{-2} h^{-1} \hat{g}_\beta$, with $\hat{g}_\beta$ defined in (14). For $\theta = \rho$, $\hat{\theta} = \hat{\rho}$, and $\hat{V}_\theta = \hat{V}_\rho = \left(\sum_{i=1}^{1/h} y_{1i}^2 \sum_{i=1}^{1/h} y_{2i}^2\right)^{-1} h^{-1} \hat{g}_\rho$, where $\hat{g}_\rho$ is defined in (6).

The lower one-sided $100(1-\alpha)\%$ level confidence interval for θ based on the feasible asymptotic theory of BN-S (2004) is given by

$$IC_{Feas,1-\alpha}^{(1)} = (-\infty, \hat{\theta} - z_\alpha \sqrt{h \hat{V}_\theta}),$$

where z_α is the α-level critical value of the standard normal distribution. The two-sided $100(1-\alpha)\%$ level confidence interval for θ is given by

$$IC_{Feas,1-\alpha}^{(2)} = (\hat{\theta} - z_{1-\alpha/2} \sqrt{h \hat{V}_\theta}, \hat{\theta} + z_{1-\alpha/2} \sqrt{h \hat{V}_\theta}).$$

We present results for three nominal levels: 95% (i.e. $\alpha = 0.05$), 90% ($\alpha = 0.10$) and 99% ($\alpha = 0.01$).

We compute the actual coverage probabilities of these confidence intervals for each of the stochastic volatility models described above. We report results across 10,000 replications for five different sample

[2]The function s-exp is the usual exponential function with a linear growth function splined in at high values of its argument: s-exp$(x) = \exp(x)$ if $x \leq x_0$ and s-exp$(x) = \frac{\exp(x_0)}{\sqrt{x_0}} \sqrt{x_0 - x_0^2 + x^2}$ if $x > x_0$, with $x_0 = \log(1.5)$.

sizes: $1/h = 1152, 288, 48, 24$ and 12, corresponding to "1.25-minute", "5-minute", "15-minute", "half-hour", "1-hour", and "2-hour" returns. Table 3.1 contains results for $\alpha = 0.05$, for each of the two designs, for both one-sided and two-sided symmetric intervals. Table 3.2 contains results for $\alpha = 0.10$ whereas Table 3.3 refers to $\alpha = 0.01$. (These tables also include results for the bootstrap method but those results will be discussed later.)

We start with Table 3.1. For the two DGP's, both one-sided and two-sided intervals tend to undercover. The degree of undercoverage is especially large for larger values of h, when sampling is not too frequent. For the covariance measure and the regression coefficient, one-sided intervals tend to perform worse than two-sided intervals. The opposite is true for the correlation coefficient, which is surprising when analyzed from the viewpoint of the theory of Edgeworth expansions (the analysis based on Edgeworth expansions suggests that the error of one-sided intervals is of the order $O\left(\sqrt{h}\right)$ whereas the error of symmetric two-sided intervals is usually of the order $O\left(h\right)$). For one-sided intervals, the covariance measure is associated with the largest distortions, followed by the regression coefficient, which in turn is worse than the correlation coefficient. For two-sided intervals, this ranking is changed, with the correlation coefficient performing worst, followed by the covariance and by the regression coefficient. The degree of undercoverage can be quite substantial at the smallest sample sizes. For instance, a lower 95% nominal level for the covariance measure between the two assets for Design 1 is equal to 80.74% when we sample every two hours ($h = 1/12$). For the regression coefficient, it is equal to 86.04% and for the correlation coefficient is is equal to 91.43%. The corresponding coverage rates for two-sided intervals based on the BN-S asymptotics are 83.90%, 85.27% and 81.04% for the covariance, the regression and the correlation coefficients, respectively. For this last measure of dependence, we also report the coverage rates of confidence intervals based on the Fisher-z transform, as proposed by BN-S (2004). For one-sided intervals, the 95% interval based on the Fisher transform covers the correlation coefficient 90.28% percent of the time whereas for two-sided intervals, the actual coverage rate is equal to 85.44%. Compared to the intervals based on the raw statistic, the Fisher-z transform outperforms the raw statistic only for the two-sided intervals and not for the one-sided interval. In both cases, however, it is clear that finite sample distortions remain for the Fisher-z transform, thus motivating the use of the bootstrap and/or of alternative analytical corrections.

The results for Design 2 are qualitatively similar to those discussed for Design 1. Quantitatively, the degree of undercoverage is smaller for Design 2, which suggests that contrary to the univariate case (see Gonçalves and Meddahi (2006a)) the asymptotic theory handles well the presence of the two-factor diffusion model of Chernov et al. (2003) in one of the volatility processes. The results for Design 2 also suggest that the theory of BN-S (2004) is robust to the introduction of leverage effects.

Tables 3.2 and 3.3 show that the performance of the asymptotic theory of BN-S (2004) for the 90% and 99% confidence intervals is qualitatively similar to the performance of the 95% level intervals.

3 The bootstrap

In this section we propose bootstrap methods for smooth functions of the realized covariance matrix. The bootstrap method we consider is the i.i.d. bootstrap applied to the vector of returns.

3.1 The bootstrap realized covariance matrix

We first state the first order asymptotic validity of the bootstrap for the realized covariance matrix and smooth functions of its elements. We then specialize our results to the three statistics of interest: realized covariance, realized correlation and realized regression.

Let $x_i = vech\left(y_i y_i'\right) = \begin{pmatrix} y_{1i}^2 & y_{1i}y_{2i} & y_{2i}^2 \end{pmatrix}'$, and recall that

$$T_h \equiv V^{-1/2}\sqrt{h^{-1}}\sum_{i=1}^{1/h}\left(x_i - E\left(x_i\right)\right) \xrightarrow{d} N\left(0, I_3\right),$$

where $\sum_{i=1}^{1/h} x_i = vech\left(\hat{\Gamma}\right)$ denotes the vectorized realized covariance matrix $\hat{\Gamma} = \sum_{i=1}^{1/h} y_i y_i'$, and $V = \lim_{h\to 0} Var\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h} x_i\right)$.

We apply the i.i.d. bootstrap to x_i. In particular, let $x_i^* = x_{I_i} = \begin{pmatrix} y_{1I_i}^2 & y_{1I_i}y_{2I_i} & y_{2I_i}^2 \end{pmatrix}'$, where I_i is i.i.d. on $\{1, \ldots, 1/h\}$. Notice that this is equivalent to bootstrapping the bivariate vector of assets returns $y_i = \left(y_{1i}, y_{2i}\right)'$. Define the (scaled) vectorized bootstrap realized covariance matrix as $\sqrt{h^{-1}}\sum_{i=1}^{1/h} x_i^* = \sqrt{h^{-1}}\sum_{i=1}^{1/h} vech\left(y_i^* y_i^{*\prime}\right) \equiv \sqrt{h^{-1}} vech\left(\hat{\Gamma}^*\right)$. As usual in the bootstrap literature, we let E^* (and Var^*) denote the expectation (and the variance) with respect to bootstrap data, conditional on the original data. It is easy to show that $E^*\left(\sqrt{h^{-1}} vech\left(\hat{\Gamma}^*\right)\right) = \sqrt{h^{-1}} vech\left(\hat{\Gamma}\right)$, and

$$V^* \equiv Var^*\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h} x_i^*\right) = h^{-1}\sum_{i=1}^{1/h} x_i x_i' - \left(\sum_{i=1}^{1/h} x_i\right)\left(\sum_{i=1}^{1/h} x_i\right)'.$$

We can show that

$$V^* \to^P V + \int_0^1 vech\left(\Sigma\left(u\right)\right) vech\left(\Sigma\left(u\right)\right)' du - \left(\int_0^1 vech\left(\Sigma\left(u\right)\right) du\right) \left(\int_0^1 vech\left(\Sigma\left(u\right)\right) du\right)',$$

which is not equal to V (one exception is when $\Sigma(u) = \Sigma$ for all u). Although V^* does not consistently estimate V, the i.i.d. bootstrap is still asymptotically valid when applied to the following studentized statistic

$$T_h^* \equiv \hat{V}^{*-1/2}\sqrt{h^{-1}}\left(vech\left(\hat{\Gamma}^*\right) - vech\left(\hat{\Gamma}\right)\right),$$

where

$$\hat{V}^* = h^{-1}\sum_{i=1}^{1/h} x_i^* x_i^{*\prime} - \left(\sum_{i=1}^{1/h} x_i^*\right)\left(\sum_{i=1}^{1/h} x_i^*\right)'$$

is a consistent estimator of V^*. The following theorem states formally these results.

Theorem 3.1 *Let Assumptions 1 and 2 hold and let $\{y_i^* : i = 1, \ldots, 1/h\}$ denote a set of i.i.d. bootstrap returns. Then, as $h \to 0$,*

a) $\hat{V}^* - V^* \xrightarrow{P^*} 0$, *in probability.*

b) $sup_x \left|P^*\left(T_h^* \le x\right) - P\left(T_h \le x\right)\right| \to 0$ *in probability.*

The proofs of all the results in this section appear in Appendix B.

Several statistics of interest can be written as smooth functions of the realized covariance matrix. Examples include the realized covariance measure between two assets, the realized regression coefficient, and the realized correlation coefficient. The following theorem proves that the i.i.d. bootstrap is first order asymptotically valid when applied to smooth functions of the (appropriately centered and studentized version of) the vectorized realized covariance matrix.

Let $f(\theta) : \mathbb{R}^3 \to \mathbb{R}$ denote a real valued function with continuous derivatives, and let $\nabla f(\theta) = \begin{pmatrix} \partial f/\partial\theta_1 & \partial f/\partial\theta_2 & \partial f/\partial\theta_3 \end{pmatrix}'$ denote its gradient. We suppose that $\nabla f(\theta)$ is nonzero at θ_0, the true value of θ. The statistic of interest is defined as

$$T_{f,h} = \frac{\sqrt{h^{-1}}\left(f\left(vech\left(\hat{\Gamma}\right)\right) - f\left(vech\left(\Gamma\right)\right)\right)}{\sqrt{\hat{V}_f}},$$

where

$$\hat{V}_{f,h} = \left(\nabla' f\left(vech\left(\hat{\Gamma}\right)\right)\hat{V}\ \nabla f\left(vech\left(\hat{\Gamma}\right)\right)\right).$$

The i.i.d. bootstrap version of $T_{f,h}$ is $T^*_{f,h}$, which replaces $\hat{\Gamma}$ with $\hat{\Gamma}^*$, Γ with $\hat{\Gamma}$, and $\hat{V}_f$ with $\hat{V}^*_f = \left(\nabla' f\left(vech\left(\hat{\Gamma}^*\right)\right)\hat{V}^*\nabla f\left(vech\left(\hat{\Gamma}^*\right)\right)\right)$, which is a consistent estimator of the bootstrap asymptotic variance $V^*_f \equiv \left(\nabla' f\left(vech\left(\hat{\Gamma}\right)\right)V^*\nabla f\left(vech\left(\hat{\Gamma}\right)\right)\right)$.

Theorem 3.2 *Under the same conditions of Theorem 3.1, as $h \to 0$,*

$$\sup_x \left|P^*\left(T^*_{f,h} \le x\right) - P\left(T_{f,h} \le x\right)\right| \to 0,$$

in probability.

The next sections give explicitly the bootstrap statistics for the three cases of interest, namely the covariance measure Γ_{12}, the correlation coefficient ρ and the regression coefficient β.

3.2 The bootstrap realized covariance

The bootstrap realized covariance measure is defined as $\hat{\Gamma}^*_{12} = \sum_{i=1}^{1/h} y^*_{1i}y^*_{2i}$, which corresponds to taking $f\left(vech\left(\hat{\Gamma}\right)\right)$ with $f(\theta) = \theta_2$, with $\theta = (\theta_1, \theta_2, \theta_3)$. Thus, the bootstrap statistic is defined as

$$T^*_{\Gamma,h} \equiv \frac{\sqrt{h^{-1}}\left(\hat{\Gamma}^*_{12} - \hat{\Gamma}_{12}\right)}{\sqrt{\hat{V}^*_{\Gamma}}},$$

where

$$\hat{V}^*_{\Gamma} = h^{-1}\sum_{i=1}^{1/h} y^{*2}_{1i}y^{*2}_{2i} - \left(\sum_{i=1}^{1/h} y^*_{1i}y^*_{2i}\right)^2.$$

Theorem 3.2 above proves the first order asymptotic validity of the bootstrap when applied to $T^*_{\Gamma,h}$.

3.3 The bootstrap realized correlation

The bootstrap realized correlation coefficient $\hat{\rho}^*$ is defined in the same fashion as $\hat{\rho}$ but with the bootstrap data replacing the original data, i.e.

$$\hat{\rho}^* = \frac{\sum y^*_{1i}y^*_{2i}}{\sqrt{\sum y^{*2}_{1i}}\sqrt{\sum y^{*2}_{2i}}}.$$

The corresponding t-statistic is given by

$$T^*_{\rho,h} \equiv \frac{\sqrt{h^{-1}}\left(\hat{\rho}^* - \hat{\rho}\right)}{\sqrt{\hat{V}^*_{\rho}}},$$

where

$$\begin{aligned}\hat{V}^*_\rho &= \left(\hat{\Gamma}^*_{11}\hat{\Gamma}^*_{22}\right)^{-1}\hat{B}^*_\rho,\\ \hat{B}^*_\rho &= h^{-1}\sum x^{*2}_{\rho i}, \quad \text{and}\\ x^*_{\rho i} &= y^*_{2i}\left(y^*_{1i}-\hat{\beta}^*_{12}y^*_{2i}\right)/2 + y^*_{1i}\left(y^*_{2i}-\hat{\beta}^*_{21}y^*_{1i}\right)/2.\end{aligned}$$

Here $\hat{\beta}^*_{kl}$ denotes the bootstrap OLS regression estimator of the realized regression of y^*_k on y^*_l, for $k,l = 1,2$. We note that $\rho = f(vech(\Gamma))$, with $f(\theta) = \frac{\theta_2}{\sqrt{\theta_1\theta_3}}$. Thus, the first order asymptotic validity of the i.i.d. bootstrap for the correlation coefficient follows from Theorem 3.2.

3.4 The bootstrap realized regression

Let $\{y^*_i = (y^*_{1i}, y^*_{2i}) : i = 1, \ldots, 1/h\}$ be an i.i.d. bootstrap sample from $\{y_i\}$. The boostrap OLS estimator that we obtain by regressing y^*_{1i} on y^*_{2i} is given by

$$\hat{\beta}^*_{12} = \frac{\sum_{i=1}^{1/h} y^*_{1i}y^*_{2i}}{\sum_{i=1}^{1/h} y^{*2}_{2i}}. \tag{9}$$

The corresponding t-statistic is

$$T^*_{\beta,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\beta}^*_{12} - \hat{\beta}_{12})}{\sqrt{\hat{V}^*_\beta}}, \tag{10}$$

where

$$\hat{V}^*_\beta = \left(\sum_{i=1}^{1/h} y^{*2}_{2i}\right)^{-2} h^{-1}\sum_{i=1}^{1/h} y^{*2}_{2i}\hat{\varepsilon}^{*2}_i \equiv \left(\hat{\Gamma}^*_{22}\right)^{-2}\hat{B}^*_{1h}. \tag{11}$$

Theorem 3.2 covers the case of realized regression when $f(\theta) = \frac{\theta_2}{\theta_3}$, with $\theta = (\theta_1, \theta_2, \theta_3)'$ by noting that $\hat{\beta}_{12} = f\left(vech\left(\hat{\Gamma}\right)\right)$.

3.5 Monte Carlo results for the bootstrap

Our theoretical results suggest the first order asymptotic validity of the i.i.d. bootstrap. Thus, we can build confidence intervals for θ using the α-percentile q^*_α of the bootstrap distribution of $T^*_{\theta,h}$. As previously, we let θ denote any of the three measures of covariation, $\hat{\theta}$ its estimator and $\hat{V}_\theta$ the corresponding variance estimator. The lower one-sided $100(1-\alpha)\%$ level bootstrap confidence interval for θ is given by

$$IC^{*(1)}_{1-\alpha} = (-\infty, \hat{\theta} - q^*_\alpha\sqrt{h\hat{V}_\theta}).$$

The bootstrap allows for two-sided symmetric and equal-tailed confidence intervals. The $100(1-\alpha)\%$ level symmetric bootstrap confidence interval is given by

$$(\hat{\theta} - p^*_{1-\alpha/2}\sqrt{h\hat{V}_\theta}, \hat{\theta} + p^*_{1-\alpha/2}\sqrt{h\hat{V}_\theta}).$$

where p^*_α is the α-percentile of the bootstrap distribution of $|T^*_{\theta,h}|$. The $100(1-\alpha)\%$ level equal-tailed bootstrap confidence interval is of the form

$$(\hat{\theta} - q^*_{1-\alpha/2}\sqrt{h\hat{V}_\theta}, \hat{\theta} - q^*_{\alpha/2}\sqrt{h\hat{V}_\theta}).$$

We concentrate our discussion on Table 3.1, which contains results for 95% level confidence intervals. Tables 3.2 and 3.3 contain the corresponding results for 90% and 99% level intervals, respectively. Since the results are qualitatively similar, we do not discuss these in detail here. Our results suggest that the i.i.d. bootstrap intervals outperform the asymptotic theory based intervals for the two DGP's and for both one-sided and two-sided intervals, for all three measures of dependence. Symmetric intervals are generally better than equal-tailed intervals (this is consistent with the theory based on Edgeworth expansions) and both improve upon the first order asymptotic theory based intervals. The gains associated with the i.i.d. bootstrap can be quite substantial, especially for the smaller sample sizes, when distortions of the BN-S intervals are larger. For instance, for the regression coefficient, the coverage rate for a symmetric bootstrap interval is equal to 93.48% when $1/h = 12$, whereas it is equal to 85.27% for the feasible asymptotic theory of BN-S (2004) (the corresponding equal-tailed interval yields a coverage rate of 90.42%, better than BN-S (2004) but worse than the symmetric bootstrap interval). The gains are especially important for the two-sided intervals for the correlation coefficient, when the asymptotic theory of BN-S (2004) does worst. For $1/h = 12$, the bootstrap symmetric interval has a rate of 93.69% (the equal tailed interval is in this case even better behaved, with a rate equal to 94.60) whereas the BN-S interval based on the raw statistic has a rate of 81.04% and the interval based on the Fisher-z transform has a rate of 85.44%. For the correlation coefficient, the bootstrap essentially removes all finite sample bias associated with the first order asymptotic theory of BN-S (2004).

4 A detailed study of realized regressions

The realized regression estimator is one of the most popular measures of covariation between two assets. In this section we study in more detail the application of the i.i.d. bootstrap to realized regression. We first provide a new interpretation for the feasible approach of BN-S (2004). In particular, we establish a link between the standard Eicker-White heteroskedasticity robust variance estimator and the variance estimator proposed by BN-S (2004). We then exploit the special structure of the regression model to obtain the asymptotic distribution of the bootstrap realized regression estimator. We relate the bootstrap variance with the Eicker-White robust variance estimator. We end this section with a discussion of the second order accuracy of the i.i.d. bootstrap in this context.

4.1 The first order asymptotic theory revisited

Given Assumptions 1 and 2, and conditionally on the volatility path, we can write

$$y_{1i} = \beta_{12,i} y_{2i} + u_i, \tag{12}$$

where independently across $i = 1, \ldots, h^{-1}$,

$$u_i | y_{2i} \sim N\left(0, V_i\right),$$

with $V_i \equiv \Gamma_{11,i} - \frac{\Gamma_{12,i}^2}{\Gamma_{22,i}}$, and $\beta_{12,i} \equiv \frac{\Gamma_{12,i}}{\Gamma_{22,i}}$. Here $\Gamma_{kl,i} = \int_{(i-1)h}^{ih} \Sigma_{kl}\left(u\right) du$. Thus, the regression coefficient in the true DGP describing the relationship between y_{1i} and y_{2i} is heterogeneous (it depends on i) and the true error term in this model is heteroskedastic.

When we regress y_{1i} on y_{2i} to obtain $\hat{\beta}_{12}$, we get that

$$\hat{\beta}_{12} \xrightarrow{P} \frac{\sum_{i=1}^{1/h} E\left(y_{1i} y_{2i}\right)}{\sum_{i=1}^{1/h} E\left(y_{2i}^2\right)} = \frac{\sum_{i=1}^{1/h} \Gamma_{12,i}}{\sum_{i=1}^{1/h} \Gamma_{22,i}} = \frac{\Gamma_{12}}{\Gamma_{22}} \equiv \beta_{12}.$$

Thus, $\hat{\beta}_{12}$ does not estimate $\beta_{12,i}$ but instead β_{12}, which can be thought of as a weighted average of $\beta_{12,i}$. We can write the underlying regression model as follows:

$$y_{1i} = \beta_{12} y_{2i} + \varepsilon_i, \tag{13}$$

where

$$\varepsilon_i = \left(\beta_{12,i} - \beta_{12}\right) y_{2i} + u_i.$$

It follows that $\varepsilon_i|y_{2i} \sim N\left(\left(\beta_{12,i} - \beta_{12}\right) y_{2i}, V_i\right)$, independently across i. Moreover, noting that $E\left(y_{2i}\right) = 0$,

$$Cov\left(y_{2i}, \varepsilon_i\right) = E\left(y_{2i}\varepsilon_i\right) = \left(\beta_{12,i} - \beta_{12}\right)\Gamma_{22,i} = \Gamma_{12,i} - \beta_{12}\Gamma_{22,i},$$

which in general is not equal to zero (unless volatility is constant). However, $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right) = 0$, and therefore $\hat{\beta}_{12}$ converges in probability to β_{12}. The fact that $E\left(y_{2i}\varepsilon_i\right) \neq 0$ is crucial to understand several properties of $\hat{\beta}_{12}$ (and of its bootstrap analogue to be defined later).

To find the asymptotic distribution of $\hat{\beta}_{12}$, we can write

$$\sqrt{h^{-1}}\left(\hat{\beta}_{12} - \beta_{12}\right) = \frac{\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i}{\sum_{i=1}^{1/h} y_{2i}^2} = \left(\Gamma_{22}\right)^{-1}\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i + o_P(1).$$

The asymptotic variance of $\sqrt{h^{-1}}\hat{\beta}_{12}$ is thus of the usual sandwich form

$$V_\beta \equiv Var\left(\sqrt{h^{-1}}\hat{\beta}_{12}\right) = \left(\Gamma_{22}\right)^{-1} B \left(\Gamma_{22}\right)^{-1},$$

where $B = \lim_{h\to 0} B_h$, and $B_h = Var\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)$. Because $E\left(y_{2i}\varepsilon_i\right) \neq 0$, we have that

$$\begin{aligned} B_h &= Var\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right) = h^{-1}\sum_{i=1}^{1/h} Var\left(y_{2i}\varepsilon_i\right) \\ &= h^{-1}\sum_{i=1}^{1/h}\left(E\left(y_{2i}^2\varepsilon_i^2\right) - \left(E\left(y_{2i}\varepsilon_i\right)\right)^2\right) \\ &= h^{-1}\sum_{i=1}^{1/h} E\left(y_{2i}^2\varepsilon_i^2\right) - h^{-1}\sum_{i=1}^{1/h}\left(E\left(y_{2i}\varepsilon_i\right)\right)^2 \equiv B_{1h} - B_{2h}. \end{aligned}$$

We can easily show that

$$B = \lim_{h\to 0} B_h = \int_0^1 \left(\Sigma_{12}^2(u) + \Sigma_{11}(u)\Sigma_{22}(u) - 4\beta_{12}\Sigma_{12}(u)\Sigma_{22}(u) + 2\beta_{12}^2\Sigma_{22}^2(u)\right) du.$$

It follows that

$$S_{\beta,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\beta}_{12} - \beta_{12})}{\sqrt{V_\beta}} \to^d N(0,1),$$

where $V_\beta = \left(\Gamma_{22}\right)^{-2} B$. We can contrast this result with Proposition 1 of BN-S (2004). It is easy to check that $g_{12} = B$.

It is helpful to contrast the BN-S (2004) variance estimator of V_β (eq. (8)) with the Eicker-White heteroskedasticity-robust variance estimator that one would typically use in a cross section regression context. Let $\hat{\varepsilon}_i$ denote the OLS residual underlying the regression model (13). Then, the Eicker-White robust variance estimator of B is given by

$$\hat{B}_{1h} = h^{-1} \sum_{i=1}^{1/h} y_{2i}^2 \hat{\varepsilon}_i^2.$$

In contrast, noting that $x_{\beta i} = y_{2i}\hat{\varepsilon}_i$, BN-S (2004)'s estimator of B corresponds to

$$\begin{aligned} h^{-1}\hat{g}_\beta &= h^{-1}\sum_{i=1}^{1/h} x_{\beta i}^2 - h^{-1}\sum_{i=1}^{1/h-1} x_{\beta i}x_{\beta,i+1} \\ &= h^{-1}\sum_{i=1}^{1/h} y_{2i}^2\hat{\varepsilon}_i^2 - h^{-1}\sum_{i=1}^{1/h-1} y_{2i}\hat{\varepsilon}_i y_{2,i+1}\hat{\varepsilon}_{i+1} \equiv \hat{B}_{1h} - \hat{B}_{2h}. \end{aligned} \quad (14)$$

We can see that $h^{-1}\hat{g}_\beta = \hat{B}_{1h} - \hat{B}_{2h}$, where $\hat{B}_{1h}$ is the usual Eicker-White robust variance estimator, and $\hat{B}_{2h} = h^{-1}\sum_{i=1}^{1/h-1} y_{2i}\hat{\varepsilon}_i y_{2,i+1}\hat{\varepsilon}_{i+1}$. This extra term is needed to correct for the fact that $E(y_{2i}\varepsilon_i) \neq 0$, as we noted above. In particular, $\hat{B}_{1h} \to B_{1h}$ and $\hat{B}_{2h} \to B_{2h}$ in probability.

4.2 First order asymptotic properties of the pairwise bootstrap

The i.i.d. bootstrap applied to the vector of returns $y_i = (y_{1i}, y_{2i})'$ is equivalent to the so-called pairwise bootstrap, a popular bootstrap method in the context of cross section regression models. Freedman (1981) proves the consistency of the pairwise bootstrap for possibly heteroskedastic regression models when the dimension p of the regressor vector is fixed. Mammen (1993) treats the case where $p \to \infty$ as the sample size grows to infinity. Mammen (1993) also discusses the second order accuracy of the pairwise bootstrap in this context. His results specialized to the case where p is fixed show that the pairwise bootstrap is not only first order asymptotically valid under heteroskedasticity in the error term, but it is also second-order correct.

It is easy to check that $\hat{\beta}_{12}^*$ defined in (9) converges in probability (under the bootstrap probability measure P^*) to $\hat{\beta}_{12} = \frac{\sum_{i=1}^{1/h} E^*(y_{1i}^* y_{2i}^*)}{\sum_{i=1}^{1/h} E^*(y_{2i}^{*2})}$. The bootstrap analogue of the regression error ε_i in model (13) is thus $\varepsilon_i^* = y_{1i}^* - \hat{\beta}_{12}y_{2i}^*$, whereas the bootstrap OLS residuals are defined as $\hat{\varepsilon}_i^* = y_{1i}^* - \hat{\beta}_{12}^* y_{2i}^*$.

Our next Theorem provides the first order asymptotic properties of $\hat{\beta}_{12}^*$.

Theorem 4.1 *Under the conditions of Theorem 3.1, as $h \to 0$,*

a) $\sqrt{h^{-1}}\left(\hat{\beta}_{12}^* - \hat{\beta}_{12}\right) \to^{d^*} N\left(0, V_\beta^*\right)$, *in probability, where* $V_\beta^* = \left(\hat{\Gamma}_{22}\right)^{-2} B_h^*$.

b) $B_h^* = Var^*\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right) = h^{-1}\sum_{i=1}^{1/h} y_{2i}^2\hat{\varepsilon}_i^2 \equiv \hat{B}_{1h}$.

c) $V_\beta^* \to^P (\Gamma_{22})^{-2} B^* \neq V_\beta$, *where* $B^* = B + \int_0^1 (\Sigma_{12}(u) - \beta_{12}\Sigma_{22}(u))^2 du$.

Part (a) of Theorem 4.1 states that the bootstrap OLS estimator has a first order asymptotic normal distribution with mean zero and covariance matrix V_β^*. Its proof follows from Theorem 3.2. Parts (b) and (c) show that the pairwise bootstrap variance estimator is not consistent for V_β in the general context of stochastic volatility. One exception is when volatility is constant, in which case $B^* = B$ and $V_\beta^* \to^P V_\beta$.

To understand the form of V_β^*, note that we can write

$$\sqrt{h^{-1}}\left(\hat{\beta}_{12}^* - \hat{\beta}_{12}\right) = \left(\sum_{i=1}^{1/h} y_{2i}^{*2}\right)^{-1} \sqrt{h^{-1}} \sum_{i=1}^{1/h} y_{2i}^* \varepsilon_i^*.$$

Since $\sum_{i=1}^{1/h} y_{2i}^{*2} \to^{P^*} \sum_{i=1}^{1/h} y_{2i}^2 = \hat{\Gamma}_{22}$, in probability, it follows that

$$\sqrt{h^{-1}}\left(\hat{\beta}_{12}^* - \hat{\beta}_{12}\right) = \left(\hat{\Gamma}_{22}\right)^{-1} \sqrt{h^{-1}} \sum_{i=1}^{1/h} y_{2i}^* \varepsilon_i^* + o_{P^*}(1),$$

in probability. We can now apply a central limit theorem to $\sqrt{h^{-1}} \sum_{i=1}^{1/h} y_{2i}^* \varepsilon_i^*$ to obtain the limiting normal distribution for $\sqrt{h^{-1}}\left(\hat{\beta}_{12}^* - \hat{\beta}_{12}\right)$. It follows that

$$\sqrt{h^{-1}}\left(\hat{\beta}_{12}^* - \hat{\beta}_{12}\right) \to^{d^*} N\left(0, V_\beta^*\right),$$

in probability, where $V_\beta^* = \left(\hat{\Gamma}_{22}\right)^{-2} B_h^*$, with $B_h^* = Var^*\left(\sqrt{h^{-1}} \sum_{i=1}^{1/h} y_{2i}^* \varepsilon_i^*\right)$. Part (b) of Theorem 4.1 follows easily from the properties of the i.i.d. bootstrap. In particular, we can show that

$$\begin{aligned} B_h^* &= h^{-1} \sum_{i=1}^{1/h} Var^*\left(y_{2i}^* \varepsilon_i^*\right) = h^{-1} \sum_{i=1}^{1/h} \left(E^*\left(\left(y_{2i}^* \varepsilon_i^*\right)^2\right) - \left(E^*\left(y_{2i}^* \varepsilon_i^*\right)\right)^2\right) \\ &= h^{-2}\left[h \sum_{i=1}^{1/h} y_{2i}^2 \hat{\varepsilon}_i^2 - \left(h \sum_{i=1}^{1/h} y_{2i} \hat{\varepsilon}_{2i}\right)^2\right] = h^{-1} \sum_{i=1}^{1/h} y_{2i}^2 \hat{\varepsilon}_i^2 - \left(\sum_{i=1}^{1/h} y_{2i} \hat{\varepsilon}_{2i}\right)^2 \\ &= h^{-1} \sum_{i=1}^{1/h} y_{2i}^2 \hat{\varepsilon}_i^2, \end{aligned}$$

since $\sum_{i=1}^{1/h} y_{2i} \hat{\varepsilon}_{2i} = 0$ by construction of $\hat{\beta}_{12}$. Thus, the i.i.d. bootstrap variance of the scaled average of the bootstrap scores $y_{2i}^* \varepsilon_i^*$ is equal to $\hat{B}_{1h}$, the Eicker-White heteroskedasticity robust variance estimator of the scaled average of the scores $y_{2i}\varepsilon_i$.

Theorem 4.1 (part c) shows that the pairwise bootstrap does not in general consistently estimate the asymptotic variance of $\hat{\beta}_{12}$. An exception is when volatility is constant. This is in contrast with the existing results in the cross section regression context, where the pairwise bootstrap variance estimator of the least squares estimator is robust to heteroskedasticity in the error term. This failure of the pairwise bootstrap to provide a consistent estimator of the variance of $\hat{\beta}_{12}$ is related to the fact that, as we explained in in the previous section, we cannot in general assume that $E\left(y_{2i}\varepsilon_i\right) = 0$, unless for instance when volatility is constant. When the the scores have mean zero, i.e. $E\left(y_{2i}\varepsilon_i\right) = 0$, the Eicker-White robust variance estimator, and therefore the i.i.d. bootstrap variance estimator, are consistent estimators of the asymptotic variance of the scaled average of the scores. Both Freedman (1981) and Mammen (1993) make this assumption. The fact that $E\left(y_{2i}\varepsilon_i\right) \neq 0$ creates a bias term in $\hat{B}_{1h}$, which is eliminated with the variance estimator proposed by BN-S (2004) (see eq. (14)). Because $B_h^* = \hat{B}_{1h}$, the i.i.d. bootstrap variance estimator is not a consistent estimator of $B_h = Var\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)$. The non zero mean property of the scores in our context is crucial in understanding the differences between the realized regression and the usual cross section regression.

The i.i.d. bootstrap is nevertheless first order asymptotically valid when applied to the t-statistic T_h, as our Theorem 3.2 proves. This first order asymptotic validity occurs despite the fact that V_β^* does not consistently estimate V_β. The key aspect is that we studentize the bootstrap OLS estimator with $\hat{V}_\beta^*$ (defined in (11)), a consistent estimator of V_β^*, implying that the asymptotic variance of the bootstrap t-statistic is one.

4.3 Second order asymptotic properties of the pairwise bootstrap

In this section, we study the second order accuracy of the pairwise bootstrap for realized regressions. In particular, we compare the rates of convergence of the error of the bootstrap and the normal approximation when estimating the distribution function of $T_{\beta,h}$. This is accomplished via a comparison of the Edgeworth expansion of the distribution of $T_{\beta,h}$ derived by Dovonon, Gonçalves and Meddahi (2007) with the bootstrap Edgeworth expansion of $T_{\beta,h}^*$, which we derive here. See Gonçalves and Meddahi (2006b) and Zhang et al. (2005b) for two recent papers that have used Edgeworth expansions for realized volatility as a means to improve upon the first order asymptotic theory.

For $i = 1, 3$, we denote by $\kappa_i\left(T_{\beta,h}\right)$ the first and third order cumulant of $T_{\beta,h}$, respectively. The second order Edgeworth expansion of the distribution of $T_{\beta,h}$ is given by (see e.g. Hall, 1992, p. 47)

$$P\left(T_{\beta,h} \leq x\right) = \Phi\left(x\right) + \sqrt{h}q\left(x\right)\phi\left(x\right) + o\left(h\right),$$

where for any $x \in \mathbb{R}$, $\Phi(x)$ and $\phi(x)$ denote the cumulative distribution function and the density function of a standard normal random variable. The correction term $q(x)$ is defined as

$$q(x) = -\left(\kappa_1 + \frac{1}{6}\kappa_3\left(x^2 - 1\right)\right),$$

where κ_1 and κ_3 are the coefficients of the leading terms of $\kappa_1(T_{\beta,h})$ and $\kappa_3(T_{\beta,h})$, respectively. In particular, up to order $O\left(\sqrt{h}\right)$, as $h \to 0$, $\kappa_1(T_{\beta,h}) = \sqrt{h}\kappa_1$ and $\kappa_3(T_{\beta,h}) = \sqrt{h}\kappa_3$.

