

A NEAR-INFRARED SPECTROSCOPIC DATABASE OF HIGH-REDSHIFT QUASARS

1.1 INTRODUCTION

With the exception of a handful of very nearby objects, the inner regions of AGN cannot be resolved. Spectroscopic data is therefore invaluable to all AGN-related science. The optical region includes a number of strong emission features, including the broad lines of $H\alpha\lambda 6563$ and $H\beta\lambda 4861$ and the narrow $[O\text{III}]\lambda\lambda 4960, 5008$ doublet. As we will see in Chapter 2, the low-ionisation Balmer lines are routinely used to derive black hole masses and quasar accretion rates. As the strongest narrow emission line, $[O\text{III}]$ is used to measure the systemic redshift, and to probe quasar-driven outflows on galactic scales (see Chapter ??).

Large optical surveys have provided spectra for hundreds of thousands of AGN and quasars. With its twelfth data release in 2016, the number of quasar spectra in the Sloan Digital Sky Survey (SDSS; York et al., 2000) catalogue alone reached almost 300,000. However, the rest-frame optical region is redshifted beyond the reach of optical spectrographs at redshifts $z \gtrsim 0.4$ and, at redshifts $z \sim 2$, near-infrared spectroscopy is required in order to access the rest-frame optical lines.

The number density of quasars in the Universe rises sharply as a function of redshift, and peaks at redshifts $2 \lesssim z \lesssim 4$. The star formation rate follows a similar evolutionary path. Therefore, understanding supermassive black hole accretion over cosmic time and quasar feedback critically depends on the availability of near-infrared spectra for high-redshift quasars. Spectroscopic observations are more challenging at infrared wavelengths than in the optical. The Earth's atmosphere is both bright and highly variable at infrared wavelengths. As a result, the number of high-redshift quasars with near-infrared spectra is limited. Previous investigations of the rest-frame optical spectra of quasars at redshifts $z \sim 2$ have typically used samples of a few dozen (e.g. Shen and Liu, 2012; Shen, 2016).

In this chapter I will describe how I have constructed a database containing 462 high-redshift quasars. In later chapters, I will describe how I have used this data to derive un-biased virial black hole mass estimates for quasars at redshifts $z \gtrsim 2$ (Chapter 2) and to study quasar-drive galaxy-wide outflows (Chapter ??). The unprecedented size and quality of this dataset make a number of other exciting investigations possible, some of which are described in Section ??.

*Other references?
Sulentic?*

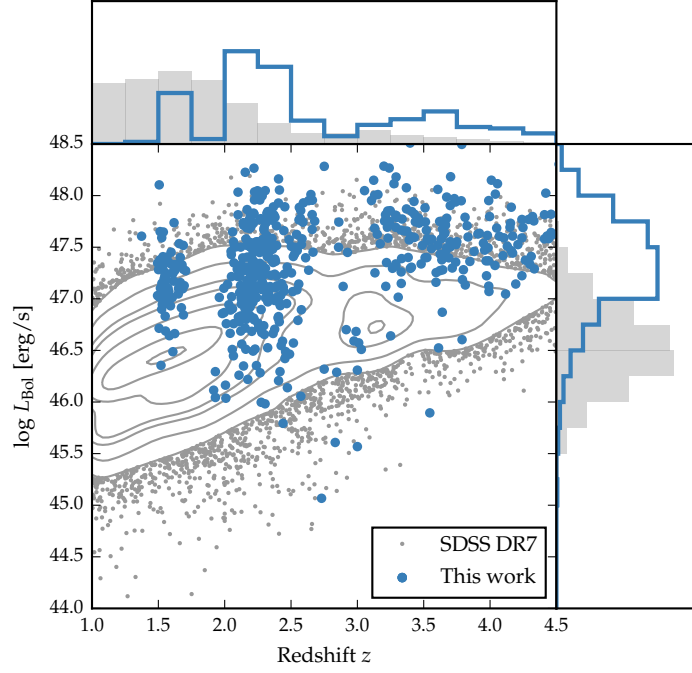


Figure 1.1: The ranges in redshift and luminosity covered by our sample, relative to the redshift-luminosity distribution of the SDSS DR7 quasar catalogue. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator. For the SDSS sample we use Hewett and Wild, (2010) redshifts and bolometric luminosities measured by Shen et al., (2011). For the quasars in our sample the redshift is defined using the peak of the $H\alpha/H\beta$ emission and the luminosity is measured in the continuum at 1350\AA and converted to a bolometric quantity using the same conversion factor employed by Shen et al., (2011). **Eight objects are missing because we do not have enough information to calculate the bolometric luminosity.**

In Fig. 1.1 we show the luminosities and redshifts of the quasar sample relative to the redshift-luminosity distribution for the Seventh Data Release (DR7; Schneider et al., 2010) of the SDSS spectroscopic quasar catalogue. Our sample spans a redshift range $1.5 < z < 4.0$ and a bolometric luminosity range $10^{45.5} - 10^{48} \text{ erg s}^{-1}$. Spectra were obtained within one or more of the JHK pass-bands and the gaps in our sample coverage at $z \sim 1.8$ and $z \sim 3$ are due to the presence of atmospheric absorption. Obtaining near-infrared spectra of adequate resolution and signal-to-noise ratio (S/N) of even moderately bright quasars remains resource intensive. As a consequence, at fixed redshift, the luminosities of the quasars are brighter than the average luminosity of the SDSS sample, although the dynamic range in luminosity is a full 1.5 decades.

Table 1.1: Summary of near-infrared spectroscopic database.

Instrument	Number
FIRE/Magellan	36
GNIRS/Gemini	29
ISAAC/VLT	13
LIRIS/WHT	21
NIRI/Gemini	31
NIRSPEC/Keck	3
SINFONI/VLT	84
SofI/NTT	111
TRIPLESPEC/ARC	38
TRIPLESPEC/Hale	60
XSHOOTER/VLT	36
Total	462

1.2 DATA

The near-infrared spectra in our database are taken from published catalogues, by downloading and reducing archival spectra, and by reducing spectra acquired in programmes led by Prof. J. Hennawi (UCSB) and Prof. X. Prochaska (UCO/LICK). As the P.I. of two programmes, I filled in an under-sampled region of the C iv blueshift parameter space by targeting quasars with the most extreme C iv blueshifts. The telescopes and instruments used to observe these spectra are summarised in Table 2.1 and the information on the individual spectra is provided in Table 1.1. In the remainder of this chapter I will describe each of these sub-samples in turn.

1.2.1 Coatman et al. (2016) Quasars

Defining sample

We selected quasars from the SDSS DR7 spectroscopic quasar catalogue with a wide range of C iv blueshifts. The sample was restricted to objects with redshifts $2.14 < z < 2.51$ (7,258 quasars), to ensure that the H β and H α emission lines fall within the H- and K-bands respectively, allowing us to observe both simultaneously with the appropriate grism configuration. Given the limited number of quasars for which near-infrared spectra could be obtained, the quasar sample was further restricted to objects that are radio-quiet (5,980 quasars), show no evidence of broad absorption lines (BALs) in their spectra (5,299 quasars), and are free from significant dust extinction. We removed radio-loud objects from our sample using the same radio-loud classification as Shen et al., (2011), and BAL quasars using the classifi-

*Does the reader
know about C iv
blueshifts?*

cations of both Shen et al., (2011) and Allen et al., (2011). The removal of quasars with significant dust extinction was achieved by identifying quasars with $i - K$ colours redder than a parametric spectral energy distributions (SED) model + SMC-like extinction curve with $E(B - V) = 0.05$ (the SED model is described in Chapter ??).

The K-magnitude was taken from the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007) Large Area Survey (ULAS). The requirement to be in the ULAS footprint and have reliable K band photometry reduced our sample of possible targets to 1,683, and the $E(B - V)$ cut left 1,204 in our sample. Finally, a flux-limit of $K < 18.5$ (AB) was applied to ensure that spectra of sufficient signal-to-noise ratio (S/N) could be obtained (412 quasars).

We were able to obtain new infra-red spectra for 19 quasars from this sample of 412 possible targets. The quasars included in this sub-sample were selected to have C IV-emission shapes which span the full range observed in the population. Reliably quantifying the distribution of C IV-emission shapes has been made possible thanks to redshift-determination algorithms (Hewett and Wild, 2010, Allen & Hewett 2016, in preparation) which are independent of the C IV-emission shape.

Observations

Near-infrared spectra were obtained with the Long-slit Intermediate Resolution Infrared Spectrograph (LIRIS; Manchado et al., 1998) mounted on the 4.2m William Herschel Telescope (WHT) at the Observatorio del Roque de los Muchachos (La Palma, Spain). Observations took place over four non-contiguous nights from 2015 March 31 to April 4. Approximately one night was lost due to poor weather and a further half-night was affected by poor transparency due to cloud. A one arcsecond slit-width was employed and the LIRIS H + K low-resolution grism was selected, which covers the spectral ranges $1.53\text{--}1.79\text{ }\mu\text{m}$ and $2.07\text{--}2.44\text{ }\mu\text{m}$ with a dispersion of $9.7\text{ }\text{\AA}/\text{pixel}$. The spatial scale of the instrument is $0.25\text{ arcsec}/\text{pixel}$. Observations were divided into 60 s sub-exposures and performed in an ABBA nodding pattern, with the object placed at two positions along the slit 12 arcsec apart. Bright A0-5V stars were observed at similar air-masses to the targets in order to provide both telluric absorption corrections and a flux calibration of the quasar spectra.

Data reduction

The raw LIRIS data frames incorporate a known ‘pixel shift’ which was first removed from all frames using the LIRIS data reduction package LIRISDR. Subsequent data reduction was undertaken with

standard IRAF¹ procedures. The flat-field images, which were taken at the beginning of each night via illumination of the dome, were averaged and normalised to remove any wavelength-dependent signature. Each individual two-dimensional spectrum was then flat-field corrected. Consecutive AB and BA pairs of two-dimensional spectra were subtracted to remove the sky background. All the subtracted AB/BA-pairs for a target were then averaged to give the final two-dimensional spectrum.

The size of the one-dimensional spectrum extraction windows, in the slit direction, varied from 6-10 pixels. To increase the S/N, optimal variance-weighted extraction with sigma clipping was employed. For the fainter objects in our sample we were unable to trace the spectrum across the dispersion axis reliably and the trace from a telluric standard-star observation, observed at a similar air mass and time, was used instead. The wavelength calibration, using argon and xenon lamp exposures, resulted in root mean square errors in the range 1.01–1.71 Å, with a mean of 1.47 Å. The telluric standard star observations were reduced using the same steps described above. The stellar continuum was divided out of the standard star spectrum, which was then divided into the quasar spectrum to remove telluric absorption features. The spectral type and magnitude of the standard star were used to flux calibrate the quasar spectrum both in a relative and absolute sense. Variable atmospheric conditions combined with the narrow slit width resulted in a significant level of uncertainty in the absolute flux calibration for the quasar observations. The use of the UKIDSS broadband magnitudes (H and K) to normalise the spectra results in a significantly improved calibration.²

1.2.2 *Shen & Liu (2012) and Shen (2016) Quasars*

Shen, (2016) and Shen and Liu, (2012) obtained near-infrared spectroscopy for a sample of 74 luminous, $1.5 < z < 3.5$ quasars selected from the SDSS DR7 quasar catalogue. Targets had to possess good optical spectra covering the C IV line and have redshifts $z \sim 1.5, 2.1$, and 3.3 to ensure that the H β -[O III] region was covered in one of the near-infrared JHK bands. Thirty-eight of the quasars were observed with TripleSpec (Wilson et al., 2004) on the Astrophysics Research Consortium (ARC) 3.5 m telescope, and 36 with the Folded-port InfraRed Echellette (FIRE; Simcoe et al., 2010) on the 6.5 m Magellan-Baade telescope. The reduction of the spectra is described in Shen, (2016) and Shen and Liu, (2012).

¹ IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

² The data reduction pipeline is available at github.com/liamcoatman/SpectraTools

1.2.3 Quasar Pairs

A large part of our catalogue was observed as part of an ongoing effort to identify quasar pairs at very close projected separations (Quasars Probing Quasars³ (QPQ); Hennawi et al., 2006a; Hennawi et al., 2010). The primary science driver of this work is to study the circum-galactic medium of the foreground quasars in absorption (Hennawi et al., 2006b). Very accurate systemic redshift measurements are a requirement and a large amount of effort has gone into obtaining near-infrared spectra which cover low-ionisation broad lines or features from the quasar narrow line region (Prochaska and Hennawi, 2009; Lau, Prochaska, and Hennawi, 2015; Hennawi et al., 2015). Twenty-nine quasars were observed with the Gemini Near-Infrared Spectrograph (GNIRS; Elias et al., 2006) on the 8.1 m Gemini North telescope, thirteen using the Infrared Spectrometer And Array Camera (ISAAC; Moorwood et al., 1998) on the European Southern Observatory (ESO) Very Large Telescope (VLT), thirty-one with the Near InfraRed Imager and Spectrometer (NIRI; Hodapp et al., 2003) also on Gemini North and thirty-six with XSHOOTER (Vernet et al., 2011), again, on the VLT.

The XSHOOTER spectra were reduced with a custom software package developed by George Becker (for details, see Lau, Prochaska, and Hennawi, 2015). The remaining data was processed with algorithms in the LowRedux⁴ package (see Prochaska and Hennawi, 2009).

1.2.4 VLT SINFONI Quasars

We performed a search of the ESO archive for high-redshift quasars observed with the SINFONI integral field spectrograph (Eisenhauer et al., 2003; Bonnet et al., 2004) at VLT/UT4. We found 79 quasars with redshifts $1.5 < z < 3.7$ which have H and/or K SINFONI spectroscopy, covering the H β and H α lines respectively. Seventy-two of the quasars are from a large programme led by L. Wisotzki (programme o83.B-0456(A)) to study the mass function and Eddington ratios of active BHs at redshifts $z \sim 2$ drawn from the Hamburg/ESO survey (Wisotzki et al., 2000). A further seven SINFONI spectra are from a programme led by J. D. Kurk (programme o90.B-0674(B)) to obtain reliable BH mass estimates from H α /H β for a sample of radio-loud/radio-quiet SDSS quasars.

The SINFONI spectra were reduced using the package EASYSINF⁵. The package, which is based on the ESO-SINFONI pipeline, is described in Williams et al., (2016).

³ www.ucolick.org/~xavier/QPQ/Quasars_Probing_Quasars

⁴ www.ucolick.org/~xavier/LowRedux

⁵ www.mrao.cam.ac.uk/~rw480/easysinf

1.2.5 ESO NTT SOFI Quasars

One quarter of the quasar catalogue derives from a large programme (programme 187.A-0645; PI: J. Hennawi) to combine near-infrared spectra from SOFI (Moorwood, Cuby, and Lidman, 1998) on the 3.6 m New Technology Telescope (NTT) with archival high-resolution optical spectra from the UV-Visual Echelle Spectrograph (UVES; Dekker et al., 2000) at VLT/UT2 and the High Resolution Echelle Spectrometer (HIRES; Vogt et al., 1994) at Keck to construct a legacy database of bright, high-redshift ($2 < z < 4$) quasars with both rest-frame optical spectra, covering the H β -[O III] complex, and high-resolution rest-frame ultraviolet spectra. The main science goal is to obtain precise systemic redshifts which are crucial for the study of absorption line systems. Observations were undertaken over 16 nights from September 2011 to March 2013. I reduced these spectra using a custom pipeline using algorithms in the LowRedux package and, in Coatman et al., (2017), published a subset of the data for the first time.

Over five nights from 2015 August 31 to September 4 we obtained near-infrared SOFI spectra for a further 26 quasars (programme 095.B-0644(A); PI: L. Coatman). These quasars were selected from the SDSS DR7 quasar catalogue using criteria very similar to those described above for the WHT sample. In particular, we selected quasars with large C IV blueshifts to improve the statistics in this region of the C IV emission-line parameter space. The spectra were reduced using the same LowRedux pipeline.

Expand section?

1.2.6 Hale TripleSpec Quasars

A further sixty quasars in our catalogue are bright SDSS quasars which were observed with the TRIPLESPEC spectrograph on the Palomar 200-inch Hale telescope (P200). The objects were observed with the same science goals as the SOFI NTT large programme. The spectra were reduced using a custom pipeline, again using algorithms in the LowRedux package.

Table 1.2: Quasars in our near-infrared spectroscopic database. Only the first 15 entries are shown. The full table (including 462 objects) is available online. Columns are as follows: (1) identifier, (2) date near-infrared spectra acquired, (3)-(4) coordinates, (5) instrument/telescope, (6) wavelength coverage, (7) velocity per pixel, (8) S/N per pixel, (9)-(11) redshifts from peak of [O III], H α , and H β - see Chapter ?? for details.

ID (1)	Date (2)	Ra (3)	Dec (4)	Instr. (5)	$\Delta\lambda$ [μm] (6)	Δv [km s^{-1}] (7)	S/N (8)	$z([\text{O III}])$ (9)	$z(\text{H}\alpha)$ (10)	$z(\text{H}\beta)$ (11)
J000039-001804	2015-09-02	+00h00m39.00s	-00d18m03.90s	Sofi/NTT	1.50-2.54	154.0	4.9		2.1412	2.1391
J000345-232353	2009-07-07	+00h03m45.00s	-23d23m53.40s	SINFONI/VLT	1.44-1.87	36.0	12.7	2.2657		2.2653
J000345-232353	2011-09-18	+00h03m45.00s	-23d23m53.40s	Sofi/NTT	1.48-1.83	63.0	36.0	2.2644		2.2776
J000451-084450	2013-07-12	+00h04m50.66s	-08d44m49.63s	XSHOOTER/VLT	0.31-2.28	15.0	10.3	3.0038		3.0052
J000451-084452	2013-08-08	+00h04m50.91s	-08d44m51.98s	XSHOOTER/VLT	0.31-2.28	15.0	5.4	2.9991		
J000500-003348	2015-09-01	+00h05m00.42s	-00d33m48.20s	Sofi/NTT	1.50-2.54	154.0	8.2		2.1842	2.1850
J000501+010221	2015-09-02	+00h05m00.53s	+01d02m20.80s	Sofi/NTT	1.50-2.54	154.0	6.8		2.1334	2.1317
J001016+001228	2015-09-04	+00h10m16.49s	+00d12m27.60s	Sofi/NTT	1.50-2.54	154.0	8.9		2.2855	2.2828
J001247+001239	2013-06-06	+00h12m47.12s	+00d12m39.49s	ISAAC/VLT	1.52-1.60	15.0	19.1			2.1618
J001708+813508	2012-08-04	+00h17m08.48s	+81d35m08.10s	TRIPLESPEC/Hale	0.94-2.80	39.0	36.5	3.3934		
J001919+010152	2015-09-04	+00h19m19.31s	+01d01m52.20s	Sofi/NTT	1.50-2.54	154.0	6.5	2.3120	2.3158	2.3154
J001955-091316	2004-11-26	+00h19m54.67s	-09d13m16.45s	GNIRS/Gemini	0.60-2.61	88.0	9.9		2.1207	2.1308
J002018-233654	2009-07-07	+00h20m18.41s	-23d36m53.80s	SINFONI/VLT	1.44-1.87	36.0	16.9	2.2975		2.2931
J002023-414639	2009-07-08	+00h20m23.38s	-41d46m38.90s	SINFONI/VLT	1.09-1.41	35.0	33.4	1.5733		1.5730
J002111-242247	2009-07-16	+00h21m10.90s	-24d22m47.20s	SINFONI/VLT	1.44-1.86	36.0	11.1	2.2622		2.2595

2.1 INTRODUCTION

Quantifying the growth-rate of massive BHs at $2 \lesssim z \lesssim 3$ would help significantly in understanding the role quasars play in galaxy evolution. Reliable estimates of BH masses are a prerequisite for such investigations.

Single-epoch virial BH mass estimates using $H\beta$ are possible up to redshifts $z \sim 0.7$, and the technique has been extended to redshifts $z \sim 1.9$ via the calibration of the broad $Mg\ II\lambda\lambda 2796,2803$ emission line (McLure and Jarvis, 2002; Onken and Kollmeier, 2008; Wang et al., 2009; Rafiee and Hall, 2011). At redshifts $z \gtrsim 2$, however, ground-based statistical studies of the quasar population generally have no access to the rest-frame optical and near-ultraviolet spectral regions. The $C\ IV\lambda\lambda 1548,1550$ emission doublet is both relatively strong in the majority of quasars and visible in modern optical spectra, such as those provided by SDSS, to redshifts exceeding $z \sim 5$. $C\ IV$ -derived BH masses have therefore become the standard (e.g. Vestergaard and Peterson, 2006; Park et al., 2013) for both individual quasars and in studies of quasar population demographics.

The luminosities of quasars at redshifts $z \gtrsim 2$ are much greater than the majority of AGN at lower redshifts for which reverberation mapping results are available. Therefore, the reliability of the existing calibration involving $C\ IV$ FWHM velocity measurements and ultraviolet luminosity is not established definitively when extrapolating to high-redshifts and luminosities. While some authors have found good agreement between BH mass-estimates based on $C\ IV$ and $H\beta$ (e.g. Vestergaard and Peterson, 2006; Assef et al., 2011; Tilton and Shull, 2013), others have questioned the consistency (e.g. Baskin and Laor, 2005; Trakhtenbrot and Netzer, 2012; Shen and Liu, 2012).

In contrast to a number of low-ionisation emission lines, such as $Mg\ II$, the $C\ IV$ emission has long been known to exhibit significant displacements to the blue (Gaskell, 1982) and more recent work (e.g. Sulentic, Marziani, and Dultzin-Hacyan, 2000; Richards et al., 2011) has established that the extent of ‘blueshifts’ in the $C\ IV$ emission correlates with a number of properties of quasar spectral energy distributions (SEDs). While the physical origin of the blueshifted emission has not been established there is a consensus that the associated gas is not tracing virial-induced velocities, that should allow a BH-mass estimate to be derived. A favoured interpretation associates the blueshifted emission with out-flowing material (see Netzer, 2015, for

a recent review), reaching velocities significantly larger than virial-induced velocities associated with the BH (e.g. Sulentic et al., 2007; Richards et al., 2011). These outflows, most likely, result from the presence of a radiation line-driven accretion-disc wind (e.g. Konigl and Kartje, 1994; Murray et al., 1995; Proga, Stone, and Kallman, 2000; Everett, 2005; Gallagher et al., 2015; Higginbottom and Proga, 2015).

In general, researchers studying quasar demographics at high-redshift adopt estimates of BH masses based on the width of C IV-emission, without reference to the blueshift of the C IV-emission (e.g. Vestergaard, 2004; Kollmeier et al., 2006; Gavignaud et al., 2008; Vestergaard et al., 2008; Vestergaard and Osmer, 2009; Kelly et al., 2010; Kelly and Shen, 2013). The systemic redshift is often assumed to be given by the peak of the C IV emission, regardless of whether there is evidence that the line is shifted or not. Figure 2.1 shows the shape of the C IV-emission in composite spectra constructed from SDSS DR7 quasars with $EW(C\text{ IV})=25\text{-}40\text{\AA}$, as a function of C IV blueshift. Quasars classified as BALs, or possessing strong associated absorbers have been excluded, and the composite-spectra shown are derived using an arithmetic mean of a minimum of 200 spectra at each blueshift. The profiles show how, at large values of blueshift ($\gtrsim 2000\text{ km s}^{-1}$) the C IV-profile is displaced to the blue by amounts comparable to the FWHM of the profile. Irrespective of the physical origin of the high-blueshift C IV-profiles, measures of the emission-line ‘width’ do not relate simply to virialized motions of the emitting gas under the gravitational influence of the BH. Excess emission-line flux in the blue wing of the C IV emission increases commonly employed measures of the line-width, notably the full-width at half maximum (FWHM) and the line dispersion (σ). As a consequence, BH-masses derived from C IV emission line velocity-widths are known to be systematically biased compared to masses from the Balmer lines (e.g. Shen et al., 2008; Shen and Liu, 2012).