Given this Edgeworth expansion, the error (conditional on Σ) incurred by the normal approximation in estimating the distribution of $T_{\beta,h}$ is given by

$$\sup_{x\in\mathbb{R}}|P(T_{\beta,h} \leq x) - \Phi(x)| = \sqrt{h}\sup_{x\in\mathbb{R}}|q(x)\phi(x)| + O(h).$$

Thus, $\sup_{x\in\mathbb{R}}|q(x)\phi(x)|$ is the contribution of order $O\left(\sqrt{h}\right)$ to the normal error.

Now consider the bootstrap. We can write a one-term Edgeworth expansion for the conditional distribution of $T^*_{\beta,h}$ as follows

$$P^*(T^*_{\beta,h} \leq x) = \Phi(x) + \sqrt{h}q^*_h(x)\phi(x) + O_P(h),$$

where q^*_h is defined as

$$q^*_h(x) = -(\kappa^*_{1,h} + \kappa^*_{3,h}(x^2 - 1)/6),$$

and where $\kappa^*_{1,h}$ and $\kappa^*_{3,h}$ are the leading terms of the first and the third order cumulants of $T^*_{\beta,h}$. In particular, $\kappa^*_1\left(T^*_{\beta,h}\right) = \sqrt{h}\kappa^*_{1,h}$ and $\kappa^*_3\left(T^*_{\beta,h}\right) = \sqrt{h}\kappa^*_{3,h}$, up to order $O\left(\sqrt{h}\right)$.

The bootstrap error implicit in the bootstrap approximation of $P(T_{\beta,h} \leq x)$ (conditional on Σ) is given by

$$\begin{aligned}
P^*\left(T^*_{\beta,h} \leq x\right) - P\left(T_{\beta,h} \leq x\right) &= \sqrt{h}\ \left(q^*_h(x) - q(x)\right)\phi(x) + O_P(h) \\
&= \sqrt{h}\left(plim_{h\to 0}q^*_h(x) - q(x)\right)\phi(x) + o_P\left(\sqrt{h}\right) \\
&= -\sqrt{h}\left[\left(\kappa^*_1 - \kappa_1\right) + \frac{1}{6}\left(\kappa^*_3 - \kappa_3\right)\left(x^2 - 1\right)\right] + o_P\left(\sqrt{h}\right),
\end{aligned}$$

where $\kappa^*_1 \equiv plim_{h\to 0}\kappa^*_{1,h}$ and $\kappa^*_3 \equiv plim_{h\to 0}\kappa^*_{3,h}$. If $\kappa^*_1 = \kappa_1$ and $\kappa^*_3 = \kappa_3$, $P^*\left(T^*_{\beta,h} \leq x\right) - P\left(T_{\beta,h} \leq x\right) = o_P\left(\sqrt{h}\right)$, and the bootstrap error is of a smaller order of magnitude than the normal error which is equal to $O\left(\sqrt{h}\right)$. If this is the case, the bootstrap is said to be second-order correct and to provide an asymptotic refinement over the standard normal approximation.

The following result gives the expressions of the leading terms of the first and third order cumulants for the original statistic and for its bootstrap analogue. We need to introduce some notation. For simplicity, we will henceforth write Σ instead of $\Sigma(u)$.

Let

$$\begin{aligned}
A_0 &= \int_0^1 \left(\Sigma_{22}\Sigma_{12} - \beta_{12}\Sigma_{22}^2\right) du, \\
A_1 &= \int_0^1 \left(2\Sigma_{12}^3 + 6\Sigma_{11}\Sigma_{12}\Sigma_{22} - 18\beta_{12}\Sigma_{12}^2\Sigma_{22} - 6\beta_{12}\Sigma_{22}^2\Sigma_{11} + 24\beta_{12}^2\Sigma_{12}\Sigma_{22}^2 - 8\beta_{12}^3\Sigma_{22}^3\right) du, \\
B &= \int_0^1 \left(\Sigma_{12}^2 + \Sigma_{11}\Sigma_{22} - 4\beta_{12}\Sigma_{12}\Sigma_{22} + 2\beta_{12}^2\Sigma_{22}^2\right) du, \\
H_1 &= \frac{4A_0}{\Gamma_{22}\sqrt{B}}, \quad \text{and} \\
H_2 &= \frac{A_1}{B^{3/2}}.
\end{aligned}$$

Similarly, let

$$\begin{aligned}
B^* &= B + \int_0^1 \left(\Sigma_{12} - \beta_{12}\Sigma_{22}\right)^2 du, \\
A_1^* &= A_1 + 2\int_0^1 \left(\Sigma_{12} - \beta_{12}\Sigma_{22}\right)^3 du, \\
H_1^* &= \frac{4A_0}{\Gamma_{22}\sqrt{B^*}}, \\
H_2^* &= \frac{A_1^*}{B^{*3/2}}.
\end{aligned}$$

In order to obtain the higher order results in this section, we add the following additional assumption. A more primitive assumption such as a multivariate analogue of Assumption V in Gonçalves and Meddahi (2006) may be sufficient to ensure Assumption 3, but we have not yet confirmed this.

Assumption 3 Let $k, l, k', l' = 1, 2$.

$$h^{-1}\sum_{i=1}^{1/h} \Gamma_{kl,i}\Gamma_{k'l',i} - \int_0^1 \Sigma_{kl}(u)\Sigma_{k'l'}(u)du = o_P(\sqrt{h}),$$

and

$$h^{-1}\sum_{i=1}^{1/h-1} \Gamma_{kl,i}\Gamma_{k'l',i+1} - \int_0^1 \Sigma_{kl}(u)\Sigma_{k'l'}(u)du = o_P(\sqrt{h}).$$

Proposition 4.1 *Under Assumptions 1, 2 and 3,*

a) $\kappa_1 = \frac{1}{2}\left(H_1 - H_2\right)$ *and* $\kappa_3 = 3H_1 - 2H_2$.

b) $\kappa_1^* = \frac{3}{4}\left(H_1^* - H_2^*\right)$ *and* $\kappa_3^* = 3\left(\frac{3}{2}H_1^* - H_2^*\right)$.

Part (a) of Proposition 4.1 is derived in Dovonon, Gonçalves and Meddahi (2007). (We reproduce the proof in Appendix C for completeness.) The proof of part (b) is in Appendix C. A comparison of the two parts reveals a disagreement between the two sets of cumulants. Notice in particular that $B \neq B^*$ contributes to this discrepancy. B here denotes the limiting variance of the scaled average of the scores whereas B^* denotes its bootstrap analogue. As we noted before, under general stochastic volatility, the pairwise bootstrap does not consistently estimate B and the bias term is exactly equal to the difference between B and B^*, i.e. $B^* - B = \int_0^1 \left(\Sigma_{12} - \beta_{12}\Sigma_{22}\right)^2 du = plim_{h\to 0} B_{2h}$, where $B_{2h} = h^{-1}\sum_{i=1}^{1/h}\left(E\left(y_{2i}\varepsilon_i\right)\right)^2$. An exception is when volatility is constant, where $B_{2h} = 0$ and therefore $B^* = B$. In this case, we also have that $A_1^* = A_1 = A_0 = 0$, implying that both the bootstrap and the normal approximations have an error of the order $O(h)$. We need a higher order expansion to be able to discriminate the two approximations. In the general stochastic volatility case, the pairwise bootstrap error is of order $O\left(\sqrt{h}\right)$, similar to the error incurred by the normal approximation.

The lack of second order refinements of the pairwise bootstrap in the context of realized regressions is in contrast with the results available in the bootstrap literature for standard regression models (see Mammen 1993). One explanation for this difference lies in the fact that $E\left(y_{2i}\varepsilon_i\right) \neq 0$, as we noted above. This implies that $T_{\beta,h}$ must rely on a variance estimator that contains a bias correction term, as proposed by BN-S (2004). Instead, in the bootstrap regression, $E^*\left(y_{2i}^*\varepsilon_i^*\right) = h\sum_{i=1}^{1/h} y_{2i}\hat{\varepsilon}_i = 0$, and therefore there is no need for the bias correction proposed by BN-S (2004). This implies that the bootstrap t-statistic $T_{\beta,h}^*$ is not of the same form as $T_{\beta,h}$, relying on a bootstrap variance estimator $\hat{V}_\beta^*$ that depends on an Eicker-White type variance estimator $\hat{B}_{1h}^*$.

5 Empirical application

A well documented empirical fact in finance is the time variability of bonds risk, as recently documented by Viceira (2007) for the US market. As suggested by the CAPM, the bond risk is often measured by its beta over the return on the market portfolio. With a positive beta, bonds are considered as risky as the market while a bond with a negative beta could be used to hedge the market risk.

Following Merton (1980) and French, Schwert and Stambaugh (1987), Viceira (2007) studies the bond risk for the US market by considering the 3-month (monthly) rolling realized beta as measured by the ratio of the realized covariance of daily log-returns on bonds and stocks and the realized volatility of the daily log-return on stocks over the same period. Following the standard practice, the number of days in a month is normalized to 22 such that the 3-month realized beta is computed considering sub-samples of 66 days. From July 1962 through December 2003, Viceira (2007) reports a strong variability of US bond CAPM betas, which may switch sign even though the average over the full sample is positive. Nevertheless, in his analysis Viceira (2007) does not discuss the precision of the realized betas as a measure of the actual covariation between bonds and stock returns.

The aim of this section is to illustrate the usefulness of our approach as a method of inference for realized covariation measures in the context of measuring the time variation of bonds risk. We consider both the US bonds market, as in Viceira (2007), and the UK bonds market.

Our data set includes the daily 7-to-10-year maturity government bond index for the US and the UK markets as released by JP Morgan from January 2, 1986 through August 24, 2007. As a proxy for the US and the UK market portfolio returns, we consider the log-return on the S&P500 and the FTSE 100 indices, respectively. The S&P500 index is designed to measure performance of the broad domestic economy through changes in the aggregate market value of 500 stocks representing all major industries. The FTSE 100 index is a capitalization-weighted index of the 100 most highly capitalized companies traded on the London Stock Exchange. Both indices are commonly used in scientific researches as well as in the finance industry as a proxy for the market portfolio. The first two series have a shorter history and therefore constrained the sample we consider in this study.

From the estimates presented in Table 3.4 (Appendix A), the full-sample beta for bonds in the US is about 0.024, slightly smaller than the UK bond beta, which is about 0.030. Both the bootstrap and the asymptotic theory based confidence intervals display support that the true values of the betas in both countries are positive.

A closer analysis of Figures 3.1 and 3.2 shows that the average positivity of the betas hides considerable time variation in both countries, a fact already documented by Viceira (2007) for the US market. Furthermore, the betas for these two countries follow similar dynamics. We can distinguish two patterns for the 3-month betas. For the period before April 1997, the betas are mostly significantly positive or, in few cases, non-significantly different from 0. This period is also characterized by betas of larger magnitude, with a maximum value of 0.500 at the end of July 1994 for the US and 0.438 in

August 1994 for the UK. The period after April 1997 is characterized by a drop of the magnitude of the bonds betas in both countries. They are often not significantly different from 0. For this whole sub-period, the betas for the US and UK bonds are significantly negative only between June 2002 and July 2003, but in these cases their magnitude is small. We conclude that bonds are riskier in the period before April 1997, while in the recent periods they appear to be non risky or at most a hedging instrument against shocks on market portfolio returns.

A comparison of the bootstrap intervals with the intervals based on the asymptotic theory of BN-S (2004) suggests that they two types of intervals tend to be similar, but there are instances where the bootstrap intervals are wider than the asymptotic theory-based intervals (see Tables 3.5 and 3.6 for a detailed comparison of the two types of intervals for a selected set of dates). This is specially true for the first part of the sample for the UK bond market, where the width of the bootstrap intervals can be much larger than the width of the BN-S (2004) intervals. In this empirical application, the gain in accuracy of the bootstrap intervals in terms of coverage probability appears to be associated with a deterioration of length of the bootstrap intervals.

6 Conclusion

This paper proposes bootstrap methods for inference on measures of multivariate volatility such as integrated covariance, integrated correlation and integrated regression coefficients. We show the first order asymptotic validity of a particular bootstrap scheme, the i.i.d. bootstrap applied to the vector of returns, for the three statistics of interest. Our simulation results show that the bootstrap outperforms the feasible first order asymptotic approach of BN-S(2004).

For the special case of the realized regression estimator, our i.i.d. bootstrap corresponds to a pairwise bootstrap as proposed by Freedman (1981) and further studied by Mammen (1993). We analyze the second order accuracy of this bootstrap method and conclude that it is not second order accurate. This contrasts with the existing literature on the pairwise bootstrap for cross section models, which shows that this method is not only robust to heteroskedasticity in the error term but it is also second order accurate. We provide a detailed analysis of the pairwise bootstrap in the context of realized regressions which allows us to highlight some key differences with respect to the usual application of the pairwise bootstrap in standard cross section regression models. These differences explain why the pairwise bootstrap does not provide second order refinements in this context.

An important characteristic of high frequency financial data that our theory ignores is the presence of microstructure effects: the prices are observed with contamination errors called noise due to the presence of bid-ask bounds, rounding errors, etc, and prices are asynchronous, i.e., the prices of two assets are often not observed at the same time. The first problem is well addressed by the literature in the univariate context, in particular, Zhang, Mykland, and Ait-Sahalia (2005a), Zhang (2006), and Barndorff-Nielsen, Hansen, Lunde and Shephard (2007) provide consistent estimators of the integrated volatility. Likewise, Hayashi and Yoshida (2005) provide a consistent estimator of the covariation of two assets when they are asynchronous, but their analysis rules out the presence of noise. Little is known when the two effects are present; see however the analysis in Zhang (2006), Griffin and Oomen (2006) and Voev and Lunde (2007). Another feature that our theory ignores is the possible presence of jumps and co-jumps. This is a difficult problem that the literature has only started recently to address (see Jacod and Todorov (2007) and Bollerslev and Todorov (2007)). The extension of our bootstrap theory to these important problems is left for future research.

Appendix A

Table 3.1: Coverage rates of nominal 95% intervals for covariation measures

	Covariance					Regression					Correlation						
	One-sided		Two-sided			One-sided		Two-sided			One-sided			Two-sided			
1/h	BN-S	Boot	BN-S	Boot		BN-S	Boot	BN-S	Boot		BN-S	Fisher	Boot	BN-S	Fisher	Boot	
				Sym	Eq-T.				Sym	Eq-T						Sym	Eq-T
						Design 1											
12	80.74	85.55	83.90	86.86	82.43	86.04	89.86	85.27	93.48	90.42	91.43	90.28	94.81	81.04	85.44	93.69	94.60
24	84.69	89.60	87.42	89.86	87.19	88.53	91.58	89.22	93.61	91.88	93.65	92.29	94.72	86.97	89.15	94.08	94.47
48	87.67	91.44	90.16	91.41	89.74	91.07	93.21	91.50	94.50	93.42	94.62	93.47	95.09	90.33	91.51	94.44	94.66
288	92.48	93.29	94.01	92.74	92.43	93.81	94.72	94.34	95.06	94.53	95.38	94.83	94.87	94.23	94.46	95.11	94.95
1152	94.19	94.22	94.85	93.34	93.05	94.46	95.13	94.81	95.15	95.02	95.39	95.15	94.93	94.87	94.95	94.88	94.83
						Design 2											
12	81.83	86.87	84.85	87.45	83.22	87.29	91.11	86.30	93.61	91.08	92.64	90.90	95.71	81.74	85.99	93.68	95.38
24	86.07	90.38	88.90	90.11	87.34	89.98	92.27	90.08	93.81	91.93	94.51	93.19	95.34	87.39	89.73	93.86	94.51
48	88.48	91.97	90.82	91.55	89.87	91.17	93.48	91.70	94.87	93.66	94.81	93.45	95.01	90.86	91.93	94.38	94.60
288	92.39	93.41	94.25	93.03	92.41	93.72	94.94	94.28	95.17	94.77	95.39	94.71	95.05	94.11	94.23	94.89	95.12
1152	94.25	93.66	94.75	92.39	92.32	94.54	95.07	94.68	95.15	95.06	95.65	95.27	94.98	94.70	94.77	95.17	95.08

Note: 10,000 replications, with 999 bootstrap replications each.

Table 3.2: Coverage rates of nominal 90% intervals for covariation measures

	Covariance					Regression					Correlation						
	One-sided		Two-sided			One-sided		Two-sided			One-sided			Two-sided			
$1/h$	BN-S	Boot	BN-S	Boot		BN-S	Boot	BN-S	Boot		BN-S	Fisher	Boot	BN-S	Fisher	Boot	
				Sym	Eq-T				Sym	Eq-T						Sym	Eq-T
							Design 1										
12	75.20	80.58	79.17	79.15	77.04	80.52	84.81	79.36	87.32	85.37	87.21	85.46	90.07	75.07	78.94	86.54	89.53
24	78.92	84.25	82.56	81.75	81.46	82.98	86.83	83.06	87.25	86.85	89.03	87.45	89.83	81.12	83.40	86.85	89.50
48	82.30	86.37	85.21	83.98	83.97	85.92	88.21	86.22	87.66	88.25	89.90	88.68	90.37	84.35	85.75	87.41	89.90
288	87.31	88.29	89.02	84.69	86.51	88.89	90.01	89.13	87.91	89.74	90.57	89.88	89.77	88.94	89.20	87.76	90.06
1152	89.31	88.44	90.00	84.62	87.26	89.34	90.30	89.66	88.31	90.06	90.88	90.42	89.97	89.74	89.86	87.79	90.06
							Design 2										
12	76.17	81.80	79.91	79.55	77.43	82.06	86.26	79.98	87.45	85.83	88.20	86.43	90.83	76.01	79.30	86.74	90.16
24	80.53	85.40	83.77	82.53	81.81	84.45	87.47	83.98	87.21	86.86	90.19	88.22	90.72	82.11	84.22	87.16	89.83
48	83.19	86.58	85.93	83.46	83.81	85.90	88.47	86.18	87.70	88.25	90.22	88.77	89.90	85.37	86.43	87.44	89.58
288	87.27	88.04	89.00	84.35	86.13	88.49	90.01	89.07	88.21	89.81	90.71	89.92	90.01	88.66	88.98	87.69	90.05
1152	89.14	88.57	90.03	84.34	86.60	89.66	90.17	89.54	87.90	90.03	90.73	90.28	90.08	89.57	89.68	87.95	90.18

Note: 10,000 replications, with 999 bootstrap replications each.

Table 3.3: Coverage rates of nominal 99% intervals for covariation measures

	Covariance					Regression					Correlation						
	One-sided		Two-sided			One-sided		Two-sided			One-sided			Two-sided			
$1/h$	BN-S	Boot	BN-S	Boot		BN-S	Boot	BN-S	Boot		BN-S	Fisher	Boot	BN-S	Fisher	Boot	
				Sym	Eq-T				Sym	Eq-T						Sym	Eq-T
							Design 1										
12	88.10	90.80	89.98	92.59	88.89	92.45	95.67	92.15	97.98	96.29	95.99	95.19	98.59	88.98	93.01	98.67	98.66
24	91.49	94.67	93.10	95.21	93.39	94.84	96.30	95.12	97.76	96.43	97.83	96.87	98.62	93.39	95.55	98.76	98.66
48	93.87	96.56	95.47	96.79	95.74	96.28	97.54	96.80	98.26	97.67	98.35	97.71	98.85	96.07	97.08	98.81	98.79
288	97.41	98.45	98.13	98.42	97.81	98.16	98.61	98.43	98.80	98.64	99.08	98.78	99.00	98.38	98.54	99.02	98.93
1152	98.48	98.65	98.97	98.37	98.10	98.84	99.02	98.93	99.06	98.90	99.25	99.03	98.96	98.77	98.89	98.95	98.89
							Design 2										
12	88.80	91.96	90.63	93.16	89.53	93.78	96.36	93.20	97.92	96.43	97.11	96.30	98.92	89.11	93.78	98.85	98.98
24	92.60	95.43	94.06	95.66	93.80	95.90	96.82	96.24	98.17	96.80	98.29	97.43	99.10	93.90	95.90	98.48	98.71
48	94.64	97.04	95.91	97.12	96.07	96.68	97.82	97.09	98.42	97.91	98.65	98.06	99.07	96.29	97.34	98.60	98.89
288	97.66	98.40	98.55	98.39	97.85	98.21	98.93	98.57	99.08	98.84	99.19	98.88	99.06	98.69	98.79	98.95	98.97
1152	98.44	98.46	98.91	98.16	97.90	98.62	98.98	98.81	98.95	98.84	99.14	98.99	99.08	98.82	98.83	99.08	99.02

Note: 10,000 replications, with 999 bootstrap replications each.

Figure 3.1: Symmetric bootstrap and BN-S (2004) asymptotic theory based 95% two-sided confidence intervals for the CAPM 3-month (monthly) rolling realized beta of US bond. April 1986 through July 2007.

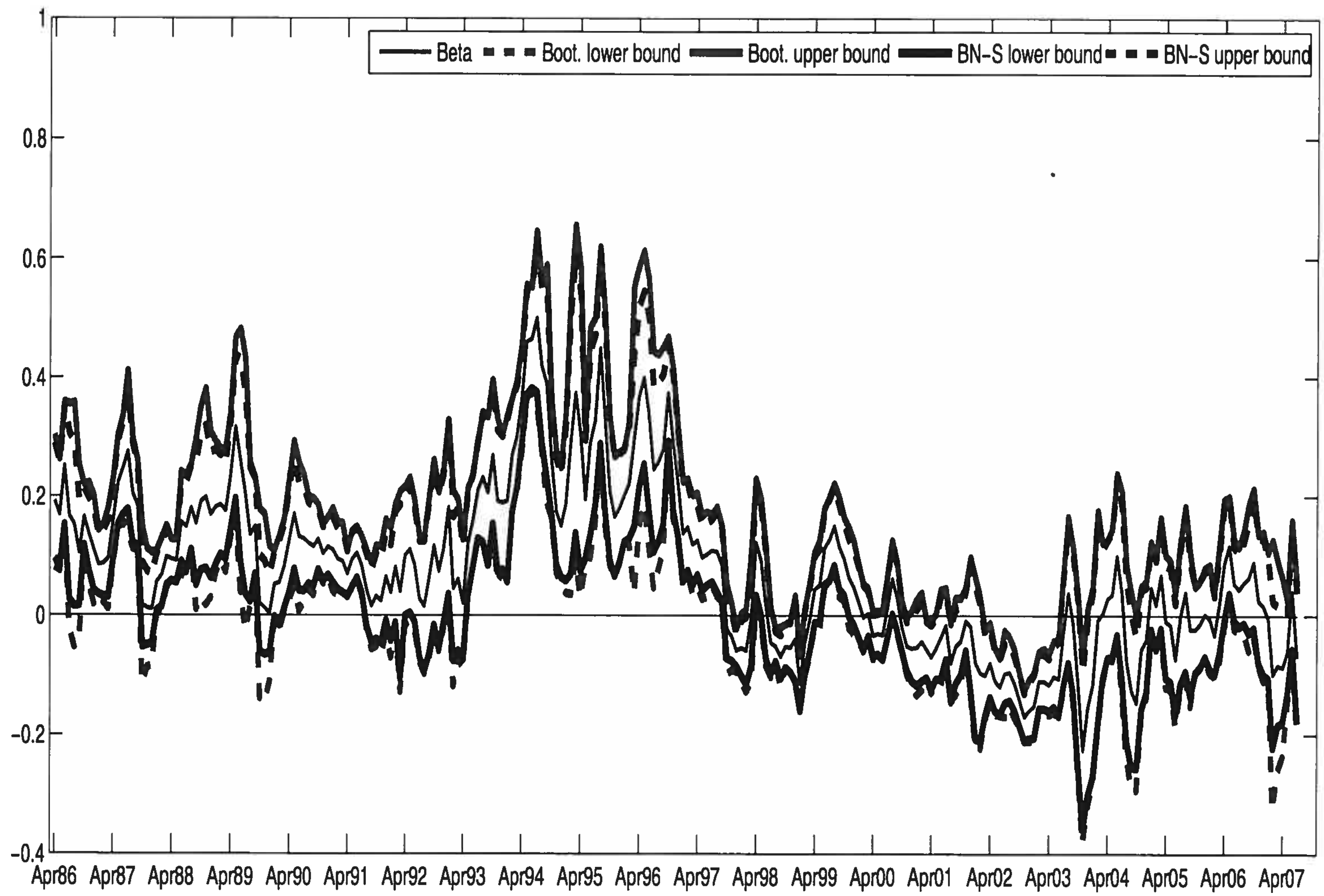


Figure 3.2: Symmetric bootstrap and BN-S (2004) asymptotic theory based 95% two-sided confidence intervals for the CAPM 3-month (monthly) rolling realized beta of UK bond. April 1986 through July 2007.

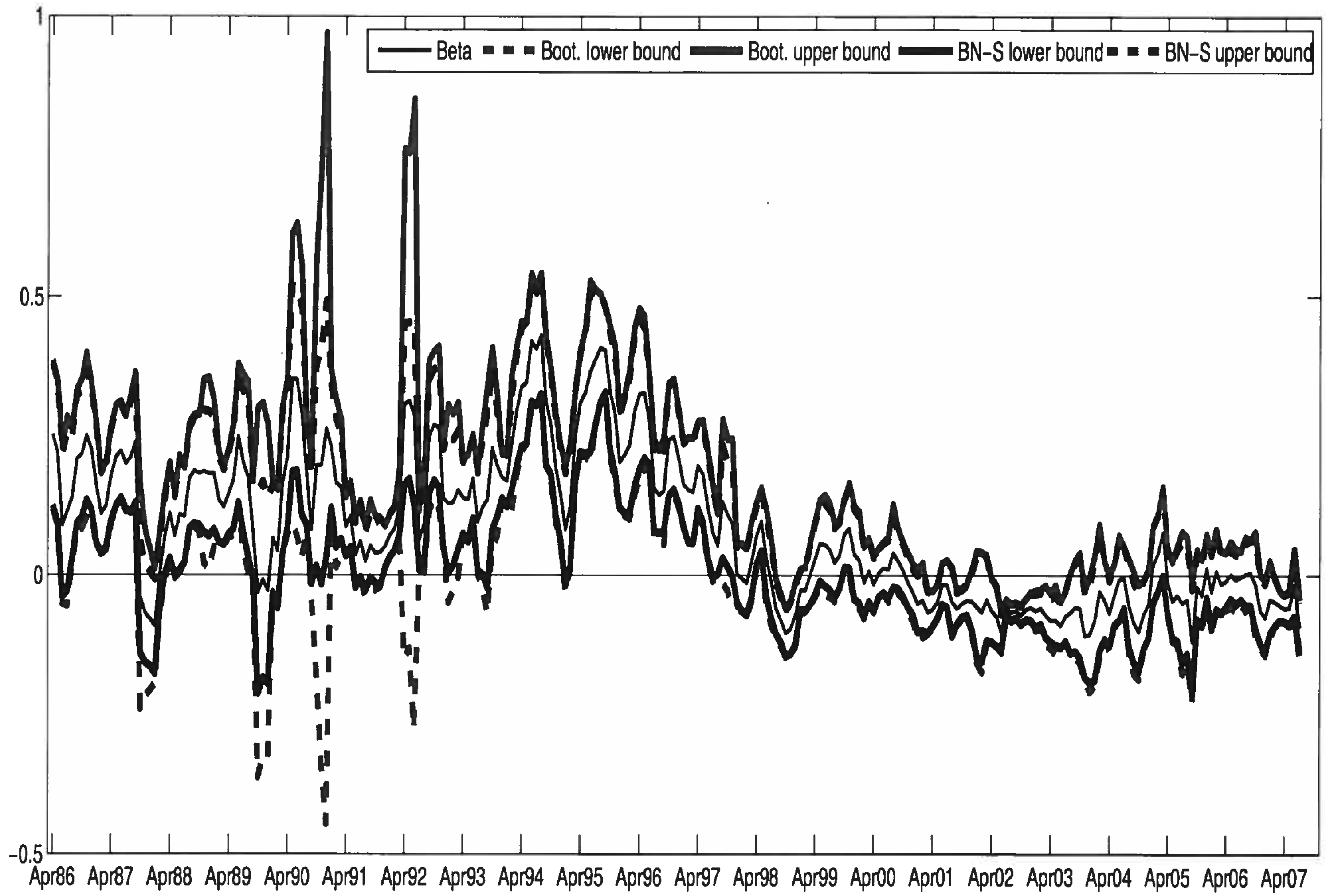


Table 3.4: Full-sample estimates of bonds betas for the US and the UK from January 2, 1986 through August 24, 2007

	Beta	BN-S 95% 2-sided CI	Boot. symm. 95% CI
US			
	0.024	$[0.010, 0.038]$	$[0.009, 0.038]$
UK			
	0.030	$[0.016, 0.045]$	$[0.015, 0.046]$

Table 3.5: Divergence between BN-S and bootstrap confidence intervals for the US

Date	Beta	BN-S	Bootstrap
31-Jul-86	0.167	$[0.027, 0.306]$	$[-0.022, 0.355]$
29-Aug-86	0.152	$[0.015, 0.289]$	$[-0.053, 0.357]$
30-Sep-86	0.106	$[0.017, 0.194]$	$[-0.041, 0.252]$
31-Jul-89	0.204	$[0.036, 0.371]$	$[-0.025, 0.432]$
29-May-92	0.111	$[0.004, 0.217]$	$[-0.010, 0.231]$
29-May-98	0.093	$[0.001, 0.184]$	$[-0.012, 0.197]$
31-Aug-00	0.062	$[0.002, 0.121]$	$[-0.002, 0.126]$
30-Jan-98	-0.054	$[-0.101, -0.008]$	$[-0.111, 0.003]$
27-Feb-98	-0.059	$[-0.115, -0.002]$	$[-0.128, 0.010]$
29-Dec-00	-0.055	$[-0.109, -0.000]$	$[-0.117, 0.008]$
31-May-01	-0.055	$[-0.107, -0.004]$	$[-0.113, 0.003]$
31-Dec-03	-0.154	$[-0.302, -0.005]$	$[-0.319, 0.011]$
29-Oct-04	-0.146	$[-0.256, -0.036]$	$[-0.293, 0.001]$

Appendix B

This Appendix contains the proofs of the results in Section 3. We first present two auxiliary lemmas.

Lemma B.1 *Let y_{ji} denote the jth component of y_i. Under Assumptions 1 and 2, for any $q_1, q_2 \geq 0$ such that $q_1 + q_2 > 0$, $h^{1-(q_1+q_2)/2} \sum_{i=1}^{1/h} |y_{1i}|^{q_1} |y_{2i}|^{q_2} = O_P(1)$.*

Proof of Lemma B.1. $\sum_{i=1}^{1/h} |y_{1i}|^{q_1} |y_{2i}|^{q_2} \leq \left(\sum_{i=1}^{1/h} |y_{1i}|^{2q_1}\right)^{1/2} \left(\sum_{i=1}^{1/h} |y_{2i}|^{2q_2}\right)^{1/2}$ by the Cauchy-Schwarz inequality. From Theorem 1 of BN-S (2004), $\sum_{i=1}^{1/h} |y_{1i}|^{2q_1} = O_P(h^{-1+q_1})$ and $\sum_{i=1}^{1/h} |y_{2i}|^{2q_2} = O_P(h^{-1+q_2})$, which proves the result.

Lemma B.2 *Let $\{y_i^* : i = 1, \ldots, 1/h\}$ denote an i.i.d. bootstrap sample of intraday returns $\{y_i : i = 1, \ldots, 1/h\}$ and assume that Assumptions 1 and 2 hold. Then, for k, l, k', $l' = 1, 2$, with probability approaching one,*

i) $\sum_{i=1}^{1/h} y_{ki}^* y_{li}^* \xrightarrow{P^*} \sum_{i=1}^{1/h} y_{ki} y_{li}$.

ii) $h^{-1} \sum_{i=1}^{1/h} y_{ki}^* y_{li}^* y_{k'i}^* y_{l'i}^* \xrightarrow{P^*} h^{-1} \sum_{i=1}^{1/h} y_{ki} y_{li} y_{k'i} y_{l'i}$.

Table 3.6: Divergence between BN-S and bootstrap confidence intervals for the UK

Date	Beta	BN-S	Bootstrap
31-Mar-88	0.070	$[0.003, 0.137]$	$[-0.010, 0.150]$
31-Oct-90	0.197	$[0.016, 0.377]$	$[-0.175, 0.568]$
31-Dec-90	0.262	$[0.031, 0.493]$	$[-0.446, 0.970]$
30-Apr-92	0.307	$[0.162, 0.452]$	$[-0.151, 0.764]$
29-May-92	0.314	$[0.173, 0.454]$	$[-0.131, 0.758]$
30-Jun-92	0.288	$[0.125, 0.450]$	$[-0.277, 0.852]$
29-Jan-93	0.129	$[0.003, 0.254]$	$[-0.049, 0.306]$
26-Feb-93	0.131	$[0.018, 0.243]$	$[-0.029, 0.290]$
31-Mar-93	0.153	$[0.046, 0.259]$	$[-0.004, 0.309]$
31-Aug-93	0.122	$[0.001, 0.242]$	$[-0.025, 0.268]$
29-Aug-97	0.054	$[0.002, 0.105]$	$[-0.003, 0.111]$
30-Sep-97	0.132	$[0.031, 0.233]$	$[-0.015, 0.279]$
31-Oct-97	0.109	$[0.015, 0.202]$	$[-0.027, 0.244]$
29-Jan-88	-0.092	$[-0.177, -0.007]$	$[-0.195, 0.012]$
31-Jan-01	-0.052	$[-0.102, -0.001]$	$[-0.111, 0.008]$
30-Sep-04	-0.085	$[-0.156, -0.014]$	$[-0.177, 0.008]$
30-Nov-06	-0.064	$[-0.122, -0.005]$	$[-0.129, 0.002]$

Proof of Lemma B.2. We show that the results hold in quadratic mean with respect to the bootstrap measure, with probability approaching one. This ensures that the bootstrap convergence also holds in probability. For (i), we have $E^*\left(\sum_{i=1}^{1/h} y_{ki}^* y_{li}^*\right) = h^{-1}E^*\left(y_{k1}^* y_{l1}^*\right) = h^{-1}h\sum_{i=1}^{1/h} y_{ki}y_{li} = \sum_{i=1}^{1/h} y_{ki}y_{li}$. Similarly,

$$Var^*\left(\sum_{i=1}^{1/h} y_{ki}^* y_{li}^*\right) = h^{-1}Var^*\left(y_{k1}^* y_{l1}^*\right) = h^{-1}\left(E^*(y_{k1}^* y_{l1}^*)^2 - (E^* y_{k1}^* y_{l1}^*)^2\right)$$

$$= h^{-1}\left(h\sum_{i=1}^{1/h}(y_{ki}y_{li})^2 - \left(h\sum_{i=1}^{1/h} y_{ki}y_{li}\right)^2\right) = \sum_{i=1}^{1/h}(y_{ki}y_{li})^2 - h\left(\sum_{i=1}^{1/h} y_{ki}y_{li}\right)^2 = o_P(1),$$

given that Lemma B.1 implies that $\sum_{i=1}^{1/h}(y_{ki}y_{li})^2 = O_P(h) = o_P(1)$ and $\sum_{i=1}^{1/h} y_{ki}y_{li} = O_P(1)$. This proves the result. The proof of (ii) follows similarly and therefore we omit the details.