Currently, the number of reverberation mapped quasars is both small (~ 50 quasars; Park et al., 2013) and, as highlighted by Richards et al., (2011), includes a restricted range of the C IV emission line shapes seen in the quasar population. In particular, the reverberation mapped objects generally possess high C IV equivalent widths and low C IV-blueshifts. Nevertheless, the derived scaling relations based on the reverberation-mapped sample are regularly applied to the quasar population with low C IV EWs and/or large C IV-blueshifts, where any non-virial outflow-related contribution to the dynamics is significant.

In recent literature, attempts have been made to minimise the influence of the systematic non-virial contribution to the C IV emission on estimates of the BH mass. Strategies include (i) significantly reducing the dependence of the derived masses on the emission-line velocity width (e.g. from the V^2 dependence predicted assuming a virialized

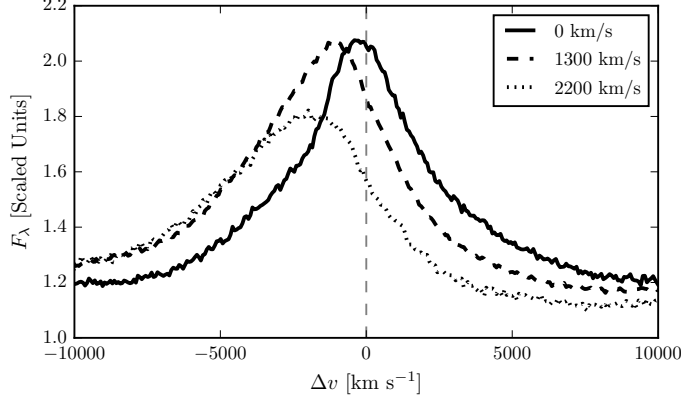


Figure 2.1: Composite spectra of the C IV-emission line as a function of C IV blueshift for SDSS DR7 quasars. Virtually the entire C IV-profile appears to shift blueward and the change in line shape is not simply an enhancement of flux in the blue wing of a still identifiable symmetric component. In order of increasing C IV blueshift, the composite spectra have FWHM 4870, 5610, and 6770 km s^{-1} and EW 33.1, 31.6, and 28.8 \AA .

broad line region to just $V^{0.56}$ in Park et al. 2013; see also Shen and Liu 2012), (ii) adopting a measure of emission-line velocity-width that is relatively insensitive to changes in the core of the emission-line profile (e.g. Denney et al., 2013) and (iii) estimating the amplitude of the non-virial contribution to the C IV emission-line via comparison with other ultraviolet emission lines (e.g. Si IV+O IV $\lambda 1400$ in Runnoe et al. 2013a and Brotherton et al. 2015). The increased number of high-quality spectra of quasars where information on both the Balmer lines in the rest-frame optical (see Chapter 1) and C IV in the ultraviolet is available enables a rather different approach. Specifically, to investigate whether, using the properties of the C IV emission line itself, it is possible to reduce, or even remove, the systematic bias in the BH-mass estimates.

2.2 QUASAR SAMPLE

The aim of the work presented in this chapter is to measure empirically the systematic bias in C IV-based virial BH mass estimates for high- z quasars as a function of the C IV emission-line blueshift. The basis for the C IV blueshift based correction is a large sample of quasars where it is possible to make a direct comparison of the C IV line-width with the line-width of the low-ionisation Balmer lines $H\alpha$ and $H\beta$, which are believed to provide reliable proxies for the virial velocity. Such an approach has not been possible hitherto as spectra that cover both the observed-frame optical (where the red-

Table 2.1: The numbers of quasars with reliable H α and H β line measurements, and the spectrographs and telescopes used to obtain the near-infrared spectra

Spectrograph	Telescope	H α Sample	H β Sample
FIRE	MAGELLAN	18	19
GNIRS	GEMINI-N	22	17
ISAAC	VLT	0	4
LIRIS	WHT	15	0
NIRI	GEMINI-N	0	12
SINFONI	VLT	2	25
SOFI	NTT	47	23
TRIPLESPEC	ARC-3.5m	33	20
TRIPLESPEC	P200	23	19
XSHOOTER	VLT	4	7
Total		164	144

shifted C iv appears) and near-infrared (where H β and H α lie) are required.

We have compiled a sample of 307 quasars at redshifts $1.5 < z < 4$ with both optical and near-infrared spectra to enable such a comparison to be performed. Reliable emission line properties were measured for 230 quasars (Section 2.3.5), with 164 possessing H α line measurements and 144 H β line measurements. This will allow us to directly compare virial BH mass estimates based on the C iv line-width with estimates based on the line-widths of the low-ionisation Balmer lines H α and H β . The sample is considerably larger than previous studies of the rest-frame optical spectra of high- z quasars (e.g. Shen and Liu, 2012). As we demonstrate in Section 2.5.3, the quasars have C iv blueshifts of up to $\sim 5000 \text{ km s}^{-1}$, and span the full range observed in the population. The near-infrared data has been described in Chapter 1 and the telescopes/spectrographs used in this chapter are summarised in Table 2.1. We have sub-divided our sample into two overlapping groups: quasars with reliable H α line measurements (the ‘H α sample’) and quasars with reliable H β measurements (the ‘H β sample’).

Optical SDSS DR7 spectra are employed for 70 quasars in the full catalogue. The SDSS DR7 catalogue contains moderate resolution $\sim 3800 - 9180 \text{ \AA}$ spectra for 105,783 quasars. Shen et al., (2011) have compiled a catalogue of properties for the SDSS DR7 quasars including, at $z > 1.5$, measurements of the broad C iv $\lambda\lambda 1548.2, 1550.8$ emission line. The SDSS DR7 spectra are moderate resolution ($R \simeq 2000$) and S/N ($S/N \simeq 20$) and cover the observed-frame wavelength interval $\sim 3800 - 9180 \text{ \AA}$. Many of the quasars in the SDSS DR7 catalogue have been re-observed as part of the Sloan Digital Sky Survey-III:

Baryon Oscillation Spectroscopic Survey (SDSS-III/BOSS; Dawson et al., 2013). As the BOSS-spectra typically have higher S/N than the SDSS DR7 spectra, we have used the BOSS spectra when available (126 quasars). We also use high-resolution optical spectra taken with VLT/UVES (11 quasars) and VLT/XSHOOTER (8 quasars), and Hamburg/ESO spectra for a further 15 quasars. The reduced and fluxed UVES spectra were made available to us by A. Dall’Aglia (a description of the reduction procedure is contained in Dall’Aglia, Wisotzki, and Worseck, (2008)). The spectral resolution of the UVES observations is very high ($R \sim 40\,000$) and the S/N of the spectra re-binned to a resolution of $\simeq 2000$ is $S/N \simeq 300$. The Hamburg/ESO optical spectra have a typical $\sim 400\text{km s}^{-1}$ spectral resolution and $S/N \gtrsim 10$ per pixel.

Near-infrared spectra, including the $H\alpha$ emission line, for a sample of 19 quasars, at redshifts $2.0 < z < 2.7$, have been obtained to complement existing SDSS optical spectra covering the C iv emission line. The 19 quasars were chosen to include a broad range of C iv line blueshifts. Our aim is to directly test the reliability of C iv-based BH mass estimates at high redshift for objects with a diverse range of C iv-line shapes. In particular, we will investigate potential systematic effects on the C iv-emission based BH masses for quasars with large, $\gtrsim 1200\text{km s}^{-1}$, C iv blueshifts, using the properties of the $H\alpha$ emission line to provide BH-mass estimates for the objects unbiased by non-virial contributions to the emission-line profile. Examining higher redshifts, our work complements other studies which attempt to improve the reliability of BH mass estimates which use the C iv line (e.g. Runnoe et al., 2013a; Denney, 2012). However, the range of C iv blueshifts in our sample is significantly more extended, which will allow us to study systematic biases in C iv-based virial BH masses more directly, i.e. as a function of the C iv blueshift. Established relations to derive BH masses from emission-line properties are employed but an advantage of our approach is that C iv and $H\alpha$ can be directly compared as a function of C iv-emission line shape. In Section ?? we describe how quasars were selected for rest-frame optical spectroscopy, before reviewing the C iv emission-line properties of high luminosity quasars in the SDSS surveys in Section 2.8. The new near-infrared spectroscopic observations are outlined in Section ?. With spectroscopic coverage of the C iv and $H\alpha$ emission lines available, the procedures used to quantify the line parameters are described in Section ?. The results of the spectral analysis in the context of BH masses and C iv-blueshift are presented in Section ?, with a discussion of the significance of the systematic trends identified included in Section 2.6. The paper concludes with a summary of the main conclusions in Section 2.7. Throughout this paper we adopt a Λ CDM cosmology with $h_0 = 0.71$, $\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. All wavelengths and equivalent width measurements are given in the quasar

rest-frame, and all emission line wavelengths are given as measured in vacuum.

2.3 SPECTRAL MEASUREMENTS

Conventionally, single-epoch virial estimates of the BH mass are a function of the line-of-sight velocity width of a broad emission line and the quasar luminosity. The velocity width is a proxy for the virial velocity in the broad line region (BLR) and, as revealed in reverberation-mapping studies, the luminosity is a proxy for the typical size of the BLR (the $R - L$ relation; e.g. Kaspi et al., 2000; Kaspi et al., 2007). Most reverberation mapping campaigns have employed $H\beta$ time-lags and velocity widths, but the line-widths of $H\alpha$ and $Mg\ II\lambda 2800$ have been shown to yield consistent BH masses (e.g. McLure and Jarvis, 2002; Greene and Ho, 2005; Onken and Kollmeier, 2008; Shen et al., 2008; Wang et al., 2009; Rafiee and Hall, 2011; Mejía-Restrepo et al., 2016). In Section 2.4.1 we verify that this is the case for the 99 quasars in our sample with measurements of both $H\alpha$ and $H\beta$ lines.

At redshifts $z > 2.2$, where the hydrogen Balmer lines and $Mg\ II$ are no longer accessible in many optical spectra, the $C\ IV\lambda 1550$ emission doublet has routinely been used to provide estimates of the virial velocity (e.g. Shen et al., 2011). As has long been recognised (Gaskell, 1982; Tytler and Fan, 1992) the $C\ IV$ emission line in many quasars includes contributions from gas that does not straightforwardly relate to virial motions within a stable BLR. A number of studies (e.g. Shen et al., 2008; Richards et al., 2011) have shown that the amplitude of the systematic shift of the $C\ IV$ emission to shorter wavelengths (relative to the systemic velocity) is strongly correlated with the properties of the emission-lines and the overall spectral energy distributions (SEDs).

In our work, a robust measure of the $C\ IV$ emission-line ‘blueshift’ provides the basis for the corrected $C\ IV$ velocity-width measurements, and hence BH masses. The effectiveness of the scheme is validated via a direct comparison of the $C\ IV$ velocity-widths to the Balmer emission velocity-widths in the same quasars. Our process is as follows. First, an accurate measure of the quasar’s systemic redshift is required, for which we adopt the centre of the Balmer emission, where the centre, λ_{half} , is the wavelength that bisects the cumulative total flux. Balmer emission centroids are available for all quasars in the catalogue but we verify that the measure is relatively unbiased through a comparison of the centroids to the wavelengths of the peak of the narrow $[O\ III]\lambda\lambda 4960, 5008$ doublet for the subset of spectra where both are available. Second, the blueshift of the $C\ IV$ emission line is determined. Again, we adopt the line centroid to provide a robust measure of the $C\ IV$ emission blueshift. The blueshift (in km s^{-1}) is

defined as $c \times (1549.48 - \lambda_{\text{half}}) / 1549.48$ where c is the velocity of light and 1549.48 \AA is the rest-frame wavelength for the C IV doublet, assuming equal contribution from both components. Positive blueshift values indicate an excess of emitting material moving towards the observer and hence out-flowing from the quasar.

Emission-line velocity widths are derived from the full-width-at-half-maximum (FWHM) of the lines but we also compute the line dispersion (calculated from the flux-weighted second moment of the velocity distribution) as some authors have claimed this provides a better estimate of the virial velocity (Denney et al., 2013).

To minimise the impact of the finite S/N of the quasar spectra and the presence of absorption features superposed on the broad emission lines we first fit a parametric model to the continuum and the emission lines. The purpose of the parametric fits is, however, simply to provide higher S/N representations of the emission features. The particular form of the model parametrizations is not important and the fits are used only to provide robust line parameters, such as the centroid λ_{half} , and FWHM, which are measured non-parametrically from the best-fitting model. The models used and the fitting procedure are described below. The issues involved in deriving parameters for broad emission lines from spectra of modest S/N – for example, subtraction of narrow line emission, subtraction of Fe II emission – have been covered comprehensively by other authors (e.g. Shen et al., 2011; Shen and Liu, 2012; Denney et al., 2013; Shen, 2016) and, as far as possible, we follow standard procedures described in the literature.

Virial BH mass estimators are calibrated using either the FWHM or dispersion (σ ; derived from the second-moment velocity) of a broad emission line (e.g. Vestergaard and Peterson, 2006; Park et al., 2013). Complications which are encountered when measuring line widths include how to model the ‘continuum’ flux, where to define the limits of the line emission, and how to deal with absorption. All of these issues are exacerbated when working with low S/N data (see Denney et al., 2013, for a discussion). In Section 2.8.2 we measured the blueshift of C IV for tens of thousands of SDSS DR7 quasar spectra. This allowed us to quantify the distribution of C IV blueshift values and hence select a subset for near-infrared observations which have C IV blueshifts spanning the full range of this distribution (Fig. 2.23). A non-parametric scheme was employed because, in comparison to recipes involving the fitting of multiple Gaussian (or other parametric) profiles, it was found to be more robust and less sensitive to the range of S/N among the spectra and to the presence of narrow absorption systems within the C IV-emission profile. In this section we will use a different approach, and measure the line properties by fitting a parametric model to the data. When working with a small number of spectra, it is possible to use careful mask-definition to minimise the effect of absorption features on the profiles used for the

parametrization. The purpose of the model fits is purely to best represent the intrinsic line profile, and no physical meaning is attached to the individual model components. We will now describe the parametric model and fitting procedure used for each emission line. The models were fit using a standard variance-weighted least squares minimisation procedure employing the Levenberg-Marquardt algorithm. Prior to the fit, the spectra were visually inspected and regions significantly affected by absorption were masked and excluded.

2.3.1 C IV

The parametrization of the C IV emission line is identical to the one described in Paper I. We first define a power-law continuum, $f(\lambda) \propto \lambda^{-\alpha}$, with the slope, α , determined using the median values of the flux in two continuum windows at 1445-1465 and 1700-1705 Å. The continuum emission is subtracted from the spectra, which is then transformed from wavelength units into units of velocity relative to the rest-frame line-transition wavelength for the C IV doublet (1549.98 Å, assuming equal contributions from both components). The parametric model is ordinarily fit within the wavelength interval 1500-1600 Å (corresponding to approximately $\pm 10\,000 \text{ km s}^{-1}$ from the rest-frame transition wavelength), a recipe that is commonly adopted (e.g. Denney et al., 2013). The line-window was extended if more than 5 per cent of the total flux in the profile was present blueward of the short wavelength limit. Narrow absorption features, which are frequently found superimposed on C IV emission (see, for example, the C IV profile of J0942+3523 in Fig. 2.4), were masked out during the fit.

The C IV emission was fit with sixth-order Gauss-Hermite (GH) polynomials, using the normalisation of van der Marel and Franx, (1993) and the functional forms of Cappellari et al., (2002). We allowed up to six components, but in many cases a lower order was sufficient (40 and 45 per cent were fit with second- and fourth-order GH polynomials respectively). GH polynomials were chosen because they are flexible enough to model the often very asymmetric C IV line profile. The flip-side of this flexibility, however, is that the model has a tendency to over-fit when spectra possess low S/N. The fits were therefore carefully checked visually and the number of components reduced if over-fitting was evident.

In Paper I we found that using the commonly employed three-Gaussian component model, rather than the GH polynomials, resulted in only marginal differences in the line parameters. Our best-fit parameters are also in good agreement with Shen et al., (2011), who employ a multi-Gaussian parametrization. In the Appendix we demonstrate that the mean difference between our FWHM measurements and the measurements of Shen et al., (2011) is just 200 km s^{-1} for the

quasars common to both samples, which is much too small to have any significant effect on our results.

We first measure and subtract the local continuum emission, by fitting a power-law to two windows on either side of the line emission, as described in Section 2.8.2. For a small number of objects, absorption features, or artefacts, in the spectrum necessitated modest adjustments to the window extents, which are specified in Table ???. The continuum-subtracted spectra are then transformed from wavelength units into units of velocity relative to the rest-frame line-transition wavelength for the C IV doublet (1549.48 Å, assuming equal contributions from both components). The parametric model is ordinarily fit within the same 1500–1600 Å window used in Section 2.8.2, which corresponds to approximately $\pm 10\,000\text{ km s}^{-1}$ from the rest-frame transition wavelength. The line-window was extended if significant flux in the profile was present blueward of the short wavelength limit. The adopted line-fitting windows, in units of velocity from the rest-frame transition wavelength, are given in Table ??.

To fit the C IV profile we employed Gauss-Hermite (GH) polynomials, using the normalisation of van der Marel and Franx, (1993) and the functional forms of Cappellari et al., (2002). We allowed up to six components in the GH polynomial model, but in many cases a lower order was sufficient; the polynomial order used for each line is given in Table ???. It is also a common practice to fit the C IV emission profile with two or three Gaussian components (e.g. Shen et al., 2011). We opted to use a GH-polynomial model primarily because it provided a significantly better fit to the most blueshifted and asymmetric C IV line (in SDSSJ0858+0152). Figure 2.2 shows how a model with three Gaussian components underestimates the flux in the blue wing and overestimates the flux in the red wing of the line profile. Using the Gaussian model rather than the GH polynomial changes the FWHM, line dispersion, and blueshift by -3, -3, and 10 per cent respectively. We have highlighted SDSSJ0858+0152 because, of all the objects in our sample, the choice of model leads to the largest change in C IV line parameters. Even in this case, however, the differences are modest.

For every other C IV line in our sample we found only marginal differences in our best-fit line parameters when, rather than using a GH polynomial model, the C IV emission was fit using a composite model of up to three Gaussians. Our best-fit parameters are also in good agreement with Shen et al., (2011), who employ a multi-Gaussian parametrization¹. The scatter between the Shen et al., (2011) results and our own is 0.1 dex about the one-to-one relation and, as expected, is larger for lines with smaller EWs.

¹ The Shen et al., (2011) parameters are derived from the SDSS DR7 spectra, whereas 16 out of 19 of our fits are to higher S/N BOSS DR12 spectra.

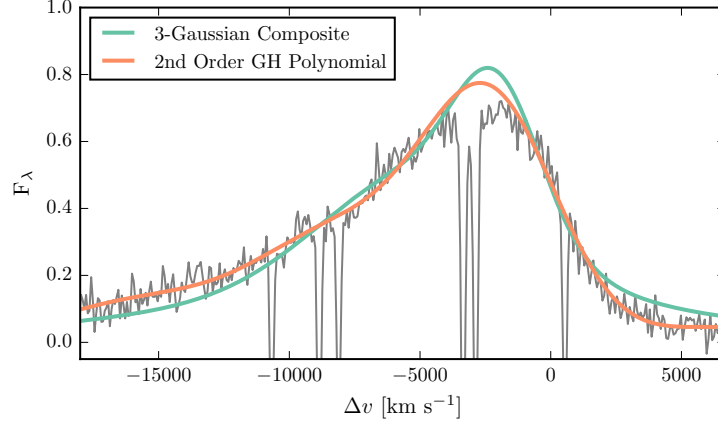


Figure 2.2: 2nd-order Gauss-Hermite (GH) polynomial and 3-component Gaussian fit to the C IV-emission line of SDSSJ0858+0152, which is the most blueshifted in our sample. We derive line parameters from the GH polynomial fit; using the Gaussian model changes the FWHM, line dispersion, and blueshift by -250, -150, and 500 km s^{-1} respectively. For the C IV lines of all other quasars in our sample the GH polynomial and Gaussian models provide equally good fits.

In this section we verify the accuracy of our parameter estimation scheme by comparing our FWHM measurements to measurements published in Shen et al., (2011), Shen and Liu, (2012) and Shen, (2016).

The parametric model we fit to the $\text{H}\beta$ /[O III] emission region was very similar to the model employed by Shen, (2016). In Fig. 2.3a we plot our $\text{H}\beta$ FWHM measurements against the measurements published in Shen, (2016), for 39 quasars in common to both samples. As expected, we observe a very tight correlation, with a scatter of 0.04 dex.

In Fig. 2.3b we plot our $\text{H}\alpha$ FWHM measurements against the measurements published in Shen and Liu, (2012), for 51 quasars in common to both samples. There is a strong correlation and, although the scatter is larger than for the $\text{H}\beta$ comparison (0.07 dex), no significant systematic bias.

Finally, in Fig. 2.3c we compare our measurements of the C IV FWHM from the 71 SDSS DR7 spectra in our sample with the measurements published in Shen et al., (2011). Shen et al., (2011) fit the C IV line profile with a composite model comprising up to three Gaussian components, whereas we used (up to order six) Gauss-Hermite polynomials. Nevertheless, there is a very strong agreement between our measurements, with a scatter of 0.05 dex.

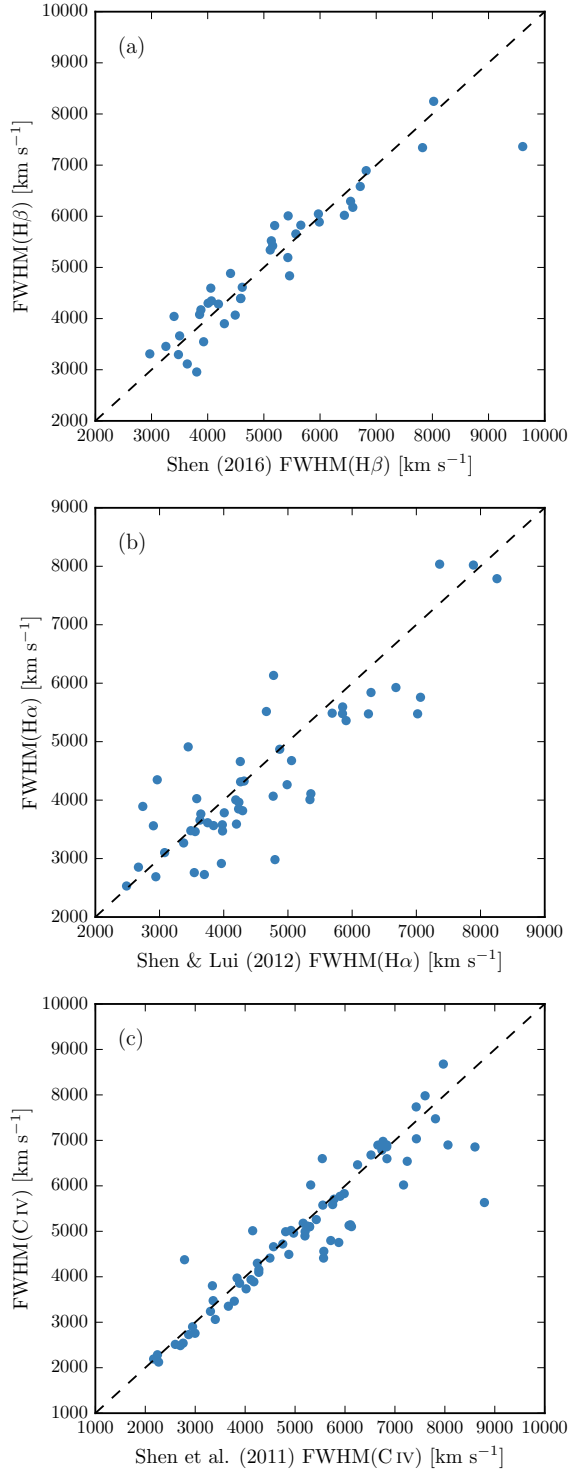


Figure 2.3: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of (a) the $\text{H}\beta$ FWHM with Shen, (2016), (b) the $\text{H}\beta$ FWHM with Shen and Liu, (2012), and (c) the C IV FWHM with Shen et al., (2011).