Proof of Theorem 3.1. The proof of (a) follows from Lemma B.2 by noting that the elements of $x_i^* x_i^{*\prime}$ are of all of the form $y_{ki}^* y_{li}^* y_{k'i}^* y_{l'i}^*$, for $k, l, k', l' = 1, 2$.

To prove (b), we first show that both $\hat{V}^*$ and V^* are non singular in large samples with probability approaching one, as the sample size grows. The probability limit of V^* follows from Theorem 4 of BN-S (2004) and is equal to

$$\begin{pmatrix} 3\int_0^1 \Sigma_{11}^2 du - \Gamma_{11}^2 & 3\int_0^1 \Sigma_{11}\Sigma_{12} du - \Gamma_{11}\Gamma_{12} & \int_0^1 (\Sigma_{11}\Sigma_{22} + 2\Sigma_{12}^2) du - \Gamma_{11}\Gamma_{22} \\ & \int_0^1 (\Sigma_{11}\Sigma_{22} + 2\Sigma_{12}^2) du - \Gamma_{12}^2 & 3\int_0^1 \Sigma_{12}\Sigma_{22} du - \Gamma_{12}\Gamma_{22} \\ & & 3\int_0^1 \Sigma_{22}^2 du - \Gamma_{22}^2 \end{pmatrix},$$

which can be written as $V + V_1$ where

$$V_1 = \begin{pmatrix} \int_0^1 \Sigma_{11}^2 du - \Gamma_{11}^2 & \int_0^1 \Sigma_{11}\Sigma_{12} du - \Gamma_{11}\Gamma_{12} & \int_0^1 \Sigma_{11}\Sigma_{22} du - \Gamma_{11}\Gamma_{22} \\ & \int_0^1 \Sigma_{11}\Sigma_{22} du - \Gamma_{12}^2 & \int_0^1 \Sigma_{12}\Sigma_{22} du - \Gamma_{12}\Gamma_{22} \\ & & \int_0^1 \Sigma_{22}^2 du - \Gamma_{22}^2 \end{pmatrix}.$$

V is the asymptotic variance of $\sqrt{h^{-1}}\sum_{i=1}^{1/h} x_i$ and it is pathwise symmetric positive definite by assumption. We show that V_1 is positive semidefinite, which guarantees the positive definiteness of $V + V_1$. For any $\lambda \in \mathbb{R}^3$, by straightforward calculation,

$$\lambda' V_1 \lambda = \int_0^1 (\lambda_1 \Sigma_{11}(u) + \lambda_2 \Sigma_{12}(u) + \lambda_3 \Sigma_{22}(u))^2 du - (\lambda_1 \Gamma_{11} + \lambda_2 \Gamma_{12} + \lambda_3 \Gamma_{22})^2 \geq 0$$

by the Jensen inequality. Thus, V_1 is positive-semidefinite and therefore V^* is positive definite.

Now, let $S_h^* = V^{*-1/2}\sqrt{h^{-1}}(\sum_{i=1}^{1/h} x_i^* - \sum_{i=1}^{1/h} x_i)$. Clearly, $T_h^* = \hat{V}^{*^{-1/2}} V^{*1/2} S_h^*$. As we just showed, $\hat{V}^{*-1} V^* \xrightarrow{P^*} I_3$, in probability. Thus, the proof of (b) follows from showing that for any $\lambda \in \mathbb{R}^3$ such that $\lambda'\lambda = 1$, $\sup_{x \in \mathbb{R}} |P^*(\sum_{i=1}^{1/h} \tilde{x}_i^* \leq x) - \Phi(x)| \xrightarrow{P} 0$, where $\tilde{x}_i^* = (\lambda' V^* \lambda)^{-1/2} \sqrt{h^{-1}} \lambda' (x_i^* - E^*(x_i^*))$, and where $\Phi(x)$ is the standard Gaussian cumulative distribution function. Clearly, $E^*\left(\sum_{i=1}^{1/h} \tilde{x}_i^*\right) = 0$ and $Var^*\left(\sum_{i=1}^{1/h} \tilde{x}_i^*\right) = 1$. Thus, by Katz's (1963) Berry-Essen Bound, for some small $\epsilon > 0$ and some constant $K > 0$,

$$\sup_{x \in \mathbb{R}} \left| P^*\left(\sum_{i=1}^{1/h} \tilde{x}_i^* \leq x\right) - \Phi(x) \right| \leq K \sum_{i=1}^{1/h} E^* |\tilde{x}_i^*|^{2+\epsilon}.$$

Next, we show that $\sum_{i=1}^{1/h} E^* |\tilde{x}_i^*|^{2+\epsilon} = o_p(1)$. We have

$$\begin{aligned} \sum_{i=1}^{1/h} E^* |\tilde{x}_i^*|^{2+\epsilon} &= h^{-1} E^* |\tilde{x}_1^*|^{2+\epsilon} = h^{-1} E^* \left| (\lambda' V^* \lambda)^{-1/2} h^{-1/2} \lambda' (x_1^* - E^*(x_1^*)) \right|^{2+\epsilon} \\ &= h^{-1} h^{-(2+\epsilon)/2} |\lambda' V^* \lambda|^{-(2+\epsilon)/2} E^* |\lambda' (x_1^* - E^*(x_1^*))|^{2+\epsilon} \\ &\leq 2^{2+\epsilon} h^{-(2+\epsilon/2)} |\lambda' V^* \lambda|^{-(1+\epsilon/2)} E^* |\lambda' x_1^*|^{2+\epsilon} \\ &\leq 2^{2+\epsilon} h^{-(2+\epsilon/2)} |\lambda' V^* \lambda|^{-(1+\epsilon/2)} E^* |x_1^*|^{2+\epsilon} \\ &= 2^{2+\epsilon} h^{-1-\epsilon/2} |\lambda' V^* \lambda|^{-(1+\epsilon/2)} \sum_{i=1}^{1/h} |x_i|^{2+\epsilon}, \end{aligned}$$

where the first inequality follows from the C_r and the Jensen inequalities and the second inequality follows from the Cauchy-Schwarz inequality and the fact that $\lambda'\lambda = 1$. We let $|z| = (z'z)^{1/2}$ for any vector z. It follows that $\sum_{i=1}^{1/h} |x_i|^{2+\epsilon} = \sum_{i=1}^{1/h} |x_i|^{2(1+\epsilon/2)} \leq \sum_{i=1}^{1/h} (y_{1i}^2 + y_{2i}^2)^{2(1+\epsilon/2)}$, since $|x_i|^2 = (y_{1i}^4 + y_{1i}^2 y_{2i}^2 + y_{2i}^4)^2 \leq y_{1i}^4 + 2y_{1i}^2 y_{2i}^2 + y_{2i}^4 = (y_{1i}^2 + y_{2i}^2)^2$. Thus, $\sum_{i=1}^{1/h} |x_i|^{2+\epsilon} \leq \sum_{i=1}^{1/h} (y_{1i}^2 + y_{2i}^2)^{2+\epsilon}$. By the Minkowski's inequality,

$$\sum_{i=1}^{1/h} |x_i|^{2+\epsilon} \leq \left\{ \left(\sum_{i=1}^{1/h} |y_{1i}|^{4+2\epsilon} \right)^{1/(2+\epsilon)} + \left(\sum_{i=1}^{1/h} |y_{2i}|^{4+2\epsilon} \right)^{1/(2+\epsilon)} \right\}^{2+\epsilon}.$$

By Lemma B.1, $\sum_{i=1}^{1/h} |x_i|^{2+\epsilon} = O_P(h^{1+\epsilon})$. Therefore, $\sum_{i=1}^{1/h} E^* |\tilde{x}_i^*|^{2+\epsilon} = O_P(h^{\epsilon/2}) = o_P(1)$.

Proof of Theorem 3.2. Since $T_h \xrightarrow{d} N(0, I_3)$, by the standard delta method, $T_{f,h} \xrightarrow{d} N(0,1)$. Similarly, by a mean value expansion, and conditionally on the original sample,

$$\sqrt{h^{-1}}\left(f(vech(\hat{\Gamma}^*)) - f(vech(\hat{\Gamma}))\right) = \sqrt{h^{-1}}\nabla' f\left(vech(\hat{\Gamma})\right)\left(vech(\hat{\Gamma}^*) - vech(\hat{\Gamma})\right) + o_{P^*}(1),$$

since $\hat{\Gamma}^* \to^{P^*} \hat{\Gamma}$ in probability. Let

$$S^*_{f,h} \equiv \frac{\sqrt{h^{-1}}\left(f(vech(\hat{\Gamma}^*)) - f(vech(\hat{\Gamma}))\right)}{\sqrt{V^*_f}},$$

with $V^*_f \equiv \nabla' f(vech(\hat{\Gamma}))V^*\nabla f(vech(\hat{\Gamma}))$. It follows that $S^*_{f,h} \to^{d^*} N(0,1)$ in probability, given Theorem 3.1 (b). Next note that $T^*_{f,h} = \sqrt{\frac{V^*_f}{\hat{V}^*_f}} S^*_{f,h}$, where $\hat{V}^*_f \to^{P^*} V^*_f$. The result follows from Polya's theorem (e.g. Serfling, 1980) given that the normal distribution is continuous.

Proof of Theorem 4.1. Part (a) follows from Theorem 3.2 with $f(\theta) = \theta_2/\theta_3$. To derive V^*_β, let $\hat{\theta}_1 = \sum_{i=1}^{1/h} y_{1i}^2$, $\hat{\theta}_2 = \sum_{i=1}^{1/h} y_{1i}y_{2i}$, $\hat{\theta}_3 = \sum_{i=1}^{1/h} y_{2i}^2$. Clearly, $\hat{\beta}_{12} = f(\hat{\theta})$ and $\nabla f(\hat{\theta}) = \begin{pmatrix} 0 & \frac{1}{\hat{\theta}_3} & -\frac{\hat{\theta}_2}{\hat{\theta}_3^2}\end{pmatrix}'$. Then V^*_β is given by $V^*_\beta = \nabla' f(\hat{\theta}))V^*\nabla f(\hat{\theta})$, with $V^* = h^{-1}\sum_{i=1}^{1/h} x_i x_i' - \left(\sum_{i=1}^{1/h} x_i\right)\left(\sum_{i=1}^{1/h} x_i\right)'$. Straightforward calculations show that
$\nabla' f(\hat{\theta}))\left(h^{-1}\sum_{i=1}^{1/h} x_i x_i'\right)\nabla f(\hat{\theta}) = \left(\hat{\Gamma}_{22}\right)^{-2}\sum_{i=1}^{1/h} y_{2i}^2\hat{\varepsilon}_i^2$ whereas $\nabla' f(\hat{\theta}))\left[\left(\sum_{i=1}^{1/h} x_i\right)\left(\sum_{i=1}^{1/h} x_i\right)'\right]\nabla f(\hat{\theta}) =$ 0. Thus $V^*_\beta = \left(\hat{\Gamma}_{22}\right)^{-2}\sum_{i=1}^{1/h} y_{2i}^2\hat{\varepsilon}_i^2$.

Part (b) is proven in the text. Part (c) follows from Theorem 4 of BN-S (2004) and the fact that $\hat{\beta}_{12} \to^P \beta_{12}$.

Appendix C

In this Appendix we prove the results appearing in Section 4. Appendix C.1 contains the proof of the asymptotic expansions of the cumulants of $T_{\beta,h}$ appearing in Proposition 4.1.(a). A number of auxiliary lemmas are also presented and proved. Appendix C.2 contains the proof of the asymptotic expansion of the bootstrap cumulants of of $T^*_{\beta,h}$ appearing in Proposition 4.1.(b) as well as some useful lemmas.

Note that the statistic of interest can be written as follows

$$T_{\beta,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\beta}_{12} - \beta_{12})}{\sqrt{\left(\sum_{i=1}^{1/h} y_{2i}^2\right)^{-2} h^{-1}\hat{g}_\beta}} = \frac{\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i}{\sqrt{h^{-1}\hat{g}_\beta}} = S_h\left(\frac{h^{-1}\hat{g}_\beta}{B_h}\right)^{-1/2},$$

where $\hat{g}_\beta$ and B_h are defined in the text, and $S_h = \frac{\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i}{\sqrt{B_h}}$.

Throughout this Appendix, we use the convention that $z_{1+1/h} = 0$ for any random variable z.

3.1 Asymptotic expansions of the cumulants of $T_{\beta,h}$

In this subsection, we first provide a set of lemmas that are useful to deriving the asymptotic expansions of the cumulants of $T_{\beta,h}$ through order $O\left(\sqrt{h}\right)$. Next, we prove these lemmas and at the end we prove

Proposition 4.1 a). We introduce the following notations.

$$
\begin{aligned}
u_i &= h^{-1}\left(y_{2i}^2\varepsilon_i^2 - E\left(y_{2i}^2\varepsilon_i^2\right)\right),\\
u_{i,i+1} &= h^{-1}\left(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1} - E\left(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}\right)\right),\\
A_{1h}^1 &= h^{-2}\sum_{i=1}^{1/h}\left(2\Gamma_{12,i}^3 - 18\beta_{12}\Gamma_{22,i}\Gamma_{12,i}^2 + 24\beta_{12}^2\Gamma_{22,i}^2\Gamma_{12,i} + 6\Gamma_{11,i}\Gamma_{22,i}\Gamma_{12,i} - 8\beta_{12}^3\Gamma_{22,i}^3 - 6\beta_{12}\Gamma_{11,i}\Gamma_{22,i}^2\right)\\
A_{1h}^2 &= h^{-2}\sum_{i=1}^{1/h}(-12\Gamma_{22,i}^3\beta_{12}^3 + 2\Gamma_{22,i}\Gamma_{22,i+1}^2\beta_{12}^3 + 2\Gamma_{22,i}^2\Gamma_{22,i+1}\beta_{12}^3 + 36\Gamma_{12,i}\Gamma_{22,i}^2\beta_{12}^2\\
&\quad -2\Gamma_{12,i+1}\Gamma_{22,i}^2\beta_{12}^2 - 2\Gamma_{12,i}\Gamma_{22,i+1}^2\beta_{12}^2 - 4\Gamma_{12,i}\Gamma_{22,i}\Gamma_{22,i+1}\beta_{12}^2 - 4\Gamma_{12,i+1}\Gamma_{22,i}\Gamma_{22,i+1}\beta_{12}^2\\
&\quad -8\Gamma_{11,i}\Gamma_{22,i}^2\beta_{12} - 28\Gamma_{12,i}^2\Gamma_{22,i}\beta_{12} + \Gamma_{12,i+1}^2\Gamma_{22,i}\beta_{12} + 4\Gamma_{12,i}\Gamma_{12,i+1}\Gamma_{22,i}\beta_{12} + \Gamma_{12,i}^2\Gamma_{22,i+1}\beta_{12}\\
&\quad +4\Gamma_{12,i}\Gamma_{12,i+1}\Gamma_{22,i+1}\beta_{12} + \Gamma_{11,i}\Gamma_{22,i}\Gamma_{22,i+1}\beta_{12} + \Gamma_{11,i+1}\Gamma_{22,i}\Gamma_{22,i+1}\beta_{12} + 4\Gamma_{12,i}^3 - \Gamma_{12,i}\Gamma_{12,i+1}^2\\
&\quad -\Gamma_{12,i}^2\Gamma_{12,i+1} + 8\Gamma_{11,i}\Gamma_{12,i}\Gamma_{22,i} - \Gamma_{11,i}\Gamma_{12,i+1}\Gamma_{22,i} - \Gamma_{11,i+1}\Gamma_{12,i}\Gamma_{22,i+1}).
\end{aligned}
$$

Similarly, let

$$
\begin{aligned}
A_{0h}^1 &= h^{-1}\sum_{i=1}^{1/h}E(y_{2i}^3\varepsilon_i), \quad A_{0h}^2 = h^{-1}\sum_{i=1}^{1/h}E(y_{2i}^2 y_{2,i+1}\varepsilon_{i+1}), \quad A_{0h}^3 = h^{-1}\sum_{i=1}^{1/h}E(y_{2,i+1}^2 y_{2i}\varepsilon_i),\\
A_{0h} &= \frac{1}{4}(2A_{0h}^1 - A_{0h}^2 - A_{0h}^3), \text{ and recall that } B_h = Var\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h}y_{2i}\varepsilon_i\right).
\end{aligned}
$$

Lemma C.3 *Let $k,l,k',l',k'',l'',m,n,m',n',m'',n'' = 1,2$ and let n_1, n_2, n_3, n_4, n_5 and n_6, be any non negative integers. Under Assumptions 1 and 2, and conditionally on the volatility path Σ,*

$$
\begin{aligned}
& h^{1-(n_1+n_2+n_3+n_4+n_5+n_6)}\sum_{i=1}^{1/h}\Gamma_{kl,i}^{n_1}\Gamma_{k'l',i}^{n_2}\Gamma_{k'l',i}^{n_3}\Gamma_{mn,i+1}^{n_4}\Gamma_{m'n',i+1}^{n_5}\Gamma_{m''n'',i+1}^{n_6}\\
\to\ & \int_0^1 \Sigma_{kl}^{n_1}(u)\Sigma_{k'l'}^{n_2}(u)\Sigma_{k'l'}^{n_3}(u)\Sigma_{mn}^{n_4}(u)\Sigma_{m'n'}^{n_5}(u)\Sigma_{m''n''}^{n_6}(u)du,
\end{aligned}
$$

as $h \to 0$.

Lemma C.4 *Under Assumptions 1 and 2, and conditionally on the volatility path, as $h \to 0$,*

- $A_{1h}^j \to A_1$, *for $j = 1,2$;*
- $B_h = h^{-1}\sum_{i=1}^{1/h}\left(\Gamma_{12,i} - 4\beta_{12}\Gamma_{22,i}\Gamma_{12,i} + 2\beta_{12}^2\Gamma_{22,i}^2 + \Gamma_{11,i}\Gamma_{22,i}\right) \to B;$
- $A_{0h}^1 = 3h^{-1}\sum_{i=1}^{1/h}(\Gamma_{12,i}\Gamma_{22,i} - \beta_{12}\Gamma_{22,i}^2) \to 3A_0;$
- $A_{0h}^2 = h^{-1}\sum_{i=1}^{1/h}(\Gamma_{12,i+1}\Gamma_{22,i} - \beta_{12}\Gamma_{22,i}\Gamma_{22,i+1}) \to A_0;$
- $A_{0h}^3 = h^{-1}\sum_{i=1}^{1/h}(\Gamma_{12,i}\Gamma_{22,i+1} - \beta_{12}\Gamma_{22,i}\Gamma_{22,i+1}) \to A_0.$

Lemma C.5 *Under Assumptions 1 and 2, and conditionally on the volatility path,*

- $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right) = 0,$
- $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^2 = hB_h,$
- $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^3 = h^2 A_{1h}^1,$
- $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^4 = 3h^2 B_h^2 + O(h),$
- $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i \sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = hA_{1h}^2,$
- $E\left(\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^2 \sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = O\left(h^2\right),$
- $E\left(\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^3 \sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = 3h^2 B_h A_{1h}^2 + O(h^3).$

Lemma C.6 *Under Assumptions 1 and 2, and conditionally on the volatility path,*

- $E\left(S_h\right) = 0,$
- $E\left(S_h^2\right) = 1,$
- $E\left(S_h^3\right) = \sqrt{h}\dfrac{A_{1h}^1}{B_h^{3/2}},$
- $E\left(S_h^4\right) = 3 + O(h),$
- $E\left(S_h\sqrt{h^{-1}}\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = \dfrac{A_{1h}^2}{\sqrt{B_h}},$
- $E\left(S_h^2\sqrt{h^{-1}}\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = O\left(\sqrt{h}\right),$
- $E\left(S_h^3\sqrt{h^{-1}}\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = 3\dfrac{A_{1h}^2}{\sqrt{B_h}} + O(h).$

Lemma C.7 *Under Assumptions 1, 2 and 3, and conditionally on the volatility path,*

$$h^{-1}\hat{g}_\beta = B_h\left(1+\frac{1}{B_h}\sum_{i=1}^{1/h}(u_i-u_{i,i+1})-\frac{4A_{0h}}{B_h\Gamma_{22}}\sum_{i=1}^{1/h}y_{2i}\varepsilon_i\right)+o_P(\sqrt{h}).$$

Proof of Lemma C.3. This result follows from the boundedness of $\Sigma_{kk}(u)$ and the Reimann integrability of $\Sigma_{kl}^{n}(u)$ for any $k,l=1,2$ and for any non negative integer n_i.

Proof of Lemma C.4. The convergence results follow from Lemma C.3. To derive the expressions of the moments, we use the fact that under our assumptions $y_1,\ldots,y_{1/h}$ are pairwise independent and $y_i\sim N(0,\Gamma_i)$ with $\Gamma_i=\int_{(i-1)h}^{ih}\Sigma(u)du$. Let C_i be the Cholesky decomposition of Γ_i. Note that $y_i\overset{d}{=}C_iu_i$: $u_i\sim iidN(0,I_2)$ where I_2 is the 2×2-identity matrix and '$\overset{d}{=}$' expresses the equivalence in distribution. Let $\Gamma_{kl,i}$ and $C_{kl,i}$ be the (k,l)-th element of Γ_i and C_i, respectively. We have that

$$C_i=\begin{pmatrix}\sqrt{\Gamma_{11,i}} & 0\\ \frac{\Gamma_{12,i}}{\sqrt{\Gamma_{11,i}}} & \sqrt{\Gamma_{22,i}-\frac{\Gamma_{12,i}^2}{\Gamma_{11,i}}}\end{pmatrix},$$

and $y_{1i}\overset{d}{=}C_{11,i}u_{1i}$ and $y_{2i}\overset{d}{=}C_{21,i}u_{1i}+C_{22,i}u_{2i}$. For the second result, let $z_i=y_{2i}\varepsilon_i-E(y_{2i}\varepsilon_i)$ and note that by definition, the $z_i's$ are i.i.d. with $Ez_i=0$. It follows that

$$B_h=Var\left(\sqrt{h^{-1}}\sum_{i=1}^{1/h}y_{2i}\varepsilon_i\right)=h^{-1}E\left(\sum_{i=1}^{1/h}(y_{2i}\varepsilon_i-E(y_{2i}\varepsilon_i))\right)^2=h^{-1}\sum_{i=1}^{1/h}E\left(z_i^2\right).$$

Now, $E\left(z_i^2\right)=E\left(y_{2i}^2\varepsilon_i^2\right)-\left(E\left(y_{2i}\varepsilon_i\right)\right)^2$. Since $\varepsilon_i=y_{1i}-\beta_{12}y_{2i}$, we get that

$$\begin{aligned}E\left(y_{2i}\varepsilon_i\right) &= E\left(y_{1i}y_{2i}\right)-\beta_{12}E\left(y_{2i}^2\right)=\Gamma_{12,i}-\beta_{12}\Gamma_{22,i},\\ E\left(y_{2i}^2\varepsilon_i^2\right) &= E\left(y_{2i}^2\left(y_{1i}-\beta_{12}y_{2i}\right)^2\right)=E\left(y_{2i}^2y_{1i}^2\right)-2\beta_{12}E\left(y_{1i}y_{2i}^3\right)+\beta_{12}^2E\left(y_{2i}^4\right).\end{aligned}$$

We now use the Cholesky decomposition to get that

$$\begin{aligned}E\left(y_{2i}^2y_{1i}^2\right) &= E\left(\left(C_{11,i}u_{1i}\right)^2\left(C_{21,i}u_{1i}+C_{22,i}u_{2i}\right)^2\right)=E\left(C_{11,i}^2u_{1i}^2\right)\left(C_{21,i}^2u_{1i}^2+2C_{21,i}C_{22,i}u_{1i}u_{2i}+C_{22,i}^2u_{2i}^2\right)\\ &= 3C_{11,i}^2C_{21,i}^2+C_{11,i}^2C_{22,i}^2=2\Gamma_{12,i}^2+\Gamma_{11,i}\Gamma_{22,i};\\ E\left(y_{1i}y_{2i}^3\right) &= E\left(\left(C_{11,i}u_{1i}\right)\left(C_{21,i}u_{1i}+C_{22,i}u_{2i}\right)^3\right)=3C_{11,i}C_{21,i}^3+3C_{11,i}C_{21,i}C_{22,i}^2=3\Gamma_{12,i}\Gamma_{22,i};\text{ and}\\ E\left(y_{2i}^4\right) &= E\left(\left(C_{21,i}u_{1i}+C_{22,i}u_{2i}\right)^4\right)=3C_{21,i}^4+6C_{21,i}^2C_{22,i}^2+3C_{22,i}^4=3\Gamma_{22,i},\text{ implying that}\\ E\left(y_{2i}^2\varepsilon_i^2\right) &= 2\Gamma_{12,i}^2+\Gamma_{11,i}\Gamma_{22,i}-6\beta_{12}\Gamma_{12,i}\Gamma_{22,i}+3\beta_{12}^2\Gamma_{22,i}^2.\end{aligned}$$

Thus,

$$E\left(y_{2i}^2\varepsilon_i^2\right)=2\Gamma_{12,i}^2+\Gamma_{11,i}\Gamma_{22,i}-6\beta_{12}\Gamma_{12,i}\Gamma_{22,i}+3\beta_{12}^2\Gamma_{22,i}^2,$$

and

$$\begin{aligned}E\left(z_i^2\right) &= 2\Gamma_{12,i}^2+\Gamma_{11,i}\Gamma_{22,i}-6\beta_{12}\Gamma_{12,i}\Gamma_{22,i}+3\beta_{12}^2\Gamma_{22,i}^2-\left(\Gamma_{12,i}-\beta_{12}\Gamma_{22,i}\right)^2\\ &= \Gamma_{12,i}^2+\Gamma_{11,i}\Gamma_{22,i}-4\beta_{12}\Gamma_{12,i}\Gamma_{22,i}+2\beta_{12}^2\Gamma_{22,i}^2,\end{aligned}$$

which implies $B_h=h^{-1}\sum_{i=1}^{1/h}\left(\Gamma_{12,i}^2+\Gamma_{11,i}\Gamma_{22,i}-4\beta_{12}\Gamma_{12,i}\Gamma_{22,i}+2\beta_{12}^2\Gamma_{22,i}^2\right)$, proving the second result. The proofs of the remaining results is similar and we omit the details.

Proof of Lemma C.5. The first result follows by definition of β_{12} whereas the second result follows by the definition of B_h. For the remaining results, write $z_i = y_{2i}\varepsilon_i - E(y_{2i}\varepsilon_i)$ and note that by definition, the $z_i's$ are i.i.d. with $Ez_i = 0$. Note also that $\sum_{i=1}^{1/h} z_i = \sum_{i=1}^{1/h} y_{2i}\varepsilon_i$ since by $\sum_{i=1}^{1/h} E\left(y_{2i}\varepsilon_i\right) = 0$. For the third result, note that

$$E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^3 = E\left(\sum_{i=1}^{1/h} z_i\right)^3 = \sum_{i,j,k=1}^{1/h} E\left(z_i z_j z_k\right) = \sum_{i=1}^{1/h} E\left(z_i^3\right).$$

We now compute $E\left(z_i^3\right)$ using the Cholesky decomposition as in the proof of Lemma C.4 to show that $\sum_{i=1}^{1/h} E\left(z_i^3\right) = h^2 A_{1h}^1$, with A_{1h}^1 as defined above. For the fourth result, note that $E\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)^4 = \sum_{i=1}^{1/h} E\left(z_i^4\right) + 3\sum_{i\neq j} E\left(z_i^2\right) E\left(z_j^2\right) = 3\left(\sum_{i=1}^{1/h} E\left(z_i^2\right)\right)^2 + O(h^3)$ and use the definition of B_h to prove the result. For the fifth result, note that

$$E\left(\left(\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right)\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) = \sum_{i=1}^{1/h} E(z_i u_i) - \sum_{i=1}^{1/h} E(z_i u_{i,i+1}) - \sum_{i=1}^{1/h} E(z_{i+1} u_{i,i+1}).$$

Useing the definitions of u_i and $u_{i,i+1}$, the result follows from simple but tedious algebra using the Cholesky decomposition. The remaining results follow similarly and therefore we omit the details.
Proof of Lemma C.6. The proof follows straightforwardly by using Lemma C.5.
Proof of Lemma C.7. Using the definition of $\hat{g}_\beta$ in the text, we can write

$$\begin{aligned} h^{-1}\hat{g}_\beta &= h^{-1}\sum_{i=1}^{1/h}\left(y_{2i}^2\varepsilon_i^2 + (\hat{\beta}_{12} - \beta_{12})^2 y_{2i}^4 - 2(\hat{\beta}_{12} - \beta_{12}) y_{2i}^3\varepsilon_i\right) \\ &\quad - h^{-1}\sum_{i=1}^{1/h}\left(y_{2i}y_{2,i+1}\varepsilon_i\varepsilon_{i+1} + (\hat{\beta}_{12} - \beta_{12})^2 y_{2i}^2 y_{2,i+1}^2 - (\hat{\beta}_{12} - \beta_{12})(y_{2i}^2 y_{2,i+1}\varepsilon_{i+1} + y_{2,i+1}^2 y_{2i}\varepsilon_i)\right). \end{aligned}$$

Adding and subtracting appropriately, it follows that

$$\begin{aligned} h^{-1}\hat{g}_\beta &= h^{-1}\sum_{i=1}^{1/h} E(y_{2i}\varepsilon_i)^2 - h^{-1}\sum_{i=1}^{1/h} E\left(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}\right) + \left(h^{-1}\sum_{i=1}^{1/h}((y_{2i}\varepsilon_i)^2 - E(y_{2i}\varepsilon_i)^2)\right) \\ &\quad - \left(h^{-1}\sum_{i=1}^{1/h}(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1} - E\left(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}\right))\right) - (\hat{\beta}_{12} - \beta_{12})h^{-1}\sum_{i=1}^{1/h} E(2y_{2i}^3\varepsilon_i) \\ &\quad + (\hat{\beta}_{12} - \beta_{12})h^{-1}\sum_{i=1}^{1/h} E(y_{2i}^2 y_{2,i+1}\varepsilon_{i+1}) + (\hat{\beta}_{12} - \beta_{12})h^{-1}\sum_{i=1}^{1/h} E(y_{2,i+1}^2 y_{2i}\varepsilon_i) + O_P(h), \\ &= B_h + h^{-1}\sum_{i=1}^{1/h}(Ey_{2i}\varepsilon_i)^2 - h^{-1}\sum_{i=1}^{1/h} E\left(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}\right) + \sum_{i=1}^{1/h}(u_i - u_{i,i+1}) - (\hat{\beta}_{12} - \beta_{12})2A_{0h}^1 \\ &\quad + (\hat{\beta}_{12} - \beta_{12})A_{0h}^2 + (\hat{\beta}_{12} - \beta_{12})A_{0h}^3 + O_P\left(h\right) \\ &= B_h + \sum_{i=1}^{1/h}(u_i - u_{i,i+1}) - \frac{2A_{0h}^1}{\Gamma_{22}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i + \frac{A_{0h}^2}{\Gamma_{22}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i + \frac{A_{0h}^3}{\Gamma_{22}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i + o_P\left(\sqrt{h}\right), \end{aligned}$$

where the remainder term is of order $o_P(\sqrt{h})$ given that $\hat{\beta}_{12} - \beta_{12} = O_P(\sqrt{h})$, $h^{-1}\sum_{i=1}^{1/h} y_{2i}^2 y_{2,i+1}^2 = O_P(1)$, and given that $h^{-1}\sum_{i=1}^{1/h}\left(y_{2i}^3\varepsilon_i - E\left(y_{2i}^3\varepsilon_i\right)\right) = O_P(\sqrt{h})$ and $h^{-1}\sum_{i=1}^{1/h}\left(y_{2,i+1}^2 y_{2i}\varepsilon_i - E\left(y_{2,i+1}^2 y_{2i}\varepsilon_i\right)\right) =$

$O_P(\sqrt{h})$ by a verifying a CLT condition. Note also that the last equality uses the fact that $\hat{\beta}_{12} - \beta_{12} = \frac{\sum_{i=1}^{1/h} y_{2i}\varepsilon_i}{\Gamma_{22}} + O_P(h)$. By Lemma C.1, $h^{-1}\sum_{i=1}^{1/h}(Ey_{2i}\varepsilon_i)^2$ and $h^{-1}\sum_{i=1}^{1/h}E(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1})$ have the same probability limit and by Assumption 3,
$h^{-1}\sum_{i=1}^{1/h}(Ey_{2i}\varepsilon_i)^2 - p\lim h^{-1}\sum_{i=1}^{1/h}(Ey_{2i}\varepsilon_i)^2 = o_P\left(\sqrt{h}\right)$ and
$h^{-1}\sum_{i=1}^{1/h}E(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}) - p\lim h^{-1}\sum_{i=1}^{1/h}E(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}) = o_P\left(\sqrt{h}\right)$.
Therefore, $h^{-1}\sum_{i=1}^{1/h}(Ey_{2i}\varepsilon_i)^2 - h^{-1}\sum_{i=1}^{1/h}E(y_{2i}\varepsilon_i y_{2,i+1}\varepsilon_{i+1}) = o_P\left(\sqrt{h}\right)$.

Proof of Proposition 4.1 (a). Given Lemma C.7, we can write

$$T_{\beta,h} = S_h\left(1 + \frac{1}{B_h}\sum_{i=1}^{1/h}(u_i - u_{i,i+1}) - \frac{4A_{0h}}{B_h\Gamma_{22}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i + o_P(\sqrt{h})\right)^{-1/2}.$$

The first and third cumulants of $T_{\beta,h}$ are given by (see e.g., Hall, 1992, p. 42) $\kappa_1(T_{\beta,h}) = E(T_{\beta,h})$ and

$$\kappa_3(T_{\beta,h}) = E(T_{\beta,h}^3) - 3E(T_{\beta,h}^2)E(T_{\beta,h}) + 2[E(T_{\beta,h})]^3.$$

Our goal is to identify the terms of order up to $O(\sqrt{h})$ of the asymptotic expansions of these two cumulants. We will first provide asymptotic expansions through order $O(\sqrt{h})$ for the first three moments of $T_{\beta,h}$. Note that for a given fixed value of k, a first-order Taylor expansion of $f(x) = (1+x)^{-k/2}$ around 0 yields $f(x) = 1 - \frac{k}{2}x + O(x^2)$. Provided that $\sum_{i=1}^{1/h}(u_i - u_{i,i+1}) = O_P(\sqrt{h})$, we have for any fixed integer k,

$$T_{\beta,h}^k = S_h^k\left(1 - \sqrt{h}\frac{k}{2}\frac{\sqrt{h^{-1}}}{B_h}\sum_{i=1}^{1/h}(u_i - u_{i,i+1}) + \sqrt{h}k\frac{2A_{0h}}{B_h\Gamma_{22}}\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}\varepsilon_i\right) + o(\sqrt{h}) = \tilde{T}_h^k + o(\sqrt{h}).$$

For $k = 1, 2, 3$, the moments of $\tilde{T}_h^k$ are given by

$$\begin{aligned}
E(\tilde{T}_h) &= E(S_h) - \sqrt{h}\frac{1}{2}\frac{1}{B_h}E\left(S_h\sqrt{h^{-1}}\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) + \sqrt{h}\frac{2A_{0h}}{\sqrt{B_h}\Gamma_{22}}E(S_h^2),\\
E(\tilde{T}_h^2) &= E(S_h^2) - \sqrt{h}\frac{1}{B_h}E\left(S_h^2\sqrt{h^{-1}}\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) + \sqrt{h}\frac{4A_{0h}}{\sqrt{B_h}\Gamma_{22}}E(S_h^3),\\
E(\tilde{T}_h^3) &= E(S_h^3) - \frac{3}{2}\sqrt{h}\frac{1}{B_h}E\left(S_h^3\sqrt{h^{-1}}\sum_{i=1}^{1/h}(u_i - u_{i,i+1})\right) + \sqrt{h}\frac{6A_{0h}}{\sqrt{B_h}\Gamma_{22}}E(S_h^4).
\end{aligned}$$

Given Lemma C.6,

$$\begin{aligned}
E(\tilde{T}_h) &= -\sqrt{h}\frac{1}{2B_h}\frac{A_{1h}^2}{\sqrt{B_h}} + \sqrt{h}\frac{2A_{0h}}{\sqrt{B_h}\Gamma_{22}},\\
E(\tilde{T}_h^2) &= 1 + O(h),\\
E(\tilde{T}_h^3) &= \sqrt{h}\frac{A_{1h}^1}{B_h^{3/2}} - \frac{3}{2B_h}\sqrt{h}\,3\times\frac{A_{1h}^2}{\sqrt{B_h}} + \sqrt{h}\frac{18A_{0h}}{\sqrt{B_h}\Gamma_{22}} + O(h).
\end{aligned}$$

Thus

$$\kappa_1(T_{\beta,h}) = \sqrt{h}\left(\underbrace{-\frac{1}{2B_h}\frac{A_{1h}^2}{\sqrt{B_h}} + \frac{2A_{0h}}{\sqrt{B_h}\Gamma_{22}}}_{\equiv\kappa_{1,h}}\right) + o(\sqrt{h}),$$

and

$$\kappa_3\left(T_{\beta,h}\right) = \sqrt{h}\left(\underbrace{\frac{A_{1h}^1}{B_h^{3/2}} - \frac{3}{B_h}\frac{A_{1h}^2}{\sqrt{B_h}} + \frac{12A_{0h}}{\sqrt{B_h}\Gamma_{22}}}_{\equiv\kappa_{3,h}}\right) + o(\sqrt{h}).$$

By Lemma C.4, we can now show that $\lim_{h\to 0}\kappa_{1,h} = -\frac{1}{2}\frac{A_1}{B^{3/2}} + \frac{1}{2}\frac{4A_0}{\sqrt{B}\Gamma_{22}} \equiv \frac{1}{2}\left(H_1 - H_2\right)$ and $\lim_{h\to 0}\kappa_{3,h} = -2\frac{A_1}{B^{3/2}} + 3\frac{4A_0}{\sqrt{B}\Gamma_{22}} \equiv 3H_1 - 2H_2$, where A_0, A_1, B, H_1 and H_2 are as defined in the text.

3.2 Asymptotic expansions of the bootstrap cumulants of $T^*_{\beta,h}$

In this section we provide the asymptotic expansions through $O_P\left(\sqrt{h}\right)$ of the first and third cumulants of the bootstrap statistic $T^*_{\beta,h}$. Let $\varepsilon_i^* = y_{1i}^* - \hat{\beta}_{12}y_{2i}^* = \hat{\varepsilon}_{I_i}$, with I_i a uniform draw from $\{1,\ldots,n\}$, and let $\hat{\varepsilon}_i^* = y_{1i}^* - \hat{\beta}_{12}^* y_{2i}^*$ be the bootstrap OLS residual. Note that

$$T^*_{\beta,h} \equiv \frac{\sqrt{h^{-1}}(\hat{\beta}^*_{12} - \hat{\beta}_{12})}{\sqrt{\left(\sum_{i=1}^{1/h} y_{2i}^{*2}\right)^{-2}\hat{B}^*_{1h}}} = \frac{\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*}{\sqrt{\hat{B}^*_{1h}}} = S_h^*\left(\frac{\hat{B}^*_{1h}}{\hat{B}_{1h}}\right)^{-1/2},$$

where $S_h^* = \frac{\sqrt{h^{-1}}\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*}{\sqrt{\hat{B}_{1h}}}$, where $\hat{B}_{1h} = h^{-1}\sum_{i=1}^{1/h} y_{2i}^2\hat{\varepsilon}_i^2$, $\hat{B}^*_{1h} = h^{-1}\sum_{i=1}^{1/h} y_{2i}^{*2}\hat{\varepsilon}_i^{*2}$, $\hat{\varepsilon}_i^* = y_{1i}^* - \hat{\beta}_{12}^* y_{2i}^*$. Let

$$\hat{A}_{0h} = h^{-1}\sum_{i=1}^{1/h} y_{2i}^3\hat{\varepsilon}_i, \qquad \hat{A}_{1h} = h^{-2}\sum_{i=1}^{1/h}(y_{2i}\hat{\varepsilon}_i)^3,$$

and let $\tilde{B}^*_{1h} = h^{-1}\sum_{i=1}^{1/h}(y_{2i}^*\varepsilon_i^*)^2$.

Proposition C.1 *Let $y_i^* \sim$ i.i.d. from $\{y_i : i = 1,\ldots,1/h\}$. Under Assumptions 1 and 2 and conditionally on Σ, as $h \to 0$,*

$$\kappa_1^*(T^*_{\beta,h}) = \sqrt{h}\underbrace{\left(-\frac{\hat{A}_{1h}}{2\hat{B}_{1h}^{3/2}} + \frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}}\hat{\Gamma}_{22}}\right)}_{\equiv\kappa^*_{1,h}} \equiv \sqrt{h}\kappa^*_{1,h},$$

$$\kappa_3^*(T^*_{\beta,h}) = \sqrt{h}\underbrace{\left(-\frac{2\hat{A}_{1h}}{\hat{B}_{1h}^{3/2}} + \frac{6\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}}\hat{\Gamma}_{22}}\right)}_{\equiv\kappa^*_{3,h}} + O_P(h) \equiv \sqrt{h}\kappa^*_{3,h} + O_P(h).$$

Proposition C.1 is used to prove Proposition 4.1 (b). The proofs of these two propositions are given after the following set of auxiliary lemmas, whose proofs follow the proofs of the propositions.

Lemma C.8 *Let $y_i^* \sim$ i.i.d. from $\{y_i : i = 1,\ldots,1/h\}$. Under Assumptions 1 and 2, and conditionally on the volatility path,*

- $E^*\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right) = 0,$

- $E^*\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right)^2 = \sum_{i=1}^{1/h}(y_{2i}\hat{\varepsilon}_i)^2 \equiv h\hat{B}_{1h},$
- $E^*\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right)^3 = \sum_{i=1}^{1/h}(y_{2i}\hat{\varepsilon}_i)^3 \equiv h^2\hat{A}_{1h},$
- $E^*\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right)^4 = 3\left(\sum_{i=1}^{1/h}(y_{2i}\hat{\varepsilon}_i)^2\right)^2 + O_P(h^3) \equiv 3h^2\left(\hat{B}_{1h}\right)^2 + O_P(h^3),$
- $E^*\left(\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right)(\tilde{B}_{1h}^* - \hat{B}_{1h})\right) = h^{-1}\sum_{i=1}^{1/h}(y_{2i}\hat{\varepsilon}_i)^3 = h\hat{A}_{1h},$
- $E^*\left(\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right)^2(\tilde{B}_{1h}^* - \hat{B}_{1h})\right) = O_P\left(h^2\right),$
- $E^*\left(\left(\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right)^3(\tilde{B}_{1h}^* - \hat{B}_{1h})\right) = 3\hat{B}_{1h}\left(\sum_{i=1}^{1/h}(y_{2i}\hat{\varepsilon}_i)^3\right) + O_P(h^3) = 3h^2\hat{B}_{1h}\hat{A}_{1h} + O_P\left(h^3\right).$

Lemma C.9 *Under Assumptions 1 and 2, and conditionally on the volatility path,*

- $E^*\left(S_h^*\right) = 0,$
- $E^*\left(S_h^{*2}\right) = 1,$
- $E^*\left(S_h^{*3}\right) = \sqrt{h}\frac{\hat{A}_{1h}}{\hat{B}_{1h}^{3/2}},$
- $E^*\left(S_h^{*4}\right) = 3 + O_P(h),$
- $E^*\left(S_h^*\sqrt{h^{-1}}(\tilde{B}_{1h}^* - \hat{B}_{1h})\right) = \frac{\hat{A}_{1h}}{\sqrt{\hat{B}_{1h}}},$
- $E^*\left(S_h^{*2}\sqrt{h^{-1}}\left(\tilde{B}_{1h}^* - \hat{B}_{1h}\right)\right) = O_P\left(\sqrt{h}\right),$
- $E^*\left(S_h^{*3}\sqrt{h^{-1}}\left(\tilde{B}_{1h}^* - \hat{B}_{1h}\right)\right) = 3\frac{\hat{A}_{1h}}{\sqrt{\hat{B}_{1h}}} + O_P(h).$

Lemma C.10 *Under Assumptions 1 and 2, and conditionally on the volatility path,*

$$\hat{B}_{1h}^* = \hat{B}_{1h}\left(1 + \frac{\tilde{B}_{1h}^* - \hat{B}_{1h}}{\hat{B}_{1h}} - \frac{2\hat{A}_{0h}}{\hat{B}_{1h}\hat{\Gamma}_{22}}\sum_{i=1}^{1/h} y_{2i}^*\varepsilon_i^*\right) + O_{P^*}(h),$$

in probability.

Proof of Proposition C.1. By Lemma C.10,

$$T^*_{\beta,h} = S^*_{\beta,h}\left(1+\frac{\tilde{B}^*_{1h}-\hat{B}_{1h}}{\hat{B}_{1h}} - \frac{2\hat{A}_{0h}}{\hat{B}_{1h}\hat{\Gamma}_{22}}\sum_{i=1}^{1/h} y^*_{2i}\varepsilon^*_i + O_{P^*}(h)\right)^{-1/2}.$$

Following the proof of Proposition 4.1.(a), for any fixed integer k, we have that

$$T^{*k}_{\beta,h} = S^{*k}_h\left(1-\sqrt{h}\frac{k}{2}\frac{\sqrt{h^{-1}}}{\hat{B}_{1h}}\left(\tilde{B}^*_{1h}-\hat{B}_{1h}\right) + \sqrt{h}k\frac{\hat{A}_{0h}}{\hat{B}_{1h}\hat{\Gamma}_{22}}\sqrt{h^{-1}}\sum_{i=1}^{1/h} y^*_{2i}\varepsilon^*_i\right) + O_P(h) \equiv \tilde{T}^{*k}_{\beta,h} + O_P(h).$$

For $k = 1, 2, 3$, the moments of $\tilde{T}^{*k}_h$ are given by

$$\begin{aligned}
E^*(\tilde{T}^*_{\beta,h}) &= 0 - \sqrt{h}\frac{1}{2}\frac{1}{\hat{B}_{1h}}E^*\left(S^*_h\sqrt{h^{-1}}\left(\tilde{B}^*_{1h}-\hat{B}_{1h}\right)\right) + \sqrt{h}\frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}E^*(S^{*^2}_h),\\
E^*(\tilde{T}^{*^2}_{\beta,h}) &= 1 - \sqrt{h}\frac{1}{\hat{B}_{1h}}E^*\left(S^{*^2}_h\sqrt{h^{-1}}\left(\tilde{B}^*_{1h}-\hat{B}_{1h}\right)\right) + \sqrt{h}\frac{2\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}E^*(S^{*^3}_h),\\
E^*(\tilde{T}^{*^3}_{\beta,h}) &= E(S^{*^3}_h) - \sqrt{h}\frac{3}{2}\frac{1}{\hat{B}_{1h}}E^*\left(S^{*^3}_h\sqrt{h^{-1}}\left(\tilde{B}^*_{1h}-\hat{B}_{1h}\right)\right) + \sqrt{h}\frac{3\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}E^*(S^{*^4}_h).
\end{aligned}$$

Lemma C.9 implies that

$$\begin{aligned}
E^*(\tilde{T}^*_{\beta,h}) &= -\sqrt{h}\frac{1}{2}\frac{1}{\hat{B}_{1h}}\frac{\hat{A}_{1h}}{\sqrt{\hat{B}_{1h}}} + \sqrt{h}\frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}} = \sqrt{h}\left(-\frac{1}{2}\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} + \frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right)\\
E^*\left(\tilde{T}^*_{\beta,h}\right) &= 1 + O_P(h),\\
E^*(\tilde{T}^{*^3}_{\beta,h}) &= \sqrt{h}\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} - \sqrt{h}\frac{9}{2}\frac{1}{\hat{B}_{1h}}\frac{\hat{A}_{1h}}{\sqrt{\hat{B}_{1h}}} + \sqrt{h}\frac{9\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}} = \sqrt{h}\left(-\frac{7}{2}\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} + 9\frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right).
\end{aligned}$$

Thus

$$\kappa^*_1(T^*_{\beta,h}) = E^*(\tilde{T}^*_{\beta,h}) = \sqrt{h}\left(-\frac{1}{2}\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} + \frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right) \equiv \sqrt{h}\kappa^*_{1,h},$$

and

$$\begin{aligned}
\kappa^*_3(T^*_{\beta,h}) &= E^*(\tilde{T}^{*^3}_{\beta,h}) - 3E^*(\tilde{T}^{*^2}_{\beta,h})E^*(\tilde{T}^*_{\beta,h}) + 2[E^*(\tilde{T}^*_{\beta,h})]^3\\
&= \sqrt{h}\left(-\frac{7}{2}\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} + 9\frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right) - 3\sqrt{h}\left(-\frac{1}{2}\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} + \frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right) + O_P(h)\\
&= \sqrt{h}\left(-2\frac{\hat{A}_{1h}}{\hat{B}^{3/2}_{1h}} + 6\frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right) + O_P(h).
\end{aligned}$$

Proof of Proposition 4.1 b). By Theorem 4 by BN-S (2004), and because $\hat{\beta}_{12} \xrightarrow{P} \beta_{12}$, we have that $\hat{B}_{1h} \xrightarrow{P} B^*$, and $\hat{A}_{0h} \xrightarrow{P} 3\int_0^1\left(\Sigma_{12}(u)\Sigma_{12}(u) - \beta_{12}\Sigma^2_{22}(u)\right)du = 3A_0$. Similarly, we can show that

$$\hat{A}_{1h} = h^{-2}\sum_{i=1}^{1/h}(\varepsilon_i y_{2i})^3 + o_P(1) = h^{-2}\sum_{i=1}^{1/h}E\left((\varepsilon_i y_{2i})^3\right) + R_h + o_P(1),$$

where $R_h = h^{-2}\sum_{i=1}^{1/h}(\varepsilon_i y_{2i})^3 - E\left((\varepsilon_i y_{2i})^3\right)$. $E(R_h) = 0$ and by straightforward calculations, $Var\left(h^{-2}\sum_{i=1}^{1/h}(\varepsilon_i y_{2i})^3\right) = O(h) = o(1)$, which implies that $R_h = o_P(1)$. By tedious but simple algebra we can verify that

$$h^{-2}\sum_{i=1}^{1/h} E\left((\varepsilon_i y_{2i})^3\right) = h^{-2}\sum_{i=1}^{1/h}\left(\begin{array}{c} 6\Gamma_{12,i}^3 + 9\Gamma_{11,i}\Gamma_{12,i}\Gamma_{22,i} - 36\beta_{12}\Gamma_{12,i}^2\Gamma_{22,i} \\ -9\beta_{12}\Gamma_{11,i}\Gamma_{22,i}^2 + 45\beta_{12}^2\Gamma_{12,i}\Gamma_{22,i}^2 - 15\beta_{12}^3\Gamma_{22,i}^3 \end{array}\right).$$

By Lemma C.3, this last expression converges to

$$\int_0^1 (6\Sigma_{12}^3 + 9\Sigma_{11}\Sigma_{12}\Sigma_{22} - 36\beta_{12}\Sigma_{12}^2\Sigma_{22} - 9\beta_{12}\Sigma_{11}\Sigma_{22}^2 + 45\beta_{12}^2\Sigma_{12}\Sigma_{22}^2 - 15\beta_{12}^3\Sigma_{22}^3)du = \frac{3}{2}A_1^*,$$

proving that $\hat{A}_{1h} \to^P \frac{3}{2}A_1^*$. Thus, using Proposition C.1, we get that

$$p\lim \kappa_{1,h}^* = p\lim\left(-\frac{1}{2}\frac{\hat{A}_{1h}}{\hat{B}_{1h}^{3/2}} + \frac{\hat{A}_{0h}}{\sqrt{\hat{B}_{1h}\hat{\Gamma}_{22}}}\right) = -\frac{1}{2}\frac{\frac{3}{2}A_1^*}{B^{*3/2}} + \frac{3A_0}{\sqrt{B^*}\Gamma_{22}} = \frac{3}{4}\left(\frac{4A_0}{\sqrt{B^*}\Gamma_{22}} - \frac{A_1^*}{B^{*3/2}}\right) \equiv \frac{3}{4}(H_1^* - H_2^*).$$

Similarly,

$$p\lim \kappa_{3,h}^* = \left(-2\frac{\frac{3}{2}A_1^*}{B^{*3/2}} + 6\frac{3A_0}{\sqrt{B^*}\Gamma_{22}}\right) = \left(\frac{3*3}{2}H_1^* - 3H_2^*\right) = 3\left(\frac{3}{2}H_1^* - H_2^*\right).$$

Proof of Lemma C.8. The first result follows by noting that $E^*\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right) = h^{-1}h\sum_{i=1}^{1/h}\hat{\varepsilon}_i y_{2i} = 0$ by the first order OLS equations. Note in particular that $E^*(\varepsilon_i^* y_{2i}^*) = 0$. The second result follows by using the independence between $y_{2i}^*\varepsilon_i^*$ and $y_{2j}^*\varepsilon_j^*$ for $i \neq j$ and noting that

$$E^*\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right)^2 = \sum_{i,j=1}^{1/h} E^*\left(\varepsilon_i^* y_{2i}^*\varepsilon_j^* y_{2j}^*\right) = \sum_{i=1}^{1/h} E^*\left(\varepsilon_i^* y_{2i}^*\right)^2 = \sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^2 \equiv h\hat{B}_{1h}.$$

The third result follows similarly. In particular,

$$E^*\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right)^3 = \sum_{i=1}^{1/h} E^*(\varepsilon_i^* y_{2i}^*)^3 = \sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^3 \equiv h^2\hat{A}_{1h}.$$

Similarly,

$$\begin{aligned} E^*\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right)^4 &= E^*\left(\sum_{i=1}^{1/h}(\varepsilon_i^* y_{2i}^*)^4 + 3\sum_{i\neq j:i,j=1}^{1/h}(\varepsilon_i^* y_{2i}^*)^2(\varepsilon_j^* y_{2j}^*)^2\right) \\ &= \sum_{i=1}^{1/h} E^*(\varepsilon_i^* y_{2i}^*)^4 + 3\sum_{i\neq j:i,j=1}^{1/h} E^*(\varepsilon_i^* y_{2i}^*)^2 E^*(\varepsilon_j^* y_{2j}^*)^2 \\ &= \sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^4 + 3\left(\sum_{i=1}^{1/h} E^*(\varepsilon_i^* y_{2i}^*)^2\right)^2 - 3\sum_{i=1}^{1/h}\left(E^*(\varepsilon_i^* y_{2i}^*)^2\right)^2 \\ &= 3\left(\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^2\right)^2 + O_P(h) \equiv 3h^2\hat{B}_{1h} + O_P(h), \end{aligned}$$

where we have used Lemma B.1 to obtain the order of the remainder term. For the remaining results, note that $E^*\left(\tilde{B}_{1h}^*\right) = h^{-1}\sum_{i=1}^{1/h} E^*(\varepsilon_i^* y_{2i}^*)^2 = h^{-1}\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^2 = \hat{B}_{1h}$, which allows us to write $\tilde{B}_{1h}^* - \hat{B}_{1h} = \sum_{i=1}^{1/h}\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2 - E^*\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2\right)\right)$. It follows that

$$\begin{aligned} E^*\left(\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right)\left(\tilde{B}_{1h}^* - \hat{B}_{1h}\right)\right) &= \sum_{i=1}^{1/h} E^*\left(\varepsilon_i^* y_{2i}^*\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2 - h^{-1}E^*(\varepsilon_i^* y_{2i}^*)^2\right)\right) \\ &= h^{-1}\sum_{i=1}^{1/h} E^*\left((\varepsilon_i^* y_{2i}^*)^3\right) = h^{-1}\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^3 \equiv h\hat{A}_{1h}. \end{aligned}$$

Next,

$$\begin{aligned} E^*\left(\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right)^2\left(\tilde{B}_{1h}^* - \hat{B}_{1h}\right)\right) &= \sum_{i=1}^{1/h} E^*\left((\varepsilon_i^* y_{2i}^*)^2\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2 - h^{-1}E^*(\varepsilon_i^* y_{2i}^*)^2\right)\right) \\ &= h^{-1}\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^4 - \left(\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^2\right)^2 = O_P(h^2). \end{aligned}$$

Finally,

$$\begin{aligned} E^*\left(\left(\sum_{i=1}^{1/h}\varepsilon_i^* y_{2i}^*\right)^3\left(\tilde{B}_{1h}^* - \hat{B}_{1h}\right)\right) &= \sum_{i=1}^{1/h} E^*\left[(\varepsilon_i^* y_{2i}^*)^3\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2 - h^{-1}E^*(\varepsilon_i^* y_{2i}^*)^2\right)\right] \\ &\quad +3\sum_{i\neq j:i,j=1}^{1/h} E^*\left[(\varepsilon_i^* y_{2i}^*)^2(\varepsilon_j^* y_{2j}^*)\left(h^{-1}(\varepsilon_j^* y_{2j}^*)^2 - h^{-1}E^*(\varepsilon_j^* y_{2j}^*)^2\right)\right] \\ &= \sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^3\left(h^{-1}(\hat{\varepsilon}_i y_{2i})^2 - \sum_{k=1}^{1/h}(\hat{\varepsilon}_k y_{2k})^2\right) \\ &\quad +3\left(\sum_{i=1}^{1/h} E^*(\varepsilon_i^* y_{2i}^*)^2\right) \\ &\quad \times\left(\sum_{i=1}^{1/h} E^*\left((\varepsilon_i^* y_{2i}^*)\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2 - \sum_{k=1}^{1/h}(\hat{\varepsilon}_k y_{2k})^2\right)\right)\right) \\ &\quad -3\sum_{i=1}^{1/h}\left(E^*\left((\varepsilon_i^* y_{2i}^*)^3\left(h^{-1}(\varepsilon_i^* y_{2i}^*)^2 - \sum_{k=1}^{1/h}(\hat{\varepsilon}_k y_{2k})^2\right)\right)\right) \\ &= O_P(h^3) + 3h^{-1}\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^2\sum_{i=1}^{1/h}(\hat{\varepsilon}_i y_{2i})^3 + O_P(h^3) \\ &= 3\hat{B}_{1h}h^2\hat{A}_{1h} + O_P\left(h^3\right). \end{aligned}$$

Proof of Lemma C.9. The proof follows easily from Lemma C.8.

Proof of Lemma C.10. Since $\hat{\beta}^*_{12} - \hat{\beta}_{12} = O_{P^*}\left(\sqrt{h}\right)$, in probability, it follows that

$$
\begin{aligned}
\hat{B}^*_{1h} &= h^{-1}\sum_{i=1}^{1/h} y^{*^2}_{2i}\hat{\varepsilon}^{*^2}_i = h^{-1}\sum_{i=1}^{1/h} y^{*^2}_{2i}(y^*_{1i} - \hat{\beta}^*_{12}y^*_{2i})^2 \\
&= h^{-1}\sum_{i=1}^{1/h} y^{*^2}_{2i}(y^*_{1i} - \hat{\beta}_{12}y^*_{2i})^2 - 2(\hat{\beta}^*_{12} - \hat{\beta}_{12})h^{-1}\sum_{i=1}^{1/h} y^{*^3}_{2i}(y^*_{1i} - \hat{\beta}_{12}y^*_{2i}) + O_{P^*}(h) \\
&= h^{-1}\sum_{i=1}^{1/h} y^{*^2}_{2i}(y^*_{1i} - \hat{\beta}_{12}y^*_{2i})^2 - 2\frac{\sum_{i=1}^{1/h} y^*_{2i}\varepsilon^*_i}{\hat{\Gamma}^*_{22}}h^{-1}\sum_{i=1}^{1/h} y^{*^3}_{2i}(y^*_{1i} - \hat{\beta}_{12}y^*_{2i}) + O_{P^*}(h).
\end{aligned}
$$

By a CLT for i.i.d random variables, we can prove that

$$
\begin{aligned}
h^{-1}\sum_{i=1}^{1/h} y^{*^2}_{2i}(y^*_{1i} - \hat{\beta}_{12}y^*_{2i})^2 - h^{-1}\sum_{i=1}^{1/h} y^2_{2i}(y_{1i} - \hat{\beta}_{12}y_{2i})^2 &= O_{P^*}(\sqrt{h}), \\
h^{-1}\sum_{i=1}^{1/h} y^{*^3}_{2i}(y^*_{1i} - \hat{\beta}_{12}y^*_{2i}) - h^{-1}\sum_{i=1}^{1/h} y^3_{2i}(y_{1i} - \hat{\beta}_{12}y_{2i}) &= O_{P^*}(\sqrt{h}), \quad \text{and} \\
\hat{\Gamma}^*_{22} - \hat{\Gamma}_{22} &= O_{P^*}(\sqrt{h}),
\end{aligned}
$$

in probability. Adding and subtracting appropriately gives the result.

References

[1] Andersen, T.G., L. Benzoni and J. Lund, 2002. "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239-1284.

[2] Andersen, T.G. and T. Bollerslev, 1998. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

[3] Andersen, T., Bollerslev, T., Diebold, F.X. and H. Ebens, 2001. "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43-76.

[4] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys, 2003. "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 529-626.

[5] Andersen, T.G., Bollerslev, T., Diebold, F.X. and J. Wu, 2005a. "Realized Beta: Persistence and Predictability," in T. Fomby (ed.) Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series in Honor of R.F. Engle and C.W.J. Granger , Volume B, forthcoming.

[6] Andersen, T.G., Bollerslev, T., Diebold, F.X. and J. Wu, 2005b. "A Framework for Exploring the Macroeconomic Determinants of Systematic Risk," *American Economic Review*, 95, 398-404.

[7] Andersen, T.G., T. Bollerslev, and N. Meddahi, 2005. "Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities." *Econometrica*, 73, 279 - 296.

[8] Barndorff-Nielsen, O. and N. Shephard, 2002. "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society*, Series B, 64, 253-280.

[9] Barndorff-Nielsen, O. and N. Shephard, 2003. "Realised power variation and stochastic volatility models," *Bernoulli*, volume 9, 2003, 243-265 and 1109-1111.

[10] Barndorff-Nielsen, O. and N. Shephard, 2004. "Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72, 885-925.

[11] Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard, 2007. "Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise," working paper, Oxford University.

[12] Bollerslev, T. and V. Todorov, 2007. "Jumps and Betas: A New Theoretical Framework for Disentangling and Estimating Systematic Risks," working paper, Northwestern University.

[13] Chernov, M., R. Gallant, E. Ghysels, and G. Tauchen, 2003. "Alternative models for stock price dynamics," *Journal of Econometrics*, 116, 225-257.

[14] Dovonon, P., S. Gonçalves, and N. Meddahi, 2007. "Edgeworth expansions for realized multivariate volatility measures," Université de Montréal and Imperial College London, mimeo.

[15] Freedman, D. A., 1981. "Bootstrapping regression models," *Annals of Statistics*, Vol. 9, No. 6, 1218-1228.

[16] Griffin, J., and R. Oomen, 2006. "Covariance measurement in the presence of non-synchronous trading and market microstructure noise," working paper, Warwick University.

[17] Gonçalves, S. and N. Meddahi, 2006a. "Bootstrapping realized volatility," Université de Montréal, mimeo.

[18] Gonçalves, S. and N. Meddahi, 2006b. "Edgeworth Corrections for Realized Volatility", *Econometric Reviews*, forthcoming.

[19] Hall, P., 1992. The bootstrap and Edgeworth expansion. Springer-Verlag, New York.

[20] Hayashi, T., and N. Yoshida, 2005. "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli* 11, 359–379.

[21] Huang, X. and G. Tauchen, 2005. "The relative contribution of jumps to total price variance," *Journal of Financial Econometrics*, 3, 456-499.

[22] Jacod, J., 1994. "Limit of random measures associated with the increments of a Brownian semimartingale," Preprint number 120, Laboratoire de Probabilitités, Université Pierre et Marie Curie, Paris.

[23] Jacod, J. and P. Protter, 1998. "Asymptotic error distributions for the Euler method for stochastic differential equations," *Annals of Probability* 26, 267-307.

[24] Jacod, J. and V. Todorov, 2007. "Testing for Common Arrivals of Jumps for Discretely Observed Multidimensional Processes," working paper, Northwestern University.

[25] Katz, M.L., 1963. "Note on the Berry-Esseen theorem," *Annals of Mathematical Statistics* 34, 1107-1108.

[26] Liu, R.Y., 1988. "Bootstrap procedure under some non-i.i.d. models," *Annals of Statistics* 16, 1696-1708.

[27] Mammen, E., 1993. "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics* 21, 255-285.

[28] Serfling, R.J., 1980. Approximation theorems of mathematical statistics, Wiley, New York.

[29] Viceira, L.M., 2007. "Bond Risk, Bond Return Volatility, and the Term Structure of Interest Rates," working paper, Harvard Business School.

[30] Voev, V. and A. Lunde, 2007. "Integrated covariance estimation using high-frequency data in the presence of noise," *Journal of Financial Econometrics*, 5, 68 - 104.

[31] White, H., 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838 (1980).

[32] Zhang, L., 2006. "Estimating covariation: Epps effect, microstructure noise," working paper, University of Illinois at Chicago.

[33] Zhang, L, P.A. Mykland, and Y. Ait-Sahalia, 2005a. "A tale of two time-scales: determining integrated volatility with noisy high frequency data," *Journal of the American Statistical Association*, 100, 1394-1411.

[34] Zhang, L., Mykland, P. and Y. Aït-Sahalia, 2005b. "Edgeworth expansions for realized volatility and related estimators," working paper, Princeton University.

Chapter 4

Three-Step Euclidean Likelihood Estimators with Moment Conditions Misspecification

1 Introduction

The moment-condition based inferences have been popularized by the unifying generalized method of moments (GMM) theory proposed by Hansen (1982). Inferences by GMM are computationally convenient and under fairly general regularity conditions, the GMM estimators are asymptotically normally distributed. The two-step efficient GMM estimator's asymptotic variance ties the semi parametric efficiency bound provided by Chamberlain (1987) and furthermore, in the case of global model misspecification, Hall and Inoue (2003) show that this estimator is $\sqrt{n}$-consistent and asymptotically normally distributed when cross sectional data are considered. In spite of these appealing properties, several studies have reported the GMM inference's lackluster performance in finite samples (see e.g. Altonji and Segal (1996), Andersen and Sørensen (1996), Hall and Horowitz (1996) and Brown and Newey (2002)). This finite sample performance limitation paves the way for an increasing research for alternatives to the GMM. The GMM alternative estimators include the continuously updated GMM (CU) estimator proposed by Hansen, Heaton and Yaron (1996), the maximum empirical likelihood estimator (EL) proposed by Qin and Lawless (1994) and the exponential tilting estimator (ET) by Kitamura and Stutzer (1997). These alternatives estimators are included in both the generalized empirical likelihood (GEL) class of estimators proposed by Newey and Smith (2004) and the minimum discrepancy (MD) class of estimators proposed by Corcoran (1998). Even though they all share the same first order asymptotic distribution, all these GMM alternative estimators are more computationally costly. The CU estimator is a solution of an optimization problem whose objective function often possesses multiple modes (Hansen, Heaton and Yaron (1996)) making the CU estimator less desirable (Schennach (2007)). Both EL and ET are solutions of saddle point problems and can be obtained through a grid search that involves optimization problems solving at several points in the parameter space. When large parameter vector is considered, these saddle point problems are computationally cumbersome.

Among these alternative estimators, as shown by Newey and Smith (2004), the EL estimator has the most desirable finite sample bias property. Newey and Smith (2004) also propose a bias corrected version of EL which is higher order efficient. These results hold in correctly specified moment condition models. A moment condition model is globally misspecified if the true data generating process deviates from these moment conditions such that no values in the parameter space solves the population moment conditions. In the case of global misspecification, the estimators listed above could behave very differently. Schennach (2007) establishes that, when the moment condition model is not correctly specified, the EL estimator ceases to be $\sqrt{n}$-consistent. In contrast, the ET estimator is $\sqrt{n}$-consistent and asymptotically normal under global misspecification. The exponentially tilted empirical likelihood estimator (ETEL) proposed by Schennach (2007) combines the desirable properties of both ET and EL. The ETEL estimator has a small sample bias of the same order of magnitude as the EL estimator and is $\sqrt{n}$-consistent and asymptotically normally distributed even in the case of global misspecification. Still, the ETEL estimator is as computationally costly as EL and ET.