2.3.2 $H\alpha$

A power-law continuum is fit using two continuum windows at 6000-6250 and 6800-7000 Å. The continuum-subtracted flux is then fit in the wavelength interval 6400-6800 Å. We adopt a rest-frame transition wavelength of 6564.89 Å to transform wavelengths into equivalent Doppler velocities. The broad component of $H\alpha$ is fit using one or two Gaussians, constrained to have a minimum FWHM of 1200km s^{-1} . When two Gaussians are used, the velocity centroids are constrained to be the same.

The emission-line profiles of both $H\beta$ and $H\alpha$ frequently include a significant narrow component from the physically more extended narrow line region (NLR). Additional Gaussian components were included in our parametric model to fit the narrow component of $H\alpha$ as well as $[\text{N II}]\lambda\lambda 6548, 6584$ and $[\text{S II}]\lambda\lambda 6717, 6731$. This resulted in a better fit to the observed flux in 50 per cent of cases. We impose a 1200km s^{-1} upper limit on the FWHM of all narrow lines and the amplitudes of all components must be non-negative. The relative flux ratio of the two $[\text{N II}]$ components is also fixed at the expected value of 2.96. In 70 per cent of the spectra the $[\text{O III}]\lambda\lambda 4960, 5008$ doublet is detected at moderate S/N in the $H\beta$ region. In these cases the peak of the $[\text{O III}]$ is used to fix the velocity offsets and the FWHMs of the narrow line components in the $H\alpha$ region. For spectra where the $[\text{O III}]$ doublet does not constrain the velocity and FWHM accurately, the narrow emission in the $H\alpha$ and $H\beta$ regions are fitted independently but, for each region, the individual narrow-line velocity offsets and the FWHMs are constrained to be identical. In these objects the narrow line contribution is generally weak, and so does not have a large effect on the line parameters we measure for the broad component.

The model described above is very similar to the one described in Shen, (2016), Shen and Liu, (2012) and Shen et al., (2011), the only major difference being that we do not fit the $H\alpha$ and $H\beta$ emission regions simultaneously. In Appendix ??, we compare our $H\alpha$ line measurements for the subset of our sample taken from Shen, (2016) and Shen and Liu, (2012), and find a scatter of just 0.07 dex.

We employ the same continuum subtraction and fitting method as for C iv, with the continuum and fitting windows as given in Table ?. We adopt a rest-frame transition wavelength of 6564.89 Å to transform wavelengths into equivalent Doppler velocities. We used a simple model with up to three broad Gaussian components to fit the $H\alpha$ emission line. We opted against parametrizing the $H\alpha$ line using a GH polynomial because the extra degrees of freedom in this model did not improve the quality of the fits². Upon inspection of the residuals from the fit, we also found no evidence that additional model com-

² The emission line parameters and subsequent analysis do not depend on whether line parameters from multiple-Gaussian or GH-polynomial model fits are used.

ponents for narrow $H\alpha$, $N\text{II}\lambda\lambda 6548, 6584$ and $S\text{II}\lambda\lambda 6717, 6731$ were required. Furthermore, narrow $O\text{III}\lambda\lambda 4960, 5008$ emission is relatively weak in these spectra.

The sole exception is the $H\alpha$ line in the spectrum of SDSSJ0738+2710. In addition to having the narrowest $H\alpha$ line, this spectra also has the strongest narrow $O\text{III}$ component ($EW = 63 \text{ \AA}$), which suggests that a contribution from the narrow-line region might be important. Introducing a single Gaussian for the narrow emission, while retaining a double Gaussian for the broad emission, the FWHM of the broad component increases to 3400 km s^{-1} (compared to 1580 km s^{-1} without the narrow component). For consistency, the parameters quoted in Table ?? are from the model with no narrow component. However, because the properties derived from the emission line width (the BH mass and the mass-normalised accretion rate) are strongly biased by the probable contribution from the narrow-line region, SDSSJ0738+2710 is excluded from the analysis in Section ??.

2.3.3 $H\beta$ and $[O\text{III}]$

Emission from optical Fe II is generally strong in the vicinity of $H\beta$. We therefore fit a combination of a power-law continuum and an optical Fe II template – taken from Boroson and Green, (1992) – to two windows at 4435-4700 and 5100-5535 \AA . The Fe II template is convolved with a Gaussian, and the width of this Gaussian, along with the normalisation and velocity offset of the Fe II template, are free variables in the pseudo-continuum fit. We use the same model to fit the broad and narrow components of $H\beta$ as was used with $H\alpha$. Each line in the $[O\text{III}]$ doublet is fit with two Gaussians, to model both the systemic and any outflow contributions. The peak flux ratio of the $[O\text{III}]$ 4960 \AA and 5008 \AA lines is fixed at 1:3. As for the fit to the narrow lines in the spectral region around $H\alpha$, the width and velocity offsets of all the narrow components are set to be equal, and an upper limit of 1200 km s^{-1} is placed on the FWHM.

Open science. Provide: Fitting region Continuum region Model Redshift used in fit Number of Gaussians / order

2.3.4 Fitting procedure

Model parameters were derived using a standard variance-weighted least-squares minimisation procedure employing the Levenberg-Marquardt algorithm. Prior to the fit, the spectra were inspected visually and regions significantly affected by absorption or of low S/N were masked out.

In Fig. 2.4 we present our parametric fits to the C IV , $H\alpha$ and $H\beta$ emission lines in a handful of quasars, which have been chosen to illustrate the range of spectrum S/N and line shapes in the sample.

The Doppler velocities have been shifted so that the H α emission line centroid is at 0 km s⁻¹. The y-axes of the data-minus-model residual plots have deliberately been scaled by the spectrum flux errors. The mean reduced chi-squared values in our H α , H β and C IV fits are 1.69, 1.62, and 1.77 respectively and, in general, there are no strong features observable in the spectrum minus model residuals. The only significant features seen in the residual C IV spectra correspond to the location of narrow absorption lines which were excluded in the fitting procedure.

Table 2.2 includes the line parameters of our best-fitting model for each line. The reported line-width measures are corrected for instrumental broadening by subtracting the resolution of the spectrograph in quadrature. The spectrograph resolutions, which we estimate from the line widths in the observed sky spectra, range from 25 km s⁻¹ for XSHOOTER to 477 km s⁻¹ for the low-resolution LIRIS grism and are therefore small relative to the quasar broad line widths.

2.3.5 Spectra removed from sample

Through visual inspection we flagged and discarded the spectra of quasars for which reliable emission line parameters could not be obtained.

First, we flagged emission lines in spectra that possessed insufficient S/N. A single minimum S/N threshold was not entirely effective and, instead, spectra were flagged when it was judged conservatively that no meaningful constraints could be placed on the velocity centroid and/or width of the emission-line.

Second, we flagged emission lines where significant regions of the continuum and/or emission line fell outside of the wavelength coverage of the spectra. Reliable continuum definition and subtraction is not straightforward for emission lines so affected.

Third, we flagged C IV emission lines because of strong, narrow absorption close to the peak of the line where reliable interpolation across the absorption, using our parametric model, was not possible.

The number of spectra that are removed by each cut is given in Table 2.3 and the distribution in redshift and luminosity is shown in Fig. 2.5. Unsurprisingly, there is a preferential removal of intrinsically faint quasars, whose spectra can be of poorer S/N, and a loss of quasars at redshifts $z \sim 2.6$ where the H α emission falls at the edge of the K-band. H β is much weaker than H α , and the H β spectra are generally of lower S/N. As a result, the fraction of H β spectra that are flagged – 39 per cent – is particularly high.

Table 2.2: The format of the table containing the emission line properties from our parametric model fits. The table is available in machine-readable form in the online journal.

	Units	Description
NAME		Catalogue name
FWHM_BROAD_HA	km s^{-1}	FWHM of broad H α line
FWHM_BROAD_HA_ERR	km s^{-1}	
SIGMA_BROAD_HA	km s^{-1}	Dispersion of broad H α line
SIGMA_BROAD_HA_ERR	km s^{-1}	
Z_BROAD_HA		Redshift from broad H α line
FWHM_BROAD_HB	km s^{-1}	FWHM of broad H β line
FWHM_BROAD_HB_ERR	km s^{-1}	
SIGMA_BROAD_HB	km s^{-1}	Dispersion of broad H β line
SIGMA_BROAD_HB_ERR	km s^{-1}	
Z_BROAD_HB		Redshift from broad H β line
FWHM_CIV	km s^{-1}	FWHM of C iv doublet
FWHM_CIV_ERR	km s^{-1}	
SIGMA_CIV	km s^{-1}	Dispersion of C iv doublet
SIGMA_CIV_ERR	km s^{-1}	
BLUESHIFT_CIV_HA	km s^{-1}	Blueshift of C iv relative to H α
BLUESHIFT_CIV_HA_ERR	km s^{-1}	
BLUESHIFT_CIV_HB	km s^{-1}	Blueshift of C iv relative to H β
BLUESHIFT_CIV_HB_ERR	km s^{-1}	
LOGL ₅₁₀₀	erg s^{-1}	Luminosity at 5100Å
LOGL ₁₃₅₀	erg s^{-1}	Luminosity at 1350Å

Table 2.3: The number of spectra removed from our sample by the cuts described in Section 2.3.5.

		H α sample	H β sample
	Total	194	279
H α /H β	Wavelength	6	27
	S/N	8	83
C iv	Wavelength	6	5
	S/N	4	12
	Absorption	6	8
	Total remaining	164	144

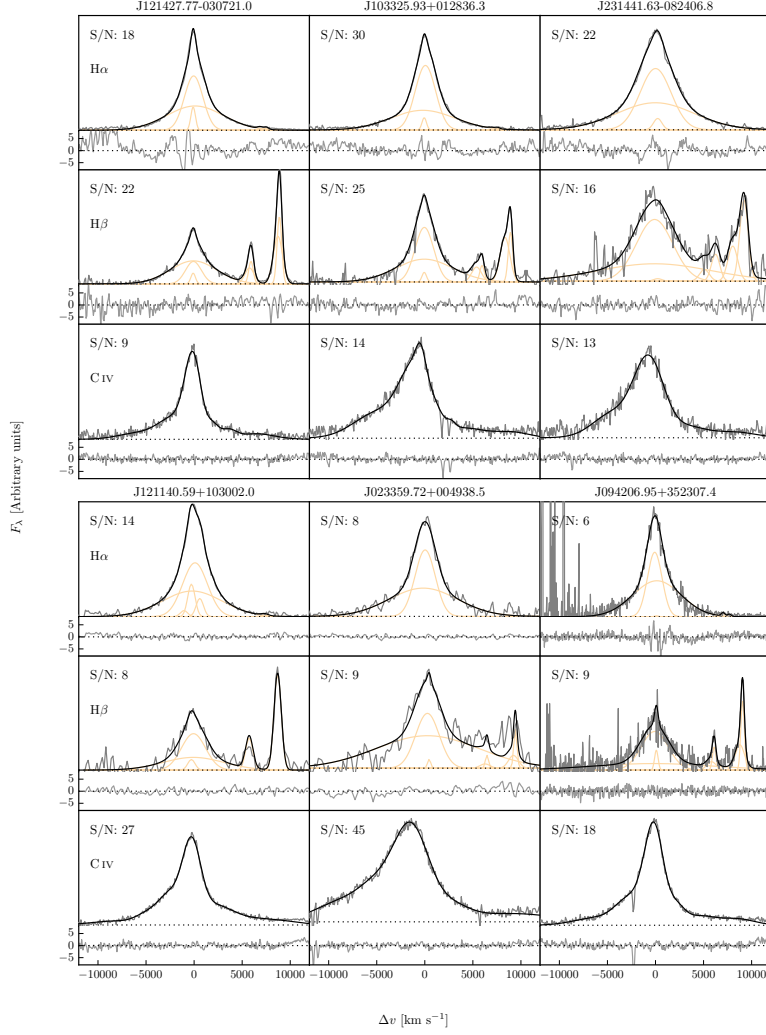


Figure 2.4: Model fits to continuum-subtracted H α , H β , and C IV emission in six quasars, chosen to represent the range of S/N (indicated in the figure and given per 150km s^{-1} pixel in the continuum) and line shapes present in the catalogue. The data is shown in grey, the best-fitting parametric model in black, and the individual model components in orange. The centroid of the broad H α emission is used to set the redshift, and Δv is the velocity shift from the line rest-frame transition wavelength. Below each fit we plot the data minus model residuals, scaled by the errors on the fluxes.

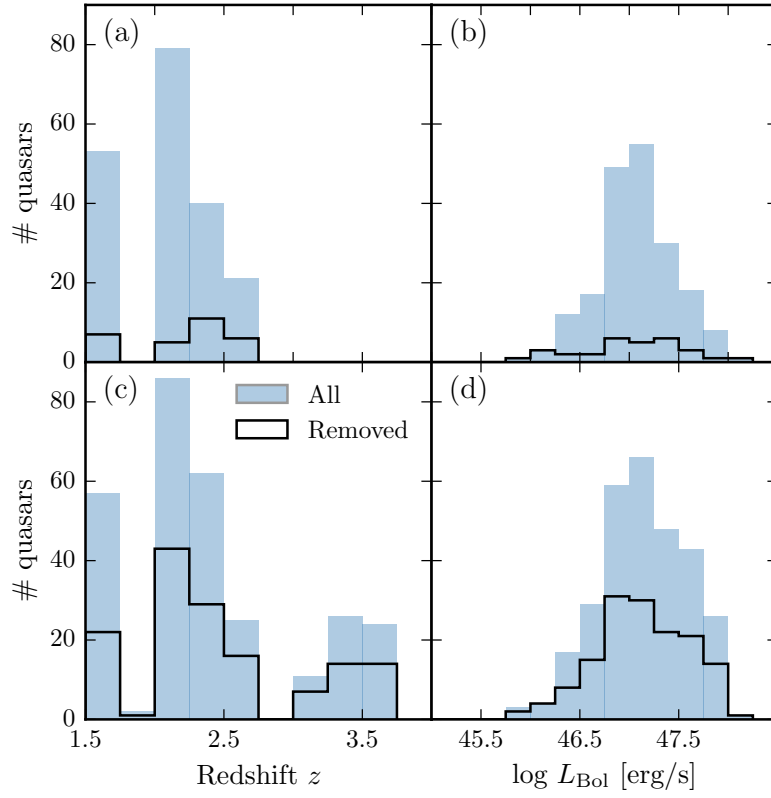


Figure 2.5: The redshift and luminosity distributions of the spectra removed from our H α /C IV (a, b) and H β /C IV (c, d) samples.

2.3.6 Emission-line parameter uncertainties

The 1σ error bars calculated from the covariance matrix in least-squares minimisation will underestimate the true uncertainties on the line parameters, since they do not account for systematic errors such as the significant uncertainty introduced in the continuum subtraction procedure. To calculate more realistic uncertainties on our fitted variables we employed a Monte Carlo approach. One thousand artificial spectra were synthesised, with the flux at each wavelength drawn from a Normal distribution (mean equal to the measured flux and standard deviation equal to the known error). Our emission-line fitting recipe was then implemented on each of these mock spectra. The uncertainty in each parameter is given by the spread in the best-fitting values from the one thousand realisations of the fitting routine. In some cases the standard deviation of the parameter distribution was biased by extreme values caused by bad fits³. We therefore chose to measure the spread in the parameter distribution by fitting a composite model with two Gaussian components – one to model uncertainty in the parameter and the other any possible outlier component. The uncertainty in each line parameter was then taken to be the width of the narrower Gaussian.

The 1σ error bars calculated from the covariance matrix in least-squares minimisation will underestimate the true uncertainties on the line parameters, since they do not account for systematic errors such as the significant uncertainty introduced in the continuum subtraction procedure. To calculate more realistic uncertainties on our fitted variables we employed a Monte Carlo approach. Artificial spectra were synthesised, with the flux at each wavelength drawn from a Normal distribution (mean equal to the measured flux and standard deviation equal to the known error). Our emission-line fitting recipe was then implemented on five thousand artificial spectra. Our parameter uncertainties are defined as the standard deviation of the best-fitting parameter values from these five thousand realisations. The uncertainty on the monochromatic continuum luminosity at 5100 \AA was estimated via a very similar method – using the error on the UKIDSS H-band magnitude to run a number of realisations of our SED-fitting routine. The uncertainties on all derived quantities, such as the BH mass, are propagated through by assuming that the uncertainties are uncorrelated and independent.

Because of its sensitivity to the flux in the wings of the line profile, care must be taken to define an appropriate range over which to measure the line dispersion. This is particularly true of Lorentzian-like profiles with extended wings. In spectra of only moderate S/N the line limits are difficult to determine unambiguously, which introduces an extra degree of uncertainty in line dispersion measurements.

³ In the analysis of the real spectra such fits are identified via visual inspection.

In common with previous work (e.g. Vestergaard and Peterson, 2006), by default, the dispersion was calculated within $\pm 10\,000\text{ km s}^{-1}$ of the line centre, but this was extended when appropriate to avoid excluding a significant amount of line flux.

2.3.7 Contemporaneity of spectra

The epochs of the near-infrared and optical spectra can differ by many years. For example, the NTT SOFI spectra were taken ~ 14 years after the SDSS spectra, and the VLT SINFONI spectra 20 years or more after the Hamburg/ESO observations⁴. If the broad emission line profiles varied significantly on these time-scales the relation between the C IV and Balmer line-width measurements could be blurred.

Cases do exist of dramatic changes in quasar spectra over short time-scales, but this phenomenon is rare (MacLeod et al., 2016). In our spectroscopic catalogue there are 112 SDSS DR7 quasars which are re-observed in BOSS and included in the DR12 quasar catalogue. The mean time elapsed between the two sets of observations is ~ 8 years. The root-mean-square difference in the C IV FWHM measured from the BOSS and SDSS spectra is a modest $\simeq 500\text{ km s}^{-1}$. Differences in the S/N of the spectra will make a substantial contribution and the scatter due to true variations in the C IV velocity-width will be significantly smaller than 500 km s^{-1} . We conclude therefore that any intrinsic changes with time do not materially affect the emission line measurements.

2.3.8 Quasar monochromatic luminosity

Computing virial BH masses also requires the quasar luminosity in an emission-line free region of the continuum adjacent to the broad line being used. The luminosity is used as a proxy for the size of the BLR. The monochromatic continuum flux is generally measured at 1350 \AA for C IV and 5100 \AA for H α and H β .

Relative flux-calibration of the infrared spectra as a function of wavelength has been achieved, to $\simeq 10$ per cent, through observations of appropriate flux standards. The absolute flux levels, however, can be in error by large factors due to variable atmospheric conditions combined with the narrow slit widths. For the majority of the quasars we have, therefore, established the absolute flux scale for each near-infrared spectrum using the same quasar SED-model fitting scheme employed in Paper I. The SED model, described in Maddox et al., (2012), gives a very good fit to the SDSS and UKIDSS magnitudes of SDSS DR7 quasars, reproducing the individual magnitudes with a $\sigma < 0.1\text{ mag}$. For 207 quasars, (Y)JHK passband magnitudes from the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007)

⁴ Time differences in the quasar rest-frame are reduced by a factor of $(1+z)$.

Large Area Survey, the Two Micron All Sky Survey (2MASS; Skrutskie et al., 2006) and the Visible and Infrared Survey Telescope for Astronomy (VISTA) Hemisphere Survey (VHS; McMahon et al., 2013) and Kilo-Degree Infrared Galaxy (VIKING; Edge et al., 2013) survey are available. The SED model was fit to the infrared magnitudes; integrating the SED model through the pass-band transmission functions, to give model magnitudes, and performing a variance weighted least-squares fit to the observed magnitudes. The flux at 5100 \AA was then taken from the normalised model.

For 19 of the remaining 23 quasars, where near-infrared photometry was not available, the quasar SED model was fit to the SDSS spectra, the flux calibration of which are known to be excellent. The fit was done using a simple variance-weighted chi-squared minimisation procedure in emission line-free intervals of the optical spectra. The model includes a reddening, $E(B - V)$, based on a Small Magellanic Cloud-like extinction curve, and an overall normalisation of the model as free parameters. In practice, the quasars possess only very modest reddenings, with $E(B - V) \simeq 0.0-0.1$. The flux at 5100 \AA was then, again, taken from the normalised SED model. For the four remaining quasars, which possess neither near-infrared photometry nor SDSS DR7 spectra, we fit the SED model to the BOSS DR12 spectra. To avoid the known issues in the flux calibration of the BOSS DR12 quasar spectra at observed-frame blue wavelengths (Lee et al., 2013), our fitting was confined to rest-frame wavelengths long-ward of 1275 \AA .

Comparison of the 5100 \AA luminosity, computed using the photometry- and spectrum-based methods for 177 quasars, showed a scatter of just ~ 0.1 dex. We therefore assume 0.1 dex to be the measurement uncertainty on the 5100 \AA luminosities.

For 34 quasars in the catalogue the optical spectra come from surveys other than SDSS/BOSS and optical magnitudes from recent epochs are not available. In order to obtain an estimate of the luminosity at 1350 \AA for the 30 quasars, we take the standard Maddox et al., (2012) quasar SED model, normalised to the near-infrared magnitudes, and read off the flux at 1350 \AA .

For all the catalogue quasars, the optical and near-infrared spectra as well as the near-infrared photometry were obtained at different epochs, with rest-frame time differences of up to ~ 5 years. Intrinsic quasar photometric variability in the rest-frame ultraviolet and optical will therefore add additional scatter of ~ 0.2 mag (e.g. MacLeod et al., 2010) to the derived 1350 \AA - and 5100 \AA -luminosities.

Given that the luminosity enters into the calculation of BH-mass only as the square-root, the uncertainty on the luminosities does not make a large contribution to the uncertainties in the BH mass estimates.