Antoine, Bonnal and Renault (2007) propose the three-step Euclidean likelihood estimator (3S) which is computationally less demanding than both EL and ETEL with the same desirable bias property. The 3S estimator is higher order equivalent to EL in the sense that these two estimators lie in the same $O_P(n^{-3/2})$ neighbourhood of each other. By definition, the 3S estimator solves an efficient two-step GMM first order condition-like. The particularity of this equation being that both the Jacobian and the variance matrices are efficiently estimated by the Euclidean likelihood implied probabilities all evaluated at the efficient two-step GMM estimate. Even though the Euclidean likelihood implied probabilities are asymptotically nonnegative, in finite sample they may be negative. Antoine, Bonnal and Renault (2007) propose a shrinkage device which yields nonnegative Euclidean likelihood implied probabilities. Moreover, they suggest that the shrunk implied probabilities could be used to estimate the Jacobian and the variance in the 3S estimator estimation but they do not conduct any specific study on the asymptotic behaviour of the resulting estimator.

When these modified implied probabilities are used to estimate the Jacobian and the variance matrices in the first order-like equation, we call the resulting estimator the shrunk three-step Euclidean likelihood (s3S) estimator.

This paper makes three main contributions. First, we formally introduce the s3S estimator and prove that it is higher order equivalent to the EL estimator when the moment condition model is well specified. In particular, we strengthen the shrinkage factor proposed by Antoine, Bonnal and Renault (2007) to make the s3S estimator robust to model misspecification. Our second contribution is related to the efficient use of the information content of overidentifying moment conditions in the aim to perform inference about population mean η of any integrable function $g(x)$ of a random variable x. Specifically, we propose a computationally less costly algorithm that yields estimates of η which are higher order equivalent to its empirical likelihood estimate.

Let $E\psi(x_i, \theta) = 0$ be an overidentifying moment restrictions in which θ is the parameter of interest. Because there are more restrictions than components in θ, this moment condition is also informative about the distribution of the random variable x and therefore may be useful for inference on η. (See Back and Brown (1993) and Qin and Lawless (1994).) In particular, when the implied probabilities resulting from the estimation of θ are used to weight the the observations $g(x_i)'s$, the resulting estimator is more efficient than the naive sample mean. We show in particular in this paper that when the empirical likelihood implied probability functions are evaluated at the s3S estimator (and not at the empirical likelihood estimator itself which is computationally more costly to obtain) the resulting weights can be used to construct estimator of the population mean η which is higher order equivalent to its empirical likelihood estimator. We also show that the same quality of inference on η could be achieved if the implied probabilities are assessed at any estimator higher order equivalent to the EL estimator, in particular, the 3S and the ETEL estimators.

Third, we study the 3S and the s3S estimators under global misspecification in cross-sectional data framework. Inference under misspecification is getting more and more attention in econometrics literature. White (1982) studies the quasi maximum likelihood estimator when the distributional assumptions are misspecified. Hall (2000) examines the implications of model misspecification for the heteroskedasticity and autocorrelation consistent covariance matrix estimator and the GMM overidentifying restrictions test. Hall and Inoue (2003) study the GMM estimators under global misspecification while Schennach (2007) analyzes the EL and ETEL under global misspecification. In the moment condition-based inference framework, the GMM overidentification test or the Sargan test for overidentifying restrictions could reject or validate the model. In the case of rejection, if no theory is available for inferences, empirical researchers could have to drop parsimonious, robust and competitive models for forecasting for other less attractive models that pass all the overidentification tests with less predictive ability. The situation could even be more ambiguous. Hall and Inoue (2003) report several empirical researches in the literature in which inference by the usual asymptotic distributions have been performed even though the data have rejected the overidentifying restrictions. In this paper, we provide global misspecification robust inference for the 3S estimator. We show that, in the case of moment misspecification, this estimator stays $\sqrt{n}$-consistent and is asymptotically normally distributed. We also provide a shrinkage factor that makes the s3S estimator $\sqrt{n}$-consistent and asymptotically normally distributed in the case of moment misspecification. Its model misspecification robust asymptotic distribution is also provided. This third contribution of the paper also reveals that as ETEL, both the 3S and the s3S estimators are $\sqrt{n}$-consistent and asymptotically normally distributed under global misspecification. Because they are in addition easier to compute, they can be considered as two appealing alternatives to the EL and the ETEL estimators as well.

The remainder of the paper is organized as follows. Section 2 describes the model and estimators and gives some results about the higher order equivalence of the s3S estimator and the EL estimator when moment conditions are well specified. This section also presents the algorithm that we propose for higher order EL-equivalent inferences about population means. In Section 3 we derive asymptotic results for 3S and s3S under moment misspecification. Our Monte Carlo experiments are introduced in Section 4 followed by Section 5 which concludes. All proofs are gathered in the Appendix.

2 The model and estimators

The statistical model we consider in this paper is one with finite number of moment restrictions. To describe it, let x_i $(i = 1, \dots, n)$ be independent realizations of a random vector x and $\psi(x, \theta)$ a known q-vector of functions of the data observation x and the parameter θ which may lie in a compact parameter set $\Theta \subset \mathbb{R}^p$ $(q \geq p)$. We assume in this section that the moment restriction model is well specified in the sense that it exists a true parameter value θ_0 satisfying the moment condition

$$E\psi_i(\theta_0) = 0. \tag{1}$$

where $\psi_i(\theta) \equiv \psi(x_i, \theta)$.

In such a moment condition model, the most popular estimator is the efficient two-step GMM estimator proposed by Hansen (1982). Let $\overline{\psi}(\theta) = \sum_{i=1}^{n} \psi(x_i, \theta)/n$, $\Omega_n(\theta) = \sum_{i=1}^{n} \psi_i(\theta)\psi_i'(\theta)/n$ and also, let $\tilde{\theta}$ be some first step preliminary (possibly asymptotically inefficient) GMM estimator of θ.

The efficient two-step GMM estimator is

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \overline{\psi}'(\theta)\Omega_n^{-1}(\tilde{\theta})\overline{\psi}(\theta).$$

Under some standard regularity assumptions, the two-step GMM estimator is asymptotically normally distributed and semiparametrically efficient (Chamberlain (1987)).

To describe the CU, EL and ET estimators, it of some interest to first introduce the class of minimum discrepancy (MD) estimators and the class of generalized empirical likelihood (GEL) estimators of which they are particular examples. The class of minimum discrepancy (MD) estimator was formulated by Corcoran (1998). Let h be a real-valued convex function of a scalar π. The minimum discrepancy estimator based on the discrepancy function h is

$$\hat{\theta}^h = \arg\min_{\theta,\,\pi_i} \sum_{i=1}^{n} h(\pi_i), \quad \text{subject to} \quad \sum_{i=1}^{n} \pi_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} \pi_i \psi_i(\theta) = 0.$$

When $h(\pi) = -ln(n\pi)$, the corresponding estimator is known as the maximum empirical likelihood estimator (EL) (Qin and Lawless (1994)) and $h(\pi) = n\pi ln(n\pi)$ yields the exponential tilting estimator (ET). Since this last discrepancy function corresponds to the Kullback-Leibler Information Criterion (KLIC), ET is also known as the KLIC estimator. When $h(\pi) = (1/2)[(n\pi)^2 - 1]/n$, the corresponding estimator is known as the Euclidean empirical likelihood (EEL) estimator which also corresponds to the continuously updating (CU) estimator proposed by Hansen, Heaton and Yaron (1996). It is worth noting that this quadratic discrepancy function belongs to the family of Cressie-Read power divergence statistics introduced by Cressie and Read (1984). For $\lambda \in \mathbb{R} \setminus \{0, 1\}$, the power-divergence statistics is given by $h_\lambda(\pi) = [\lambda(\lambda-1)]^{-1}[(n\pi)^{1-\lambda} - 1]/n$. The quadratic discrepancy function corresponds to h_{-1}.

Let $\hat{\pi}_i(\hat{\theta}^h)$ $(i = 1, \ldots, n)$ be the solutions for π_i $(i = 1, \ldots, n)$ for this optimization program. $\hat{\pi}_i(\hat{\theta}^h)$ $(i = 1, \ldots, n)$ are interpreted as the empirical distribution of the random variable x on the drawn sample x_i $(i = 1, \ldots, n)$ and thus are called implied probabilities. They are useful to construct more efficient empirical estimates of data generating process (see e.g. Back and Brown (1993), Qin and Lawless (1994), Imbens, Spady and Johnson (1998) and Newey and Smith (2004)).

Newey and Smith (2004) propose the generalized empirical likelihood (GEL) class estimators. Let ρ be a concave function of a scalar v defined on ϑ, an open interval containing zero. The GEL estimator based on ρ is

$$\hat{\theta}^{gel} = \min_{\theta} \sup_{\lambda \in \Lambda_T(\theta)} \sum_{t=1}^{T} \rho(\lambda' \psi_t(\theta)),$$

where $\widehat{\Lambda}_T(\theta) = \{\lambda : \lambda'\psi_t(\theta) \in \vartheta, t = 1, \ldots, T\}$. The GEL estimators corresponding to $\rho(v) = ln(1-v)$, $\rho(v) = -\exp v$ and $\rho(v) = -v^2/2 - v$ are the EL, ET and CU estimators, respectively. Newey and Smith (2004) show that the MD estimator obtained from any power divergence statistics has an equivalence in the GEL class estimators. This result have been generalized by Ragusa (2005) to the whole class of MD estimators.

Among the GEL estimators, Newey and Smith (2004) show that EL has the most desirable finite sample bias. However, EL ceases, as shown by Schennach (2007), to be $\sqrt{n}$-consistent in the case of model misspecification. In contrast, under some regularity conditions, ET is $\sqrt{n}$-consistent in the case of model misspecification (Imbens (1997)). Taking advantage from the bias interest of EL and the robustness of ET, Schennach (2007) proposes the exponentially tilted empirical likelihood (ETEL) estimator $\hat{\theta}^{etel}$ given by

$$\hat{\theta}^{etel} = \arg\min_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} \tilde{h}(\hat{\pi}_i(\theta)),$$

where $\hat{\pi}_i(\theta)$ $(i = 1, \ldots, n)$ solve

$$\min_{\pi_i} n^{-1} \sum_{i=1}^{n} h(\pi_i) \quad \text{subject to} \sum_{i=1}^{n} \pi_i \psi_i(\theta) = 0 \ and \ \sum_{i=1}^{n} \pi_i = 1$$

and where $\tilde{h}(\pi) = -ln(n\pi)$ and $h(\pi) = n\pi ln(n\pi)$.

Schennach (2007) shows that the ETEL estimator has the same $O(n^{-1})$ bias as EL and therefore is better than ET in terms of finite sample bias. In addition, it also stays $\sqrt{n}$-consistent in the case of model misspecification.

More recently, Antoine, Bonnal and Renault (2007) have proposed the three-step Euclidean likelihood (3S) estimator which is computationally less demanding than ETEL or any MD estimator as it involves only two quadratic optimization problems and a GMM first order condition-like resolution. The 3S estimator is the solution of

$$\left[\sum_{i=1}^{n} \pi_i(\hat{\theta}) \frac{\partial \psi_i}{\partial \theta'}(\hat{\theta})\right] \left[\sum_{i=1}^{n} \pi_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta})\right]^{-1} \sum_{i=1}^{n} \psi_i(\theta) = 0 \tag{2}$$

where $\hat{\theta}$ is the efficient two-step GMM estimator and

$$\begin{aligned}\pi_i(\theta) &= n^{-1} - n^{-1}(\psi_i(\theta) - \overline{\psi}(\theta))' V_n^{-1}(\theta)\overline{\psi}(\theta),\\ V_n(\theta) &= n^{-1}\sum_{t=1}^{n}\psi_i(\theta)(\psi_i(\theta) - \overline{\psi}(\theta))'.\end{aligned} \tag{3}$$

$\pi_i(\theta)$ $(i = 1, \dots, n)$ are the implied probabilities yield by the quadratic discrepancy function evaluated at θ. In Equation (3), the variance and the Jacobian of $\psi_i(\theta)$ at θ_0 are estimated using $\pi_i(\hat{\theta})$'s as weights and are more efficient than sample means which use uniform weights. This efficiency results from the fact that the Euclidian likelihood implied probabilities provide population expectation estimates using the overidentifying moment conditions as control variables.

However, the nonnegativity of Euclidean likelihood implied probabilities function as given by Equations (3) is not guaranteed. Nonnegative implied probabilities are desirable to allow for probability interpretation in the usual sense. In addition, they are useful in sampling methods that take advantage from the information content of the moment conditions (Brown and Newey (2002)). The use of the shrinkage factor correction proposed by Antoine, Bonnal and Renault (2007) avoids negative implied probabilities. Because both corrected and non corrected implied probabilities are higher order asymptotically equivalent, the resulting estimators from each of them are asymptotically equivalent at least at the first order. The corrected implied probabilities $\tilde{\pi}_i(.)$ $(i = 1, \dots, n)$ are defined as convex combination of $\pi_i(.)$ and the uniform weight $1/n$ and are nonnegative by construction

$$\tilde{\pi}_i(\theta) = \frac{1}{1+\epsilon_n(\theta)}\pi_i(\theta) + \frac{\epsilon_n(\theta)}{1+\epsilon_n(\theta)}\frac{1}{n} \tag{4}$$

where the the shrinkage factor $\epsilon_n(\theta)$ converges in probability to 0 while guaranteing the nonnegativity of $\tilde{\pi}_i(\theta)$ as well. Antoine, Bonnal and Renault (2007) propose as shrinkage factor

$$\epsilon_n^0(\theta) = -n \min\left[\min_{1\le i\le n} \pi_i(\theta), 0\right]. \tag{5}$$

However, in the case of model misspecification that we discuss in the next section, this shrinkage coefficient will diverge to infinity (as soon as $\psi_i(\theta)$ has an unbounded support) but with an unknown rate. For a theoretical interest that we will discuss later, we propose the shrinkage factor $\epsilon_n^1(\theta)$

$$\epsilon_n(\theta) = \epsilon_n^1(\theta) = \sqrt{n}\epsilon_n^0(\theta).$$

$\epsilon_n^1(\theta)$ has the same benefits as $\epsilon_n^0(\theta)$ in correctly specified models. One can easily verify that $\epsilon_n^1(\theta)$ yields nonnegative corrected implied probabilities and, from Theorem 2.2 by Antoine, Bonnal and

Renault (2007) it converges in probability to 0. Since $\tilde{\pi}_i(.)$ $(i = 1, \dots, n)$ are obtained through a shrinkage procedure, we will call them shrunk implied probabilities.

Referring back to the 3S estimator as defined by Equation (2), as the optimal weights are not guaranteed to be non-negative, this may affect the accuracy of the Jacobian or the variance estimates and therefore make the resulting 3S estimator behave poorly in finite sample. This motivates the use of the shrunk implied probabilities in (2). We call the resulting estimator the shrunk three-step Euclidian likelihood (s3S) estimator.

By analogy to the three-step Euclidean likelihood estimator, we define the shrunk three-step Euclidian likelihood estimator as the solution of

$$\left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \frac{\partial \psi_i}{\partial \theta'}(\hat{\theta})\right] \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta})\right]^{-1} \sum_{i=1}^{n} \psi_i(\theta) = 0, \tag{6}$$

where $\hat{\theta}$ is the efficient two-step GMM estimator.

2.1 Asymptotic higher order equivalence of the EL and the s3S estimators

Under some standard regularity conditions, Antoine, Bonnal and Renault (2007) show that the 3S estimator is higher order equivalent to EL. These conditions include the identification of the true parameter value θ_0 by the moment restrictions in (1). This identification condition imply in particular that the moment conditions model is well specified. Specifically, they show that $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$, where $\hat{\theta}^{3s}$ and $\hat{\theta}^{el}$ denote the three-step Euclidean likelihood and the empirical likelihood estimators, respectively. As the ETEL estimator is also proven to be equivalent to EL up to $O_P(n^{-1})$, all three share the same $O(n^{-1})$ bias. The following result shows that the shrunk three-step Euclidean likelihood estimator $\hat{\theta}^{s3s}$ is also higher order equivalent to the empirical likelihood estimator $\hat{\theta}^{el}$. The following assumptions are needed. For brevity, we only highlight in the text those assumptions that are relevant to the exposition and relegate the remainder to the appendix.

Assumption 1 *i)* θ_0 *is an interior point of* Θ, *a compact subset of* $\mathbb{R}^p$.

ii) $\psi_i(.)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0.

iii) $E\psi_i(\theta) = 0 \Leftrightarrow \theta = \theta_0$.

iv) $\Omega(\theta_0) = E\psi_i(\theta_0)\psi_i'(\theta_0)$ *is a nonsingular matrix.*

v) $J^0 = E\partial\psi_i(\theta_0)/\partial\theta'$ *is of rank* p.

vi) $J^{0\prime}\Omega^{-1}(\theta_0)E\psi_i(\theta) = 0 \Leftrightarrow \theta = \theta_0$.

vii) The shrunk three-step Euclidean likelihood estimator is well defined, i.e., there is a sequence $\{\hat{\theta}_{n=1}^{\infty}\}$ *that solves* (6).

viii) $E\sup_{\theta\in\Theta}\|\psi_i(\theta)\|^{\alpha} < \infty$ *for some* $\alpha > 2$ *and* $E\sup_{\theta\in\mathcal{N}}\|\partial\psi_i(\theta)/\partial\theta'\| < \infty$.

Assumption 1 provides sufficient conditions for consistency and asymptotic normality of both the efficient two-step GMM estimator $\hat{\theta}$ and the empirical likelihood estimator $\hat{\theta}^{el}$. Assumption 1-vi) is an identification condition insuring the consistency of both $\hat{\theta}^{3s}$ and $\hat{\theta}^{s3s}$.

Theorem 2.1 *If Assumption 1, and Assumption 9 in the appendix hold, then* $\hat{\theta}^{s3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$.

The details of the proof of Theorem 2.1 are reported to the appendix. We show that $\hat{\theta}^{s3s} - \hat{\theta}^{3s} = O_P(n^{-3/2})$ and deduce the stated order of magnitude by relying on the fact that $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$. This result, typically shows that the shrunk three-step Euclidean likelihood estimator has the same first order asymptotic distribution as the empirical likelihood estimator and both have the same $O(n^{-1})$ bias as well.

Next, we show how the closeness of $\hat{\theta}^{3s}$, $\hat{\theta}^{s3s}$ and $\hat{\theta}^{el}$ can be exploited to make easier inferences about population means.

2.2 Inference about population means

When moment conditions in Equation (1) overidentify the parameter of interest, they are also informative about the data generating process distribution (see Back and Brown (1993)). For any integrable function $g(x)$, the implied probabilities can be used to perform inference about $\eta = Eg(x)$. Particularly, for any minimum discrepancy estimator based on power divergence statistics, $\hat{\theta}^{pd}$, Antoine, Bonnal and Renault (2007) show that $\hat{\eta} = \sum_{i=1}^{n}\pi_i^{pd}(\hat{\theta}^{pd})g(x_i)$ is an estimator of η more efficient than the sample mean of $g(x)$ which assigns uniform weighs to the observations.

Moreover, $\hat{\eta}$ is an efficient estimator of η. To see this, let us consider the following augmented moment restrictions

$$E(\psi_i'(\theta), (g(x_i) - \eta)')' = 0 \tag{7}$$

and $\hat{\beta}^{pd}$ the minimum power divergence estimator of $\beta = (\theta', \eta')'$ based on these augmented moment restrictions. Antoine, Bonnal and Renault (2007) show that $\hat{\beta}^{pd}$ corresponds in its η-argument to $\hat{\eta}$ and in its θ-argument to the minimum power divergence estimator of θ, $\hat{\theta}^{pd}$, based on the non augmented moment restrictions $E\psi_i(\theta) = 0$. Because $\hat{\beta}^{pd}$ is an efficient estimator, so is $\hat{\eta}$ of which estimation takes advantage from the non extended moment conditions. We extend this result to the whole class of minimum discrepancy estimators.

Let $\hat{\theta}^h$ be a Minimum Discrepancy estimator based on (1) with a discrepancy function h (the h-Minimum Discrepancy estimator) and $\pi_i^h(\hat{\theta}_h)$ $(i = 1, \dots, n)$ its corresponding implied probabilities.

Theorem 2.2 *Let $\hat{\beta}^h = (\hat{\theta}^{1h'}, \hat{\eta}^{h'})'$ be the h-Minimum Discrepancy estimator of $\beta = (\theta', \eta')'$ based on the augmented moment restrictions in Equation (7). If $\psi_i(.)$ is differentiable on Θ with probability one and all of the necessary conditions of Lagrange Theorem for constrained optimization are fulfilled, then*

$$\hat{\eta}^h = \sum_{i=1}^{n} \pi_i^h(\hat{\theta}^h) g(x_i) \quad \text{and} \quad \hat{\theta}^{1h} = \hat{\theta}^h.$$

This result shows that there is no need to solve for the augmented moment conditions program to get the h-minimum discrepancy estimator $\hat{\eta}^h$ of η. Actually, one just has to get the h-minimum discrepancy estimator of θ based on (1) and the resulting implied probabilities help to compute $\hat{\eta}^h$ which, in turn, as a minimum discrepancy estimator, is more efficient than the sample mean as soon as the restrictions in (1) are overidentifying for the true parameter value θ_0.

Let now $\hat{\eta}^{el}$ be the minimum discrepancy estimator of η obtained from $\pi_i^{el}(\hat{\theta}^{el})$ $(i = 1, \dots, n)$, the EL implied probabilities evaluated at the EL estimator of θ by (1). It is known that $\hat{\eta}^{el}$ will have a more desirable higher order properties over the other $\hat{\eta}^h$ in terms of bias (Newey and Smith (2004)).

The aim of the following result is to provide an estimator of η computationally less costly than $\hat{\eta}^{el}$ but higher order equivalent.

For any $\theta \in \Theta$, let $\pi_i^{el}(\theta)$ be the implied probabilities obtained at θ by the empirical likelihood discrepancy function. Theorem 2.3 below shows that any estimator $\hat{\theta}$ of θ_0 which is in a $O_P(n^{-3/2})$ neighborhood of $\hat{\theta}^{el}$ leads to $\hat{\eta} = \sum_{i=1}^{n} \pi_i^{el}(\hat{\theta}) g(x_i)$ sharing with $\hat{\eta}^{el}$ the same higher order bias. The following assumption is needed.

Assumption 2 *There exists a measurable function $b(x)$ such that, in a neighbourhood $\mathcal{N}$ of θ_0 and for any $k = 1, 2, \dots, q$, $s, u = 1, 2, \dots, p$, $|\psi_k(x, \theta)| \|g(x)\| < b(x)$, $|\partial^2 \psi_k(x, \theta)/\partial\theta_s \theta_u| \|g(x)\| < b(x)$ and $E\{b(x)\} < \infty$ and $E\|\partial\psi_k(x, \theta_0)/\partial\theta_s\| \|g(x)\| < \infty$, ψ_k is the k-th component of ψ_i.*

Theorem 2.3 *Let $\hat{\theta}$ be any estimator of θ_0 such that $\hat{\theta} - \hat{\theta}^{el} = O_P(n^{-3/2})$ and $\hat{\eta} = \sum_{i=1}^{n} \pi_i^{el}(\hat{\theta})g(x_i)$. If Assumptions 1 and 2 are satisfied, then $\hat{\eta} - \hat{\eta}^{el} = O_P(n^{-3/2})$.*

From this result, either ETEL, 3S or s3S estimator will have its empirical likelihood implied probabilities leading to estimate of η that is higher order equivalent to $\hat{\eta}^{el}$. The empirical likelihood implied probabilities at $\hat{\theta} \in \Theta$, $\pi_i^{el}(\hat{\theta})$ $(i = 1, \ldots, n)$, are given by $\pi_i^{el}(\hat{\theta}) = 1/n[1 + \hat{\lambda}'\psi_i(\theta)]$ with

$$\hat{\lambda} = \arg \min_{\lambda \in \hat{\Lambda}(\hat{\theta})} -\sum_{i=1}^{n} \log[1 + \lambda'\psi_i(\hat{\theta})]/n,$$

see Qin and Lawless (1994).

Therefore, once the 3S or s3S estimator is computed, one can easily, by a single optimization, get $\pi_i^{el}(\hat{\theta}^{3s})$ or $\pi_i^{el}(\hat{\theta}^{s3s})$ $(i = 1, \ldots, n)$. $\hat{\eta} = \sum_{i=1}^{n} \pi_i^{el}(\hat{\theta}^{3s})g(x_i)$ or $\hat{\eta} = \sum_{i=1}^{n} \pi_i^{el}(\hat{\theta}^{s3s})g(x_i)$ are both estimates of η, higher order equivalent to $\hat{\eta}^{el}$. In this procedure, one no longer needs to solve for the saddle point program that lead to EL or ETEL estimators to get an estimator for η which is higher order equivalent to the empirical likelihood estimator.

The next section studies the 3S and the s3S estimators in the case of model misspecification.

3 The limiting behaviour of the 3S and s3S estimators in misspecified models

In this section, we study the behaviour of the three-step Euclidean (3S) likelihood estimator and the shrunk three-step Euclidean likelihood (s3S) estimator in misspecified models. Following Hall (2000), Hall and Inoue (2003) and Schennach (2007), we consider a moment restriction model as given in (1) as misspecified, when there is no value of θ at which the population moment condition is satisfied. In the literature, this case is commonly referred to as non-local or global misspecification. Hall and Inoue (2003) study the two-step GMM estimator under global misspecification. Specifically, when data are independent and identically distributed, Hall and Inoue (2003) show that the two-step GMM estimator is $\sqrt{n}$-consistent and asymptotically normally distributed. Our work relies on Hall and Inoue (2003) results that we extend to the 3S and the s3S estimators. Certain assumptions are required to analyze the large sample properties of these estimators. As in the last section and for brevity, we only highlight in the text those assumptions that are relevant to the exposition and relegate the remainder to the appendix.

Assumption 3 *x_i forms an i.i.d. sequence.*

Let $\mu(\theta) = E\psi(x_i, \theta)$ and $\psi_i(\theta) = \psi(x_i, \theta)$.

Assumption 4 *i) $\mu : \Theta \to \mathbb{R}^q$ such that $\|\mu(\theta)\| > 0$ for all $\theta \in \Theta$.*

ii) W_n is a positive semidefinite matrix that converges in probability to the positive definite matrix of constants W.

iii) (Identification) There exists $\theta_* \in \Theta$ such that $Q_0(\theta_*) < Q_0(\theta)$ for any $\theta \in \Theta \setminus \{\theta_*\}$ where $Q_0(\theta) = E\psi_i'(\theta) W E\psi_i(\theta)$.

As in Hall (2000) and Hall and Inoue (2003), Assumption 4-i) captures the global model misspecification. Assumption 4-iii) is the identification condition for a misspecified model. It states that the GMM population objective function given by $Q_0(\theta)$ is minimized at only one point, θ_*, in the parameter set Θ. θ_* is often referred to as the pseudo true parameter value. In a well specified model, θ_* corresponds to the true parameter value θ_0 and $Q_0(\theta_0) = 0$.

Let $\bar{\theta} = \arg\min_{\theta \in \Theta} \bar{\psi}(\theta)' W_n \bar{\psi}(\theta)$ be the GMM estimator defined by the weighting matrix W_n. Under Assumptions 3, 4 and Assumption 12 given in the appendix, Lemma 1 by Hall (2000) applies and $\bar{\theta}$ is consistent for θ_*. This result includes the two-step GMM estimator $\hat{\theta}$ under mild further assumptions. The problem that arises with the two-step GMM estimator is that the weighting matrix it relies on depends on a first step GMM estimator $\tilde{\theta}$ which needs to be consistent. Usually, $\tilde{\theta}$ is obtained by a non random positive definite weighting matrix W^1. We introduce in Appendix B the specific regularity conditions that guarantee the consistency and asymptotic normality of $\tilde{\theta}$ and $\hat{\theta}$.

To describe the asymptotic behaviour of the three-step Euclidean likelihood and the shrunk three-step Euclidean likelihood estimators, we need to introduce some notation. For $\theta \in \Theta$, let

$$
\begin{aligned}
\bar{G}(\theta) &= \textstyle\sum_{i=1}^n \pi_i(\theta)\{\partial\psi_i'(\theta)/\partial\theta\}, \quad & \bar{G}^s(\theta) &= \textstyle\sum_{i=1}^n \tilde{\pi}_i(\theta)\{\partial\psi_i'(\theta)/\partial\theta\}, \\
\bar{M}(\theta) &= \textstyle\sum_{i=1}^n \pi_i(\theta)\psi_i(\theta)\psi_i'(\theta), & \bar{M}^s(\theta) &= \textstyle\sum_{i=1}^n \tilde{\pi}_i(\theta)\psi_i(\theta)\psi_i'(\theta), \\
G(\theta) &= E(\partial\psi_i'(\theta)/\partial\theta) - Cov\{\psi_i'(\theta)V^{-1}(\theta)E(\psi_i(\theta)), (\partial\psi_i'(\theta)/\partial\theta)\}, \\
M(\theta) &= E\psi_i(\theta)\psi_i'(\theta) - Cov\{\psi_i'(\theta)V^{-1}(\theta)E(\psi_i(\theta)), \psi_i(\theta)\psi_i'(\theta)\}
\end{aligned}
$$

$\pi_i(\theta)$, $\tilde{\pi}_i(\theta)$ are defined as in Equations (3) and (4) and $V(\theta) = Var(\psi_i(\theta))$.

The three-step Euclidean likelihood estimator $\hat{\theta}^{3s}$ is the solution of

$$\bar{G}(\hat{\theta})\bar{M}(\hat{\theta})^{-1}\bar{\psi}(\theta) = 0 \tag{8}$$

and the shrunk three-step Euclidean likelihood estimator $\hat{\theta}^{s3s}$ is the solution of

$$\bar{G}^s(\hat{\theta})\bar{M}^s(\hat{\theta})^{-1}\bar{\psi}(\theta) = 0, \tag{9}$$

where $\hat{\theta}$ is the two-step GMM estimator. The following assumptions are necessary to show the consistency of $\hat{\theta}^{3s}$ and $\hat{\theta}^{s3s}$.

Assumption 5 *i) $M(\theta_*)$ is nonsingular and for $\theta \in \Theta$, $G(\theta_*)M(\theta_*)^{-1}E\psi_i(\theta) = 0 \Leftrightarrow \theta = \theta_{**}$.*
ii) The three-step Euclidean likelihood estimator is well defined, i.e., there is a sequence $\{\hat{\theta}_n^{3s}\}_{n=1}^{\infty}$ such that $\bar{G}(\hat{\theta})\bar{M}(\hat{\theta})^{-1}\bar{\psi}(\hat{\theta}^{3s}) = 0$ a.s.

Assumption 5-i) is the identification condition for misspecified model for the 3S estimator problem. Typically, it states that the population version of Equation (8) has a unique solution, θ_{**}, in the parameter set Θ. θ_{**} is the pseudo true value for the three-step Euclidean likelihood estimator $\hat{\theta}^{3s}$. Obviously θ_{**} depends on both θ_* and W. However, we will not explicitly mention this dependence for sake of simplicity. In the next two theorems, we assume that Assumption 4 holds for the two-step GMM estimator $\hat{\theta}$.

Theorem 3.1 *If Assumptions 3-5, and Assumptions 12-13 in Appendix hold, then $\hat{\theta}^{3s} \xrightarrow{P} \theta_{**}$.*

The shrinkage factor makes the analysis of the shrunk three-step Euclidean likelihood estimator more difficult in the case of misspecified model. The shrinkage factor that we use is $\epsilon_n^1(\theta) = -\sqrt{n}\min\{0, \min_{1\le i\le n}[1 - \bar{\psi}'(\theta)V_n^{-1}(\theta)(\psi_i(\theta) - \bar{\psi}(\theta))]\}$. In the case of correctly specified model, $\bar{\psi}(\theta_0)$ converges to 0 and under some regularity assumption, $\epsilon_n^1(\theta_0)$ also converges to 0. However, when the model is misspecified, $\bar{\psi}(\theta)$ does not converge to 0 for any value of θ. For a large n, $\epsilon_n^1(\theta) \overset{a}{\sim} -\sqrt{n}\min\{0, \min_{1\le i\le n}[1 - E\{\psi_i'(\theta)\}V^{-1}(\theta)(\psi_i(\theta) - E\psi_i(\theta))]\}$. For our analysis, we need to have an insight of the order of magnitude of $\epsilon_n^1(\hat{\theta})$. Let $l_i = \inf_{\theta\in\bar{\mathcal{N}}_*}\psi_i'(\theta)V^{-1}(\theta)E\psi_i(\theta)$, where $\bar{\mathcal{N}}_*$ is closed neighbourhood of θ_* included in Θ. We make the following assumption.

Assumption 6 *i) $\forall a, b \in \mathbb{R}$, $a \neq b$, $Prob[l_i \in (a, b)] \neq 0$.*
ii) $E\{\psi_i(\theta_*)\psi_i'(\theta_*)\}$ is nonsingular and $E[\partial\psi_i'(\theta_*)/\partial\theta][E\psi_i(\theta_*)\psi_i'(\theta_*)]^{-1}E\psi_i(\theta) = 0 \Leftrightarrow \theta = \theta_{**}$, for

$\theta \in \Theta$.

iii) The shrunk three-step Euclidean likelihood estimator is well defined, i.e., there is a sequence $\{\hat{\theta}_n^{s3s}\}_{n=1}^{\infty}$ *such that* $\bar{G}^s(\hat{\theta})\bar{M}^s(\hat{\theta})^{-1}\bar{\psi}(\hat{\theta}^{s3s}) = 0$ *a.s.*

Assumption 6.i) allows l_i to lie in any interval on the real line with probability different from 0. Typically, l_i could be normally distributed. Under this assumption and some regularity conditions, $-\min\{0, \min_{1\leq i\leq n}[1-\bar{\psi}'(\hat{\theta})V_n^{-1}(\hat{\theta})(\psi_i(\hat{\theta})-\bar{\psi}(\hat{\theta}))]\}$ diverges to infinity and the factor $\sqrt{n}$ gives an idea about the divergence rate of $\epsilon_n^1(\hat{\theta})$. If Assumption 6.i) holds, $\sum_{i=1}^n \pi_i(\hat{\theta})y_i \overset{a}{\sim} \sum_{i=1}^n y_i/n$, where y_i is any measurable function of x_i. Therefore, Equation (9) is equivalent, up to some negligible terms, to $(\sum_{i=1}^n\{\partial\psi_i(\hat{\theta})/\partial\theta'\}/n)(\sum_{i=1}^n\{\psi_i(\hat{\theta})\psi_i'(\hat{\theta})\}/n)^{-1}\sum_{i=1}^n\psi_i(\theta) = 0$. The identification condition for the s3S estimator given by Assumption 6.ii) is related to the population version of this last equation.

Theorem 3.2 *If Assumptions 3, 4, 6, and Assumptions 12-13 in Appendix hold, and that* $\hat{\theta} - \theta_* = O_P(n^{-1/2})$*, where* $\hat{\theta}$ *is the two-step GMM estimator, then* $\hat{\theta}^{s3s} \xrightarrow{P} \theta_{**}$.

We, next provide asymptotic distribution for both the three-step Euclidean likelihood and the shrunk three-step Euclidean likelihood estimators in misspecified models. Since these estimators rely on the two-step GMM estimator, the asymptotic distribution derived by Hall and Inoue (2003) for the two-step GMM in misspecified models are useful for our asymptotic theory. We recall their results that we also specialize for our use.

3.1 Asymptotic distribution of the two-step GMM estimator in misspecified models

The first step GMM estimator $\tilde{\theta}$ solves

$$\frac{\partial\bar{\psi}'}{\partial\theta}(\tilde{\theta})W^1\bar{\psi}(\tilde{\theta}) = 0, \tag{10}$$

where W^1 is, usually, a non-random weighting matrix. Often, in empirical works, the identity matrix is used as weighting matrix. We treat it here as non-random. Under Assumption 3 and Assumptions 10, 12 as given in Appendix, the results by Hall and Inoue (2003) apply and $\tilde{\theta} - \theta_*^1 = O_P(n^{-1/2})$, θ_*^1 being the unique solution of the population analogue of Equation (10).

Actually, a simple Taylor expansion of the first order condition in (10) around θ_*^1 yields

$$0 = \frac{\partial\bar{\psi}'}{\partial\theta}(\theta_*^1)W^1\bar{\psi}(\theta_*^1) + \left[\frac{\partial\bar{\psi}'}{\partial\theta}(\theta_*^1)W^1\frac{\partial\bar{\psi}}{\partial\theta'}(\theta_*^1) + (\bar{\psi}'(\theta_*^1)W^1 \otimes I_p)\bar{J}^{(2)}(\theta_*^1)\right](\tilde{\theta} - \theta_*^1) + O_P(n^{-1}), \tag{11}$$

where I_p is the $p \times p$-identity matrix and

$$\bar{J}^{(2)}(\theta) = \frac{\partial}{\partial \theta'} vec\left(\frac{\partial \bar{\psi}}{\partial \theta'}(\theta)\right).$$

Let $\Omega(\theta) = E\psi_i(\theta)\psi_i'(\theta)$ and

$$\begin{array}{rclrcl} J_i(\theta) & = & \partial\psi_i(\theta)/\partial\theta', & J^{(2)}(\theta) & = & E[(\partial/\partial\theta')vec\{\partial\psi_i(\theta)/\partial\theta'\}], \\ \bar{J}(\theta) & = & \partial\bar{\psi}(\theta)/\partial\theta', & \bar{H}_1(\theta) & = & \bar{J}'(\theta)W^1(\bar{J}(\theta) + (\bar{\psi}'(\theta)W^1 \otimes I_p)\bar{J}^{(2)}(\theta), \\ J(\theta) & = & EJ_i(\theta), & H_1(\theta) & = & J'(\theta)W^1J(\theta) + (E(\psi_i'(\theta))W^1 \otimes I_p)J^{(2)}(\theta). \end{array}$$

Since $\bar{H}_1(\theta)$ is a quadratic function of sample mean, $\bar{H}_1(\theta)$ is $\sqrt{n}$-consistent for its probability limit $H_1(\theta)$ meaning that $\bar{H}_1(\theta) - H_1(\theta) = O_P(n^{-1/2})$. Therefore,

$$\tilde{\theta} - \theta_*^1 = -H_1^{-1}(\theta_*^1)\bar{J}'(\theta_*^1)W^1\bar{\psi}(\theta_*^1) + O_P(n^{-1}). \tag{12}$$

On the other hand, the two-step GMM estimator solves the first order condition

$$\bar{J}'(\hat{\theta})W_n(\tilde{\theta})\bar{\psi}(\hat{\theta}) = 0, \tag{13}$$

where $W_n(\theta) = [\sum_{i=1}^n \psi_i(\theta)\psi_i'(\theta)/n]^{-1} \equiv \Omega_n^{-1}(\theta)$. The stochastic nature of the weighting matrix adds a layer of complexity of the expansion of the two-step GMM estimator.

We first expand $\Omega_n(\tilde{\theta})$ around θ_*^1 and then we deduce an expansion of $W_n(\tilde{\theta})$. This latter, ultimately allows to get an expansion for $\hat{\theta}$. We have

$$\Omega_n(\tilde{\theta}) = \Omega_n(\theta_*^1) + R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial\theta'}(\theta_*^1)(\tilde{\theta} - \theta_*^1)\right) + O_P(n^{-1}),$$

where $R_{k,l}(X)$ reshapes the kl-vector X into a $k \times l$-matrix, column-wise.

Let

$$\begin{array}{rcl} \psi_i^* & = & \psi_i(\theta_*^1), \\ J^* & = & J(\theta_*^1), \\ W^{-1} & = & E\psi_i^*\psi_i^{*\prime} - R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial\theta'}(\theta_*^1)[H_1^{-1}(\theta_*^1)J^{*\prime}W^1E\psi_i^*]\right), \\ \xi_i(\theta_*^1) & = & \psi_i^*\psi_i^{*\prime} - \Omega(\theta_1^*) + R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial\theta'}(\theta_*^1)H_1^{-1}(\theta_*^1)[(\bar{J}'(\theta_*^1) - J^{*\prime})W^1E\psi_i^* + J^{*\prime}W^1(\bar{\psi}(\theta_*^1) - E\psi_i^*)]\right), \\ \xi_{w,i}(\theta_*^1) & = & -W\xi_i(\theta_*^1)W. \end{array}$$

From the expression of $\tilde{\theta} - \theta_*^1$ given by Equation (12) and up to some arrangements, we have

$$\Omega_n(\tilde{\theta}) = W^{-1} + \frac{1}{n}\sum_{i=1}^n \xi_i(\theta_*^1) + O_P(n^{-1}).$$

Clearly, $E\xi_i(\theta_*^1) = 0$ and $\sum_{i=1}^n \xi_i(\theta_*^1)/n = O_P(n^{-1/2})$. Furthermore,

$$W_n(\tilde{\theta}) - W = \Omega_n^{-1}(\tilde{\theta}) - W = -\Omega_n^{-1}(\tilde{\theta})(\Omega_n(\tilde{\theta}) - W^{-1})W.$$

Thus

$$W_n(\tilde{\theta}) - W = \frac{1}{n}\sum_{i=1}^{n}\{-W\xi_i(\theta_*^1)W\} + O_P(n^{-1})$$

or equivalently,

$$W_n(\tilde{\theta}) - W = \frac{1}{n}\sum_{i=1}^{n}\xi_{w,i}(\theta_*^1) + O_P(n^{-1}). \tag{14}$$

Thanks to Assumption 3 and Assumptions 11, 12 as given in Appendix, we can expand the first order condition for $\hat{\theta}$ in (13) as follows

$$0 = \bar{J}'(\theta_*)W_n(\tilde{\theta})\bar{\psi}(\theta_*) + \left[\bar{J}'(\theta_*)W_n(\tilde{\theta})\bar{J}(\theta_*) + (\bar{\psi}'(\theta_*)W_n(\tilde{\theta}) \otimes I_p)\bar{J}^{(2)}(\theta_*)\right](\hat{\theta} - \theta_*) + O_P(n^{-1}),$$

Let

$$\begin{aligned}
\psi_{*i} &= \psi_i(\theta_*), \\
\mu_* &= E\psi_{*i}, \\
J_* &= J(\theta_*), \\
\bar{H}(\theta) &= \bar{J}'(\theta)W_n(\tilde{\theta})\bar{J}(\theta) + (\bar{\psi}'(\theta)W_n(\tilde{\theta}) \otimes I_p)\bar{J}^{(2)}(\theta), \\
H(\theta) &= J'(\theta)WJ(\theta) + \{E(\psi_i'(\theta))W \otimes I_p\}J^{(2)}(\theta).
\end{aligned}$$

Because $\bar{H}(\theta)$ is a polynomial function of sample mean, $\bar{H}(\theta_*)$ is $\sqrt{n}$-consistent for its probability limit $H(\theta_*)$ meaning that $\bar{H}(\theta_*) - H(\theta_*) = O_P(n^{-1/2})$. Therefore,

$$\hat{\theta} - \theta_* = -H^{-1}(\theta_*)\bar{J}'(\theta_*)W_n(\tilde{\theta})\bar{\psi}(\theta_*) + O_P(n^{-1}).$$

Thus $\hat{\theta} - \theta_*$ can be written

$$\hat{\theta} - \theta_* = -H^{-1}(\theta_*)\{[\bar{J}'(\theta_*) - J_*']W\mu_* + J_*'[W_n(\tilde{\theta}) - W]\mu_* + J_*'W[\bar{\psi}(\theta_*) - \mu_*]\} + O_P(n^{-1}). \tag{15}$$

From Equations (14) and (15), $\hat{\theta} - \theta_*$ is asymptotically equivalent to a sample mean of centered random vectors which are i.i.d as is $x_i : i = 1, \ldots$. Assuming, as it is the case here that these vectors have finite variance, the central limit theorem applies and $\sqrt{n}(\hat{\theta} - \theta_*) = O_P(1)$ as it is asymptotically Gaussian. This is a result of Hall and Inoue (2003).

The main reason of this usual Gaussian asymptotic behaviour of the two-step efficient GMM estimator is the cross sectional nature of the random variables as they are assumed to be i.i.d. This result falls down in the time series context where the lag dependence is not finite and the moment conditions are globally misspecified. In such a case, as shown by Hall and Inoue (2003) (see also Hall (2000)), the optimal weight for the two-step efficient GMM estimator dictates its rate of convergence to the GMM estimator which therefore may no longer be $\sqrt{n}$-consistent or even asymptotically Gaussian.

3.2 Asymptotic distributions of the three-step Euclidean likelihood estimators

In this section, we derive the asymptotic distribution of both the 3S and the s3S estimators under global misspecification. We find that they are $\sqrt{n}$-consistent and are asymptotically characterized by a normal distribution. The asymptotic normality of the 3S estimator is not surprising as its estimating equation is a smooth function of sample mean and the efficient two-step GMM estimator is also asymptotically Gaussian.

Besides, the estimating equation of the s3S estimators is not a smooth function of sample means. This makes less apparent the reason of its asymptotic normal behaviour. Let us consider again the shrunk implied probabilities as introduced by (4)

$$\begin{aligned}\tilde{\pi}_i(\theta) &= \frac{1}{1+\epsilon_n(\theta)}\pi_i(\theta) + \frac{\epsilon_n(\theta)}{1+\epsilon_n(\theta)}\frac{1}{n} \\ &= \frac{1}{n} - \frac{1}{1+\epsilon_n(\theta)}\frac{1}{n}\left(\psi_i(\theta) - \bar{\psi}_i(\theta)\right)' V_n^{-1}(\theta)\bar{\psi}(\theta).\end{aligned}$$

In a correctly specified model, the term $\frac{1}{n}\left(\psi_i(\theta) - \bar{\psi}_i(\theta)\right)' V_n^{-1}(\theta)\bar{\psi}(\theta)$ correct the uniform weight $\frac{1}{n}$ to deliver population means estimates which are more efficient than the sample mean by using the information content of the moment conditions. $\epsilon_n(\theta)$ adjusts for non-negative weights in finite sample and vanishes asymptotically. However, in misspecified models and as pointed out by Schennach (2007), this shrinkage factor does not vanish asymptotically. Nevertheless, we can see that under mild assumptions $\epsilon_n(\theta)$ diverge to infinity. The key idea to conserve asymptotic normality for the s3S estimator is to accelerate the divergence of this factor such that the discontinuous part of its estimating equation appears negligible compared to the smooth part of this function. In this paper, we use

$$\epsilon_n^1(\theta) = \sqrt{n}\epsilon_n^0(\theta),$$

$\epsilon_n^0(\theta)$ is given by Equation (5). However, we should notice that any shrinkage factor $\epsilon_{\alpha,n}(\theta) = n^\alpha \epsilon_n^0(\theta)$ with $\alpha \geq 1/2$ could lead to the same result.

The three-step Euclidean likelihood estimator $\hat{\theta}^{3s}$ solves (8) and, by the mean value expansion of (8) around θ_{**}, we have

$$\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\frac{\partial\bar{\psi}}{\partial\theta'}(\bar{\theta})(\hat{\theta}^{3s} - \theta_{**}) = -\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}),$$

where $\bar{\theta} \in (\hat{\theta}^{3s}, \theta_{**})$.

To show that $\sqrt{n}(\hat{\theta}^{3s} - \theta_{**})$ is asymptotically normally distributed, we just have to show that the right hand side of the last equation properly scaled is asymptotically Gaussian and the term multiplying the $\hat{\theta}^{3s} - \theta_{**}$ in the left hand side is asymptotically non singular.

The following assumptions are also useful.

Assumption 7 *i)* $\theta_{**} \in Int(\Theta)$.

ii) There exists a measurable function $b(x)$ *such that, in a neighbourhood of* θ_{**}, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b(x)$, *for all* $k = 1, 2, \ldots, q$ *and* $s = 1, 2, \ldots, p$ *and* $Eb(x) < \infty$.

iii) $D_* = G(\theta_*)M^{-1}(\theta_*)J(\theta_{**})$ *is nonsingular.*

iv) $Var z_{3,i} < \infty$, *where* $z_{3,i} = \{vec'J_i(\theta_*), [\psi_i(\theta_*) \otimes vecJ_i(\theta_*)]', [\psi_i(\theta_*) \otimes vec\psi_i(\theta_*), \psi_i'(\theta_*)]', \psi_i(\theta_{**})\}'$.

Let

$$\begin{aligned}
\mu_{**} =& E\psi_i(\theta_{**}),\\
m_* =& M^{-1}(\theta_*),\\
g_* =& G(\theta_*),\\
Y_{1n} =& R_{p,q}\left(\frac{\partial vec[G]}{\partial\theta'}(\theta_*)(\hat{\theta} - \theta_*)\right) - g_* m_* R_{q,q}\left(\frac{\partial vec[M]}{\partial\theta'}(\theta_*)(\hat{\theta} - \theta_*)\right).
\end{aligned}$$

Theorem 3.3 *If Assumptions 3, 5, 8 and Assumptions 11, 12, and 13 given in appendix hold, then*

$$\sqrt{n}\begin{pmatrix} \bar{\psi}(\theta_{**}) - \mu_{**} \\ [\bar{G}(\theta_*) - G(\theta_*)]m_*\mu_{**} \\ [\bar{M}(\theta_*) - M(\theta_*)]m_*\mu_{**} \\ Y_{1n}m_*\mu_{**} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} \end{pmatrix}\right),$$

$$\sqrt{n}\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) \xrightarrow{d} \mathcal{N}(0, \Omega_*)$$

and

$$\sqrt{n}(\hat{\theta}^{s3s} - \theta_{**}) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where $\Sigma = D_*^{-1}\Omega_* D_*^{-1'}$ *and*

$$\begin{aligned}
\Omega_* =& g_* m_* \Omega_{11} m_* g_*' + \Omega_{22} + g_* m_* \Omega_{33} m_* g_*' + \Omega_{44} + g_* m_* \Omega_{12} + \Omega_{21} m_* g_*' - g_* m_* (\Omega_{13} + \Omega_{31}) m_* g_*' \\
&+ g_* m_* \Omega_{14} + \Omega_{41} m_* g_*' - g_* m_* \Omega_{32} - \Omega_{23} m_* g_*' + \Omega_{24} + \Omega_{42} - g_* m_* \Omega_{34} - \Omega_{43} m_* g_*'.
\end{aligned}$$

Next, we derive the asymptotic distribution of the shrunk three-step estimator $\hat{\theta}^{s3s}$. Under Assumptions 3, 6 and Assumptions 11, 12, and 13 in Appendix, as we show in the proof of Theorem

3.2,

$$\frac{1}{1+\epsilon_n^1(\hat{\theta})} = o_P(n^{-1/2}).$$

As a result,

$$\sum_{i=1}^{n} \bar{\pi}_i(\hat{\theta}) J_i'(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} J_i'(\hat{\theta}) + o_P(n^{-1/2}) \quad \text{and} \quad \sum_{i=1}^{n} \bar{\pi}_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \psi_i(\hat{\theta})\psi_i'(\hat{\theta}) + o_P(n^{-1/2}).$$

Hence, $\bar{G}s(\hat{\theta})\bar{M}^{s^{-1}}(\hat{\theta})\bar{\psi}(\hat{\theta}^{s3s}) = 0 = \bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}^{s3s}) + o_P(n^{-1/2})$. By a mean value expansion of $\bar{\psi}(\theta)$ around θ_{**}, we have

$$\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{J}(\bar{\theta})(\hat{\theta}^{s3s} - \theta_{**}) = -\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) + o_P(n^{-1/2}),$$

where $\bar{\theta} \in (\hat{\theta}^{s3s}, \theta_{**})$.

As for the 3S estimator, we get the asymptotic normality for the s3S estimator by establishing that the right hand side of the last equation scaled by square root of n is asymptotically Gaussian and by insuring that the factor of $\hat{\theta}^{s3s} - \theta_{**}$ is asymptotically non singular.

We make the following assumptions.

Assumption 8 *i)* $\theta_{**} \in Int(\Theta)$.

ii) There exists a measurable function $b(x)$ *such that, in a neighbourhood of* θ_{**}, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b(x)$, *for all* $k = 1, 2, \dots, q$ *and* $s = 1, 2, \dots, p$ *and* $Eb(x) < \infty$.

iii) $D_*^s = J'(\theta_*)\Omega^{-1}(\theta_*)J(\theta_{**})$ *is nonsingular.*

iv) Let $Var\,(vec' J_i(\theta_*), vec'[\psi_i(\theta_*)\psi_i'(\theta_*)], \psi_i(\theta_{**}))' < \infty$.

Let

$$\begin{aligned}
\mu_{**} &= E\psi_i(\theta_{**}),\\
\omega_* &= \Omega^{-1}(\theta_*),\\
j_* &= J'(\theta_*),\\
Y_{2n} &= R_{p,q}\{J^{(2)}(\theta_*)(\hat{\theta}-\theta_*)\} - J'(\theta_*)\Omega^{-1}(\theta_*)R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial\theta'}(\theta_*)(\hat{\theta}-\theta_*)\right).
\end{aligned}$$

Theorem 3.4 *If Assumptions 3, 5, 8 and Assumptions 11, 12, and 13 given in appendix hold, then*

$$\sqrt{n}\begin{pmatrix} \bar{\psi}(\theta_{**}) - \mu_{**} \\ [\bar{J}'(\theta_*) - J'(\theta_*)]\omega_*\mu_{**} \\ [\Omega_n(\theta_*) - \Omega(\theta_*)]\omega_*\mu_{**} \\ Y_{2T}\omega_*\mu_{**} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \Omega_{11}^s & \Omega_{12}^s & \Omega_{13}^s & \Omega_{14}^s \\ \Omega_{21}^s & \Omega_{22}^s & \Omega_{23}^s & \Omega_{24}^s \\ \Omega_{31}^s & \Omega_{32}^s & \Omega_{33}^s & \Omega_{34}^s \\ \Omega_{41}^s & \Omega_{42}^s & \Omega_{43}^s & \Omega_{44}^s \end{pmatrix}\right),$$

$$\sqrt{n}\bar{J}'(\hat{\theta})\Omega^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) \xrightarrow{d} \mathcal{N}(0,\Omega_*^s)$$

and

$$\sqrt{n}(\hat{\theta}^{s3s} - \theta_{**}) \xrightarrow{d} \mathcal{N}(0,\Sigma^s),$$

where $\Sigma^s = D_*^{s^{-1}}\Omega_*^s D_*^{s^{-1'}}$ *and*

$$\Omega_*^s = j_*\omega_*\Omega_{11}^s\omega_*j_*' + \Omega_{22}^s + j_*\omega_*\Omega_{33}^s\omega_*j_*' + \Omega_{44}^s + j_*\omega_*\Omega_{12}^s + \Omega_{21}^s\omega_*j_*' - j_*\omega_*(\Omega_{13}^s + \Omega_{31}^s)\omega_*j_*'$$
$$+ j_*\omega_*\Omega_{14}^s + \Omega_{41}^s\omega_*j_*' - j_*\omega_*\Omega_{32}^s - \Omega_{23}^s\omega_*j_*' + \Omega_{24}^s + \Omega_{42}^s - j_*\omega_*\Omega_{34}^s - \Omega_{43}^s\omega_*j_*'.$$

Theorems 3.3 and 3.4 show that both the three-step Euclidean and the shrunk three-step Euclidean likelihood estimators are $\sqrt{n}$-consistent and are asymptotically normally distributed in misspecified models. Note that these results contain analogue results for correctly specified models as special cases. In correctly specified models and for both estimators, $\mu_{**} = 0$ and $\theta_* = \theta_{**} = \theta_0$, where θ_0 is the true parameter value. In addition, $D_* = D_*^s$ and both estimators have the same asymptotic distribution as the efficient two-step GMM estimator. Because the asymptotic distributions they yield are also valid in well specified models, we claim that Theorems 3.3 and 3.4 provide model misspecification robust inference for the three-step and the shrunk three-step Euclidean likelihood estimators, respectively.

Furthermore, these results also show that these estimators have very interesting properties with respect to the alternative most useful moment condition-based estimators. In well specified models, they have the same higher order bias as the EL and ETEL estimators while in misspecified models, they stay $\sqrt{n}$-consistent for they pseudo true values as do the ET and ETEL estimators. Moreover, they are computationally more tractable than all of the estimators in the class of minimum discrepancy estimators.

4 Simulations

Throughout these Monte Carlo experiments, we compare seven alternatives moment condition based estimators. We consider the Euclidean empirical likelihood (EEL), the empirical likelihood (EL), the exponential tilting (ET), the exponentially tilted empirical likelihood (ETEL), the three-step Euclidean likelihood (3S) and the shrunk three-step Euclidean likelihood estimators. We specifically consider two variants of the shrunk three-step Euclidean likelihood estimator corresponding to two different shrinkage factors. The first one is obtained with $\epsilon_n(\theta) = \epsilon_n^0(\theta)$ as shrinkage factor. This estimator is referred to as s3S0. The second is obtained with $\epsilon_n(\theta) = \epsilon_n^1(\theta) = \sqrt{n}\epsilon_n^0(\theta)$, the shrinkage factor we base our asymptotic theory on in Section 3. We refer to the shrunk three-step Euclidean likelihood estimator yielded by this shrinkage factor as s3S.

We first compare the finite sample bias of these seven estimators. For this aim, we use the same Monte Carlo design as Schennach (2007) which is a slightly expanded version of the Monte Carlo design used by Hall and Horowitz (1996), Imbens, Spady and Johnson (1998) and Kitamura (2001). The moment conditions of our Monte Carlo design are the following

$$E\psi(x_i,\theta) = 0: \quad \psi(x_i,\theta) = [r(x_i,\theta) \quad r(x_i,\theta)x_{i1} \quad r(x_i,\theta)(x_{i3}-1) \quad \ldots \quad r(x_i,\theta)x_{iK}]',$$

where $r(x_i,\theta) = \exp(-.72 - (x_{i1} + x_{i2})\theta + 3x_{i2}) - 1$. The true parameter value $\theta_0 = 3.0$ and $(x_{i1}, x_{i2})' \sim \mathcal{N}(0, 0.16 I_2)$ and $x_{ik} \sim \chi_1^2$, for $k = 1, \ldots, K$.

Instead of solving the saddle point problem that yields the EEL estimator, we rather compute the continuously updated estimator proposed by Hansen, Heaton and Yaron (1996) which is known to be equal to the EEL estimator (see Newey and Smith (2004) and Antoine, Bonnal and Renault (2007)). We compute the EL, ET and ETEL estimators by solving the saddle point problems that provide them respectively and we obtain the 3S, s3S0 and s3S estimators by the three-step procedures as described in Section 2.

We conduct this experiment with $K = 4$ and $K = 10$, respectively by replicating 10,000 samples of size $n = 200$. The EEL estimator algorithm fails to converge in about 3% of the simulated samples and the 3S estimator computation procedure fails in only 10 samples in this experiment. The cases where the 3S estimator fails to converge are related to the negativity of some of the implied probabilities used for the Jacobian and the variance estimation. This confirm the lack of stability that one can suspect for the 3S estimator in finite samples. The samples in which at least one estimator's computation fails to converge have been replaced.

Table 4.1 reports the simulated bias of the estimators we consider. The 3S, s3S0 and s3S estimators behave, in terms of bias, rather like EL and ETEL confirming the result in the literature about the $O_P(n^{-1})$-equivalence of EL, ETEL and 3S estimators and also our result about the higher order equivalence between s3S and EL estimators. The EEL estimator appears to be the worst in terms of bias. One can also mention that the finite sample bias of all of the estimators increase with the number of moment conditions. The largest increasing occurs for EEL followed by ET. The other estimators that are higher order to EL have a lower increasing in their bias. The s3S even yields the lowest increase.

Our second Monte Carlo illustration is related to the $\sqrt{n}$-consistency results under misspecification. The Monte Carlo design we use is the same as the one used by Schennach (2007) to allow for direct comparison. The moment conditions are the following

$$E\psi(x_i, \theta) = 0: \quad \psi(x_i, \theta) = (x_i - \theta \quad (x_i - \theta)^2 - 1)'$$

where $x_i \sim \mathcal{N}(0, 1)$ for a correctly specified model *(Model C)* and $x_i \sim \mathcal{N}(0, (0.8)^2)$ for a misspecified model *(Model M)*.

In Model C, the true parameter value is $\theta_0 = 0$ and in Model M, the pseudo true value is $\theta_* = 0$ for all of the estimators we consider. We replicate 10,000 samples of size $n = 1,000$ and 2,000 samples of size $n = 5,000$ for both Model C and Model M.

Table 4.1: The simulated bias of the EEL, 3S, s3S0, s3S, EL, ETEL and ET estimators

	EEL	3S	s3S0	s3S	EL	ETEL	ET
$K = 4$	0.289	0.059	0.060	0.052	0.064	0.060	0.104
$K = 10$	0.526	0.096	0.093	0.045	0.137	0.109	0.238

Table 4.2: The simulated standard deviations of EEL, 3S, s3S0, s3S, EL, ETEL and ET estimators for Models C and M

		EEL	3S	s3S0	s3S	EL	ETEL	ET
$n = 1,000$								
	Model C	0.032	0.032	0.032	0.032	0.032	0.032	0.032
	Model M	0.028	0.032	0.032	0.032	0.055	0.038	0.031
$n = 5,000$								
	Model C	0.014	0.014	0.014	0.014	0.014	0.014	0.014
	Model M	0.012	0.014	0.014	0.014	0.060	0.018	0.014

Table 4.2 displays the simulated standard errors for all of the estimators. In the correctly specified model, the simulated standard errors are the same for all of the estimators. This confirms that the

estimators have the same asymptotic distribution as predicted by the theory. The cumulative distribution functions plotted by Figures 4.2 and 4.3 also confirm this theoretical result. For the misspecified model, the simulated standard errors of the 3S and the s3S estimators shrink by approximately $\sqrt{5}$ from $n = 1,000$ to $n = 5,000$ and confirm our theoretical prediction for these estimators in misspecified models. Even though we do not study the behaviour of the EEL estimator in misspecified models, our simulation results suggest that this estimator may stay $\sqrt{n}$-consistent in misspecified models. The same observation is valid for the s3S0 estimator. The results for EL, ET and ETEL estimators confirm the findings by Schennach (2007). While the simulated standard errors of ET and ETEL shrink by approximately $\sqrt{5}$, the simulated standard errors of the EL estimator seem not to shrink providing evidence against the $\sqrt{n}$-consistency of the EL estimator for an asymptotic distribution in the case of model misspecification.

Figure 4.1: Simulated cumulative distribution function of EL, ET, ETEL, s3S, EEL, 3S, s3S0 estimators with $K = 4$ and $K = 10$, $n = 200$

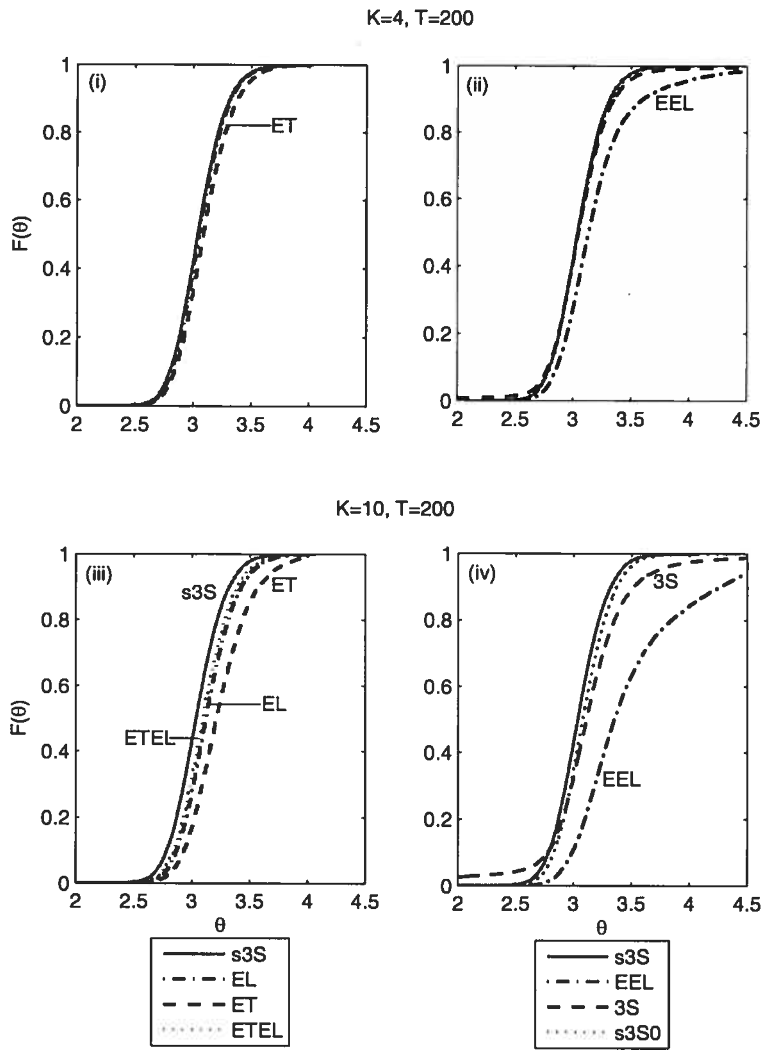


Figure 4.2: Simulated cumulative distribution function of EL, ET, ETEL, s3S, EEL, 3S, s3S0 estimators from Model C (i-ii) and Model M (iii-iv). $n = 1,000$

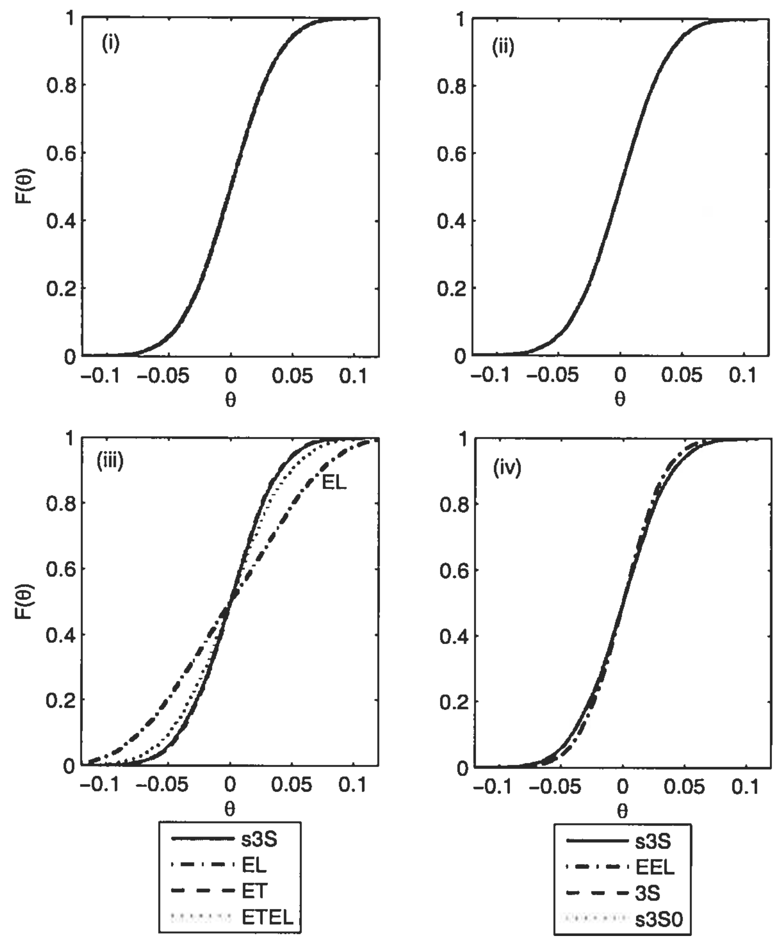
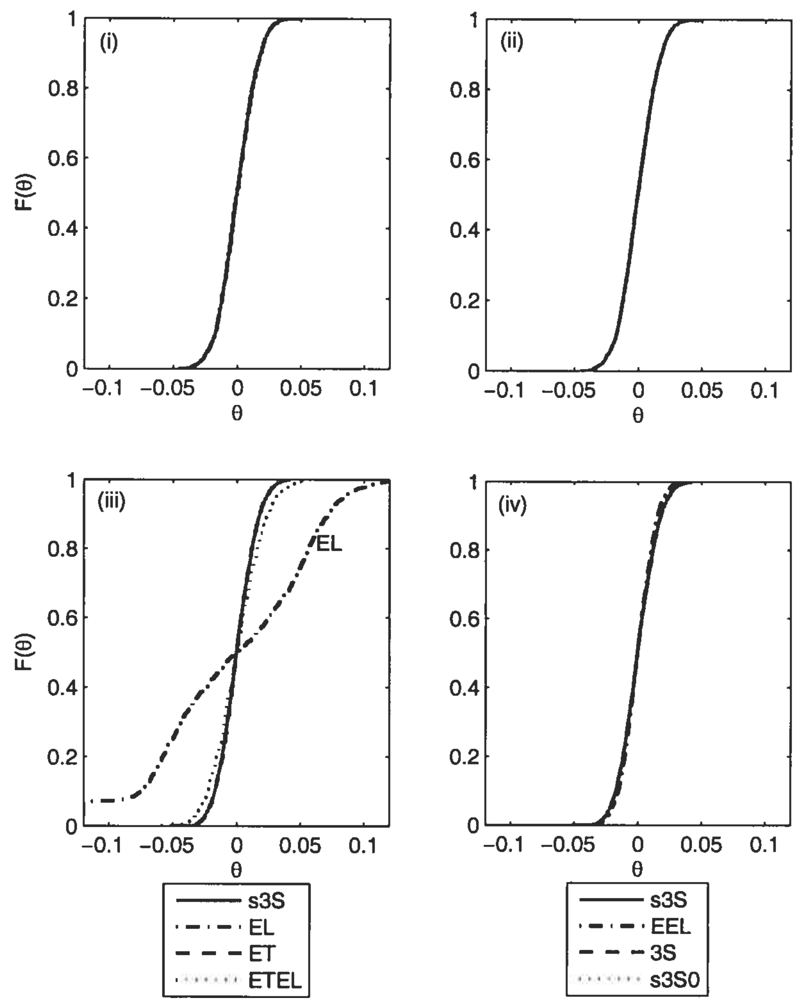


Figure 4.3: Simulated cumulative distribution function of EL, ET, ETEL, s3S, EEL, 3S, s3S0 estimators from Model C (i-ii) and Model M (iii-iv). $T = 5,000$



5 Conclusion

This paper explores some properties of the computationally appealing three-step Euclidean likelihood (3S) estimator and proposes the shrunk three-step Euclidean likelihood (s3S) estimator. In correctly specified models, as the 3S estimator, we show that the s3S estimator is equivalent to the EL estimator up to $O_P(n^{-1})$ thus they have the same higher order bias. We also provide a useful algorithm that yields more accurate (in terms of higher order bias) population means estimates when overidentifying moment conditions are available for the data generating process.