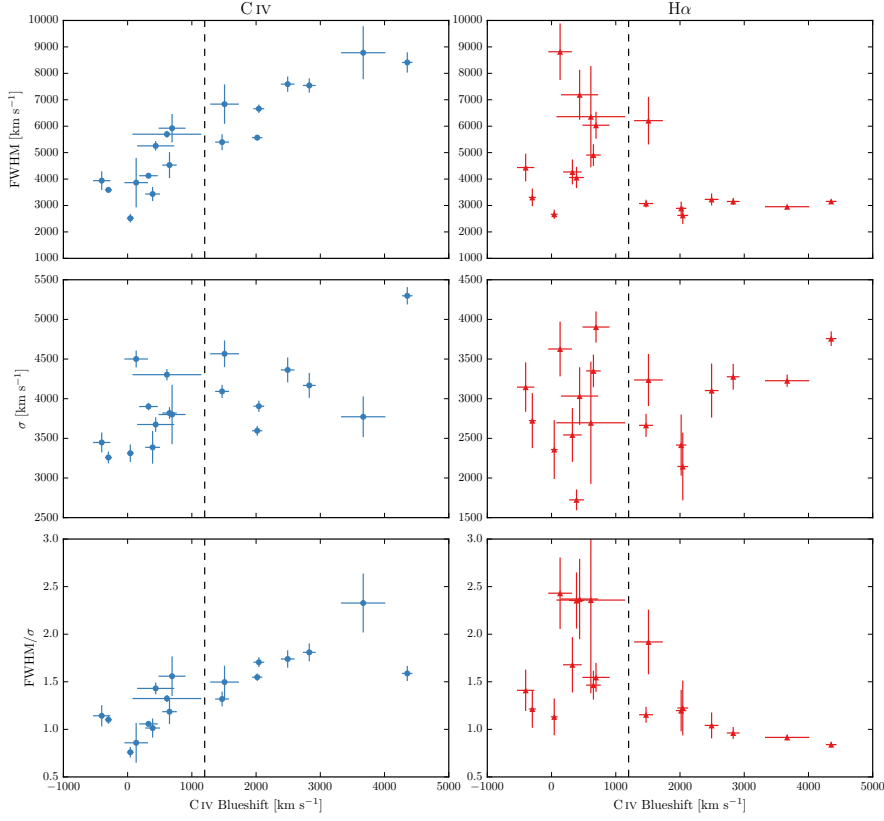


Figure 2.6: The FWHM, dispersion (σ) and shape (FWHM/σ) of C IV and H α as a function of the C IV blueshift. The vertical line demarcates the ‘high’ and ‘low’ C IV blueshift regimes discussed in the text. At high blueshift it is clear that BH masses estimated from the C IV FWHM (as is typically done at the redshifts considered) will be significantly larger than those estimated from the H α FWHM.

The luminosity at 1350\AA was taken from the spectral fits of Shen et al., (2011). The quasar rest-frame continuum at 5100\AA often lies at the edge, or beyond, the wavelength coverage of the LIRIS spectra. Monochromatic 5100\AA luminosities were therefore calculated from the fit of our parametric quasar model (described in Maddox et al., (2012)) to the UKIDSS H- broadband magnitude for each quasar. The model fits to the quasars are excellent, with residuals in SDSS and UKIDSS passbands under 10 per cent.

2.3.9 Characterising the emission-line profiles

Our definition of the C IV blueshift differs slightly in two ways from the values plotted for the quasar population in Fig. 2.23. Firstly, we use the peak of our parametric model fit to the H α line to define the

quasar systemic-redshift⁵. Secondly, the centre of the C iv line is now defined as the wavelength that bisects the cumulative total flux of our best-fit GH-polynomial model rather than of the data.

There has been a considerable degree of attention paid to the effectiveness of different velocity-width measures of the C iv-emission; specifically, the line FWHM and the dispersion, σ , derived from the second-moment velocity (e.g. Assef et al., 2011; Denney et al., 2013). The FWHM and line dispersion trace different parts of the broad line velocity field, with the FWHM relatively more sensitive to any low-velocity core present and the line dispersion relatively more sensitive to the high velocity wings. The shape of the line can be characterised by the ratio FWHM/σ . $\text{FWHM}/\sigma \simeq 2.35$ for a Gaussian profile, while $\text{FWHM}/\sigma \simeq 1$ for a peakier Lorentzian profile⁶. In practice, the line dispersion is almost certainly a more robust velocity indicator when the assumptions underlying the virial-origin of the emission-line velocity width are true and the spectral S/N and resolution are adequate. This was demonstrated by Denney et al., (2013) for a sample of quasars possessing a significantly smaller range in C iv-blueshift than investigated here.

In reality, however, as highlighted by Denney, (2012), contributions to the C iv-emission line profile from gas where virial motions do not dominate can be significant. Looking to the future, the results of the new reverberation-mapping projects (Shen et al., 2015; King et al., 2015) will show what fraction of the C iv-emission line, as a function of velocity, does reverberate for quasars with an extended range of C iv-emission shapes. The derivation of quantitative corrections to transform velocity-width measures from single-epoch to reverberation-only line profiles should then be possible.

As such information is not yet available, there is a strong rationale for investigating whether the systematic changes in the C iv-emission line profile can be used to improve the single-epoch BH-mass estimates derived using the C iv line. In the left panels of Fig. 2.6 we show how the C iv FWHM, line dispersion, σ , and line shape, FWHM/σ , vary as a function of the blueshift. The C iv FWHM is correlated with the blueshift, with the median FWHM of quasars with the largest blueshifts a factor of 2-3 higher than quasars with only moderate blueshifts. The dispersion, however, does not show a similarly strong systematic variation.

Without knowledge of the C iv-blueshifts, the dynamic range present in the FWHM and line dispersion measurements accords with the expectations from the study of Denney et al., (2013); the factor of $\simeq 4$ spread in the FWHM measurements indicating greater sensitivity to

⁵ The H α -derived redshifts are very closely in agreement with those from the forthcoming Allen & Hewett redshifts, which are plotted in Fig. 2.23c.

⁶ Strictly $\text{FWHM}/\sigma \rightarrow 0$ for a Lorentzian profile, but values close to unity are typical when the dispersion is calculated over a velocity range, $\simeq \pm 10\,000\text{km s}^{-1}$, used to parametrize broad emission lines in quasar spectra.

the emission-line profile shape than is the case for the dispersion, which varies by a factor of only $\simeq 2$. Adopting a value of 1200 km s^{-1} to define ‘low’ and ‘high’ blueshift, the median C IV-emission dispersion for the low and high-blueshift samples differ by only 10 per cent. It follows, therefore, that while the dispersion provides a relatively line-profile independent measure of the velocity width for quasars where the underlying assumption regarding the virial-origin of the velocity width applies, quasars where the assumption is not true can be assigned apparently normal velocity-widths and hence potentially incorrect BH-masses.

To emphasise this point, in Fig. 2.7 we overlay the C IV line profiles of SDSSJ1236+1129 and SDSSJ1525+2928, whose dispersions (Table ??) are indistinguishable (4168 ± 271 and $4303 \pm 128 \text{ km s}^{-1}$ respectively). Notwithstanding the very similar dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread of the C IV emission in both quasars.

The analysis here, building on earlier work (including Shen and Liu, 2012; Sulentic et al., 2007), confirms a link between C IV emission-line shape and blueshift, raising the prospect of developing a blueshift-dependent correction to single-epoch BH-mass estimates based on the C IV line. Expressed in another way, we are interested in testing if the significant systematic change in line shape as a function of C IV blueshift can be used to provide improved single-epoch BH-masses from the C IV emission line. The tightness of the correlation we observe between the C IV FWHM and blueshift implies that such an approach may be more effective than using the C IV emission-line velocity dispersion without reference to blueshifts. A further practical advantage is that, given the typical S/N of current survey-quality spectra, virial BH mass estimates for high-redshift quasars are usually based on the FWHM rather than the dispersion (e.g. Shen et al., 2011), which, being strongly affected by the continuum placement, is often found to be difficult to measure robustly (e.g. Mejía-Restrepo et al., 2016). As a first step towards the goal, below (Sec. 2.5.5) we investigate the apparent systematic trends in the H α FWHM and line shape as a function of C IV blueshift (shown in the right of Fig. 2.6).

2.4 AN EMPIRICAL CORRECTION TO C IV-BASED VIRIAL BH-MASS ESTIMATES

2.4.1 H α /H β FWHM comparison

BH-mass calibrations which use the width of the broad H β emission line as a proxy for the virial velocity are widely regarded as the most reliable, since most reverberation mapping employs the H β line and the R – L relation has been established using H β . When H β is not

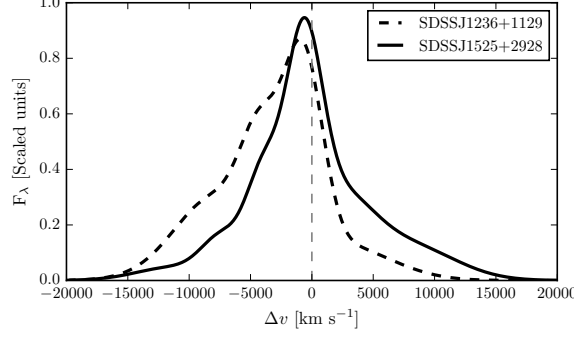


Figure 2.7: Comparison of the C IV line profiles of SDSSJ1236+1129 and SDSSJ1525+0426. Notwithstanding the essentially identical dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread of the C IV emission in both quasars.

available, H α has been shown to be a reliable substitute (e.g. Greene and Ho, 2005; Shen et al., 2011; Shen and Liu, 2012).

In our sample, we have 99 quasars with reliable measurements of both H α and H β lines. The 99 objects include 21 quasars which were excluded from the main 308-object catalogue because adequate measurements of the C IV FWHM and blueshift could not be acquired. The line widths are compared in Fig. 2.8 and, as expected, a tight correlation is observed. Greene and Ho, (2005), using a sample of 162 quasars with high S/N SDSS spectra at $z < 0.35$, established the following relation between the H α and H β FWHMs

$$\text{FWHM}(\text{H}_\alpha) = (1.07 \pm 0.07) \times 10^3 \left(\frac{\text{FWHM}(\text{H}_\beta)}{10^3 \text{ km s}^{-1}} \right)^{(1.03 \pm 0.03)} \quad (2.1)$$

The relation is shown as the dashed line in Fig. 2.8. The root-mean-square scatter about this relation is 0.07 dex, compared to the ~ 0.1 dex found by Greene and Ho, (2005). However, we find a systematic offset, in the sense that the H β line-widths we measure are on average larger by 270 km s^{-1} than predicted by the Greene and Ho, (2005) relation. As our sample covers higher redshifts and luminosities than the sample in Greene and Ho, (2005), we derive a new relation between the H α and H β FWHMs.

We assume a relation of the same form used by Greene and Ho, (2005), i.e. a simple power-law, and infer the model parameters by fitting a linear model (with slope α and intercept β) in log-log space. The fit is performed within a Bayesian framework described by Hogg, Bovy, and Lang, (2010). Each data point is treated as being drawn from a distribution function that is a convolution of the projection of the point's covariance tensor, of variance Σ_i^2 , with a Gaussian of

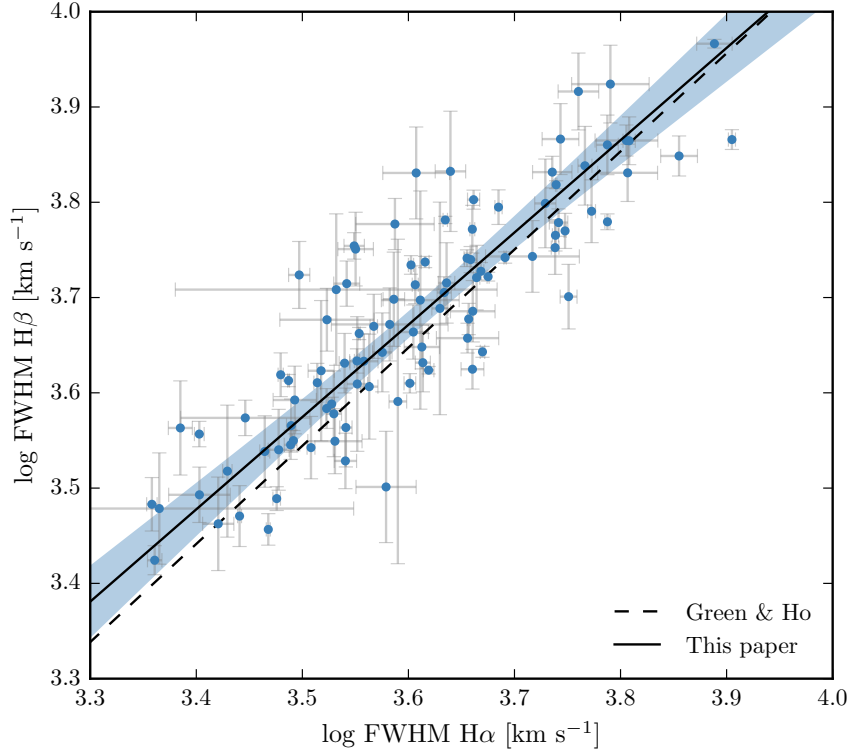


Figure 2.8: Comparison of H α and H β FWHM measurements for 99 quasars. The solid line is our best-fitting power-law model, and the blue-shaded region shows the 2- σ uncertainties on the model parameters. The dashed line is the relation found by Greene and Ho, (2005) using a sample of $z < 0.35$ SDSS AGN.

variance V representing the intrinsic variance in the data. The log-likelihood is then given by

$$\ln \mathcal{L} = - \sum_{i=1}^N \frac{1}{2} \ln [2\pi (\Sigma_i^2 + V)] - \sum_{i=1}^N \frac{\Delta_i^2}{2[\Sigma_i^2 + V]} \quad (2.2)$$

where Δ_i is the orthogonal displacement of each data point from the linear relationship. An advantage of this approach is that it allows a proper treatment of the measurement errors on both variables, which in this case are comparably large. The model also makes the reasonable assumption that there is an intrinsic scatter in the relationship between the variables that is independent of the measurement errors. Following the suggestion by Hogg, Bovy, and Lang, (2010), the linear model was parametrized in terms of (θ, b_\perp) , where θ is the angle the line makes with the horizontal axis and b_\perp is the perpendicular distance from the line to the origin. Uniform priors were placed on these parameters, and the Jeffreys prior (the inverse variance) was placed on the intrinsic variance. The posterior distribution was sampled using a Markov Chain Monte Carlo (MCMC) method using the Python package emcee (Foreman-Mackey et al., 2013).

The one- and two-dimensional posterior distributions are shown in Fig. 2.9. The solid line in Fig. 2.8 is the maximum likelihood solution

$$\text{FWHM}(H_\beta) = (1.23 \pm 0.10) \times 10^3 \left(\frac{\text{FWHM}(H_\alpha)}{10^3 \text{ km s}^{-1}} \right)^{0.97 \pm 0.05} \quad (2.3)$$

and the shaded region shows the 2σ uncertainties on the model parameters.

As discussed above, our relation is displaced to slightly higher $H\beta$ FWHM than the Greene and Ho, (2005) relation – the offset is 210 km s^{-1} for a quasar with $H\alpha$ FWHM 4500 km s^{-1} . We infer a power-law index that, although slightly shallower, is consistent with the Greene and Ho, (2005) index within the quoted uncertainties. The intrinsic scatter in the data, σ_I , we infer from the fit is 0.04 dex. This is smaller than the total scatter seen in Fig. 2.8 (0.06 dex), which suggests that measurement errors make a significant contribution to the total scatter in the relation.

For 19 of the 99 quasars with $H\beta$ and $H\alpha$ emission profiles, one of the two Gaussians used to reproduce the $H\beta$ profiles has a FWHM greater than 20000 km s^{-1} and a fractional contribution to the total $H\beta$ broad line flux of >0.3 (Marziani et al., 2009; Marziani et al., 2013). Such a broad component is not seen in the $H\alpha$ profiles and the very broad $H\beta$ -component may be an artifact of the fitting scheme. A particular issue for $H\beta$ is the presence of Fe II emission, often at a significant level. Furthermore, additional lines could be contributing to the underlying continuum (e.g. the $\text{He I } \lambda\lambda 4922, 5017$ doublet; Véron, Gonçalves, and Véron-Cetty, 2002; Zamfir et al., 2010).

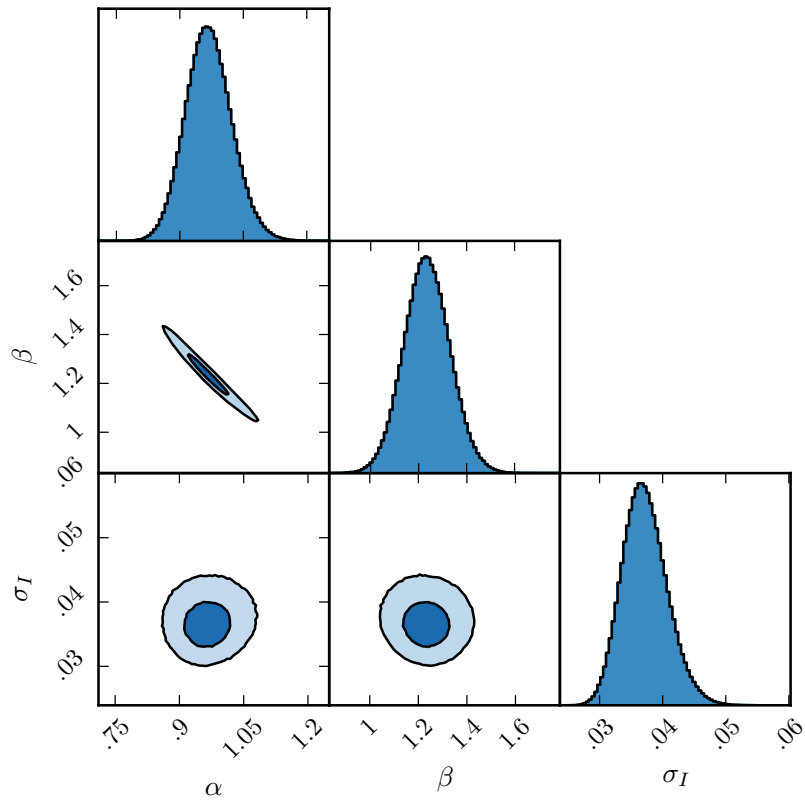


Figure 2.9: One- and two-dimensional projections of the MCMC sampling of the posterior distribution from the fit in Fig. 2.8. α is the power-law index, 10^β is the normalisation, and σ_I is the intrinsic scatter. In the two-dimensional projections, 1- and 2- σ contours are shown.

In Sec. 2.4.3 we use the whole of the H β profile to derive an unbiased BH mass. If, instead, the FWHM is calculated from the narrower of the two Gaussian components rather than the composite profile, then the H β FWHM decreases by 630km s^{-1} on average. The H α FWHM, which are calibrated against the H β FWHM, will also decrease by the same amount on average. This will enhance the C iv FWHM relative to the H α /H β FWHM by ~ 15 per cent and increase the size of the correction which must be applied to the C iv-based BH masses by ~ 30 per cent.

Virial BH mass estimators are typically based on the width of H β . However, the H α and H β emission is believed to originate from the same gas and the transformation between the emission-line velocity widths is expected to be well defined. Greene and Ho, (2005), using a sample of 162 quasars with high S/N SDSS spectra, established the following relation between the H α and H β FWHM:

$$\text{FWHM}(\text{H}\beta) = (1.07 \pm 0.07) \times 10^3 \left(\frac{\text{FWHM}(\text{H}\alpha)}{10^3 \text{ km s}^{-1}} \right)^{(1.03 \pm 0.03)} \quad (2.4)$$

Greene and Ho, (2005) found the root-mean-square scatter about this relation to be ~ 0.1 dex. We do not have a sufficient number of robust H β line measurements to test this relation directly. However, we are in the process of acquiring a much larger sample of quasars with near-infrared spectra covering H α and H β at similar redshifts and luminosities to the sample presented here. The H α and H β line widths of this sample are in excellent agreement with the Greene and Ho, (2005) relation. To indirectly test the H α /H β line width relation for the sample presented here, we first constructed mean composite spectra in the H α and H β emission line regions to increase the S/N. The individual rest-frame spectra (defined using the wavelength of the H α centroid) were interpolated on to a common wavelength grid. The spectra were then normalised using the continuum flux under the line centre, which was found by linearly interpolating between two emission-line free windows on either side of the line. Figure. 2.10 shows the composite H α and H β line regions overlaid and, as expected, the line profiles are closely matched.

2.4.2 Measuring the quasar systemic redshift

An accurate measure of the quasar's systemic redshift is required in order for the blueshift of the C iv emission line to be determined. Balmer emission centroids, where the centroid, λ_{half} , is the wavelength that bisects the cumulative total flux, are available for all quasars in the catalogue and so we use this to define the systemic redshift.

For 83 and 120 quasars in the H α and H β samples respectively narrow [O III] emission is also detected. In the model fit to the H β region

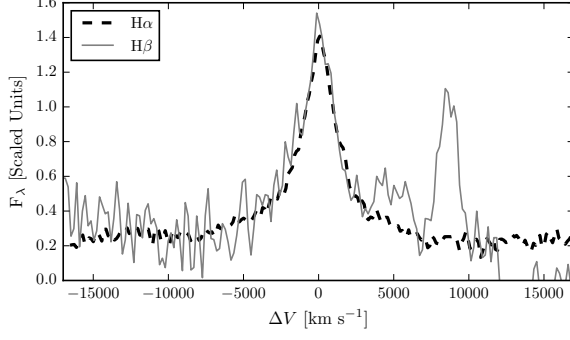


Figure 2.10: The H α and H β emission line regions in the median composite spectrum, shown as function of the velocity shift from the respective predicted line peak wavelengths. The line fluxes have been scaled in order for the profile shapes to be readily compared. The H α and H β line profiles are very similar, which suggests a tight correlation between the H α and H β line widths. Quasar narrow-line emission from O III λ 5008.2 is visible but, overall, the O III λ 4960,5008 emission is relatively weak in these spectra.

the velocity centroids of the broad H β -line and the core component of the [O III] emission were deliberately determined separately. We find the intrinsic difference in the velocity centroids of the Balmer broad emission and the narrow [O III] emission to have a dispersion of 360 km s^{-1} , which is very similar to the value found by Shen et al., (2016). However, the median velocity centroid of the narrow component of the [O III] emission is blueshifted by 270 km s^{-1} relative to the centroid of the broad Balmer line. Applying our parametric model fitting routine to the composite spectrum from Hewett and Wild, (2010), which is constructed using relatively low redshift SDSS quasars with $L_{\text{Bol}} \sim 10^{44} \text{ erg s}^{-1}$, the centroids of the broad component of H β and the narrow component of [O III] are found to be at essentially identical velocities, suggesting that the blueshifting of narrow [O III] could be luminosity dependent. Regardless, since both the systematic offset and the scatter are small in comparison to the dynamic range in C IV blueshifts, the blueshift-based empirical correction we will derive does not depend on whether the broad Balmer emission or the [O III] centroid is used to define the systemic redshift.

2.4.3 Balmer/C IV line widths as a function of C IV-blueshift

A fundamental assumption on which single-epoch virial BH-mass estimates are based is that the widths of the broad emission lines are directly related to the virial motions of the emitting clouds moving in the gravitational potential of the central BH. However, the C IV line profiles of the quasars in our sample with the largest C IV blueshifts

indicate that non-virial motions, very likely due to outflows, are having a significant effect on the observed C iv emission velocity profile (e.g. Gaskell, 1982; Baskin and Laor, 2005; Sulentic et al., 2007; Richards et al., 2011; Wang et al., 2013). As shown in Fig. 2.1, at fixed emission-line EW, virtually the entire C iv-profile appears to shift blueward and the change in line shape is not simply an enhancement of flux in the blue wing of a still identifiable symmetric component. While gravity almost certainly plays a key role, determining the escape velocity for out-flowing material for example, it is clear that the virial assumption, on which single-epoch BH-mass measurements are predicated, is not straightforwardly applicable for the C iv-emission line in quasars exhibiting large blueshifts.