We also study the 3S and the s3S in misspecified models. As a result, we provide global model misspecification robust asymptotic distributions for these estimators. Moreover, this study reveals that even in misspecified models, these estimators stay $\sqrt{n}$-consistent and asymptotically normally distributed. These properties make both estimators two useful and particularly attractive alternatives to the EL estimator which is not $\sqrt{n}$-consistent in misspecified models and also to the ETEL estimator which is harder to compute.

By some Monte Carlo experiments, we evaluate the relative finite sample performance of these estimators. These experiments suggest that the s3S estimator can behave better than the 3S estimator particularly when not all of the implied probabilities are nonnegative. These experiments also validate the fact that both the 3S and the s3S estimators are $\sqrt{n}$-consistent while the EL estimator is not.

One possible development of this work that we plan for future research is to study the bias corrected version of the 3S and the s3S estimators. Because in correctly specified models they are higher order equivalent to the EL estimator, we can use the EL higher order bias derived by Newey and Smith (2004) to correct them. This could lead to interesting discussions about their higher order efficiency.

A Proofs of results in Section 2:

Assumption 9 *Let*

$$g_n(\theta) = \{\sum_{i=1}^{n} \pi_i(\hat{\theta})(\partial\psi_i'(\hat{\theta})/\partial\theta)\}\{\sum_{i=1}^{n} \pi_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta})\}^{-1}\bar{\psi}(\theta)$$

and $\mathcal{N}(\epsilon) = \{\theta : \|\theta - \theta_0\| < \epsilon\}$.
i) For some $\epsilon > 0$, g_n *has partial derivatives* $\mathcal{D}_n(\theta) = \partial g_n(\theta)/\partial\theta'$ *on* $\mathcal{N}(\epsilon)$ *such that, for all* $\delta > 0$,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \text{Prob}\{\sup_{\theta\in\mathcal{N}(\epsilon)} \|\mathcal{D}_n(\theta) - \mathcal{D}_n(\theta_0)\| > \delta\} = 0.$$

ii) There exists a measurable function $b(x)$ *such that, in a neighbourhood of* θ_0 *and for all* $k, l, r = 1, 2, \ldots, q$, $s = 1, 2, \ldots, p$, $|\psi_k(x,\theta)\psi_l(x,\theta)\psi_r(x,\theta)| < b(x)$, $|\partial\psi_k(x,\theta)/\partial\theta_r| < b(x)$ *and* $E\{b(x)^2\} < \infty$.

Assumption 9-i) is an asymptotic continuity condition on the gradient of g_n. This condition is required for the Theorem 1 by Robinson (1988) that we rely on. The point ii) of the same assumption is the usual dominance conditions for uniform convergence.

Lemma A.1 *Let* h *be a continuous function on a compact set* Θ *such that* $\forall\theta \in \Theta$, $h(\theta) = 0 \Leftrightarrow \theta = \theta_0$. *Let* h_n *be a sequel of functions defined on* Θ *and* $\hat{\theta}_n$ *be a sequel of values in* Θ *such that* $h_n(\hat{\theta}_n) = 0$ *a.s. If* $\sup_{\theta\in\Theta} \|h_n(\theta) - h(\theta)\| \xrightarrow{P} 0$, *then* $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Proof: Let $\mathcal{N}$ be a open neighborhood of θ_0 and $\mathcal{N}^c$ its complement. Since h is continuous on Θ, it is also continuous on $\Theta \cap \mathcal{N}^c$ which is compact. Let $\epsilon = \min_{\theta\in\Theta\cap\mathcal{N}^c} \|h(\theta)\|$. Since $\|h(.)\|$ is continuous on the compact set $\Theta \cap \mathcal{N}^c$, there exists $\theta^* \in \Theta \cap \mathcal{N}^c$ such that $\epsilon = \|h(\theta^*)\|$. Clearly, $\epsilon > 0$ since $\theta^* \neq \theta_0$. On the other hand, for the uniform convergence hypothesis, with probability approaching one, $\|h(\hat{\theta}_T)\| = \|h_T(\hat{\theta}_T) - h(\hat{\theta}_T)\| < \epsilon$. By definition of ϵ, $\hat{\theta}_T \notin \mathcal{N}^c$ and then $\hat{\theta}_T \in \mathcal{N}$ □

Proof of Theorem 2.1. We show that $\hat{\theta}^{3s} - \hat{\theta}^{s3s} = O_P(n^{-3/2})$ and we use the result by Antoine, Bonnal and Renault (2007), namely that $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$ to deduce that $\hat{\theta}^{s3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$. Our proof for $\hat{\theta}^{3s} - \hat{\theta}^{s3s} = O_P(n^{-3/2})$ relies on the result in Theorem 1 by Robinson (1988).

$$g_n(\theta) = \left[\sum_{i=1}^{n} \pi_i(\hat{\theta})J_i'(\hat{\theta})\right]\left[\sum_{i=1}^{n} \pi_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta})\right]^{-1}\bar{\psi}(\theta)$$

and

$$g_n^s(\theta) = \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta})J_i'(\hat{\theta})\right]\left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta})\right]^{-1}\bar{\psi}(\theta),$$

where $J_i(\theta) = \partial\psi_i(\theta)/\partial\theta'$. By definition, $g_n(\hat{\theta}^{3s}) = 0$ and $g_n^s(\hat{\theta}^{s3s}) = 0$.
From Theorem 4.1 by Antoine, Bonnal and Renault (2007), $\hat{\theta}^{3s} = \theta_0 + o_P(1)$. By the dominance condition in Assumption 9-ii), $\partial g_n(\theta_0)/\partial\theta' = \mathcal{D}_0 + o_P(1)$, where $\mathcal{D}_0$ is the nonsingular matrix $J^{0'}\Omega^{-1}(\theta_0)J^0$. To apply the Theorem 1 by Robinson (1988), we need to show that $\hat{\theta}^{s3s} = \theta_0 + o_P(1)$ before we can conclude, thanks to Assumption 9-i), that

$$\hat{\theta}^{3s} - \hat{\theta}^{s3s} = O_P(\|g_n(\hat{\theta}^{s3s}) - g_n^s(\hat{\theta}^{s3s})\|). \tag{16}$$

Let us show that $\hat{\theta}^{s3s} = \theta_0 + o_P(1)$. We use Lemma A.1.
Since $\hat{\theta}$ is the two-step GMM estimator, $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$ and by Antoine, Bonnal and Renault (2007) that $\text{Prob}(\min_{1\leq i\leq n} \pi_i(\hat{\theta}) \geq 0) \to 1$ as $n \to \infty$. Hence, for any $\epsilon > 0$, there exists $n_0 \geq 0$ such that for any $n \geq n_0$, $\text{Prob}(\min_{1\leq i\leq n} \pi_i(\hat{\theta}) \geq 0) \geq 1 - \epsilon$ i.e. $\text{Prob}(\epsilon_n^0(\hat{\theta}) = 0) \geq 1 - \epsilon$ i.e. $\text{Prob}(n\epsilon_n^0(\hat{\theta}) = 0) \geq 1 - \epsilon$. In other words, $\text{Prob}(n\epsilon_n(\hat{\theta}) = 0) \to 1$ as $n \to \infty$. Therefore,

$$\sqrt{n}\epsilon_n^1(\theta) \xrightarrow{P} 0. \tag{17}$$

On the other hand, for any $k = 1, 2, \ldots, q$ and $s = 1, 2, \ldots, p$,

$$\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) - \frac{1}{1+\epsilon_n^1(\hat{\theta})} \bar{\psi}'(\hat{\theta}) V_n^{-1}(\hat{\theta}) \frac{1}{n} \sum_{i=1}^{n} (\psi_i(\hat{\theta}) - \bar{\psi}(\hat{\theta})) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta})$$

and for any $k, l = 1, 2, \ldots, q$,

$$\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_{i,k}(\hat{\theta}) \psi_{i,l}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \psi_{i,k}(\hat{\theta}) \psi_{i,l}(\hat{\theta}) - \frac{1}{1+\epsilon_n^1(\hat{\theta})} \bar{\psi}'(\hat{\theta}) V_n^{-1}(\hat{\theta}) \frac{1}{n} \sum_{i=1}^{n} (\psi_i(\hat{\theta}) - \bar{\psi}(\hat{\theta})) \psi_{i,k}(\hat{\theta}) \psi_{i,l}(\hat{\theta}).$$

By our dominance conditions in Assumption 9-ii), and the result in Equation (17), $\bar{\psi}(\hat{\theta}) = O_P(n^{-1/2})$ and for any $k, l = 1, 2, \ldots, q$ and $s = 1, 2, \ldots, p$,

$$\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) = E \frac{\partial \psi_{i,k}}{\partial \theta_s}(\theta_0) + o_P(1) \quad \text{and} \quad \sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_{i,k}(\hat{\theta}) \psi_{i,l}(\hat{\theta}) = E \psi_{i,k}(\theta_0) \psi_{i,l}(\theta_0) + o_P(1).$$

Therefore,

$$Z_n \equiv \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) J_i'(\hat{\theta}) \right] \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) \right]^{-1} \xrightarrow{P} Z(\theta_0) \equiv J^{0'} \Omega^{-1}(\theta_0).$$

Let $g(\theta) = Z(\theta_0) E\psi_i(\theta)$. By the identification conditions by Assumption 1-vi), $g(\theta) = 0$ only at θ_0. $g_n(\theta) - g(\theta) = Z_n(\bar{\psi}(\theta) - E\psi_i(\theta)) + (Z_n - Z) E\psi_i(\theta)$. By the Cauchy-Schwarz inequality,

$$\|g_n(\theta) - g(\theta)\| \leq \|Z_n\| \sup_{\theta \in \Theta} \|\bar{\psi}(\theta) - E\psi_i(\theta)\| + \|Z_n - Z\| E \sup_{\theta \in \Theta} \|\psi_i(\theta)\|.$$

By Lemma 2.4 by Newey and McFadden (1994), $\sup_{\theta \in \Theta} \|\bar{\psi}(\theta) - E\psi_i(\theta)\| \xrightarrow{P} 0$. Because $Z_n - Z \xrightarrow{P} 0$, we deduce that $\sup_{\theta \in \Theta} \|g_n(\theta) - g(\theta)\| \xrightarrow{P} 0$. Lemma A.1 therefore, applies and $\hat{\theta}^{s3s} \xrightarrow{P} \theta_0$.

As a result, we can apply Theorem 1 by Robinson (1988) and the asymptotic stochastic order in (16) is valid. We have

$$\begin{aligned}
\hat{\theta}^{3s} - \hat{\theta}^{s3s} \leq & O_P \left\{ \left\| \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) J_i'(\hat{\theta}) \right] \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) \right]^{-1} \right.\right. \\
& \left.\left. - \left[\sum_{i=1}^{n} \pi_i(\hat{\theta}) J_i'(\hat{\theta}) \right] \left[\sum_{i=1}^{n} \pi_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) \right]^{-1} \right\| \|\bar{\psi}(\hat{\theta}^{s3s})\| \right\} \\
\leq & O_P \left\{ \left\| \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) J_i'(\hat{\theta}) - \sum_{i=1}^{n} \pi_i(\hat{\theta}) J_i'(\hat{\theta}) \right] \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) \right]^{-1} \right.\right. \\
& \left.\left. - \left[\sum_{i=1}^{n} \pi_i(\hat{\theta}) J_i'(\hat{\theta}) \right] \left[\left[\sum_{i=1}^{n} \pi_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) \right]^{-1} - \left[\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) \right]^{-1} \right] \right\| \|\bar{\psi}(\hat{\theta}^{s3s})\| \right\}.
\end{aligned}$$

Under our regularity assumptions, $\sum_{i=1}^{n} \pi_n(\hat{\theta}) J_i(\hat{\theta}) \xrightarrow{P} J^0$. Moreover, for any $k = 1, 2, \ldots, q$ and $s = 1, 2, \ldots, p$,

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) - \sum_{i=1}^{n} \pi_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) &= \frac{\epsilon_n^1(\hat{\theta})}{1+\epsilon_n^1(\hat{\theta})} \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) - \sum_{i=1}^{n} \pi_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) \right] \\
&= \frac{\epsilon_n^1(\hat{\theta})}{1+\epsilon_n^1(\hat{\theta})} \bar{\psi}'(\hat{\theta}) \left[-V_n^{-1}(\hat{\theta}) \frac{1}{n} \sum_{i=1}^{n} [\psi_i(\hat{\theta}) - \bar{\psi}(\hat{\theta})] \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) \right] \\
&= \epsilon_n^1(\hat{\theta}) \frac{1}{1+\epsilon_n^1(\hat{\theta})} \bar{\psi}'(\hat{\theta}) \left[-V_n^{-1}(\hat{\theta}) \frac{1}{n} \sum_{i=1}^{n} [\psi_i(\hat{\theta}) - \bar{\psi}(\hat{\theta})] \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) \right] \\
&= o_P(n^{-1/2}) O_P(1) O_P(n^{-1/2}) O_P(1) = o_P(n^{-1}).
\end{aligned}$$

Thus $\sum_{i=1}^n \tilde{\pi}_i(\hat{\theta})J_i'(\hat{\theta}) - \sum_{i=1}^n \pi_i(\hat{\theta})J_i'(\hat{\theta}) = o_P(n^{-1})$.

Similarly, with $M_n = \sum_{i=1}^n \tilde{\pi}_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta})$ and $N_n = \sum_{i=1}^n \pi_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta})$, $M_n - N_n = o_P(n^{-1})$. On the other hand, since $M_n^{-1} - N_n^{-1} = -M_n^{-1}(M_n - N_n)N_n^{-1}$, we deduce that $M_n^{-1} - N_n^{-1} = o_P(n^{-1})$. Furthermore, because $J^{0'}\Omega^{-1}(\theta_0)J^0$ is nonsingular and, by a mean value expansion, we can easily deduce that $\hat{\theta}^{s3s} - \theta_0 = O_P(n^{-1/2})$ and our dominance conditions insure that $\bar{\psi}(\hat{\theta}^{s3s}) = O_P(n^{-1/2})$. Therefore, $\hat{\theta}^{3s} - \hat{\theta}^{s3s} = O_P(n^{-3/2})$. Since, by Antoine, Bonnal and Renault (2007), $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$, we also have $\hat{\theta}^{s3s} - \hat{\theta}^{el} = O_P(n^{-3/2})$ □

Proof of Theorem 2.2. Let $\mathcal{L}_0$ be the Lagrangian associated to the problem:

$$\min_{\theta,\pi_1,\dots,\pi_n} \sum_{i=1}^n h(\pi_i) \quad \text{subject to} \quad \sum_{i=1}^n \pi_i = 1 \quad \sum_{i=1}^n \pi_i\psi_i(\theta) = 0$$

and $\hat{\theta}^h$, $\pi_1^h(\hat{\theta}^h), \dots, \pi_n^h(\hat{\theta}^h)$ be its solution. $\mathcal{L}_0 = \sum_{i=1}^n h(\pi_i) - \lambda(\sum_{i=1}^n \pi_i - 1) - \beta'\sum_{i=1}^n \psi_i(\theta)$. As the Lagrange Theorem's necessary conditions are fulfilled, the solution together with the Lagrange multipliers $\hat{\lambda}^h$ and $\hat{\beta}^h$ solves the first order conditions (in the corresponding arguments) given by:

$$\left\{\begin{array}{rl} h_\pi(\pi_i) - \lambda - \beta'\psi_i(\theta) = & 0 \quad \forall i = 1,\cdots,n. \\ \sum_{i=1}^n \pi_i = & 1 \\ \sum_{i=1}^n \pi_i\psi_i(\theta) = & 0 \\ \beta'\sum_{i=1}^n \pi_i(\partial\psi_i(\theta)/\partial\theta') = & 0 \end{array}\right. \tag{18}$$

Let us now consider the augmented problem

$$\min_{\theta,\eta,\pi_1,\dots,\pi_n} \sum_{i=1}^n h(\pi_i) \quad \text{subject to} \quad \sum_{i=1}^n \pi_i = 1 \quad \sum_{i=1}^n \pi_i\psi_i(\theta) = 0 \quad \sum_{i=1}^n \pi_i(g(x_i) - \eta) = 0$$

and let $\mathcal{L}_1$ be the associated Lagrangian and $(\hat{\theta}^{1h}, \pi_1^{1h}(\hat{\theta}^{1h}), \dots, \pi_n^{1h}(\hat{\theta}^{1h}))$ its solution. $\mathcal{L}_1 = \sum_{i=1}^n h(\pi_i) - \lambda(\sum_{i=1}^n \pi_i - 1) - \beta'\sum_{i=1}^n \psi_i(\theta) - \beta_1'\sum_{i=1}^n \pi_i\psi_n(\theta)$ and the first order conditions are:

$$\left\{\begin{array}{rl} h_\pi(\pi_i) - \lambda - \beta'\psi_i(\theta) - \beta_1'(g(x_i) - \eta) = & 0 \quad \forall i = 1,\cdots,n. \\ \sum_{i=1}^n \pi_i = & 1 \\ \sum_{i=1}^n \pi_i\psi_i(\theta) = & 0 \\ \beta'\sum_{i=1}^n \pi_i(\partial\psi_i(\theta)/\partial\theta') = & 0 \\ \sum_{i=1}^n \pi_i(g(x_i) - \eta) = & 0 \\ \beta_1'\sum_{i=1}^n \pi_i = & 0 \end{array}\right.$$

Note that the last two equations in this first order condition give $\hat{\beta}_1 = 0$ and $\hat{\eta} = \sum_{i=1}^n \pi_i^{1h}(\hat{\theta}^{1h})g(x_i)$ and the solution for θ, $\pi_1, \dots, \pi_n$, λ and β solves the first three equations. With $\beta_1 = 0$, these first three equations are equivalent to (18). Because the objective function in both optimization problems are the same and do not depend on β_1 nor η, we deduce that both the augmented and non-augmented optimization problems have the same solutions in the arguments they share, i.e. $\hat{\theta}^{1h} = \hat{\theta}^h$, $\pi_1^{1h}(\hat{\theta}^{1h}) = \pi_1^h(\hat{\theta}^h)$, ..., $\pi_n^{1h}(\hat{\theta}^{1h}) = \pi_n^h(\hat{\theta}^{1h})$. Since $\hat{\eta} = \sum_{i=1}^n \pi_i^{1h}(\hat{\theta}^{1h})g(x_i)$, we also have $\hat{\eta} = \sum_{i=1}^n \pi_i^h(\hat{\theta}^h)g(x_i)$ □

Proof of Theorem 2.3. Under Assumptions 1 and 2, by Newey and Smith (2004), $\hat{\theta}^{el} - \theta_0 = O_P(n^{-1/2})$. Since $\hat{\theta} - \hat{\theta}^{el} = O_P(n^{-3/2})$, $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$ and by a mean value expansion around θ_0, $\bar{\psi}(\hat{\theta}) = O_P(n^{-1/2})$ as well as $\bar{\psi}(\hat{\theta}^{el}) = O_P(n^{-1/2})$. On the other hand, for any $\theta \in \Theta$, the empirical likelihood implied probabilities are given by

$$\pi_i^{el}(\theta) = \frac{1}{n}\frac{1}{1 + \lambda_\theta'\psi_i(\theta)},$$

where λ_θ is a solution of the optimization program $\min_{\lambda\in\Lambda(\theta)}\{-\sum_{i=1}^n \log(1 + \lambda'\psi_i(\theta))/n\}$, $\widehat{\Lambda}(\theta) = \{\lambda \in \mathbb{R}^q : \lambda'\psi_i(\theta) > -1, \forall\theta \in \Theta\}$. By Lemma A2 by Newey and Smith (2004), $\hat{\lambda} \equiv \lambda_{\hat{\theta}} = O_P(n^{-1/2})$ and $\hat{\lambda}^{el} \equiv \lambda_{\hat{\theta}^{el}} = O_P(n^{-1/2})$. Moreover, since $E\sup_{\theta\in\Theta}\|\psi_i(\theta)\|^\alpha < \infty$, $\max_{1\le i\le n}\sup_{\theta\in\Theta}\|\psi_i(\theta)\| = O_P(n^{1/\alpha})$ and therefore,

$\|\bar{\lambda}\| \max_{1\le i\le n} \sup_{\theta\in\Theta} \|\psi_i(\theta)\| = o_P(1)$, for $\bar{\lambda} = \hat{\lambda}, \hat{\lambda}^{el}$. On the other hand,

$$\hat{\eta} - \hat{\eta}^{el} = \sum_{i=1}^{n}[\pi_i^{el}(\hat{\theta}) - \pi_i^{el}(\hat{\theta}^{el})]g(x_i) = \sum_{i=1}^{n}\left[\frac{1}{n(1+\hat{\lambda}'\psi_i(\hat{\theta}))} - \frac{1}{n(1+\hat{\lambda}^{el\prime}\psi_i(\hat{\theta}^{el}))}\right]g(x_i).$$

By some re-arrangement,

$$\hat{\eta} - \hat{\eta}^{el} = (\hat{\lambda}^{el} - \hat{\lambda})'\frac{1}{n}\sum_{i=1}^{n}\frac{\psi_i(\hat{\theta}^{el})}{[1+\hat{\lambda}'\psi_i(\hat{\theta})][1+\hat{\lambda}^{el\prime}\psi_i(\hat{\theta}^{el})]}g(x_i) - \hat{\lambda}'\frac{1}{n}\sum_{i=1}^{n}\frac{\psi_i(\hat{\theta}) - \psi_i(\hat{\theta}^{el})}{[1+\hat{\lambda}'\psi_i(\hat{\theta})][1+\hat{\lambda}^{el\prime}\psi_i(\hat{\theta}^{el})]}g(x_i)$$

$$= (\hat{\lambda}^{el} - \hat{\lambda})'\left(1 + O_P(\|\hat{\lambda}\| \max_{1\le i\le n}\sup_{\theta\in\Theta}\|\psi_i(\theta)\|)\right) \times \left((1 + O_P(\|\hat{\lambda}^{el}\| \max_{1\le i\le n}\sup_{\theta\in\Theta}\|\psi_i(\theta)\|)\right)\frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\theta}^{el})g(x_i)$$

$$- \hat{\lambda}'\left(1 + O_P(\|\hat{\lambda}\| \max_{1\le i\le n}\sup_{\theta\Theta}\|\psi_i(\theta)\|)\right) \times \left(1 + O_P(\|\hat{\lambda}^{el}\| \max_{1\le i\le n}\sup_{\theta\in\Theta}\|\psi_i(\theta)\|)\right)\frac{1}{n}\sum_{i=1}^{n}(\psi_i(\hat{\theta}) - \psi_i(\hat{\theta}^{el}))g(x_i).$$

As $\|\hat{\lambda}\| \max_{1\le i\le n}\sup_{\theta\in\Theta}\|\psi_i(\theta)\| = o_P(1)$ and $\|\hat{\lambda}^{el}\| \max_{1\le i\le n}\sup_{\theta\in\Theta}\|\psi_i(\theta)\|) = o_P(1)$,

$$\hat{\eta} - \hat{\eta}^{el} = (\hat{\lambda}^{el} - \hat{\lambda})'(1 + o_P(1))\frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\theta}^{el})g(x_i) - \hat{\lambda}'(1 + o_P(1))\frac{1}{n}\sum_{i=1}^{n}(\psi_i(\hat{\theta}) - \psi_i(\hat{\theta}^{el}))g(x_i). \tag{19}$$

Let $f(\theta, \lambda) = \sum_{i=1}^{n}\psi_i(\theta)/n(1+\lambda'\psi_i(\theta))$. By definition, $f(\hat{\theta}, \hat{\lambda}) = f(\hat{\theta}^{el}, \hat{\lambda}^{el}) = 0$.
By a mean value expansion,

$$f(\hat{\theta}^{el}, \hat{\lambda}^{el}) = f(\hat{\theta}, \hat{\lambda}) + (\partial f(\bar{\theta}, \bar{\lambda})/\partial\theta')(\hat{\theta}^{el} - \hat{\theta}) + (\partial f(\bar{\theta}, \bar{\lambda})/\partial\lambda')(\hat{\lambda}^{el} - \hat{\lambda}),$$

$\bar{\theta} \in (\hat{\theta}, \hat{\theta}^{el})$ and $\bar{\lambda} \in (\hat{\lambda}, \hat{\lambda}^{el})$. Thus,

$$(\partial f(\bar{\theta}, \bar{\lambda})/\partial\lambda')(\hat{\lambda}^{el} - \hat{\lambda}) = -(\partial f(\bar{\theta}, \bar{\lambda})/\partial\theta')(\hat{\theta}^{el} - \hat{\theta}).$$

Since $(\partial f(\bar{\theta}, \bar{\lambda})/\partial\lambda') \xrightarrow{P} \Omega(\theta_0)$ (see Newey and Smith (2004)), $(\partial f(\bar{\theta}, \bar{\lambda})/\partial\lambda')$ is nonsingular with probability approaching one. For large n, we can write

$$\hat{\lambda}^{el} - \hat{\lambda} = -[(\partial f(\bar{\theta}, \bar{\lambda})/\partial\lambda')]^{-1}(\partial f(\bar{\theta}, \bar{\lambda})/\partial\theta')(\hat{\theta}^{el} - \hat{\theta}) = O_P(n^{-3/2}).$$

By our dominance conditions in Assumption 2, $\sum_{i=1}^{n}\psi_i(\hat{\theta}^{el})g(x_i)/n$ is bounded in probability. Therefore, the first term in the RHS of (19) has $O_P(n^{-3/2})$ as order of magnitude as n grows.
On the other hand, by Taylor expansions, we have, with g_u denoting the u-th component of g,

$$\frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\theta})g_u(x_i) = \frac{1}{n}\sum_{i=1}^{n}\psi_i(\theta_0)g_u(x_i) + \frac{1}{n}\sum_{i=1}^{n}(\partial\psi_i(\theta_0)/\partial\theta')g_u(x_i)(\hat{\theta} - \theta_0) + O_P(\|\hat{\theta} - \theta_0\|^2)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\theta}^{el})g_u(x_i) = \frac{1}{n}\sum_{i=1}^{n}\psi_i(\theta_0)g_u(x_i) + \frac{1}{n}\sum_{i=1}^{n}(\partial\psi_i(\theta_0)/\partial\theta')g_u(x_i)(\hat{\theta}^{el} - \theta_0) + O_P(\|\hat{\theta}^{el} - \theta_0\|^2).$$

By subtracting these equations side by side, we have

$$\sum_{i=1}^{n}\psi_i(\hat{\theta})g_u(x_i)/n - \sum_{i=1}^{n}\psi_i(\hat{\theta}^{el})g_u(x_i)/n = \sum_{i=1}^{n}(\partial\psi_i(\theta_0)/\partial\theta')g_u(x_i)(\hat{\theta} - \hat{\theta}^{el})/n + O_P(n^{-1}).$$

By the law of large number, $\sum_{i=1}^{n}(\partial\psi_i/\partial\theta')(\theta_0)g_u(x_i)/n$ is bounded in probability thus, $\hat{\lambda}'(1+o_P(1))\sum_{i=1}^{n}[\psi_i(\hat{\theta}) - \psi_i(\hat{\theta}^{el})]g(x_i) = O_P(n^{-3/2})$. Therefore, the second term in the RHS of (19) also, has $O_P(n^{-3/2})$ as order of magnitude as n grows. Consequently, $\hat{\eta} - \hat{\eta}^{el} = O_P(n^{-3/2})$ □

B Regularity conditions for the GMM estimators $\tilde{\theta}$ and $\hat{\theta}$

The first step GMM estimator $\tilde{\theta}$ is defined as

$$\arg\min_{\theta\in\Theta} \bar{\psi}'(\theta)W^1\bar{\psi}(\theta).$$

The following assumption insures the consistency and asymptotic normality of $\tilde{\theta}$ in the case of global misspecification as formalized by Assumption 4-i).

Assumption 10 *i)* $\|\mu(\theta)\| > 0$ *for all* $\theta \in \Theta$.
ii) W^1 *is a symmetric positive definite matrix.*
iii) There exists $\theta_*^1 \in \Theta$ *such that* $Q_0^1(\theta_*^1) < Q_0^1(\theta)$ *for all* $\theta \in \Theta \setminus \{\theta_*^1\}$, *where* $Q_0^1(\theta) = E\psi_i'(\theta)W^1E\psi_i(\theta)$.
iv) $\theta_*^1 \in Int(\Theta)$.
v) $\psi(x,.)$ *is twice continuously differentiable on* $Int(\Theta)$ *and* $\partial\psi(.,\theta)/\partial\theta'$ *and* $(\partial/\partial\theta')vec[\partial\psi(.,\theta)/\partial\theta']$ *are measurable for each* $\theta \in Int(\Theta)$.
vi) There exists a measurable function $b_1(x)$ *such that* $|\psi_k(x,\theta)| < b_1(x)$, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b_1(x)$, $|\partial^2\psi_k(x,\theta)/\partial\theta_s\partial\theta_u| < b_1(x)$ *in a neighbourhood of* θ_*^1, *for all* $k = 1, 2, \dots, q$ *and* $s, u = 1, 2, \dots, p$ *and* $E\{b(x)^2\} < \infty$.
vii) $H_1(\theta_*^1) = J'(\theta_*^1)W^1J(\theta_*^1) - (E\psi_i'(\theta_*^1)W^1 \otimes I_p)J^{(2)}(\theta_*^1)$ *is nonsingular.*
viii) $Var z_{1,i} < \infty$, *where* $z_{1,i} = (\psi_i'(\theta_*^1), vec'\{\partial\psi_i(\theta_*^1)/\partial\theta'\})'$.

Assumptions 10-(i-iii) imply Assumption 4. Under Assumptions 3, 12 and 10-(i-iii), the result by Hall (2000) implies that $\tilde{\theta} \xrightarrow{P} \theta_*^1$. If Assumptions 3, 12 and 10 hold, $\sqrt{n}(\tilde{\theta} - \theta_*^1) \xrightarrow{d} \mathcal{N}(0, \omega_1)$. One can refer to Hall and Inoue (2003) for an explicit expression for ω_1. These conditions also imply that $\Omega_n(\tilde{\theta}) = \sum_{i=1}^n \psi_i(\tilde{\theta})\psi_i'(\tilde{\theta})/n$ is consistent for $E\psi_i(\theta_*^1)\psi_i'(\theta_*^1)$. We will explicitly assume, next that this probability limit is nonsingular. This additional assumption guarantees the two-step GMM estimator computation in large sample.

Assumption 11 *i) Assumption 10 holds.*
ii) $E\{\psi_i(\theta_*^1)\psi_i'(\theta_*^1)\}$ *is nonsingular.*
iii) There exists $\theta_* \in \Theta$ *such that* $Q_0(\theta_*) < Q_0(\theta)$ *for all* $\theta \in \Theta \setminus \{\theta_*\}$, *where* $Q_0(\theta) = E\psi_i'(\theta)W_nE\psi_i(\theta)$.
iv) $\theta_* \in Int(\Theta)$.
v) There exists a measurable function $b_2(x)$ *such that* $|\psi_k(x,\theta)| < b_2(x)$, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b_2(x)$, $|\partial^2\psi_k(x,\theta)/\partial\theta_s\partial\theta_u| < b_2(x)$ *in a neighbourhood of* θ_*, *for all* $k = 1, 2, \dots, q$ *and* $s, u = 1, 2, \dots, p$ *and* $E\{b_2(x)^2\} < \infty$.
vi) $H(\theta_*) = J'(\theta_*)WJ(\theta_*) - (E\psi_i'(\theta_*)W^1 \otimes I_p)J^{(2)}(\theta_*)$ *is nonsingular, where* $W = \{E\psi_i(\theta_*^1)\psi_i'(\theta_*^1)\}^{-1}$.
viii) $Var z_{2,i} < \infty$, *where* $z_{2,i} = (\psi_i'(\theta_*), vec'\psi_i(\theta_*^1)\psi_i'(\theta_*^1), vec'\{\partial\psi_i(\theta_*)/\partial\theta'\})'$.

Assumptions 11-(i-iii) imply Assumption 4. Under Assumptions 3, 12 and 11-(i-iii), the result by Hall (2000) implies that $\hat{\theta} \xrightarrow{P} \theta_*$. If Assumptions 3, 12, 11 hold, $\sqrt{n}(\hat{\theta} - \theta_*) \xrightarrow{d} \mathcal{N}(0, \omega_2)$. One can refer to Hall and Inoue (2003) for an explicit expression for ω_2.

C Proofs of results in Section 3:

Assumption 12 *i)* Θ *is compact.*
ii) $\psi(.,\theta)$ *is measurable for each* $\theta \in \Theta$ *and* $\psi_i(.)$ *is continuous with probability one on* Θ.
iii) $E\left[\sup_{\theta\in\Theta} \|\psi_i(\theta)\|\right] < \infty$.

Assumption 13 *i)* $\psi(x,.)$ *is differentiable with probability one on* Θ.
ii) There exists a measurable function $b(x)$ *such that, in a neighbourhood of* θ_* *and for all* $k, l, r = 1, 2, \dots, q$, $s = 1, 2, \dots, p$, $|\psi_k(x,\theta)\psi_l(x,\theta)\psi_r(x,\theta)| < b(x)$, $|\psi_l(x,\theta)(\partial\psi_k(x,\theta)/\partial\theta_s)| < b(x)$, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b(x)$ *and* $E\{b(x)\} < \infty$.