The main aim of this paper is to investigate potential systematic trends in C iv-based single-epoch virial BH masses as a function of the C iv blueshift. Calibrations using H β (and therefore also H α) are generally accepted to be the most reliable, since most reverberation mapping employs the H β line and the R – L relation has been established using H β . Therefore, we will test the reliability of the C iv-based estimates by comparing C iv line profiles to H α profiles in the same quasars.

In this section we directly compare the C iv and H α /H β line widths as a function of the C iv blueshift. Because virial BH mass estimates are generally based on the H β FWHM, we first convert our H α FWHM measurements to equivalent H β FWHM using Eq. 2.3. In Fig. ??a and b we show the C iv FWHM relative to both the (H β -scaled) H α FWHM and the H β FWHM, as a function of the C iv blueshift.

Employing the same Bayesian fitting framework described in Section 2.4.1, we fit independent linear models to the C iv FWHM relative to the H α and H β FWHM as a function of the C iv blueshift. As before, our model has an additional parameter representing any intrinsic scatter in the relationship between the variables which is independent of measurement errors. We also tested a model where some fraction of the data points (which is free to vary) are drawn from an outlier distribution, represented by a broad Gaussian centered on the mean of the data. We found, however, that the inferred outlier fraction was very low (0.004, corresponding to ~ 0.7 data points) and so did not include such a component in our model.

In Fig. 2.13 we show the one- and two-dimensional projections of the posterior distribution from the linear fit to the FWHM C iv/H α ratio. The projections from the FWHM C iv/H β fit (not shown) have very similar appearances. In Fig. ??a we plot the maximum likelihood model and the 2σ uncertainties on the model parameters. The maximum likelihood line is given by

$$\text{FWHM}(\text{C iv, Corr.}) = \frac{\text{FWHM}(\text{C iv, Meas.})}{(0.41 \pm 0.02) \left(\frac{\text{C iv Blueshift}}{10^3 \text{ km s}^{-1}} \right) + (0.62 \pm 0.04)} \quad (2.5)$$

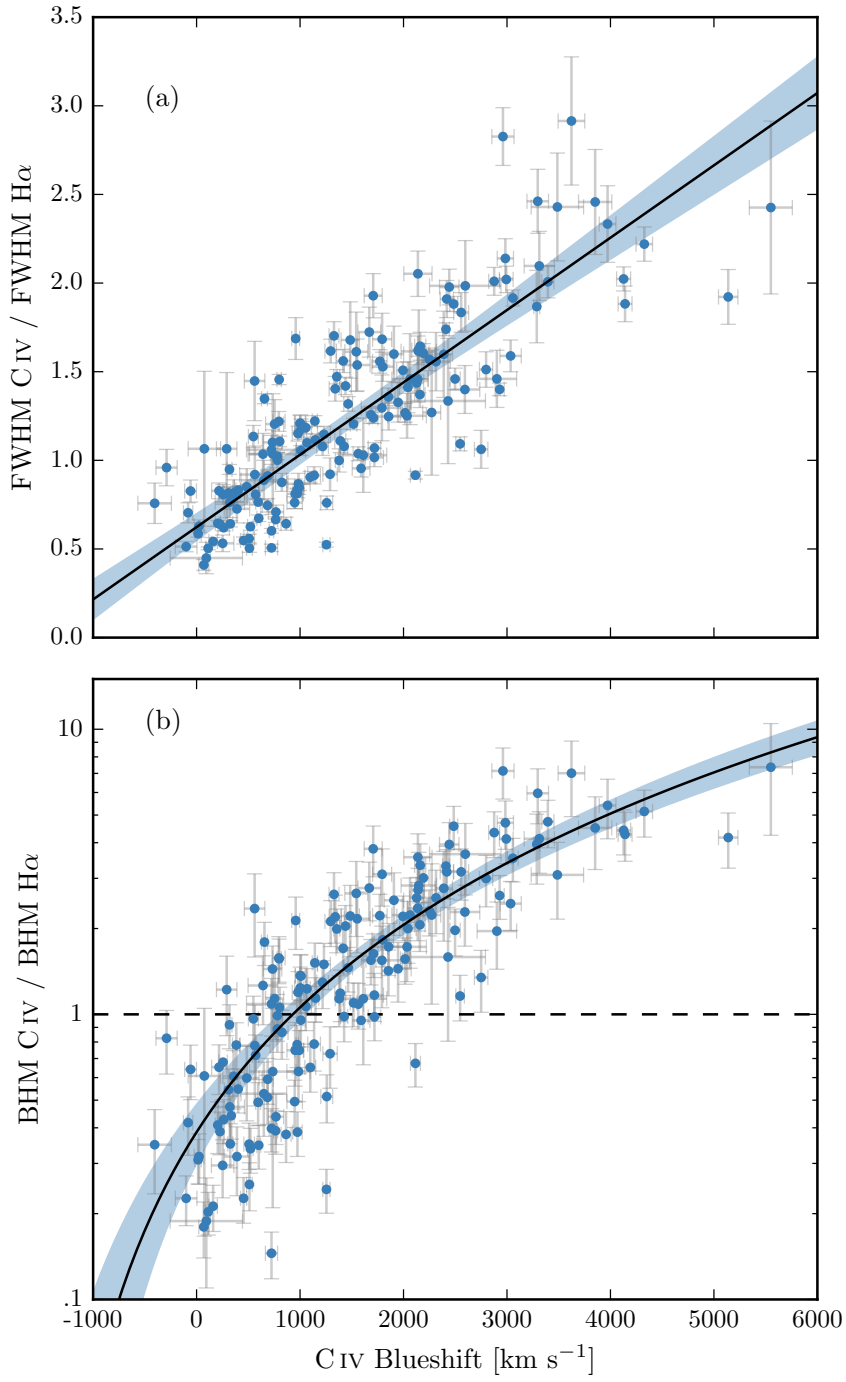


Figure 2.11: C iv FWHM relative to H α FWHM (a), and C iv based BH mass (BHM) compared to H α based mass (b), both as a function of the C iv blueshift. The black line is our best-fit linear model, and the shaded region shows the 2- σ uncertainties on the slope and intercept. The H α FWHM have been scaled to match the H β FWHM using Eq. 2.3.

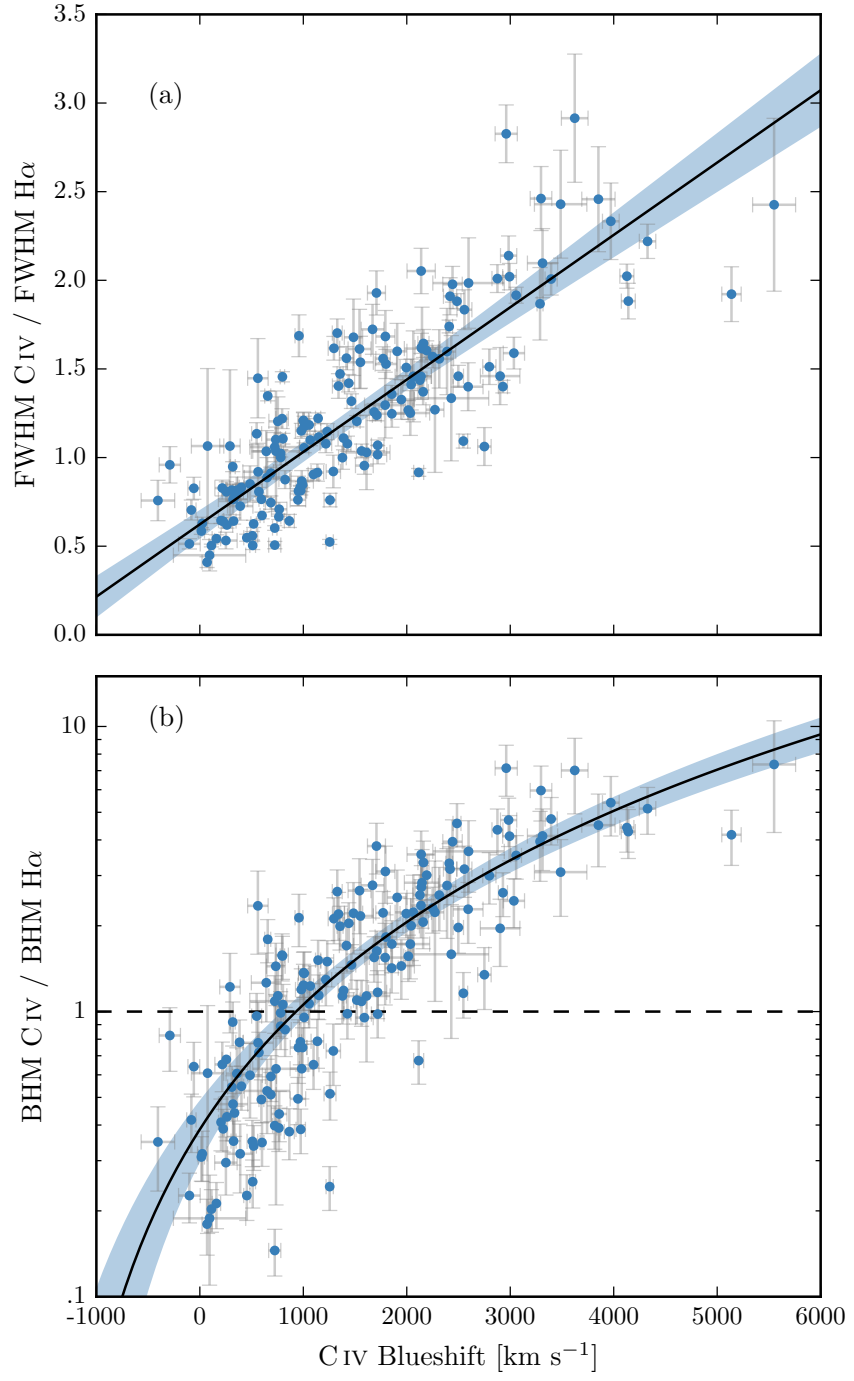


Figure 2.12: C IV FWHM relative to H β FWHM (a), and C IV based BH mass (BHM) compared to H β based mass (b), both as a function of the C IV blueshift.

for the C IV/H α fit and

$$\text{FWHM}(\text{C IV, Corr.}) = \frac{\text{FWHM}(\text{C IV, Meas.})}{(0.36 \pm 0.03) \left(\frac{\text{C IV Blueshift}}{10^3 \text{ km s}^{-1}} \right) + (0.61 \pm 0.04)} \quad (2.6)$$

for the C IV/H β fit. The intercepts of the two relations are consistent, while the difference between the slopes is only marginally inconsistent given the quoted uncertainties.

The intrinsic scatter in the data about the linear relation we infer is 0.23 ± 0.02 and 0.25 ± 0.02 for the H α and H β fits respectively. The intrinsic scatter for the H α fit is represented by the Normal probability density distribution shown in Fig. 2.14. In the same figure we show the distribution of the orthogonal displacement of each data point from the best-fitting linear relationship. The two distributions are well-matched, which demonstrates that our model is a good representation of the data and the measurement errors on the data points are small relative to the intrinsic scatter.

The overall (intrinsic and measurement) scatter about the best-fitting model is slightly higher when the C IV line-widths are compared to H β (0.12 dex) than when compared to H α (0.10 dex). This is likely due, at least in part, to the generally higher S/N of the H α emission. In addition, contributions from the strong [O III] doublet in the vicinity of H β make de-blending the H β emission more uncertain. As a consequence, for quasars where H α and H β are both measured, the mean uncertainty on the H α FWHM is 130 km s^{-1} , compared to 340 km s^{-1} for H β .

In the next section we use both the H α and H β lines to calculate unbiased BH masses. We use the H α measurements to derive an empirical C IV blueshift based correction to the C IV masses (Eq. 2.8) because of the issues related to the accurate modelling of the H β -profile just described. An extra advantage, which is evident in Fig. ??, is that the H α sample has a better C IV blueshift coverage. However, as can be seen from the similarity of Equations 2.5 and 2.6, our results would not change significantly were we instead to use the H β sample.

2.4.4 C IV-derived BH masses at low C IV blueshift

The H α and C IV FWHM (dispersion) of the 10 quasars with C IV blueshifts $< 1200 \text{ km s}^{-1}$ are linearly correlated, as expected if the dynamics of the BLR clouds are dominated by virial motions. The median C IV/H α FWHM (dispersion) ratio is 0.91 (1.22) with standard deviation 0.17 (0.28). Thus, the dispersion-based comparison results in a median C IV/H α consistent with the value of $\simeq 1.4$ from assuming a virial origin for the emission but with a relatively large standard deviation. As predicted in Section ??, the FWHM-based comparison results in a systematically lower median C IV/H α . However, the corre-

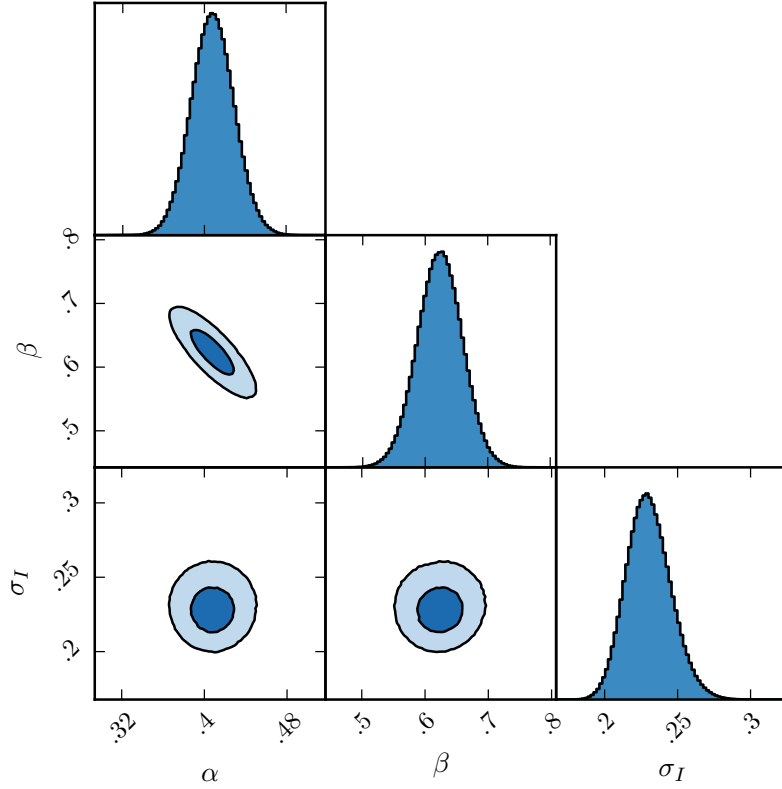


Figure 2.13: One- and two-dimensional projections of the MCMC sample of the posterior distribution for a linear fit to the FWHM C IV/H α ratio as a function of the C IV blueshift. In the two-dimensional projections we show 1- and 2- σ contours. The posterior distribution for the linear fit to the FWHM C IV/H β ratio, which we do not show, has a very similar appearance.

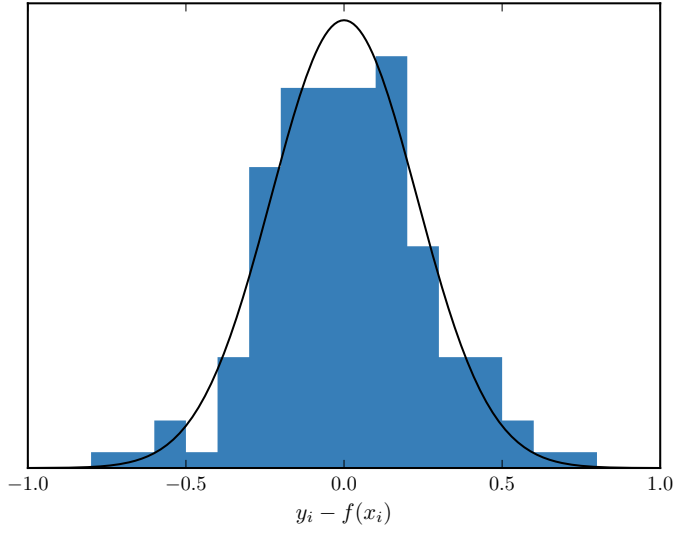


Figure 2.14: The distribution of the orthogonal displacement of each data point from the best-fitting linear relationship in the fit to $\text{FWHM}(\text{C IV})/\text{FWHM}(\text{H}\alpha)$ as a function of the C IV blueshift (blue histogram). The black curve is a Normal distribution with a width equal to the intrinsic scatter in the population inferred from the fit. The two distributions are well-matched, which demonstrates that our model is a good representation of the data and the measurement errors on the data points are small relative to the intrinsic scatter.

lation between the C iv and H α FWHMs is significantly tighter, lending support to the proposal that corrections to BH-mass estimates based on the C iv emission line properties may be possible.

Virial BH masses were calculated using the widely adopted Vestergaard and Peterson, (2006) calibrations. The Vestergaard and Peterson, (2006) C iv FWHM calibration uses the monochromatic continuum luminosity at 1350 Å to predict the BLR radius and corresponds to ($a = 6.66$, $b = 2$, $c = 0.53$) in Eq. 2.7. The calibration coefficient $a = 6.73$ in their equivalent dispersion-based relation. For the H β calibration, Vestergaard and Peterson, (2006) use the monochromatic continuum luminosity at 5100 Å and calibration coefficients corresponding to ($a = 6.91$, $b = 2$, $c = 0.5$). BH masses are computed using the line and continuum properties given in Tables ?? and ??, and we convert our H α emission-line velocity-width measures to predicted H β widths using Eq. 2.4.

As a direct consequence of the empirically small C iv/H α FWHM ratio, the C iv-derived BH mass estimates are systematically lower than the corresponding H α -derived masses when the blueshift is small. This can be seen in Fig 2.15, where for every quasar with a C iv blueshift $< 1200 \text{ km s}^{-1}$ (i.e. to the left of the dashed line), the C iv-derived BH mass is smaller than the corresponding H α -derived mass. The median fractional difference between the two estimates is 0.60.

For the 10 quasars with low C iv blueshifts, we looked for correlations of the C iv/H α FWHM ratio with other spectral properties. We found marginal evidence for an anti-correlation with the H α FWHM (Spearman coefficient 0.58 with p-value 0.08). Among the quasars with H α FWHM $> 4000 \text{ km s}^{-1}$ we found the mean C iv/H α FWHM ratio to be 0.83, compared to 1.01 for the quasars with H α FWHM $< 4000 \text{ km s}^{-1}$. Similar trends have been observed at low- z ; in a sample of Boroson and Green, (1992) quasars, Baskin and Laor, (2005) found the C iv line to be broader than H β when the H β FWHM $< 4000 \text{ km s}^{-1}$ and narrower when the H β FWHM $> 4000 \text{ km s}^{-1}$.

2.4.5 C iv based virial BH mass estimates

Single-epoch virial BH mass estimates normally take the form

$$M_{\text{BH}} = 10^a \left(\frac{\Delta V}{1000 \text{ km s}^{-1}} \right)^b \left[\frac{L_\lambda}{10^{44} \text{ erg s}^{-1}} \right]^c \quad (2.7)$$

where ΔV is a measure of the line width (from either the FWHM or dispersion), L_λ is the monochromatic continuum luminosity at wavelength λ , and a , b , and c are coefficients, determined via calibration against a sample of AGN with reverberation-mapping BH mass estimates. Several calibrations have been derived using different lines

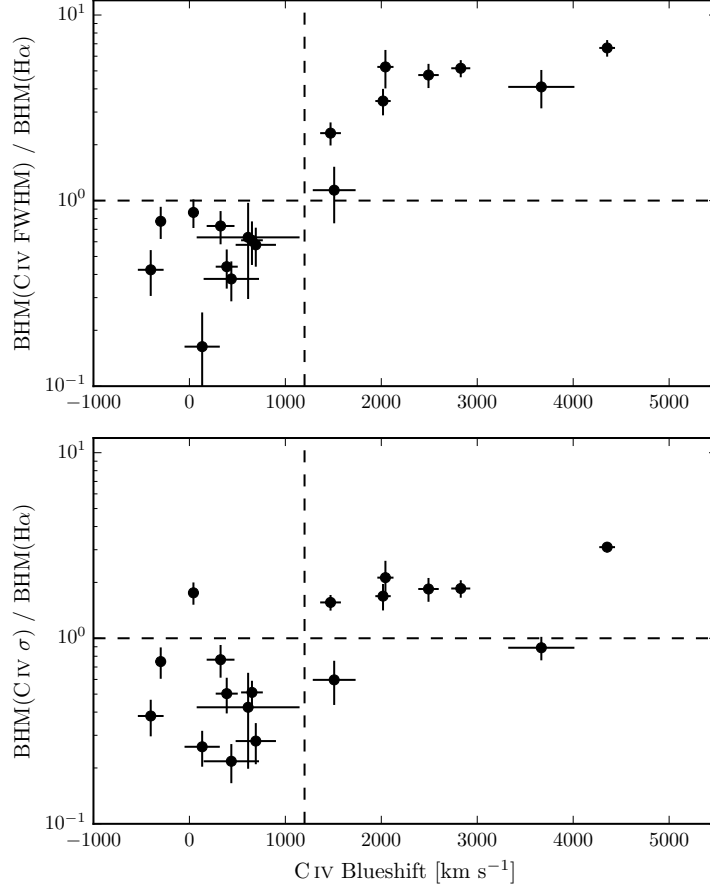


Figure 2.15: Comparison of virial BH mass estimates based on the C IV FWHM (*top*) and dispersion σ (*bottom*) and H α FWHM as a function of the C IV blueshift. The horizontal line indicates agreement between C IV and H α BH masses, and the vertical line demarcates the ‘high’ and ‘low’ C IV blueshift regimes discussed in the text. The BH masses of quasars with moderate C IV blueshifts are underestimated when using the C IV FWHM, while the masses of quasars with large blueshifts are severely overestimated. This situation cannot be corrected by changing the exponent on the FWHM (e.g. Rafiee and Hall, 2011; Park et al., 2013) or the overall scaling in standard virial BH mass relations.

(e.g. H β , Mg II, C IV) and different measures of the line width (FWHM or dispersion) (e.g. Vestergaard, 2002; McLure and Jarvis, 2002; Vestergaard and Peterson, 2006; McGill et al., 2008; Wang et al., 2009; Rafiee and Hall, 2011; Park et al., 2013).

Reverberation mapping measurements of nearby AGN have revealed the BLR to be stratified, with high-ionisation lines, including C IV, emitted closer to the BH than low-ionisation lines, including H α and H β (e.g. Onken and Peterson, 2002). Vestergaard and Peterson, (2006) found that the C IV-emitting region is at approximately half the radius of the H β /H α emitting region. Given the $\Delta V \propto R_{\text{BLR}}^{-0.5}$ virial relation, this leads to the prediction that the C IV line widths should be $\simeq 1.4$ times broader than H α for a given BH mass. More recently, Denney, (2012) found that there is a significant contribution from gas at larger radii to the C IV emission line, enhancing the profile at lower-velocity and leading to smaller FWHM or dispersion values. The ratio of the line widths is therefore predicted to be lower than the factor of $\simeq 1.4$.