Proposition C.1 *Let* $J_i(\theta) = \partial\psi_i(\theta)/\partial\theta'$, $a_i(\theta) = \psi_i(\theta)\otimes(vec'\psi_i(\theta)\psi'(\theta), vec'J_i(\theta))'$, *and* $V(\theta) = Var(\psi_i(\theta))$. *If* $Var\, a_i < \infty$, $Var\, vecJ_i(\theta) < \infty$ *and* $V(\theta)$ *is nonsingular, then*

$$\begin{aligned}
&\text{plim}\bar{G}(\theta) \equiv G(\theta) = EJ_i'(\theta) - Cov\{\psi_i'(\theta)V^{-1}(\theta)E(\psi_i(\theta)), J_i'(\theta)\},\\
&\sqrt{n}(\bar{G}(\theta) - G(\theta)) = O_P(1),\\
&\text{plim}\bar{M}(\theta) \equiv M(\theta) = E\psi_i(\theta)\psi_i'(\theta) - Cov\{\psi_i'(\theta)V^{-1}(\theta)E(\psi_i(\theta)), \psi_i(\theta)\psi_i'(\theta)\},\\
&\sqrt{n}(\bar{M}(\theta) - M(\theta)) = O_P(1).
\end{aligned}$$

Proof of Proposition C.1. By the law of large numbers for independent and identically distributed random vectors, $V_n(\theta)$ is consistent for $V(\theta)$ and therefore with probability one as n goes to infinity, $V_n(\theta)$ is nonsingular. Because $a_i(\theta)$ is an i.i.d sequence with finite variance, the central limit theorem applies and $\sum_{i=1}^n\{a_i(\theta) - Ea_i(\theta)\}/\sqrt{n}$ is zero-mean asymptotically normally distributed.

$$\begin{aligned}
\bar{G}(\theta) &= \sum_{i=1}^n \pi_i(\theta)J_i'(\theta) = \frac{1}{n}\sum_{i=1}^n\{1 - (\psi_i(\theta) - \bar{\psi}(\theta))'V_n^{-1}(\theta)\bar{\psi}(\theta)\}J_i'(\theta)\\
&= (1 + \{\bar{\psi}'(\theta)V_n^{-1}(\theta)\bar{\psi}(\theta)\})\frac{1}{n}\sum_{i=1}^n J_i'(\theta) - \frac{1}{n}\sum_{i=1}^n\{\bar{\psi}'(\theta)V_n^{-1}(\theta)\psi_i(\theta)\}J_i'(\theta).
\end{aligned}$$

We have $\sum_{i=1}^n J_i'(\theta)/n = \sum_{i=1}^n[J_i'(\theta) - E\{J_i'(\theta)\}]/n + E\{J_i'(\theta)\}$.
By the central limit theorem, $\sum_{i=1}^n[J_i'(\theta) - E\{J_i'(\theta)\}]/n = O_P(n^{-1/2})$ therefore,

$$\begin{aligned}
(1 + \{\bar{\psi}'(\theta)V_n^{-1}(\theta)\bar{\psi}(\theta)\})\tfrac{1}{n}\textstyle\sum_{i=1}^n J_i'(\theta) = \;& (1 + \{E\psi'(\theta)V^{-1}(\theta)E\psi(\theta)\})E\{J_i'(\theta)\}\\
&+(1 + \{E\psi'(\theta)V^{-1}(\theta)E\psi(\theta)\})\tfrac{1}{n}\textstyle\sum_{i=1}^n[J_i'(\theta) - E\{J_i'(\theta)\}] + o_P(\tfrac{1}{\sqrt{n}}).
\end{aligned}$$

The (k,l)-component of $\sum_{i=1}^n\{\bar{\psi}'(\theta)V_n^{-1}(\theta)\psi_i(\theta)\}J_i'(\theta)/n$ is $\bar{\psi}'(\theta)V_n^{-1}(\theta)\sum_{i=1}^n\psi_i(\theta)J_{i,kl}'(\theta)/n$.
Let $\bar{\psi} \equiv \bar{\psi}(\theta)$, $\psi_i \equiv \psi_i(\theta)$, $V_n \equiv V_n(\theta)$ and $V^{-1} \equiv V^{-1}(\theta)$.

$$\bar{\psi}'V_n^{-1}\sum_{i=1}^n\psi_iJ_{i,kl}'(\theta)/n = E\psi_i'V^{-1}E\{\psi_iJ_{i,kl}'(\theta)\} + E\psi_i'V^{-1}\sum_{i=1}^n\{\psi_iJ_{i,kl}'(\theta) - E\psi_iJ_{i,kl}'(\theta)\}/n + o_P(\frac{1}{\sqrt{n}}).$$

Thus

$$\sqrt{n}(\bar{G}(\theta) - G(\theta)) = (1 + \{E\psi'(\theta)V^{-1}(\theta)E\psi(\theta)\})\sum_{i=1}^n[J_i'(\theta) - E\{J_i'(\theta)\}]/\sqrt{n} + B_n(\theta)/\sqrt{n} + o_P(1),$$

where $B_n(\theta)$ is a $p\times q$-matrix with its (k,l)-component given by $E\psi'(\theta)V^{-1}(\theta)\sum_{i=1}^n\{\psi_i(\theta)J_{i,kl}'(\theta) - E\psi_i(\theta)J_{i,kl}'(\theta)\}$. By the central limit theorem, each component of $\sqrt{n}(\bar{G}(\theta) - G(\theta))$ is asymptotically normally distributed. Therefore, $\sqrt{n}(\bar{G}(\theta) - G(\theta)) = O_P(1)$. Similarly, we also have $\sqrt{n}(\bar{M}(\theta) - M(\theta)) = O_P(1)$ □

Proof of Theorem 3.1. Under Assumption 4, the two-step GMM estimator $\hat{\theta}$ is consistent for θ_* and Assumptions 12-13 allow Lemma 4.3 by Newey and McFadden (1994) to apply and $\bar{G}(\hat{\theta}) \xrightarrow{P} G(\theta_*)$ and $\bar{M}(\hat{\theta}) \xrightarrow{P} M(\theta_*)$ so that $\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) \xrightarrow{P} G(\theta_*)M(\theta_*)^{-1}$.
For $\theta \in \Theta$, let $h_n(\theta) = \bar{K}(\hat{\theta})\bar{M}(\hat{\theta})^{-1}\bar{\psi}(\theta)$ and $h(\theta) = G(\theta_*)M^{-1}(\theta_*)E\psi_i(\theta)$. By definition, $h_n(\hat{\theta}^{3s}) = 0$ and, by Assumption 5 , $h(\theta) = 0 \Leftrightarrow \theta = \theta_{**}$, for $\theta \in \Theta$. To apply the consistency result by Lemma A.1, we establish the sufficient condition given by $\sup_{\theta\in\Theta}\|h_T(\theta) - h(\theta)\| \xrightarrow{P} 0$.

$$\begin{aligned}
\|h_T(\theta) - h(\theta)\| &= \left\|\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta) - G(\theta_*)M^{-1}(\theta_*)E\psi_i(\theta)\right\|\\
&= \Big\|\left(\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) - G(\theta_*)M^{-1}(\theta_*)\right)(\bar{\psi}(\theta) - E\psi_i(\theta))\\
&\quad + \left(\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) - G(\theta_*)M^{-1}(\theta_*)\right)E\psi_i(\theta) + G(\theta_*)M^{-1}(\theta_*)(\bar{\psi}(\theta) - E\psi_i(\theta))\Big\|\\
&\le \left\|\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) - G(\theta_*)M^{-1}(\theta_*)\right\|\left[\|\bar{\psi}(\theta) - E\psi_i(\theta)\| + \|E\psi_i(\theta)\|\right]\\
&\quad + \|G(\theta_*)\|\|M(\theta_*)\|^{-1}\|\bar{\psi}(\theta) - E\psi_i(\theta)\|.
\end{aligned}$$

Clearly, $\sup_{\theta\in\Theta}\|E\psi_i(\theta)\| \leq E\sup_{\theta\in\Theta}\|\psi_i(\theta)\| < \infty$ by Assumption 12. By the same assumption, Lemma 4.2 by Newey and McFadden (1994) applies and $\sup_{\theta\in\Theta}\|\bar{\psi}(\theta) - E\psi_i(\theta)\| \xrightarrow{P} 0$. Therefore, from Lemma A.1, we deduce that $\hat{\theta}^{3s} \xrightarrow{P} \theta_{**}$ □

Lemma C.1 *Let $x_i, i = 1,2,\ldots,n$ be an i.i.d random sample and let $y(x_i,\theta)$ be a measurable real valued function of x_i and θ, continuous with probability one at each $\theta \in \bar{\mathcal{N}}$, where $\bar{\mathcal{N}}$ is a compact subset of Θ. Let $\bar{\theta}$ be a random vector that lies in $\mathcal{N}$ with probability approaching one a n goes to infinity.*
If $\text{Prob}[\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta) \in (a,b)] \neq 0$ for any a and b on the real line such that $a \neq b$, then for any $M > 0$, $\text{Prob}\{\max_{1\leq i\leq n} y(x_i,\bar{\theta}) > M\} \to 1$ as $n \to \infty$.

Proof: Because $\bar{\theta} \in \bar{\mathcal{N}}$ with probability approaching one as n grows to infinity, for large n and for any $i = 1,\ldots,n$, $\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta) \leq y(x_i,\bar{\theta})$ with probability one. Therefore, with probability approaching one as n grows, $\max_{1\leq i\leq n}[\inf_{\theta\in\Theta} y(x_i,\theta)] \leq \max_{1\leq i\leq n} y(x_i,\bar{\theta})$. Then, for $M > 0$, $\text{Prob}\{\max_{1\leq i\leq n}[\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta)] > M\} \leq \text{Prob}\{\max_{1\leq i\leq n} y(x_i,\bar{\theta}) > M\}$. As $x_i, i = 1,\ldots,n$ are *i.i.d*, so are $\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta), i = 1,\ldots,n$ and hence,

$$\begin{aligned}\text{Prob}\left\{\max_{1\leq i\leq n}[\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta)] > M\right\} =& 1 - \text{Prob}\left\{\max_{1\leq i\leq n}[\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta)] \leq M\right\}\\ =& 1 - \text{Prob}\left\{\inf_{\theta\in\bar{\mathcal{N}}} y(x_i,\theta) \leq M; \forall i = 1,\ldots,n\right\}\\ =& 1 - \left\{Pr[\inf_{\theta\in\bar{\mathcal{N}}} y(x_1,\theta) \leq M]\right\}^n.\end{aligned}$$

Since $\text{Prob}\{\inf_{\theta\in\bar{\mathcal{N}}} y(x_1,\theta) \in (a,b)\} \neq 0$ for any $a \neq b$, $0 < \text{Prob}\{\inf_{\theta\in\bar{\mathcal{N}}} y(x_1,\theta) \leq M\} < 1$. Thus, $\lim_{n\to\infty}\{\text{Prob}\{\inf_{\theta\in\bar{\mathcal{N}}} y(x_1,\theta) \leq M\}\}^n = 0$. Thus, $\text{Prob}\{\max_{1\leq i\leq n} y(x_i,\theta) > M\} \to 1$ as $n \to \infty$ □

Lemma C.2 *Under Assumptions 3, 6 and 13, if the GMM estimator $\hat{\theta}$ is such that $\hat{\theta} - \theta_* = O_P(n^{-1/2})$, then $\text{Prob}\{\epsilon_n^0(\hat{\theta}) > M\} \to 1$, for all $M > 0$, where $\epsilon_n^0(\theta) = -n\min[\min_{1\leq i\leq n}\pi_i(\theta), 0]$.*

Proof: By definition, $\epsilon_n^0(\hat{\theta}) = \max\{\max_{1\leq i\leq n} -n\pi_i(\hat{\theta}); 0\}$. As a result, $\{\epsilon_n^0(\hat{\theta}) > M \geq 0\}$ is equivalent to $\{\max_{1\leq i\leq n}[-T\pi_t(\hat{\theta})] > M\}$ which, by the definition of $\pi_i(\theta)$ (see Equation (3)), is equivalent to

$$\left\{\max_{1\leq i\leq n}[\{(\psi_i(\hat{\theta}) - \bar{\psi}(\theta))'V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})\} - 1] > M\right\}.$$

On the other hand, since $\hat{\theta}$ is consistent for θ_*, by the dominance conditions in Assumption 13, $\bar{\psi}'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})$ converges in probability to a fixed scalar c. Then, to complete the proof, it is sufficient to show that:

$$\text{Prob}\left\{\max_{1\leq i\leq n}\psi_i'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) > M\right\} \to 1, \quad \forall M.$$

Moreover,

$$\psi_i'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) = \psi_i'(\hat{\theta})\left[V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E\psi_i(\theta_*)\right] + \psi_i'(\hat{\theta})V^{-1}(\theta_*)E\psi_i(\theta_*).$$

Because $\hat{\theta}$ is $\sqrt{n}$-consistent, by Assumption 13, $V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E\psi_i(\theta_*) = O_P(n^{-1/2})$. By Assumption 13, $E\sup_{\theta\in\bar{\mathcal{N}}_*}\|\psi_i(\theta)\|^2 < \infty$, where $\bar{\mathcal{N}}_*$ is a closed neighbourhood of θ_* included in Θ. By Lemma 4 in Owen (1990) and Lemma D.2. in Kitamura, Tripathi and Ahn (2004), (see also Lemma A.1 by Bonnal and Renault (CIRANO working paper 2004s-18) $\max_{1\leq i\leq n}\sup_{\theta\in\bar{\mathcal{N}}_*}\|\psi_i(\theta)\| = o_P(n^{1/2})$. Hence, for n large enough and by Cauchy-Schwartz inequality,

$$\begin{aligned}\left|\psi_i'(\hat{\theta})\left[V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E\psi_i(\theta_*)\right]\right| \leq& \frac{1}{\sqrt{n}}\max_{1\leq i\leq n}\sup_{\theta\in\bar{\mathcal{N}}_*}\|\psi_i(\theta)\|\\ &\times\sqrt{n}\|V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E\psi_i(\theta_*)\| = o_P(1).\end{aligned}$$

Then, $\psi_i'(\hat\theta)\left[V_n^{-1}(\hat\theta)\bar\psi(\hat\theta) - V^{-1}(\theta_*)E\psi_i(\theta_*)\right] = o_P(1)$ uniformly over $i = 1, \ldots, n$. For this, it suffices to show that $\text{Prob}\{\max_{1\le i\le n}\psi_i'(\hat\theta)V^{-1}(\theta_*)E\psi_i(\theta_*) > M\} \to 1$ as $n \to \infty$, for all M. By Assumption 6, we can apply Lemma C.1 with $y(x_i, \theta) = \psi_i'(\theta)V^{-1}(\theta_*)E\psi_i(\theta_*)$ and the result follows □

Proof of Theorem 3.2. Let $y(x_i, \theta) = \partial\psi_i'(\theta)/\partial\theta$ or $\psi_i(\theta)\psi_i'(\theta)$.

$$\begin{aligned}\sum_{t=1}^{T}\bar\pi_i(\hat\theta)y(x_i,\hat\theta) =& \sum_{i=1}^{n}\left[\frac{1}{1+\epsilon_n^1(\hat\theta)}\pi_i(\hat\theta)y(x_i,\hat\theta) + \frac{\epsilon_n^1(\hat\theta)}{1+\epsilon_n^1(\hat\theta)}\frac{1}{n}y(x_i,\hat\theta)\right]\\ =&\frac{1}{1+\epsilon_n^1(\hat\theta)}\sum_{i=1}^{n}\pi_i(\hat\theta)y(x_i,\hat\theta) + \frac{\epsilon_n^1(\hat\theta)}{1+\epsilon_n^1(\hat\theta)}\frac{1}{n}\sum_{i=1}^{n}y(x_i,\hat\theta)\\ =&\frac{1}{n}\sum_{i=1}^{n}y(x_i,\hat\theta) - \frac{1}{1+\epsilon_n^1(\hat\theta)}\frac{1}{n}\sum_{i=1}^{n}\{(\psi_i(\hat\theta)-\bar\psi(\hat\theta))'V_n^{-1}(\hat\theta)\bar\psi(\hat\theta)\}y(x_i,\hat\theta).\end{aligned}$$

By Lemma 4.3 by Newey and McFadden (1994), Under some regularity conditions,

$$\sum_{i=1}^{n}\{(\psi_i(\hat\theta)-\bar\psi(\hat\theta))'V_n^{-1}(\hat\theta)\bar\psi(\hat\theta)y(x_i,\hat\theta)/n = O_P(1).$$

Besides, Lemma C.2 insures that $\sqrt{n}/(1+\epsilon_n^1(\hat\theta)) = 1/(n^{-1/2}+\epsilon_n^0(\hat\theta)) = o_P(1)$ as $\epsilon_n^0(\hat\theta)$ diverges to infinity. Therefore, $\sum_{i=1}^n \bar\pi_i(\hat\theta)y(x_i,\hat\theta) = \sum_{i=1}^n y(x_i,\hat\theta)/n + o_P(n^{-1/2})$. Specifically,

$$\sum_{i=1}^{n}\bar\pi_i(\hat\theta)\frac{\partial\psi_i'}{\partial\theta}(\hat\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\psi_i'}{\partial\theta}(\hat\theta) + o_P(n^{-1/2}) \quad\text{and}\quad \sum_{i=1}^{n}\bar\pi_i(\hat\theta)\psi_i(\hat\theta)\psi_i'(\hat\theta) = \frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat\theta)\psi_i'(\hat\theta) + o_P(n^{-1/2}).$$

Because $\hat\theta$ is $\sqrt{n}$-consistent and by the regularity conditions in Assumption 13 we have

$$\sum_{i=1}^{n}\bar\pi_i(\hat\theta)\frac{\partial\psi_i'}{\partial\theta}(\hat\theta) = E\frac{\partial\psi_i'}{\partial\theta}(\theta_*) + O_P(n^{-1/2}) \quad\text{and}\quad \sum_{i=1}^{n}\bar\pi_i(\hat\theta)\psi_i(\hat\theta)\psi_i'(\hat\theta) = E\psi_i(\theta_*)\psi_i'(\theta_*) + O_P(n^{-1/2}).$$

Then,

$$Z_n(\hat\theta) \equiv \left[\sum_{i=1}^{n}\bar\pi_i(\hat\theta)\frac{\partial\psi_i'}{\partial\theta}(\hat\theta)\right]\left[\sum_{i=1}^{n}\bar\pi_i(\hat\theta)\psi_i(\hat\theta)\psi_i'(\hat\theta)\right]^{-1} \xrightarrow{P} Z(\theta_*) \equiv E\frac{\partial\psi_i'}{\partial\theta}(\theta_*)\left[E\psi_i(\theta_*)\psi_i'(\theta_*)\right]^{-1}.$$

Next, we show that $\hat\theta^{s3s} \xrightarrow{P} \theta_{**}$ using Lemma A.1. We need to show that $\sup_{\theta\in\Theta}\|h_n(\theta) - h(\theta)\| \xrightarrow{P} 0$ with $h_n(\theta) = Z_n(\hat\theta)\bar\psi(\theta)$ and $h(\theta) = Z(\theta_*)E\psi_i(\theta)$. Obviously, $\|h_n(\theta) - h(\theta)\| \le \|Z_n(\hat\theta)\|\|\bar\psi(\theta) - E\psi_i(\theta)\| + \|Z_n(\hat\theta) - Z(\theta_*)\|\|E\psi_i(\theta)\|$. By the same arguments as in the proof of Theorem 3.1 we can deduce that $\sup_{\theta\in\Theta}\|h_n(\theta) - h(\theta)\| \xrightarrow{P} 0$ and therefore, $\hat\theta^{s3s} \xrightarrow{P} \theta_{**}$ □

Proof of Theorem 3.3. We show that $\sqrt{n}\bar G(\hat\theta)\bar M^{-1}(\hat\theta)\bar\psi(\theta_{**})$ is asymptotically normally distributed. Clearly,

$$\begin{aligned}\bar G(\hat\theta)\bar M^{-1}(\hat\theta)\bar\psi(\theta_{**}) =&[\bar G(\hat\theta) - \bar G(\theta_*)]\bar M^{-1}(\hat\theta)\bar\psi(\theta_{**}) + \bar G(\theta_*)[\bar M^{-1}(\hat\theta) - \bar M^{-1}(\theta_*)]\bar\psi(\theta_{**})\\ &+ \bar G(\theta_*)\bar M^{-1}(\theta_*)[\bar\psi(\theta_{**}) - E\psi_i(\theta_{**})] + [\bar G(\theta_*) - G(\theta_*)]\bar M^{-1}(\theta_*)E\psi_i(\theta_{**})\\ &+ G(\theta_*)[\bar M^{-1}(\theta_*) - M^{-1}(\theta_*)]E\psi_i(\theta_{**}),\end{aligned}\tag{20}$$

where $G(\theta) = \text{plim}\bar G(\theta)$ and $M(\theta) = \text{plim}\bar M(\theta)$.

A Taylor expansion of $\bar G(\hat\theta)$ around θ_* yields

$$\bar G(\hat\theta) = \bar G(\theta_*) + R_{p,q}\{(\partial/\partial\theta')[vec\bar G(\theta_*)](\hat\theta - \theta_*)\} + O_P(n^{-1}).$$

Let $z_{3,i} = \{vec'J_i(\theta_*), [\psi_i(\theta_*)\otimes vecJ_i(\theta_*)]', [\psi_i(\theta_*)\otimes vec\psi_i(\theta_*), \psi_i'(\theta_*)]', \psi_i(\theta_{**})\}'$.

Since $Varz_{3,i} < \infty$, by Proposition C.1, $\bar{G}(\theta_*) - G(\theta_*) = O_P(n^{-1/2})$ and our dominance assumptions also guarantee that $(\partial/\partial\theta')[vec\bar{G}(\theta_*)] - (\partial/\partial\theta')[vecG(\theta_*)] = o_P(1)$ thus

$$\bar{G}(\hat{\theta}) = \bar{G}(\theta_*) + R_{p,q}\left(\frac{\partial vec[G]}{\partial\theta'}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_P(n^{-1}). \tag{21}$$

Similarly, $\bar{M}(\theta) - M(\theta) = O_P(n^{-1/2})$ and

$$\bar{M}(\hat{\theta}) = \bar{M}(\theta_*) + R_{q,q}\left(\frac{\partial vec[M]}{\partial\theta'}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_P(n^{-1}).$$

Furthermore,
$\bar{M}^{-1}(\theta_*) - M^{-1}(\theta_*) = -\bar{M}^{-1}(\theta_*)(\bar{M}(\theta_*) - M(\theta_*))M^{-1}(\theta_*) = -M^{-1}(\theta_*)(\bar{M}(\theta_*) - M(\theta_*))M^{-1}(\theta_*) + O_P(n^{-1})$
and similarly,

$$\bar{M}^{-1}(\hat{\theta}) - \bar{M}^{-1}(\theta_*) = -M^{-1}(\theta_*)R_{q,q}\left(\frac{\partial vec[M]}{\partial\theta'}(\theta_*)(\hat{\theta} - \theta_*)\right)M^{-1}(\theta_*) + O_P(n^{-1}). \tag{22}$$

Thus, by reporting (21) and (22) in (20), we have

$$\begin{aligned}\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) =& Y_{1n}m_*\mu_{**} + g_*m_*[\bar{\psi}(\theta_{**}) - E\psi_i(\theta_{**})] + [\bar{G}(\theta_*) - G(\theta_*)]m_*\mu_{**} \\ &- g_*m_*[\bar{M}(\theta_*) - M(\theta_*)]m_*\mu_{**} + O_P(n^{-1}).\end{aligned} \tag{23}$$

By Proposition C.1,

$$\sqrt{n}(\bar{G}(\theta_*) - G(\theta_*)) = (1 + \{\mu_*'V^{-1}(\theta_*)\mu_*\})\sum_{i=1}^{n}[J_i'(\theta_*) - E\{J_i'(\theta_*)\}]/\sqrt{n} + B_n^1(\theta_*)/\sqrt{n} + o_P(1)$$

and

$$\sqrt{n}(\bar{M}(\theta_*) - M(\theta_*)) = (1 + \{\mu_*'V^{-1}(\theta_*)\mu_*\})\sum_{i=1}^{n}[\psi_{*i}\psi_{*i}' - E\{\psi_{*i}\psi_{*i}'\}]/\sqrt{n} + B_n^2(\theta_*)/\sqrt{n} + o_P(1),$$

where $B_n^1(\theta_*)$ is a $p \times q$-matrix with its (k,l)-component given by

$$\mu_*'V^{-1}(\theta_*)\sum_{i=1}^{n}\{\psi_i(\theta_*)J_{i,kl}'(\theta) - E\psi_i(\theta_*)J_{i,kl}'(\theta)\}$$

and $B_n^2(\theta_*)$ is a $q \times q$-matrix with its (k,l)-component given by

$$\mu_*'V^{-1}(\theta_*)\sum_{i=1}^{n}\phi_{i*},$$

where $\phi_{i*} = \{\psi_{i*}\psi_{i*,k}\psi_{i*,l} - E\psi_i(\theta_*)\psi_{i*,k}\psi_{i*,l}\}$.
We can easily deduce that $\sqrt{n}\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ is asymptotically Gaussian by the central limit theorem. □

Proof of Theorem 3.4. We show that $\sqrt{n}\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ is asymptotically normally distributed. We have

$$\begin{aligned}\bar{J}'(\hat{\theta})\Omega^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) =& [\bar{J}'(\hat{\theta}) - \bar{J}'(\theta_*)]\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) + \bar{J}'(\theta_*)[\Omega_n^{-1}(\hat{\theta}) - \Omega_n^{-1}(\theta_*)]\bar{\psi}(\theta_{**}) \\ &+ \bar{J}'(\theta_*)\Omega_n^{-1}(\theta_*)[\bar{\psi}(\theta_{**}) - E\psi_i(\theta_{**})] + [\bar{J}'(\theta_*) - J'(\theta_*)]\Omega_n^{-1}(\theta_*)E\psi_i(\theta_{**}) \\ &+ J'(\theta_*)[\Omega_n^{-1}(\theta_*) - \Omega^{-1}(\theta_*)]E\psi_i(\theta_{**}).\end{aligned}$$

On the other hand, by a Taylor expansion and the fact the $\hat{\theta}$ is $\sqrt{n}$-consistent,

$$\bar{J}'(\hat{\theta}) = J'(\theta_*) + R_{p,q}\left(J^{(2)}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_P(n^{-1})$$

and

$$\Omega_n(\hat{\theta}) = \Omega_n(\theta_*) + R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial\theta'}(\theta_*)(\hat{\theta}-\theta_*)\right) + O_P(n^{-1}).$$

From the expansion of $\hat{\theta} - \theta_*$ given by Equation (15), we can write

$$\begin{aligned} \bar{J}'(\hat{\theta})\Omega^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) =& Y_{2n}\omega_*\mu_{**} + j_*\omega_*[\bar{\psi}(\theta_{**}) - E\psi_i(\theta_{**})] + [\bar{J}'(\theta_*) - J'(\theta_*)]\omega_*\mu_{**} \\ & - j_*\omega_*[\Omega_n(\theta_*) - \Omega(\theta_*)]\omega_*\mu_{**} + O_P(n^{-1}). \end{aligned} \tag{24}$$

Therefore, $\sqrt{n}\bar{J}'(\hat{\theta})\Omega^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ is asymptotically Gaussian by the central limit theorem. □

References

[1] Altonji, J. and L. Segal, 1996. "Small Sample Bias in GMM estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14, 353-366.

[2] Antoine, B., H. Bonnal and E. Renault, 2007. "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood," *Journal of Econometrics*, 138, 461-487.

[3] Back, K. and D. P. Brown, 1993. "Implied Probabilities in GMM estimators," *Econometrica*, 61, 971-975.

[4] Baggerly, K. A., 1998. "Empirical Likelihood as Goodness-of-fit Measure," *Biometrika*, 85, 535-547.

[5] Davidson, J., 1994. "Stochastic Limit Theory," *Oxford University Press*

[6] Donald, G. S., G. W. Imbens and W. K. Newey, 2002. "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55-93.

[7] Hall, A. R., 2000. "Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test," *Econometrica*, 68, 1517-1527.

[8] Hall, A. R. and A. Inoue, 2003. "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361-394.

[9] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 1029-1054.

[10] Hansen, L. P., P. Heaton and A. Yaron, 1996. "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 362-280.

[11] Imbens, G. W., 1997. "One-Step Estimators in Overidentified Generalized Method of Moments Estimator," *Review of Economic Studies*, 64, 359-383.

[12] Imbens, W. G. and R. H. Spady, 2002. "Confidence Intervals in Generalized Method of Moments Models," *Journal of Econometrics*, 107, 87-98.

[13] Imbens, W. G., R. H. Spady and P. Johnson, 1998. "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333-357.

[14] Kitamura, Y., 1997. "Empirical Likelihood Methods with Weakly Dependent Processes," *The Annals of Statistics*, 25, 2084-2102.

[15] Kitamura, Y. and M. Stutzer, 1997. "An Information-Theorethic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861-874.

[16] Kitamura, Y., G. Tripathi and H. Ahn, 2004. "Empirical Likelihood-based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667-1714.

[17] Mykland, P. A., 1995. "Dual Likelihood," *The Annals of Statistics*, 23, 396-421.

[18] Newey, K. W. and D. McFadden, 1994. "Large Sample Estimation and Hypothesis," *Handbook of Econometrics*, IV, Edited by R. F. Engle and D. L. McFadden, 2112-2245.

[19] Newey, K. W. and R. J. Smith, 2004. "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

[20] Newey, K. W. and K. D. West, 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

[21] Owen, A., 1990. "Empirical Likelihood Ratio Confident Regions," *The Annals of Statistics*, 18, 90-120.

[22] Owen, A., 2001. "Empirical Likelihood", *Chapman & Hall.*

[23] Qin, J. and J. Lawless, 1994. "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300-325.

[24] Ragusa, G., 2004. "Alternatives to GMM: Properties of Minimum Divergence Estimators," *working Paper,* University of California, San Diego.

[25] Ramalho, J. S. and R. J. Smith, 2002. "Generalized Empirical Likelihood Non-Nested Tests," *Journal of Econometrics*, 107, 99-125.

[26] Robinson, P. M., 1988. "The Stochastic Difference Between Econometric Statistics," *Econometrica*, 56, 531-548.

[27] Schennach, S. M., 2007. "Point Estimation with Exponentially Tilted Empirical Likelihood," *The Annals of Statistics*, 35, 634-672.

Conclusion générale

Cette thèse étudie les modèles de volatilité multivariée ainsi que les méthodes d'inférence fondées sur les conditions de moment.

Dans le premier essai, nous étendons le modèle à facteurs de volatilité stochastique de Doz et Renault (2006) à la prise en compte de l'effet de levier et de l'effet de skewness conditionnels reconnus présents dans les rendements. Nous proposons aussi des conditions de moment pour l'estimation de ce modèle par la méthode des moments généralisée (GMM). Nous appliquons ce modèle aux rendements journaliers excédentaires de 24 indices sectoriels du marché financier du Royaume Uni. La modélisation des effets de levier et de skewness augmente l'efficacité de l'estimateur des paramètres de volatilité. Les résultats suggèrent que la compatibilité avec les asymétries fait obtenir une persistance plus faible pour la volatilité et nous permettent aussi de documenter une relation entre l'effet de skewness dans les rendements et leur volatilité.

Le deuxième essai se rapporte aux tests de facteur hétéroscédastiques pour les processus multivariés de rendements. Spécifiquement, le test proposé par Engle et Kozicki (1993) est fondé sur les conditions de moment résultant de la représentation factorielle et est une application du test des restrictions suridentifiantes du GMM (Hansen (1982)). Cet essai montre que ces conditions de moment ne garantissent pas les hypothèses standard d'application de la théorie de test par GMM. Nous montrons en particulier que l'identification au premier ordre des paramètres n'est pas assurée. Nous proposons une théorie générale qui fournit la distribution asymptotique de la statistique du test de suridentification du GMM dans une situation où les paramètres qui ne sont pas identifiables au premier ordre le sont au deuxième ordre. Une application de cette nouvelle théorie nous permet en particulier de corriger le test de Engle et Kozicki (1993).

Dans le troisième essai, nous proposons des méthodes de bootstrap pour la matrice de covariance réalisée évalué sur les données de haute fréquence. Ces méthodes s'appliquent aussi aux fonctions de la matrice de covariance telles que la covariance réalisée, la corrélation réalisée et le coefficient de régression réalisé. Il est à noter que le coefficient de régression réalisé inclus des statistiques aussi pertinentes pour l'analyse financière que les bêtas introduits par la théorie du capital asset pricing model (CAPM) pour l'évaluation du risque systématique des titres financiers.

Les méthodes de bootstrap que nous introduisons se veulent être une alternative pour l'approximation asymptotique de Barndorff-Nielsen et Shephard (2004). Les expériences de Monte Carlo que nous effectuonssuggèrent que la méthode de bootstrap que nous proposons fonctionne mieux, particulièrement lorsque les données sont générées à une fréquence faible. Nous observons aussi à travers des développements d'Edgeworth que le boostrap i.i.d. ne conduit pas à des raffinements d'ordre supérieur pour le coefficient de régression. Ceci est contraire aux résultats de Freedman (1981) et Mammen (1993), qui ont montré que le bootstrap par couples est supérieur à la distribution asymptotique normale pour les modèles de régression de coupes transversales avec hétéroscédasticité dans l'erreur. La raison principale des différences obtenues réside dans la nature des scores servant à la normalisation de la statistique de bootstrap. Dans le domaine des données en coupe instantanée qui est celui de Freedman (1981) et Mammen (1993), aussi bien les scores de la régression originelle que ceux de la régression de bootstrap sont d'espérance nulle. Grâce à cela, les facteurs de normalisation tendent vers la même limite en probabilité. Ceci est crucial dans l'obtention du raffinement à l'ordre supérieur par le bootstrap. Tel n'est pas le cas dans la configuration des processus de diffusion. L'espérance du score dans la régression originelle est non nulle tandis que l'espérance du score dans la régression de bootstrap est nulle. Cette différence force des normalisations qui ne sont pas analogues et qui ne convergent pas non plus en probabilité vers la même limite. Ceci s'avère coûteux pour la performance à l'ordre supérieur du bootstrap i.i.d. et explique aussi les différences avec Freedman (1981) et Mammen (1993).

Le quatrième essai porte sur les développements récents des méthodes d'inférence basées sur les conditions de moment. Cet essai propose un algorithme relativement simple permettant d'obtenir des estimateurs de moyennes de population de faible biais en échantillon fini grâce aux conditions de moment suridentifiantes. Une deuxième contribution de cet essai est de dériver les distributions asymptotiques robustes à la mauvaise spécification des conditions de moment pour l'estimateur de vraisemblance euclidienne à trois étapes proposé par Antoine, Bonnal et Renault (2007). Nous considerons aussi une variante de cet estimateur, l'estimateur de vraisemblance empirique euclidienne à trois étapes corrigé ou "shrunk three-step Euclidian likelihood estimator", qui utilise des probabilités impliquées positives. Dans la littérature, seul l'estimateur de maximum de vraisemblance empirique via minimum d'entropie (exponentially tilted empirical likelihood estimator) (ETEL) proposé par Schennach (2007) produit des biais d'échantillon fini aussi faibles que ces deux estimateurs quand les conditions de moment sont bien spécifiées et converge à la vitesse usuelle vers une distribution normale en cas de mauvaise spécification des conditions de moment. Il convient toutefois de souligner que l'estimateur ETEL est relativement beaucoup plus difficile à calculer que les estimateurs 3S et s3S.