We calculate virial BH mass estimates from C IV, H α and H β using the widely-adopted Vestergaard and Peterson, (2006) scaling relations (their equations 5 and 7 for H α /H β and C IV respectively). In Figs. ??c and d the C IV-based estimates are compared to the H α /H β estimates as a function of the C IV blueshift. There is a strong systematic error in the C IV-based masses as a function of blueshift, which is a direct consequence of the FWHM trend described in the previous section. The C IV emission-based BH-masses are in error by a factor of more than five at 3000km s^{-1} in C IV emission blueshift and the overestimate of the BH-masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts, $\gtrsim 5000\text{km s}^{-1}$.

The virial product is the product of the virial velocity squared and the BLR radius (e.g. Shen, 2013), and is proportional to the BH mass. We use the corrected C IV FWHM given by Eq. 2.5 as an indicator of the virial velocity, and adopt the same $R - L$ relation for the 1350 \AA continuum luminosity as Vestergaard and Peterson, (2006) (i.e. $R \propto L^{0.53}$). To find the constant scaling factor necessary to transform the virial product in to a BH mass we compute the inverse-variance weighted mean difference between the virial products and the H α -based masses. The virial BH mass can then be expressed in terms of the corrected C IV FWHM and monochromatic continuum luminosity at 1350 \AA

$$\text{MBH}(\text{C IV, Corr.}) = 10^{6.71} \left(\frac{\text{FWHM}(\text{C IV, Corr.})}{10^3 \text{ km s}^{-1}} \right)^2 \left(\frac{-L_{\text{c}}(1350\text{ \AA})}{10^{44} \text{ erg s}^{-1}} \right)^{0.53} \quad (2.8)$$

Given measured C IV emission line FWHM and blueshift, equations 4 and 6 can then be used to provide an unbiased estimate of the quasar BH mass.

In Section 2.3.9 we have shown that the C iv emission at large C iv-blueshift is not dominated by virial-induced motions due to the BH. The empirically derived increase in the C iv emission FWHM with blueshift leads directly to an overestimate of BH-mass if the trend with blueshift is not taken into account. The availability of the H α -spectra for the sample allows the quantification of the bias in inferred BH-mass under the assumption that the H α emission line provides a reliable BH mass.

Figure 2.15 shows the ratio of the C iv- and H α -FWHM derived BH masses as a function of the C iv blueshift. We see that for quasars with C iv blueshifts $> 2000 \text{ km s}^{-1}$, the C iv-based masses overestimate the H α -based masses by as much as a factor of ~ 5 .

The existence of a trend in the C iv-dispersion values with C iv blueshift is evident from inspection of the bottom left panel of Fig. 2.6 but the systematic trend relative to the spread at fixed blueshift is significantly smaller than when using C iv FWHM. Similarly, Fig. 2.15 shows, at most, only a weak increase in the ratio of C iv- and H α -derived masses. Without knowledge of the C iv blueshifts the distribution of C iv- and H α -dispersion based BH masses could be taken to be reassuring. Including the C iv-blueshift information, however, demonstrates that any such interpretation is inherently flawed because the origin of the C iv emission velocity width is not due to virial-motions for a significant range of C iv blueshift. To reiterate the point made above (Sec. 2.3.9), we believe that using a greater knowledge of the line profile (i.e. both the FWHM and blueshift) is a better motivated (and more practical) approach to obtaining more reliable virial BH mass estimates from the C iv line.

The number of objects in our sample is small but an important factor contributing to the significant correlation evident in the FWHM version of Fig. 2.15 is a change in the emission-line shape of H α as the C iv-blueshift increases. By comparing the distributions of the H α FWHM and dispersion as a function of C iv-blueshift (shown in the right-panels of Fig. 2.6), there is trend for the H α lines to become peakier (with FWHM/ σ approaching unity) as the C iv blueshift increases. Whether the size of the true systematic bias in BH masses inferred from C iv-emission FWHM is as large as shown in Fig. 2.15 will depend on the future parametrization of the reverberation-component present in H β (and H α) profiles for quasars with high luminosities and large C iv blueshifts.

In summary, Fig. 2.15 illustrates the extent to which key derived physical parameters, including the BH mass and L/L_{Edd} , could be systematically in error when C iv-FWHM measures are used without incorporating the information from the C iv blueshifts. Other authors have proposed empirical corrections to C iv-based BH masses based on similar systematic trends seen in the C iv line shape (Denney, 2012) and the continuum-subtracted peak flux ratio of the ultraviolet

emission-line blend of Si IV+O IV (at 1400 Å) to that of C IV (Runnoe et al., 2013a). In Section 2.6.1 we apply these corrections the quasars in our sample, and discuss the effect they have on the systematic bias seen in Fig. 2.15.

2.5 PRACTICAL APPLICATION OF THE C IV-BASED CORRECTION TO VIRIAL BH-MASS ESTIMATES

2.5.1 Recipe for unbiased C IV based BH masses

Equations 4 and 6 together provide an un-biased estimate of the virial BH mass given the FWHM and blueshift of C IV, together with the continuum luminosity at 1350 Å. The FWHM is readily obtained, either directly from the data, or, via the fitting of a parametric model to the C IV-emission line. The blueshift – defined as the bisector of the cumulative line flux – is also straightforward to measure and our preferred procedure is described in Section 2.3.1. The only potential complication arises in establishing the quasar systemic redshift and hence defining the zero-point for the C IV-blueshift measurement, since both the blueshift and the systemic redshift cannot be determined from C IV alone. In practice, when rest-frame optical lines are accessible, as is the case for the quasar sample here, an accurate systemic redshift can be obtained. The [O III] doublet and the Balmer lines all have velocity centroids very close to systemic, and the same is true for the broad Mg II doublet. For quasars at very high redshifts, $z \sim 6$, systemic redshifts can also be derived using the [C II] 158 μm emission in the sub-millimetre band (e.g. Venemans et al., 2016). However, in general, for example in determining the BH-masses of quasars at redshifts $z > 2$, if only the rest-frame ultraviolet region is available determining a reliable systemic redshift is non-trivial.

Shen et al., (2016) and our own work shows that there is an intrinsic variation of $\sigma \simeq 220 \text{ km s}^{-1}$ in the velocity centroids of the broad-line region relative to a systemic-frame defined by the quasar narrow-line regions. As we showed in Paper I, the SDSS DR7 pipeline redshifts are not sufficiently reliable to measure the C IV blueshift accurately because, in part, the C IV emission line itself contributes to the determination of the quasar redshifts (see figure 1 in Paper I). The redshift-determination scheme of Hewett and Wild, (2010) provided much improved redshifts for the SDSS DR7 quasar catalogue, not least because the redshift estimates for the majority of quasars were derived using emission-lines other than the C IV-line itself. The redshifts for quasars in the SDSS DR10 and DR12 catalogues (Pâris et al., 2014; Pâris et al., 2017) possess errors of $\simeq 500\text{--}750 \text{ km s}^{-1}$ (Pâris et al., 2012; Font-Ribera et al., 2013). The impact of low spectrum S/N for fainter quasars in all the SDSS data releases increases the uncertainty further. Table 2.4 includes the values for the fractional error in the corrected

Table 2.4: The fractional error on the corrected BH mass as a function of C iv blueshift for different uncertainties in the quasar systemic redshift.

δv (km s ⁻¹)	C iv blueshift (km s ⁻¹)			
	0	1000	2000	4000
250	0.33	0.20	0.14	0.09
500	0.65	0.39	0.28	0.18
1000	1.30	0.79	0.57	0.36

BH-mass that result from a given error in the determination of the systemic rest-frame. For example, the fractional error in the corrected BH mass is 0.39 for a quasar with a 1000 km s⁻¹ C iv blueshift when there is a 500 km s⁻¹ uncertainty in the quasar systemic redshift.

Of potentially more significance for studies of BH-masses as a function of quasar and host-galaxy properties are redshift errors that depend on the form of the quasar ultraviolet SED. The redshifts from Hewett and Wild, (2010) still suffer from systematic errors that are correlated with the shape, and particularly the blueshift, of the C iv emission line. For the Hewett and Wild, (2010) redshifts, and ultraviolet emission-line based redshifts in general, quasars with large C iv EW and modest blueshifts have relatively small ($\simeq 300$ km s⁻¹) SED-dependent redshift errors. Redshift uncertainties as large as $\simeq 1000$ km s⁻¹ for such quasars are unusual and the large relative error in the corrected C iv BH-mass given in Table 2.4 is pessimistic.

Conversely, systematic redshift errors are greatest for quasars with large blueshifts, reaching ~ 750 km s⁻¹ in the extreme for the Hewett and Wild, (2010) values. The associated error in the corrected C iv BH-masses is, however, mitigated somewhat due to the smaller gradient of the MBH(C iv)/MBH(Balmer) relation at large C iv blueshift (see Fig. ??). A definitive quantification of any systematic SED-dependent errors present in the quasar redshifts contained in the SDSS DR12 catalogue is not yet available but the principal component analysis (PCA) based redshift estimates are expected to be largely free of SED-dependent systematics. Given the importance of generating more accurate redshifts for the SDSS DR7 and DR12 quasar samples we will publish a catalogue of more accurate redshifts in due course (see Section 2.7).

2.5.2 Systematic trends in residuals

The scatter about the best-fitting line in the C iv/H α FWHM versus C iv-blueshift relation is ~ 0.1 dex, an order of magnitude smaller than the size of the C iv-blueshift dependent systematic but, nevertheless, still significant. With a view to reducing the scatter further,

we searched for measurable parameters which correlate with the scatter at fixed C IV blueshift, including the luminosity, redshift, [O III] equivalent width (EW), and Fe II EW. The only significant correlation we find is with the H α FWHM (Fig. 2.16). Quasars with broad H α lines tend to lie below the relation while quasars with narrow H α tend to lie above it. One possibility is that this correlation is simply due to random scatter (either intrinsic or measurement error) in the H α FWHM which, with the other quasar properties fixed, would naturally produce a correlation between FWHM(C IV)/FWHM(H α) and FWHM(H α). However, the fact that we see no such correlation between the model residuals and the C IV FWHM suggests that the H α FWHM correlation could be revealing something more fundamental. The H α /H β FWHM is part of ‘eigenvector 1’ (EV1), the first eigenvector in a principal component analysis which originated from the work of Boroson and Green, (1992). While a number of parameters have been considered within the EV1 context (e.g. Brotherton and Francis, 1999), Fig. 2.16 suggests that part of the scatter between the Balmer and C IV velocity widths might be attributed to differences in the spectral properties which are correlated with EV1 (Marziani et al., 2013).

The residuals and the H α FWHM also correlate with the shape of the line (FWHM/ σ , where σ is the dispersion, derived from the second moment velocity; e.g. Kollatschny and Zetzl, 2011; Kollatschny and Zetzl, 2013). The narrow lines are, on average, ‘peakier’ (with FWHM/ $\sigma \simeq 1$) than the broader lines (with FWHM/ $\sigma \simeq 2$). The origin of the Balmer-line shape correlation is not clear but one possibility is an orientation-dependence of the H α FWHM (e.g. Shen and Ho, 2014). In this scenario quasars with broader emission lines are more likely to be in an edge-on orientation relative to our line of sight.

At radio wavelengths, the morphology of the radio structure, parametrized in terms of ‘core dominance’ is believed, at least in a statistical sense, to be a proxy for the orientation of the accretion disk (e.g. Jackson and Browne, 1991). We matched our sample to the FIRST radio catalogue (White et al., 1997) in an attempt to identify orientation-dependent signatures. Following Shen et al., (2011), we classified quasars with matches within 5 arcseconds as core-dominated, while, if multiple matches were found within 30 arcseconds, quasars were classified as lobe-dominated. Twenty core- quasars and six lobe-dominated quasars resulted but no statistically significant differences in the H α line-widths of the two samples were found. It should be noted that the sub-sample of radio-detected quasars is small and the effectiveness of the test is further compromised by the lack of radio-detected quasars at large blueshifts (see figure 14 of Richards et al., 2011, for example).

There are currently very few reverberation-mapping measurements of quasars with large C IV blueshifts. Looking to the future, the results of the large on-going statistical reverberation mapping projects

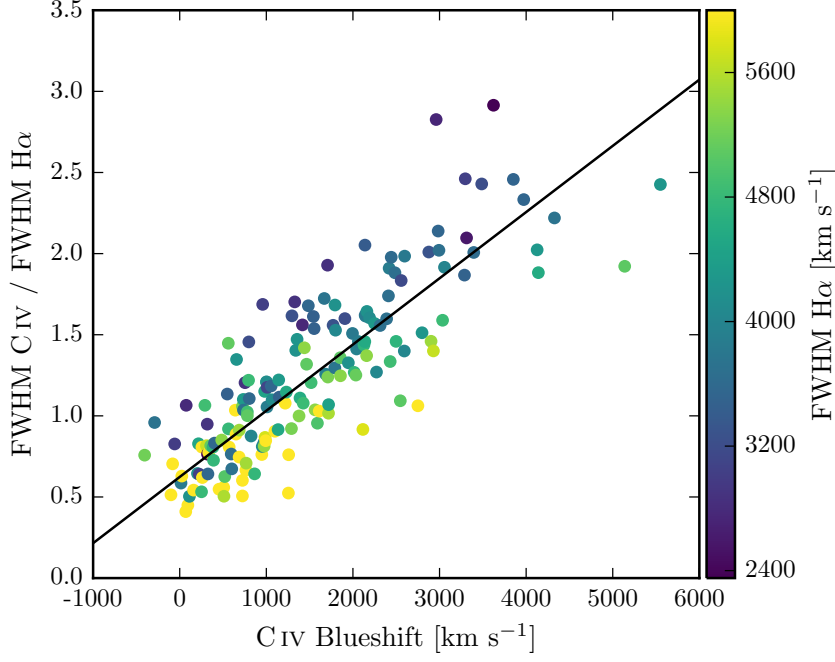


Figure 2.16: Same as Fig. ??a, with the marker colour representing the $H\alpha$ FWHM. At fixed C IV blueshift, there is a clear $H\alpha$ FWHM dependent systematic in the model residuals.

(e.g. Shen et al., 2015; King et al., 2015) for luminous quasars at high-redshift will shed new light on the Balmer line emitting region of the BLR for quasars with a range of C IV blueshifts and lead to a greater understanding of the relation between the Balmer line profile and the BH mass.

2.5.3 Effectiveness of the C IV blueshift based correction to BH masses

Figure 2.23 demonstrates that our sample has an excellent coverage of the EW-blueshift parameter space in relation to SDSS DR7 quasars at redshifts $1.6 < z < 3.0$. The systematic offset to higher C IV blueshifts for our catalogue relative to the SDSS quasars as a whole is a result of the higher mean luminosity relative to the SDSS sample (Fig. 1.1). Our sample includes 21 quasars with C IV blueshifts $> 3000 \text{ km s}^{-1}$, and extends to $\sim 5000 \text{ km s}^{-1}$, i.e. at the very extreme of what is observed in this redshift and luminosity range. Our investigation thus demonstrates that the C IV-blueshift based correction derived in this paper is applicable to very high blueshifts. Conversely, there are no quasars in our catalogue with C IV blueshifts $\lesssim 0 \text{ km s}^{-1}$ and we caution against extrapolating the correction formula to negative blueshifts.

Figure 2.18 compares the C IV- and $H\alpha$ -based BH masses before and after applying the blueshift-based correction to the C IV FWHM. Before the correction, the correlation between the C IV- and $H\alpha$ -based

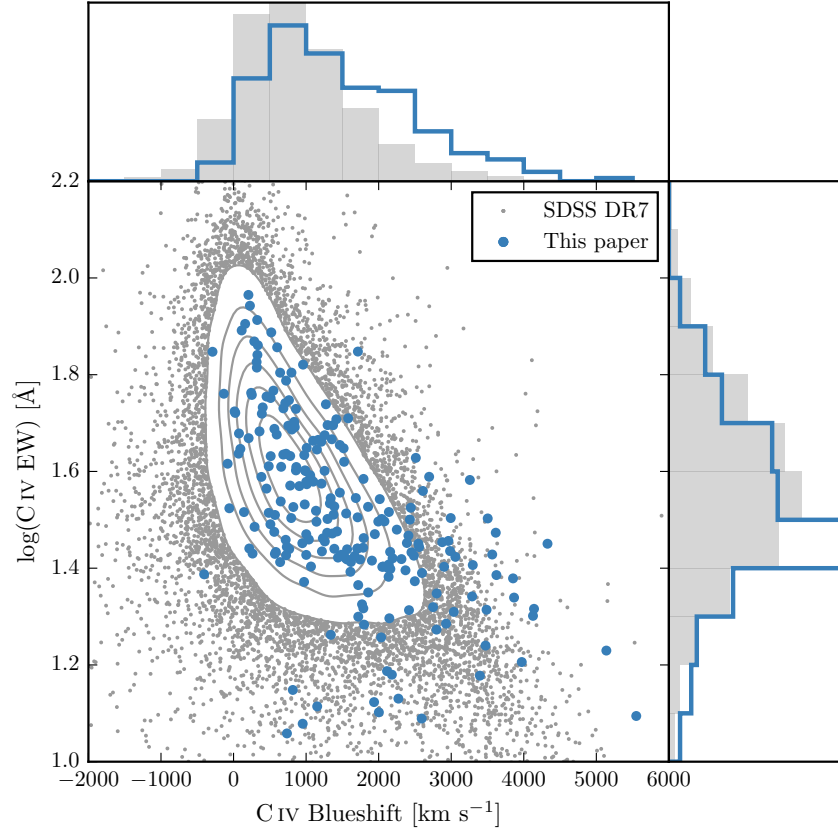


Figure 2.17: Rest-frame EW versus blueshift of the broad C iv-emission line for 32,157 SDSS DR7 quasars at $1.6 < z < 3.0$ (grey) and our sample (blue). For the SDSS quasars, the systemic redshifts used to calculate the blueshifts are from Hewett and Wild, (2010) and C iv emission properties are described in Paper I. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator. Our sample has very good coverage; the shift to high blueshifts is a result of the high luminosity of our sample in relation to the SDSS sample and the correlation between luminosity and blueshift.

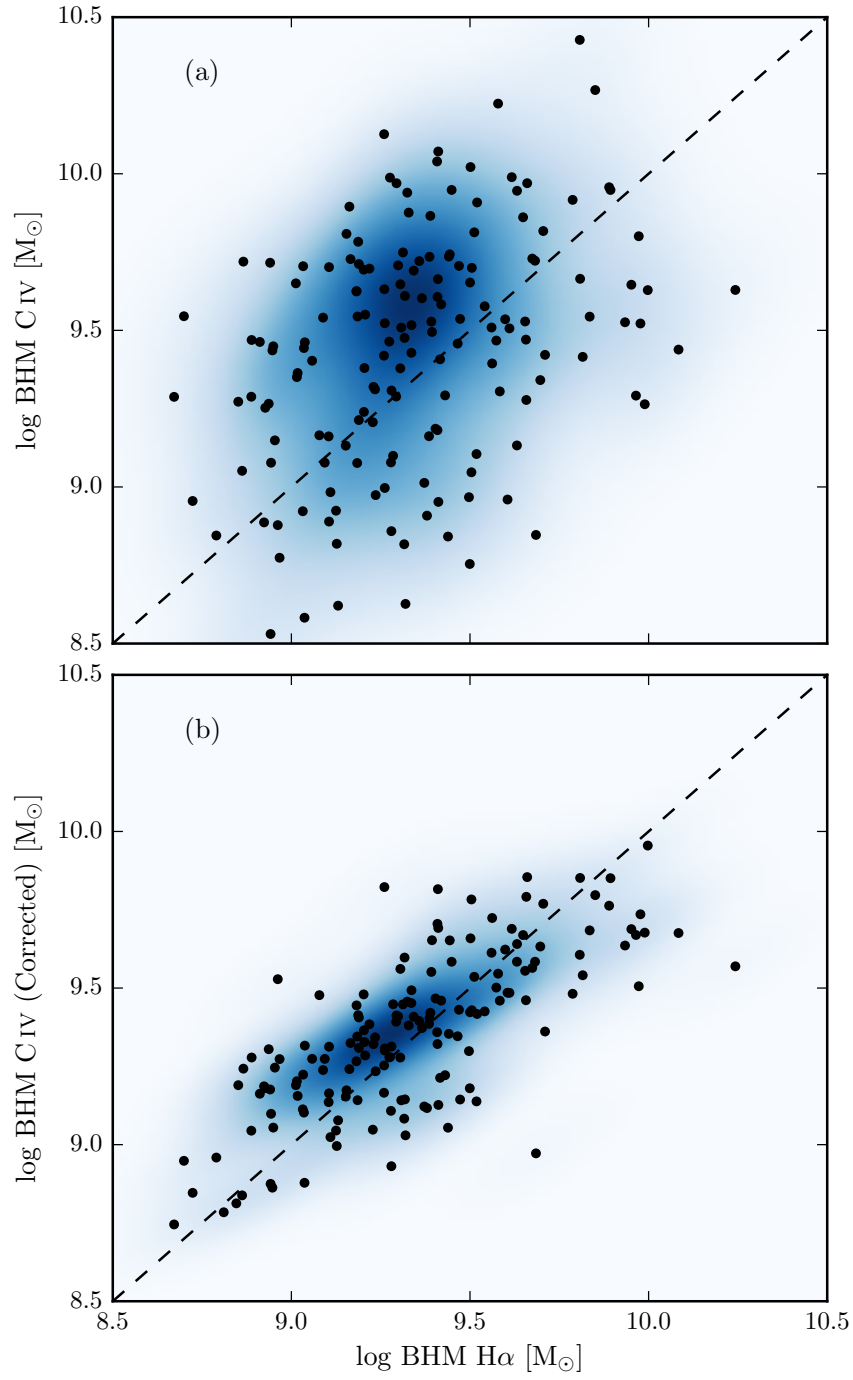


Figure 2.18: Comparison of the C iv- and H α -based BH masses before (a) and after (b) applying the C iv blueshift-based correction to the C iv FWHM. The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. The correction to the C iv BH masses decreases the scatter by from 0.4 to 0.2 dex.

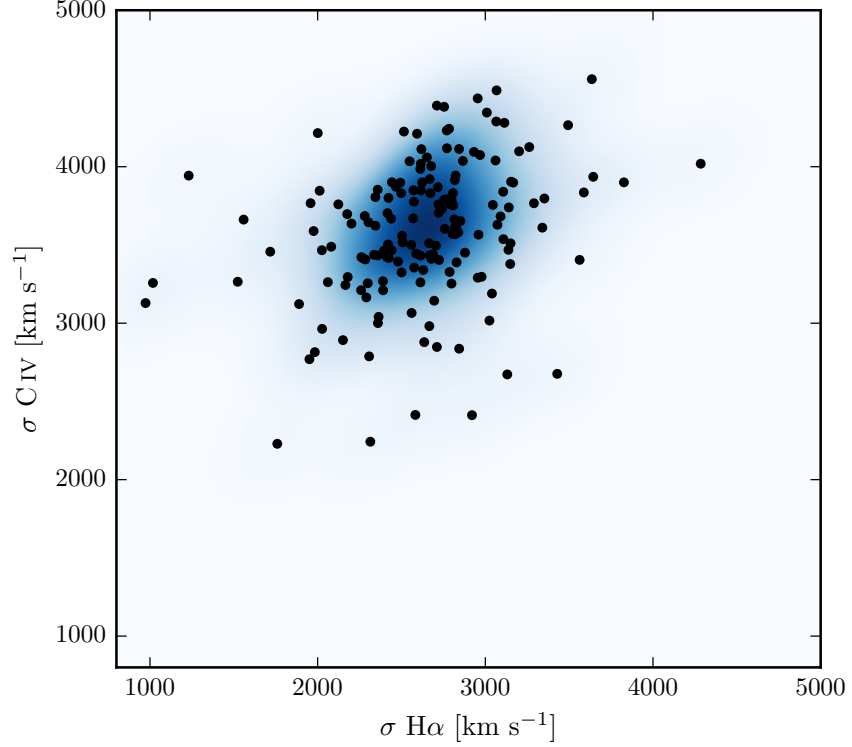


Figure 2.19: Comparison of the C IV and H α line dispersion, σ . The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. Estimating a reliable BH mass from the C IV FWHM and blueshift line is substantially more effective than using the C IV line dispersion with, or without, the line blueshift. The C IV dispersion values are larger than the corresponding H α measurements by a factor of 1.4 on average, which is consistent with reverberation mapping measurements (Vestergaard and Peterson, 2006).

BH masses is very weak, and the scatter between the masses is 0.4 dex. After correcting the C IV FWHM for the non-virial contribution, the correlation improves dramatically. The scatter between the corrected C IV-based masses and the H α -based masses is reduced to 0.2 dex. The scatter is 0.24 dex at low C IV blueshifts ($\sim 0 \text{ km s}^{-1}$) and 0.10 dex at high blueshifts ($\sim 3000 \text{ km s}^{-1}$).

There has been a considerable amount of attention regarding the relative merits of using the FWHM or dispersion to characterise the velocity width (e.g. Denney et al., 2013). As we showed in Paper I, the line dispersion is relatively insensitive to the blueshift and shape of the C IV line. Therefore, without the blueshift information, using the line dispersion would yield a more accurate BH mass than the FWHM (Fig. 2.19). The correlation between the H α and C IV line dispersion is, however, weak. The Pearson coefficient for the correlation is 0.36 (and just 0.15 when the H β measurements are used in place

of $H\alpha$). Furthermore, there is little dynamic range in the line dispersion: the scatter is just 480 and 460 km s^{-1} for $H\alpha$ and C iv respectively. The observation suggests that the line dispersion does not fully trace the dynamic range in BH mass present in the quasar population. At least part of the reason is that the line dispersion is difficult to measure reliably in current survey-quality data, particularly because of the sensitivity to flux ascribed to the wings of the emission line (e.g. Mejía-Restrepo et al., 2016). Figures 2.18 and 2.19 demonstrate that estimating a reliable BH mass from the C iv FWHM and blueshift line is substantially more effective than using the C iv line dispersion with, or without, the line blueshift.

2.5.4 Comparison to previous prescriptions

In Fig. 2.20 we compare the C iv blueshift-based correction presented in this paper to various prescriptions which have been proposed in the literature to derive BH masses from the C iv line which are consistent with the masses derived from the Balmer lines. In each case we compare the corrected C iv-based masses to the $H\alpha$ -based masses as a function of the C iv blueshift. The correction proposed by Runnoe et al., (2013a) is based on the spectral region at rest-frame wavelengths of $\sim 1400 \text{ \AA}$ (see below). Therefore, our analysis is based on the 123 quasars which satisfy this requirement.

In Fig 2.20a the C iv BH masses have been corrected using the C iv shape (FWHM/ σ) based correction proposed by Denney, (2012). The correction is not applicable at large C iv blueshifts, since it was calibrated on a sample of low-luminosity AGN which does not include any such objects. Therefore, while the consistency between the $H\alpha$ - and C iv-based masses at low C iv blueshifts is improved, at high C iv blueshifts the C iv-based masses remain seriously overestimated.

Runnoe et al., (2013a) used the continuum-subtracted peak flux ratio of the ultraviolet emission-line blend of Si iv+O iv (at 1400 \AA) relative to C iv to correct for non-virial contributions to the C iv velocity-width. Following Runnoe et al., (2013a), we measure the peak flux by fitting a model with four Gaussian components (two for each emission line) to the continuum-subtracted flux. As is evident from Fig. 2.23, a correlation exists between the blueshift and equivalent width of C iv: C iv emission which is strongly blueshifted is typically weak. The Si iv+O iv emission-line blend, however, shows significantly less systematic variation. Therefore, the Si iv+O iv-based correction is quite effective in practice: the systematic bias in the C iv BH masses at large C iv blueshifts is reduced to a factor of ~ 2 (Fig. 2.20b). However, the C iv based masses are still systematically overestimated at large C iv blueshifts.

In contrast to the widely-used Vestergaard and Peterson, (2006) C iv-based virial BH mass calibration, the more recent Park et al.,

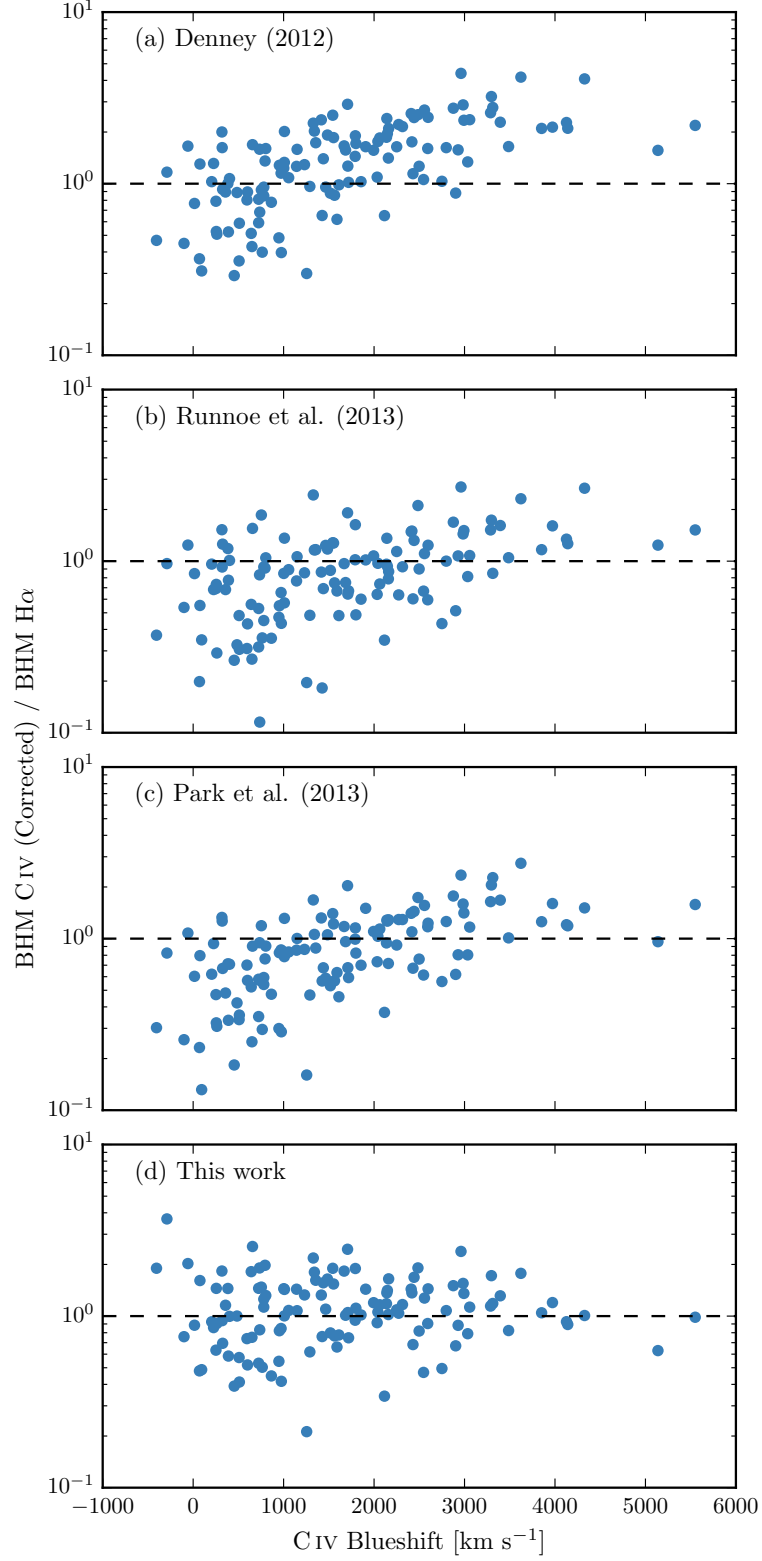


Figure 2.20: Comparison of BH mass estimates derived from C IV and H α as a function of the C IV blueshift. Corrections to the C IV-based masses have been applied based on the shape (FWHM/ σ) of the C IV emission line (a; Denney, 2012), the peak flux ratio of the Si IV+O IV blend relative to C IV (b; Runnoe et al., 2013a), by significantly reducing the dependence of the derived BH mass on the C IV velocity-width (c; Park et al., 2013), and based on the C IV blueshift (d; this paper).

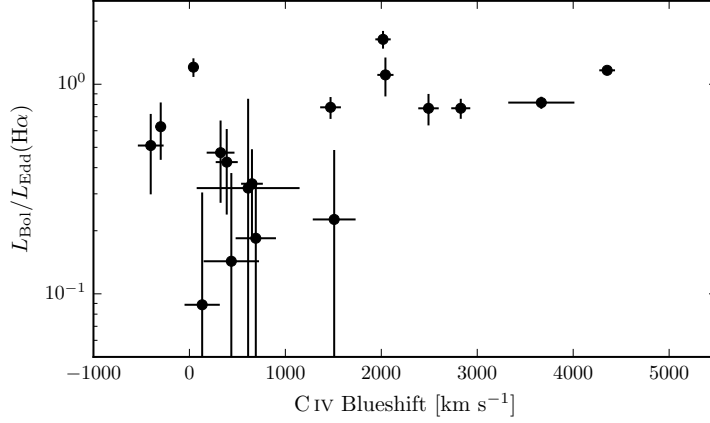


Figure 2.21: $H\alpha$ -derived Eddington ratio versus C IV blueshift. At blueshift $\gtrsim 2000 \text{ km s}^{-1}$ all quasars have high accretion rates ($L/L_{\text{Edd}} \simeq 1$). This is in agreement with Kratzer and Richards, (2015), but in contrast to what one would derive from naive use of C IV-based BH mass scaling relations.

(2013) calibration significantly reduces the dependence of the derived masses on the emission-line velocity width (from the V^2 dependence predicted assuming a virialized BLR to just $V^{0.56}$). As a consequence, the C IV based masses of the quasars with large C IV blueshifts are much reduced (Fig. 2.20c). However, the systematic error in the C IV-based BH masses as a function of C IV blueshift remains.

An alternate virial BH-mass calibration is proposed by Park et al., (2013), using an improved sample of AGN with reverberation mapped masses. A major difference from the calibration of Vestergaard and Peterson, (2006) is that Park et al., (2013), recognising the poor correlation sometimes observed between the C IV and $H\beta$ FWHM, allow the exponent on the velocity width (b in Eq. 2.7) to vary. Calibrating Eq. 2.7 against reverberation BH masses, they find a best-fit value of $b = 0.56$, which is much less than the $b = 2.0$ in the strict virial regime. As a result, the derived BH masses are much less sensitive to the C IV-emission line properties.

As a comparison, the C IV-based masses shown in Fig 2.20d have been corrected using to the C IV blueshift-based procedure presented in this paper. No systematic in the BH masses as a function of the C IV blueshift is evident.

2.5.5 Population trends with C IV blueshift

Even with the caveat regarding the small sample size, the differences in the $H\alpha$ emission-profile as a function of C IV-blueshift (Fig. ??) appear to be systematic. At C IV-blueshift $< 1200 \text{ km s}^{-1}$, the $H\alpha$ FWHM range is $\simeq 2700 - 8800 \text{ km s}^{-1}$, with mean $\simeq 5200 \text{ km s}^{-1}$. However,

amongst the six quasars with C iv-blueshift $>2000 \text{ km s}^{-1}$, the mean H α FWHM $=3000 \text{ km s}^{-1}$, with a scatter of just 200 km s^{-1} . The apparent trend of peakier H α -emission, with FWHM/ σ close to unity, at large C iv-blueshift is enhanced by the modest increase in H α EW with blueshift (Table ??). Amongst the low-C iv-blueshift population there are in addition quasars with broader and more Gaussian-like H α line profiles, with FWHM/ $\sigma \simeq 2$.

The change in the H α emission-line profiles as a function of C iv-blueshift means that the H α -FWHM derived BH masses at high-blueshift are smaller than the sample mean. We transformed the observed luminosity into a mass-normalised accretion rate (Eddington ratio). To convert the monochromatic luminosity, which is observed, in to a bolometric luminosity we use the bolometric correction factor given by Richards et al., (2006) ($L_{\text{bol}} = 9.26 L_{5100}$). Although there is evidence that the bolometric correction factor is a function of the luminosity, as well as of other parameters including the C iv blueshift (Krawczyk et al., 2013), the differences are small over the parameter range covered by our sample, and for simplicity we adopt a constant factor.

The results, shown in Fig. 2.21, show that at large blueshifts quasars are accreting at around their Eddington limits (Fig. 2.21). The converse is, however, not true, i.e. not all quasars with high Eddington ratios possess large C iv blueshifts (see Baskin and Laor, 2005).

2.6 DISCUSSION

2.6.1 Biases in single-epoch C iv-based BH-mass estimates

The C iv line profiles of the quasars with the largest C iv blueshifts (in the bottom right of Fig. 2.23) demonstrate that non-virial motions are having a significant effect on the C iv BLR dynamics. At fixed emission-line EW, almost the entire C iv-profile appears to shift blueward and the change in line shape is not simply an enhancement of flux in the blue wing of a symmetric component. It is clear that the virial assumption, on which single-epoch BH-mass measurements are predicated, is not straightforwardly applicable for the C iv-emission line in quasars exhibiting large blueshifts.

Quantitatively, the C iv-shape change is most apparent from the FWHM values, which are strongly correlated with the C iv-blueshift. This trend has previously been identified, by comparison with Mg II at lower-redshifts (Shen et al., 2008; Shen et al., 2011) and H β at higher redshifts (Shen and Liu, 2012). We find that virial BH mass estimates based on the C iv FWHM will overestimate the true mass by a factor of ~ 5 for objects exhibiting the largest C iv blueshifts.

In contrast, the C iv line dispersion does not show a similarly strong dependence on the blueshift. This is a result of the shape of the

C iv line profile being dependent on the blueshift; the low-blueshift profiles are peakier ($\text{FWHM}/\sigma \simeq 1$) than the high-blueshift profiles ($\text{FWHM}/\sigma \simeq 2$). Denney, (2012) found the level of contamination in single-epoch spectra from non-reverberating gas to be correlated with the shape (FWHM/σ) of the C iv profile. Their investigation was based on a sample of reverberation mapped quasars, which have a narrow range of C iv-emission line shapes, including the absence of any objects with large C iv blueshifts. The FWHM/σ -based correction to the C iv FWHM proposed by Denney, (2012) increases the inferred BH masses by ~ 0.2 dex at the low end of our C iv blueshift distribution, thereby reducing part of the systematic trend in the BH mass (Fig. 2.22). However, it is not applicable at the high C iv blueshift end, where velocity widths are likely dominated by non-virial motions. Based on the typical line shape of C iv in these high-blueshift quasars ($\text{FWHM}/\sigma \simeq 2$), the Denney, (2012) correction decreases the predicted BH masses by ~ 0.3 dex, which has only a moderate decreasing effect on the strong systematic (Fig. 2.22).

While the C iv-line dispersion is largely independent of the blueshift, it does not follow that dispersion-based BH-mass estimates are correct, because the underlying assumption regarding the virial-origin of the C iv emission profile breaks down at large blueshifts. Furthermore, given the difficulty in obtaining reliable dispersion measurements from survey-quality spectra with limited S/N (e.g. Mejía-Restrepo et al., 2016), virial BH-mass estimates for existing large samples of high-redshift quasars are usually based on the FWHM (e.g. Shen et al., 2011). Our work therefore suggests that a viable recipe for obtaining more reliable BH mass estimates for large numbers of quasars at high redshift is to measure both the FWHM and the blueshift, which together can be used to derive a FWHM corrected for the non-virial contribution.

Although we do not have enough quasars in our sample to derive a reliable quantitative correction to BH-masses as a function of C iv blueshift, we are assembling a large sample of quasar spectra with coverage of both C iv and $\text{H}\beta/\text{H}\alpha$ in order to empirically validate such a correction. Both Runnoe et al., (2013a) and Shen and Liu, (2012) considered a similar approach, but concluded that a low-ionisation broad line (e.g. Mg II), or features from the quasar NLR, are required to determine the systemic redshift and hence the C iv blueshift. Empirical tests of the reliability of the improved Hewett and Wild, (2010) redshifts for the SDSS DR7 quasars (Shen et al., 2016) and the availability of the largely SED-independent principal component analysis redshifts for DR10 and DR12 (Pâris et al., 2014; Pâris et al., 2017) already allow meaningful corrections to BH-mass estimates for quasars exhibiting large C iv-blueshifts. Our intention is, however, to determine a definitive correction formula using the redshifts from Allen & Hewett (2017, in preparation) for both DR7 and DR12.

Given the difficulty of measuring reliable C iv blueshifts, Runnoe et al., (2013a) opted instead to use the continuum-subtracted peak flux ratio of the ultraviolet emission-line blend of Si iv+O iv (at 1400 Å) to that of C iv to correct for non-virial contributions to the C iv velocity width. This parameter was chosen because it showed the strongest correlation with the FWHM C iv/H β residuals, as well as with the strengths of optical O iii and Fe ii. The strengths of optical O iii and Fe ii, being parameters in EV1, are also correlated with the C iv blueshift (Sulentic et al., 2007). As the C iv blueshift increases the EW decreases systematically (Fig. 2.23). The Si iv+O iv emission-line blend, however, shows significantly less systematic variation. Therefore, while the Runnoe et al., (2013a) Si iv+O iv-based correction is effective in practice, the C iv blueshift measurement provides a more direct measure of the non-virial contributions to the C iv velocity width.

In our sample, we find the 1400 Å/C iv peak flux ratio to be strongly correlated to the C iv blueshift (the Spearman coefficient for the correlation is 0.82, p-value 2E-5). As such, the correction to C iv-based virial masses proposed by Runnoe et al., (2013a, their equation 3) removes a large part of the systematic in the H α /C iv FWHM residuals with the C iv blueshift (Fig. 2.22); the median C iv/H α FWHM ratio at large C iv blueshifts ($>1200\text{km s}^{-1}$) is reduced from 4.5 to 1.5. However, at low ($<1200\text{km s}^{-1}$) C iv blueshifts, the trend for C iv to predict lower BH masses persists, and the scatter between the C iv- and H α -based masses increases by a factor of two. In accordance with our expectations, we find the FWHM C iv/H α residuals to be more tightly correlated to the C iv blueshift (Spearman coefficient 0.82, p-value 3E-5) than to the 1400 Å/C iv peak flux ratio (Spearman coefficient 0.72, p-value 7E-4).

2.6.2 Possible systematic trends in H α BH-mass estimates

In Section 2.5.5, we found that the quasars with large C iv blueshifts have systematically narrower H α FWHM (Fig. 2.6). Using the standard virial BH mass calibrations this implies that the high C iv blueshift population have high accretion rates ($L/L_{\text{Edd}} \simeq 1$; Fig. 2.21). This interpretation requires some caution since the emission-line shape (characterized by the value of FWHM/σ) of H α is also changing as a function of the C iv blueshift (Fig. 2.6). At low C iv blueshifts there are a range of shapes, but all of the quasars exhibiting large C iv blueshifts have peaky H α profiles with $\text{FWHM}/\sigma \simeq 1$. This raises the question of whether the H α FWHM is a reliable proxy for the virial-induced velocity dispersion for the full range of H α line shapes we have in our sample.

When calibrating the virial-product to masses derived independently using the BH mass – stellar velocity dispersion ($M_{\odot} - \sigma$) relation, Collin et al., (2006) find that the scaling factor, f , is a factor ~ 2

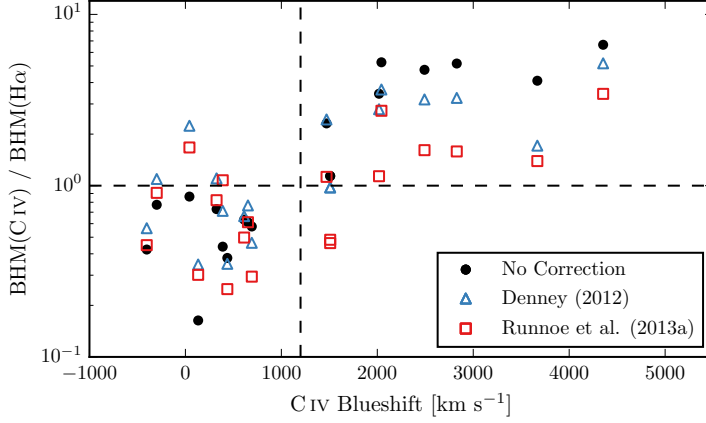


Figure 2.22: Comparison of BH mass estimates derived from the FWHM of C IV and H α as a function of the C IV blueshift (*black circles*), and after applying corrections to the C IV-derived mass based on the line shape (*blue triangles*; Denney, 2012) and the peak flux ratio of the Si IV+O IV blend relative to C IV (*red squares*; Runnoe et al., 2013a). While the shape-based correction improves the consistency between BH mass estimates in the low-blueshift population, only the Si IV+O IV/C IV peak flux-based correction is effective at high blueshifts (although a weak systematic remains).

larger for their Population ‘1’ sources (with $\text{FWHM}/\sigma < 2.35$ and essentially equivalent to population A of Sulentic and co-workers and to the high-blueshift quasars here) than for their Population 2 (with $\text{FWHM}/\sigma > 2.35$). For single-epoch BH-mass estimates, assuming a constant value of f , as is normally done (e.g. Vestergaard and Peterson, 2006), means that Population 1 masses will be underestimated and Population 2 will be overestimated. In the context of this result from Collin et al., (2006), our high-blueshift objects all possess peaky H α -lines and, while our quasar sample probes much higher luminosities and masses, the true BH-masses may also be underestimated. Adopting such an interpretation, the amplitude of the trend seen in Fig. 2.15 might not be so pronounced.

As mentioned in Section ?? and discussed in Richards et al., (2011), quasars with current reverberation mapping measurements have a restricted range of C IV-line shapes. There are currently very few reverberation-mapping measurements of quasars with large C IV blueshifts but the results of the large on-going statistical reverberation mapping projects (e.g. Shen et al., 2015) for luminous quasars at high-redshift will go some way to establishing whether the quasar broad line regions producing Balmer emission look the same for objects with very different C IV-emission blueshifts.

Although the EV1-trends (Sulentic et al., 2000; Shen and Ho, 2014) are most likely driven by the accretion rate, orientation may also have a role to play in determining the observed properties of the BLR. Shen

and Ho, (2014) argue that a large part of the scatter observed in the $H\beta$ FWHM relates not to a spread in BH masses, but rather to the orientation of the BLR relative to the line-of-sight. For this to be true, the BLR would need to be in a flattened disc-like geometry, in which case the observed line width would increase with the inclination of the disc relative to the line of sight. Brotherton, Singh, and Runnoe, (2015) found that the core-dominance of radio-loud quasars, which is believed to be a reliable proxy for orientation, at least in a statistical sense, is significantly correlated with the $H\beta$ FWHM and hence with the BH-mass estimates. This raises the question of whether the narrow $H\alpha$ emission lines observed in the quasars with the largest C iv blueshifts could be an orientation effect. However, there is no evidence that the C iv blueshift is dependent on the orientation (inferred from the radio core-dominance; Richards et al., 2011; Runnoe et al., 2014). Furthermore, Leighly, (2004) showed that the He II $\lambda 1640$ emission-line properties of quasars with large C iv blueshifts are more consistent with differences in the SED rather than differences in the orientation. Collin et al., (2006) showed that orientation effects were also sub-dominant to the Eddington ratio in determining the shape of the $H\beta$ line and the $H\alpha$ line shape trend we observe is consistent with the finding of Marziani et al., (2003) that the $H\beta$ emission profiles of high/low Eddington ratio low- z quasars and type 1 Seyfert nuclei are well fit by Lorentzian and double Gaussian profiles respectively. Overall, therefore, orientation does not appear to be the dominant effect in determining the C iv blueshift and correlated changes in the $H\alpha$ line profile.

2.6.3 Accretion-rate trends in the quasar population

The blueshifting of C iv is usually interpreted as evidence for strong outflows resulting from the presence of a radiation-driven accretion-disc wind. Richards et al., (2002) found that quasars with large C iv blueshifts have weak He II. This is evidence for weak soft X-ray continuum emission (Leighly, 2004), which would allow a strong line-driven wind to form. The strength of such a wind is predicted to be related to the quasar far-ultraviolet SED, which, in turn, could be related to the mass-accretion rate. This picture is therefore consistent with our finding that the quasars with large C iv blueshifts have high accretion rates.

All of the objects in our sample which exhibit large C iv blueshifts would be classified as population A in the Sulentic et al., (2000) scheme based on the $H\alpha$ FWHM (see Section ??). Our results therefore support the idea of the Sulentic et al., (2000) A/B division being driven by the Eddington ratio, with population A sources possessing higher accretion rates. However, we also observe a number of quasars which have high Eddington ratios but do not have line profiles suggestive

of strong outflows in the C iv BLR. This suggests that a high accretion rate is a necessary but not sufficient condition for the existence of outflows (Baskin and Laor, 2005).

The two-dimensional nature of the C iv emission line parametrization and the apparent anti-correlation between C iv EW and C iv blueshift suggests that the quasar population exhibits a continuum of properties. As such, more accurate C iv blueshift measurements for SDSS-quasars should allow an improved mapping between the C iv-emission properties and key physical parameters of the quasars. This includes improving our understanding of the origin of quasars with exceptionally weak, blueshifted C iv emission (weak emission line quasars; Luo et al., 2015) which could be exotic versions of wind-dominated quasars (Plotkin et al., 2015).

2.6.4 *The BAL parent population*

Classical high-ionization BAL (HiBAL) quasars are also predominantly Population A objects in the scheme of Sulentic et al., (2000). There are no HiBAL quasars in our sample by design (Section ??) but it is generally accepted that quasars which show high-ionisation BALs are likely to be radiating with relatively high L/L_{Edd} (e.g. Zhang et al., 2014). We therefore propose that the subset of the quasar population that exhibits large C iv-emission blueshifts, with high-EW and narrow-H α emission lines, may be directly related to the HiBAL quasar population – perhaps even the ‘parent’ population (Richards, 2006). A prediction of such a linkage is that near-infrared observations of the rest-frame optical spectra of HiBAL quasars will show strong, relatively narrow, Balmer emission lines, very similar to those of the quasars with high C iv-blueshifts presented in this paper (see Runnoe et al., 2013b, for such a study).

2.6.5 *The frequency of quasars with high accretion rates*

Quantifying the frequency of quasars producing outflows as a function of key parameters, e.g. quasar luminosity, BH-mass, redshift,... will be important to constrain models of quasar-galaxy evolution. At fixed BH mass, the intrinsic and the observed fraction of quasars exhibiting properties that depend on the Eddington ratio can differ significantly. As an illustration, we consider the implications for the intrinsic fraction of quasars possessing large C iv blueshifts given the observed numbers in the $m_i < 19.1$ flux-limited sub-sample of the SDSS DR7 quasar catalogue, from which our quasar sample is effectively drawn (Section ??). In order to estimate the size of the selection effect, we considered the detection probability for a much-simplified quasar population. We assume that all quasars with C iv blueshifts $> 1200 \text{ km s}^{-1}$ have enhanced accretion rates relative to the ‘normal’

population (with C iv blueshifts $< 1200 \text{ km s}^{-1}$). If the accretion rate of the high-blueshift population is double the rate of the low-blueshift population (which is true in an average sense – see Fig. 2.21), then the high-blueshift population will be brighter by $\simeq 0.75$ magnitude. Under the assumption that the BH mass distribution is independent of the C iv blueshift, the high-blueshift population will then be over-represented in a flux-limited sample. To estimate the size of the bias, we need to know how many more quasars, at redshifts $2 < z < 2.5$, there are with $m_i < 19.1 + 0.75 = 19.85$ relative to $m_i < 19.1$. This is the fraction of the population which, as a consequence of having enhanced accretion rates, are boosted above the survey flux limit. The main colour-selected SDSS DR7 quasar catalogue extends only to $m_i = 19.1$ and, assuming the luminosity function is continuous⁷ we thus use the number counts at $m_i < 19.1$ and $m_i < 18.35$, which differ by a factor of $\simeq 4$.

At redshifts $2 < z < 2.5$, there are 3,834 quasars with C iv blueshifts $< 1200 \text{ km s}^{-1}$ and 2,484 with blueshifts $> 1200 \text{ km s}^{-1}$ in the SDSS DR7 $m_i < 19.1$ quasar sample, a ratio of $\sim 2:1$. The above calculation, although much idealised, suggests that the intrinsic fraction of high-blueshift quasars is a factor of four smaller than in the flux-limited sample (i.e. ~ 15 per cent of the ultraviolet-selected non-BAL quasar population).

2.7 CONCLUSIONS

The main results of this paper are as follows:

- We have analysed the spectra of 230 high-luminosity ($10^{45.5} - 10^{48} \text{ erg s}^{-1}$), redshift $1.5 < z < 4.0$ quasars for which spectra of the Balmer emission lines and the C iv emission line exist. The large number of quasars in our spectroscopic catalogue and the wide range in C iv blueshifts the quasars possess has allowed us to directly investigate biases in C iv-based BH mass estimates which stem from non-virial contributions to the C iv emission as a function of the C iv blueshift, which, in turn, depends directly on the form of the quasar ultraviolet SEDs (Richards et al., 2011).
- The C iv emission-based BH-masses are systematically in error by a factor of more than five at 3000 km s^{-1} in C iv emission blueshift and the overestimate of the BH-masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts, $\gtrsim 5000 \text{ km s}^{-1}$.
- We have derived an empirical correction formula for BH-mass estimates based on the C iv emission line FWHM and blueshift.

⁷ The luminosity function and number-counts vary only smoothly (e.g. Ross et al., 2013) for the magnitude and redshift range used here.

The correction may be applied using equations 4 and 6 in Section 2.4.3. The large SED-dependent systematic error in C iv-based BH-masses is removed using the correction formulae. The remaining scatter between the corrected C iv-based masses and the H α -based masses is 0.24 dex at low C iv blueshifts ($\sim 0 \text{ km s}^{-1}$) and 0.10 dex at high blueshifts ($\sim 3000 \text{ km s}^{-1}$). This is a significant improvement on the 0.40 dex scatter observed between the un-corrected C iv and H α BH masses. The correction depends only on the C iv line properties - i.e. the FWHM and blueshift - and allows single-epoch virial BH mass estimates to be made from optical spectra, such as those provided by the SDSS, out to redshifts exceeding $z \sim 5$.

As discussed in Section 2.5.1, uncertainties in redshift estimation and hence the definition of the systemic rest-frame for quasars impact on the accuracy of the corrected BH-masses. Using published redshift estimates, notably those from Hewett and Wild, (2010) for the SDSS DR7 quasars and the BOSS PCA-based redshifts from Pâris et al., (2017) for SDSS DR12, the correction formula given in Section 2.4.3 produces significant improvements to C iv-based BH mass estimates. In a forthcoming work, Allen & Hewett (in preparation) will present a new redshift-estimation algorithm that produces redshifts independent of the C iv blueshift and other variations in the ultraviolet SEDs of luminous quasars. Allen & Hewett will publish improved redshifts for all quasars in the SDSS DR7 and DR12 which will reduce SED-dependent systematic errors below the apparent inherent dispersion of $\simeq 220 \text{ km s}^{-1}$ associated with broad emission line redshifts (Shen et al., 2016). At the same time we will publish catalogues of unbiased BH masses for both SDSS DR7 and DR12 based on the Allen & Hewett redshifts. The components from the mean-field independent component analysis (see Allen et al., 2013, for an application to astronomical spectra) used in the Allen & Hewett redshift algorithm will also be published. With these components, if a rest-frame ultraviolet spectrum is available, it will be straightforward to determine the systemic redshift, via a simple optimisation procedure, and hence calculate the C iv blueshift.

We have presented an analysis of biases in C iv-derived virial BH masses of high-luminosity ($L_{\text{bol}} \sim 47 \text{ erg s}^{-1}$) quasars at redshifts ~ 2.5 from the Sloan Digital Sky Survey. Many authors have reported a large scatter between C iv- and H α /H β -based masses, and part of this scatter has been shown to correlate with the C iv blueshift (Shen and Liu, 2012). The blueshifting of C iv is usually interpreted as evidence for strong outflows which, most likely, result from the presence of a radiation line-driven accretion-disc wind. Our study is the first to examine this bias systematically across the full range of C iv-emission line blueshifts observed in the SDSS sample. In particular, we have used rest-frame optical spectra of 19 quasars in the redshift range

$2 < z < 2.7$ to directly compare C IV and H α emission properties as a function of the C IV blueshift. We reach the following conclusions:

- A strong correlation between C IV-velocity width and blueshift is found and at large blueshifts, $>2000 \text{ km s}^{-1}$, the velocity-widths are dominated by non-virial motions. This suggests that the assumption that velocity-widths arise from virial-induced motions, on which single-epoch BH-mass measurements are predicated, is not straightforwardly applicable to these high-blueshift quasars.
- We use the H α emission line to provide BH-mass estimates that are unbiased by non-virial contributions to the velocity width. We find that the C IV-based BH masses of quasars with low C IV blueshifts are systematically underestimated (by a factor of ~ 1.7) whereas the masses of quasars with large blueshifts are severely overestimated (by a factor of ~ 5).
- We find a systematic change in the shape of the H α line profile as a function of the C IV blueshift. Specifically, the H α line profiles of the quasars with high C IV blueshifts are all ‘peaky’ with FWHM/σ close to unity.
- We suggest that the high C IV blueshift quasars are high Eddington-ratio objects that are inherently rare (comprising $\sim 15\%$ of the UV-selected sample), but are being boosted in number by a factor of ~ 4 in the flux-limited SDSS sample.

With a relatively small sample of 19 quasars we have been able to uncover systematic trends in the C IV and H α emission line shapes as a function of the C IV blueshift. This confirms the prospect of developing a blueshift-dependent correction to C IV-based single-epoch BH-mass estimates using a larger samples of luminous quasars with both rest-frame UV and rest-frame optical spectroscopy. We are currently in the process of assembling such a sample, which will contain ~ 300 luminous quasars, 80 per cent at redshifts $z \geq 2$. A new SED-independent redshift-estimation algorithm (Allen & Hewett 2017, in preparation) makes it possible to quantify the distribution of C IV-blueshifts in the observed quasar population as a whole, thereby allowing us to make empirical corrections to C IV-based BH-masses for all luminous, high-redshift SDSS/BOSS quasars.

2.8 C IV BLUESHIFTS IN THE QUASAR POPULATION

Recognition that the C IV emission line in quasars can exhibit significant asymmetric structure, with an excess of flux to the blue of the predicted rest-frame transition wavelength, extends back to Gaskell, (1982). Significant progress in understanding the relationship between

changes in C iv-emission shape and quasar properties has come about through studies in which near-infrared spectra of the hydrogen Balmer lines have been obtained. Such studies typically involve samples of modest size and the location of the Balmer lines provides a reliable estimate of the quasar systemic redshifts; recent examples include Shen and Liu, (2012) and Marziani et al., (2016). In Section ?? we adopt the same approach to estimate systemic redshifts for the quasar sample presented here with near-infrared spectra. However, improvements in the estimation of systemic redshifts from ultraviolet quasar spectra means that it is now possible to quantify the distribution of C iv-blueshifts in the observed quasar population as a whole.

2.8.1 Quasar systemic redshifts

Historically, the parametrisation of the C iv emission-line properties for quasars in large surveys has not proved straightforward because the C iv emission line has itself been used in the determination of the quasar redshifts. The SDSS provided the first catalogue of tens of thousands of redshift $z > 1.6$ quasars with spectra of adequate velocity resolution and S/N that effective statistical studies of the rest-frame ultraviolet emission-line properties, including line-shape, have proved possible.

The comprehensive compilation of quasar properties for the SDSS DR7 quasars by Shen et al., (2011) provides a natural starting point for population studies. In Fig. 2.23a we plot the C iv-blueshift versus C iv-emission equivalent width (EW) using the SDSS pipeline redshifts and the blueshifts calculated by Shen et al., (2011). A strong trend in the blueshift values as a function of line EW is not evident in Fig. 2.23a; structure in the parameter space is being masked because the C iv emission line is itself being used in the determination of the quasar redshifts.

The redshift-determination scheme of Hewett and Wild, (2010) provided much improved redshifts, not least because the redshift estimates for the majority of quasars were derived using emission-lines other than the C iv-line itself. Figure 2.23b shows SDSS DR7 quasars in the same C iv parameter space as Figure 2.23a, but now using Hewett and Wild, (2010) redshifts. The improved redshift estimates are predominantly responsible for the differences seen in Fig. 2.23a and b; the appearance in Fig. 2.23b of the extension to high blueshift for quasars with low C iv EW is particularly evident.

The large systematic variation in the C iv emission-line profile within the population is evident from figures 11 and 12 of Richards et al., (2011). The plots and analysis in Richards et al., (2011) employ the quasar redshifts from Hewett and Wild, (2010) but, as is evident from the figures, the systematic variation in the C iv shape is correlated with changes in the quasar SEDs, including the strengths of

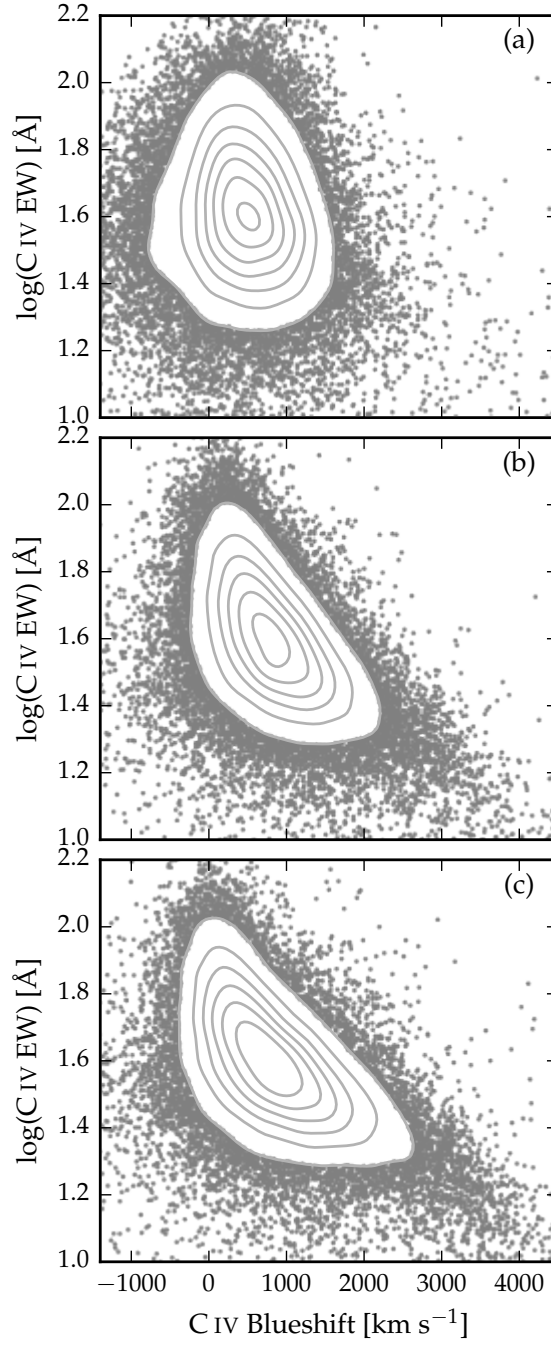


Figure 2.23: Rest-frame EW versus blueshift of the broad C iv-emission line for 32,157 SDSS DR7 quasars at $1.6 < z < 3.0$. Panel (a) uses C iv line parameters from Shen et al., (2011) and SDSS pipeline systemic redshifts. Panels (b) and (c) use systemic redshifts from Hewett and Wild, (2010) and Allen & Hewett (2017, in preparation) respectively, and C iv line measurements described in Sec. 2.8.2. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator.

the Si III λ 1892 and C III λ 1908 emission lines in the rest-frame ultraviolet. As a consequence, the redshifts from Hewett and Wild, (2010) still suffer from systematic errors that are correlated with the shape, and particularly the blueshift, of the C IV emission line. The nature of the systematic variations in the quasar ultraviolet SEDs are such that for quasars with close-to symmetric C IV profiles and line centroids close to the systemic redshift, the Hewett and Wild, (2010) redshifts result in C IV blueshifts that are overestimated by a few hundred km s⁻¹, whereas, for quasars with strong blue-asymmetric C IV profiles and line centroids displaced significantly to the blue of the systemic redshift, the C IV blueshifts are underestimated by, in the most extreme cases, up to 1200 km s⁻¹.

Figure 2.23c shows the C IV emission line parameters calculated using a new redshift-estimation algorithm (Allen & Hewett 2017, in preparation) that takes account of the quasar SED variations, producing redshifts independent of the large systematic shape changes seen in the C IV emission line. The low-ionization emission lines visible in the rest-frame ultraviolet (over wavelengths from Mg II λ 2796,2803 down to the O I λ 1304+Si II λ 1307 blend) using the new redshift-algorithm are located at rest-frame wavelengths in excellent agreement with the systemic redshift defined using the rest-frame narrow-line optical O III doublet and broad-line H β and H α .

The systematic trends seen in Fig. 2.23b, in particular the extension to high blueshift at low C IV EW, become more apparent in Fig. 2.23c, as expected from consideration of the known SED-related errors in the redshifts from Hewett and Wild, (2010). A population of quasars with only modest blueshifts and low EW is also apparently still present.

2.8.2 C IV emission line blueshift measurements

The differences in the distribution of C IV emission line properties seen in the three panels of Fig. 2.23 are due primarily to the change in the systemic redshift estimates. It is also necessary, however, to obtain a measure of the C IV emission line ‘location’ in order to calculate the blueshifts. When working with moderately-sized samples, parametric fits to the emission-line profile may be undertaken using careful mask-definition to minimise the effect of absorption features on the profiles used for the parametrization, and this is the approach we follow below in Section ??.

Effective analysis of the tens of thousands of spectra from SDSS DR7, and now DR12, however, requires a more robust scheme to determine a C IV-blueshift estimate that is not very sensitive to the range of S/N among the spectra or the presence of narrow absorption systems within the C IV-emission profile. Shen et al., (2011) provide a discussion (their section 3) of the factors that effect the measurement of broad emission lines in quasar

spectra of modest S/N. Their careful analysis of the C IV emission properties employed the results of parametric fits of three Gaussians to the spectra. Our own experiments in quantifying the C IV emission properties of SDSS spectra showed that a simple non-parametric measure of the C IV emission location reduced the number of outliers significantly. Visual inspection of spectra demonstrated that the improvement is due primarily to the identification of, and interpolation over, associated and outflow absorption systems, which forms part of the non-parametric measurement scheme.

We therefore chose to use a non-parametric scheme to measure the blueshift of the C IV line, which we will now describe. A continuum is first defined as a power-law of wavelength, $f(\lambda) \propto \lambda^{-\alpha}$, with the slope, α , determined using the median⁸ values of the flux in two continuum windows at 1445–1465 and 1700–1705 Å (the same wavelengths as adopted by Shen et al., (2011)). The C IV emission line is taken to lie within the wavelength interval 1500–1600 Å, a recipe that is commonly adopted (e.g. Shen et al., 2011; Denney et al., 2013). To reduce the impact of narrow absorption systems on the emission-line profile a ‘pseudo continuum’ is defined by applying a 41-pixel median filter to the quasar spectrum. Pixels within the C IV profile that lie more than 2σ below the pseudo-continuum are deemed to be affected by absorption and added to an ‘absorber’-mask. Two pixels on either side of each such pixel are also included in the mask. For each masked pixel, the flux values in the spectrum are replaced by values from the pseudo-continuum.

The wavelength that bisects the cumulative total line flux, λ_{half} , is recorded and the blueshift (in km s^{-1}) defined as $c \times (1549.48 - \lambda_{\text{half}}) / 1549.48$ where c is the velocity of light and 1549.48 Å is the rest-frame wavelength for the C IV doublet⁹. Positive blueshift values indicate an excess of emitting material moving towards the observer and hence outflowing from the quasar. Hewett and Wild, (2010) redshifts are used to define the quasar rest-frame.

2.8.3 Sample selection - C IV properties

The primary aim of the paper is to investigate the potential systematic effects on the C IV-emission based BH masses for quasars with large, $\gtrsim 1200 \text{ km s}^{-1}$, C IV blueshifts, using the properties of the H α emission line to provide BH-mass estimates for the objects unbiased by non-virial contributions to the emission-line profile. The orange symbols in Fig. 2.23 show the C IV parameters of our quasar targets for which near-infrared spectra of adequate S/N were obtained. These

⁸ The median is used to improve the robustness of the continuum estimate from the relatively small wavelength intervals.

⁹ The adopted C IV rest-frame wavelength assumes an optically thick BLR, in which case the contribution from each component is equal. Adopting a 2:1 ratio (appropriate for an optically thin BLR) changes the blueshifts by $\sim 80 \text{ km s}^{-1}$.

quasars were selected using our non-parametric blueshift measures (based on the Hewett and Wild, (2010) redshifts). The sample of 19 quasars spans the full dynamic range in C IV-parameters based on the Hewett and Wild, (2010) systemic redshifts and the coverage is in fact even more complete when using the forthcoming SED-independent redshifts from Allen & Hewett (2017, in preparation). As is evident from the sparsity of quasars with large C IV blueshifts when the SDSS pipeline systemic redshifts are used (Fig. 2.23a), improvements in the estimation of systemic redshifts from ultraviolet spectra have been a crucial factor in allowing us to reliably select a sample of quasars with a range of C IV blueshifts. In subsequent sections we re-derive the systemic redshifts and C IV blueshifts for this sample using parametric fits to the H α and C IV emission (the former from our near-infrared observations). Thus, while the systematic trends in BH masses inferred from measurements of the C IV emission line depend on the distribution of C IV emission line properties within the quasar population, the results of our analysis of the H α and C IV emission line properties are independent of the redshifts used to produce the panels in Fig. 2.23.

